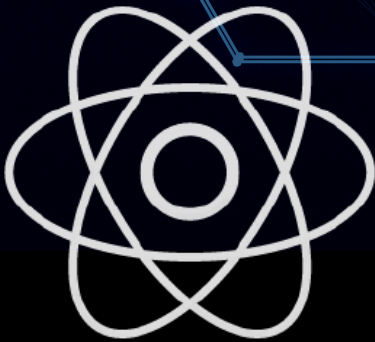


JOURNAL OF ENGINEERING RESEARCH & SCIENCES

JENRS



www.jenrs.com
ISSN: 2831-4085

Volume 1 Issue 3
March 2022



EDITORIAL BOARD

Editor-in-Chief

Prof. Paul Andrew
Universidade De São Paulo, Brazil

Editorial Board Members

Dr. Jianhang Shi

Department of Chemical and Biomolecular Engineering, The Ohio State University, USA

Dr. Sonal Agrawal

Rush Alzheimer's Disease Center, Rush University Medical Center, USA

Dr. Namita Lokare

Department of Research and Development, Valencell Inc., USA

Dr. Dongliang Liu

Department of Surgery, Baylor College of Medicine, USA

Dr. Xuejun Qian

Great Lakes Bioenergy Research Center & Plant Biology Department, Michigan State University, USA

Dr. Jianhui Li

Molecular Biophysics and Biochemistry, Yale University, USA

Dr. Atm Golam Bari

Department of Computer Science & Engineering, University of South Florida, USA

Dr. Lixin Wang

Department of Computer Science, Columbus State University, USA

Dr. Prabhash Dadhich

Biomedical Research, CellfBio, USA

Dr. Żywiołek Justyna

Faculty of Management, Czestochowa University of Technology, Poland

Prof. Kamran Iqbal

Department of Systems Engineering, University of Arkansas Little Rock, USA

Dr. Ramcharan Singh Angom

Biochemistry and Molecular Biology, Mayo Clinic, USA

Dr. Qichun Zhang

Department of Computer Science, University of Bradford, UK

Dr. Mingsen Pan

University of Texas at Arlington, USA

Editorial

In the ever-evolving landscape of research and innovation, diverse disciplines converge to address pressing challenges and unearth novel solutions. In this issue 19 papers were accepted. This editorial presents a glimpse into a myriad of recent research endeavours spanning fields as varied as oil extraction optimization, machine learning in agriculture, vocational education reform, and robotic disinfection. From the depths of theoretical physics to the intricacies of practical engineering, each study contributes to advancing knowledge, solving real-world problems, and shaping the future of their respective domains. Through collaboration, experimentation, and relentless inquiry, researchers around the globe strive to push the boundaries of possibility and pave the way for a brighter tomorrow.

Oil extraction, a cornerstone of many economies, demands continual optimization to ensure sustainability and cost-effectiveness. In this vein, recent research delves into the realm of oil field development, particularly focusing on sucker rod pump units (SRPU), a prevalent mechanized method in the industry. Despite their ubiquity, inefficient management has led to escalating operating costs. However, promising advancements emerge from the integration of NOISE technology for early equipment diagnosis and frequency converters for motor control. Developed at the Institute of Management Systems of the National Academy of Sciences of Azerbaijan, the "NOISE Control, Diagnostics and Management Complex" showcases notable improvements in efficiency, presenting a method for calculating energy savings. Such innovations hold significant promise for the oil and gas sector, offering tangible pathways towards enhanced operational efficiency and cost savings [1].

The agricultural landscape is inherently linked to soil health, with vital properties dictating crop productivity and ecosystem stability. Addressing this nexus, recent research endeavours to predict soil properties using machine learning approaches. Focusing on critical factors like calcium, phosphorus, and pH, the study employs multiple regression models to forecast soil attributes. Notably, gradient boosting emerges as a standout performer, showcasing superior accuracy in property prediction. These findings hold immense implications for agricultural practices, empowering farmers with predictive insights to optimize soil management strategies. By harnessing the power of machine learning, stakeholders can enhance agricultural productivity, mitigate environmental degradation, and foster sustainable land stewardship [2].

Vocational and technical education institutions play a pivotal role in nurturing skilled workforces and driving economic development. However, despite their importance, these institutions face multifaceted challenges ranging from management issues to curriculum deficiencies. Recent research delves into the depths of these challenges, employing qualitative methodologies to identify solution proposals. From establishing effective control mechanisms to strengthening education-employment-production linkages, the study offers actionable insights to enhance vocational education quality and relevance. By addressing management, program, and implementation concerns, stakeholders can unlock the full potential of vocational education, fostering inclusive growth and socioeconomic empowerment [3].

In the realm of image processing, the quest for optimal enhancement techniques is perpetual. Recent research contributes to this pursuit by conducting a comparative analysis of various image enhancement methods, particularly focusing on local domain techniques. Employing three-dimensional image quality statistics such as Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index (SSIM), the study evaluates the efficacy of different enhancement approaches. By elucidating the conditions for choosing the most

suitable method based on quantitative metrics, the research provides valuable insights for image processing practitioners and researchers alike [4].

The relationship between healthcare expenditure and economic growth is a topic of significant interest and debate. Recent research delves into this complex interplay, particularly focusing on Southern and Western Sub-Saharan Africa. Utilizing advanced panel data analyses, including mean group and dynamic fixed effects models, the study scrutinizes the short and long-term effects of total health spending on economic growth. Notably, findings reveal a positive correlation between health spending, life expectancy, and economic growth, underscoring the importance of investment in healthcare infrastructure as a driver of economic prosperity in the region [5].

Diffraction analysis plays a pivotal role in various scientific and engineering domains, from optics to acoustics. Recent research introduces a novel method of fundamental components to tackle the problem of diffraction on non-ideal boundaries. Distinguishing itself from traditional heuristic approaches, the proposed method incorporates an adjustment procedure to enhance accuracy while maintaining formula compactness. By refining heuristic formulas for diffraction on impedance half-planes, the study offers insights into advancing high-speed solvers and elucidating numerical solution interpretations [6].

Human pose estimation is integral to numerous applications, from surveillance to augmented reality. Recent research conducts a comparative study of 2D human pose estimation using various versions of Residual Networks (ResNet) on the Human 3.6M Dataset. Through fine-tuning and quantitative evaluation, the study elucidates the performance of different ResNet architectures in estimating human poses. By analyzing error rates and training epochs, the research provides valuable insights for optimizing human pose estimation algorithms in real-world applications [7].

Effective traffic management relies on accurate detection and categorization of traffic panels. Recent research proposes a novel categorization method for traffic panels, classifying them into symbol-based, text-based, and supplementary/additional categories. Through detailed analysis of text-based panels, the study highlights challenges such as language diversity and dataset scarcity. By gathering a novel dataset for Persian text-based traffic panels and utilizing advanced algorithms like tiny YOLOv3, the research advances the state-of-the-art in traffic panel detection, paving the way for improved transportation infrastructure [8].

Fault diagnosis in complex systems poses significant challenges, often requiring adaptive and reliable models. Recent research introduces an ensemble deep auto-encoder method for intelligent fault diagnosis. By leveraging tuning algorithms and hypersphere classification, the proposed model exhibits flexibility and reliability in adapting to changing operating conditions. Moreover, by integrating power spectrum analysis and Layer-wise Relevance Propagation, the model enhances diagnostic reliability and interpretability. Through comprehensive validation and comparison, the research demonstrates the efficacy of the ensemble approach in fault diagnosis applications [9].

High Altitude Platform (HAP) networks hold promise for extending wireless connectivity, but challenges such as handover failure and resource congestion must be addressed. Recent research proposes a novel call admission control scheme using deadline, channel, and tolerance aware scheduling (DCTS) algorithm. By ensuring tight latency guarantees and mitigating packet loss due to deadline violations, the DCTS system enhances quality of service in HAP wireless networks. Through extensive testing and comparison with existing algorithms, the research showcases the effectiveness of the proposed approach in managing network resources and optimizing real-time communication [10].

Buckling analysis is critical for assessing structural stability in engineering applications. Recent research presents an exact trigonometric solution for the buckling problem of a three-dimensional rectangular plate subjected to compressive load. By formulating total potential energy and deriving rotation-deflection equations, the study establishes a critical buckling load formula. Through theoretical analysis and comparison with previous studies, the research offers valuable insights into buckling behaviour and the applicability of different plate theories [11].

Additive manufacturing holds immense potential for transforming manufacturing processes across industries. Recent research investigates existing additive manufacturing techniques and modeling approaches, particularly focusing on thermal finite element analysis. By conducting numerical simulations and evaluating thermal structures, the study sheds light on solid and powder material properties and boundary conditions. Through comprehensive analysis, the research provides valuable insights into optimizing additive manufacturing processes and enhancing component quality [12].

Micro-forming represents a promising avenue for manufacturing small-scale components with precision. Recent research delves into micro-forming techniques, particularly focusing on deep drawing processes and their industrial applications. By exploring methods such as the Limit dome height test and Nakajima test, the study highlights challenges and advancements in micro-forming technology. Through systematic investigation, the research aims to contribute to the development of automated micro-forming machines capable of high-volume production [13].

POLYMATH software offers versatile tools for solving mathematical problems in chemical engineering. Recent research demonstrates the utility of POLYMATH through a comprehensive problem set spanning transport phenomena, heat transfer, and reaction kinetics. While predominantly used for educational purposes, the software's capabilities extend to process industries, providing solutions for differential equations, nonlinear equations, and regression problems. By incorporating novel solution approaches and enabling log-log plots, the research underscores the versatility and applicability of POLYMATH in chemical engineering analysis [14].

The center of gravity plays a crucial role in the stability and performance of various systems, from graphic computer images to sports equipment and transportation vehicles. Recent research investigates the experimental determination of the center of gravity to optimize stability and performance. By leveraging experimental methods and computational techniques, the study aims to enhance stability in diverse applications, including graphic images, sports equipment, and transportation vehicles. Through precise center of gravity adjustments, stakeholders can improve performance, safety, and efficiency in various domains [15].

Disinfecting mobile robots represent a novel solution for maintaining cleanliness and hygiene in diverse environments. Recent research focuses on designing an omnidirectional mobile robot equipped with UV light tubes for disinfection purposes. By integrating controllers, cameras, and actuators, the proposed robot navigates autonomously, avoiding obstacles while disinfecting surfaces with UV light. Through feedback mechanisms and encoder inputs, the robot achieves precise motion control, ensuring thorough disinfection in targeted areas. The research contributes to the development of advanced robotic systems for enhancing cleanliness and safety in various settings [16].

Key points detection and matching are essential tasks in computer vision applications, including object recognition and 3D reconstruction. Recent research introduces a robust key points descriptor, termed UFAHB, utilizing a cascade of Upright FAST-Harris Filter and Binary

Robust Independent Elementary Feature descriptor. Through comprehensive performance evaluation, the study demonstrates the robustness and efficiency of the UFAHB descriptor across diverse imaging conditions. By addressing challenges such as illumination variation and geometric transformations, the research advances the state-of-the-art in local key points matching, facilitating applications in image processing and computer vision [17].

Anaerobic digestion holds promise for converting diverse feedstock into biogas, offering renewable energy and waste management solutions. Recent research provides insights into utilizing various feedstock for anaerobic digestion, considering factors such as sustainability, availability, and nutrient content. By evaluating conventional and non-conventional digester types, the study aims to optimize biogas production while addressing environmental and economic considerations. Through systematic analysis, the research contributes to advancing biogas technology and promoting sustainable energy practices [18].

Riverine fisheries play a crucial role in food security and ecosystem health, necessitating effective management strategies. Recent research investigates the length-weight relationships (LWRs) and condition factors of dominant fish species in Nyangweta tributary of River Kuja, Kenya. Through comprehensive analysis of fish populations and environmental variables, the study elucidates growth patterns and physiological conditions in key species. By establishing strong linear relationships and assessing condition factors, the research offers valuable insights for fisheries management and conservation efforts in riverine ecosystems [19].

In a world marked by constant change and complexity, research remains a beacon of hope and progress. The diverse array of studies showcased in this editorial underscore the breadth and depth of human ingenuity, from unravelling the mysteries of the universe to enhancing the efficiency of everyday processes. Whether it's harnessing machine learning for sustainable agriculture, designing autonomous robots for disinfection, or refining mathematical models for chemical engineering analysis, each endeavour contributes to the collective pursuit of knowledge and innovation. As we navigate the challenges and opportunities of the 21st century, let us continue to celebrate and support the endeavours of researchers worldwide, for their discoveries hold the key to a brighter, more resilient future for all [1].

References:

- [1] A.T. Abbas, G.G. Agaverdi, R.A. Haji, A.Y. Gabil, R.M. Huseyn, Y.A. Nikolaevich, K.M. Ilgizovich, "Ways to Increase the Efficiency of Sucker Rod Pump Units in Oil Production," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 1–8, 2022, doi:10.55708/js0103001.
- [2] V. Kumar, J.S. Malhotra, S. Sharma, P. Bhardwaj, "Soil Properties Prediction for Agriculture using Machine Learning Techniques," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 9–18, 2022, doi:10.55708/js0103002.
- [3] C. Akıllı, İ.B. Arabacı, E. Kırçıl, "An Overview of Solutions Regarding the Problems in Vocational and Technical Education - Example of Elazığ Province," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 19–27, 2022, doi:10.55708/js0103003.
- [4] A.V. Deshpande, "An Overview on Various Techniques used for Correct Interpretation of Roadway Symbols," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 28–38, 2022, doi:10.55708/js0103004.
- [5] F.O. Yusufu, B.O. Awoyemi, K. Akomolafe, "The Effects of Total Health Expenditure on Economic Growth in Southern and Western sub-Saharan Africa," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 39–51, 2022, doi:10.55708/js0103005.

- [6] M. Vesnik, "Physical Interpretation of the Solution to the Problem of Diffraction on a Half-plane with Non-Ideal Boundary Conditions," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 52–58, 2022, doi:10.55708/js0103006.
- [7] H.-Y. -Tran, T.-M. Bui, T.-L. Pham, V.-H. Le, "An Evaluation of 2D Human Pose Estimation based on ResNet Backbone," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 59–67, 2022, doi:10.55708/js0103007.
- [8] S. Kheirinejad, N. Riahi, R. Azmi, "Text-Based Traffic Panels Detection using the Tiny YOLOv3 Algorithm," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 68–80, 2022, doi:10.55708/js0103008.
- [9] K. Zhang, Y. Wang, H. Qu, "Bearing Fault Diagnosis Based on Ensemble Depth Explainable Encoder Classification Model with Arithmetic Optimized Tuning," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 81–97, 2022, doi:10.55708/js0103009.
- [10] H.M. Hussien, S.F. Meko, K. Katzis, L.P. Mfupe, E.T. Bekele, "Call Admission Control for Real-Time Applications of TVWS Wireless Access from HAP," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 98–105, 2022, doi:10.55708/js0103010.
- [11] O.F. Chukwudi, O.T. Edozie, N.-D. Chidobere, "Buckling Analysis of a Three-Dimensional Rectangular Plates Material Based on Exact Trigonometric Plate Theory," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 106–115, 2022, doi:10.55708/js0103011.
- [12] V. Kohale, S. Jawade, G. Kakandikar, "A Review- Modelling Approach and Numerical Analysis of Additive Manufacturing," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 116–125, 2022, doi:10.55708/js0103012.
- [13] N. Tiwari, G. Kakandikar, O. Kulkarni, "Micro Forming and its Application: A Critical Review," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 126–132, 2022, doi:10.55708/js0103013.
- [14] A.M. Abubakar, B. Iliyasu, Z.M. Sarkinbaka, "Detailed Overview on POLYMATH Software for Chemical Engineering Analysis," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 133–147, 2022, doi:10.55708/js0103014.
- [15] J. Je, E. Jekal, "Experimental Methodology to Find the Center of Gravity of a Solid," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 148–152, 2022, doi:10.55708/js0103015.
- [16] W. Qaisar, M.T. Riaz, A. Basit, Y. Naseem, Z. Nazir, "Disinfecting Omnidirectional Mobile Robot with Vision Capabilities," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 153–163, 2022, doi:10.55708/js0103016.
- [17] A.D. Mohammed, O.O. Adeniyi, S.A. Muhammed, M.A. Saddiq, E. Ayobami, "Cascaded Keypoint Detection and Description for Object Recognition," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 164–169, 2022, doi:10.55708/js0103017.
- [18] A.M. Abubakar, "Biodigester and Feedstock Type: Characteristic, Selection, and Global Biogas Production," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 170–187, 2022, doi:10.55708/js0103018.
- [19] F.M. Ondemo, A. Getabu, Z. Gichana, J.O. Omweno, "Length-Weight Relationships (LWRs) and Condition Factor of Seven Fish Species in River Nyangweta Tributary, Kenya," *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 193–199, 2022, doi:10.55708/js0103020.

Editor-in-chief

Prof. Paul Andrew

CONTENTS

<i>Ways to Increase the Efficiency of Sucker Rod Pump Units in Oil Production</i> Aliiev Telman Abbas, Guluyev Gambar Agaverdi, Rzayev Asif Haji, Aliyev Yaver Gabil, Rezvan Mahammad Huseyn, Yashin Anton Nikolaevich, Khakimyanov Marat Ilgizovich	01
<i>Soil Properties Prediction for Agriculture using Machine Learning Techniques</i> Vijay Kumar, Jai Singh Malhotra, Saurav Sharma, Parth Bhardwaj	09
<i>An Overview of Solutions Regarding the Problems in Vocational and Technical Education – Example of Elazığ Province</i> Ceyda Akıllı, İmam Bakır Arabacı, Engin Kırçıl	19
<i>An Overview on Various Techniques used for Correct Interpretation of Roadway Symbols</i> Abhinav Vinod Deshpande	28
<i>The Effects of Total Health Expenditure on Economic Growth in Southern and Western sub-Saharan Africa</i> Faith Onechojo Yusufu, Bosede Olanike Awoyemi, Kehinde Akomolafe	39
<i>Physical Interpretation of the Solution to the Problem of Diffraction on a Half-plane with Non-Ideal Boundary Conditions</i> Michael Vesnik	52
<i>An Evaluation of 2D Human Pose Estimation based on ResNet Backbone</i> Hai-Yen -Tran, Trung-Minh Bui, Thi-Loan Pham, Van-Hung Le	59
<i>Text-Based Traffic Panels Detection using the Tiny YOLOv3 Algorithm</i> Saba Kheirinejad, Noushin Riahi, Reza Azmi	68
<i>Bearing Fault Diagnosis Based on Ensemble Depth Explainable Encoder Classification Model with Arithmetic Optimized Tuning</i> Kaibi Zhang, Yanyan Wang, Hongchun Qu	81
<i>Call Admission Control for Real-Time Applications of TVWS Wireless Access from HAP</i> Habib M. Hussien, Sultan F. Meko, Konstantinos Katzis, Luzango P. Mfupe, Ephrem T. Bekele	98
<i>Buckling Analysis of a Three-Dimensional Rectangular Plates Material Based on Exact Trigonometric Plate Theory</i> Onyeka Festus Chukwudi, Okeke Thompson Edozie, Nwa-David Chidobere	106
<i>A Review- Modelling Approach and Numerical Analysis of Additive Manufacturing</i> Vaishnavi Kohale* , Samidha Jawade, Ganesh Kakandikar	116

<i>Micro Forming and its Application: A Critical Review</i> Neha Tiwari, Ganesh Kakandikar, Omkar Kulkarni	126
<i>Detailed Overview on POLYMATH Software for Chemical Engineering Analysis</i> Abdulhalim Musa Abubakar, Bello Iliyasu, Zakiyyu Muhammad Sarkinbaka	133
<i>Experimental Methodology to Find the Center of Gravity of a Solid</i> Joohoon Je, Eunsung Jekal	148
<i>Disinfecting Omnidirectional Mobile Robot with Vision Capabilities</i> Waqas Qaisar, Muhammad Tanveer Riaz , Abdul Basit, Yasir Naseem, Zohaib Nazir	153
<i>Cascaded Keypoint Detection and Description for Object Recognition</i> Abdulmalik Danlami Mohammed, Ojerinde Oluwaseun Adeniyi, Saliu Adam Muhammed, Mohammed Abubakar Saddiq, Ekundayo Ayobami	164
<i>Biodigester and Feedstock Type: Characteristic, Selection, and Global Biogas Production</i> Abdulhalim Musa Abubakar	170
<i>Physiological and biochemical responses of the pistachio (Pistacia vera L.) cultivar "Mateur" as affected by regulated deficit irrigation</i> Dhiaa Miri, Walid Abidi, Rawaa Akrimi	Withdrawn
<i>Length-Weight Relationships (LWRs) and Condition Factor of Seven Fish Species in River Nyangweta Tributary, Kenya</i> Fredrick Mang'era Ondemo, Albert Getabu, Zipporah Gichana, Job Ombiro Omweno	193

Ways to Increase the Efficiency of Sucker Rod Pump Units in Oil Production

Aliev Telman Abbas¹, Guluyev Gambar Agaverdi¹, Rzayev Asif Haji¹, Aliyev Yaver Gabil¹, Rezvan Mahammad Huseyn¹, Yashin Anton Nikolaevich², Khakimyanov Marat Ilgizovich^{*2}

¹Laboratory "Control, diagnostics and management system for oil extracting facilities", Institute of Control Systems of ANAS, Baku, 1141, Azerbaijan

²Department of Electrical Engineering and Electrical Equipment, Ufa State Petroleum Technological University, Ufa, 450000, Russia

*Corresponding author: Khakimyanov Marat Ilgizovich, Pervomayskaya st., 14, Ufa, Rep. Bashkortostan, 450000, Russia, Contact No: 89033125413 & Email: hakimyanovmi@gmail.com

ABSTRACT: The article shows that in the middle and late stages of oil field development, of all mechanized methods, the most common is oil extraction by sucker rod pump units (SRPU). However, due to inefficient management of them, their operating costs are increasing. The article presents the results of improving the efficiency of SRPU due to the use of NOISE technology for early diagnosis of the technical condition of equipment and frequency converters for controlling its induction motor. The results of the implementation of the "NOISE Control, Diagnostics and Management Complex for oil wells operated by SRPU" created at the Institute of Management Systems of the National Academy of Sciences of Azerbaijan are presented. A method for calculating the saving of electrical energy in the control of SRPU frequency converters has been developed. The results of calculation of electric energy savings for the oil and gas production facility Bibi-Heybat Oil of the Republic of Azerbaijan are also presented.

KEYWORDS: Oil production, Sucker rod pump units, Electric Drives, Induction motor, Efficiency, NOISE Technology, Robust Technology, Frequency Converters

1. Introduction

Currently, most oil fields are in the middle and late stages of development, where the reservoir pressure is not sufficient to lift the liquid to the surface of the earth. At the same time, the reservoir pressure is maintained by pumping water into the reservoir, which contributes to an increase in the water content in the total liquid. According to published data, the water content in the liquid extracted from an oil well in the USA is about 90% [1], and in the Russian Federation it is about 86% [2]. At the same time, they are forced to use mechanized methods of oil production. Currently, the main methods of mechanized mining are the use of sucker rod pump units (SRPU) and installations of electric submersible pumps (ESP). The published data indicate that in the Russian Federation 41% of wells are operated by SRPU and 54% by ESP [3], and in the USA more than 85% of wells with mechanized production are equipped with SRPU [4]. They are the main pumping system that provides mechanical energy for oil

production [5]. Advantages of SPRU are simplicity, reliability and the possibility of application in a wide range of operating conditions [6].

However, with the use of mechanized mining, due to inefficient management of them, the operating costs of production increase dramatically [7] and, consequently, the production benefits become lower [8]. Thus, at the same time, one of the important issues is to increase the efficiency of SPRU. Research conducted at the Institute of Control Systems of ANAS of Azerbaijan for many years, as well as experience in oil and gas production fields, have shown that the main sources of operating costs are:

- the costs of repairing ground and underground equipment of wells operated with SPRU due to untimely diagnostics and inadequate identification of technical conditions that lead to unnecessary downtime and material costs;
- electricity supply costs.

2. Problem Statement

Thus, one of the possible ways to increase the efficiency of SPRU by reducing the cost of repairing ground and underground equipment of wells using NOISE technology for diagnostics and adequate identification of technical conditions and power supply costs using frequency converters to control an induction motor is given. The task is also set to develop a methodology for calculating the savings of electrical energy when using a frequency-controlled drive of the SPRU.

3. Solving the problem. NOISE technology for early diagnosis and adequate identification of the technical condition of rod deep-pumping units

In the paper robust technology and control system of oil wells SPRU [9], a sequence of calculations is proposed, the totality of which constitutes the procedure for normalizing correlation functions:

1) According to the expression

$$R_{gg}(\mu) = \frac{1}{N} \sum_{i=1}^N g(i\Delta t)g((i+\mu)\Delta t);$$

estimates of the correlation function of the noisy signal are calculated for $\mu = 1, 2, 3, 4, \dots$

2) According to the expression

$$D_g = \frac{1}{N} \sum_{i=1}^N g(i\Delta t)g(i\Delta t);$$

the estimate of the variance of the noisy signal is calculated.

3) According to the expression

$$D_\varepsilon = \frac{1}{N} \sum_{i=1}^N [g^2(i\Delta t) + g(i\Delta t)g((i+2)\Delta t) - 2g(i\Delta t)g((i+1)\Delta t)];$$

an estimate of the interference variance D_ε is calculated.

4) According to the expression

$$R_{UU}(\mu=0) = D_g - D_\varepsilon;$$

the variance value $R_{UU}(\mu=0)$ of the useful signal $U_p(i\Delta t)$ is calculated.

5) According to the expression

$$r_{gg}^R(\mu) = \begin{cases} \frac{R_{gg}(\mu=0)}{D_g} = 1 & \text{by } \mu = 0 \\ \frac{R_{gg}(\mu)}{D_g - D_\varepsilon} = 1 & \text{by } \mu \neq 0 \end{cases};$$

estimates of the normalized correlation function $r_{gg}^R(\mu)$ of the noisy signal $g(i\Delta t)$ are calculated at $\mu = 0, 1, 2, 3, \dots$

The proposed procedure makes it possible to calculate robust estimates (free from errors caused by interference in the signal composition) of the normalized autocorrelation function $r_{gg}^R(\mu)$ of the noisy signal $g(i\Delta t)$.

The main idea of robust analysis is to develop estimates that are resistant to outliers and at the same time show a result close enough to classical methods on data that do not contain atypical observations.

Applying what has been said to the measured output signal of the force sensor of the column of rods on the suspension of the SPRU, you can make the following procedure for calculating informative signs to identify the technical condition of the SPRU:

1. The swing period T_{ST} SPRU is divided into 8 equal time intervals ΔT_{ST} (For example, if $T_{ST} = 1024$ samples of discrete values of the force signal are taken during the swing period of the SPRU, then ΔT_{ST} will be equal to 128 samples);

2. The normalized correlation function $r_{gg}^R(\mu)$ of the noisy force signal $g(i\Delta t)$ is calculated;

3. The estimates of $r_{gg}^R(\mu=0)$, $r_{gg}^R(\mu=1\Delta T_{ST})$, $r_{gg}^R(\mu=2\Delta T_{ST})$, ..., $r_{gg}^R(\mu=7\Delta T_{ST})$ are calculated using the expressions:

$$\begin{aligned} \Delta r_{gg}^R(\mu=1\Delta T_{ST}) &= r_{gg}^R(\mu=0) - r_{gg}^R(\mu=1\Delta T_{ST}), \\ \Delta r_{gg}^R(\mu=3\Delta T_{ST}) &= r_{gg}^R(\mu=2\Delta T_{ST}) - r_{gg}^R(\mu=3\Delta T_{ST}), \\ \Delta r_{gg}^R(\mu=5\Delta T_{ST}) &= r_{gg}^R(\mu=4\Delta T_{ST}) - r_{gg}^R(\mu=5\Delta T_{ST}), \\ \Delta r_{gg}^R(\mu=7\Delta T_{ST}) &= r_{gg}^R(\mu=6\Delta T_{ST}) - r_{gg}^R(\mu=7\Delta T_{ST}). \end{aligned}$$

4. The minimum value of the normalized correlation function of the noisy signal $r_{gg}^{R_{min}}(\mu)$ and the corresponding value of μ_{min} are calculated.

5. Informative features are determined in the form of the following coefficients.

$$\begin{aligned} K_{N1}^R &= \Delta r_{gg}^R(\mu=1\Delta T_{ST}); K_{N2}^R = \Delta r_{gg}^R(\mu=3\Delta T_{ST}); \\ K_{N3}^R &= \Delta r_{gg}^R(\mu=5\Delta T_{ST}); K_{N4}^R = \Delta r_{gg}^R(\mu=7\Delta T_{ST}); \\ K_{N5}^R &= \frac{\Delta r_{gg}^R(\mu=1\Delta T_{ST})}{\Delta r_{gg}^R(\mu=3\Delta T_{ST})}; K_{N6}^R = \frac{\Delta r_{gg}^R(\mu=1\Delta T_{ST})}{\Delta r_{gg}^R(\mu=5\Delta T_{ST})}; \\ K_{N7}^R &= \frac{\Delta r_{gg}^R(\mu=1\Delta T_{ST})}{\Delta r_{gg}^R(\mu=7\Delta T_{ST})}; K_{N8}^R = \frac{\Delta r_{gg}^R(\mu=3\Delta T_{ST})}{\Delta r_{gg}^R(\mu=5\Delta T_{ST})}; \\ K_{N9}^R &= \frac{\Delta r_{gg}^R(\mu=3\Delta T_{ST})}{\Delta r_{gg}^R(\mu=7\Delta T_{ST})}; K_{N10}^R = \frac{\Delta r_{gg}^R(\mu=5\Delta T_{ST})}{\Delta r_{gg}^R(\mu=7\Delta T_{ST})}; \\ K_{N11}^R &= \Delta r_{gg}^{R_{min}}(\mu). \end{aligned}$$

Below, the possibility of using the coefficients $K^{R_{N1}}-K^{R_{N11}}$ as informative signs for identifying the technical condition of the SPRU will be shown.

The obtained informative signs in the form of coefficients $K^{R_{N1}}-K^{R_{N11}}$ were used in the complex of NOISE control systems as informative signs for identifying the technical condition of the SPRU, for the oil and gas production facility Bibi-Heybat Oil [10].

4. The system of NOISE control, diagnostics and control of SPRU with the use of frequency converters

To ensure the effective operation of the SPRU, it is necessary to carry out continuous monitoring, identification of the technical condition and control (change in the swing number of the SPRU) of the object in real time, which easily implemented based on frequency converters.

Figure 1 shows a block diagram of the NOISE monitoring, diagnostics and control system of the SPRU using frequency converters and a view of the station on the installation. The block diagram consists of three levels:

1. The level of a deep-pumping unit consisting of a plunger deep pump "1"; plunger "2"; pumping pipes "3"; rods "4"; polished rod "5"; balancer head "6"; balancer "7"; connecting rod "8"; crank counterweight "9"; gearbox "10"; V-belt transmission "11"; induction motor "12"; balancing counterweight "13"; force sensor "14"; wellhead pressure sensor "15"; rotation angle sensor "16".

2. The level of the control station, consisting of a SPRU controller according to a dynamogram for collecting information from force sensors "14", wellhead pressure "15" and rotation angle "16"; a frequency converter for controlling the speed of the electric motor; a radio modem with an antenna for information exchange of the SRPU control system with a centralized control center using the MODBUS-RTU protocol.

3. The level of the centralized control center of the oil field consists of an industrial computer and a radio modem with an antenna and serves up to 200 wells [11].

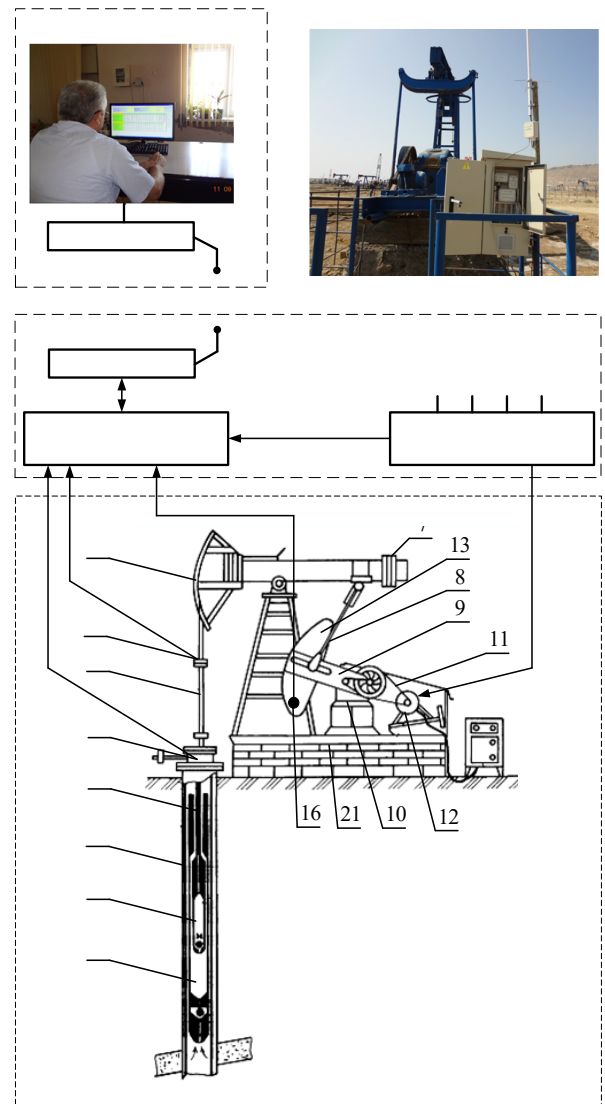


Figure 1: The block diagram of the NOISE control, diagnostics and control system of the SPRU with the use of frequency converters and the type of station on the installation

The SPRU control system uses the technology of dynamogram identification based on the estimates of the normalized correlation functions of the force signal $U_p(i\Delta t)$. To implement them, first of all, based on the duration of the interval of the swing period T_{ST} , the sampling step of the signal was determined [12]. For most oil wells, the duration of the T_{ST} time varies in the range of $5 = 20$ seconds. At the same time, we experimentally found that in order to obtain estimates of the robust correlation function $r_{gg}^R(\mu)$ with the necessary accuracy, it is sufficient to discretize the signal of the effort with a frequency $f = 50 - 100$ Hz. Experimentally, it was found that during the swing period, any slight change in the technical condition of the SPRU is reflected in the estimates of robust informative signs $K^{R_{N1}}-K^{R_{N11}}$. As a result of the SPRU operation process, the corresponding reference combinations of these coefficients are formed and stored for its various technical conditions. Identification of the technical condition of the SPRU is reduced to the formation of the current combination of informative features and their comparison with the corresponding

combinations of reference informative features $K^{R_{N1}}-K^{R_{N11}}$. Due to this, there is no need for a visual interpretation of the dynamogram to determine the current technical condition of the SPRU.

The signals from the frequency converter are sent to the induction motor and the SPRU Dynamometry controller. Further, after starting the installation in operating mode, the information is sent from the force sensors, wellhead pressure and rotation angle to the SPRU Dynamometry controller, to which a radio modem with an antenna is connected.

5. Improving the efficiency of SPRU by reducing the cost of power supply using frequency converters

As an electric drive of SPRU, induction motors (IM) with increasing starting torque of 0.4 kV, power from 7.5 to 55 kW, and synchronous rotation speed of 500 ...1500 rpm are mainly used. Electric motors (EM) of the 5A, 6A, and AIR series are widely used [13].

Induction motors are widely used in almost of industries (compressor and pumping stations, ventilation systems, conveyors, elevators, cranes, submersible pumps, etc.). The main advantages of induction motors are simplicity of design, high reliability, low cost, etc.

Interest in the use of IM has increased even more due to the appearance of a wide range of affordable and reliable frequency converters for regulating the rotation speed of electric motors in a large range [14]. Various energy-saving control stations (CS) for RDPU with the use of frequency converters began to be created, which provide significant advantages in comparison with conventional CS. Thus, the introduction and operation of CS RDPU in the fields of the Bibi-Heybat oil and gas production department showed its effectiveness. The sources of economic efficiency are the following factors [15,16]:

- saving of consumed electric energy by more than 30%;
- saving costs for the repair of wells and oil production equipment by more 20% due to the reduction of mechanical elements wear;
- reduction of production losses by more 2% due to the reduction of well downtime during repairs;
- a significant indicator of the system's performance is an increase in the inter-repair period of wells by more than 20%.

However, the calculation of electric energy savings raises doubts among consumers due to the lack of a sound calculating methodology. The article proposes a methodology for calculating the saving of electric energy when using a control station of SPRU with variable frequency drive.

6. The method of calculating the savings of electrical energy in the control of SPRU with a frequency-controlled drive

When calculating the consumed electric energy by oil-production equipment, it is necessary to take into account the active input (P_1) and output (P_2) powers of the frequency converter according to the following formulas [17], [18]:

$$P_1 = \sqrt{3} \cdot U_1 \cdot I_1 \cdot \cos \varphi_1, \quad (1)$$

$$P_2 = \sqrt{3} \cdot U_2 \cdot I_2 \cdot \cos \varphi_2, \quad (2)$$

where P_1 , U_1 , I_1 , $\cos \varphi_1$ – power, voltage, current, and power factor at the inverter input, P_2 , U_2 , I_2 , $\cos \varphi_2$ – power, voltage, current and power factor at the inverter output.

It is known that the value of the power factor at the inverter output $\cos \varphi_2$ varies depending on the degree of motor load from 0.17 to 0.85 [17], [18]. The values of the power factor at the inverter output, depending on the load for a common industrial induction motor, are shown in Table 1.

Table 1: Power Factor Dependence at the Inverter Output from the Load for a Common Industrial Induction Motor

Load, %	Power factor $\cos \varphi_2$	Power factor $\text{tg} \varphi_2$
0	0.17	5.80
25	0.55	1.52
50	0.73	0.94
75	0.80	0.75
100	0.85	0.62

The values of the power factor should be taken into account in the calculation of the energy consumed [19], [20], [21]. However, the data in Table 1 are given with great discreteness, which makes it difficult to use them in calculations. For this purpose, using piecewise linear approximation, the values $\cos \varphi_2$ for intermediate values of the IM load were determined (Table 2). The values of the power factor at the inverter output, depending on the load for a common industrial induction motor, are shown in Figure 2.

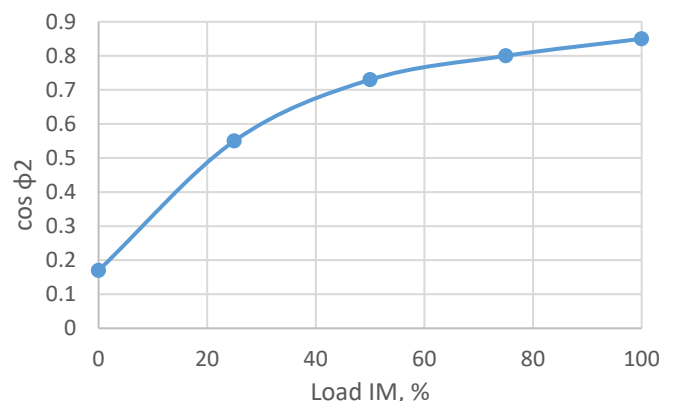


Figure 2: Dependence of Power Factor on Induction Motor Load

It should be noted that the values of $\cos\varphi_2$ were calculated using different formulas for different intervals:

$$\begin{aligned} f(x) &= 0.0152 \cdot x + 0.17; (0 \leq x < 25\%); \\ f(x) &= 0.0072 \cdot x + 0.37; (25\% \leq x < 50\%); \\ f(x) &= 0.0028 \cdot x + 0.59; (50\% \leq x < 75\%); \\ f(x) &= 0.0020 \cdot x + 0.65; (75\% \leq x < 100\%). \end{aligned}$$

Given that the inherent electrical energy consumption of the inverter is much less than the energy transmitted in the load, it can be neglected and taken:

$$P_1 = P_2;$$

$$\sqrt{3} \cdot U_1 \cdot I_1 \cdot \cos\varphi_1 = \sqrt{3} \cdot U_2 \cdot I_2 \cdot \cos\varphi_2;$$

consequently

$$I_1 = \frac{U_2 \cdot I_2 \cdot \cos\varphi_2}{U_1 \cdot \cos\varphi_1}. \quad (3)$$

Table 2: The Values $\cos\varphi_2$ Obtained Using the Piecewise Linear Approximation

Load IM, %	$\cos\varphi_2$	Load IM, %	$\cos\varphi_2$	Load IM, %	$\cos\varphi_2$
0	0.170	25	0.550	75	0.800
1	0.185	26	0.557	76	0.802
2	0.200	27	0.564	77	0.804
3	0.215	28	0.571	78	0.806
4	0.230	29	0.578	79	0.808
5	0.246	30	0.586	80	0.810
6	0.261	31	0.593	81	0.812
7	0.276	32	0.600	82	0.814
8	0.291	33	0.607	83	0.816
9	0.306	34	0.614	84	0.818
10	0.322	35	0.622	85	0.820
11	0.337	36	0.629	86	0.822
12	0.352	37	0.636	87	0.824
13	0.367	38	0.643	88	0.826
14	0.382	39	0.650	89	0.828
15	0.398	40	0.658	90	0.830
16	0.413	41	0.665	91	0.832
17	0.428	42	0.672	92	0.834
18	0.443	43	0.679	93	0.836
19	0.458	44	0.686	94	0.838
20	0.474	45	0.694	95	0.840
21	0.489	46	0.701	96	0.842
22	0.504	47	0.708	97	0.844
23	0.519	48	0.715	98	0.846
24	0.534	49	0.722	99	0.848
				100	0.850

It is also known that

1) the power factor at the input of the inverter, taking into account the total harmonic distortion, is taken equal to 0.89, that is, $\cos\varphi_2 = 0,89$;

2) with frequency regulation, the ratio of voltage to frequency is assumed constant. Since the input of the inverter

$$\frac{U_1}{F_1} = \text{const} = k,$$

for the value U_2 we get:

$$U_2 = F_2 \cdot k.$$

Then (3) can be rewritten in the form:

$$I_1 = \frac{F_2 \cdot k \cdot I_2 \cdot \cos\varphi_2}{F_1 \cdot k \cdot 0.89} = \frac{F_2 \cdot I_2 \cdot \cos\varphi_2}{F_1 \cdot 0.89}. \quad (4)$$

Electric energy saving will be due to the difference in currents:

$$\Delta I = I_2 - I_1. \quad (5)$$

Putting (4) in (5) we get:

$$\Delta I = I_2 \cdot \left(1 - \frac{F_2 \cdot \cos\varphi_2}{F_1 \cdot 0.89} \right). \quad (6)$$

Electric energy saving in N hours will be in kWh:

$$\begin{aligned} EES_N &= \sqrt{3} \cdot N \cdot U_1 \cdot \Delta I \cdot 0.89 \cdot 10^{-3} = \\ &= 1.5394 \cdot N \cdot U_1 \cdot \Delta I \cdot 10^{-3}. \end{aligned} \quad (7)$$

For example, when powering the SPRU electrical motor from a network with a voltage $U_1=380$ V and a frequency 50 Hz, we obtain:

$$\Delta I = I_2 \cdot \left(1 - \frac{F_2 \cdot \cos\varphi_2}{50 \cdot 0.89} \right) = I_2 \cdot \left(1 - \frac{F_2 \cdot \cos\varphi_2}{44.5} \right); \quad (8)$$

$$EES_N = EES \cdot N = 0.585 \cdot \Delta I \cdot N. \quad (9)$$

Using the obtained expressions, the value of the electric energy saving parameters were calculated for the induction motor operation modes ($P_{NOM} = 37$ kW, $\cos\varphi_{NOM} = 0,85$, $I_{NOM} = 66$ A) with different load, the calculation results are summarized in Table 3 and Figure 3.

The general algorithm for calculating the values of electric energy saving parameters:

1 The inverter output current I_2 is measured, i.e. the current consumed by the motor.

2 The motor load value is calculated as the ratio of the actual output current I_2 to the nominal current I_{NOM} . The value of the motor load from Table II determines the value of the power factor $\cos\phi_2$.

3 Using (6), the difference between the input and output currents ΔI is determined.

4 Using (7) and using the known input voltage U_1 , the value of electric energy saving EES per hour is determined.

Table 3: Electric Energy Saving Parameters Depending on the Operation Mode of an Induction Motor

Load IM, %	I_2 , A	$\cos\phi_2$	Current difference, ΔI , A	Annual savings, kWh
100% rotation speed, $F_1=50$ Hz, $U_1=380$ V, $\cos\phi_1=0.89$, $F_2=50$ Hz				
100	66.0	0.85	2.97	15220
75	49.5	0.80	5.01	25685
50	33.0	0.73	5.93	30441
25	16.5	0.55	6.30	32344
100% rotation speed, $F_1=50$ Hz, $U_1=380$ V, $\cos\phi_1=0.89$, $F_2=25$ Hz				
100	66.0	0.85	34.5	176942
75	49.5	0.80	27.3	139841
50	33.0	0.73	19.5	99886

25	16.5	0.55	11.4	58505
----	------	------	------	-------

The particular algorithm for calculating the values of the parameters of saving electrical energy for the SPRU, where the power supply is carried out with a voltage $U_1=380$ V and a frequency of 50Hz. This algorithm differs from the general one in that formulas are used here, respectively (6), (7), (8), (9).

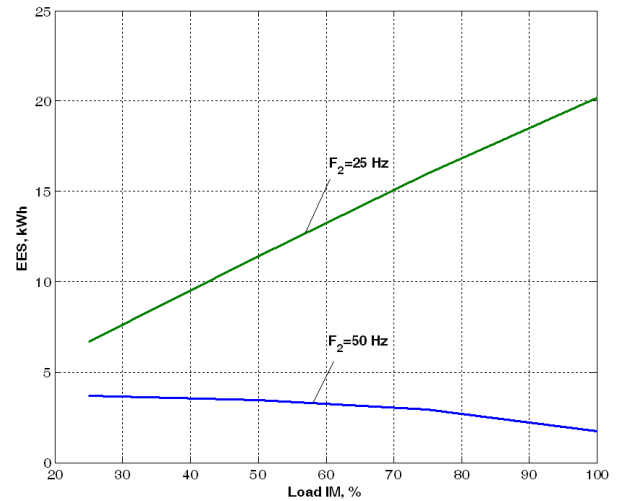


Figure 3: Electric Energy saving of an Induction Motor

Table 4: Calculation of Energy Savings for 20 Wells of the Bibi-Heybat Oil and Gas Production Unit Equipped with Developed Control Stations with Inverters (August 2015)

Well number	P_{FC} , kW	P_{IM} , kW	I_{2max} , A	F_2 , Hz	$\cos\phi_1$	U_1 , V	I_1 , A	Load IM, %	$\cos\phi_2$	U_2 , V	I_2 , A	ΔI , A	SEE, kWh
3882	22	22	48.4	47	0.89	380	10.77	34.2	0.614	357	16.6	5.83	3.41
3097	30	30	66.0	49	0.89	380	20.06	41.5	0.665	372	27.4	7.34	4.29
3845	22	22	48.4	45	0.89	380	15.74	45.8	0.701	342	22.2	6.46	3.78
3098	30	30	66	50	0.89	380	12.97	29.8	0.586	380	19.7	6.73	3.94
3658	22	22	48.4	45	0.89	380	9.883	33.2	0.607	342	16.1	6.22	3.64
3844	22	22	48.4	45	0.89	380	14.0	42.5	0.672	342	20.6	6.6	3.86
3814	30	30	66.0	45	0.89	380	19.03	42.4	0.672	342	28	8.97	5.25
3617	30	30	66.0	49	0.89	380	15.14	33.9	0.614	372	22.4	7.26	4.25
3841	18.5	22	48.4	30	0.89	380	10.11	44.6	0.694	228	21.6	11.5	6.73
1245	18.5	18.5	40.7	49	0.89	380	9.398	34.1	0.614	372	13.9	4.5	2.63
3261	18.5	15	33.0	42	0.89	380	14.83	62.4	0.763	319	20.6	5.77	3.37
3146	18.5	22	48.4	45	0.89	380	19.11	52.8	0.738	342	25.6	6.49	3.80
3677	30	30	66.0	49	0.89	380	5.707	17.7	0.443	372	11.7	5.99	3.51
3826	22	22	48.4	48	0.89	380	21.51	55.3	0.744	364	26.8	5.29	3.10
811	18.5	18.5	40.7	45	0.89	380	9.712	37.1	0.636	342	15.1	5.39	3.15
1150	18.5	18.5	40.7	45	0.89	380	10.71	39.5	0.658	342	16.1	5.39	3.15
1192	18.5	18.5	40.7	38	0.89	380	12.72	50.1	0.73	288	20.4	7.68	4.50
3279	18.5	18.5	40.7	46	0.89	380	16.11	52.0	0.735	349	21.2	5.09	2.98
557	18.5	18.5	40.7	48	0.89	380	9.594	35.1	0.622	364	14.3	4.71	2.75
3929	18.5	22	48.4	40	0.89	380	7.326	29.1	0.578	304	14.1	6.77	3.96
Total savings of all wells in 1 hour, kWh													76.1
Total working time in a year (when working 11 months a year), h													8040
Total annual energy savings, thousand kWh													612.2

Total annual savings in money (at an electric power cost of 0.05 \$ / kWh), \$	30608
--	-------

Currently, the Institute of Control Systems of the National Academy of Sciences of Azerbaijan (ANAS) has developed a CS based on frequency converters of the Altivar 71 and 930 type, which is manufactured by the special design bureau "Cybernetics" of ANAS. These SUS are being widely implemented in the oil fields of Azerbaijan. At the moment, more than 400 wells are equipped with frequency converters based control stations as part of a measurement, monitoring, diagnostics and management complex. At the same time, the electric energy savings are calculated according to the proposed methodology. Table 4 shows the results of calculating energy savings for 20 wells of the Bibi-Heybat Oil and Gas Production Unit equipped with the developed CS.

7. Conclusion

1. Using the normalized correlation functions, the normalized coefficients K_{N1} , K_{N2} , ..., K_{N11} , free from the indicated interferences, are determined.

2. In each field for wells with the same depth, the range of change in the values of the same coefficients $K_{N1} - K_{N11}$ under the same technical conditions is insignificant (no more than 5-10%), which allows using them as informative parameters for identifying the technical condition of the SPRU.

3. A robust remote control system has developed for the SPRU in real-time based on a frequency-controlled driver.

4. A methodology for calculating the savings of electrical energy in the control of SPRU frequency converters has compiled. The results of the calculation also reveal the available additional energy saving reserves, such as:

- the need for accurate balancing counterweights of SPRU;
- measures to increase power factor $\cos\varphi$;
- the need to maintain the operating mode with motor load within 70...75% of the nominal and the frequency within 80...85% of the nominal.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] G.P. Karmakar, Chandrima Chacraborty. "Improved oil recovery using polymer gels: a review," *Indian Journal of Chemical Technology*, 2006, no 13, pp. 162-167.
- [2] V.M. Maksimov. "On the current state of oil production, recovery factor oil and methods of increasing oil recovery," *Drilling and oil*, 2011, no 2, pp. 12-16.
- [3] "Marketing research of the sucker rod pumps market. Analytical report," *Research.Techart*. 2010.
- [4] A.B. Neely, H.E. Tolbert, "Experience with Pump-Off Control in the Permian Basin," *SPE Technical Conference and Journal of Petroleum Technology*, 1988, no 14345, pp.645-648.
- [5] Li, K., Han, Y., Wang, T., 2018. "A novel prediction method for down-hole working conditions of the beam pumping unit based on 8-directions chain codes and online sequential extreme learning machine," *Journal of Petroleum Science and Engineerin*, vol. 160, pp. 285-301.
- [6] A.H. Rzayev, M.H. Rezvan, M.I. Khakimyanov, I.N. Shafikov, "State-of-the-art of the automation of artificial lift units in the CIS," *Transaction of Azerbaijan National Academy of Sciences, Series of Physical-Technical and Mathematical Sciences: Informatics and Control Problems*, vol. XXXIII, no 6, 2013, pp. 177 - 187.
- [7] Zhang, R., Wang, Z., Wang, X., et al., 2018, "Integrated Diagnostics Method and Application of Ground and Downhole Working Condition in Rod Pumping Well," *Journal of Applied Science and Engineering*, vol. 24, no 4, pp. 615-624.
- [8] Han, D., 2010. "Discussions on concepts, countermeasures and technical routes for the secondary development of high water-cut oil fields," *PetroleumExplorationandDevelopment*, vol. 37, no 5, pp. 583-591.
- [9] T.A. Aliev, D.A. Iskenderov, G.A. Guluyev, A.H. Rzayev, M.H. Rezvan, "Results of introducing the control, diagnostics and management complex for oil wells operated by sucker rod pumps at "Bibi Heybat Oil" oil and gas producing company," *Azerbaijan Oil Econ*, no 6, 2014, pp. 37 - 41.
- [10] T.A. Aliev, A.H. Rzayev, G.A. Guluyev, T.A. Alizada, N.E. Rzayev, "Robust technology and system for management of sucker rod pumping units in oil wells," in *Mechanical Systems and Signal Processing*, vol. 99, 2018, pp. 47 - 56.
- [11] F.A. Gizatullin, M.I. Khakimyanov, F.F. Khusainov, "Features of electric drive sucker rod pumps for oil production," *Journal of Physics: Conference Series*, vol. 944, 2018, pp. 1-6.
- [12] A.H. Rzayev, M.H. Rezvan, M.I. Khakimyanov, "Automation artificial lift systems in the CIS," *Oil and Gas Business: electronic scientific journal*, vol. 5, no 6, pp. 19 - 29. [Online]. Available: <http://ogbus.ru/article/view/avtomatizaciya-ustanovok-mexanizirovannoj-dobychi-nefti-na-territorii-sng/22792>. Accessed: Jan, 2020.
- [13] T. Neuberger, S.B. Weston, "Variable Frequency Drives: Energy Savings for Pumping Applications," *Eaton Corporation*, no. IA04008002E /Z12581, 2012.
- [14] E. Al-Bassam, R. Alasser, "Measurable energy savings of installing variable frequency drives for cooling towers' fans, compared to dual speed motors," *Energy and Buildings*, vol. 67, 2013, pp. 261-266.
- [15] S. Prachyl, "Variable Frequency Drives and Energy Savings," *Siemens Industry Inc. Drive Technologies*, 2010.
- [16] C.M. Burt, X. Piao, F. Gaudi, B. Busch, N.F. Taufik, "Electric motor efficiency under variable frequencies and loads," *Journal of irrigation and drainage engineering*, vol. 134, no 2, 2008, pp. 129 - 136.
- [17] W.H. Wolfle, W.G. Hurley, "Power factor correction and harmonic filtering for AC/DC converters," in *IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02*, vol. 4, 2002, pp. 3238-3243.
- [18] A.H.do Nascimento, "Power factor correction and harmonic filtering planning in electrical distribution network," *Journal of Control, Automation and Electrical Systems*, vol. 27, no 4, 2016, pp. 441-451.
- [19] C. Liu, Y. He, K. Dai, Y. Kang, "Industrial power distribution system harmonic resonance problem and solution with shunt active power filter," in *IECON 2017- 43rd Annual Conference of the IEEE Industrial Electronics Society*, 2017, pp. 1243-1248.
- [20] T. Aliev, O. Nusratov, G.A. Guluev, As.G. Rzaev, F. Pashayev, M.G. Rezvan "Position-Binary Technology for Processing the Signals of a Load for Identification of the Technical State of Deep-Well

Sucker-Rod Pumps," *Measurement Techniques*, vol. 61(3), 2018, pp. 1–6.

- [21] A.M. Zyuzev, M.V. Bubnov "Sucker-rod pumping unit balance diagnostics by wattmeter card," *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, vol. 330, 2019, pp. 178–187.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



ALIEV TELMAN ABBAS graduated from the Azerbaijan Industrial Institute in 1958. In 1966, he defended his PhD thesis at the Bauman Moscow State Technical University. In 1977, he defended his doctoral dissertation on the specialty "Technical cybernetics and information theory". In 1985, he received the title of professor. In 1983 he was elected a corresponding member, and in 2001 - a full member of the National Academy of Sciences of Azerbaijan. He is the author of NOISE signal analysis technologies for early and adequate diagnostics of the technical condition of objects.



GULUYEV GAMBAR AGAVERDI graduated from the Azerbaijan Polytechnic Institute in 1973 with a degree in automation and telemechanic. In 2009 year, he defended his Ph.D. thesis at the Institute of Control Systems of ANAS. Doctor of Technical Sciences since 2015 years. The main direction of research his is the creation of monitoring, control systems that improve the quality of identification of technical objects and systems, and increase their profitability.



RZAYEV ASIF HAJI graduated from the Azerbaijan Institute of Oil and Chemistry in 1971 with a degree in electrical engineering. In 1983, he defended his Ph.D. thesis at the Azerbaijan Institute of Oil and Chemistry. Doctor of Technical Sciences since 2018. The main direction of research is the creation of digital devices for control, management and systems for automating the processes of oil production, transportation and storage.



ALIYEV YAVER GABIL graduated from Azerbaijan Polytechnic Institute in 1980 with a degree in electrical engineering. He is a senior researcher and leading specialist.
The main direction of the research is the creation of controllers oil pumps and measure oil production.



REZVAN MAHAMMAD HUSEYN graduated from Azerbaijan State University in 1986 with a degree of mathematics. He is a senior researcher and leading specialist.
The main direction of the research is the creation of software for controller's oil pumps and measure oil production.



YASHIN ANTON NIKOLAEVICH has done his bachelor's degree from Ufa State Petroleum Technological University in 2018. He has done his master's degree from Ufa State Petroleum Technological University in 2020.

Research experience related to the study of the electric drive in the oil industry.



KHAKIMYANOV MARAT ILGIZOVICH has done his engineering degree from Ufa State Petroleum Technological University in 1999. He has completed his Doctor of Technical Sciences degree in Ufa State Aviation Technical University in 2019.

Research experience related to the study of the electric drive in the oil industry.

Soil Properties Prediction for Agriculture using Machine Learning Techniques

Vijay Kumar*, Jai Singh Malhotra, Saurav Sharma, Parth Bhardwaj

Department of Computer Science and Engineering, National Institute of Technology Hamirpur, Hamirpur, 177005, India

*Corresponding author: Vijay Kumar, NIT Hamirpur, vijaykumarchahar@gmail.com

ABSTRACT: Information about soil properties help the farmers to do effective and efficient farming, and yield more crops with less usage of resources. An attempt has been made in this paper to predict the soil properties using machine learning approaches. The main properties of soil prediction are Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand. These properties greatly affect the production of crops. Four well-known machine learning models, namely, multiple linear regression, random forest regression, support vector machine, and gradient boosting, are used for prediction of these soil properties. The performance of these models is evaluated on Africa Soil Property Prediction dataset. Experimental results reveal that the gradient boosting outperforms the other models in terms of coefficient of determination. Gradient boosting is able to predict all the soil properties accurately except phosphorus. It will be helpful for the farmers to know the properties of the soil in their particular terrain.

KEYWORDS Machine Learning, Agriculture, Prediction, Soil Properties

1. Introduction

India has a 1.27 billion population, which is second-most in the entire world. It is the seventh-largest country in the world with an area of 3.288 million sq km. Indians are very much dependent on agriculture. It is the largest source of livelihood in India. In rural households, 70% of people are primarily dependent on agriculture, with about 82% of farmers being small and marginal. In 2020-21, total food grain production was estimated at 308.65 million tonnes (MT). India is the largest producer (25% of global production), the consumer (27% of world consumption), and the importer (14%) of pulses in the world. India's annual milk production was 165 MT (2017-18), making India the largest producer of milk, jute, and pulses, with the world's second-largest cattle population of 190 million in 2012 [1]. With merely 2.4% arable land resources and 4% water resources [2], Indian agriculture is feeding nearly 1.3 billion people, which implicates huge pressure on land and other natural resources for continuous productivity [3].

After the green revolution (which started in the 1960s), India made significant progress in agriculture production, which became possible due to modernization. With the development in technology, farmers have been provided with advanced farming techniques, better seeds (High Yielding Variety (HYV) seeds), mechanized farm tools, chemical fertilizers, facilities of irrigation, and electrical energy [4]. Since the green revolution, there has been excessive use of chemical fertilizers which has increased the crop productivity manifold. However, it has turned into a problem as overuse

of these chemical fertilizers has been detrimental for crop productivity and soil fertility. Fertilizer recommendations rarely match soil needs which has caused overuse of these chemical products [3].

So, there is a need for accurate fertilizer recommendations for the farmer and accurately analyzing soil properties is the first step for that. Indian Agricultural Research Institute (ICAR) recommends soil test-based, balanced and integrated nutrient management through conjunctive use of both inorganic and organic sources of plant nutrients to reduce the use of chemical fertilizers, preventing deterioration of soil health, environment and contamination of groundwater [5].

This paper aims to study the ability of various machine learning techniques to accurately predict the soil properties relevant for agriculture using spectroscopy data. Over the last 20 years, soil spectroscopy has become a powerful technique for analyzing relative to the traditionally used chemical methods, particularly in the infrared range. Spectroscopy is known as a fast, economical, quantitative, and eco-friendly technique, which can be used in the fields as well as in the laboratory to provide hyperspectral data with narrow and numerous data [6], [7]. In this paper, the different properties of soil like Calcium, Phosphorus, pH, Soil Organic Carbon and Sand are predicted by using machine learning models. In [8], it is found consistently higher performance of machine learning methods over simpler approaches in spectroscopy. In [9], it is reported the decline in the use of some models such as Support Vector Machines (SVM) and multivariate adaptive regression spline, giving

way to more advanced alternatives such as Random Forest (RF). In this paper, machine learning algorithms such as Multivariate Regression, Random Forest Regression, Support Vector Machine, and Gradient boosting with a different degree of accuracy are used for comparative analysis. The dataset is split into a training and testing dataset (80% training data and 20% testing data) [10]. The machine learning models are trained on the training data. After a model is trained, the testing data is used to check the accuracy of the trained model. Here, the coefficient of determination (COD) is calculated to check the working of the models after being trained. After training the models, the best working model is deployed to predict the properties of the soil (Calcium, Phosphorous, pH, Soil Organic Carbon, and Sand). These predicted values of the soil properties are going to be helpful in choosing the different suitable fertilizers.

The remaining structure of this paper is as follows. Section 2 presents the materials and methods used for soil prediction. Experimental results and discussion are mentioned in Section 3. Section 4 presents the concluding remarks.

2. Materials and Methods

In this section, the dataset and techniques used for soil prediction are briefly described.

2.1. Data Set

A collection of 1,886 soil sample measures is used for performance comparison of machine learning models. The soil was collected from a variety of locations in Africa. Each data point consists of 3,594 features :

1. **PIDN:** unique soil sample identifier
2. **SOC:** Soil organic carbon
3. **pH:** pH values
4. **Ca:** Mehlich-3 extractable Calcium
5. **P:** Mehlich-3 extractable Phosphorus
6. **Sand:** Sand content
7. **m7497.96 - m599.76:** There are 3,578 mid-infrared absorbance measurements. For example, the "m599.76" column is the absorbance at wavenumber 599.76 cm⁻¹.
8. **Depth:** Depth of the soil sample (2 categories: 1. "Topsoil", 2. "Subsoil")
9. **BSA:** Average long-term Black Sky Albedo measurements from MODIS satellite images (BSAN = near-infrared, BSAS = shortwave, BSAV = visible)
10. **CTI:** Compound topographic index calculated from Shuttle Radar Topography Mission elevation data
11. **ELEV:** Shuttle Radar Topography Mission elevation data
12. **EVI:** Average long-term Enhanced Vegetation Index from MODIS satellite images

13. **LST:** Average long-term Land Surface Temperatures from MODIS satellite images (LSTD=day time temperature, LSTN night time temperature)
14. **Ref:** Average long-term Reflectance measurements from MODIS satellite images (Ref1 = blue, Ref2 = red, Ref3 = near-infrared, Ref7 = mid-infrared)
15. **Reli:** Topographic Relief calculated from Shuttle Radar Topography mission elevation data
16. **TMAP TMFI:** Average long-term Tropical Rainfall Monitoring Mission data (TMAP = Mean Annual Precipitation, TMFI = Modified Fournier Index)

The five main target variables for predictions are: Soil Organic Carbon(SOC), pH, Calcium, Phosphorus, and Sand. The data has not been altered and is in the original measurements. Thus, it includes both positive as well as negative values. The dataset was available at Kaggle.com under a competition named "Africa Soil Property Prediction Competition" [11].

2.2. Techniques Used

2.2.1. Linear Regression

Regression is an approach to supervised learning. It can be used to model continuous variables or make predictions. Some examples of application of linear regression algorithm are: prediction of the price of real estate, forecasting of sales, prediction of students' exam scores, forecasting of movements in the price of a stock in the stock exchange. In Regression, we have the labeled datasets and the output variable value is determined by input variable values. The most simple form of regression is linear regression where the attempt is made to fit a straight line (straight hyper-plane) to the dataset and it is possible when the relationship between the variables of the dataset is linear as shown in Figure 1

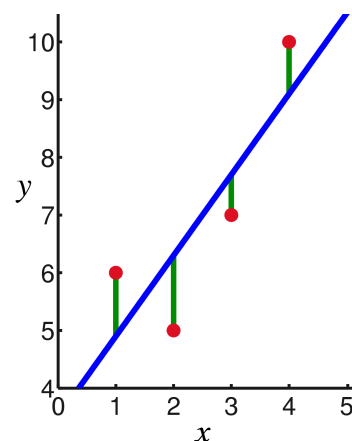


Figure 1: Linear Regression

Advantage of linear regression is, that it is easy to understand and it is also easy to avoid overfitting by regularization. SGD is used to update linear models with new data. Linear Regression is a good fit if it is known that the relationship between covariates and response variables is linear.

It shifts focus from statistical modeling to data analysis and preprocessing. Linear Regression is good for learning about the data analysis process. However, it is not a recommended method for most practical applications because it oversimplifies real-world problems [12]-[14].

2.2.2. Multiple Linear Regression

A simple linear regression model has a dependent variable guided by a single independent variable. However, real-life problems are more complex. Generally, one dependent variable depends on multiple factors. For example, the price of a house depends on many factors like the neighborhood it is situated in, its area, number of rooms, attached facilities, availability of nearby facilities like airport/railways/shopping centres, etc. In simple linear regression, there is a one-to-one relationship between the input variable and the output variable. But in multiple linear regression, there is a many-to-one relationship between a number of independent (input/predictor) variables and one dependent (output/response) variable. Adding more input variables does not mean the regression will be better or will offer better predictions.

This technique gives a deep insight into the relationship between the set of independent variables and dependent variables. It also gives insight into relationships among the independent variables. This is achieved through multiple regression, tabulation techniques, and partial correlation. It models complex real-world problems in a practical and realistic way.

However, it suffers from high computational complexity, requires knowledge and expertise on statistical techniques and statistical modeling. The sample size for statistical modeling needs to be high to get a higher confidence level on analysis outcome. Also, it often gets too difficult to do a meaningful analysis and interpretation of the outputs of the statistical model [12]-[14].

2.2.3. Decision Tree

A Decision Tree is a Supervised Machine Learning approach to solve classification and regression problems by continuously splitting data based on a certain parameter. The decisions are in the leaves and the data is split in the nodes as shown in Figure 2. In the Classification Tree, the decision variable is categorical (outcome in the form of Yes/No) and in the Regression tree, the decision variable is continuous. Decision Tree is suitable for regression as well as classification problems. It offers ease in interpretation, handles categorical and quantitative values, is capable of filling missing values in attributes with the most probable value and assures a high performance due to efficiency of the tree traversal algorithm. Decision Tree might encounter the problem of overfitting for which Random Forest is the solution which is based on an ensemble modeling approach[15]. Disadvantages of decision tree are: unstable, difficult to control size of the tree, prone to sampling error and locally optimal solution. Decision Trees can be used in predicting the future use of library books and tumor prognosis problems [12].

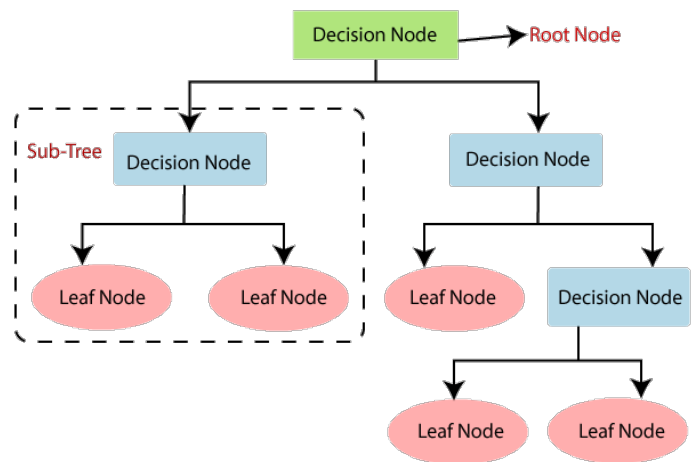


Figure 2: Decision Tree

2.2.4. Random Forest

Random forests or random decision forests is an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned as shown in Figure 3. Random decision forests correct for decision tree's habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance [16]-[18].

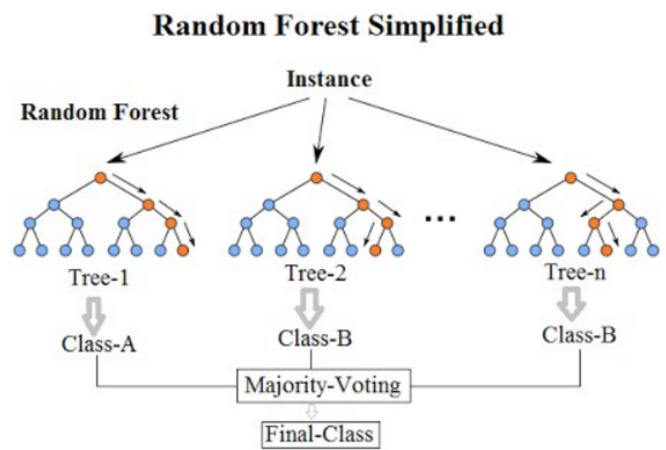


Figure 3: Random Forrest

2.2.5. Gradient Boosting

Gradient boosting is used in regression and classification tasks. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees as shown in Figure 4. When a decision tree is a weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest. A gradient-boosted model is built in a stage-wise fashion, as in other boosting methods, but it generalizes the other methods

by allowing optimization of an arbitrary differential loss function [19] [18]

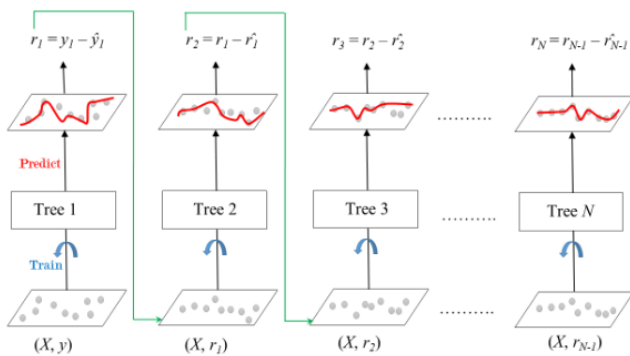


Figure 4: Gradient Boosting

2.2.6. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model, which can be used for both classification and regression. It is a non-probabilistic binary linear classifier. It was developed in 1993 at Bell laboratories. It is one of the most robust learning frameworks. SVM maps training samples into a sample space to maximize the width between two categories. New samples are mapped to a space and they are classified on the base of which side of the gap they are, as shown in Figure 5 [20].

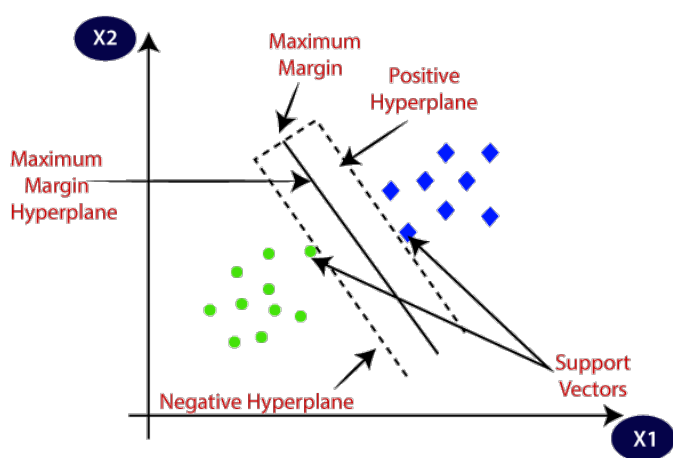


Figure 5: Support Vector Machine

2.3. Methodology used for soil prediction

2.3.1. Preparing the data

- 1. Removing irrelevant variables :** Removing the unnecessary /irrelevant variables like PIDN are removed during the preprocessing phase. These variables in the data set might lead to the model not working properly due to a possible false correlation
- 2. Normalizing data :** Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the data set to use a common scale without distorting differences in the

ranges of values or losing information. Normalization is also required for some algorithms to model the data correctly. For eg. assume your input data set contains one column with values ranging from 0 to 1, and another column with values ranging from 10,000 to 100,000. The great difference in the scale of the numbers could cause problems when you attempt to combine the values as features during modeling. Normalization avoids these problems by creating new values that maintain the general distribution and ratios in the source data, while keeping values within a scale applied across all numeric columns used in the data set.

- 3. Splitting data into train and test sets :** The train-test split procedure is used to estimate the performance of machine learning algorithms. They are used to make predictions on data. It is a fast and easy procedure to perform, the results of which allow you to compare the performance of machine learning algorithms for your predictive modeling problem. Although it is simple to use and interpret. Some classification problems do not have a balanced number of examples for each class label. As such, it is desirable to split the dataset into train and test sets in a way that preserves the same proportions of examples in each class as observed in the original dataset.

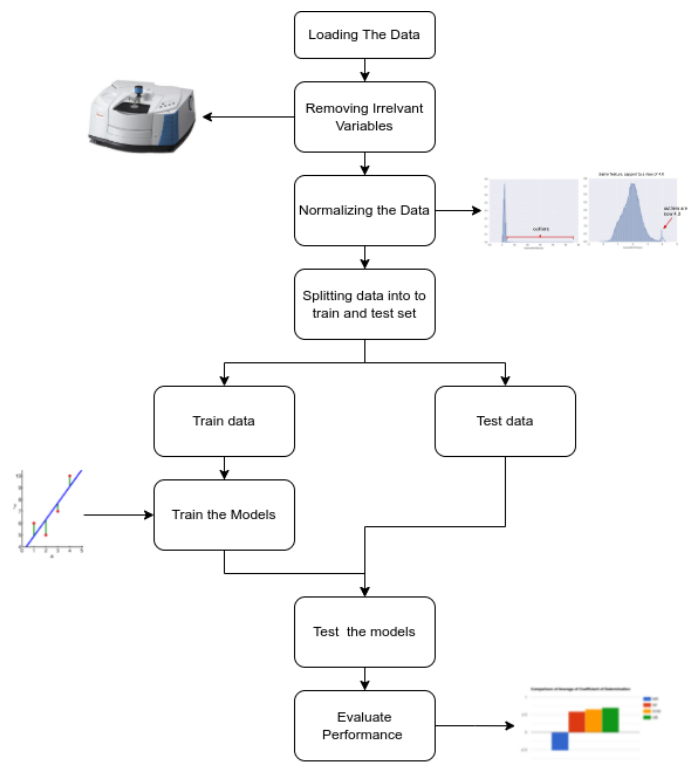


Figure 6: Methodology used for soil prediction

3. Results and Discussion

In this section, the performance evaluation of machine learning models is evaluated.

3.1. Performance measure

Coefficient of determination as metric is used for comparing the performance of different machine learning models trained on the same data set. In statistics, the coefficient of determination (R^2 or r^2 and pronounced "R squared"), is the proportion of the variation in the dependent variable that is predictable from the independent variable(s). The formula is defined as [21]:

$$R^2 = 1 - RSS/TSS$$

Where RSS represents sum of squares residuals and TSS represents Total sum of squares.

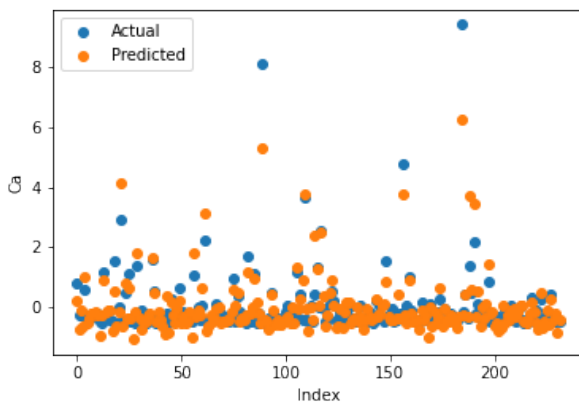


Figure 7: Actual versus predicted values of Calcium using multiple linear regression

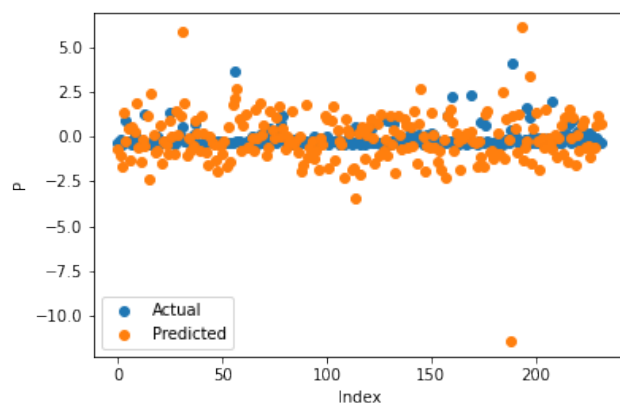


Figure 8: Actual versus predicted values of Phosphorous using multiple linear regression

Table 1: Multivariate Regression Coefficient of Determination

Calcium	0.809695
Phosphorous	-5.556849
pH	0.618799
Soil Organic Carbon	0.851829
Sand	0.691344

3.2. Multiple Linear Regression

Figures 7-11 depict the actual and predicted values of soil properties using multiple linear regression. It is observed from the figures that the predicted values attained from multiple linear regression are almost similar to the actual values for calcium and soil carbon. Table 1 shows the results obtained from multiple linear regression. The results reveal that the value of the coefficient of determination is higher for calcium and carbon.

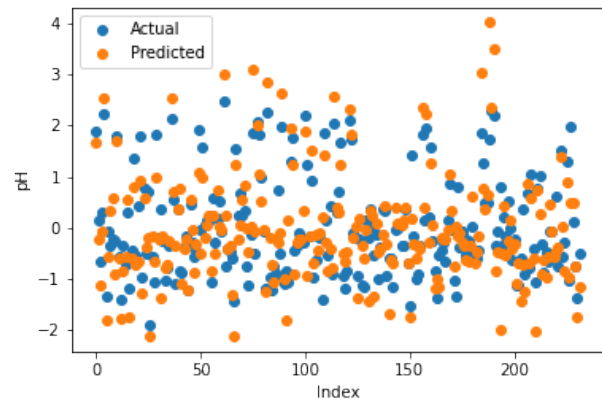


Figure 9: Actual versus predicted values of pH using multiple linear regression

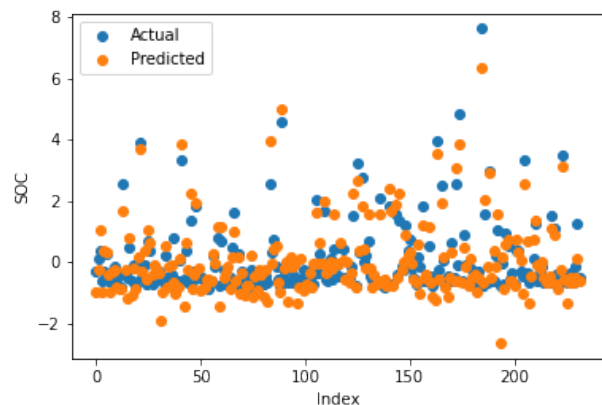


Figure 10: Actual versus predicted values of Soil Organic Content using multiple linear regression

3.3. Random Forrest Regression

Figures 12-16 show the actual and predicted values of soil properties using random forest. It can be found from the figures that the predicted values attained from random forest are almost similar to the actual values for calcium, sand, and soil carbon. Table 2 represents the values of coefficient of determination obtained from random forest. The results reveal that the value of coefficient of determination is higher for calcium, carbon, and sand.

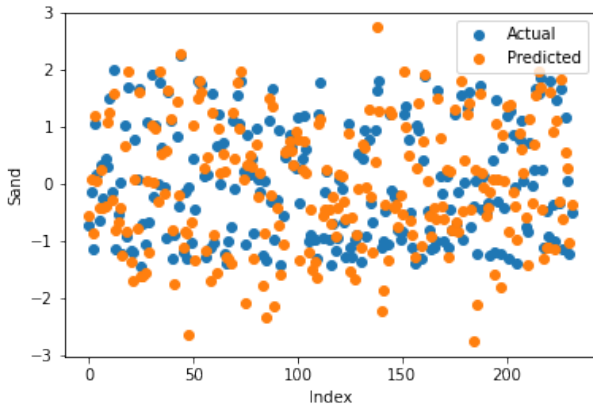


Figure 11: Actual versus predicted values of Sand using multiple linear regression

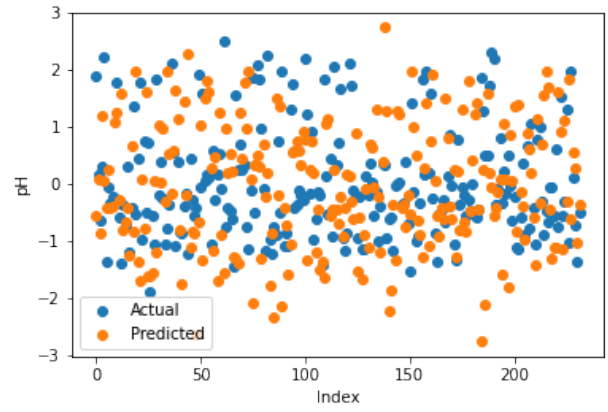


Figure 14: Actual versus predicted values of pH using random forest regression

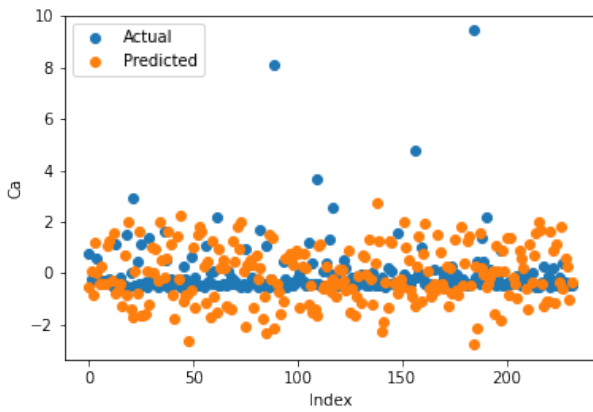


Figure 12: Actual versus predicted values of Calcium using random forest regression

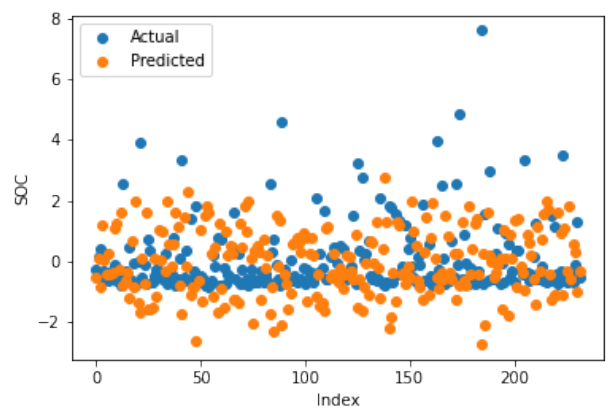


Figure 15: Actual versus predicted values of Soil Organic Content using random forest regression

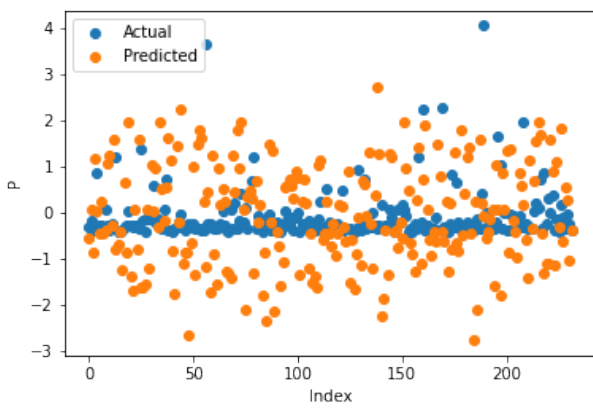


Figure 13: Actual versus predicted values of Phosphorous using random forest regression

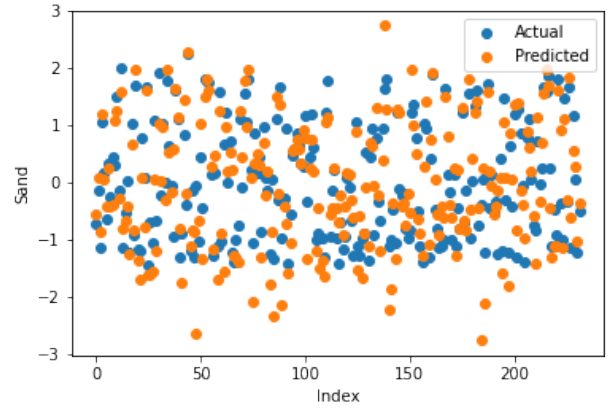


Figure 16: Actual versus predicted values of Sand using random forest regression

Table 2: Random Forrest Regression Coefficient of Determination

Calcium	0.842057
Phosphorous	-0.179327
pH	0.647040
Soil Organic Carbon	0.812501
Sand	0.822526

3.4. Support Vector Machine Regression

Figures 17-21 illustrate the actual and predicted values of soil properties using support vector machine. These figures reveal that the predicted values attained from support vector machine are almost similar to the actual values for calcium, sand, pH, and soil carbon. Table 3 represents the values of the coefficient of determination obtained from support vector machine. The results reveal that the value of the coefficient of determination is higher for calcium, carbon, pH, and sand.

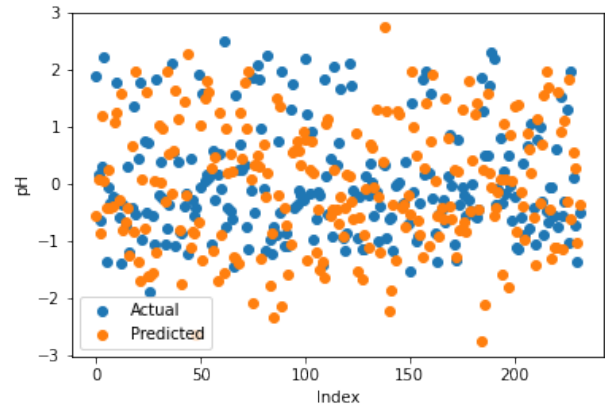


Figure 19: Actual versus predicted values of pH using support vector machine regression

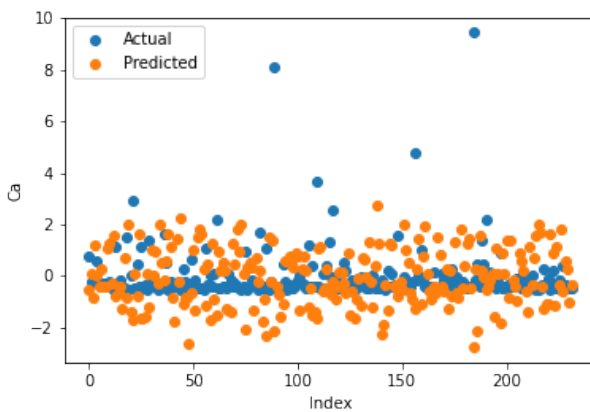


Figure 17: Actual versus predicted values of Calcium using support vector machine regression

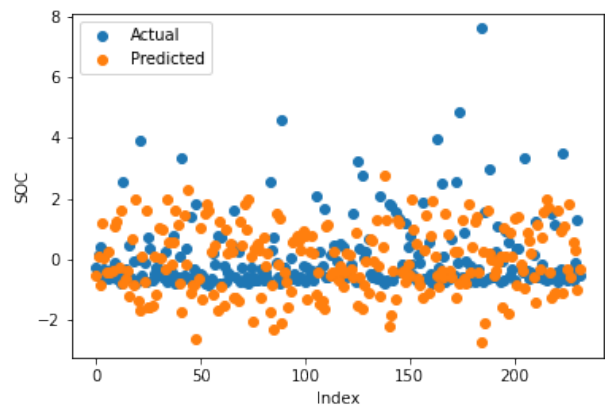


Figure 20: Actual versus predicted values of Soil Organic Content using support vector machine regression

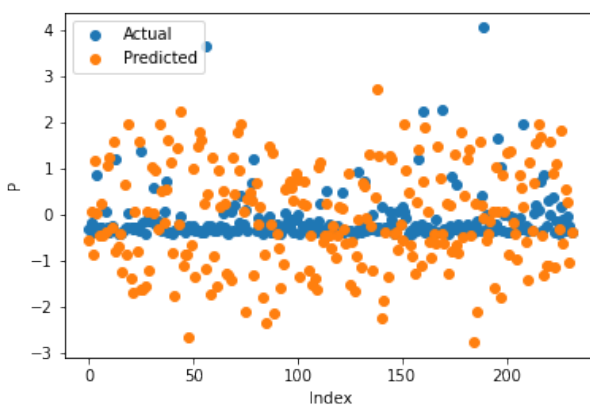


Figure 18: Actual versus predicted values of Phosphorous using support vector machine regression

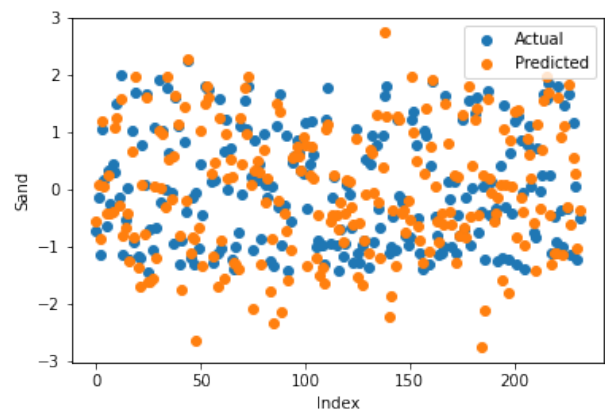


Figure 21: Actual versus predicted values of Sand using support vector machine regression

Table 3: Support Vector Regression Coefficient of Determination

Calcium	0.738632
Phosphorous	0.196469
pH	0.746837
Soil Organic Carbon	0.783020
Sand	0.840441

3.5. Stochastic Gradient Boosting Regression

Figures 22-26 show the actual and predicted values of soil properties using gradient boosting. These figures reveal that the predicted values attained from gradient boosting are almost similar to the actual values for calcium, sand, pH, and soil carbon. Table 4 represents the values of the coefficient of determination obtained from gradient boosting. The results reveal that the value of the coefficient of determination is higher for calcium, carbon, pH, and sand.

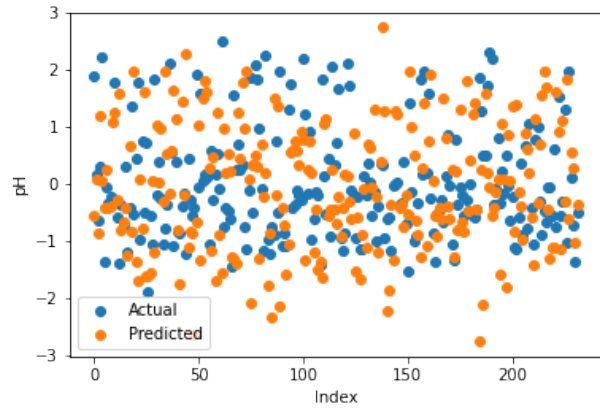


Figure 24: Actual versus predicted values of pH using stochastic gradient boosting regression

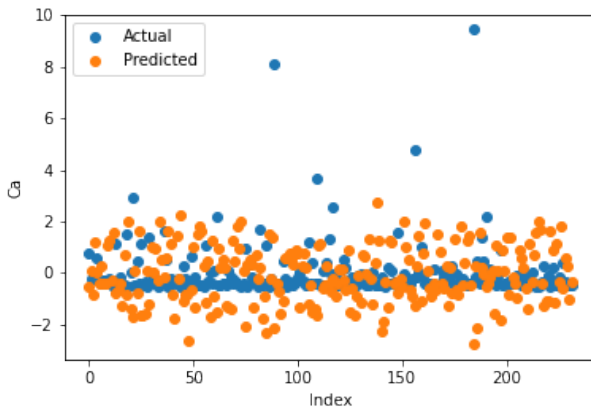


Figure 22: Actual versus predicted values of Calcium using stochastic gradient boosting regression

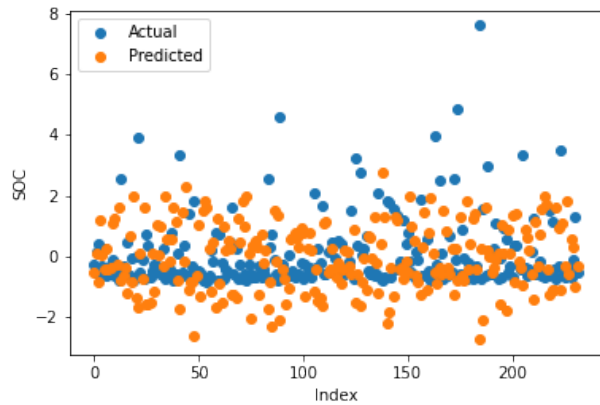


Figure 25: Actual versus predicted values of Soil Organic Content using stochastic gradient boosting regression

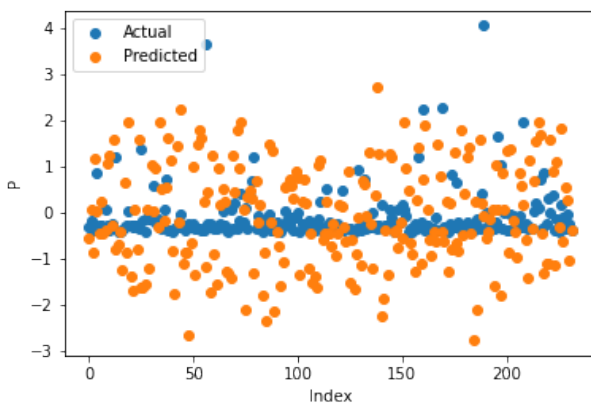


Figure 23: Actual versus predicted values of Phosphorous using stochastic gradient boosting regression

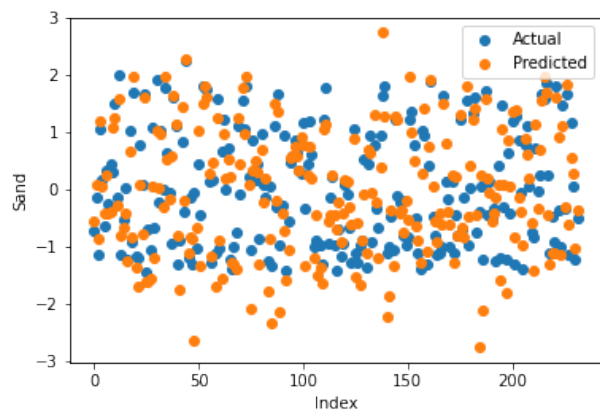


Figure 26: Actual versus predicted values of Sand using stochastic gradient boosting regression

Table 4: Gradient Boosting Coefficient of Determination

Calcium	0.868045
Phosphorous	0.122908
pH	0.741945
Soil Organic Carbon	0.911778
Sand	0.863650

3.6. Discussion

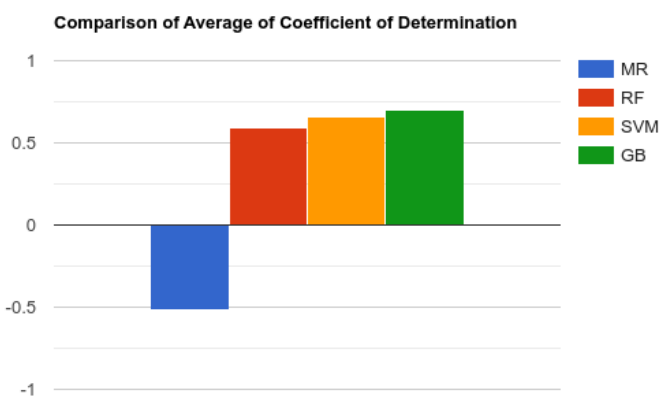
Table 5: Comparative analysis of machine learning model in terms of Coefficient of Determination

Component	MR	RF	SVM	GB
C	0.8096	0.842057	0.7386	0.8680
P	-5.5568	-0.1793	0.1964	0.1229
pH	0.6187	0.6470	0.7468	0.7419
SOC	0.8518	0.8125	0.7830	0.9117
Sand	0.6913	0.8225	0.8404	0.8636

Table 6: Average performance obtained from different machine learning models

Multivariate Regression	-0.5170364
Random Forrest Regression	0.5889594
Support Vector Regression	0.661079
Gradient Boosting	0.701665

Figure 27: Comparison of Average of Coefficient of Determination



In Table 5 MR is Multiple Linear Regression, RF is random forest regression, SVM is Support Vector Machine regression and GB is gradient boosting regression. Here we are comparing the performance of each technique in predicting an individual component in soil.

Table 5 shows the performance comparison of four machine learning models in terms of the Coefficient of determination for every component of soil. The results reveal that gradient boosting performs better than others in terms of C, SOC, and Sand. Table 6 depicts the average performance of machine learning models. It is observed from the table that gradient boosting outperforms the others in terms of the Coefficient of determination.

We find that there is very little correlation of the spectroscopy data and remote sensing data with the with the

amount of phosphorous in the soil, even being negative in case of multiple linear regression and random forest. While both gradient boosting and support vector machines give a very weak positive correlation. Possible ways to fix this problem is either getting more data or trying out deep learning techniques. Gradient boosting outperforms all other methods in Predicting calcium, soil organic carbon and sand, while coming at a close second in predicting pH. In a real world deployment, a hybrid approach can be used for evaluating each component. Also, there's a possibility of creating hybrid models that use consensus of multiple models. This can be done by taking weighted averages of output of different models may mitigate problem of over fitting and improve accuracy.

4. Conclusion

This paper studied the machine learning techniques to predict soil properties for precision agriculture. Four machine learning techniques were used to evaluate the soil properties such as Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand. These techniques were trained and tested on the Africa Soil Property Prediction dataset. It is observed from the results that stochastic gradient boosting performed better than the other techniques. Stochastic gradient boosting was able to predict Phosphorous better than multiple linear regression and Random Forest. Support vector regression was best at predicting the phosphorous component. It can be seen that there is a potential to use spectroscopy as an alternative method of soil component analysis. Deep learning and hybrid models may be used for predicting soil properties in an effective and efficient manner. The main limitation of our study is the use a small number of soil components for prediction. This study can be extended by using a large dataset and other models.

References

- [1] Food, A. O. of United Nations, "India at a glance", <https://www.fao.org/india/fao-in-india/india-at-a-glance/en/>.
- [2] J. v. B. Ephraim Nkonya, Alisher Mirzabaev, *Economics of Land Degradation and Improvement – A Global Assessment for Sustainable Development*, Spinger, 2016.
- [3] B. S. B. Arun Kumar, A. Kumar, "Prospective of Indian agriculture: highly vulnerable to huge unproductivity and unsustainability", *CURRENT SCIENCE*, vol. 119, no. 7, pp. 1079–1080, 2020, doi: <http://dx.doi.org/10.1002/andp.19053221004>.
- [4] P. Yadav, "Agricultural situation in india", 2014.
- [5] M. of Agriculture, "Overuse of fertilizer", <https://pib.gov.in/Pressreleaseshare.aspx?PRID=1696465>.
- [6] I. Barra, S. Haefele, R. Sakrabani, F. Kebede, "Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances -a review", *TrAC Trends in Analytical Chemistry*, vol. 135, 2020, doi:10.1016/j.trac.2020.116166.
- [7] B. Patra, R. Pal, P. Rajamani, P. Surya, "Mineralogical composition and c/n contents in soil and water among betel vineyards of coastal odisha, india", *SN Applied Sciences*, vol. 2, 2020, doi: 10.1007/s42452-020-2631-5.
- [8] I. Barra, S. M. Haefele, R. Sakrabani, F. Kebede, "Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances-a review", *TrAC Trends in Analytical Chemistry*, vol. 135, p. 116166, 2021, doi:<https://doi.org/10.1016/j.trac.2020.116166>.

- [9] H. Yu, D. Liu, G. Chen, B. Wan, S. Wang, B. Yang, "A neural network ensemble method for precision fertilization modeling", *Mathematical and Computer Modelling*, vol. 51, no. 11, pp. 1375–1382, 2010, doi:<https://doi.org/10.1016/j.mcm.2009.10.028>, mathematical and Computer Modelling in Agriculture.
- [10] I. Sydorenko, "Training data", <https://labelyourdata.com/articles/machine-learning-and-training-data>, 2021.
- [11] "Africa soil property prediction challenge", <https://www.kaggle.com/c/afsis-soil-properties/data>.
- [12] S. Ray, "A quick review of machine learning algorithms", "2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)", pp. 35–39, 2019, doi:10.1109/COMITCon.2019.8862451.
- [13] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press, 2 ed., 2009, doi:10.1017/CBO9780511815867.
- [14] X. Yan, *Linear Regression Analysis: Theory and Computing*, World Scientific Publishing Company Pte Limited, 2009.
- [15] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. Mclachlan, A. Ng, B. Liu, P. S. Yu, et al., "Top 10 algorithms in data mining", *Knowledge and Information Systems*, vol. 14, no. 1, p. 1–37, 2007, doi:10.1007/s10115-007-0114-2.
- [16] T. K. Ho, "Random decision forests", "Proceedings of 3rd International Conference on Document Analysis and Recognition", vol. 1, pp. 278–282 vol.1, 1995, doi:10.1109/ICDAR.1995.598994.
- [17] T. K. Ho, "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998, doi:10.1109/34.709601.
- [18] T. Hastie, J. Friedman, R. Tibshirani, *The Elements of statistical learning: data mining, inference, and prediction*, Springer, 2017.
- [19] S. M. Piryonesi, T. E. El-Diraby, "Data analytics in asset management: Cost-effective prediction of the pavement condition index", *Journal of Infrastructure Systems*, vol. 26, no. 1, p. 04019036, 2020, doi:10.1061/(asce)is.1943-555x.0000512.
- [20] C. Cortes, V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, p. 273–297, 1995, doi:10.1007/bf00994018.
- [21] N. J. Nagelkerke, et al., "A note on a general definition of the coefficient of determination", *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.



Vijay Kumar is Assistant Professor in Computer Science and Engineering Department, NIT Hamirpur. He received his Ph.D. degree from NIT Kurukshetra. Previously, he received M.Tech. and B.Tech. degrees from GJUST, Hisar and M.M. Engineering College, Mullana, respectively.

He has 4 years of teaching and research experience in the Thapar Institute of Engineering & Technology, Patiala. Prior, he has 8 years of teaching experience in various reputed institutes. He completed 2 DST SERB sponsored research projects. Presently, he is working on 1 CSIR sponsored research project. He has published more than 100 research papers in International Journals/Conferences.



Jai Singh Malhotra is pursuing bachelors degree from National Institute of Technology in Computer Science and Engineering (expected may 2022).



Saurav Sharma is pursuing bachelors degree from National Institute of Technology in Computer Science and Engineering (expected may 2022).



Parth Bhardwaj is pursuing bachelors degree from National Institute of Technology in Computer Science and Engineering (expected may 2022).

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

An Overview of Solutions Regarding the Problems in Vocational and Technical Education - Example of Elazığ Province

Ceyda Akıllı *, İmam Bakır Arabacı, Engin Kırçıl

Firat University, Educational Management, Institute of Educational Sciences, Elazığ, 23100, Turkey

*Corresponding author: Ceyda AKILLI, Firat University, +905389454300 & ceyda_akilli@hotmail.com

Corresponding author ORCID: 0000-0003-1928-1124

ABSTRACT: Vocational and technical education institutions help to train individuals in a well-equipped manner, increase their employability level, and provide workforce in areas that countries need. Vocational and technical education is of great importance in terms of ensuring the development of countries at the national and international level, training qualified intermediate staff and increasing employment opportunities. Despite the innovations and projects carried out in the vocational and technical education process; educators and students face many problems in the process. The aim of this research is to examine the solution proposals for the problems experienced in vocational and technical education institutions in terms of management, program and application. In the research, the situation analysis design, which is among the qualitative research methods, was used. Focus group interview technique was used in order to examine the subject discussed in the research in detail. A semi-structured interview form was used as a data collection tool. As a result of the research, the main solution to the problems experienced in the management dimension in vocational and technical education institutions is to establish an effective control mechanism, to strengthen the education-employment-production relationship, and to take deterrent measures for the implementation of the Vocational Education Law No. 3308. As a result of the research, solution suggestions for the problems experienced in the dimension of the curriculum in vocational education are to adapt to the new generation teaching methods and techniques, to offer internship opportunities to students abroad, and to have the qualified manpower needed by the domestic and national defense industry. Finally, solution suggestions for the problems experienced in the implementation phase in vocational and technical education institutions are to make school and field preferences in certain time periods during the student placement processes, to limit these processes and to allocate sufficient quotas to the relevant institutions.

KEYWORDS: Solution Proposals, Technical Education, Vocational Education

1. Introduction

In today's world, the ability of countries to develop in economic, social, cultural and technological dimensions is directly proportional to their well-equipped and qualified human resources. It can be said that the main factor that determines the training of human resources in a well-equipped manner, providing a competitive environment and reducing unemployment is the importance given to vocational and technical education institutions. Vocational and technical education institutions help to

train individuals in a well-equipped manner, increase their employability level, and provide workforce in areas that countries need. Labor force is the part of the social population that has economic activity [1]. It is among the main duties of vocational and technical education to raise a sufficient number and quality of labor force in every field that our country needs [2]. Vocational education is the acquisition of knowledge, skills and application abilities required by a profession with social validity in order for individuals to continue their lives [3]. Vocational education is the process of reconciling

individuals' cognitive, affective, social, economic and personal development with employability skills [4]. Vocational education can be defined as "an education process consisting of a balanced coordination of individual, profession and education parameters" [5]; on the other hand, it can also be seen as "using the profession as a tool while developing the skills of the individual and increasing the knowledge and skills in this process" [6].

Countries provide the workforce they need through vocational education and update their vocational education systems in line with changing needs [7]. The requirements of the period in this update process are shaped within the framework of technological developments and informational transformations [8]. While developed countries proactively change their vocational education programs in accordance with the needs of the age; it is seen that underdeveloped or developing countries do not realize this change process [9]. The main objectives of vocational technical education are; to educate and train individuals as qualified human resources for many employment areas, especially industry, trade and service sectors, and to provide the basic education necessary for transition to higher education institutions, which are the continuation of their vocational branches [10]. It is aimed to specialize individuals in their own branches by carrying out theoretical and practical training in vocational and technical education institutions [11].

Turkey is among the countries with a young population with a high workforce capacity. One of the most important conditions for our country to benefit from the demographic opportunity it has; young people who will join the workforce must go through a training process where they can have the necessary equipment to be employed in a way that can come to the fore in national and international competitive markets [12]. Vocational education institutions undertake the responsibility of raising qualified, equipped, entrepreneurial and productive generations that can be employed in the sectors needed in today's world [13]. Education programs of institutions providing vocational and technical education are structured in line with the needs of the sectors and 21st century skills [11]. The most important issues to be considered in terms of institutions providing vocational and technical education are; it is the practical implementation of enriched trainings in the light of sectoral needs with on-the-job learning methods and the provision of international equivalence of diplomas [14]. Vocational and technical education needs to be redesigned and redesigned with a dynamic structure in

order to achieve the aim of raising qualified personnel who can contribute to production [13]. For this reason, the need for educational reforms to update vocational and technical education processes is increasing [15]. Many researches are carried out, policies are formed and projects are carried out in order to ensure the quality in vocational and technical education, to establish the technological infrastructure and to ensure modernization. In the recent past, the most important decisions for the improvement of vocational and technical education given in formal education institutions were taken at the 12th, 15th and 16th National Education Councils, and at the 15th National Education Council convened in 1996, it was recommended to update the education systems with a focus on vocational and technical education. In our country, which is aware of the necessity of having well-equipped manpower for economic development, the search continues to bring the quality of vocational and technical education to the next level, and many projects are carried out in this context [16]. Among these projects are the "School-Industry Joint (OSANOR) Project" that supports sectoral collaborations, the "Strengthening the Vocational Education and Training System (MEGEP) Project" that encourages vocational education to a systemic innovation, and the Erasmus+ projects that enable international learning mobility. The Vocational Education Law No. 3308 was established to regulate the principles regarding the training of apprentices, journeymen and masters, and vocational training to be carried out in schools and businesses. Within the scope of this Law, provincial vocational education boards are established in provinces in order to take decisions on the planning, development and evaluation of vocational education to be carried out in vocational and technical education schools, institutions and enterprises, also to provide opinions and recommendations to the governorship. The main purpose of the law is to transform vocational and technical education into a modern process enriched with technology; also to improve the personal rights of vocational education personnel and students.

Despite the innovations and projects carried out in the vocational and technical education process; Educators and students face many problems in vocational and technical education activities carried out in our country. First of all, it has been determined by many experts that there are negative prejudices against vocational education in our country [17]. The main reason for this situation is the inadequacy of vocational training and consultancy services [18]. In addition, there is a conflict between general education and vocational education due to social

and environmental factors, and families do not prefer to have their students study in vocational education institutions. The reason for this situation may be negative perceptions towards vocational education, not regular transitions between school levels, diversity in high school types, and insufficient emphasis on culture courses in vocational education. The high costs of vocational education and the inflexibility of education programs are among the obstacles to vocational and technical education [18]. The disconnection between vocational education institutions and sectors, the lack of practical training and the inability to carry out the apprenticeship system in a functional way are among the problem areas [19]. The fact that less talented students are admitted to vocational education institutions compared to general high schools and that schools cannot follow the technology used in the industry reduces the preferability of vocational and technical education institutions [20].

The aim of this research is to examine the solution proposals for the problems experienced in vocational and technical education institutions at the formal education level in terms of management, program and application. The research questions created in this context are as follows:

1. What are the solutions for the problems experienced in the management dimension in vocational and technical education institutions?
2. What are the solutions for the problems experienced in the dimension of the curriculum in vocational and technical education institutions?
3. What are the solutions for the problems experienced in the application dimension in vocational and technical education institutions?

2. Method

In the method section, there is information about the model of the research, the data collection tool and the study group.

2.1. Model of the Research

In the research, the situation analysis design, which is among the qualitative research methods, was used. The most basic feature of the situation analysis design is the in-depth investigation of one or more situations [21]. In this context, all factors related to a situation are handled in a holistic approach and how certain factors are affected by the relevant situation. Situation analysis is an empirical research method used in situations where the boundaries between the phenomenon and the context it is attached to are not clear and more than one source of evidence or data is available [22]. Situation analysis is a

qualitative research approach in which the researcher examines one or more situations in depth with the help of data collection tools and defines situations and themes related to the situation [23]. The research is based on the cross-sectional survey model. Survey research, which is widely used in social sciences, is research that includes the opinions and attitudes of individuals about a phenomenon and event [24]. Focus group interview technique was used in order to examine the subject discussed in the research in detail. Focus group interviews are a data collection method frequently used in the field of social sciences [25]. Its practicality and usefulness are factors that are effective in the frequent use of focus group interviews [26]. The main element of focus group interviews is that they consist of few participants and participant responses are affected by group dynamics [27].

2.2. Data Collection Tool

A semi-structured interview form was used as a data collection tool. The form was developed by the researcher with the help of an expert. The most important convenience provided by the semi-structured interview technique to the researcher is that it provides more systematic and comparable information since the interview is carried out in accordance with the pre-prepared interview protocol [21]. In this research, a 3-item interview form focusing on the problems experienced in formal education institutions serving within the scope of vocational and technical education and the solution proposals developed for these problems was used. Interviewing is used as the shortest way to learn the knowledge, thoughts, attitudes and behaviors of individuals on various issues and their possible causes [28]. The main purpose of using an interview technique is not usually to test a hypothesis; it is trying to understand other people's experiences and how they make sense of these experiences [29]. The interview technique used in qualitative research is to use the perspectives of the people researched, to reveal their world of meaning, to see the world through their eyes [24].

2.3. Study Group

The study group of the research consists of a focus group of 20 people, consisting of school principals, workshop heads and teachers working in vocational and technical education institutions in the province of Elazığ. The study group of the research was formed based on the non-probability maximum variation method. The aim here is to reflect the sample diversity created in a way that addresses the research questions

in a multi-faceted manner. The aim is not to provide diversity to generalize; it is to try to find out whether there are common facts or differences among various situations and to reveal different dimensions of the problem according to diversity [21]. The demographic distributions of the participants in the study group are as seen in Table 1.

Table 1: Demographic data of the study group

Variables	Frequency (f)	Percentile (%)
Title		
School Principal	5	25
Studio chief	10	50
Teacher	5	25
Seniority Year		
1- 5	2	10
6- 10	5	25
11- 15	11	55
16 and above	2	10
Age		
23-30	3	15
31- 40	6	30
41-50	7	35
51- 60	3	15
61 and above	1	5
Total	20	100

As seen in Table 1, the demographic characteristics of the participants were examined in terms of title, seniority and age variables. In the title variable, 50% of the participants are workshop chiefs. The main reason for this situation is that the workshop chiefs are at the forefront in the implementation phase, they are specialized in their fields and their sectoral cooperation is strong. 25% of the participants are school principals working in vocational and technical education institutions and 25% are technical teachers working in these institutions.

In the variable of seniority years, the majority of the participants consisted of educators with a seniority between 11 and 15 years ($f=11$). This shows that the majority of the participants have served long enough to specialize in their fields. In this way, it is thought that vocational and technical education institutions are competent in knowing their problems closely and producing solutions to these problems. 25% of the participants have 6-10 years, 10% have 1-5 years, 10% have 16 years or more years of service.

When examined in terms of the age variable, the majority of the participants are in the 41-50 age category ($f=7$). 30% of the participants are in the age range of 31-40, 15% are in the age range of 23-30, 15% are in the age range of 51-50. There are 5 participants aged 61 and over in the study group. When the age

categories of the participants are examined, it is seen that they are generally in the middle age group.

2.4. Data Analysis

Research data were interpreted in line with descriptive analysis. Interpretation is the process of giving meaning to the analyzed data [28]. Descriptive analysis is carried out by transferring the answers of different individuals to the questions posed within the scope of the research, without changing them, in the form of quotations [21]. Descriptive analysis; it consists of the processes of establishing the theoretical framework, analyzing the data according to this framework, defining the findings and interpreting the findings [30]. While analyzing the data, in order to ensure confidentiality and impartiality, the participants in the focus group interview were given codes as K-1, K-2, ..etc. The answers given by the working group to the questions posed in the focus group interview were recorded and resolved, and analyzed by categorizing them according to the problem areas. These categories were ordered according to the number of citations of the participants, and percentiles and frequencies were determined. Problem areas were strengthened by showing striking citations from the participants as examples. During the data analysis process, categories and patterns were taken into consideration, the frequency of the variables was interpreted, the relationships between the variables were taken into account, and the variables were grouped according to their characteristics. The answers have been interpreted in a way that reflects the reality as it is, by establishing logical meaning chains. Because the most important problem of qualitative research is whether the meanings and results reached are correct, valid and repeatable [31].

3. Findings

The research findings were examined in terms of management, curriculum, and application dimensions.

3.1. Suggestions Regarding the Problems Experienced in the Dimension of Management in Vocational and Technical Education Institutions

The first question posed to the participants in the focus group interview was "What are your suggestions for the management problems in vocational and technical education institutions?" has been. The categories created for the answers given by the participants are given in Table 2.

Table 2: Suggestions regarding the problems experienced in the management dimension in vocational and technical education institutions

Category	Frequency (f)	Percentile (%)
1. Establishment of an effective control mechanism related to vocational technical education.	10	25,65
2. Strengthening the education-employment-production relation in vocational and technical education.	8	20,52
3. Taking the necessary deterrent measures for the implementation of the Vocational Education Law No. 3308.	5	12,83
4. Carrying out incentive activities in order to increase the value attributed to vocational and technical education.	5	12,83
5. Improving the institutional capacity by preparing projects in cooperation with the funding institution/organization.	4	10,25
6. Development of educational environments and human resources.	2	5,12
7. Selection of the branch managers responsible for the vocational and technical education branch in the provincial national education directorates from among the technical teachers.	2	5,12
8. Student transfers are not weekly; scheduling monthly.	1	2,26
9. Developing the professional competence of teachers.	1	2,56
10. Carrying out necessary studies for vocational high schools to issue Vocational Qualification Certificates.	1	2,56
Total	39	100

As can be seen in Table 2, the majority of the participants suggest the establishment of an effective control mechanism for the problems experienced in the management dimension of the vocational and technical education process (f=10). Auditing is a process of examining whether organizational activities are carried out in accordance with predetermined objectives. Participants state that the teaching activities carried out in vocational and technical education institutions are not effectively supervised and this situation creates an authority gap.

“The biggest problem of vocational education is the control gap. The fact that the inspectors are far from the technical field further deepens the problem. Managers need to take steps in this regard.” K-12

As the participant with the code K-12 stated, the fact that individuals who have command of the field in the audit processes of vocational and technical education institutions and manage the process effectively will be able to reveal the problems experienced in the institutions and the activities carried out in an objective way.

Another solution proposal stated by the participants in the focus group meeting for the problems experienced in the management dimension is to strengthen the education-employment-production relationship (f=8). Stating that the most important elements of vocational and technical education are student employment and contributing to national production, the participants state that the organic bond between these elements should be strengthened.

“The main purpose of vocational and technical education is to train human resources, to employ them and to ensure sustainability in production. Unfortunately, I think that employment and production dimensions are ignored in vocational education. At this point, managers should assume the biggest role and build a bridge between sectors.” K-3

As the K-3 coded participant stated, the main purpose of vocational and technical education organizations is to increase the employment power of individuals, to train needed intermediate staff and to support national production. The main way to achieve organizational goals is through effective leadership and management processes. It is recommended to take steps to strengthen the employment and production dimensions of institutions so that vocational and technical education can achieve its organizational goals.

Participants suggested that necessary measures be taken for the implementation of the Vocational Education Law No. 3308 in order to prevent the problems experienced in the management dimension (f=5).

“The most important power of vocational high schools is the law numbered 3308. Unfortunately, the implementation phase is unsuccessful. Deterrent measures should be taken for the implementation of the articles in the law. K-9

Vocational Education Law No. 3308; it aims to regulate the principles regarding the training of

apprentices, journeymen and masters, and vocational training in schools and businesses. It is thought that the problems experienced in vocational and technical education institutions will be greatly reduced with the implementation of the Law No. 3308, which determines the principles regarding vocational education and regulates the practices. The implementation and follow-up of the articles in the law should be carried out by the managers.

Participants stated that necessary studies should be carried out (f=5) to change the perceptions of stakeholders regarding vocational and technical education institutions in a positive way and that institutional capacity should be increased by carrying out project-based studies (f=4). They stated that it can be prevented.

3.2. Suggestions Regarding the Problems Experienced in the Dimension of Curriculum in Vocational and Technical Education Institutions

Within the scope of the research, the participants were asked, "What are your suggestions for the problems experienced in the scope of the curriculum in vocational and technical education institutions?" question was posed. The categories created for the answers given by the participants are given in Table 3.

Table 3: Suggestions regarding the problems experienced in the dimension of teaching programs in vocational and technical education institutions

Category	Frequency (f)	Percentile (%)
1. Adapting to new generation teaching methods and techniques.	14	24,56
2. Providing students with internship opportunities abroad.	10	17,54
3. Having the qualified manpower needed by the domestic and national defense industry.	10	17,54
4. To train professional staff needed by business people who invest abroad.	9	15,78
5. Coordinator teachers spend more time with students in appropriate enterprises while vocational training planning is carried out in enterprises.	5	8,77
6. Subjecting vocational teachers to practical training by individuals from the sector regarding their own fields.	4	7,05
7. Improving the content of applied training hours in vocational technical education institutions.	3	5,26

8. Increasing access to guidance in vocational and technical education.	2	3,50
Total	57	100

Participants mostly stated that "adaptation to new generation teaching methods and techniques" is required as a solution proposal for the problems experienced in the dimension of teaching programs in vocational and technical education institutions (f=14).

"The subjects and practices taught in vocational high schools and vocational training centers should be a new generation. Every school should have a 3D printer, but because the budget is not enough, outdated education is given in many schools." K-1

The participant coded K-1 stated that new generation teaching methods should be used, but the institutional infrastructure required for this is not sufficient. Vocational and technical education centers are responsible for training human resources in line with the needs of the age. In order to train students with 21st century skills in accordance with the professions of the future, a contemporary education program should be adopted and practical training should be carried out in this direction. However, due to problems such as technological infrastructure, lack of budget and institutional capacity, the trainings cannot achieve the desired effect. As a solution to this situation, it is suggested that managers can provide funds to their institutions through projects, make use of simulation applications and eliminate existing deficiencies through technological opportunities.

"The curriculum is old and inadequate. In cases where the budget cannot be provided, applied trainings can be carried out through the internet and technology." K-15

Participants think that providing students with internship opportunities abroad (f=10) and having the qualified manpower needed by the domestic and national defense industry (f=10) can be a solution to the problems experienced in the dimension of the curriculum.

"The internship abroad of the students studying in vocational education institutions develops their vision, enables them to specialize in their professions, and improves their self-confidence. In this way, negative

opinions towards vocational high schools can be eliminated.” K-7

“Great strides are being made in the field of defense industry in our country. In order for us to be able to provide support, we need to specialize in areas such as UAV and SIHA.”

K-13

As stated in K-7 and K-13 coded participants, vocational and technical education institutions should be utilized effectively in order for our country to develop nationally and internationally and to strengthen economic development. Curriculum should be updated in line with new generation needs and have an enriched content. Participants also suggested that the curriculum should be renewed in line with the training of professional staff needed by business people who invest abroad (f=9) and that coordinator teachers should spend more time with students (f=5).

3.3. Suggestions Regarding the Problems Experienced in the Application Dimension in Vocational and Technical Education Institutions

The third question posed to the participants in the focus group meeting was “What are your suggestions for the problems experienced in practice in vocational and technical education institutions?” has been. The categories created for the answers given by the participants are given in Table 4.

Table 4: Suggestions regarding the problems experienced in the application dimension in vocational and technical education institutions

Category	Frequency (f)	Percentile (%)
1. When placing students in the 9th grades, school and field preferences are taken at the same time and the student is placed in an early period.	8	21,62
2. Limitation of field and branch change operations.	6	16,21
3. Allocating sufficient quotas to vocational technical education schools.	5	13,51
4. Ensuring the planning of private vocational high schools in a way that does not allow the activities of official vocational high schools to be restricted.	5	13,51
5. Limiting the placement of students subject to mainstreaming to vocational and technical education schools, taking into account the student's disability.	4	10,81

6. The opening of vocational secondary schools.	3	8,10
7. Vocational education centers and vocational technical education schools do not graduate with the same diploma.	2	5,40
8. Providing vocational education guidance in the 8th grades of secondary schools.	2	5,40
9. Ensuring the implementation of the protocol articles in the works and transactions related to the project schools established within the framework of the protocol signed between TOBB University of Economics and MEB.	1	2,70
10. Ensuring the establishment of an Agriculture and Livestock Vocational High School in each province.	1	2,70
Total	37	100

As can be seen in Table 4, the majority of the participants stated that while placing students in the 9th grades as a solution to the problems experienced in the vocational and training centers, the choice of school and field should be taken at the same time and the placement of the student should be ensured in the early period (f=8).

“The choice of school and field should be made at the same time. We are late in choosing a field and this causes the child to specialize in his branch late” K-20

As can be understood from the participant's views, making field choices late may cause students to have problems with their task specialization, and this situation poses an obstacle for students in the application phase. The participants also stated that field and branch changes should not be made continuously and that it is essential to limit this issue (f=5).

“Until the end of the 1st term of the 9th grade of the field change procedures; branch change procedures must be carried out until the end of the 1st semester of the 10th grade. Thus, arbitrary transactions are prevented.”

K-14

Participants also stated that sufficient quotas should be allocated to vocational technical education schools (f=5) and that plans for private vocational high schools should be provided in a way that would not allow the activities of official vocational high schools to be restricted (f=5).

“The quotas of secondary education institutions providing general academic education should be limited and sufficient

quotas should be allocated to vocational and technical education schools.” K-10

“While planning for private vocational high schools, taking the needs of the province into account, giving the necessary opening approvals and ensuring that these schools are opened in a way that does not allow the activities of official vocational high schools to be restricted is the most valuable solution to prevent the problems experienced during the implementation phase.” K-19

As can be understood from the participant statements above, the participants think that vocational and technical education institutions do not have sufficient quotas and that private vocational high schools may restrict the activities of official vocational high schools. As a solution to this situation, they stated that quota plans should be made during the implementation phase and local needs should not be ignored.

4. Conclusions and Discussion

Vocational and technical education is of great importance in terms of ensuring the development of countries at national and international level, strengthening the competitive market, training qualified intermediate staff and increasing employment opportunities. However, it is known that, as in every organization, there are some problems in vocational and technical education institutions [32,33]. At the point of solving problems; Suggestions produced by individuals working in relevant institutions and experiencing the problems personally should be taken into account.

As a result of the research, the main solution to the problems experienced in the management dimension in vocational and technical education institutions is the establishment of an effective control mechanism. Auditing is a reflective management process that enables organizations to achieve their goals and creates a sense of trust among stakeholders [34]. In this context, it is very important to establish a strong control mechanism at the stage of examining the achievement of the objectives of vocational and technical education institutions. Strengthening the education-employment-production relationship in vocational and technical education is another suggestion that can prevent the problems experienced in the management dimension. In the management activities carried out in vocational and technical education institutions; it is necessary to pay attention to what the job market expects from vocational education and to focus on domestic production [35].

Another element of solving the problems experienced in vocational and technical education institutions in terms of management is to take deterrent measures for the implementation of the Vocational Education Law No. 3308. Law no. 3308 is a turning point in terms of vocational and technical education [16]. The proper implementation of this law is primarily the primary duty of individuals working at the management level. Incentive activities to be carried out by the managers at the point of preventing the negative prejudices that individuals attribute to vocational education are another measure that can prevent the problems experienced in vocational education.

Within the scope of the research, solution proposals for the problems experienced in the dimension of teaching programs in vocational education were examined. Adapting to new generation teaching methods and techniques, providing students with internship opportunities abroad, and having the qualified manpower needed by the domestic and national defense industry are among the solutions suggested by the participants. Highlighting the technical dimension in vocational education and the fact that teaching methods and techniques are not up-to-date are the biggest obstacles to the success of these schools [36]. In order to improve the skills of students studying in vocational and technical education institutions, modern teaching methods should be used, and they should be introduced to educational activities carried out abroad and existing professions [37].

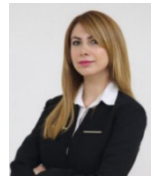
The opinions of the participants on the solution proposals for the problems experienced in the implementation phase in vocational and technical education institutions were examined. In the student placement processes, making school and field preferences in certain time periods, limiting these processes and allocating sufficient quotas to the relevant institutions are among the suggestions made by the participants. The most important problem in front of the applications carried out in vocational education is the decrease in the tendency of students to prefer these schools [38]. In this context, directing students to vocational education and making school and branch changes in a systematic and limited way constitute a solution to the problems encountered during the implementation phase.

References

- [1] C. Bekiroğlu, “Türkiye’de İşsizlik Sorununun Çözülmesinde Uygulanan Ekonomi Politikalarının Analizi,” (Master Thesis, Kadir Has University, 2010).

- [2] M. Çelik et al., "Sınavsız Geçişin Karamanoğlu Mehmetbey Üniversitesi Bünyesindeki Meslek Yüksekokullarındaki Eğitim, Öğrenci ve Personel Kalitesine Etkisi Üzerine Bir Araştırma" *Sözlü Bildiri, Uluslararası Türkiye Eğitim Araştırmaları Kongresi*, Mayıs 2009.
- [3] İ. Egin, *Mesleki ve teknik eğitimde program geliştirme*, Nobel Yayınları, Ankara, 2000.
- [4] M. Donnelly, *Vocational education*, EBSCO Research Starters EBSCO Publishing Inc, 2008.
- [5] H. M. Süer, "Ticaret Meslek Liselerinde Muhasebe Eğitimi, Muhasebe Eğitiminden Beklentiler ve Karşılaşılan Sorunlar," (Master Thesis, İstanbul Üniversitesi, 2007).
- [6] İ. Sezgin, "Yeni mesleki-teknik eğitim sistemi ve okul-endüstri ilişkileri," *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, vol. 2, pp. 210-228, 1987.
- [7] F. Koç, "Mesleki eğitimde okul sanayi iş birliği," *Milli Eğitim Dergisi*, vol. 141, no. 5, pp.64-66, 1999.
- [8] İ. H. Bülbül, "Mesleki ve teknik eğitimde öğretim stratejileri ve yeni teknolojilere entegrasyonu," *16. Milli Eğitim Şurası Hazırlık Dokümanı*. Ankara, pp. 27-32, 1998.
- [9] A. Gür, "Rekabet Gücünün Artırılmasında Nitelikli İşgücü İhtiyacı ve Mesleki Eğitim," (Ph. D. Thesis, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 2011).
- [10] İ. Eşme, "Konferans Açılış Bildirgesi Sunumu," Uluslararası Mesleki ve Teknik Eğitim Konferansı, YÖK, Ankara, 2007.
- [11] İ. E. Göktürk et al., "Sosyal bilimler meslek yüksekokullarının eğitim sürecinde; uygulama açısından karşılaşılan sorunlar ve çözüm önerileri," *Electronic Journal of Vocational Colleges*. Aralık UMYOS Özel Sayı, 2013.
- [12] İ. Aktaşlı et al., "Mesleki ve Teknik Eğitimde Güncellenmiş Durum Analizi," *Eğitim Reformu Girişimi Sözlü Bildiri*, İstanbul, Nisan 2012.
- [13] V. Özcan, "Türkiye'de Mesleki ve Teknik Eğitim ile İstihdam İlişkisi," (Master Thesis, Gediz Üniversitesi, 2014).
- [14] N. Şener, E. Torun, "Meslek Yüksekokullarında İstihdam Sorunu," *Sözlü Bildiri*. Mayıs 2009. 1.Ulusal arası 5. Ulusal Meslek Yüksek Okulları Sempozyumu, Konya, 2009.
- [15] M. Fariás, M.P. Sevilla, "Effectiveness of vocational high schools in students' access to and persistence in postsecondary vocational education," *Res High Educ*, vol.56, pp. 693-718, 2015, doi:10.1007/s11162-015-9370-2.
- [16] O. C. Adıgüzel, Ş. Berk, "Mesleki ve teknik ortaöğretimde yeni arayışlar: Yeterliğe dayalı modüler sistemin değerlendirilmesi," *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, vol. 4, no.1, pp. 220-236, 2009.
- [17] H. Binici, N. Arı, "Mesleki ve teknik eğitimde arayışlar," *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, vol. 24, no.3, pp. 383-396, 2004.
- [18] A. Şimşek, *Türkiye'de mesleki ve teknik eğitimin yeniden yapılandırılması*. TÜSİAD Lebib Yalkın Yayınları, İstanbul, 1999.
- [19] A Barabasch, S. Petrick, S. "Multi-level policy transfer in Turkey and its impact on the development of the vocational education and training (VET) sector," *Globalisation, Societies and Education*, vol. 10, no.1, pp. 119-143, 2012, doi: 10.1080/14767724.2012.646904.
- [20] M. Sönmez, "Türkiye'de mesleki ve teknik örgün öğretimin sorunları ve yeniden yapılandırılma zorunluluğu," *Eğitim ve Bilim*, vol. 33, no. 147, pp. 71-84, 2008.
- [21] A. Yıldırım, H. Şimşek, *Sosyal bilimlerde nitel araştırma yöntemleri*, Seçkin Yayıncılık, Ankara, 2013.
- [22] R. Yin, *The case study research, design and methods*, Sage Publications, Newbury Park, 1994.
- [23] J. W. Creswell, *Qualitative inquiry and research design: Choosing among five approaches*, Sage Publications, Thousand Oaks, 2010.
- [24] A. Tanrıoğen, *Bilimsel araştırma yöntemleri*, Anı Yayıncılık, Ankara, 2012.
- [25] G. Kamberelis, G. Dimitriadis, *Focus groups: Strategic articulations of pedagogy, politics, and inquiry*, Sage Publications, Thousand Oaks, 2005.
- [26] M. Savin-Baden, C. H. Major, *Qualitative research: The essential guide to theory and practice*. Taylor and Francis Group, London, Routledge, 2013.
- [27] D. L. Morgan, R. A. Krueger, *When to use focus groups and why*, Sage Publications, Newbury Park, 1993.
- [28] N. Karasar, *Bilimsel araştırma yöntemi*, Nobel Yayınevi, Ankara, 2012.
- [29] A. Türnüklü, "Eğitimbilim araştırmalarında etkin olarak kullanılabilir nitel bir araştırma örneği: Görüşme," *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, vol. 24, no.24, pp.543-559, 2000.
- [30] R. Altunışık et al., *Sosyal bilimlerde araştırma yöntemleri*, Sakarya Kitapevi, Sakarya, 2010.
- [31] B. Miles, A. Hubberman, *Qualitative data analysis*, Sage Publications, London, 1994.
- [32] H. İşler, "Avrupa Birliği'ne Uyum Sürecinde Türkiye'deki Mesleki ve Teknik Ortaöğretim Sisteminin Avrupa Birliği'ne Uygunluğu," (Ph. D. Thesis, Dokuz Eylül Üniversitesi, 2006).
- [33] S. Önal, "3308 Sayılı Yasa Kapsamında İşletmelerde Yapılan Meslek Eğitimi Uygulamalarının Değerlendirilmesi-Endüstri Meslek Lisesi Örneği," (Master Thesis, Ankara Üniversitesi, 1997).
- [34] J. K. A. Card, "Secondary Teacher's Attitudes Toward Important and Existing Classroom Observation Practices That Promote Teacher Effectiveness," (Master's Thesis, Dowling College, 2006).
- [35] T. Karmel, "Vocational education and training in Australian schools," *The Australian Educational Researcher*, vol. 34, no. 3, pp.101-117, 2007.
- [36] K. Clarke, J. Polesel, "Strong on retention, weak on outcomes: the impact of vocational education and training in schools," *Discourse: Studies in the Cultural Politics of Education*, vol. 34, no.2, pp. 259-273, 2013, doi: 10.1080/01596306.2013.770251.
- [37] M. Klatt, "Understanding youth transition system in Poland through the analysis of partnerships between vocational upper secondary schools and industry," *Journal of Youth Studies*, vol. 18, no. 9, pp. 1186-1203, 2015, doi: 10.1080/13676261.2015.1020938.
- [38] C. Erden Özsoy, "Mesleki eğitim - istihdam ilişkisi: Türkiye'de mesleki eğitimin kalite ve kantitesi üzerine düşünceler," *Electronic Journal of Vocational Colleges- Aralık 4. UMYOS Özel Sayısı*, pp.173- 181, 2015.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Dr. Ceyda AKILLI completed her master's degree in Educational Administration at the Department of Firat University in 2017 and completed her doctorate in the same field in 2022. She works as an expert in European Union projects.



Prof. Dr. İmam Bakır ARABACI works as a professor in the field of Educational Administration at Firat University. He has studies in many areas such as total quality management and strategic planning.



Engin KIRÇIL completed his master's degree in educational administration. He served as director and assistant director. He works in the field of projects.

An Overview on Various Techniques used for Correct Interpretation of Roadway Symbols

Abhinav Vinod Deshpande *

School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu (T.N.), India-632014

*Corresponding author: Abhinav Vinod Deshpande, Email: avd.a.deshpande@gmail.com

ABSTRACT: In this study paper, a comparative analysis of various image enhancements in local domain techniques was made based on three-dimensional image quality statistics namely Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Image Quality Model (SSIM) to determine image quality. especially among them so that the process of image enhancement in the local domain is carried out. The conditions for choosing the best method are to have a small number of Mean Squared Error (MSE) and a high number of Peak Signal to Noise Ratio (PSNR) and Structural Properties to measure Image Quality (SSIM).

KEYWORDS: Traffic Sign Recognition, Image Denoising, Filtering

1. Introduction

Car companies are booming due to the increase in private car production and growth in their consumption. Adding to the increase in competitors also encourages the increase in the number of companies. It will simplify the process of thinking of other novel methods using road identification. It helps to provide a safe driving environment [1-5]. The speed of a moving vehicle can be determined by using road signs that set the speed limit. Its speed can be controlled by using them. They find use that includes places that use good travel. Many road signs cover these areas [6 - 10]. When driving, the driver usually ignores the speed limit. This conveys the message of the speed of the car. The chances of an accident can be reduced and thus ensure the safety of the driver. Now, the time has come to design an architecture that automatically identifies Speed Limit symbols. It also warns him to stop the car [11]-[15]. The Speed Limit signal was performed following the test. Candidates entitled Maternal Districts are found in the list. The marker is categorized by identifying the regions of interest [16]-[20]. Colorful and common symptoms can be seen in order to fulfill a motive [21]-[25].

1.1. Sign's interpretation methods

The original problem was divided into a problem category corresponding to what were two coded sections in the ECOC matrix. Class dividers are assigned to all real

classes using these methods. Considering the research work found in [26 - 30], the color version under discretization of the visible spectrum was compared with other markers. Discriminatory regional groups were considered under different categories. Using standard street map templates, they were taught in offline mode thus using a one-size-fits-all principal to maximize diversity [1-5]. Color range adjustment was used to compare the actual color of the road sign that was differentiated and thus became more robust thus laying the basis for defining the level of variation [6 - 10]. Assessing the actual performance is very difficult in this case [11 - 15]. Performance testing of multiple machine learning methods is not possible due to the lack of a suitable website or strategy [16 - 20]. Several attempts have been made to identify road signs despite the inclusion of manuscripts [21 - 25]. Rapid growth is reflected in field-based research work using road signs [26 - 30]. Research conducted in this field in the light of the past few decades clearly shows the concepts. Different types of ideas and new strategies are used [1-5]. The road signal circuits that are most commonly used are the acquisition module [6 - 10]. The viewing module offers different classes and shows a given road sign category [11 - 15].

2. Literature Survey

It is a difficult task when it comes to comparing actual performance as there is no such type of validated data or

system to test the performance of a few machine learning methods that try to find a solution to the problem of road signs, although there is literature [16 - 20]. locating and monitoring road signs for the last few decades. Numerous novel ideas and practical methods have been suggested [21 - 25]. Typically, the acquisition component seeks out potential traffic signal circuits while the phase in which the traffic signal is a part is determined by the awareness component. The traditional methods used for the purpose of obtaining sign boards can be divided into three main classes, namely, color-based methods, shape-based methods and smooth window-based methods. One might consider the process of recognizing the image of a particular sign among all the other members, the function of Herculean when it comes to seeing the image of a sign board in a group of other sign boards. It has been found that a variety of techniques are available in the field of image classification of well-known symbols from common methods such as template matching in complex machine learning techniques. One of the most important and critical algorithms used to perform the task of separating multiple image board images can be attributed to the well-known Vector Support Machine (SVM) algorithm. When authorized transcripts [26 - 30] are considered, one can obtain a value attached to the automatic acquisition and display of the signature board that supports Support Vector Machines (SVMs) in combination with Gaussian kernels. However, the system was required to separate candidate blogs in the status category before recognition. As a sequence, only the pixels that were part of the symbol were used to form the element vector. In [1-5], alternatives to those currently available in the context of automatically detecting and seeing signal boards have been made public. It can be misunderstood that in this study the emphasis was on improving the accuracy of the most common methods of automatically detecting and identifying signals and thus leading to a reduction in the number of vectored entries that appear to be most useful. and the duration of testing the latest templates [6-10]. The SVM method of distinguishing traffic signal recognition was provided [11-15] while in [16 - 20] an effective strategy that facilitates the process of detecting sloping speed signals by removing rotating elements with the help of the Fourier adjective. -based wave launch. The different categories of sign board images were divided with the help of Support Vector Machines (SVMs) which combines the formation of binary trees. In [21 - 25], structural-based classification was constructed using SVM. Representing the features, the two types of features include an unlimited framework

and a temporary production proposed by some scientists. One can conclude that conducting a comprehensive assessment process satisfies our main goal which involves specific processes that include the identification of road signs that fall into different categories including multi-layer geometric patterns and controls of road signs that reflect control. speed of moving vehicles in a safe and healthy environment. One might consider the process of recognizing the image of a particular sign among all the other members, the function of Herculean when it comes to seeing the image of a sign board in a group of other sign boards. It has been found that a variety of techniques are available in the field of image classification of well-known symbols from common methods such as template matching in complex machine learning techniques. One of the most important and critical algorithms used to perform the task of separating multiple image board images can be attributed to the well-known Vector Support Machine (SVM) algorithm.

3. Driver Fatigue and Driver Mental Workload

Driver fatigue is a major cause of road accidents. Fatigue is also a structure that includes factors such as day time, waking time, activity duration and monotony, and safety-related outcomes [26 - 30]. Fatigue can be caused by drowsiness (drowsiness), boredom, and mental or physical fatigue. From these exciting features, drowsiness is considered the most important factor in fatigue when used in driving situations. Drowsiness of drivers has contributed to road accidents between professionals [1-5] and the average number of driving [6 - 10]. Accidents caused by the driver's drowsiness can have the same death rate as alcohol-related accidents [11 - 15]. Many factors contribute to sleep, such as long working hours [16 - 20], lack of sleep [21 - 25] and medical conditions [26 - 30]. Insomnia is very common in some communities, including junior doctors [1-5], underwater sailors [6 - 10] and 'miners' flight [11-15]. Chronic sleep deprivation is a known risk factor in driving [16 - 20]. It was also confirmed that the 24 h circadian rhythm was marked with peaks and holes in the levels of awareness as evidence of studies that included both independent sleep measures and objectives [21 - 25]. Work-related factors also contribute to driver drowsiness [26 - 30]. These factors may include driving duration [1-5] and monotony [6 - 10] as experience in highway driving [11 - 15]. The effects of drowsiness are reflected in the reduced number of nutritional warnings [16 - 20]. In driving conditions this leads to significant changes in driver performance such as reduced speed, distance between vehicles and road

maintenance [21 - 25] all increasing the risk of road accidents [26 - 30]. With growing evidence linking driver drowsiness to road accident accidents, the industry has responded by investing in driver monitoring tools aimed at reducing this risk [1-5].

4. Tools and Techniques to Monitor Driver Fatigue

These tools use a variety of methods including those based on (a) continuous driving time, (b) direct driver performance (e.g., steering) or (c) physical response (e.g., eye metrics). Among the latter, features of the eyes and eyelids have been used to reduce drowsiness [6 - 10]. One of these tools, the Optalert Alertness Monitoring System (OAMS) [11 - 15] uses infra-red (IR) reflectance oculography to monitor the movement of the eyelid.

The system uses an IR emitter and a spectrum embedded in the spectrum to continuously measure the blink of an eye, from where sleep levels are detected. OAMS has been used to diagnose and monitor driver drowsiness in mines [16 - 20] and in the road transport industry [21 - 25] and in driver drowsiness [26 -33]. OAMS has been used in commercial settings primarily to provide sleep data in a central monitoring area. If the driver's drowsiness reaches a certain level of risk, intervention (e.g., compulsory leave) may be made. Such applications come with high cost of monitoring and use. Moreover, such interventions are more dependent on the detection of fatigue than on prevention. The current study examines the benefits of responding to real-time drowsiness in drivers to avoid fatigue [1-5] reports observational data showing a decrease in moderate sleep apnea related to driver response during scheduled discharge. of OAMS in many mining areas [6 - 10].

These data suggest that, at the level of group variability, a real-time driver response may reduce their drowsiness, compared to OAMS monitoring activity alone. They do not; however, provide insight into how each driver might respond to such a response. Our research focuses on this individual response and improved understanding of the use of real-time feedback for individual drivers that could lead to a cheaper strategy to improve the driving results of at-risk individuals. It also aims to evaluate the effectiveness of OAMS in improving driver performance. Focused on Australian Army Reserve staff. For many of the participants, the Army training weekend is usually preceded by a full week of operation and usually a long journey to the Army Reserve work area. In addition,

members often report to work during the day, which includes commuting to and from work.

4.1. Definitions of various Image Quality Metrics

4.1.1. Mean Squared Error (MSE)

Mean Squared Error (MSE) or Mean Squared Deviation (MSD) measurement or non-observable value measurement method is defined as the square measure of error square the difference between the estimated values and the estimated value.

4.1.2. Peak Signal to Noise Ratio (PSNR)

The Peak Signal to Noise Ratio (PSNR) is defined as the ratio between the maximum potential for a given signal and the potential for destructive noise affecting the reliability of its representation.

4.1.3. Structural Similarity for measuring the Image Quality (SSIM)

Image Quality Assignment (SSIM) is defined as the method used to measure the similarity between two images and is also used to predict the assumed quality of digital television and cinema images and other digital forms. photos and videos.

5. New Technique suggested by researchers

A block diagram of our proposed research methodology is shown in Figure (Figure 1) below:

In our test, a traffic signal image site is developed that includes different types of traffic signals of different colors, shapes, sizes and light variations depending on the surrounding weather conditions such as sun, cloud, rain, fog, snow, smoke. and dark weather etc.

After traffic signal images are captured on a mobile sign traffic website with the help of an external mobile camera installed in the car or in some cases it may be mounted on the body itself and in some cases, the driver may wear dual camera mounted on both sides of the camera. pre-embedded images, detection of road signs followed by a process of identifying a sign board followed by a road sign. process. recognition and appropriate classification. The embedded image is found on the image website and is subject to some previous processing functions such as extracting audio and enhancing the image on a local background. Images are obtained using the help of other methods used to separate the image. Featured images are then subjected to a process where the whole image is divided into multiple images or a set of

identical images using the appropriate classification process. After the sorting process, the images are subjected to a feature removal process in which a set of relevant features was removed from the image using specific feature removal techniques. Images are then categorized with the help of a suitable classification algorithm commonly used in the field of work-related applications where an accurate and accurate description of the sign boards is required to ensure the safety of the vehicle driver. for the purpose of separating images.

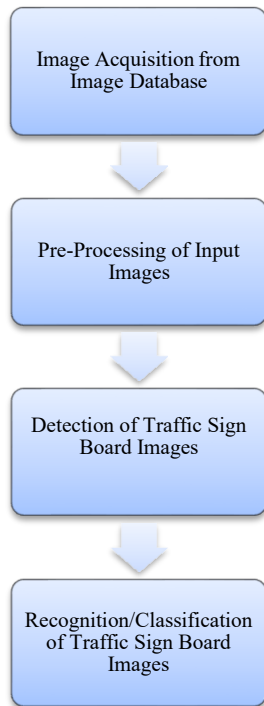


Figure 1: Our Traffic Sign Recognition (TSR) System Block Diagram

6. Results obtained practically as well as Discussion

In this experiment, a traffic signal image was prepared that included images of different road signs of different categories of different colors, shapes, sizes and light variations depending on the surrounding weather conditions such as sun, cloud, rain, fog, snow, density. and dusty weather etc. There are a total of 679 road markings on the 18-dimensional training image website and a total of 48 road markings available at the test site. The training image website is divided into two main categories, namely, non- textual image symbols containing only the color, composition and available text details as well as graphic imagery images containing the color, shape and text information contained in it. Examples of different traffic signals available on our Test and Training photography site are included (Figures 2 and 3) on this page.



Figure 2: Text as well as Non-Text information containing Roadway Symbols from Test data



Figure 3: Text as well as Non-Text information containing Roadway Symbols from Training data

Image filtering techniques with the help of Arithmetic Mean Filter, Circular Averaging Filter (Pillbox), Harmonic Mean Filter, Box Filter, Gaussian Filter were used and various parameters provide a clear idea of the image quality considered. they are numbered. There are a number of parameters that determine the quality of the included image provided when Medium Error (MSE), Peak Signal to Noise Ratio (PSNR) and Image Quality Correction (SSIM) are the same. A table format with numerical values of the above parameters (Table 1) will be very useful in determining which best algorithm to use, for example, PSNR and SSIM values should be higher while MSE values should be lower than a given algorithm. The algorithm that eliminates the above criteria will be

used in the research project for this project. The main goal of this project is to focus on comparative work that combines various techniques for removing noise from digital images using standard performance indicators that allow the researcher to select and apply the best algorithm for future work. Processes that include digital image enhancement, image classification using appropriate techniques, removal of useful features in images followed by image classification using a separator algorithm that should be set for future work. The strategies integrated into performing these specific tasks will also address the comparative process outlined in the above section of this paper. The strategies selected for the proposed activity will be taken from these comparative activities using the best and most accurate method of almost 100% and the main goal will be to achieve the best results compared to other strategies found in the literature section. to ensure high fidelity, efficiency and durability in sound present in two-dimensional images. The different types of algorithms used in the above test are discussed in this paper.

Fast Local Laplacian (FLLF) filter method is applied to the resulting image to extract the output image. The results obtained after performing the Phase 1 Fast Local Laplacian (FLLF) process are shown as shown in Figure 4 (Figures 4 and 5) below:



Figure 5: Fast Local Laplacian Filtering (FLLF) Type 1 Results from Training image database



Figure 4: Fast Local Laplacian Filtering (FLLF) Type 1 Results from Testing image database.

7. Various Image Enhancement in the spatial domain techniques

7.1. Fast Local Laplacian Filtering (FLLF) Type 1

Fast Laplacian Local Filter Method (FLLF) Type 1 is used for the purpose of enhancing the input color provided by enhancing the color contrast of the color image. In this process, the provided input image is first inserted into the workspace by setting the filter parameters to increase the data to less than 0.4. Then, the

7.2. Fast Local Laplacian Filtering (FLLF) Type 2

The calculation value required for digital image processing is very high in the case of Fast Local Laplacian (FLLF) Type 2 filtering compared to the various algorithms used to enhance the input color rendering provided. The algorithm measurement process is performed with the help of the Num Intensity Levels parameter that separates the pixel width from the color input of the image in the number of samples. Performance speed and image quality of the input color provided is measured using this parameter. In this process, the inserted image of the actual color is first introduced into the workspace and displayed. The information contained within a given image is processed using a sigma value parameter and the brightness increases using an alpha value parameter which also effectively enhances the input of the input color image. Killing speed can be increased by using a small number of samples but also usually produces art objects that are clearly visible to the naked eye especially in areas with flat contrast. Work can be timed using only 20 power levels.

The resulting image is subject to processing and is displayed as a final output. Results obtained after performing the Laplacian Rapid Screening (FLF) procedure Type 2 is shown as shown in Figure (Figures 6 and 7) below:



Figure 6: Fast Local Laplacian Filtering (FLLF) Type 2 Results from Testing image database



Figure 9: Laplacian Local Fast Filtering (FLLF) Type 3 Results from Training image database



Figure 7: Fast Local Laplacian Filtering (FLLF) Type 2 Results from Training image database

7.3. Fast Local Laplacian Filtering (FLLF) Type 3

In this process, the color input provided is subjected to processing with the help of the Fast Local Laplacian Filtering (FLLF) Type 3 strategy and is timed using 100 degrees of firmness as the results obtained after performing the whole process are much better compared to the results. obtained during the sample image provided thus a large number of samples are used although the time required to process the image is very high. The resulting image is subject to processing at 100 power levels and is displayed. The results obtained after performing the Laplacian Rapid Scheme (FLF) Type 3 are shown as shown in Figure 8 (Figures 8 and 9) below:

7.4. Fast Local Laplacian Filtering (FLLF) Type 4

In the Fast Local Laplacian (FLLF) type 4 filtering method, the color variation of the included color image is enhanced using the color mode parameter. First, the inserted image is imported into the workspace and the actual image size is reduced and displayed. Filter parameters are adjusted in such a way that details smaller than 0.3 in normal range 0 to 1 can be amplified dramatically. In this process, the embedded color image is enhanced by enhancing the local light intensity. The results obtained after performing the Laplacian Rapid Transmission Process (FLLF) Type 4 are shown as shown in Figure 10 (Figures 10 and 11) below:



Figure 8: Laplacian Local Fast Filtering (FLLF) Type 3 Results from Testing image database



Figure 10: Fast Local Laplacian Filtering (FLLF) Type 4 Results from Testing image database

Figure 12: Fast Local Laplacian Filtering (FLLF) Type 5 Results from Testing image database



Figure 11: Fast Local Laplacian Filtering (FLLF) Type 4 Results from Training image database

Figure 13: Fast Local Laplacian Filtering (FLLF) Type 5 Results from Training image database

7.5. Fast Local Laplacian Filtering (FLLF) Type 5

In this Fast Local Laplacian (FLLF) type 5 filtering process, the given color image is enhanced by enhancing the local color brightness. Colors are most saturated when Color Mode is set to split instead of light despite the fact that an equal amount of light is applied to each image while testing processes are applied to the input images. The results obtained after performing the Laplacian Rapid Screening (FLF) procedure Type 5 are shown as shown in Figure (Figures 12 and 13) below:

7.6. Fast Local Laplacian Filtering (FLLF) Type 6

In the Fast Local Laplacian (FLLF) Type 6 filter process, the input color is first applied to the workspace and then displayed. It is converted into a floating-point image format for the purpose of adding external sound in an easy way. Gaussian sound with zero meaning and variation as 0.001 is added to the external input image. The data size is fine-tuned following the set value of the slip for the purpose of the application. Then, the Edge-sensitive filter is applied to a specific color source. There is an amazing improvement in the provided image of Peak Signal to Noise Ratio (PSNR). It is also noted that the details are smooth and there is no change in the sharp contrast at the edges of the input color image. Results

obtained after performing the Laplacian Rapid Sorting Process (FLLF) Type 6 are shown as shown in Figure (Figures 14 and 15) below:



Figure 14: Fast Local Laplacian Filtering (FLLF) Type 6 Results from Testing image database



Figure 15: Fast Local Laplacian Filtering (FLLF) Type 6 Results from Training image database

7.7. Fast Local Laplacian Filtering (FLLF) Type 7

In this Laplacian Quick Area (FLLF) Filtering Process 7, the information that is naturally present within a given image is corrected without interfering with sharpening the edges present in the image. First, a color-coded image is inserted into the workspace and displayed. The amplitude of the data is set to be smooth and the slide value is also set to the maximum value of the application. It is generally recognized that the filter produces excellent quality results with a small number of hardness levels when the alpha number is greater than 1. The degree of stiffness is set to a minimum value to ensure that the input image is processed at a fast speed. . Then, a filter is applied. The output image is displayed after the complete image filtering process has been performed at the end of the test process. The results obtained after performing the Rapid Laplacian Local Filtering Process (FLLF) Type 7 are shown as shown in Figure 16 (Figures 16 and 17) below:



Figure 16: Fast Local Laplacian Filtering (FLLF) Type 7 Results from Testing image database

Table 1: Comparison of various image enhancement in the spatial domain techniques in our Test data based upon Mean Squared Error, PSNR and SSIM numerical values

1	2016	Fast LocalLaplacianFiltering (FLLF) Type 1	298.540	23.409	0.811
2	2016	Fast LocalLaplacianFiltering (FLLF) Type 2	133.472	26.911	0.878
3	2016	Fast LocalLaplacianFiltering (FLLF) Type 3	156.974	26.212	0.868
4	2016	Fast LocalLaplacianFiltering (FLLF) Type 4	841.632	18.920	0.644
5	2016	Fast LocalLaplacianFiltering (FLLF) Type 5	809.782	19.089	0.649
6	2016	Fast Local Laplacian Filtering	141.266	26.654	0.887

		(FLLF) Type 6			
7	2016	Fast LocalLaplacianFiltering (FLLF) Type 7	0.0002	35.445	0.972
8	2006	Entropy Filtering	18.16306875	-12.562758333333	-0.00484375
9	2006	StandardDeviation Filtering	335.46449375	-24.801495833333	-0.00370625
10	-	Pretrained Neural Network based Image Denoising	1.25E-5	47.180	0.998

Table 3: Comparison of various image enhancement in the spatial domain techniques in our Test data against our proposed technique depending on various image quality metrics like MSE, PSNR and SSIM numerical values

1	2006	Pretrained Neural Network based Image Denoising	1.25E- 5	47.180	0.998
2	2006	Fast LocalLaplacianFiltering (FLLF) Type 7	0.0002	35.445	0.972
3	Our Proposed Technique	Pretrained Neural Network based ImageDenoising + Fast Local Laplacian Filtering (FLLF) Type 7	0.0000	Infinity	1.0000

8. Uniqueness of Propounded Strategy

The main objective of our proposed research project is to identify the best way to remove natural noise from an image among various image removal techniques based on different image quality metrics such as Mean Squared Error (MSE), Peak Signal to Noise Rate (PSNR) and Image Quality. SSIM). The best strategy is identified by partners by considering that it has the smallest number of Mean Squared Error (MSE) numbers and has the largest value. Higher Signal Number to Noise Ratio (PSNR) and Parallel Image Quality (SSIM). After the Mean Squared Error (MSE) number, Peak Signal to Noise Ratio (PSNR) and Measurement Properties to measure Image Quality (SSIM) were completed, it was noted that the value of the Mean Squared Error (MSE) was obtained by 0.0000 and Peak. Signal to Noise Ratio (PSNR) and SSIM value will be Infinity and 1.0000. Pretrained Neural Network based Image Denoising and Laplacian Rapid Screen Filter (FLLF) Type 7 were used in road sign board images. The results obtained after combining the techniques of Pretrained Neural Network based Image Denoising and Laplacian Rapid Filtering (FLLF) Type 7 of the traffic sign board images available on our test image site are shown below:



Figure 18 Pretrained Neural Network based Image Denoising + Fast Local Laplacian Filtering (FLLF) Type 7 Results from the Testing image database

9. Concluding Remarks

In this research paper, comparisons of various image enhancements to local domain techniques were made to select the best method between different strategies based on calculating different parameters that determine the quality of road input features such as Mean Squared Error (MSE), Peak Signal to Noise. Rate (PSNR) and Image Quality (SSIM) and after a thorough comparison, concluded that if the techniques of Pretrained Neural Network based Image Denoising and Laplacian Faster Filter (FLLF) Type 7 exist. well integrated, with Peak

Signal to Noise Ratio (PSNR) and similar Image Quality Assurance Form (SSIM) obtained as Infinity and 1.0000 respectively with a relatively small error rate. (MSE) value 0.0000 compared to other site image testing strategies. The combination of these two methods mentioned in the above section of our research paper will be used for the Enlargement of Input Photographs available for the Examination and Training of Photographic Boards for future research work.

Conflict of Interest

The author declares no conflict of interest.

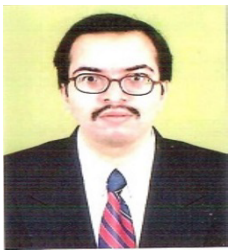
References

- [1] Huaping Liu, Yulong Liu, Fuchun Sun, "Traffic sign recognition using group sparse coding", Elsevier- Information Sciences, Volume 266, January 2021, pp. 75-89.
- [2] Zhan-Li Sun, Han Wang, Wai-Shing Lau, Gerald Seet, Danwei Wang, "Application of BW-ELM model on traffic sign recognition", Elsevier- Neurocomputing, Volume 128, October 2021, pp. 153-159.
- [3] Fatin Zaklouta, Bogdan Stanculescu, "Real-time traffic sign recognition in three stages", Elsevier- Robotics and Autonomous Systems, Volume 62, August 2021, pp. 16-24.
- [4] Shuihua Wang, Hangrong Pan, Chenyang Zhang, Yingli Tian, "RGB-D image-based detection of stairs, pedestrian crosswalks and traffic Signs", Elsevier- Journal of Visual Communication and Image Retrieval, Volume 25, November 2020, pp. 263-272.
- [5] Jonathan J. Kay, Peter T. Savolainen, Timothy J. Gates, Tapan K. Datta, "Driver behaviour during bicycle passing manoeuvres in response to aShare the Road sign treatment", Elsevier-Accident Analysis and Prevention, Volume 70, April 2021, pp. 92-99.
- [6] Jesmin Khan, Sharif Bhuiyan, Reza Adhami, "Hierarchical clustering of EMD based interest points for road sign detection", Elsevier- Optics & Laser Technology, Volume 57, October 2021, pp. 271-283.
- [7] Zong-Yao Chen, Wei-Chao Lin, Shih-Wen Ke, Chih-Fong Tsai, "Evolutionary feature and instance selection for traffic sign recognition", Elsevier- Computers in Industry, Volume 74, September 2020, pp. 201-211.
- [8] Samuele Salti, Alioscia Petrelli, Federico Tombari, Nicola Fioraio, Luigi Di Stefano, "Traffic sign detection via interest region extraction", Elsevier- Pattern Recognition, Volume 48, June 2020, pp. 1039-1049.
- [9] J.M. Lillo-Castellano, I. Mora-Jiménez, C. Figuera-Pozuelo, J.L. Rojo-Álvarez, "Traffic sign segmentation and classification using statistical learning methods", Elsevier- Neurocomputing, Volume 153, November 2021, pp. 286-299.
- [10] Haojie Li, Fuming Sun, Lijuan Liu, Ling Wang, "A novel traffic sign detection method via colour segmentation and robust shape matching", Elsevier- Neurocomputing, Volume 169, May 2021, pp. 77-88.
- [11] Zhenyu An, Zhenwei Shi, Ying Wu, Changshui Zhang, "A novel unsupervised approach to discovering regions of interest in traffic images", Elsevier- Pattern Recognition, Volume 48, February 2020, pp. 2581-2591.
- [12] Yingying Zhu, Chengquan Zhang, Duoyou Zhou, Xinggang Wang, Xiang Bai, Wenyu Liu, "Traffic sign detection and recognition using fully convolutional Network guided proposals", Elsevier- Neurocomputing, Volume 214, July 2021, pp. 758- 766.
- [13] Ayoub Ellahyani, Mohamed El Ansari, Ilyas El Jaafari, "Traffic sign detection and recognition based on random forests", Elsevier-Applied Soft Computing, Volume 46, February 2020, pp. 805-815.
- [14] Selcan Kaplan Berkaya, Huseyin Gunduz, Ozgur Ozsen, Cuneyt Akinlar, Serkan Gunal, "On circular traffic sign detection and recognition", Elsevier- Expert Systems with Applications, Volume 48, 2020, pp. 67-75.
- [15] Yongtao Yu, Jonathan Li, Chenglu Wen, Haiyan Guan, HuanLuo, Cheng Wang, "Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data", Elsevier- ISPRS Journal of Photogrammetry and Remote Sensing, Volume 113, January 2021, pp. 106-123.
- [16] Hamed Habibi Aghdam, Elnaz Jahani Heravi, Domenec Puig, "A practical approach for detection and classification of traffic signs using Convolutional Neural Networks", Elsevier- Robotics and Autonomous Systems, Volume 84, July 2020, pp. 97-112.
- [17] Mario Soilán, Belen Riveiro, Joaquin Martinez-Sanchez, Pedro Arias, "Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory", Elsevier- ISPRS Journal of Photogrammetry and Remote Sensing, Volume 114, February 2020, pp. 92-101.
- [18] Y. Ouerhani, A. Alfalou, M. Desthieux, C. Brosseau, "Advanced driver assistance system: Road sign identification using VIAPIX system and a correlation technique", Elsevier- Optics and Lasers in Engineering, Volume 89, May 2021, pp. 184-194.
- [19] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs", IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 3, June 2020, pp. 1360-1369.
- [20] Nadra Ben Romdhane, Hazar Mliki, Mohamed Hammami, "An Improved Traffic Signs Recognition and Tracking Method for Driver Assistance System", in the Proceedings of IEEE International Conference on Information Security (ICIS'2020), June 26- 29, 2020, Okayama, Japan.
- [21] Ruben Laguna, Ruben Barrientos, L. Felipe Blazquez, Luis J. Miguel, "Traffic sign recognition application based on image processing techniques", in the Proceedings of the Elsevier-Proceedings of the 19th World Congress, The International Federation of Automatic Control, Cape Town, South Africa. August 24-29, 2021, pp. 104-109.
- [22] X. Baro, S. Escalera, J. Vitria, O. Pujol, P. Radeva, "Traffic sign recognition using evolutionary adaboost detection and forest- ecoc classification", IEEE Transactions on Intelligent Transport Systems, Volume 10, Issue 1, 2020, pp. 113-126.
- [23] S.M. Bascon, J.A. Rodriguez, S.L. Arroyo, A.F. Caballero, F. Lopez-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs", Computer Vision Image Understanding, Volume 114, Issue 3, 2021, pp. 373-383.
- [24] C.C. Chang, Y.P. Hsieh, "A fast VQ codebook search with initialization and search order", Information Sciences, Volume 183, Issue 1, 2021, pp. 132-139.
- [25] Ciresan, U. Meier, J. Mascim, J. Schmidhuber, "A committee of neural networks for traffic signs classification", in Proceedings of International Joint Conference on Neural Networks (IJCNN), July 2020, pp. 1918-1921.
- [26] S. Escalera, O. Pujol, P. Radeva, "Traffic sign recognition system with b-correction", Elsevier- Machine Vision Applications, Volume 21, Issue 2, 2020, pp. 99-111.
- [27] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition", IEEE Transactions on Intelligent Transportation Systems, Volume 11, Issue 4, 2021, pp. 917-930.

- [28] Y. Gu, T. Yendo, M.P. Tehrani, T. Fujii, M. Tanimoto, "Traffic sign detection in dual-focal active camera system", in the Proceedings of IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 1054-1059.
- [29] S. Gunal, R. Edizkan, "Subspace based feature selection for pattern recognition", Elsevier-Information Sciences, Volume 178, 2021, pp. 3716-3726.
- [30] Y. Huang, K. Huang, Y. Yu, T. Tan, "Salient coding for image classification", in the Proceedings of Computer Vision and Pattern Recognition (CVPR), 2021, pp. 1753-1760.
- [31] S. K. Chung, "A Phase Tracking System for Three Phase Utility Interface Inverters", IEEE Trans. On Power Electronics, Vol. 15, pp. 431-438, May 2000.
- [32] Hassanabad, A. H., & Nazeipur, D. Design and Simulation of a Control System for Investors in Wind Turbines. Vol. 6, No. 3, 2021, pp. 31-36.
- [33] Jun-min, L. K. P. Z., & Yang, X. U. A. N. (2003). Harmonics detection for three-phase circuits based on resampling theory and mean filtering. Proceedings of the CSEE.

topics in the concerned field. I was also being awarded Honorary Doctor of Letters (D.Litt.) in the branch of Electronics Engineering by the California Public University (CPU), CA, USA.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Mr. Abhinav V. Deshpande

I have done B.E. in Electronics & Telecommunication Engineering from G. H. Rasoni College of Engineering (GHRCE), Nagpur, Maharashtra (M.H.), India-440016 in the year 2010 with an aggregate of 63% and M. Tech. in Electronics Engineering from the same college with a CGPA of 7.90 on a scale of 10.00. I was awarded Ph. D. in Electronics Engineering from the School of Electronics Engineering (SENSE) at Vellore Institute of Technology (VIT), Vellore, Tamil Nadu (T.N.), India-632014. I also worked as an Assistant Professor on Contract Basis in the Department of Electronics & Telecommunication Engineering at Prof. Ram Meghe Institute of Technology & Research (PRMITR), Badnera, Amravati, Maharashtra (M.H.), India-444701 for a total duration of 1 year from the period starting from 16/06/2014 up to 07/07/2015. I have published 58 research papers in various and reputed International Journals and Conferences. I am currently having a total number of 38 citations on my research papers and I am also having 3 as the h-index and 1 as the i-10 index. I have successfully completed 2 research projects during the course of B.E., M. Tech. and Ph.D. I am also having memberships of 15 different and reputed professional organizations. I also received 17 awards from various and reputed professional organizations. I have also published 1 Book by the Lambert Academic Publishing (LAP) in Saarbrucken, Germany on 10/08/2012. I have also attended 40 various Workshops on different

The Effects of Total Health Expenditure on Economic Growth in Southern and Western sub-Saharan Africa

Faith Onechojo Yusufu ^{1,2,*}, Bosede Olanike Awoyemi ², Kehinde Akomolafe ²

¹Department of Economics, University of Jos, Plateau State, Nigeria

²Department of Economics, Afe-Babalola University Ado-Ekiti, Ekiti State, Nigeria

*Corresponding author: Faith Onechojo Yusufu, Email address: onechojoy@unijos.edu.ng or faithyusufu54@gmail.com, +2347032732838
Corresponding author ORCID: 0000-0002-8531-0467

ABSTRACT: The purpose of this work was to investigate the effects of total health spending on the growth of the economy in Southern and Western Sub-Saharan Africa. The mean group, dynamic fixed effect, and pooled mean group/ARDL (Autoregressive Distributed Lags) panel data analyses were used to scrutinize the short and long-term effects of total per capita health spending on the growth of the economy. The short-run finding reveals that total health per capital expenditure and life expectancy at birth (LEB) has an upbeat effect on the growth of the economy (LGDP PC) in Southern Africa at all relevant levels. At all significant levels, total health per capital expenditure and life expectancy at birth (LEB) both have an upbeat outcome on the growth of the economy (LGDP PC). The short-run analysis shows that current health per capital spending and government spending has a positive effect on the growth of the economy (LGDP PC) at a 5% significant level in West Africa, while total and current health per capital expenditure has an upbeat effect on economic growth (LGDP PC) at 5% significant level in the long run. Because per-capita health spending has a beneficial impact on the growing economy, more monies should be dedicated to the health sector to increase the quality of healthcare operations.

KEYWORDS: Economic growth, GDP per capita, Public spending, ARDL (Autoregressive Distributed Lags).

1. Introduction

Inadequate investment from both the corporate and public sectors, makes improving health outcomes in Africa difficult. Most African governments find it difficult to devote at least 15% of their annual budget to health, as recommended by the World Bank [1]. This can be attributed to rising healthcare expenditures and the economic crisis, which has put additional strain on healthcare spending. Infrastructure, equipment, and qualified health care practitioners are all lacking in African countries' healthcare systems. The average overall health spending in African countries in 2010 was US\$ 135 per capita, compared to US\$ 3150 in high-income countries. Furthermore, health spending per capita in North America, Europe, Latin America and the Caribbean, and Sub-Saharan Africa in 2016 was approximately \$ 9,031, \$3,183, \$63, and \$84 correspondingly [1]. This demonstrates how distant African countries are from meeting the bare minimum of healthcare requirements and from meeting the healthcare needs of Africans in particular [2]. This has resulted in higher rates of mother

and newborn mortality, HIV/AIDS, and other deadly diseases in Africa. These figures show that Sub-Saharan African countries must increase healthcare investment to keep up with the rest of the globe.

Out-of-pocket payments, which constitute 40% or more of total health expenditure; in most African nations; are the most regressive form of health expenditure [1]. Some forms of health care have been hampered as a result, and individuals who cannot afford them are denied access. In most African countries, only the middle and upper classes can afford basic health care. This has resulted in countless deaths and impoverishments, as well as a loss of human capital and a continual decline in African life expectancy. While some previous studies believe that medical care has significantly contributed to observed mortality decreases due to enhanced combat of infectious diseases, others consider that increased medical care does not unambiguously lead to an increase in life expectancy [3]. Furthermore, the appropriate allocation of scarce resources to diverse sections of the country is hampered by a slew of impediments such as

bias and theft of cash. This has made it extremely difficult for African countries to meet the Millennium Development Goals (MDGs) of lowering infant mortality, increasing maternal health, and combatting HIV/AIDS, malaria, and other diseases, particularly in rural regions.

For example, in Nigeria for 2020, the health division obtained a capital spending allowance of N46 billion, which was N2 billion less than the country's allowance for education, and it was an N4.15 billion less than the capital spending allowance for health in 2019, which was N50.15 billion, and despite the relative budgetary allocation for the health sector, it hasn't reflected in the health status of Nigerians as most rural communities have no health centers. One of the causes for insufficient health resource allowance in Sub-Saharan Africa is a lack of focus on the function of health expenditure. Mismanagement of healthcare resources and a deficient healthcare system are also among the issues. Healthy people are more productive and make more money, whereas sickness and disability hurt earnings, particularly in developing nations where most jobs demand manual labor [4]. The objective of this study was to look at the effects of total health spending on the growth of the economy in southern and western Sub-Saharan Africa's regions.

2. Literature Review

2.1. Total spending on health

This is the total of broad government and private health spending in a particular year, designed in national currency units at current prices. It is calculated the same as the aggregate of all financing agents who manage funds for the purchase of healthcare goods and services. The average overall health spending in African countries in 2010 was US\$ 135 per capita, compared to US\$ 3150 in high-income countries. Even though a lot of African nations have improved the proportion of overall public spending billed to health, total health funding remains a key hindrance to successful health care performance. Furthermore, the slowing of economic growth and high public debt has limited fiscal flexibility for public funding of healthcare, amid the average debt-to-GDP ratio rising by 15% proportion points from 2010 to 2017. Overall healthcare spending in Africa has maintained in a narrow range of 5-6 % of GDP on average from the year 2000 to the year 2015, despite nearly doubling in per capita conditions ranging from \$150-292 (in constant PPP dollars). Scarce governmental resources and unpredictability of foreign help have resulted in excessive private out-of-pocket spending, pushing many individuals into poverty.

Figure 2.1 illustrates the level of total current health spending expressed as a percentage of GDP for certain Sub-Saharan nations, demonstrating that there are no significant increases or declines between 2000 and 2018.

Current health spending estimates include healthcare items and services spent each year. This measure eliminates capital health spending such as information machinery and vaccine supplies for emergencies etc.

2.2. Determinants of the growth of Health Expenditure

There are many determinants of the intensification of health care spending. Some of them are;

2.2.1. The Role of the Government

Governments take part in an imperative function in health care finance by creating and providing adequate assets via public budgets and other contributing machinery, combining assets devoted to health advancement, overseeing the assets allocation process, and acquiring health services from different dealers. Governmental duty for public health goes further than charitable activities and services to incorporate mandatory immunization legislation, quarantine, and regulatory authority.

The state corporation functions by inspiring residents to do things that ease their fitness like juggling or by eating healthy meals. Creating a good health finance system is one of the most important instruments for demonstrating leaders' obligation and political willpower, as well as their capacity to transform these pledges into outcomes [5].

2.2.2. The price of the health care

The higher the price, the greater the amount of money required to make it available. Nigeria, for example, spent more than ten billion dollars on the COVID vaccine. Although in Africa, domestic government funds financed 44% of current health spending (meaning "the ultimate utilization of health goods and services") in 2016, out-of-pocket payments accounted for more than 37 percent of total African health spending, which can be attributed to health care unit pricing. South Africa initiated the first signal of the latest customer charge obliteration initiatives in Africa in 1994. The UN Secretary-General and African Union advocated for healthcare with no charge for pregnant women and children below the age of five years in 2010.

Countries such as Rwanda, for example, have a budget that assures the health sector receives more than 20% of the financing, as opposed to the Abuja Declaration's 15%, which many African countries have yet to implement.

In addition to health spending, additional drivers such as capita per GDP, urbanization, vaccination, and adequate drinking water improved life expectancy, child (below 5 years) mortality, and crude death. Unemployment and HIV prevalence, on the other hand, diminish life expectancy and increase infant mortality, child (below 5 years) mortality, and crude death. Health spending is a

significant factor in achieving better health outcomes in Sub-Saharan African countries. As a result, intensifying the amount of health spending devoted to the health division results in improved health. Furthermore, amending policies to boost capita per GDP, vaccination, urbanization, and basic drinking water provision, as well as initiatives, to minimize unemployment and HIV occurrence, provide a better health outcome [2]

2.2.3. Foreign aid

Foreign aid is very crucial in developing countries. The reliance of middle and low-income countries on foreign aid, as well as trade openness, has increased government spending [6]. The quantity of government-to-government international aid specifically for health has expanded considerably since the beginning of the decade. Aid now directly funds approximately 10% of Africa's healthcare expenditure. However, it is becoming increasingly obvious that this additional spending has minimal impact on health in the world's poorest regions. Very little progress has been achieved in achieving the Millennium Development Goals for Health.

Development Goals and far too many people still pay for health care out of pocket. Part of the failure can be traced back to the existing paradigm of official foreign aid, in which governments in affluent nations give enormous quantities of money to governments in impoverished countries with the assumption that it will be used wisely. Unfortunately, due to corruption and other forms of mismanagement, very little of this money is used to provide medical care. In Sub-Saharan Africa, health care is a necessity rather than a luxury. To stimulate economic growth in Sub-Saharan Africa, it is therefore required to construct effective and efficient health care programs, raise health expenditure, make better use of the young population, and create a better climate for foreign direct investment [7].

2.3. Theoretical Framework

2.3.1. Keynesian standard dynamic model of growth

In this model Keynes recognizes that increased government spending raises aggregate demand and increases consumption.

In equation (1), the balance equation establishes the equality of the national income to the sum of all expenditures as;

$$Y(t) = C(t) + I(t) + G(t) \quad (1)$$

The model assumes that consumption expenditure in a given period is determined by the level of income in that same period. Consumption expenditure is treated as an endogenous variable equal to the quantity of domestic consumption of some portion of national income, and ultimate consumption is income independent. As a result,

consumer spending has increased. $C(t)$ is described by the linear equation of the economic multiplier.

$$C(t) = a(t)Y(t) + b(t) \quad (2)$$

where $a(t)$ is the multiplier factor that explains the marginal propensity to consume ($0 < a(t) < 1$), and the function $b(t) > 0$ is the autonomous consumption that is independent of income?

The expression $a(t)Y(t)$ Consumption that is unrelated to income is defined. Exogenous variables in the static model include investment spending and government expenditure. According to [8], in the dynamic Keynesian model, investment expenditure is viewed as endogenous and is supposed to be dependent on income level. The investment expenditure $I(t)$ is determined by the rate of change in national income and represents private-sector spending. The equation of the economic accelerator describes this assumption.

2.3.2. Effects of total health expenditure and per-capita health expenditure on economic growth

In [9], the author examined the impact of government spending on general administration, defense, education, and health on Nigeria's GDP using time series data generated from Central Bank of Nigeria (CBN) statistical bulletins spanning the years 1983 to 2016. In the multiple regression analysis, the Ordinary Least Squares (OLS) estimation approach was applied. The findings revealed that spending on General Administration has an upbeat and significant association with the growth of the economy; spending on Defense has a downbeat but significant association with GDP; spending on Education has an upbeat and highly significant relationship with economic growth, and spending on Health has an upbeat but insignificant outcome on GDP [10]. Investigated the validity of Wagner's hypothesis in elucidating health spending in Botswana. The first school of thinking holds that health spending stimulates the economy, whereas the second holds that health spending drives the economy. The third school of thought holds that there is a feedback effect between health spending and the economy, but the fourth holds that there is no causality between the two variables at all. However, this analysis discovered that there is no causal association between health expenditure and GDP in Botswana, hence denying Wagner's theory's relevance.

In [11], it is used the ordinary least squares (OLS) multiple regression approaches to examine the impact of government spending on economic growth in Nigeria. The findings revealed that government total capital spending, total recurrent expenditures, and government expenditure on education and power hurt economic growth and are significant in explaining this relationship,

whereas rising government expenditure on transportation and communication and health increases economic growth.

In [12], the VECM approach to empirically examine the influence of government spending on economic growth in Tanzania from 1990 to 2015. All variables showed long-run cointegration, according to the results. Furthermore, the findings revealed that government spending, foreign direct investment inflows, gross capital formation, and inflation have positive and significant long-run and short-run relationships with the Tanzanian growth economy. [13] Used panel data from 30 Sub-Saharan African countries from 1970 to 2010 to look into the extent to which population health affects economic performance. The authors looked at the connection between population health capital and the growth of the economy in Sub-Saharan Africa using a theoretical model based on Solow growth model augmentation and a panel cointegration econometric approach. They discovered that the health condition of the population has not considerably determined the growth of the economy. The consequence of HIV=AIDS, however, resulted in a significant downbeat outcome of population health on the growth of the economy.

In [14], it is used annual data from 1990 to 2015 to estimate the relationship between health expenditure (HE), environmental contamination, and the growth of the economy in Sub-Saharan African countries. The ARDL estimation method was used to model the long run and short run, and the VECM Granger causality test was used to check the direction of causality. To begin, the ARDL test results show that economic expansion has a favorable impact on HE in the long run. According to the findings, a 1% increase in per capita GDP results in a 0.332 percent increase in health expenditure.

3. Methodology

Descriptive statistical tools (table 1-3); which are correlation and summary statistic and unit root test was used in the study in order to clearly understand the data set used in this study. The panel data analysis using the mean group, dynamic fixed effect, and pooled mean group/ARDL (Autoregressive Distributed Lags) was used and the data covered the period 2000-2019. The investigation confirmed that the variables are non-stationary based on the results of IPS, Fisher ADF, and PP in table 2, which reveal that the variables are integrated of order 1 and significant at the 5% level. As a result, they are amenable to ARDL analysis.

3.1. Model Specification

Examine the effects of total health expenditure and per health expenditure on economic growth in Southern Africa and Western Africa

$GDP_PC = f (THE, CAH, CHE, PHE, PRHE, GE, LEB, PGR, LF)$

However, the models are specified in empirical forms as;

$$nGDP_{it} = \gamma_0 + \gamma_1 \ln THE_{it} + \gamma_2 \ln CHE_{it} + \gamma_3 \ln GE_{it} + \gamma_4 \ln PGR_{it} + \gamma_5 \ln LEB_{it} + \gamma_6 \ln CAH_{it} + \gamma_7 \ln PRHE_{it} + \gamma_8 \ln PUHE_{it} + \gamma_9 \ln LF_{it} + \varepsilon_{1it} \dots \dots \dots (3)$$

“ γ_s ” represent the coefficients of the regression equation, “ γ_0 ” are constants and “ ε_{1it} ” are the error term where; the variables under consideration are GDP per capita (GDP PC), capital health expenditure as a percentage of GDP (CAH GDP), current health expenditure as a percentage of GDP (CHE GDP), total health expenditure as a percentage of GDP (THE GDP), total health expenditure per capita (THE PC), private health expenditure as a percentage of GDP (PRHE GDP), public health expenditure per capita (PUHE PC), public health expenditure as a percentage of GDP (PUHE G (LF).

3.2. Summary Statistics for Southern Africa and Western Africa

The basic statistical summary of the series under consideration for Southern Africa and West Africa includes the observation, mean, minimum, maximum, standard deviation, and observations which are summarized in *table 1*. From *table 1*, in Southern Africa, the mean GDP_PC is 2898.817 and the standard deviation is greater than the mean with a value of 2593.325 which indicates wide variability. The minimum value for GDP_PC is 272.991 which is the smallest value in the series and the maximum value of GDP_PC is 10892.54 which is the highest value in the series. The mean value of CAH_GDP is 0.312916 and the standard deviation is 0.26757 which indicates that the value shows no evidence of variability. The minimum value for CAH_GDP is 0.002817 and is the smallest value in the series and the maximum value of CAH_GDP is 1.196823 which is the highest value in the series. The mean value of CHE_GDP is 5.896041 and the standard deviation is 1.83989 which indicates that the value does not show evidence of variability. The minimum value for CHE_GDP is 1.908599 and is the smallest value in the series and the maximum value of CHE_GDP is 11.57911 which is the highest value in the series.

Moreover, the mean value of THE_GDP is 5.306437 and the standard deviation is less than the mean with a value of 2.48785 which indicates that the value is close to the mean. The minimum value for THE_GDP is 0 which is the smallest value in the series and the maximum value of THE_GDP is 11.57911 which is the highest value in the series. The mean value of THE_PC is 68.23239 and the standard deviation is greater than the mean with a value of 281.552 this indicates that the value shows evidence of wide variability from the mean. The minimum value for THE_PC is 0 which is the smallest value in the series and the maximum value of THE_PC is 1552.573 which is the

highest value in the series. The mean value of PRHE_GDP is 2.908607 and the standard deviation is less than the mean with a value of 1.82036 this indicates that the value is close to the mean. The minimum value for PRHE_GDP is 0 which is the smallest value in the series and the maximum value of PRHE_GDP is 9.49238 which is the highest value in the series. The mean value of PUHE_GDP is 2.664256 and the standard deviation is less than the mean with a value of 1.28062, this indicates that the value is close to the mean. The minimum value for PUHE_GDP is 0.438216 which is the smallest value in the series and the maximum value of PUHE_GDP is 5.826442 which is the highest value in the series. Also, the mean value of PRHE_PC is 72.24588 and the standard deviation is greater than the mean with a value of 82.8003 this indicates that there is wide variability of the mean. The minimum value for PRHE_PC is 1.775926 which is the smallest value in the series and the maximum value of PRHE_PC is 409.989 which is the highest value in the series.

Furthermore, the mean value of PGR is 1.894686 and the standard deviation is 1.28693 which indicates that the value is less than the mean. The minimum value for PGR is -2.62866 is the smallest value in the series and the maximum value of PGR is 5.604957 which is the highest value in the series. Also, the mean value of LEB is 57.67017 and the standard deviation is 10.1628 which indicate that the value is close to the mean as the mean is greater than the value of the standard deviation. The minimum value for LEB is 42.595 is the smallest value in the series and the maximum value of LEB is 78.49756 which is the highest value in the series. Lastly, the mean value of LF is 50.08885 and the standard deviation is 16.1664 which indicate that the value is close to the mean. The minimum value for LF is 25.11 is the smallest value in the series and the maximum value of LF is 74.58 which is the highest value in the series.

In West Africa, the mean GDP_PC is 2134.786 and the standard deviation is greater than the mean with a value of 3232.933 which indicates variability. The minimum value for GDP_PC is 194.873 which is the smallest value in the series and the maximum value of GDP_PC is 20532.98 which is the highest value in the series. The mean value of CAH_GDP is 0.47801 and the standard deviation is 0.60799 which indicates that the value shows slight evidence of variability. The minimum value for CAH_GDP is 0 and is the smallest value in the series and the maximum value of CAH_GDP is 3.028301 which is the highest value in the series. The mean value of CHE_GDP is 5.588473 and the standard deviation is 2.04079 which indicate that the value does not show evidence of variability. The minimum value for CHE_GDP is 2.49142 and is the smallest value in the series and the maximum value of CHE_GDP is 12.6201 which is the highest value in the series. Also, the mean value of THE_GDP is 5.30905

and the standard deviation is less than the mean with a value of 2.33325 this indicates that the value is close to the mean. The minimum value for THE_GDP is 0 which is the smallest value in the series and the maximum value of THE_GDP is 12.6201 which is the highest value in the series.

The mean value of THE_PC is 83.51476 and the standard deviation is greater than the mean with a value of 147.991 this indicates that the value shows evidence of wide variability from the mean. The minimum value for THE_PC is 0 which is the smallest value in the series and the maximum value of THE_PC is 829.9979 which is the highest value in the series. The mean value of PRHE_GDP is 3.675908 and the standard deviation is less than the mean with a value of 1.87557 this indicates that the value is close to the mean. The minimum value for PRHE_GDP is 0 which is the smallest value in the series and the maximum value of PRHE_GDP is 9.199446 which is the highest value in the series. The mean value of PUHE_GDP is 1.719096 and the standard deviation is less than the mean with a value of 1.22588, this indicates that the value is close to the mean. The minimum value for PUHE_GDP is 0.240667 which is the smallest value in the series and the maximum value of PUHE_GDP is 6.048561 which is the highest value in the series. Also, the mean value of PRHE_PC is 38.80165 and the standard deviation is greater than the mean with a value of 52.9876 this indicates that there is wide variability of the mean. The minimum value for PRHE_PC is 3.470781 which is the smallest value in the series and the maximum value of PRHE_PC is 276.8144 which is the highest value in the series. Furthermore, the mean value of PGR is 2.588873 and the standard deviation is 0.72641 which is less than the mean indicating that the standard deviation is close to the mean. The minimum value for PGR is -1.13152 is the smallest value in the series and the maximum value of PGR is 5.363199 which is the highest value in the series. Also, the mean value of LEB is 58.96343 and the standard deviation of 6.73067 indicates that the value is close to the mean. The minimum value for LEB is 47.713 which is the smallest value in the series and the maximum value of LEB is 74.30976 which is the highest value in the series. Lastly, the mean value of LF is 46.6044 and the standard deviation is 12.07798 which indicates that the value is close to the mean. The minimum value for LF is 23.74 is the smallest value in the series and the maximum value of LF is 72.13 which is the highest value in the series.

3.3. Unit root test

The investigation confirmed that the variables are non-stationary based on the results of IPS, Fisher ADF, and PP in Table 2, which reveal that the variables are integrated of order 1 and significant at the 5% level. As a result, they are amenable to ARDL analysis.

3.4. Correlation Matrix

This section discusses the degree of association and the possible relationship that exists between the variables with 290 numbers of observations. It also shows how variables are related and ascertain whether or not the explanatory variables are highly correlated as shown in *table 3*.

Table 3 shows that there is a positive relationship between the dependent variable GDP_PC which is a point of interest to this study and CAH_GDP, CHE_GDP, THE_PC, PUHE_PC, PRHE_PC, PGR, LEB and LGE. The correlation between CAH_GDP and GDP_PC is 0.6055, indicating that they are positively and strongly correlated. The correlation between CHE_GDP, CHE_PC and GDP_PC is 0.5855 and 0.491 respectively, indicating that there is a strong and positive correlation among the variables. Also, the correlation between THE_PC and GDP_PC is 0.7709 which implies that increase in THE_GDP is positively correlated to GDP_PC.

Similarly, the correlation between PUHE_PC, PRHE_PC and GDP_PC is 0.6735 and 0.6185 respectively, indicating that PUHE_PC and PRHE_PC are positively and strongly correlated to GDP_PC. In the same vein, correlation between PGR, LEB and GDP_PC is 0.5973 and 0.4666 respectively, which indicates that they are positively and strongly correlated, this implies that PGR and LEB are strongly correlated to GDP_PC. Lastly, the correlation between LGE and GDP_PC is 0.8459 indicating that there is a positive and strong correlation among the variables, this implies that LGE is strongly correlated to GDP_PC. Overall, GDP_PC has a positive and strong relationship with the variables of interest. There is also no problem of multicollinearity among the variables, indicating that the variables are amenable to further analysis.

4. Data Analysis and Presentation of Results

4.1. The effect of total per-capita health expenditure on economic growth across the Southern and Western regions in sub-Saharan Africa

This section presents the estimated coefficients of total per capita health expenditure (THE PC), current health expenditure per capita (CHE PC), and life expectancy at birth (LEB) on GDP per capita (LGDP PC) used in this study using Mean Group (MG), Dynamic Fixed Effect (DFE), and Pooled Mean Group (PMG) (PMG). The study, however, used the Hausman test to select the appropriate model from Mean Group (MG), Dynamic Fixed Effect (DFE), and Pooled Mean Group (PMG). To choose between MG, DFE, and PMG, Hausman's test is used. If the test results are significant, it suggests MG or DFE; otherwise, PMG will be explored.

4.1.1. The effect of total per-capita health expenditure on economic growth across the Southern African region

Mean Group Effect of Total Health Expenditure per Capita on Economic Growth in Southern Africa

Table 4 displays the long and short-term effects of the mean group effect of total health expenditure per capita on economic growth in Southern Africa. The short-run result demonstrates According to the Z-statistics, the ECT value is negative and significant at the 5% level of significance, indicating convergence to equilibrium. This demonstrates that short-run inconsistencies are being addressed and incorporated into the long-run relationship. In the short run, both total health per capita spending (LTHE PC) and life expectancy at birth (LEB) have a positive influence on LGDP PC at all levels of significance, indicating that a unit increase in THE PC and LEB will boost LGDP PC by 9% and 3%, respectively. Furthermore, THE PC and LEB exhibit long-run positive impacts on LGDP PC at all levels of significance, implying that a unit increase in THE PC and LEB increases LGDP PC by 77.7 and 10.8 percent, respectively. According to the findings, an upbeat in per capita health spending and life expectancy at birth will boost the growth of economy in Southern Africa in the short term. Similarly, an upbeat in life expectancy at birth and per capita health expenditure has a long-term favorable influence on Southern African growth of the economy.

4.1.2. Dynamic Fixed Effect of total health expenditure per-capita on economic growth In Southern Africa

Table 5 shows that both total health per capita expenditure (THE PC) and Life expectancy at birth (LEB) have positive effects on LGDP PC at all levels of significance in the short term, indicating that a unit increase in THE PC and LEB increases LGDP PC by 32 and 0.02 percent, respectively. Furthermore, in the long run, THE PC has a positive influence on LGDP PC at all levels of significance, whereas LCHE PC hurts LGDP PC, implying that a percentage rise in THE PC increases LGDP PC by 0.28 percent, whilst LCHE PC decreases LGDP PC by 41.7 percent.

According to the findings, increasing per capita health spending and life expectancy at birth will boost the growth of the economy in Southern Africa in the short term. Similarly, an upbeat in per capita health spending will enhance the growth of economy in Southern Africa, however, an upbeat in per capita current health spending will lower the growth of the economy in the long run.

4.1.3. Pooled Mean Group effect of total health expenditure per-capita on economic growth In Southern Africa

Table 6 shows that life expectancy at birth (LEB) and total health per capita spending (THE PC) have upbeat effects on LGDP PC at all levels of significance in the short run, indicating that a short-run relationship exists

between THE PC, LEB, and LGDP PC, such that a unit increase in THE PC and LEB increases LGDP PC by 1 and 26%, respectively. Furthermore, in the long run, THE PC and LEB have upbeat effects on LGDP PC at 1% and 5% levels of significance, implying that a 1% rise in THE PC and LEB will improve the growth of the economy by 1% and 11%, respectively. The current health expenditure has a downbeat effect on LGDP PC, implying that a 1percent (%) boost in CHE PC reduces LGDP PC by 30%. According to the findings, increasing per capita health spending and existing health spending will boost the growth of economy in Southern Africa in the short run. Similarly, in the long run, a rise in life expectancy at birth and per capita health spending has a beneficial power on the growth of the economy in Southern Africa, whereas per capita current health spending has a downbeat outcome.

4.1.4. *The Mean Group and Pooled Mean Group effects (Hausman test)*

The Hausman test outcome is provided in *table 7a*, and the Chi-Square statistics of 2.89 with a P-Value of 0.4088 indicate that the pooled mean group model is the best model for the variables. As a result, the pooled mean group's outcomes are adopted and emphasized in this study.

4.1.5. *The fixed effect and pooled mean group dynamic for trade share model (The Hausman test)*

The Hausman test outcome is provided in *table 7b*, and the Chi-Square statistics of 4.823 with a P-Value of 0.73 indicate that the pooled mean group is the best model for the variables. As a result, the pooled mean group model outcome is adopted and emphasized in this work.

4.1.6. *The effect of total per capita health expenditure on economic growth across the West African region*

Mean Group Effect of Total Health spending Per-capita on the Growth of economy in West Africa

The mean group effect of total health spending per capita on the growth of the economy in West Africa is shown in *Table 8*. Based on the Z-statistics, the short-run result shows that the ECT value (-0.04145) is negative and notable at the 5% level of significance, indicating convergence to equilibrium. The short-run study demonstrates that LGE and LCHE PC have upbeat outcomes on LGDP PC at the 5% level of significance, implying that a percentage increase in LCHE PC and LGE will boost LGDP PC by 6.2 and 9.5 percent, respectively, in the short term.

Similarly, the long run study demonstrates that THE PC and LCHE PC have upbeat and significant outcomes on LGDP PC at the 5% level of significance, implying that a percentage rise in THE PC and CHE PC will enhance the

growth of the economy by 1.2 and 8.5 percent, respectively, in the long run. The findings show that increasing government spending and per capita, current health spending in West Africa will boost the growth of economy in the short run. Similarly, a rise in per capita current health spending and per capita health spending has a long-term favorable influence on West African growth of the economy.

The Fixed Effect Dynamic of total health expenditure per-capita on the growth of the economy in West Africa

Table 9 illustrates the long-run and short-run effects of the dynamic fixed effect of total health expenditure per capita on economic growth in Southern Africa. Based on the Z-statistics, the ECT value (-0.02517) is negative and significant at the 10% level of significance, indicating convergence to equilibrium. The short-run study demonstrates that current health per capita expenditure (LCHE PC) has a positive influence on LGDP PC at all levels of significance, implying that a percentage increase in LCHE PC will boost LGDP PC by 6.6% in the short run. Similarly, the long run study reveals that total per capita expenditure (THE PC) has a positive influence on economic growth (LGDP PC) at all levels of significance, implying that a percentage increase in THE PC increases LGDP PC by 1.3% in the long run. The findings show that increasing per capita current health spending in West Africa will boost the growth of economy in the short run. Similarly, a rise in per capita health expenditure has a long-term favorable influence on West African growth of the economy.

Pooled Mean Group outcome of overall health spending per-capita on the growth of the economy

Table 10 illustrates the long-run and short-run effects of the dynamic fixed effect of total health expenditure per capita on economic growth in Southern Africa. Based on the Z-statistics, the ECT value (-0.014) is negative and significant at the 5% level of significance, indicating convergence to equilibrium. At the 5% level of significance, current health per capita expenditure has an upbeat outcome on LGDP PC in the short run, implying that a percentage increase in LCHE PC increases LGDP PC by 7.7 percent in the short run. Similarly, in the long term, total per capita spending (LTHE PC) and life expectancy at birth (LEB) have a positive effect on LGDP PC at all levels of significance, showing that a 1% rise in THE PC and LEB will boost LGDP PC by 1.5 and 2.6 percent, respectively. As a result, a rise in per capita current health spending in West Africa is likely to boost the growth of economy in the short run. Similarly, a rise in life expectancy at birth and per capita health spending has a long-term favorable influence on West African economic growth.

The Mean-Group and Pooled-Mean Group effects (The Hausman test)

The Hausman test outcome is shown in *table 11a*, and the Chi-Square statistics of 0.63 with a P-Value of 0.422 indicate that the pooled mean group model is the best model for the variables.

The fixed effect and pooled mean group dynamic (The Hausman test)

The Hausman test findings are provided in *table 11b*; the Chi-Square statistics of -1.49 with a P-Value of 0.245 indicate that the pooled mean group model is the best model for the variables. The Hausman test results show that the mean group (MG) model is the best model for the variables.

5. Conclusion and Recommendation

In the short run, current health per capital expenditure and government expenditure has a positive effect on the growth of the economy (LGDP PC) at a 5% significant level in West Africa, while total and current health per capital expenditure has an upbeat outcome on the growth of the economy (LGDP PC) at a 5% significant level in the long run. The effect of total per capita health spending on the growth of the Southern African region was embraced and underlined in this study, as were the results of the pooled mean group model. According to the findings, increasing per capita health spending and existing health spending will boost the growth of the

economy in Southern Africa in the short run. Similarly, in the long run, an increase in life expectancy at birth and per capita health expenditure has a beneficial influence on economic growth in Southern Africa; however, a rise in per capita current health spending has a downbeat effect. The Hausman test outcome illustrates that the mean group (MG) model is the best model for the variables. According to the findings, increasing per capita current health spending will boost the growth of the economy in West Africa in the short run. Similarly, a boost in per capita health spending has a beneficial long-term effect. Furthermore, the findings indicated that a boost in per capita current health spending will boost the growth of the economy in West Africa in the short run, and an increase in per capita health spending has a favorable outcome on the growth of West Africa economy in the long run. Health spending per capita has a beneficial outcome on the growth of the economy; hence more funds should be committed to the health care industry to assure the quality of health services obtained by each patient.

6. Limitation of study

Owing to the magnitude of this study, finance, geographical distance and time was a limitation to the researcher as primary data could also be used for a research of this nature. Further studies can be carried out by taking each country in sub-Saharan Africa individually and studying each of the countries extensively by using primary data.

Table 1: Summary Statistics for Southern Africa and West Africa

Variable	Mean	Min	Max	Observations
SOUTHERN AFRICA				
GDP_PC	2898.817(2593.325)	272.991	10892.54	N = 200
CAH_GDP	0.312916(0.26757)	0.002817	1.196823	N = 55
CHE_GDP	5.896041(1.83989)	1.908599	11.57911	N = 180
THE_GDP	5.306437(2.48785)	0	11.57911	N = 200
THE_PC	183.5202(281.552)	0	1552.573	N = 200
PRHE_GDP	2.908607(1.82036)	0	9.49238	N = 200
PUHE_GDP	2.664256(1.28062)	0.438216	5.826442	N = 180
PRHE_PC	72.24588(82.8003)	1.775926	409.989	N = 180
PGR	1.894686(1.28693)	-2.62866	5.604957	N = 200
LEB	57.67017(10.1628)	42.595	78.49756	N = 200
LF	50.08885(16.1664)	25.11	74.58	N = 200
WEST AFRICA				
GDP_PC	2134.786(3232.933)	194.873	20532.98	N = 319
CAH_GDP	0.47801 (0.60799)	0	3.028301	N = 123
CHE_GDP	5.588473(2.04079)	2.49142	12.6201	N = 304
THE_GDP	5.30905 (2.33325)	0	12.6201	N = 320
THE_PC	83.51476(147.991)	0	829.9979	N = 320

PRHE_GDP	3.675908(1.87557)	0	9.199446	N = 320
PUHE_GDP	1.719096(1.22588)	0.240667	6.048561	N = 304
PRHE_PC	38.80165(52.9876)	3.470781	276.8144	N = 304
PGR	2.588873(0.72641)	-1.13152	5.363199	N = 320
LEB	58.96343(6.73067)	47.713	74.30976	N = 320
LF	46.6044(12.07798)	23.74	72.13	N = 320

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 2: Unit root Analysis

Variables	Fisher PP		Fisher ADF		IPS		Remark
	Level	First difference	Level	First difference	Level	First difference	
LGDP_PC	116.630 (0.0524)	370.19 (0.000)	86.399(0.645)	211.37 (0.000)	3.721(0.999)	-7.012 (0.000)	I (1)
CHE_GDP	110.597(0.091)	1195.32 (0.000)	87.006 (0.628)	271.78 (0.000)	0.377 (0.647)	-9.560 (0.000)	I (1)
CAH_GDP	64.483 (0.003)	169.27 (0.000)	43.722 (0.176)	61.42 (0.000)	-0.226 (0.411)	-2.791 (0.003)	I (1)
PUHE_PC	72.091 (0.938)	665.46 (0.000)	67.737 (0.973)	304.55 (0.000)	1.956 (0.975)	-11.194 (0.000)	I (1)
PRHE_PC	78.886 (0.871)	671.65 (0.000)	86.057(0.655)	259.45 (0.000)	0.389 (0.651)	-9.901 (0.000)	I (1)
CHE_PC	68.458(0.939)	483.33 (0.000)	75.862(0.819)	249.49 (0.000)	0.512 (0.696)	-8.787 (0.000)	I (1)
LLEB	51.399(0.999)	146.13 (0.000)	73.616 (0.920)	1098.14 (0.000)	3.203 (0.999)	-28.576 (0.000)	I (1)
LGE	166.615 (0.100)	-12.504 (0.000)	109.008(0.084)	324.49 (0.000)	1.461(0.072)	-12.504 (0.000)	I (1)

Table 3: Correlation Matrix

	GDP_PC	CAH_GDP	CHE_GDP	CHE_PC	THE_PC	PUHE_PC	PRHE_PC	PGR	LEB	LGE
GDP_PC	1.000									
CAH_GDP	0.606	1.000								
CHE_GDP	0.586	0.067	1.000							
CHE_PC	0.491	0.024	0.119	1.000						
THE_PC	0.771	-0.025	0.054	-0.001	1.000					
PUHE_PC	0.674	-0.043	0.093	0.007	0.971	1.000				
PRHE_PC	0.619	0.017	-0.035	-0.018	0.870	0.726	1.000			
PGR	0.597	0.064	-0.090	-0.132	-0.467	-0.497	-0.317	1.000		
LEB	0.467	0.002	-0.021	0.238	0.549	0.544	0.455	-0.551	1.000	
LGE	0.846	-0.030	0.167	0.130	0.313	0.301	0.277	-0.188	0.222	1.000

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 4: The Estimation of Mean Group: Error Correction Form (MG saved as Estimate outcome)

D.LGDP_PC	Coef.	Std. Err.	Z	P> z
SHORT RUN				
ECT	-0.23003	0.109746	-2.1	0.02
LGE D1.	-0.21825	0.187384	-1.16	0.244
LEB D1.	0.0959187	0.293352	3.27	0.000***
LTHE_PC D1.	0.0343	0.000739	4.64	0.000***
LCHE_PC D1.	0.222112	0.241819	0.92	0.358
Constant	-0.27208	0.848047	-0.32	0.748
LONG RUN				

LEB	0.77793	0.83487	3.31	0.000***
LTHE_PC	0.107556	0.032549	3.3	0.000***
LCHE_PC	2.127554	1.721452	1.24	0.216

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 5: The Fixed Effects Dynamic Regression: Estimated Error Correction Form (DFE saved as Estimate outcome)

	Coef.	Std. Err.	Z	P> z
SHORT RUN				
ECT	-0.01736	0.009332	-1.86	0.034**
LGE D1.	0.007078	0.026787	0.26	0.792
LEB D1.	0.319815	0.137688	2.32	0.02**
LTHE_PC D1.	0.000202	7.69E-05	2.63	0.009***
LCHE_PC D1.	0.012031	0.020047	0.6	0.548
Constant	0.141981	0.208745	0.68	0.496
LONG-RUN				
LEB	0.074634	2.650113	0.03	0.978
LTHE_PC	0.002756	0.000419	6.57	0.000***
LCHE_PC	-0.41654	0.076889	-5.42	0.000***

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 6: Pooled Mean Group Regression (PMG saved as Estimate outcome)

D.LGDP_PC	Coef.	Std. Err.	Z	P> z
SHORT RUN				
ECT	-0.037	0.010	-3.560	0.000
LGE D1.	-0.054	0.037	-1.480	0.139
LEB D1.	0.2609	1.500	1.740	0.082*
LTHE_PC D1.	0.010	0.000	7.140	0.000***
LCHE_PC D1.	-0.097	0.121	-0.800	0.423
Constant	0.054	0.024	2.250	0.024
LONG RUN				
LEB	0.111	0.546	2.040	0.042**
LTHE_PC	0.010	0.000	4.080	0.000***
LCHE_PC	-0.301	0.096	-3.160	0.002***
Log Likelihood	446.9231			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 7a: The Mean Group and Pooled Mean Group effects (Hausman test)

	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	Mg	PMG	Difference	S.E.
LEB	0.77793	1.110957	-0.33303	16.0969
LTHE_PC	0.107556	0.000797	0.106759	0.226766
LCHE_PC	2.127554	-0.30146	2.429014	2.557174
chi2(3)	2.89			
Prob>chi2	0.4088			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 7b: The fixed effect and pooled mean group dynamic (The Hausman test)

	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	DFE	PMG	Difference	S.E.
LEB	0.074634	1.110957	-1.03632	.

LTHE_PC	0.002756	0.000797	0.001959	.
LCHE_PC	-0.41654	-0.30146	-0.11508	.
chi2(3)	4.8223			
Prob>chi2	0.7300			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 8: The Estimation of Mean Group: Error Correction Form (Estimate results saved as MG)

D.LGDP_PC	Coef.	Std. Err	. z	P> z
SHORT RUN				
ECT	-0.04145	0.017833	-2.32	0.01
LGE D1.	0.061986	0.032891	1.89	0.038**
LEB D1.	0.035544	0.120284	0.3	0.768
LTHE_PC D1.	0.000128	0.002387	0.05	0.957
LCHE_PC D1.	0.094725	0.044426	2.13	0.031**
Constant	-0.69615	0.591278	-1.18	0.239
LONG RUN				
LEB	0.55827	0.49864	1.12	0.263
LTHE_PC	0.011608	0.016502	2.52	0.012**
LCHE_PC	0.084721	0.00799	2.31	0.022**

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 9: Fixed Effects Dynamic Regression: Estimated Error Correction Form (Estimate results saved as DFE)

	Coef.	Std. Err.	Z	P> z
SHORT RUN				
ECT	-0.02517	0.014125	-1.78	0.075*
LGE D1.	0.02083	0.019392	1.07	0.283
LEB D1.	-0.00247	0.008446	-0.29	0.77
LTHE_PC D1.	4.36E-05	0.000135	0.32	0.747
LCHE_PC D1.	0.065686	0.02014	3.26	0.001***
Constant	-0.00357	0.069703	-0.05	0.959
LONG RUN				
LEB	0.032263	0.062401	0.52	0.605
LTHE_PC	0.012712	0.003163	4.02	0.000***
LCHE_PC	-0.16407	0.439308	-0.37	0.709

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 10: Pooled Mean Group Regression (PMG saved as Estimate outcome)

D LGDP_PC	Coef.	Std. Err.	Z	P> z
SHORT RUN				
ECT	-0.014	0.007	-2.100	0.025**
LGE D1.	0.009	0.036	0.260	0.796
LEB D1.	0.012	0.044	0.280	0.778
LTHE_PC D1.	0.000	0.001	0.010	0.993
LCHE_PC D1.	0.077	0.033	2.320	0.020**
_cons	-0.186	0.232	-0.800	0.424
LONG RUN				
LEB	0.258	0.067	3.880	0.000***
LTHE_PC	0.015	0.005	3.100	0.002***
LCHE_PC	0.316	0.248	1.270	0.203
Likelihood	677.5314			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 11a: The Mean Group and Pooled Mean Group effects (The Hausman test)

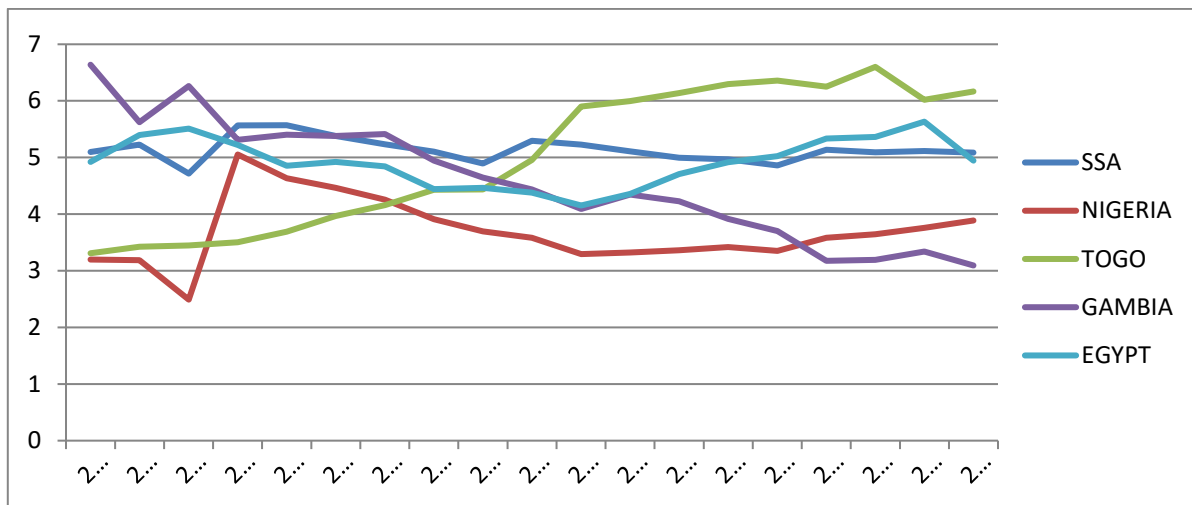
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	Mg	Pmg	Difference	S.E.
LEB	0.55827	0.258243	0.300027	0.679098
LTHE_PC	0.011608	-0.01477	0.026378	0.022074
LCHE_PC	0.084721	0.315681	-0.23096	0.950672
chi2(3)	0.63			
Prob>chi2	0.422			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database

Table 11b: The fixed effect and pooled mean group dynamic (The Hausman test)

	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	DFE	Pmg	Difference	S.E.
LEB	0.032263	0.258243	-0.22598	.
LTHE_PC	-0.01271	-0.01477	0.002058	.
LCHE_PC	-0.16407	0.315681	-0.47975	.
chi2(3)	-1.49			
chi2<prob	0.245			

Source: Author's Computation 2021, data from World Development Indicator (WDI) Database



Source: Author's compilation with data from World Development Indicator (WDI) 2021
Figure 1 Total current health expenditure 2000-2018

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] World Health Organisation, "State of financing in the African region Systems", *World Health Organization. Regional Office for Africa*, no. January, pp. 1-25, 2013.
- [2] X. Weibo, B. Yimer, "The Effect of Healthcare Expenditure on the Health Outcomes in Sub-Saharan African Countries", *Asian Journal of Economics, Business and Accounting*, vol. 12, no. 4, pp. 1-22, 2019, doi:10.9734/ajeba/2019/v12i430158.
- [3] N. M. Odhiambo, "Health expenditure and economic growth in sub-Saharan Africa: an empirical investigation", *Development Studies Research*, vol. 8, no. 1, pp. 73-81, 2021, doi:10.1080/21665095.2021.1892500.
- [4] B. Yimer Ali, R. Moracha Ogeto, "Healthcare Expenditure and Economic Growth in Sub-Saharan Africa", *Asian Journal of Economics, Business and Accounting*, vol. 13, no. 2, pp. 1-7, 2020, doi:10.9734/ajeba/2019/v13i230170.
- [5] S. M. Piabuo, J. C. Tieguhong, "Health expenditure and economic growth - a review of the literature and an analysis between the economic community for central African states (CEMAC) and selected African countries", *Health Economics Review*, vol. 7, no. 1, pp. 1-13, 2017, doi:10.1186/s13561-017-0159-1.
- [6] M. O. F. Commerce, "The Determinants of Government Expenditure in South Africa by Glenda Maluleke" no. November, 2016.
- [7] B. Aboubacar, D. Xu, "The Impact of Health Expenditure on the Economic Growth in Sub-Saharan Africa", *Theoretical Economics*

- Letters*, vol. 07, no. 03, pp. 615–622, 2017, doi:10.4236/tel.2017.73046.
- [8] V. V. Tarasova, V. E. Tarasov, "ECONOMIC GROWTH MODEL WITH CONSTANT PACE AND DYNAMIC MEMORY" vol. 2, no. 2, pp. 40–45, 2017, doi:10.20861/2304-2338-2017-84-001.
- [9] E. F. Nwaolisa, "THE IMPACT OF GOVERNMENT EXPENDITURE ON NIGERIA ECONOMIC GROWTH: A FURTHER DISAGGREGATED APPROACH" vol. 5, no. 2, 2017.
- [10] K. Tsaurai, "IS WAGNER ' S THEORY RELEVANT IN EXPLAINING HEALTH EXPENDITURE DYNAMICS IN BOTSWANA ?" vol. 3, no. 4, pp. 107–115, 2014.
- [11] O. E. Akpokerere, E. J. Ighoroje, "The Effect of Government Expenditure on Economic Growth in Nigeria : A Disaggregated Analysis from 1977 to 2009" vol. 4, no. 1, pp. 76–82, 2013.
- [12] R. S. Juma, H. Ouyang, J. Cai, "The effect of government expenditure on economic growth: The case of Tanzania using VECM approach" vol. 9, no. 2, pp. 144–155, 2018.
- [13] P. B. Frimpong, G. Adu, "Population Health and Economic Growth in Sub-Saharan Africa: A Panel Cointegration Analysis", *Journal of African Business*, vol. 15, no. 1, pp. 36–48, 2014, doi:10.1080/15228916.2014.881227.
- [14] S. Zaidi, K. Saidi, "Environmental pollution, health expenditure and economic growth in the Sub-Saharan Africa countries: Panel ARDL approach", *Sustainable Cities and Society*, vol. 41, pp. 833–840, 2018, doi:10.1016/j.scs.2018.04.034.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Faith O. Yusufu has done her bachelor's degree from Covenant University ota, Ogun state Nigeria in 2015. She has done her master's degree from University of Jos, Plateau state Nigeria in 2018. She is currently a PhD student in Afe-Babalola University Ado Ekiti, Ekiti state Nigeria. She presently lectures in the Economics Department of University of Jos as an Assistant lecturer and has published in a number of journals both locally and internationally.

Dr. Bosede O. Awoyemi earned her Bachelor's degree in Economics in 2006 from the University of Ado-Ekiti in Nigeria. In 2010, she received her Master's degree in Economics and her PhD in Economics from the University of Ibadan in Nigeria. In 2019-2021, she worked as a research fellow in the UNECA's macroeconomics and governance division. She is also a lecturer at Afe Babalola University Ado- Ekiti's Department of Economics, where she teaches microeconomics, health economics, and other economics-related courses. She has over 25 articles published in reputable journals. She serves as an external reviewer for a number of referred journals.

Dr. Kehinde J. Akomolafe earned his Bachelor's and Master's degree in Economics in Ekiti state university in 2004 and 2007 respectively and his PhD in Economics from the University of Ibadan Nigeria in 2016. He is also a lecturer at Afe-Babalola University Ado-Ekiti's Department of Economics. He has many articles published in reputable journals.

Physical Interpretation of the Solution to the Problem of Diffraction on a Half-plane with Non-Ideal Boundary Conditions

Michael Vesnik *

Kotelnikov Institute of Radioengineering and Electronics, Russian Academy of Sciences, Moscow, 125009, Russia

*Corresponding author: Michael Vesnik, vesnik@cplire.ru

ABSTRACT: The recently proposed method of fundamental components is employed to develop a technique for obtaining a heuristic solution to the problem of diffraction on a half-plane with non-ideal boundary conditions. The difference between the new method and traditional heuristic approaches, such as the geometric theory of diffraction and the method of edge waves, is the presence of an adjustment procedure which allows increasing the accuracy while maintaining the compactness of the formulas. For the case of the problem of diffraction on an impedance half-plane, heuristic formulas are constructed. Then they are refined using a verification solution. A quantification of accuracy is carried out, and a physical interpretation of the solution is presented. The prospects for applying this approach to constructing high-speed solvers and carrying out the physical interpretation of numerical solutions are discussed.

KEYWORDS: Electromagnetic Diffraction, Heuristic Approaches, Impedance Boundary Conditions

1. Introduction

Topical practical problems, such as the study of scattering on targets with low radar visibility, the propagation of electromagnetic waves in urban areas, etc., require the use of special solvers. These solvers, in turn, require efficient (i.e., fast and accurate) methods for solving diffraction problems [1].

In the topical problems mentioned above, which can be classified as diffraction by polygons and polyhedra, there are many edges with various shapes and boundary conditions. Computer resources may not be enough to apply numerical methods, and it is unlikely to obtain rigorous analytical solutions for all of them. One has to apply heuristic formulas [2 – 5]. The recently developed Method of Fundamental Components (MFC) [6 – 8] offers an efficient solution to this problem. Heuristic formulas obtained employing MFC are compact, fast, and, unlike known heuristic approaches, accurate in an entire range of parameters [8].

The MFC heuristic solution is based on a set of fundamental components [8]. At the first stage it is the solution of the simplest problems of diffraction. Further, these solutions are refined using other basic components

and a verification solution (usually, numerical or numerical-analytical one).

The MFC application in order to obtain heuristic formulas describing the solution to the problem of diffraction on a semitransparent half-plane is based on the following algorithm [6, 8, 9]:

- Development of a verification solution.
- Selection of primary heuristic formulas from the number of known solutions.
- Carrying out the adjustment procedure in order to refine the primary heuristic formulas by comparing them with the verification solution.

Further, when constructing the heuristic solution of the MFC, we will follow this sequence of actions.

One of the fundamental components is the singular diffraction coefficient for a semi-infinite scatterer [6, 8, 9]. Obtaining heuristic formulas for two-dimensional semi-infinite scatterers, even in spite of the existence of rigorous solutions, is a topical problem [9]. In this study we will analyze heuristic formulas for the problem of diffraction on an impedance half-plane and give a physical

interpretation of the verification by means of the Wiener-Hopf method (WHM) solution.

The present study is devoted to the technique of constructing heuristic solutions to diffraction problems on two-dimensional semi-infinite objects with non-ideal boundary conditions. Such solutions can be further applied to construct heuristic solutions of diffraction problems on three-dimensional objects. To build heuristic solutions, one can use any reliable solution, mostly numerical. The Wiener-Hopf method satisfies all the required conditions.

The previous works of the author were devoted to obtaining individual specific solutions. This article systematizes previous results, formulates a general methodology and calculates quantification of accuracy.

Let us consider two types of wave excitation: TH-polarization or TE-polarization (magnetic field vector \vec{H} or electrical field vector \vec{E} is perpendicular to the edge, correspondingly).

2. Construction of a heuristic solution for TH-polarization

2.1. Development of a verification solution

Consider the problem of electromagnetic wave diffraction on an impedance half-plane. The boundary conditions for the case of TH-polarization (vector \vec{H} is perpendicular to the edge) have the form [9 – 16]:

$$\begin{cases} H_{x+} - H_{x-} = -Z^{-1}E_z \\ E_{z+} = E_{z-} = E_z \text{ when } y = 0 \end{cases}, \quad (1)$$

where impedance $Z=iX$, X is a variable parameter, which the reflection coefficient depends on, and i is an imaginary unit. Here the signs "+" and "-" correspond to the regions $y>0$ and $y<0$, respectively (see Figure 1).

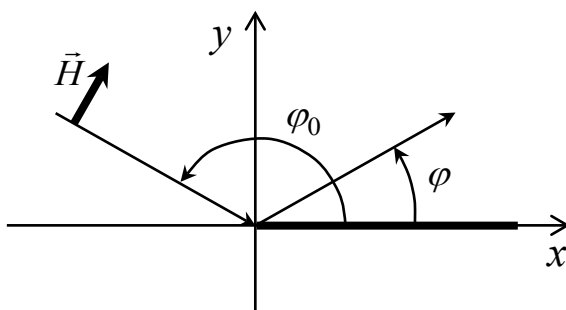


Figure 1: The geometry of the problem.

The reflection and transmission coefficients R_{TH} and T_{TH} for an unbounded half-plane with boundary conditions (1) on the surface can be written as follows:

$$R_{TH}(X, \varphi_0) = \frac{-W_0}{W_0 + 2i \sin(\varphi_0)X}, \quad W_0 = 120\pi, \quad (2)$$

$$T_{TH} = 1 + R_{TH}. \quad (3)$$

The verification solution:

$$fr(X, \varphi, \varphi_0) \quad (4)$$

is obtained employing the Wiener-Hopf method [9 – 13].

2.2. Selection of primary heuristic formulas

We choose primary heuristic formulas as in [8]:

$$fg(R, T, \varphi, \varphi_0) = \frac{1}{2} \left(\frac{1-T}{-\cos \frac{\varphi-\varphi_0}{2}} + \frac{R}{-\cos \frac{\varphi+\varphi_0}{2}} \right), \quad (5)$$

$$fg^0(R, T, \varphi, \varphi_0) = \frac{1}{2} \left[\frac{(1-T) \sin \frac{\varphi-\varphi_0}{2}}{-\cos \frac{\varphi-\varphi_0}{2}} + \frac{R \sin \frac{\varphi+\varphi_0}{2}}{-\cos \frac{\varphi+\varphi_0}{2}} \right]. \quad (6)$$

In (5) and (6): $R=R_{TH}$, $T=T_{TH}$. When $X=0$: $R_{TH}=-1$, $T_{TH}=0$, and expressions (5), (6) are transformed into known expressions f, f^0 from [17 – 19]. Formula (5) corresponds to the generalized diffraction coefficients (GDC) approximation, and formula (6) corresponds to the physical optics diffraction coefficients (PODC) approximation [8].

2.3. Carrying out the adjustment procedure

The adjustment procedure consists in comparing the heuristic formula (5) with the Wiener-Hopf solution (4) [9]. At the first stage, we compare the functions (4) and (5) and empirically find the semitransparency function depending on (X, φ) [10, 11], which allows us to bring the primary formula closer to a rigorous solution:

$$cxf(X, \varphi) = 1 - x(X) \left\{ 1 - \cos \left[\frac{\pi - \varphi}{1 + x(X)} \right] \right\}, \quad (7)$$

$$x(X) = 1 - \exp \left(-\frac{X}{W_0} \right), \quad (8)$$

$$fh(R, T, \varphi, \varphi_0) = fg(R, T, \varphi, \varphi_0) \frac{cxf(X, \varphi_s)}{cxf(X, \varphi)}. \quad (9)$$

Here, $cxf(X, \varphi_s)$ is the function $cxf(X, \varphi)$ value at the singularity point angle $\varphi_s = \pi - \varphi_0$, that is disposed on the boundary between light and shadow.

3. Construction of a heuristic solution for TE-polarization

3.1. Development of a verification solution

The boundary conditions for the case of the incidence of a TE-polarized wave (vector \vec{E} is perpendicular to the edge) on an impedance half-plane have the form [9 – 16]:

$$\begin{cases} Z(H_{z+} - H_{z-}) = E_x \\ E_{x+} = E_{x-} = E_x \text{ when } y = 0 \end{cases} \quad (10)$$

The geometry of the problem is depicted in Figure 2.

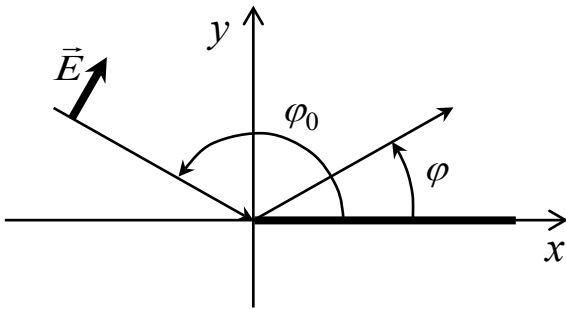


Figure 2: The geometry of the problem

The reflection and transmission coefficients R_{TE} and T_{TE} for an unbounded half-plane with boundary conditions (10) on the surface can be written as follows:

$$R_{TE}(X, \varphi_0) = \frac{W_0}{W_0 + 2i[\sin(\varphi_0)]^{-1}X}, \quad W_0 = 120\pi, \quad (11)$$

$$T_{TE} = 1 - R_{TE}. \quad (12)$$

The verification solution:

$$gr(X, \varphi, \varphi_0). \quad (13)$$

is obtained employing the Wiener-Hopf method [9 – 13].

3.2. Selection of primary heuristic formulas

Primary heuristic formulas we choose as in [8]:

$$gf(R, T, \varphi, \varphi_0) = \frac{1}{2} \left(\frac{1-T}{-\cos \frac{\varphi-\varphi_0}{2}} + \frac{R}{-\cos \frac{\varphi+\varphi_0}{2}} \right), \quad (14)$$

$$gf^0(R, T, \varphi, \varphi_0) = \frac{1}{2} \left[\frac{(1-T)\sin \frac{\varphi-\varphi_0}{2}}{-\cos \frac{\varphi-\varphi_0}{2}} + \frac{R \sin \frac{\varphi+\varphi_0}{2}}{-\cos \frac{\varphi+\varphi_0}{2}} \right]. \quad (15)$$

In (14) and (15): $R=R_{TE}$, $T=T_{TE}$. When $X=0$: $R_{TH}=1$, $T_{TH}=0$, and expressions (14), (15) are transformed into known expressions g , g^0 from [17 – 19]. The difference between formulas (14), (15) and formulas (5), (6) is that we substitute different values of R and T into each one. Formula (14) corresponds to the GDC approximation, and formula (15) corresponds to the PODC approximation [8].

3.3. Carrying out the adjustment procedure

Just as for the case of the TH polarization described in sub-section 2.3, we empirically select the semitransparency function depending on (X, φ) [10, 11] for the TE- polarization:

$$\begin{aligned} cxg(X, \varphi) &= [\sin(\varphi/2)]^{x(X)}, \\ x(X) &= 1 - \exp(-X/W_0). \end{aligned} \quad (16)$$

Heuristic formula for TE- polarization:

$$gh(R, T, \varphi, \varphi_0) = gf(R, T, \varphi, \varphi_0) \frac{cxg(X, \varphi)}{cxg(X, \varphi_s)}. \quad (17)$$

Here also $cxg(X, \varphi_s)$ is the function $cxg(X, \varphi)$ value at the singularity point angle $\varphi_s = \pi - \varphi_0$, that is disposed on the boundary between light and shadow.

4. Calculation results

Figure 3 depicts the calculation results for TH-polarization at $\varphi=90^\circ$ and different values of X .

A study of the behavior of the functions whose graphs are depicted in Figure 3 showed that as the parameter X changes from 0 to infinity, the shapes of curves GDC (5) and PODC (6) remain unchanged, while curves (4) and (9) shift from the curve (5) to curve (6).

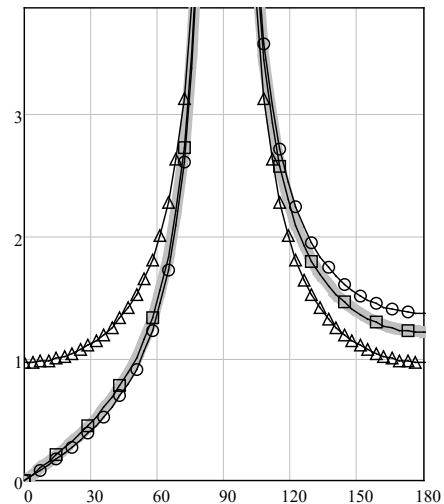


Figure 3 (a): $X=50$, $x(X)=0.124$

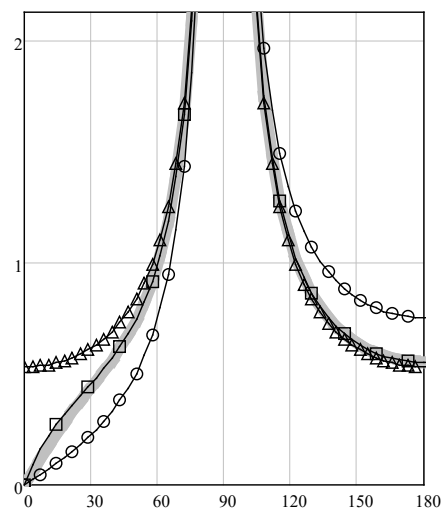


Figure 3 (b): $X=300$, $x(X)=0.549$

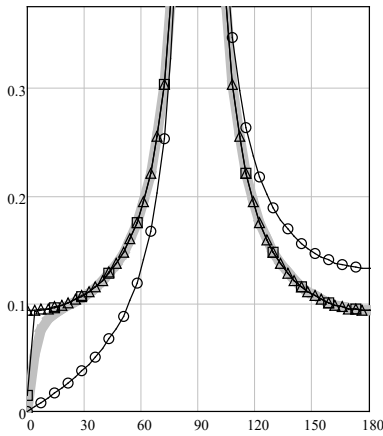


Figure 3 (c): Ez X=2000 x(X)=0.995

Figure 3: The TH-polarization scattering diagrams modules for $\varphi_0 = 90^\circ$. The angle φ is plotted horizontally. The scattering diagrams (SD) modules are plotted vertically. The solid gray line is WHM solution $fr(X, \varphi, \varphi_0)$ (4), circles are the GDC formula (5), triangles are the PODC formula (6), squares are the heuristic solution $fh(R, T, \varphi, \varphi_0)$ (9).

A study of the behavior of the functions whose graphs are depicted in Figure 4 showed that as the parameter X changes from 0 to infinity, the shapes of curves GDC (14) and PODC (15) remain unchanged, while curves (13) and (17) shift from the curve (14) to curve (15).

Figure 4 illustrates the calculation results for TE-polarization at $X=100, \varphi=90^\circ$.

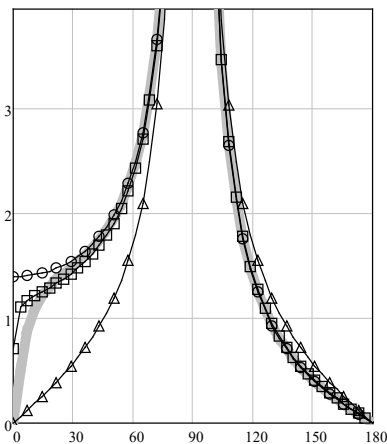


Figure 4 (a): X=30, x(X)=0.076

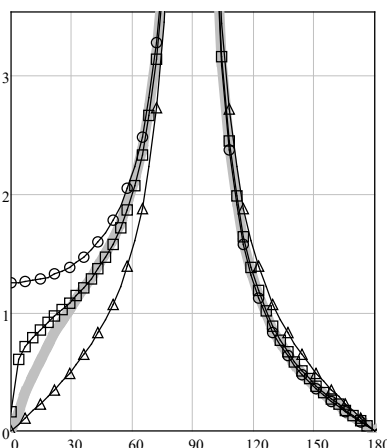


Figure 4 (b): X=100, x(X)=0.233

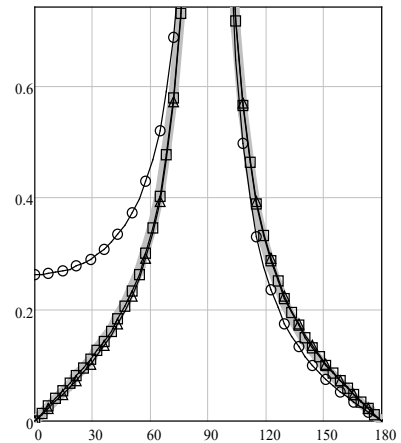


Figure 4 (c): X=1000, x(X)=0.93

Figure 4: The TE-polarization scattering diagrams modules for $\varphi_0 = 90^\circ$. The angle φ is plotted horizontally. The scattering diagram (SD) modules are plotted vertically. The solid gray line is WHM solution $gr(X, \varphi, \varphi_0)$ (13), the circles are the GDC formula (14), triangles are the PODC formula (15), squares are the heuristic solution $gh(R, T, \varphi, \varphi_0)$ (17).

5. Quantification of accuracy

Quantification of accuracy was carried out employing formulas, which we will call "functions L ":

$$Lfh(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{fr(X, \varphi, \varphi_0)}{fh(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi, \quad (18)$$

$$Lfg(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{fr(X, \varphi, \varphi_0)}{fg(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi, \quad (19)$$

$$Lfg^0(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{fr(X, \varphi, \varphi_0)}{fg^0(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi, \quad (20)$$

$$Lgh(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{gr(X, \varphi, \varphi_0)}{gh(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi, \quad (21)$$

$$Lgf(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{gr(X, \varphi, \varphi_0)}{gf(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi, \quad (22)$$

$$Lgf^0(R, T, \varphi, \varphi_0) = \lg \left(\frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \left| \frac{gr(X, \varphi, \varphi_0)}{gf^0(R, T, \varphi, \varphi_0)} \right| - 1 \right) d\varphi. \quad (23)$$

Each of the functions L is a quantitative integral estimate of the accuracy of the function in the denominator of the integrand on a logarithmic scale. The smaller the value of a function L is, the closer the corresponding function in the denominator of the integrand is to the verification solution.

Graphs quantitatively describing the deviations of heuristic approximations from the rigorous solution for TH- polarization are depicted in Figure 5.

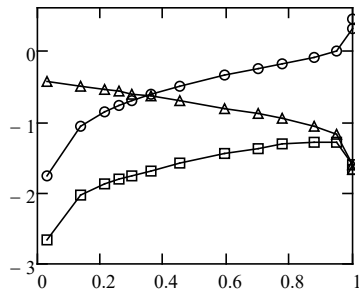


Figure 5: The TH-polarization accuracy quantification graphs for $\varphi_0=90^\circ$. The values of the function $x(X)$ are plotted horizontally. The amplitudes of the functions describing the modulus of the mean deviation are plotted along the vertical. The squares are the heuristic solution $Lfh(R, T, \varphi, \varphi_0)$ (18), the circles are the $Lfg(R, T, \varphi, \varphi_0)$ (19), the triangles are the $Lfg^0(R, T, \varphi, \varphi_0)$ (20).

Graphs quantitatively describing the deviations of heuristic approximations from the rigorous solution for TE- polarization are depicted in Figure 6.

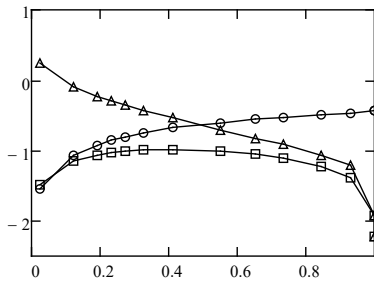


Figure 6: The TE-polarization accuracy quantification graphs for $\varphi_0=90^\circ$. The values of the function $x(X)$ are plotted horizontally. The amplitudes of the functions describing the modulus of the mean deviation are plotted along the vertical. The squares are the heuristic solution $Lgh(R, T, \varphi, \varphi_0)$ (21), the circles are the $Lgf(R, T, \varphi, \varphi_0)$ (22), the triangles are the $Lgf^0(R, T, \varphi, \varphi_0)$ (23).

6. Physical interpretation of the solution

As X changes from 0 to ∞ , the following changes occur: the expression $fh(R, T, \varphi, \varphi_0)$ (9) transforms from the formula $fg(R, T, \varphi, \varphi_0)$ (5) to the formula $fg^0(R, T, \varphi, \varphi_0)$ (6), and the expression $gh(R, T, \varphi, \varphi_0)$ (17) transforms from the formula $gf(R, T, \varphi, \varphi_0)$ (14) to the formula $gf^0(R, T, \varphi, \varphi_0)$ (15).

This means that the greater the transparency of the object is, the more the formulas for the GDC approximation (5), (14) are transformed into formulas for the PODC approximation (6), (15). This property can be used in the future to build primary heuristic formulas for semi-transparent objects.

Thus, none of the approximations GDC or PODC taken separately gives satisfactory solution in the entire range of values of the parameter X [8, 9, 11].

7. Non-transparent half-plane

As an additional illustration of the capabilities of the MFC, we present the calculated data for solving the problem of diffraction of a TH-polarized electromagnetic

wave on a non-transparent half-plane [11, 20]. Generalized bilateral impedance boundary conditions (GBIBC), for which boundary conditions (1) are a special case, lead to the following expressions for the diffraction coefficient:

$$fgH(R, T, \varphi, \varphi_0) = \frac{1}{2} \left(\frac{1}{-\cos \frac{\varphi - \varphi_0}{2}} + \frac{1 - 2 \arg(R)/\pi + C}{-\cos \frac{\varphi + \varphi_0}{2}} \right). \quad (24)$$

Here R is the reflection coefficient depending on the parameter X , similar to X [11, 20], C is a small parameter for additional accuracy correction. In further work in accordance with the MFC concept on the heuristic formula (24), in order to apply it in a specific practical problem, the parameter C should either be distributed over other components of the heuristic formula (24), or turned into a basic component that has a physical meaning. In accordance with the GBIBC, in (24) for all X values [20]: $|R|=1, T=0$, i.e. the half-plane is completely non-transparent, and only the phase of the reflection coefficient R changes. The graphs of the behavior of the function (24) are depicted in Figure 7.

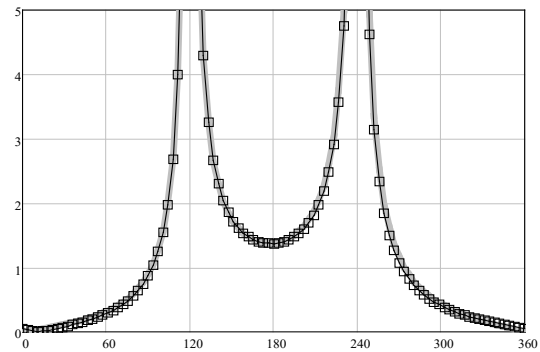


Figure 7 (a): $X1=20, \varphi_0=60, C=0$

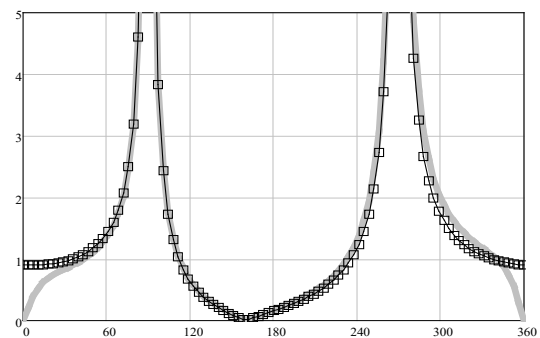


Figure 7 (b): $X1=400, \varphi_0=90, C=0$

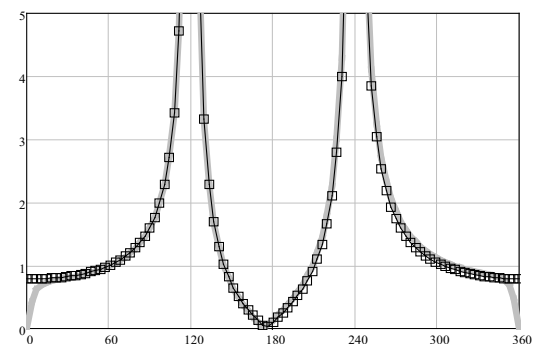


Figure 7 (c): $X1=2000, \varphi_0=60, C=0$

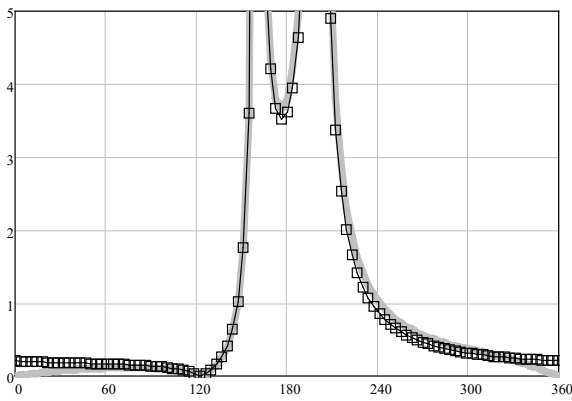


Figure 7 (d): $X1=200, \varphi_0=20, C=0.0457$

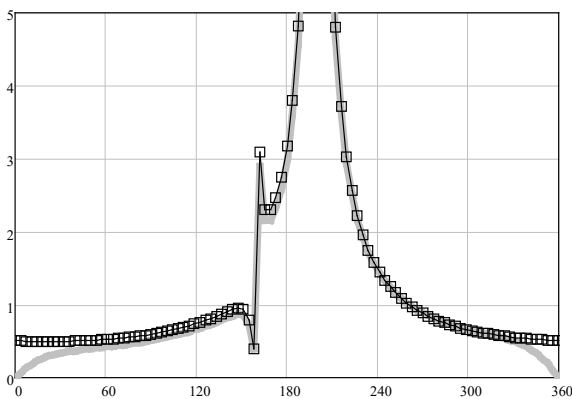


Figure 7 (e): $X1=500, \varphi_0=20, C=0.012$

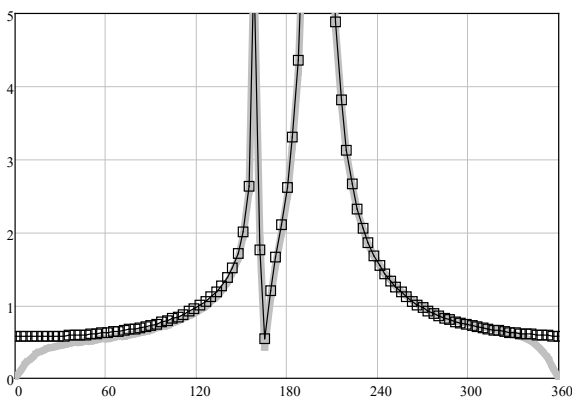


Figure 7 (f): $X1=700, \varphi_0=20, C=0.012$

Figure 7: The TH-polarization scattering diagrams modules. The angle φ is plotted horizontally. The scattering diagrams (SD) modules are plotted vertically. The solid gray line is WHM solution [11, 20], the squares are the heuristic solution $f_g H(R, T, \varphi, \varphi_0)$ (24).

7.1. Physical interpretation of the heuristic solution of the problem of diffraction on a non-transparent half-plane

For TH-polarization, as a result of the change in the phase of the reflection coefficient R as the parameter similar to X changes from 0 to ∞ , the transformation takes place: $f \rightarrow g$ [11, 20].

8. Discussion

In this article, heuristic formulas are obtained for solving problems of diffraction on a semitransparent half-plane. A refinement over the GDC and PODC approximations was found employing the method of

fundamental components (MFC). Heuristic formulas similar to GDC have been proposed many years ago [21, 22], although they were applied without an adjustment procedure.

Obtaining refined heuristic solutions for two-dimensional scatterers with non-ideal boundary conditions is a topical problem [23–26]. Such solutions are obtained even in the presence of reliable numerical solutions that can be used for verification of the mentioned heuristic formulas. The purpose of obtaining refined heuristic solutions is to use them in solvers designed for resource-intensive practical problems.

Together with heuristic approaches that describe the shape of the contour and edge profile, as well as the influence of the far zone condition [6, 8], the approach that describes the influence of boundary conditions complements the set of fundamental components that allows you to build effective heuristic formulas for three-dimensional objects of complex shape and with various boundary conditions.

In addition to refining the primary heuristic formulas, the MFC can be used to correct the numerical solution based on the experimental results. Such a need may arise in those cases when, due to inaccurate input data, the numerical calculation diverges from the experimental results. Engineering approaches of this kind are already widely used, however, unlike MFC, they are usually not based on primary heuristic formulas which have a physical meaning that helps to understand the reasons for the discrepancies between theory and experiment.

The MFC approach can be used to calculate diffraction for large 3D objects, for which it is impossible to obtain a numerical solution. The procedure for adjusting the heuristic solution is carried out on a smaller size scatterer. Having achieved the required accuracy, one can further apply the heuristic formula for a larger scatterer, which can no longer be calculated by rigorous numerical methods. In this case, the performance of heuristic formulas will not change, and the accuracy will even increase compared with the smaller size scatterer.

The direction of future research is to apply the technique proposed in this article to objects of different shapes and with different boundary conditions. In the future, such solutions can significantly increase the speed of solvers and give them new opportunities for solving practical problems.

9. Conclusions

The article proposes a technique for improving the accuracy of primary heuristic formulas employing the adjustment procedure. The technique has been verified for boundary conditions of certain types. However, this approach can be just as well applied to a half-plane with arbitrary boundary conditions, provided that a verification solution is available.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The author is grateful to Dr. Sci. S.E. Bankov for providing programs for verification and analytical expressions for R , T .

References

- [1] U. Jakobus, "Overview of hybrid methods in FEKO: Theory and applications," 2010 International Conference on Electromagnetics in Advanced Applications, 2010, pp. 434-437, doi: 10.1109/ICEAA.2010.5653159
- [2] Hönl, H.; Maue, A.W.; Westpfahl, K. *Handbuch der Physik*; Flüggé, S., Ed.; Springer: Berlin/Heidelberg, Germany, 1961; Mir: Moscow, Russia, 1964; Volume 25/1, p. 218.
- [3] Kravtsov, Y.A.; Zhu, N.Y. *Theory of Diffraction: Heuristic Approaches*; Alpha Science Int.: Oxford, UK, 2010
- [4] James, G.L. *Geometrical Theory of Diffraction for Electromagnetic Waves*; Peter Peregrinus Ltd.: London, UK, 1976.
- [5] Borovikov, V.A.; Kinber, B.E. *Geometrical Theory of Diffraction*; IEEE: London, UK, 1994.
- [6] M.V. Vesnik. *The Method of the Generalized Eikonal. New Approaches in the Diffraction Theory* (Walter de Gruyter, Berlin, 2015).
- [7] M.V. Vesnik MV. Physical Interpretation of Strict Solutions of Diffraction Problems by Heuristic Relations. *Journal of Mathematical Sciences*, 239, 751–770 (2019). <https://doi.org/10.1007/s10958-019-04324-8>
- [8] M.V. Vesnik, "New Possibilities for Constructing Heuristic Solutions to Problems of Electromagnetic Diffraction", *Eng* 2022, 3, 27-41. <https://doi.org/10.3390/eng3010004>
- [9] M.V. Vesnik, S.E. Bankov, "Heuristic solution to the problem of diffraction of a TE-polarized electromagnetic wave on a semitransparent half-plane", *Waves in Random and Complex Media*, July 2021, DOI 10.1080/17455030.2021.1951888
- [10] M.V. Vesnik, "Using the Method of Base Components for a Heuristic Solution to the Diffraction Problem on a Half-Plane with Nonideal Boundary Conditions", *J. Commun. Technol. Electron.* 2019, 64, 1211–1217
- [11] M.V. Vesnik, "Physical Interpretation of the Solution to the Problem of Diffraction on the Impedance Half-plane", *2020 7th All-Russian Microwave Conference (RMC)*, Moscow, Russia, 2020, pp. 200-202, DOI: 10.1109/RMC50626.2020.9312342
- [12] S.E. Bankov, *Integrated Microwave Optics*, (Fizmatlit, Moscow, 2018).
- [13] B. Noble, *Methods based on the Wiener – Hopf Technique for the Solution of Partial Differential Equations*, London, 1958.
- [14] Senior, T.B.A. Half plane edge diffraction. *Radio Sci.* 1975, 10, 645.
- [15] Senior, T.B.A. Diffraction tensors for imperfectly conducting edges. *Radio Sci.* 1975, 10, 911.
- [16] Voitovich, N.N.; Katsenelenbaum, B.Z.; Korshunova, E.N.; Pangonis, L.I.; Pereyaslavets, M.L.; Sivov, A.N.; Shatrov, A.D. *Electromagnetics of Antennas with Semitransparent Surfaces. A Method of Constructive Synthesis*; Nauka: Moscow, Russia, 1989.
- [17] P.Y. Ufimtsev, *Method of Edge Waves in the Physical Theory of Diffraction* (Sovetskoe Radio, Moscow, 1962; US Air Force Foreign Technology Division, 1-1154, 1962).
- [18] Ufimtsev, P.Y. *Theory of Edge Diffraction in Electromagnetics*; Tech Science Press: Encino, CA, USA, 2003. ISBN 0-9657001-7-8.
- [19] Ufimtsev, P.Y. *Fundamentals of the Physical Theory of Diffraction*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007.
- [20] S.E. Bankov, M.V. Vesnik, V.F. Kravchenko, "Heuristic Solution to the Diffraction Problem on a Superconducting Half-Plane", *Journal of Communications Technology and Electronics*, 2020, Vol. 65, No. 4, pp. 398–405.
- [21] Raman CV, Krishnan KS. The Diffraction of Light by Metallic

Screens, *Proc. R. Soc. Lond. A* 1927 Vol.116 254-267; DOI: 10.1098/rspa.1927.0135. Published 1 October 1927

- [22] Shmoys J. Diffraction by a Half-Plane with a Special Impedance Variation, *IRE Trans. on Antennas and Propagation* Vol.7, No.5, December 1959, pp.88 – 90
- [23] Luebbers RJ. "Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path loss", *IEEE Transactions on Antennas and Propagation*, vol. AP-32, pp. 70–76, Jan. 1984
- [24] Holm PD. A New Heuristic UTD Diffraction Coefficient for Nonperfectly Conducting Wedges, *IEEE Transactions on Antennas and Propagation*, vol. 48, NO. 8, August 2000
- [25] El-Sallabi HM, Rekanos IT, Vainikainen P. "A New Heuristic Diffraction Coefficient for Lossy Dielectric Wedges at Normal Incidence", *IEEE Antennas and Wireless Propagation Letters*, Vol. 1, 2002, pp. 165-168
- [26] Soni S, Bhattacharya A. «New heuristic diffraction coefficient for modeling of wireless channel», *Progress In Electromagnetics Research C*, Vol. 12, 125 – 137, 2010

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



MICHAEL V. VESNIK has done his bachelor's degree from the Moscow Institute of Electronic machine Building. He has done his master's degree from the Physical faculty of Moscow State University. He has completed his PhD degree in Kotelnikov IRE RAS in 2006. He has completed his Dr. Sci degree in Kotelnikov IRE RAS in 2019.

Research interests: electromagnetic diffraction, elastic waves diffraction, diffraction theory, heuristic approaches, geometrical theory of diffraction, physical theory of diffraction, diffraction on polygons and polyhedra, wave propagation in urban environment, impedance boundary conditions, antenna theory.

An Evaluation of 2D Human Pose Estimation based on ResNet Backbone

Hai-Yen -Tran¹, Trung-Minh Bui², Thi-Loan Pham³, Van-Hung Le^{*,2}

¹Tan Trao University, Tuyen Quang, 22000, Vietnam

²Vietnam Academy of Dance, HaNoi, 100000, Vietnam

³Hai Duong College, HaiDuong, 02203, Vietnam

*Corresponding author: Van-Hung Le, Tuyen Quang province, Email: van-hung.le@mica.edu.vn

Corresponding author ORCID: <https://orcid.org/0000-0003-4302-0581>

ABSTRACT: 2D Human Pose Estimation (2D-HPE) has been widely applied in many practical applications in life such as sports analysis, medical fall detection, human-robot interaction, using Convolutional Neural Networks (CNNs), which has achieved many good results. In particular, the 2D-HPE results are intermediate in the 3D Human Pose Estimation (3D-HPE) process. In this paper, we perform a study to compare the results of 2D-HPE using versions of Residual Network (ResNet/RN) (RN-10, RN-18, RN-50, RN-101, RN-152) on HUman 3.6M Dataset (HU-3.6M-D). We transformed the original 3D annotation data of the Human 3.6M dataset to a 2D human pose. The estimated models are fine-tuning based on two protocols of the HU-3.6M-D with the same input parameters in the RN versions. The best estimate has an error of **34.96** pixels with Protocol #1 and **28.48** pixels with Protocol #3 when training with 10 epochs, increasing the number of training epochs reduces the estimation error (15.8 pixels of Protocol #1, 12.4 pixels of Protocol #3). The results of quantitative evaluation, comparison, analysis, and illustration in the paper.

KEYWORDS 2D Human Pose Estimation, Residual Networks backbone, Human 3.6M Dataset, Convolutional Neural Networks

1. Introduction

Human pose estimation is defined as the process of localizing joints of humans in the 2D or 3D space (also known as keypoints - elbows, wrists, etc). Estimating human pose from the captured images/video has two research directions: 2D-HPE and 3D-HPE. If the output is a human pose on images or videos then this problem is called 2D-HPE. If the output is a human pose on 3D space then is called 3D-HPE. Therefore, a lot of research on this issue in the last 5 years. The results of human pose estimation are applied in many fields such as sports analysis [1, 2]; medical fall event detection [3]; identification and analysis in traditional martial arts [4]; robot interaction, construction of actions and movements of people in the game [1]. The 2D-HPE is an intermediate result for the 3D-HPE. The 3D-HPE result is highly dependent on the 2D-HPE result when based on the approach of [5]. To build a complete system, it is necessary to evaluate and compare the results at each step as in the studies of [6]-[7] for building a System on Chip. The authors made a test scheduling the algorithms on Chip.

Currently, many studies on 3D-HPE use 2D-HPE results on color images as an intermediary to estimate 3D human pose [8]-[9]. These studies are often grouped into the "2D

to 3D Lifting Approaches" [10].

Estimating 2D human pose based on deep learning has two methods: The first is the regression methods, which applied a deep network to learn joints location from the input ground-truth joints on the images to body joints or parameters of human body models/human skeleton to predict the key points on the human; The second method predicts the approximate locations of body parts. Deep learning network has achieved remarkable results for the estimation task. In which, all skeletal keypoints are regressed based on the ground-truth heatmaps (2D keypoints) by 2D Gaussian kernels [11]-[12]. In particular, the 2D keypoint estimation from the heatmap is shown in stacked hourglass networks [13] as start-of-the-art. However, it still faces many challenges such as heavy occlusion, partially visible human body. RN [14] is one of the backbones with the best results in feature extraction of ImageNet datasets and is used in many CNNs to detect, segment, recognize the objects, and estimate pose (as presented in Figure 1th [15]). In this paper, we experiment to compare the estimation of 2D human pose based on studies using the CNNs to estimate 2D human pose according to the regression methods. We use different versions of RN for 2D-HPE. The training model is based on RN-10, RN-18, RN-50, RN-101, RN-152. The

results of 2D human pose prediction are evaluated on the benchmark HU-3.6M-D, which is a widely used and challenging dataset, the body parts of the human are obscured. To get the 2D human pose annotation data of the HU-3.6M-D for the 2D-HPE, we perform an inverse transformation from the 3D pose annotation of human in the Real-World Coordinate System (R-WCS) of MOCAP system to 2D pose annotation of human according to image coordinates based on the set of intrinsic parameters provided for calibration of the image data. The results are presented in the following part of the paper.

In this paper, we have some contributions as follows:

- We have fine-tuned different versions of the RN with the size (224×224) of input data to estimate the 2D human pose in the RGB image.
- We have fine-tuned the estimated model on the HU-3.6M-D, with 2D pose ground-truth determined based on 3D pose annotation data and the intrinsic parameters of the camera.
- We evaluate the estimated results based on the absolute estimated coordinates between the original data and the estimated data. From there, choose the best RN version with input data of 224×224 for 2D-HPE on the RGB image, and will have good results in 3D-HPE.

The paper is organized as follows. Section 1 introduces several backbones for detecting and estimating people on images. Section 2 presents the related studies on 2D-HPE methods. Section 3 presents the main idea and versions of RN. Section 4 shows and discusses the experimental results of 2D human pose estimation, and Section 5 concludes the paper and future works.

2. Related Works

RN [14] is a backbone applied to many CNNs for feature extraction and object prediction in the first step such as Fast R-CNN [16], Faster R-CNN [17], Mask R-CNN [18], etc. Figure 1 shows the RN as the backbone in the Mask R-CNN network architecture. RN [14] is more efficient than other backbones like AlexNet [19], VGG [20], [21].

2D-HPE from RGB image data using CNN can be done by two methods [10]: regression methods, body part detection methods.

The regression methods use the CNNs model to learn joints location from the input ground-truth joints on the images to body joints or parameters of human body models/human skeleton to predict the key points on the human. In [22], the author proposed a Deep Neural Network (DNN) based on the cascade technique for regressing the location of body joints. The proposed CNN includes seven layers, the input image size of CNN is resized to 220×220 pixels. The cascade of pose regressors technique is applied to train the multi-layer prediction model. The first stage is the cascade starts with the initial position predicted over the entire input image. In the next stage, Deep Neural Networks regressors are trained to predict a displacement of

the joint locations with the correct locations in the previous stage. Thus, the currently predicted pose is refined based on each subsequent stage. In [23], a strategy of compositional pose regression based on the RN-50 [14]. The bones are parameterized and bone-based representation that contains human skeleton information and skeleton structure but did not use joint-based representation. The loss function is calculated based on each part of the human body, the joints are defined based on a constant origin point in the image coordinate system J_0 . Each bone has a directed vector pointing from it to its parent. In [24], it is proposed a regression method that used two Soft-argmax functions (Block-A and Block-B) for 2D human pose estimation from images, Block-A provides refined features and Block-B provides skeleton-part and active context maps. Two blocks are used to build one prediction block. Block-A used a residual separable convolution, the input feature maps are transformed into part-based detection maps and context maps by Block-B.

As for the body-part detection methods, a body-part detector is trained to predict the locations of human joints. In [13], it is proposed the stacked hourglass architecture for the training model to predict the positions of body joints on the heatmap in which the 2D annotation is used to generate the heatmap by 2D Gaussian heatmap method. The stacked hourglass repeats the bottom-up and top-down processing with intermediate supervision with the eight hourglasses. This CNN used the convolutional and max-pooling layers at a very low resolution and used then the top-down sequence of upsampling (the nearest neighbor upsampling of the lower resolution) and a combination of features across scales. The results of 2D-HPE on the MPII dataset based on the (PCKh@0.5) measurement are 90.9%, 99.0%, 97.0% are the results of the FLIC dataset on the (PCK@0.2) measurement.

In [25], it is proposed a two-branch CNN model, the body part detection is predicted from heatmaps by using the 2D keypoints annotation to generate the ground-truth confidence maps. The confidence maps are predicted by the first branch and the part affinity fields are predicted by the second branch. The part affinity fields are a novel feature representation of both location and orientation information across the limb's active region.

Most of the above studies were evaluated on the COCO [26], MPII [27] datasets and evaluated on the Percentage of Correct Keypoints $PCK - \%$ measure. This measure is usually based on the estimated joint length with the root joint length, without taking into account the absolute estimates of the 2D keypoints (estimated absolute coordinates and ground-truth coordinates).

3. 2D-HPE Based on The RN and Its Variations

2D-HPE is an intermediate result to estimate 3D human posture according to the method: 2D to 3D lifting methods and model-based methods [10]. Therefore, the 2D-HPE results have a great influence on the 3D-HPE results. The RN is applied in many studies on human pose estimation and

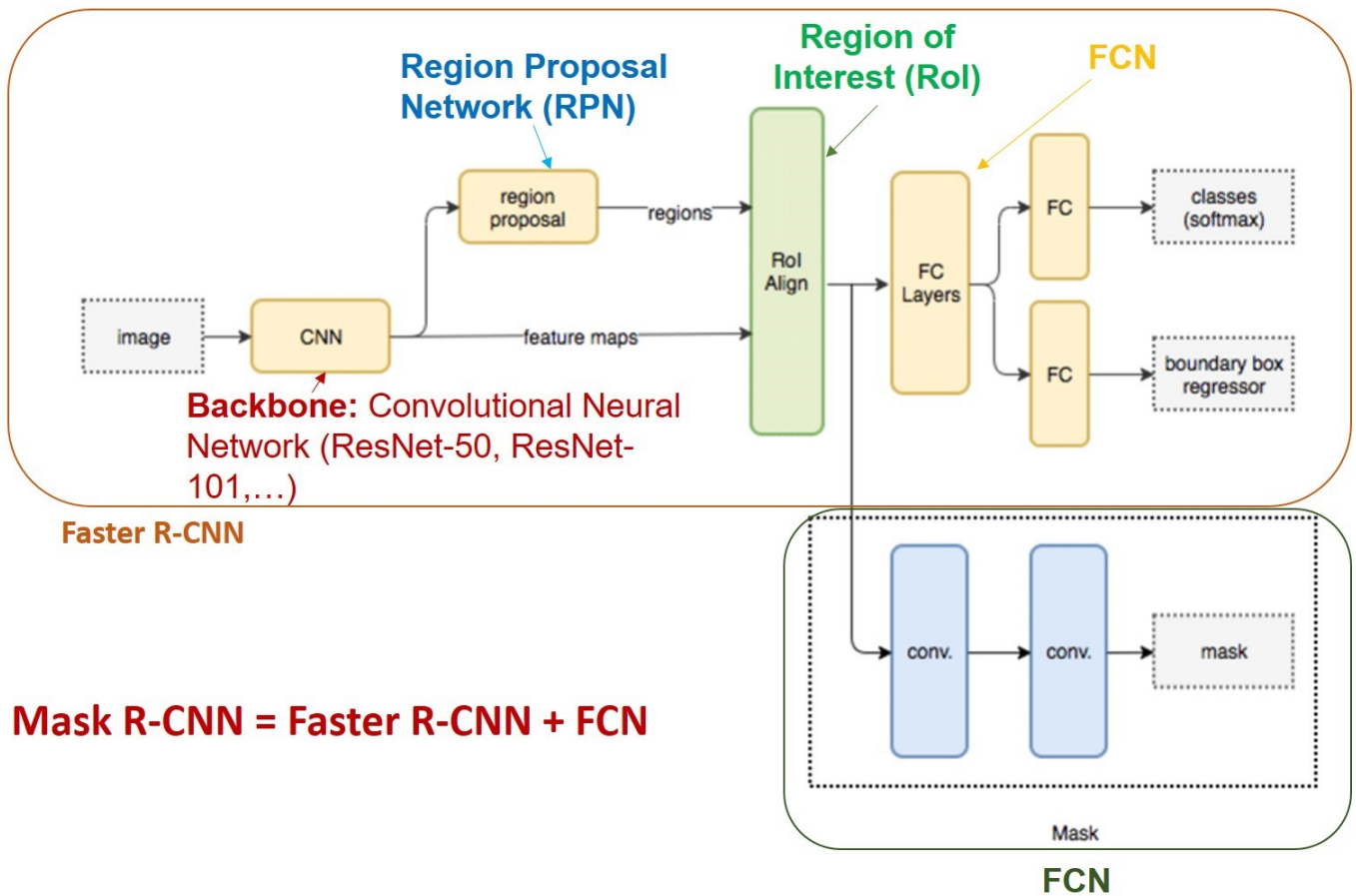


Figure 1: The human instance segmentation model on the image based on the Mask R-CNN architecture. Mask R-CNN is generated based on the combination of Faster R-CNN [17] and FCN.

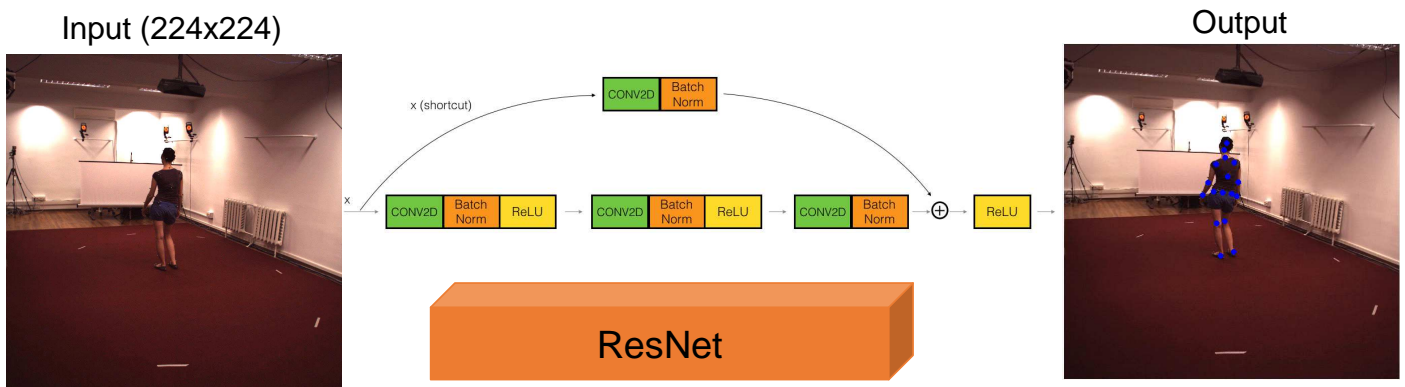


Figure 2: Illustrating the architecture of the RN for 2D-HPE.

gives good results [28]-[29]. That is the motivation for us to carry out this study to select a model with good 2D pose estimation results. We compared results from different versions of the RN (RN-10, RN-18, RN-34, RN-50, RN-101, RN-152) [14] to select the best results.

Residual Network (ResNet/RN) was introduced in 2015 and the 1st place in the 2015 ILSVRC challenge with an error rate is only 3.57%. Currently, there are many variations of RN architecture with a different number of layers. The named RN is followed by a number indicating the RN architecture with a certain number of layers. The RN have the number with each version RN-10 (10 Conv layers), RN-18 (18 Conv layers), RN-34 (34 Conv layers), RN-50 (50 Conv

layers), RN-101 (101 Conv layers), RN-152 (152 Conv layers), as shown in Fig. 3.

RN is a DNN designed to work with hundreds or thousands of convolutional layers. When building a CNN network with many convolutional layers, the Vanishing Gradient phenomenon occurs, leading to bad model training results. The Vanishing Gradient phenomenon is presented as follows: The training process in DNN often uses Backpropagation Algorithm [30]. The main idea of this algorithm is that the output of the current layer is the input of the next layer and computes the corresponding cost function gradient for each parameter (weight) of the network. The Gradient Descent is then used to update those parameters.

The above process will be repeated until the parameters of the network are converged. Normally we would have a hyper-parameter (the number of epochs - the number of times the training set is traversed once and the weights updated) that defines the number of iterations to perform this process. If the number of loops is too small, then the network may not give good results, and vice versa, the training time will be longer if the number of loops is too large. However, in practice Gradients will often have smaller values at lower layers. As a result, the updates performed by Gradients Descent do not change much of the weights of those layers and make them not converged and the network does not work well. This phenomenon is called "Vanishing Gradients". RN proposed to use a uniform "identity shortcut connection" to traverse one or more layers, illustrated in Fig. 4.

a curved arrow starting at the beginning and ending at the end of the residual block as Fig. 4. In other words, it will add an input x value to the output of the layer, which will counteract the zero derivatives, since x is still added. With Hx being the predicted value, Fx is the label, the desired output Hx to be equal to or approximately Fx .

When the input of the network is the same as the output of the network, RN uses identity block, otherwise use the convolutional block, as presented Fig. 5.

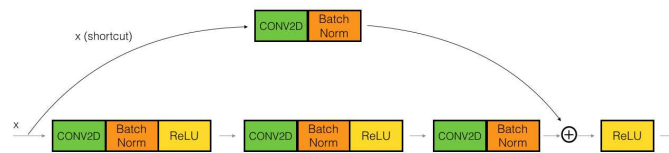


Figure 5: Illustrating convolutional block of RN.

In this paper, RN is a backbone for 2D-HPE and feature extraction. Recently, RN version 2 (v2) [14] is an improved version of RN version 1 (v1) for classification performance. The residual block [31] of RN v2 has two changes: A stack of $1 \times 1-3 \times 3-1 \times 1$ at the steps BN, ReLU, Conv2D is used; the Batch normalization and ReLU activation that comes before 2D convolution. Figure 6 shows the difference between RN v1 and RN v2.

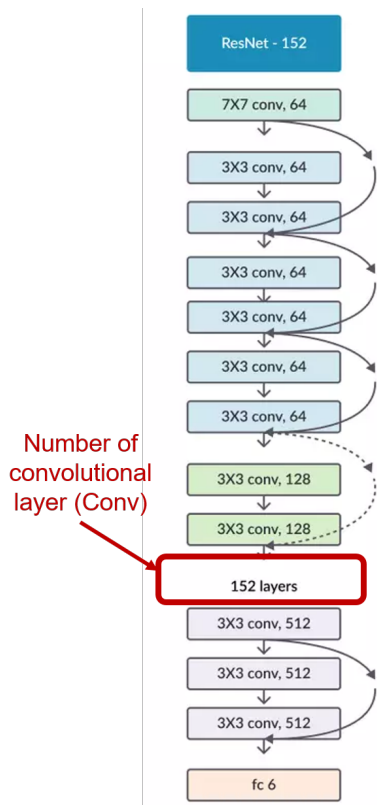


Figure 3: Illustrating of RN-152 architecture.

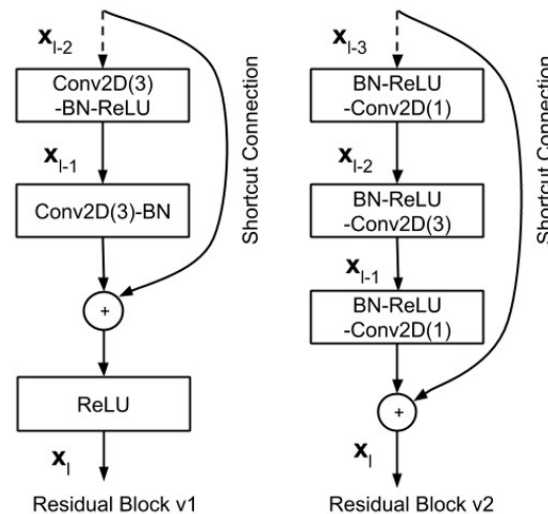


Figure 6: A comparison of residual blocks between RN v1 and RN v2 [31].

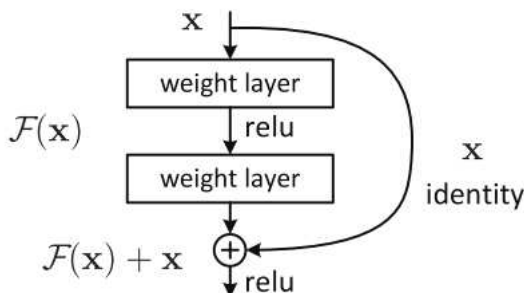


Figure 4: A Residual Block across two layers of RN.

RN is like other CNNs, includes convolution, pooling, activation, and fully-connected layer. In RN, there appears

4. Experimental Results

4.1. Dataset

To fine-tune, generate and evaluate the model and the estimated model, we use the benchmark HU-3.6M-D [32]. HU-3.6M-D is the indoor dataset for the evaluation of 3D-HPE from single-view of the cameras or multi-view of the cameras (the data is collected in a Lab environment from 4 different perspectives). This dataset is captured from 11 subjects/people (6 males and 5 females), the people perform with six types of action (upper body directions move-

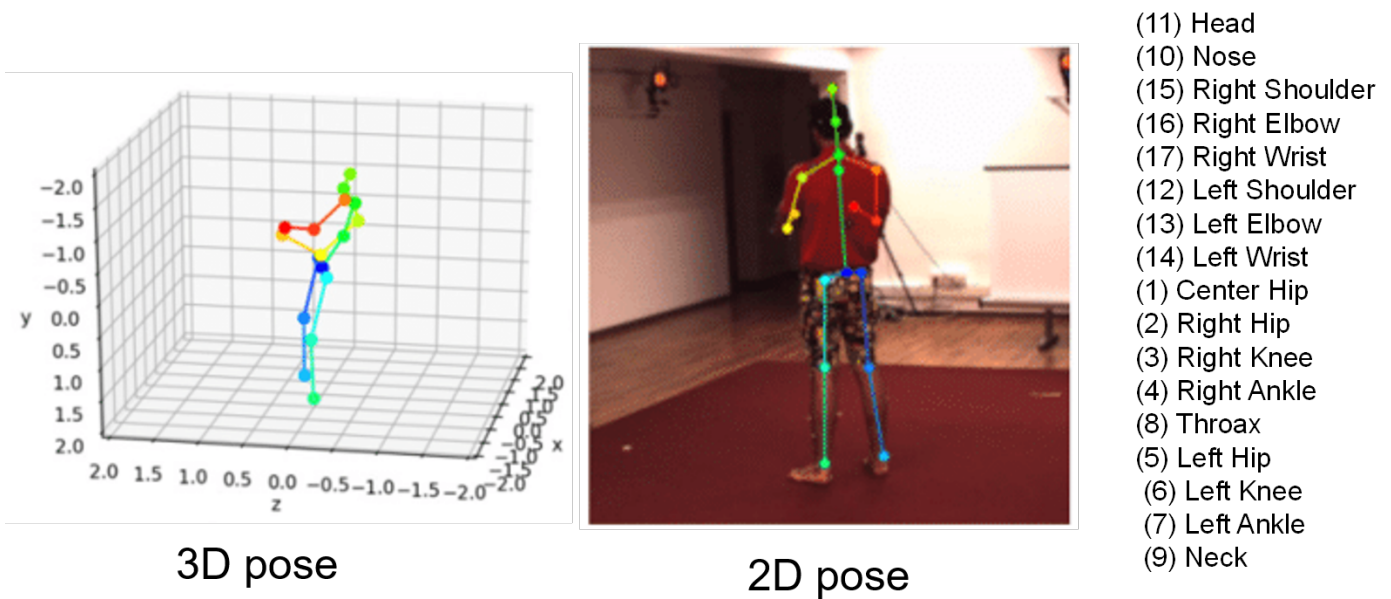


Figure 7: An illustration of human pose in HU-3.6M-D.

ment, full body upright variations, walking variations, variations while seated on a chair, sitting on the floor, various movements) which includes 16 daily activities (directions, discussion, greeting, posing, purchases, taking photo, waiting, walking, walking dog, walking pair, eating, phone talk, sitting, smoking, sitting down, miscellaneous). The frames are captured from TOF (Time-of-Flight) cameras, the data frame rate of the cameras is from 25 to 50 Hz. This dataset contains about 3.6 million images (1,464,216 frames for training - 5 people (2 female and 3 male), 646,180 frames for validation - 2 people (1 female and 1 male), 1,467,684 frames for testing - 4 people (2 female and 2 male)), 3.6 million 3D human pose annotations captured by the marker-based MoCap system. 3D human pose annotation of HU-3.6M-D consists of 17 key points arranged in order as shown in Fig. 7.

3D human pose annotations of HU-3.6M-D are annotated based on the Mocap system. The coordinate system of this data is the R-WCS. To evaluate the estimation results, we convert this data to the Camera Coordinate System (CCS). We based on the parameter set of the cameras and the formula for converting data from 2D to 3D of Nicolas [33] by Eq. 1.

$$\begin{aligned}
 P3D_c.x &= \frac{x_d - cx * depth_{x_d, y_d}}{fx} \\
 P3D_c.y &= \frac{y_d - cy * depth_{x_d, y_d}}{fy} \\
 P3D_c.z &= depth_{x_d, y_d}
 \end{aligned}
 \tag{1}$$

where fx, fy, cx and cy are the intrinsics of the depth camera. $P3D_c$ is the coordinate of the keypoint in the CCS.

Before evaluating the results of the 2D posture estimation, we re-projected the 3D human pose annotation from

the R-WCS to the CCS using Eq. 2.

$$P3D_c = P3D_w - T * R^{-1}
 \tag{2}$$

where R and T are the rotation and translation parameters to transform from the R-WCS to the CCS. $P3D_w$ is the coordinate of the keypoint in the R-WCS. We also projected to 2D human pose annotation using Eq. 3.

$$\begin{aligned}
 P2D.x &= \frac{P3D_c.x * fx}{P3D_c.z} + cx \\
 P2D.y &= \frac{P3D_c.y * fy}{P3D_c.z} + cy
 \end{aligned}
 \tag{3}$$

where $P2D$ is the coordinate of the keypoint in the image.

The source code and HU-3.6M-D have 2D annotation data as shown in link ¹.

The authors have divided the HU-3.6M-D into 3 protocols to train and test the estimation models. Protocol #1 includes Subject #1, Subject #5, Subject #6, and Subject #7 for the training model, and Subject #9 and Subject #11 for the testing model. Protocol #2 is divided similarly to Protocol #1. However, the predictions are further post-processed by a rigid transformation before comparing to the ground-truth. Protocol #3 includes Subject #1, Subject #5, Subject #6, Subject #7, and Subject #9 for the training model, and Subject #11 for the testing model. This dataset is saved in path ².

4.2. Implementation

The input data of our network includes color/RGB image data and 2D human pose annotation. All images are resized to (224×224) before being fed to the network.

In this paper, the loss function for training the estimation model includes two parts: L_1, L_2 . We used the loss

¹https://drive.google.com/drive/folders/1s3VmcZL8M2EK2M_Ese1-EWBVEZ5brM7Q?usp=sharing

²<http://vision.imar.ro/human3.6m/>

function L_1 and Adam optimizer for the training process. First, we initialized the loss function L_1 for 2D coordinates predicted from RN. Then, we computed the loss function L_2 from the predicted 2D data. The loss function L of the whole training process is calculated as Eq. 4.

$$L = \alpha * L_1 + \beta * L_2 \quad (4)$$

We set α and β to 0.1 to bring the 2D error (in pixels) into a similar range. The mean error was used to calculate the loss functions. We trained each network for 10 epochs, with the batch size being 32, Adam optimizer with the learning rate being 0.001, the number of the worker being 4.

In this paper, we used a PC with GPU GTX 970, 4GB for fine-tuning, training, testing the RN and its variations. The source code of fine-tuning, training, testing and development process was developed in Python language (≥ 3.6 version) with the support of the OpenCV-Python, Pytorch/Torch (≥ 1.1 version), CUDA/cuDNN 11.2 libraries. In addition, the support of some other libraries is required such as Numpy, Scipy, Pillow, Cython, Matplotlib, Scikit-image, TensorFlow $\geq 1.3.0$, Keras $\geq 2.0.8$, H5py, Imgaug, IPython. The source code for fine-tuning, training, testing is shown in link ³.

4.3. Evaluation Measure

To evaluate 2D-HPE, we evaluate in two phases. The first is to evaluate 2D human pose estimation results based on Eq. 5. It is the average distance between the 2D keypoint of the 2D ground-truth and the estimated 2D keypoint when using the trained model based on RN, the distance is calculated as the L_2 error value on the test set in pixels.

$$Err_{avg} = \frac{1}{N} \sum_1^N \frac{1}{J} \sum_1^J L_2 p_i, \tilde{p}_i \quad (5)$$

where N and J are the numbers of frames and number of joints ($J = 17$) respectively, \tilde{p}_i and p_i are predicted and ground-truth coordinates of i^{th} joint of the hand, L_2 is the Euclidean distance between two points.

4.4. Results and discussions

The pre-trained model of RN and its variants are shown in link ⁴. In this paper, we only evaluate the 2D-HPE on Protocol #1 and Protocol #3 of the HU-3.6M-D. The average error (Err_{avg}) between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #1 of the HU-3.6M-D is shown in Tab. 1.

The average error (Err_{avg}) on the validation set following each epoch of Protocol #1 of the HU-3.6M-D is shown in Fig. 8.

In table 1, the average error of the RN-10 is 34.96 pixels, which is the best result on Protocol #1. The average error (Err_{avg}) between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #3 of the HU-3.6M-D is shown in Tab. 2.

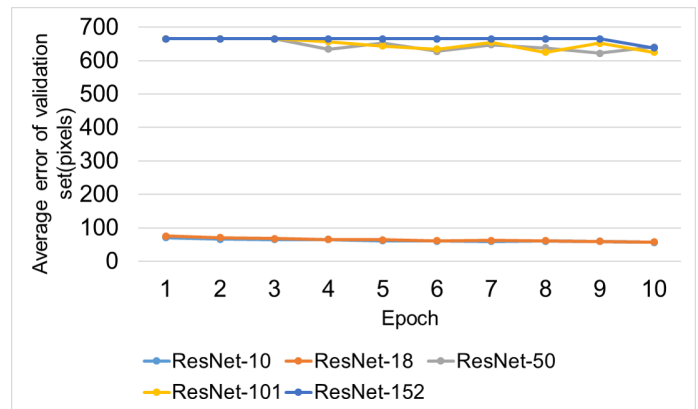


Figure 8: The average error between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #1 on the validation set.

The average error (Err_{avg}) on the validation set following each epoch of Protocol #3 of the HU-3.6M-D is shown in Fig. 9.

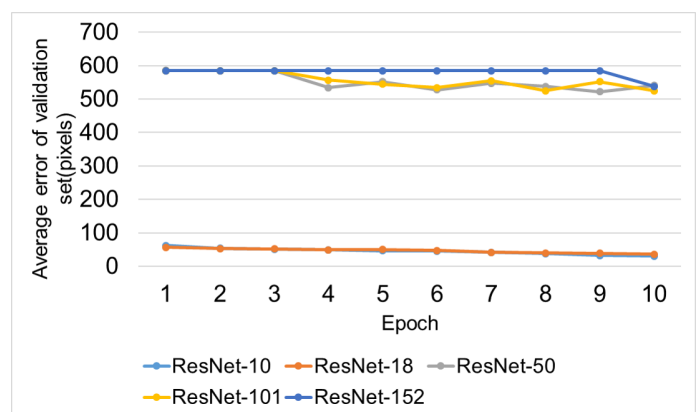


Figure 9: The average error between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #3 on the validation set.

Table 1: The average error (Err_{avg}) (IP1) between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #1 of the HU-3.6M-D.

CNNs	RN	RN	RN	RN	RN
/ Average Error (Err_{avg}) (Pixels)	-10	-18	-50	-101	-152
	34.96	38.58	669.11	628.95	652.34

In table 2, the average error of the RN-10 is 28.48 pixels, which is the best result on Protocol #3. Figure 10 illustrates a 2D-HPE result on the image. The blue skeleton is the ground-truth of the human pose, the red skeleton is the estimated human pose. When we do the training RN-10 with 50 epochs, the average error (Err_{avg}) on the test set of Protocol #1 and Protocol #3 is **15.8 pixels**, **12.4 pixels**, respectively. Thus, increasing the number of training epochs reduces the estimation error.

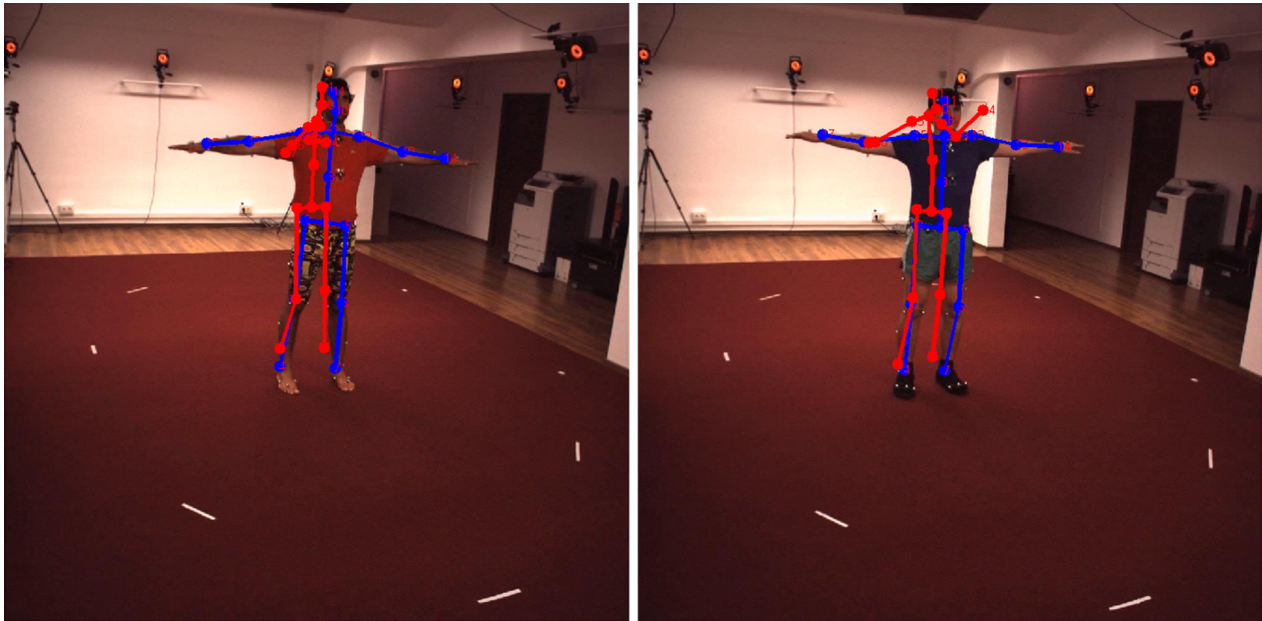
In this paper, the RN-10 has better results than RN-18, RN-50, RN-101, RN-152 networks when training 10 epochs,

³https://drive.google.com/drive/folders/1-Hu2842xWDtZWB0762iT_vIvYcuaAR7V?usp=sharing

⁴<https://drive.google.com/drive/folders/1pXkTmHAjFDNK3VaFdcGH614LF8pG4QXM?usp=sharing>

Table 2: The average error (Err_{avg}) between the 2D keypoint annotation and the estimated 2D keypoint of Protocol #3 of the HU-3.6M-D.

CNNs	RN	RN	RN	RN	RN
-10	-18	-50	-101	-152	
Average Error (Err_{avg}) (Pixels)	28.48	29.35	578.99	602.44	593.48



Subject 9

Subject 11

Figure 10: Illustrating a 2D-HPE result on the image of Protocol #1 on Subject #9, Subject #11.

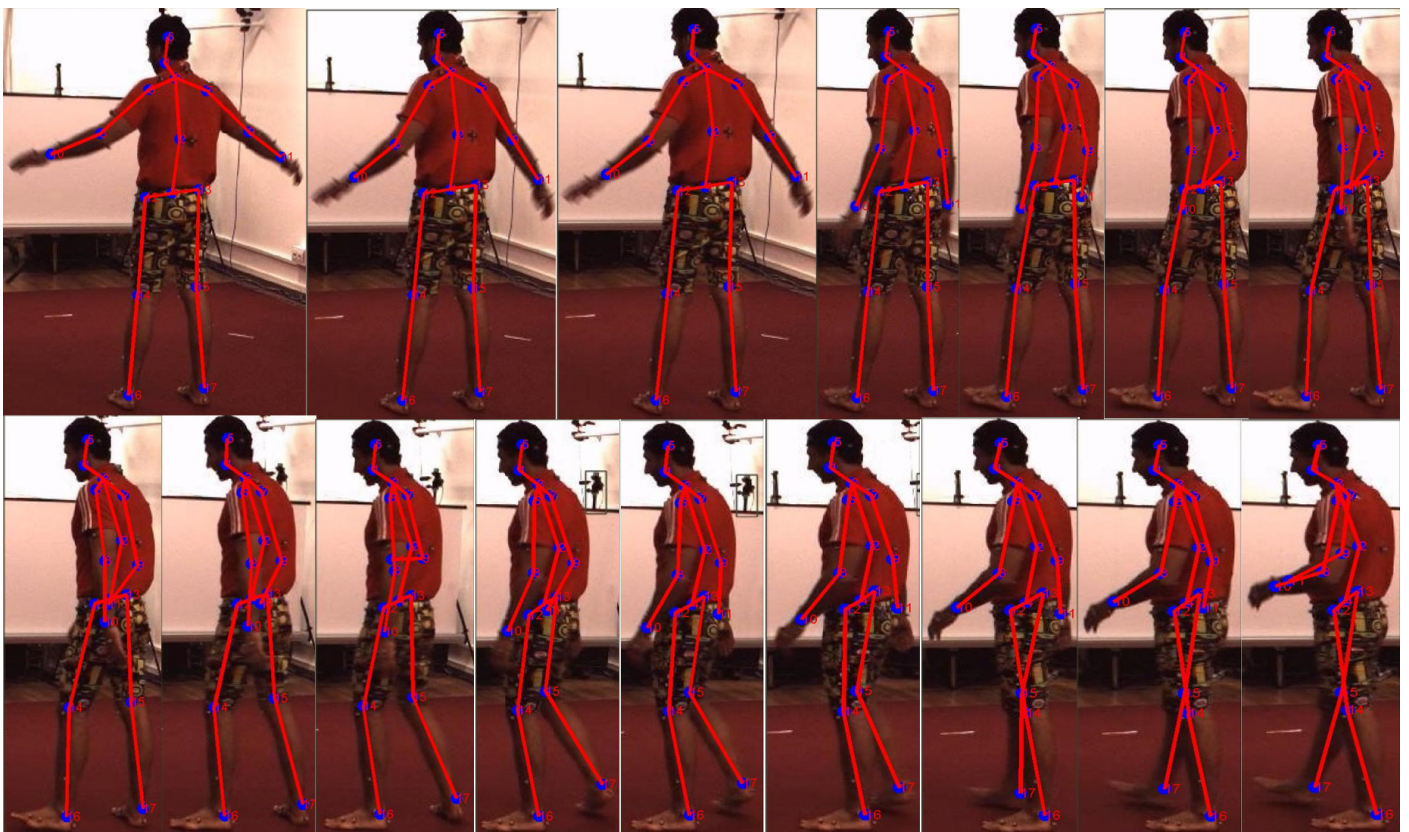


Figure 11: Illustrating a 2D-HPE result on the image of Protocol #1 on Subject #9 of RN-10.

RN-10 network is a smaller CNN than other networks, which proves that a smaller number of convolutional layers will make the network converge faster. This is also consistent with the explanation that smaller networks will learn more efficiently than large CNNs [34].

Figure 11 shows the result sequences of 2D-HPE of RN-10 on Protocol #1 of Subject #9. The estimated 2D keypoints are blue-green nodes, the joints between the estimated 2D keypoints are the red lines.

5. Conclusions and Future Works

In this paper, we have performed a comparative study for 2D-HPE based on versions of RN (RN-10, RN-18, RN-50, RN-101, RN-152) on HU-3.6M-D with two evaluations Protocols (Protocol #1, Protocol #3). We have transformed 3D human pose annotation data to 2D human pose annotation. The average error of the RN-10 is 34.96 pixels, 28.48 pixels, respectively, which is the best result on Protocol #1, Protocol #3. The results are evaluated and shown in detail and visually on the images. Therefore, RN-10 is a good CNN for estimating 2D human pose on images, this result can be used to estimate 3D human pose. In the future, we will use the human pose estimation results of RN-10 for 3D-HPE to compare with the studies of reference [35] and [9], which have the best results currently on the 3D-HPE.

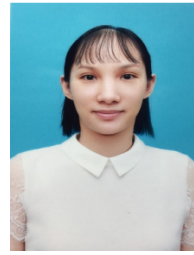
Conflict of Interest: The paper is our research, not related to any organization or individual. It is part of a series of studies by 2D, 3D human pose estimation.

Acknowledgement: This research is funded by Tan Trao University in TuyenQuang, Viet Nam.

References

- [1] N. S. Willett, H. V. Shin, Z. Jin, W. Li, A. Finkelstein, "Pose2Pose: Pose Selection and Transfer for 2D Character Animation", *International Conference on Intelligent User Interfaces*, Proceedings IUI", pp. 88–99, 2020, doi:10.1145/3377325.3377505.
- [2] H. Zhang, C. Sciutto, M. Agrawala, K. Fatahalian, "Vid2Player: Controllable Video Sprites That Behave and Appear Like Professional Tennis Players", *ACM Transactions on Graphics*, vol. 40, no. 3, pp. 1–16, 2021, doi:10.1145/3448978.
- [3] H. G. Weiming Chen, Zijie Jiang, X. Ni, "Fall Detection Based on Key Points of of human-skeleton using openpose", *Symmetry*, 2020.
- [4] N. T. Thanh, L. V. Hung, P. T. Cong, "An evaluation of pose estimation in video of traditional martial arts presentation", *Journal of Research and Development on Information and Communication Technology*, vol. 2019, no. 2, pp. 114–126, 2019, doi:10.32913/mic-ict-research.v2019.n2.864.
- [5] X. Zhou, Q. Huang, X. Sun, X. Xue, Y. Wei, "Towards 3d human pose estimation in the wild: A weakly-supervised approach", *The IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [6] G. Chandrasekaran, S. Periyasamy, K. Panjappagounder Rajamanickam, "Minimization of test time in system on chip using artificial intelligence-based test scheduling techniques", *Neural Computing and Applications*, vol. 32, no. 9, pp. 5303–5312, 2020, doi: 10.1007/s00521-019-04039-6.
- [7] G. Chandrasekaran, P. R. Karthikeyan, N. S. Kumar, V. Kumarasamy, "Test scheduling of System-on-Chip using Dragonfly and Ant Lion optimization algorithms", *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 3, pp. 4905–4917, 2021, doi:10.3233/JIFS-201691.
- [8] J. Martinez, R. Hossain, J. Romero, J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation", *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 2659–2668, 2017.
- [9] S. Li, L. Ke, K. Pratama, Y.-W. Tai, C.-K. Tang, K.-T. Cheng, "Cascaded deep monocular 3d human pose estimation with evolutionary training data", *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [10] Q. Dang, J. Yin, B. Wang, W. Zheng, "Deep learning based 2D human pose estimation: A survey", *TPAMI*, vol. 24, no. 6, pp. 663–676, 2021, doi:10.26599/TST.2018.9010100.
- [11] Z. Luo, Z. Wang, Y. Huang, L. Wang, T. Tan, E. Zhou, "Rethinking the Heatmap Regression for Bottom-up Human Pose Estimation", *CVPR*, pp. 13259–13268, 2021, doi:10.1109/cvpr46437.2021.01306.
- [12] A. Bulat, G. Tzimiropoulos, "Human pose estimation via convolutional part heatmap regression", *European Conference on Computer Vision*, vol. 9911 LNCS, pp. 717–732, 2016, doi:10.1007/978-3-319-46478-7_44.
- [13] A. Newell, K. Yang, J. Deng, "Stacked Hourglass Networks for Human Pose Estimation", *ECCV*, 2016.
- [14] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", *IEEE Conference on CVPR*, vol. 2016-December, pp. 770–778, 2016, doi:10.1109/CVPR.2016.90.
- [15] R. Zhang, L. Du, Q. Xiao, J. Liu, "Comparison of Backbones for Semantic Segmentation Network", *Journal of Physics: Conference Series*, vol. 1544, 2020, doi:10.1088/1742-6596/1544/1/012196.
- [16] R. Girshick, "Fast R-CNN", *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, pp. 1440–1448, 2015, doi:10.1109/ICCV.2015.169.
- [17] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017, doi:10.1109/TPAMI.2016.2577031.
- [18] K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN", *ICCV*, 2017.
- [19] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012.
- [20] M. Lin, Q. Chen, S. Yan, "Network in network", *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, pp. 1–10, 2014.
- [21] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", Y. Bengio, Y. LeCun, eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [22] A. Toshev, C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks", *IEEE Conference on CVPR*, 2014.
- [23] S. Liang, X. Sun, Y. Wei, "Compositional Human Pose Regression", *ICCV*, vol. 176-177, pp. 1–8, 2017, doi:10.1016/j.cviu.2018.10.006.
- [24] D. C. Luvizon, H. Tabia, D. Picard, "Human pose regression by combining indirect part detection and contextual information", *Computers and Graphics (Pergamon)*, vol. 85, pp. 15–22, 2019, doi: 10.1016/j.cag.2019.09.002.
- [25] Z. Cao, T. Simon, S. E. Wei, Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields", *IEEE Conference on CVPR*, vol. 2017-Janua, pp. 1302–1310, 2017, doi:10.1109/CVPR.2017.143.

- [26] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, "Microsoft COCO: Common objects in context", "Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)", vol. 8693 LNCS, pp. 740–755, 2014, doi:10.1007/978-3-319-10602-1_48.
- [27] M. Andriluka, L. Pishchulin, P. Gehler, B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis", "IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", 2014.
- [28] X. Xiao, W. Wan, "Human pose estimation via improved ResNet50", "4th International Conference on Smart and Sustainable City (ICSSC 2017)", vol. 148, pp. 148–162.
- [29] Y. Wang, T. Wang, "Cycle Fusion Network for Multi-Person Pose Estimation", *Journal of Physics: Conference Series*, vol. 1550, no. 3, 2020.
- [30] N. Benvenuto, F. Piazza, "On the Complex Backpropagation Algorithm", *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 967–969, 1992, doi:10.1109/78.127967.
- [31] N. V. Hieu, N. L. H. Hien, "Recognition of plant species using deep convolutional feature extraction", *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 904–910, 2020.
- [32] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments", *TPAMI*, vol. 36, no. 7, pp. 1325–1339, 2014.
- [33] N. burrus, "Kinect calibration", <http://nicolas.burrus.name/index.php/Research/KinectCalibration>.
- [34] X. Zhang, X. Zhou, M. Lin, J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices", "CVPR", pp. 6848–6856, 2018.
- [35] C. Zheng, S. Zhu, M. Mendieta, T. Yang, C. Chen, Z. Ding, "3d human pose estimation with spatial and temporal transformers", "Proceedings of the IEEE International Conference on Computer Vision (ICCV)", 2021.



learning.

Hai-Yen Tran Faculty Information Technology National Economics University (2009). She received M.Sc. degree at Faculty Information Technology Hanoi National University of Education (2013). Currently, she is a lecture of Vietnam Academy of Dance. Her research interests include computer science, deep



Thi-Loan Pham received Bachelor degree at Faculty Information Technology Hanoi Pedagogical University 2 (2007). She received M.Sc. degree at University of Engineering and Technology (2012). Currently, she is a lecture of College of HaiDuong. Her research interests include computer science, deep learning.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Van-Hung Le received M.Sc. degree at Faculty Information Technology Hanoi National University of Education (2013). He received PhD degree at International Research Institute MICA HUSTC-NRS/UMI - 2954 - INP Grenoble (2018). Currently, he is a lecture of Tan Trao University. His research interests include Computer vision, RANSAC and RANSAC variation and 3-D object detection, recognition; machine leaning, deep learning.



Trung-Minh Bui received Bachelor degree at Thainguyen of Information and Communication Technology (ICTU) (2010). He received M.Sc. degree at Thainguyen of Information and Communication Technology (ICTU) (2014). Currently, he is a lecture of Tan Trao University. His research interests include Computer science; machine leaning, deep learning.

Text-Based Traffic Panels Detection using the Tiny YOLOv3 Algorithm

Saba Kheirinejad ^{1,*}, Noushin Riahi ², Reza Azmi ²

¹Center for Ubiquitous Computing, University of Oulu, Oulu, Finland

²Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran

*Corresponding author: Saba Kheirinejad, University of Oulu, Email: saba.kheirinejad@oulu.fi

Corresponding author ORCID: 0000-0002-3998-804X

ABSTRACT: Lately, traffic panel detection has been engrossed by academia and industry. This study proposes a new categorization method for traffic panels. The traffic panels are classified into three classes: symbol-based, text-based, and supplementary/additional traffic panels. Although few types of research have investigated text-based traffic panels, this type is considered in detail in this study. However, there are many challenges in this type of traffic panel, such as having different languages in different countries, their similarity with other text panels, and the lack of suitable quality datasets. The panels need to be detected first to obtain a reasonable accuracy in recognizing the text. Since there are few public text-based traffic panels datasets, this study gathered a novel dataset for the Persian text-based traffic panels all over the streets of Tehran-Iran. This dataset includes two collections of images. The first collection has 9294 images, and the latter has 3305 images. The latter dataset is more monotonous than the first one. Thus, the latter is utilized as the main dataset, and the first is used as an additional dataset. To this end, the algorithm uses the additional dataset for pre-training and the main datasets for training the network. The tiny YOLOv3 algorithm that is fast and has low complexity compared to the YOLOv3 is used for pre-training, training, and testing the data to examine the utility and advantages of the data. The K-fold cross-validation procedure is used to estimate the model's skill on the new data. It achieves 0.973 for Precision, 0.945 for Recall, and 0.955 for F_{measure} .

KEYWORDS: Intelligent transportation system, Deep learning, Convolutional neural networks, Tiny YOLOv3, Traffic signs, Traffic panels.

1. Introduction

Countries define rules to guarantee a secure traffic system. All traffic users such as drivers and pedestrians need to obey these rules. Traffic panels/signs play a crucial role in the system. Drivers observe the panels and act based on the information that the panels provide. The drivers intentionally or unintentionally ignore the traffic panels in different circumstances (once a vehicle has high speed or the driver is distracted by something), which can cause horrible accidents and misfortune. The definite dependence on the mentally and physically vulnerable human is hazardous. To this end, vehicles armed with progressive technologies play a significant role in the security of the traffic system and protecting many human lives. Detection of the traffic panels is considered a formidable technology in the intelligent transportation system. It has various uses including driver assistant systems, unmanned autonomous vehicles, and road panel keeping [1]. The traffic panels have been divided into two categories in the previous works: I) symbol-based and II) text-based traffic panel. This study classified

the traffic panels into three classes: I) symbol-based, II) supplementary/additional, and III) text-based traffic panels.

1.1. Symbol-Based Traffic Panels

The first group utilizes specific and straightforward symbols to represent the concept of the panels. Symbol-based traffic panels are divided into several categories based on their form including rectangular, triangular, square, circular, pentagonal, and octagonal shapes, and their colors such as white, red, yellow, blue, and green. The shapes and colors with low variety determine the specific purpose of the panels, such as stop, pedestrian crossing, bicycle crossing, railway crossing, watching for children, speed limit, etc. The typical concepts of this type of traffic panel are danger, prohibition, commitment, warning, priority, end of prohibition, etc. Prohibition panels are in a circle form with a red margin, warning panels are triangular, and information panels are blue [1]. The symbol-based traffic panels have been standardized

in most countries, they have the same shapes and meanings, and there are several suitable and standard datasets for them. Many studies have mainly considered this type of panel [2]-[23]. Instances of symbol-based traffic panels can be seen in Figure 1.



Figure 1: Instances of symbol-based traffic panels.

1.2. Supplementary/Additional Traffic Panels

The traffic panels in this group complete the concept of symbol-based traffic panels. They contain ‘text,’ ‘arrow,’ ‘pictograph,’ and ‘a combination of the text and pictograph’ [24]-[30]. The shape of these panels is generally rectangular, their color is white, and they are found under the symbol-based traffic panels. There is a considerable variation of this type in Germany. However, there are a few numbers of this type in Iran. Since the number of traffic panels in this group is less than the number of traffic panels in two other groups, this group has attracted less attention. Instances of supplementary/additional traffic panels in Germany can be observed in Figure 2 and one instance in Iran can be observed respectively in Figure 3.

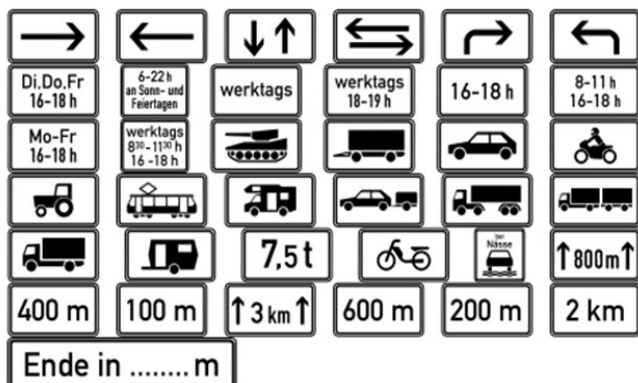


Figure 2: Instances of supplementary traffic panels in Germany [24].



Figure 3: Instances of supplementary/additional traffic panels in Iran.

1.3. Text-Based Traffic Panels

Text-based traffic panels include text and arrows. The texts deliver rich and valuable semantic information about the traffic system and have a fundamental function in the traffic system and intelligent transportation system. They illustrate the correct routes, warnings about possible dangers, and permission or prohibition of access to the road, etc. The automatic detection of these panels can alert the driver about the traffic and environment. Some systems can read the text using a synthesized voice or combine it with navigation systems like GPS. They play the assistant role and aim to increase the drivers' attention to have a safe traffic environment. Fewer studies have been done in this category than symbol-based category [31]-[44]. The reason is a multitude of challenges which are presented in the following [1].

- There are not enough and suitable public datasets.
- The texts are written in different languages in different countries.
- Sometimes, the panels are not evident in various weather situations such as snowy, rainy, sunny, foggy, cloudy, etc.
- There are different text-based boards in the street that look like the text-based traffic panels such as billboards, store name boards, and advertisements on the vehicles.
- The color of the panels and texts fade due to the sunshine for a long time and reaction between the air and color.
- Some panels have been physically damaged.
- Some panels have been obstructed by the trees, vehicles, and other objects.

As far as we know, few benchmark datasets for text-based traffic panel detection are publicly available [45, 40]. Hence, this study gathered a novel public dataset all over the city of Tehran-Iran and called it ‘Persian Text-Based Traffic Panel Dataset’ [46]. As well as the general challenges mentioned above, there are other exclusive challenges for the Persian dataset: I) As shown in Figure 4, some traffic panels are not uniform from the perspective of color. In other words, they are a combination of multiple colors. II) There are a few greens

and white rectangular panels that only includes the symbols, and there is no text inside them, as shown in Figure 5. Since these panels look like text-based traffic panels, the network might make a mistake in detecting them. III) Persian text-based traffic panels include two languages, Persian and English. Therefore, the size of the texts is small, and the characters are unreadable from a relatively long distance.

1.4. Contribution

The contributions of this study are outlined as follows. This paper provides short and valuable literature on different types of traffic panels. Different traffic panels are categorized into a novel and comprehensive manner. The newest related works for detecting traffic panels in traditional and non-traditional methods are presented for each category. This study introduces a new dataset and elaborates on data collection and traffic panel detection techniques and challenges in detail. This work explains the way of evaluation using a state-of-the-art algorithm, the Tiny YOLOv3, to analyze the efficiency and benefits of the data. In the evaluation and results section, the algorithm's performance for the dataset has been extensively discussed. The results show that the algorithm provides 0.973 for Precision, 0.945 for Recall, and 0.955 for Fmeasure.



Figure 4: Instances of the panels that are not uniform.



Figure 5: Examples of the symbol-based traffic panels looking like the text-based traffic panels.

1.5. Organization

The study is structured as follows. Recent works in different types of traffic panels are reviewed in section 2. The used algorithm is explained in section 3. The steps of the data collection and labeling are described in section 4. The evaluation and results are discussed in section 5. The discussion is presented in section 6. Ultimately, the

conclusion and perspectives about the future studies are proposed in section 7.

2. Related Work

This section discusses related works in different categories of traffic panels.

2.1. Symbol-Based Traffic Panel Detection

Detection methods for symbol-based traffic panels are generally divided into three sorts: I) colour-based methods, II) shape-based methods, and II) deep learning methods.

2.1.1. Color-Based Methods

The color-based methods usually use normalized color space such as RGB (Red, Green, Blue), YUV, and HSI (hue, saturation, and intensity), with specific thresholding. Segmentation means separating the traffic panels from the road background. Due to the damage to the color and shape of the traffic panels, it is not easy to distinguish its position from other outdoor images. Most segmentation techniques use color. Since the traffic panels are divided based on the color information from the background, it depends on the threshold of the input image in various color spaces. First, the segmentation is done using a different color range. Shape detection is done in the next step. Many researchers in this field face issues for white color and uneven lighting level [1]. In [47] red and green components' average is calculated for each pixel, and the G-R histogram is developed by displaying G-R's value. Many researchers use HSI color space because it understands the human understanding of RGB space well. While some researchers have found better results in the color space of YCbCr and YUV [48].

2.1.2. Shape-Based Methods

The particular shapes of traffic panels such as circles, triangles, and squares make them significant for distinguishing from other panels. For the shape-based methods, Hough transformation and different types of Hough transformation are used. Like detection of the general objects, a sliding window using categorization of the region is used to investigate the presence of traffic panels in the current window. Shape detection is an essential component of information in the recognition process of the traffic panels. Techniques such as connected components, dividing and merging edge detection, and clustering are utilized to compound pixels using color information-based similarity measurements to model the features at the detection stage. Attributes like aspect ratio, width, height, perimeter, and area of border-

box, are extracted experimentally to obtain candidate bubbles. The resulting bubble is confirmed at this stage. There are several features that are extractable from the bubble that can be used in the classification of the shapes better [49]. Therefore, for each segmented bubble, right, left, bottom, and top DtB vectors are produced. In [50], utilized only right and left DtB vectors to detect the shape. The task in [51] is given by the union of distance vectors from the bubble's center to the bubble's outer edge, i.e., distance vector from the center of union. These feature vectors are utilized in the categorization of the shapes using a linear support vector machine. In [52] an alternative method uses the signature form of the connected components, which is computed directly through the bubble. Referring to the angle, the signature is defined the distance from the center to the object border. The absolute value of the fast Fourier transform is used for each shape signature to counteract the direction of the panel. The absolute value of discrete Fourier transform can be a good choice as well, but is not preferred due to computational complexity [53]. In [54], the Histogram of oriented gradients algorithm is developed using integrating the color information into the feature vector to improve performance. Since the road panels have high color contrast, some studies consider the pixels as features for shape definition [3].

2.1.3. Deep Learning Methods

Late advancements in deep learning techniques have inspired researchers to exploit neural networks to detect and recognize traffic panels. Contrary to the comparative achievement of detectors with hand-crafted features, most non-deep learning-based systems are not suitable to correctly detect a vast number of traffic panels [15]. Others converted the basic image into the grayscale image using support vector machines (SVM). They exploited convolutional neural networks (CNN) with fixed and learnable layers to detect the traffic panels [55]. In utilized CNN and twin SVM hybrid model in a new study [18]. In [56], a traffic panel detection and recognition algorithm were presented using CNN. Target of the system included traffic panels, Chinese characters, English letters, and digits. The system uses a multi-task CNN as its base trained to obtain valuable features for classifying and localization of different traffic panels and texts. Authors in a newer study proposed a CNN consists of a CNN-based challenge classifier [20]. In [57] for extracting the Region of Interest (RoI), the AdaBoost classifier and local binary pattern feature detector are combined. Cascaded CNNs are used to decrease negative samples of RoI for traffic panel recognition. It is also used cascade

machine learning network with hierarchical classes for traffic panel detection [22]. In [58] it is presented a novel real-time method based on cascade deep learning and AR for a fast and precise framework for traffic panel recognition. It superimposes augmented virtual objects onto a real scene under all kinds of driving conditions, such as adverse weather circumstances. They merged the segmentation architectures Seg-Net and U-Net for detecting the traffic panels from video sequences [11]. Different Studies used CNN and mask R-CNN to detect the symbol-based traffic panel [12, 14]. Various studies used different versions of You Only Look Once (YOLO) algorithm, YOLOv3, YOLOv4, and YOLOv5 to recognize the traffic panels [16, 17, 23]. They presented a new traffic panel recognition method via incorporating a lightweight superclass detector with a refinement classifier [59]. It is presented a new dataset in Pakistan and used transfer learning to recognize the traffic panels [21].

2.2. Additional/Supplementary Traffic Panel Detection

Since the supplementary or additional panels often appear below the symbol-based traffic panels, in most detection methods, first symbol-based panels are identified, then the interested region located below the symbol-based panels are specified, at last, the additional traffic panels in the RoI are searched. To this end, it is presented an algorithm to detect additional panels besides the other panels. This method exploits Hough transform for lines and geometric constraints to find additional panels. Then uses SVM classification to verify the candidate shape [24]. They used MLP neural network to detect the rectangular shape in a region below speed-limit panels [30]. The authors presented a method to determine the shapes with 'arrow,' 'pictogram,' 'text or mixed.' Afterward, they classified the shapes with pyramid-HOG features [25]. The author utilized a Maximally Stable Extremal Regions (MSER) based approach to detect the additional panels [26]. In their next study, they found corner areas via aggregated channel features. Then, quadrangle generation and filtering technique were used to filter the variation of large aspect ratio for supplementary panels [28]. They introduced a complete pipeline to recognize the text by optical character recognition (OCR) in their earlier work. They assumed a specific additional panel, classified its layout, determined the bounding boxes of the content using regression, then used a multi-class classification step or applied a text sequence classifier if necessary [27]. It is proposed CNNs to classify the road condition and recognize additional panels that display the validation of 'when wet' additional panels [29].

2.3. Text-Based Traffic Panel Detection

Few studies have been done on text-based traffic panel detection compared to symbol-based traffic panels. This study categorized methods of text-based traffic panel detection into two categories, I) the methods that extract the features in a traditional way, and II) the methods that extract the features using deep learning algorithms.

2.3.1. The Methods That Extract the Features in a Traditional Way

In [35], the author used the traffic panels properties such as geometric constraints and color distribution to distinguish the traffic panels from the other objects. At last, they used the vertical plane criterion for traffic panel detection. In [60] considered the traffic panels in Spain. They exploited the information of HSI color space and computation of the shape classification techniques for thresholding. They detected the blue pixels by applying an appropriate threshold and used achromatic decomposition to detect the white pixels. They labeled the connected components and considered all the blob candidates in a selection process and removed some of them based on their aspect ratio and size. Xavier et al. [32] found edges of the input image using a canny edge detector. They used Hough transform to find the straight lines of the detected edges. Then, they determined the candidate regions (top, right, and left of the road) that contain the traffic panels. They utilized HSV thresholding to detect the text-based traffic panels in the candidate regions. Finally, they vertically aligned the detected panels using the random sample consensus (RANSAC) algorithm and homography techniques. In [36], the images have been taken from Google Street View Service that provides 360° panoramic views with high resolution from different locations on various streets and roads worldwide. The authors used the text localization algorithm and detected the panel characters using MSER. They used MSER to determine traffic panels as rectangles with high text density. As plenty of traffic panels that have been studied in this paper have a blue background, a blue segmentation method was applied to detect the extensive rectangular regions. In [37], specified the RoI of the text-based traffic panels by defining the vanishing point and sides of the road. Text-based traffic panels candidates were detected exploiting thresholding techniques in RoI by MSER and HSV. The number of candidates was decreased via temporal and contextual information at the end. In [38], detected the text-based traffic panels by Bag of Visual Words (BoVW) and color segmentation techniques. They applied the BoVW method on the specific parts of the images identified by

the blue and white masking. Khodayari et al. [61] processed the input images in HSV space. Then, they detected the traffic panels using the fuzzy logic method. In the fuzzy images, the properties of each pixel, such as brightness and color are presented using linguistic values. Yellow, green, light green, white, dark red, and dark blue are examples of the linguistic values assigned to each pixel. In [45], presented a dataset including videos of traffic panels recorded by a moving vehicle in municipal regions. They categorized the video sequences based on their features in a way that researchers could select the piece of data that helps their research purpose.

2.3.2. The Methods That Extract the Features Using Deep Learning Algorithms

In [31], the author proposed a new Cascaded Localization Network. They detected the candidate traffic panels in each frame on a set of continuous image frames using a set of features learned by CNN. They also collected a novel dataset for the traffic guide panel to train and appraise the framework. In [62], the author collected a text-based traffic panels dataset containing English and Chinese traffic panels. They exploited a fully convolutional network to segment the candidate traffic panels. In [42], the author proposed a new system for recognizing both symbol-based and text-based traffic panels. They applied MSER on grey and normalized RGB channels to extract the RoIs from each video frame. They trained a multi-task CNN with images labeled from street views and synthetic traffic panels and used it to recognize the panels. In [33], it is used a region proposal network to detect Korean character candidates. Then, they classified Korean characters by a classification network. In [34], the author applied MSER and OCR on given text-based traffic panels to recognize the texts. Peng et al. [43] presented a deep learning-based cascade detection model with two stages to detect the traffic panel text in natural scenes. They found the panels' RoIs using an improved Single Shot Multi-Box Detector (SSD) network. Zhang et al. [44] solved the false detection and undetected by proposing a cascaded R-CNN to reach the multi-scale features in pyramids. All layers of the cascaded network combine the output bounding box of the previous layer for joint training; other than the first layer, this technique chips for traffic panels detection. They also presented a multi-scale attention technique to get the weighted multi-scale features using softmax and dot-product, highlighting the traffic panel features and improving the detection accuracy by fining the features. In [63], the author proposed a new detection method called MSA_YOLOv3 to detect the small traffic panels precisely. They achieved

data augmentation by exploiting image mix-up technology. They introduced a multi-scale spatial pyramid pooling block into the Darknet53 network to help the network learn object features more generally. In [39], it is proposed an algorithm to detect the mixed horizontal-and-vertical-text traffic panel with the Chinese language in the street. To differentiate the traffic panels from other similar things in color in the complicated background like street scenes, the different red, green, and blue features were effectively mixed. Since the Chinese text lines are usually vertical and horizontal on text-based traffic panels, the presented technique created the text lines by the structural information and position of the characters. In [40], the author presented a novel public multi-task dataset for detecting text-based traffic panels. In [41], it is presented an end-to-end trainable deep neural network that can recognize multi-oriented text instances in adverse meteorological conditions.

3. The Tiny YOLOv3 Algorithm

Since the computer's hardware is continuously growing, CNN-based deep learning methods have advanced fast and acquired remarkable results in computer vision and machine vision scope [64]. Traffic panel detection is considered a subset of the more general scope of object detection. Object detection is a relatively old subject. However, it is still a complex issue and an active scope for many researchers despite the endeavors that have been accomplished. Various object detection algorithms have been suggested so far. Currently, the most famous algorithms are SSD [65], R-CNN [66], and YOLO. YOLO is an algorithm that has become widespread for object detection recently. They have proposed its three versions (1-3) [67, 68, 69]. The fourth and fifth versions have been presented by [70] and [71], respectively. This algorithm solves object detection as a regression problem. It classifies the objects and gives their location as an output using an end-to-end network in one step. The noteworthy thing about this algorithm is its speed. It is one of the fastest algorithms so far. YOLOv3 uses the K-means clustering algorithm to automatically choose the most suitable anchors for the dataset. This algorithm is very complicated and demands complex hardware. It has 256 layers, and 53 layers out of 256 are convolution layers, as it utilizes the Darknet-53 as the backbone network. The tiny YOLOv3 is a smaller and faster version of the YOLOv3 algorithm. It has an architecture with low complexity and a more straightforward implementation for datasets with small sizes. This algorithm is used for pre-training, training, and testing the dataset. The tiny YOLOv3 algorithm used

in this study has 46 total layers, including 13 Convolution, 11 Leaky Relu, 11 Batch Normalization, 6 Max Pooling, 3 Input, 1 Concatenate, and 1 Upsampling layer. In Table 1, the features of the tiny YOLOv3 architecture are summarized.

Table 1 : Features of the tiny YOLOv3 architecture.

Name of layers	Number of layers
Convolution	13
Batch Normalization	11
Leaky Relu	11
Max Pooling	6
Input	3
Upsampling	1
Concatenate	1

4. Dataset Collection

Before getting into the details of collecting the dataset, the characteristics of the Persian text-based traffic panels in Iran are explained. These traffic panels express a particular meaning according to their shape and color. The concepts of the colors for this type of traffic panel are as follows:

- Green: Route guide for highways and religious places
- Blue: Route guide for freeways and service guide
- Black and white: Route guide for the other ways
- Yellow: General warnings and warnings for construction or repair operations
- Orange: Guide for administrative, training, and service areas
- Brown: Guide for recreational, cultural, and tourism areas

The concept of the shapes in the text-based traffic panels are as follows:

- Horizontal rectangle: Warning for repairing and maintenance
- Vertical rectangle: Guide for imperative and service panels
- Flag rectangle: Guide for routes

Since few public datasets focus on text-based traffic panels, this study collected a novel challenging dataset containing traffic panels with the Persian and English text called 'Persian Text-Based Traffic Panels.' In this dataset, the images have been taken using diverse smartphone cameras with various specifications by crowdsourcing method. Most images have been taken all over the city of Tehran in Iran, more on highways and streets. Two collections of images were collected. First, 4000 images were collected in different situations. Instances of these images can be witnessed in Figure 6. The images were augmented to 9294, followed by labeling. The tiny YOLOv3 algorithm was used for training and testing to

examine the utility and advantages of the data. Nevertheless, this dataset did not attain suitable performance despite the positive expectation. The reason might be numerous text-based traffic panels in different situations compared to the total number of images (9294). In other words, there were many challenges in that dataset. It was compared to the dataset (an available part of the dataset) that has been used in [31]. Since most investigated text-based traffic panels are green and somewhat look like each other, this dataset is nearly monotonous and does not have many challenges. Hence, training and testing a model with this dataset is more straightforward than our dataset, gaining more accurate results. Thus, 1500 new images were taken as the main dataset. These images were more uniform than the images in the additional dataset. The images were augmented to 3305 and were labeled. This dataset was called the 'main dataset' and was utilized for training the algorithm. It is worth noting that the previous dataset was used to pre-train the algorithm and was called 'additional dataset.' Next, the dataset collection steps are explained in the following. Before getting into the details, methods of taking the images are explained. The images were taken in different situations such as:

- Some images have been taken in a static condition.
- Some images have been taken in a moving car, and they might have been zoomed (in case of being far from the traffic panel).
- The images have been taken at various distances.
- The images have been taken in different climates (e.g., sunny, cloudy, and rainy).
- The images have been taken at different times of the day with different light conditions (e.g., morning, noon, evening, and night).
- Some images have been taken from the traffic panels that have been physically damaged.
- Most of the images have been taken in the city with complex background.
- When lights of the cars and lamps in the middle of the highway and street illuminate the traffic panels).
- Some images have been taken from the traffic panels blocked by the trees, billboards, etc.
- Some images have been taken when the camera is behind the glass of the car.

4.1. Additional Dataset Collection

This dataset contains 4000 images with 12 Gigabyte sizes. Since different persons and cameras have taken the images, they are in both vertical and horizontal shapes, and their sizes are different. In other words, the images are not uniform. The images needed to be preprocessed to get uniform for use in the network. The desired size of the

input images for the network is 416×416 . If the images were directly resized to 416×416 , the objects within the image would lose their original form. To this end, the cropping technique was used, and the images were converted to square form with the same length and width. Then, a bicubic interpolation method was applied to make the images monotonous.



Figure 6: Examples of the additional dataset.

The cropping strategy has a pair of advantages: I) data augmentation and II) making the images uniform. To augment the data, the images were cropped three times, from the left side, right side, and both sides. Using this technique, the total number of images increases, and no pixels are discarded from the original images, i.e., all the pixels were optimally used. After cropping each image, three images are produced from one original image. However, in some images, cropping causes the exit of some traffic panels from the cropped images. Thus, the images with no traffic panels need to be removed, which made the number of the final images not precisely three times the number of the original images. Therefore, the total number of augmented images is 9294. Finally, a bicubic interpolation technique was applied to derive a uniform dataset with 416×416 size.

4.2. Main Dataset Collection

Due to the additional dataset challenges, a new and more uniform dataset was collected. 1500 images were collected and augmented to 3305 using the exact technique utilized for the additional dataset. Instances

from the main dataset can be observed in Figure 7. In the new dataset, the panels showing the name of the alleys and streets were ignored. Because they are smaller than the usual traffic panels, and their texts are unclear. The green and white text-based traffic panels were used more. The images that a big part of them have been blocked with the trees, cars, and the other objects were not also used.



Figure 7: Examples of the main dataset.

After data collection and applying the preprocessing steps, the images were labeled. Before explaining the labeling process, the predicted bounding box and ground-truth concepts need to be defined. The predicted bounding box is the bounding box predicted by the network. The ground-truth bounding box is the exact location of the defined traffic panel in the images. As shown in Figure 8 the predicted bounding box and ground-truth bounding box have been depicted in red and green. The predicted bounding boxes that overlap more with the ground-truth bounding boxes are more precious. Ground-truth bounding boxes are usually determined in two forms: I) the coordinates of the two corner points connected by diameter, and II) the coordinates of one corner and the length and width of the bounding box. This study uses the first one. The authors developed a code, after running the code, the dataset's images are displayed one by one. As it can be seen in Figure 9, the user needs to click twice (one clicks for left up corner and another one for right down corner) for each panel, in such a way that the bounding box covers the whole panel.

5. Evaluation and Result

Some important parameters for implementing the algorithm are summarized in Table 2. First, the network uses the additional dataset for training and the COCO dataset's weights as the pre-train weights. Then, it uses the main dataset for training and the obtained weights from the previews network as the pre-train weights. The number of the training dataset is 80 percent of the whole

dataset's images, the validation set is 10 percent of the images number in the training dataset, and the number of the test set is 20 percent of the whole images. The exact number of the main and additional datasets are explained in Table 3.



Figure 8: Red: the predicted bounding box, Green: the ground-truth bounding box.



Figure 9: A ground-truth bounding box, two yellow points are manually determined by users.

To evaluate the algorithm's accuracy, the first step is to calculate the Intersection over Union (IoU) for all the bounding boxes. The IoU of each box is computed by dividing the area of the intersection between ground-truth and predicted bounding box over the area of the union between ground-truth and predicted bounding box. This study considers the IoU threshold 0.5. The Precision, Recall, and $F_{measure}$ determining the accuracy of the algorithm are described in the following.

True Positive: TP

False Positive: FP

False Negative: FN

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F_{measure} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (3)$$

This study uses K-Fold cross-validation technique with K=5 to evaluate the algorithm’s performance. The results are illustrated in Table 4. The evaluation outcomes of Fold1 for various thresholds of IoU are depicted in Table 5. As shown, once the IoU threshold boost, Precision, Recall, and Fmeasure drop. The comparison of this method with other methods for text-based traffic panel detection can be observed in Table 6.

Table 2: Parameters’ value of the network.

Input image size	416 * 416
Anchor number	6
Class number	1, traffic panel class
Optimized function	Adam
Learning rate	0.001
Batch size	8
Epoch	100
Patience	5
Graph card	NVIDIA GeForce GTX1070

Table 3: Characteristics of the additional and main dataset.

Number of images in	Additional dataset	Main dataset
Total collection	4000	1500
Total collection after data augmentation	9294	3305
Train collection	7435	2644
Validation collection	743	264
Test collection	1859	661

Table 4: Precision, Recall, and Fmeasure for each five folds in cross validation.

Test subset	Image number	Traffic panels number	True Positive number	False Positive number	False Negative number	Precision	Recall	Fmeasure
Fold1	661	871	823	24	48	0.971	0.944	0.957
Fold2	661	880	829	27	51	0.968	0.942	0.954
Fold3	661	865	819	19	46	0.977	0.946	0.960
Fold4	661	850	807	17	43	0.979	0.949	0.963
Fold5	661	875	827	25	48	0.970	0.945	0.944
Average						0.973	0.945	0.955

Table 5: Precision, Recall, and Fmeasure for various IoU in Fold1.

IoU	Precision	Recall	Fmeasure
>0.5	0.971	0.944	0.957
>0.6	0.955	0.940	0.950
>0.7	0.818	0.935	0.873
>0.8	0.509	0.900	0.650

Table 6: Comparison of different methods for text-based traffic panel detection.

Study	Precision	Recall	Fmeasure
Rong et al. [31]	0.73	0.64	0.68
Greenhalgh et al. [37]	0.96	0.90	0.93
Zhu et al. [62]	0.93	0.94	0.93
Peng et al. [43]	0.93	0.91	0.92
This study	0.97	0.94	0.95

6. Discussion

Good accuracy in the high IoU thresholds is crucial as the detected text-based traffic panels may not contain the full text. The used algorithm in this study has some flaws mentioned in the following. As shown in Figure 10, the predicted bounding boxes do not contain the full text though IoU is higher than 0.5. Since the next step after the text-based traffic panel detection is the recognition of the texts, the predicted bounding boxes need to contain the entire text to deliver the concept. As shown in Figure 11.



Figure 10: Predicted bounding boxes that do not include the full text.



Figure 10: Instances of TP and FN detected panel.

Despite the clarity of some panels in specific sizes, the network cannot detect them. The reason is anchors' size chosen by the algorithm at the beginning. As depicted in Figure 12, the panels that were wrongly detected as the traffic panels look like some text-based traffic panels in size and background. Although the Tiny YOLOv3 has a few flaws, it is a robust algorithm and can detect traffic panels in many challenging situations. It does not wrongly detect the panels that contain several symbol-based panels (indeed, they are considered symbol-based traffic panels) and look like the text-based traffic panels, as shown in Figure 13.



Figure 11: Instances of the FP panels



Figure 12: Instances of correctly detected text-based traffic panels.

As depicted in Figure 14, other than a few cases mentioned as the algorithm's flaws, it does not incorrectly detect the other similar objects to the text-based traffic panels such as the advertisements and billboards on the sides of the roads and vehicles, store name board, etc. It

can accurately detect the text-based traffic panels in the complicated backgrounds as shown in Figure 15. It can correctly detect the lopsided text-based traffic panels, as shown in Figure 16. As shown in Figure 17, it can detect the text-based traffic panels partially blocked by objects, including vehicles, other panels, trees, etc.



Figure 13: Instances of the correctly detected text-based traffic panels and other similar panels beside them.



Figure 14: Instances of the correctly detected text-based traffic panels with complicated backgrounds.



Figure 15: Instances of the correctly detected lopsided text-based traffic panels.



Figure 16: Instances of the partially blocked text-based traffic panels that have been correctly detected.

The tiny YOLOv3 algorithm was exploited for pre-training, training, and testing the data to examine the utility and advantages of the data. The K-fold cross-validation procedure was used to estimate the model's skill on the new data. It achieves 0.973 for Precision, 0.945 for Recall, and 0.955 for F_{measure} . Since the next step of text-based traffic panel detection is recognizing the text on the traffic panel, the authors plan to enhance text-based traffic panel detection accuracy for IoU thresholds greater than 0.5 to focus on recognizing the texts in the dataset as the future study.

References

- [1] A. Gudigar, S. Chokkadi and U. Raghavendra, "A review on automatic detection and recognition of traffic sign," *Multimedia Tools and Applications*, vol. 75, no. 1, pp. 333-364, 2016.
- [2] J. Stallkamp, M. Schlipsing, J. Salmen and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *International joint conference on neural networks (IJCNN)*, 2016.
- [3] F. Zaklouta and B. Stanculescu, "Real-time traffic sign recognition in three stages," *Robotics and autonomous systems*, vol. 62, no. 1, pp. 16-24, 2014.
- [4] J. Greenhalgh and M. Mirmehdi, "Real-Time Detection and Recognition of Road Traffic Signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498 -- 1506, 2012.
- [5] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," *The 2013 international joint conference on neural networks (IJCNN)*, pp. 1-8, 2013.
- [6] R. Timofte, K. Zimmermann and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Machine vision and applications*, vol. 25, no. 3, pp. 633 - 647, 2014.
- [7] Y. Yang, H. Luo, H. Xu and F. Wu, "Towards Real-Time Traffic Sign Detection and Classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 2022 - 2031, 2016.
- [8] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-Sign Detection and Classification in the Wild," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [9] E. Ayoub, M. El Ansari and I. El Jaafari, "Traffic sign detection and recognition based on random forests," *Applied Soft Computing*, vol. 46, pp. 805 -- 815, 2016.
- [10] Y. Yuan, Z. Xiong and Q. Wang, "An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1918 - 1929, 2017.
- [11] U. Kamal, T. Islam Tonmoy, S. Das and M. K. Hasan, "Automatic Traffic Sign Detection and Recognition Using SegU-Net and a Modified Tversky Loss Function With L1-Constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1467 - 1479, 2020.
- [12] C. Gamez Serna and Y. Ruichek, "Traffic Signs Detection and Classification for European Urban Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4388-4399, 2020.
- [13] L. Wei, C. Xu, S. Li and X. Tu, "Traffic Sign Detection and Recognition Using Novel Center-Point Estimation and Local Features," *IEEE Access*, vol. 8, pp. 83611-83621, 2020.
- [14] D. Tabernik and D. Škočaj, "Deep Learning for Large-Scale Traffic-Sign Detection and Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1427 - 1440, 2020.
- [15] F. Almutairy, T. Alshaabi, J. Nelson and S. Wshah, "ARTS: Automotive Repository of Traffic Signs for the United States," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 457 - 465, 2021.
- [16] C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang and K. D. Hartomo, "Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN," *IEEE Access*, vol. 9, pp. 97228-97242, 2021.
- [17] X. Liu, X. Jiang, H. Hu, R. Ding, H. Li and C. Da, "Traffic Sign Recognition Algorithm Based on Improved YOLOv5s," in *2021 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2021.
- [18] Y. Sun and L. Chen, "Traffic Sign Recognition Based on CNN and Twin Support Vector Machine Hybrid Model," *Journal of Applied Mathematics and Physics*, vol. 9, no. 12, pp. 3122-3142, 2021.
- [19] Z. Wang, J. Wang, Y. Li and S. Wang, "Traffic Sign Recognition With Lightweight Two-Stage Model in Complex Scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1121-1131, 2022.
- [20] S. Ahmed, U. Kamal and M. K. Hasan, "DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection Under Challenging Weather Conditions," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-13, 2021.
- [21] Z. Nadeem, A. Samad, Z. Abbas and J. Massod, "Pakistani traffic-sign recognition using transfer learning," in *2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2022.
- [22] Z. Liu, M. Qi, C. Shen, Y. Fang and X. Zhao, "Cascade saccade machine learning network with hierarchical classes for traffic sign detection," *Sustainable Cities and Society*, vol. 67, p. 102700, 2021.
- [23] J. Wan, W. Ding, H. Zhu, M. Xia, Z. Huang, L. Tian, Y. Zhu and H. Wang, "An efficient small traffic sign detection method based on yolov3," *Journal of Signal Processing Systems*, vol. 93, no. 8, pp. 899--911, 2021.
- [24] D. Nienhüser, T. Gumpf, J. M. Zöllner and K. Natroshvili, "Fast and reliable recognition of supplementary traffic signs," in *Intelligent Vehicles Symposium*, 2010.
- [25] A. S. Puthon, F. Moutarde and F. Nashashibi, "Recognition of supplementary signs for correct interpretation of traffic signs," in *Intelligent Vehicles Symposium (IV)*, 2013.
- [26] T. Wenzel, S. Brueggert and J. Denzler, "Additional Traffic Sign Detection -- A Comparative Study," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015.
- [27] T. Wenzel, S. Brueggert and J. Denzler, "Towards unconstrained content recognition of additional traffic signs," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017.
- [28] T. Wenzel, S. Brueggert and J. Denzler, "Additional traffic sign detection using learned corner representations," in *Intelligent Vehicles Symposium (IV)*, 2016.
- [29] T. Weber, E. Ercelik, M. Ebert and A. Knoll, "Recognition and Evaluation of Additional Traffic Signs on the example of '80 km/h when wet'," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019.
- [30] O. Hamdoun, A. Bargeton, F. Moutarde, B. Bradai and L. Chanussot, "Recognition of End-of-Speed-Limit and Supplementary Signs for Improved Speed Limit Support," in *15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting*, 2008.
- [31] X. Rong, C. Yi and Y. Tian, "Recognizing text-based traffic guide panels with cascaded localization network," in *European Conference on Computer Vision*, 2016.
- [32] O. de G. Henrique, F. A. da Silva, D. R. Pereira, L. L. de Almeida, A. O. Artero, A. F. Bonora and V. H. C. de Albuquerque, "Automatic Detection and Recognition of Text-Based Traffic Signs from images," *IEEE Latin America Transactions*, vol. 16, no. 12, pp. 2947 - 2953, 2018.
- [33] H. Eun, J. Kim, J. Kim and C. Kim, "Fast Korean Text Detection and Recognition in Traffic Guide Signs," in *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018.
- [34] R. Jain and D. Gianchandani, "A Hybrid Approach for Detection and Recognition of Traffic Text Sign using MSER and OCR," in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2018 2nd International Conference on, 2018.
- [35] W. Wu, X. Chen and J. Yang, "Detection of text on road signs from video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 4, pp. 378 -- 390, 2005.
- [36] Á. González, L. M. Bergasa, J. J. Yebes and J. Almazán, "Text recognition on traffic panels from street-level imagery," in *Intelligent Vehicles Symposium*, 2012.
- [37] J. Greenhalgh and M. Mirmehdi, "Recognizing Text-Based Traffic Signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 13, pp. 1360 -- 1369, 2015.
- [38] A. Gonzalez, L. M. Bergasa and J. J. Yebes, "Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 228 -- 238, 2014.

- [39] J. Guo, R. You and L. Huang, "Mixed Vertical-and-Horizontal-Text Traffic Sign Detection and Recognition for Street-Level Scene," *IEEE Access*, vol. 8, pp. 69413-69425, 2020.
- [40] K. S. Boujemaa, M. Akallouch, I. Berrada, K. Fardousse and A. Bouhoute, "ATTICA: A Dataset for Arabic Text-Based Traffic Panels Detection," *IEEE Access*, vol. 9, pp. 93937-93947, 2021.
- [41] R. Bagi, T. Dutta, N. Nigam, D. Verma and H. P. Gupta, "Met-MLTS: Leveraging Smartphones for End-to-End Spotting of Multilingual Oriented Scene Texts and Traffic Signs in Adverse Meteorological Conditions," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-10, 2021.
- [42] H. Luo, Y. Yang, B. Tong, F. Wu and B. Fan, "Traffic sign recognition using a multi-task convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1100-1111, 2017.
- [43] X. Peng, X. Chen and C. Liu, "Real-time Traffic Sign Text Detection Based on Deep Learning," in *Materials Science and Engineering*, 2020.
- [44] J. Zhang, Z. Xie, J. Sun, X. Zou and J. Wang, "A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection," *IEEE Access*, vol. 8, pp. 29742-29754, 2020.
- [45] Korghond, N. Khazaei and R. Safabakhsh, "AUT-UTP: Urban traffic panel detection and recognition dataset," in *24th Iranian Conference on Electrical Engineering (ICEE)*, 2016.
- [46] S. Kheirinejad, N. Riahi and R. Azmi, "Persian Text Based Traffic sign Detection with Convolutional Neural Network: A New Dataset," in *10th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2020.
- [47] A. Soetedjo and K. Yamada, "A new approach on red color thresholding for traffic sign recognition system," *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 19, no. 5, pp. 457-465, 2007.
- [48] A. Mogelmoose, M. M. Trivedi and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484 - 1497, 2012.
- [49] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gómez-Moreno and F. López-Ferreras, "Road-Sign Detection and Recognition Based on Support Vector Machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264 -278, 2007.
- [50] Y. Li, S. Pankanti and W. Guan, "Real-time traffic sign detection: an evaluation study," in *2010 20th International Conference on Pattern Recognition*, 2010.
- [51] C. G. Kiran, L. V. Prabhu and K. Rajeev, "Traffic sign detection and pattern recognition using support vector machine," in *2009 Seventh International Conference on Advances in Pattern Recognition*, 2009.
- [52] S. M. Bascón, J. A. Rodríguez, S. L. Arroyo, A. F. Caballero and F. López-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs," *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 373-383, 2010.
- [53] P. Gil-Jimenez, H. Gomez-Moreno, P. Siegmann, S. Lafuente-Arroyo and S. Maldonado-Bascón, "Traffic sign shape classification based on Support Vector Machines and the FFT of the signature of blobs," in *2007 IEEE Intelligent Vehicles Symposium*, 2007.
- [54] I. M. Creusen, R. G. Wijnhoven, E. Herbschleb and P. H. N.de, "Color exploitation in hog-based traffic sign detection," in *2010 IEEE International Conference on Image Processing*, 2010.
- [55] Y. Wu, Y. Liu, J. Li, H. Liu and X. Hu, "Traffic sign detection based on convolutional neural networks," in *International joint conference on neural networks (IJCNN)*, 2013.
- [56] R. Qian, B. Zhang, Y. Yue, Z. Wang and F. Coenen, "Robust chinese traffic sign detection and recognition with deep convolutional neural network," in *11th International Conference on Natural Computation (ICNC)*, 2015.
- [57] Z. Di, J. Zhang, D. Zhang, M. Bao, J. Cheng and K. Tang, "Traffic sign detection based on cascaded convolutional neural networks," in *17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2016.
- [58] L. Abdi and A. Meddeb, "Deep learning traffic sign detection, recognition and augmentation," in *Proceedings of the Symposium on Applied Computing*, 2017.
- [59] Z. Wang, J. Wang, Y. Li and S. Wang, "Traffic Sign Recognition With Lightweight Two-Stage Model in Complex Scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1121-1131, 2022.
- [60] A. V. Reina, R. L. Sastre, S. L. Arroyo and P. G. Jiménez, "Adaptive traffic road sign panels text extraction," in *International Conference on Signal Processing, Robotics and Automation*, 2006.
- [61] A. Khodayari and M. Yousefi, "Design an Intelligent Driver Assistance System Based On Traffic Sign Detection with Persian Context," *International Journal of Automotive Engineering*, vol. 6, no. 2, pp. 2138 -2147, 2016.
- [62] Y. Zhu, M. Liao, M. Yang and W. Liu, "Cascaded Segmentation-Detection Networks for Text-Based Traffic Sign Detection," *IEEE Transaction Intelligent Transportation Systems*, vol. 19, no. 1, pp. 209 - -219, 2018.
- [63] H. Zhang, L. Qin, J. Li, Y. Guo, Y. Zhou, J. Zhang and Z. Xu, "Real-Time Detection Method for Small Traffic Signs Based on Yolov3," *IEEE Access*, vol. 8, pp. 64145-64156, 2020.
- [64] Z. Yi, S. Yongliang and Z. Jun, "An improved tiny-yolov3 pedestrian detection algorithm," *Optik*, vol. 183, pp. 17 -23, 2019.
- [65] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, . C. Y. Fu and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016.
- [66] X. Wang, A. Shrivastava and A. Gupta, "A-fast-rcnn: Hard positive generation via adversary for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [67] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE conference on computer vision and pattern recognition*, 2016.
- [68] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *IEEE conference on computer vision and pattern recognition*, 2017.
- [69] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," in *arXiv preprint arXiv*, 2018.
- [70] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [71] J. Nelson and J. Solawetz, "Responding to the Controversy about YOLOv5," <https://blog.roboflow.com/yolov4-versus-yolov5/>, 2020.
- [72] M. Andreas, . M. M. Trivedi and . T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1484 -- 1497, 2012.
- [73] X. Baró, S. Escalera, J. Vitria, O. Pujol and P. Radeva, "Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 113 - 126, 2009.
- [74] M. Benallal and J. Meunier, "Real-time color segmentation of road signs," in *CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology*, 2003.
- [75] G. H. de Oliveira, F. A. da Silva, D. R. Pereira, L. L. de Almeida, A. O. Artero, A. F. Bonora and V. H. C. de Albuquerque, "Automatic Detection and Recognition of Text-Based Traffic Signs from images," *IEEE Latin America Transactions*, vol. 16, no. 12, pp. 2947-2953, 2018.
- [76] A.-S. Puthon, F. Moutarde and F. Nashashibi, "Recognition of supplementary signs for correct interpretation of traffic signs," in *2013 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, 2013.
- [77] A. Ruta, F. Porikli, S. Watanabe and Y. Li, "In-vehicle camera traffic sign detection and recognition," *Machine Vision and Applications*, vol. 22, no. 2, pp. 359 - 375, 2011.
- [78] T. Wenzel, S. Brueggert and J. Denzler, "Towards unconstrained content recognition of additional traffic signs," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017.
- [79] M. Zadeh, M. T. Kasvand and C. Y. Suen, "Localization and recognition of traffic signs for automated vehicle control systems," *Intelligent Transportation Systems*, vol. 3207, pp. 272-282, 1998.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Saba Kheirinejad received the M.Sc. degree in Artificial Intelligence and Robotics from Alzahra University, Tehran, Iran. She is currently pursuing a Ph.D. degree with the Centre for Ubiquitous Computing, University of Oulu, Finland. Her current research interests include ubiquitous computing, human-computer interaction, and digital health.



Noushin Riahi received the B.S. degree from Isfahan University of Technology in 1986 and the M.S. and Ph.D. degree in Electrical and Electronic Engineering from Sharif University of Technology, Tehran, Iran in 1990 and 1998. Her research interests include speech and sound processing, opinion mining, recommender systems, social network processing, and biological signal processing.

Currently, she is associate professor at Alzahra University, computer group. Topics taught by her include speech processing, advance computer architecture, digital signal processing, natural language processing, signals and systems theory, microprocessors, digital design, etc.



Dr. Reza Azmi received his BS degree in Electrical Engineering from Amirkabir University of Technology (Tehran polytechnic) in 1990 and his MS and PHD degrees in electrical and computer Engineering from Tarbiat Modares University, Tehran, Iran in 1993 and 1999 respectively. Since 2001 he has joined computer department of Faculty of engineering at Alzahra University. He was

member of multimedia systems working group in ITRC, Optical Character Recognition working group in supreme council of information and communication technology, he was manager and technical member of many industrial projects and four US international patents. Dr. Azmi is founder and co-founder in three startup companies. He has received numerous awards for being selected as a top industrial researcher.

Bearing Fault Diagnosis Based on Ensemble Depth Explainable Encoder Classification Model with Arithmetic Optimized Tuning

Kaibi Zhang, Yanyan Wang, Hongchun Qu*

Chongqing University of Posts and Telecommunications, College of Automation, Chongqing, 400065, China

*Corresponding author: Hongchun Qu, ORCID: <https://orcid.org/0000-0001-7623-2383>, Email: hcchyu@gmail.com

ABSTRACT: In a dynamic and complex bearing operating environment, current auto-encoder-based deep models for fault diagnosis are having difficulties in adaptation, which usually leads to a decline in accuracy. Besides, the opaqueness of the decision process by such deep models might reduce the reliability of the diagnostic results, which is not conducive to the subsequent optimization of the model. In this work, an ensemble deep auto-encoder method is developed and tested for intelligent fault diagnosis. To mitigate the influence of the changing operating environment on the diagnostic accuracy of the model, a tuning algorithm is used to adaptively adjust the parameters of the model, and a hypersphere classification algorithm is used to separately train different types of fault data. The encoder components in the ensemble model are automatically updated based on the diagnostic accuracy of the base encoder model under different operating conditions. To improve the reliability of the diagnosis results, the power spectrum analysis and Layer-wise Relevance Propagation algorithm are combined to explain the diagnosis results. The model was validated on three public datasets and compared with individual encoder methods as well as other common fault diagnosis algorithms. The results confirm that the model proposed is flexible enough to cope with changes in operating conditions and has better diagnostic and generalizing capabilities.

KEYWORDS: Fault diagnosis, Ensemble model, Dynamic composition, Deep auto-encoder, Layer-wise Relevance Propagation

1. Introduction

Rolling bearings, whose health status affects the state of the running equipment, is one of the most common parts in industrial machinery [1]. Traditional fault diagnostic methods based on signal processing usually require employees to have not only complete knowledge of relevant industries [2], but also the ability of signal process and analysis [3].

As an intelligent diagnosis method that can automatically learn feature representation, deep auto-encoder has received extensive attention from scholars because it can reduce the requirement for practitioners and improve the accuracy of fault diagnosis when used for fault diagnosis.

Deep auto-encoder models originally used for fault diagnosis were typically stacked from individual-based

encoder models, such as sparse auto-encoders (SAE)[4], compression auto-encoders (CAE) [5], denoising auto-encoders (DAE) [6], or their variants [7-9]. This kind of model can give full play to its advantages and obtain excellent diagnostic results when dealing with relatively simple data and less noise impact. However, bearings usually work in complex and noisy environments, so those traditional models often fail to accurately diagnose the fault signals with multiple jamming signals collected in real production environments.

To deal with such complex fault signals, some researchers combined different auto-encoders. For example, DAE and CAE are put together to form a new deep auto-encoder [10], or the integration of the three basic models SAE, DAE and CAE with different weight [11]. Compared with individual encoder models, these models have higher diagnostic accuracy when handling fault

signals with noises. However, bearings are always in a changing environment during actual operation, and the diagnostic accuracy of such models are unstable due to their static structures and parameters. Some researchers have used tuning algorithms, such as particle swarm optimization [12-15] and cuckoo optimization[16], to optimize parameters under different working conditions. The diagnostic performance of these optimized models under different working conditions has been improved to a certain extent. However, the changing operating environment for bearings may also produce other factors that are detrimental to the diagnostic capability of the model, such as the unbalanced distribution of fault data samples [17]. Therefore, how to accurately detect various fault signals in the changing operating environment is still a major challenge in the field of fault diagnosis.

In addition, although the deep auto-encoder can provide high precision diagnosis results, the diagnosis results may not be trusted by experts at some point due to the opacity of its decision process. Only when users understand the reasons behind the model's diagnostic behavior, can they fully trust the model and make a reliable decision according to the model's diagnostic results [18]. Besides, it is difficult to optimize and migrate the model because the information of its training and decision-making process is usually hard to be reserved. One way to interpret deep models is to introduce additional modules into the model [19,20] that can directly output the diagnostic reasons during the fault diagnosis process. However, this approach will make the model more complex and requires more training time and datasets [21]. Another way is to use the ex-post interpretation model, which is retrospective after the decision is made [22, 23]. As a post-hoc interpretation model designed for computer vision, Layer-wise Relevance Propagation (LRP) [24-27] has been used to interpret fault diagnosis results based on time-domain data and convolution network. However, many current deep models for fault diagnosis are trained and validated based on frequency-domain data. To the best of our knowledge, LRP has not been used for diagnostic models based on ensemble deep auto-encoder and frequency-domain data.

Therefore, in this paper, we propose an ensemble deep auto-encoder model, ALEDA, to improve the accuracy of fault diagnosis models under various operating conditions and the confidence of diagnostic results. Firstly, based on the arithmetic optimization algorithm[28], the parameters of the encoder model are optimized, which realizes the adaptive adjustment of the model parameters under different working conditions. Secondly, based on the hypersphere algorithm[29, 30], each type of data is trained separately, which alleviates the problem of unbalanced distribution of fault data samples caused by

heterogeneities of working conditions. Then, the encoders are combined according to the diagnostic accuracy of the basic models optimized in specific condition, which is helpful to enhance the adaptability in a changing environment. Finally, the diagnostic results of the model are explained by using power spectral analysis and Layer-wise Relevance Propagation algorithm. It not only improves the reliability of the diagnosis results, but also provides enlightenment for interpreting the diagnosis results of the fault diagnosis model based on the frequency domain data. The model is validated on three public datasets and compared with other state-of-the-art fault diagnosis algorithms. The results demonstrate that the proposed model can flexibly respond to changes in working environments and has better capabilities of diagnosis and generalization.

The rest of this paper is organized as follows. Section 2 briefly introduces the basic theory of related methods. Section 3 describes the proposed model in detail. Section 4 verifies the effectiveness of the proposed model on three datasets, and analyzes and discusses the experimental results. Finally, conclusions are given in Section 5.

2. Related work

2.1. Arithmetic Optimization Algorithm

Compared with those classic optimization algorithms, the Arithmetic Optimization Algorithm[28] (AOA) is a new type of optimization algorithm whose effectiveness has not been verified in fault diagnosis models. And it is a group-based meta-heuristic optimization algorithm mainly including two stages of exploration and exploitation. In the exploration stage, the multiplication and division search strategy are mainly used to explore the search range to find the best solution. In the exploitation stage, the addition and subtraction search strategy are mainly used to optimize the solutions obtained in the previous phase. The algorithm defines two coefficients, one is *OPC* (Optimization Phase Control parameter), used to control the phase of the algorithm (1); the other is *OSC* (Optimized speed control parameter), used to control the updating speed of particle position (2).

$$OPC(i) = Min_OPC + i \times \frac{Max_OPC - Min_OPC}{Max_i} \quad (1)$$

$$OSC(i) = 1 - \frac{C - i^{1/\beta}}{M - i^{1/\beta}} \quad (2)$$

where i is the current iteration. Max_i represents the maximum number of iterations of AOA algorithm. Max_OPC indicates the maximum value of *OPC*, which is set to 0.9. Min_OPC represent the minimum values of *OPC*, which is set to 0.2. And β defines the development precision over iterations, which is fixed at 0.5.

The number k_1, k_2, k_3 is randomly selected between 0 and 1. If $k_1 > OPC(i)$, the algorithm enters the exploration

stage. And if $k_2 < 0.5$, the position of the particle is updated according to (3), otherwise according to (4).

$$X_{i,j}(i+1) = best(x_j) \div (OSC(i) + 0.01) \times ((ub_j - lb_j) \times \mu + lb_j), k_2 < 0.5 \quad (3)$$

$$X_{i,j}(i+1) = best(x_j) \times OSC(i) \times ((ub_j - lb_j) \times \mu + lb_j), otherwise \quad (4)$$

Where ub_j and lb_j are used to limit the optimization range of j_{th} parameter, μ is used to control the speed of position updates in the search phase, which is set to 0.5.

And when $k_1 < OPC(i)$, the algorithm enters the exploration stage. And if $k_3 < 0.5$, the position of the particle is updated according to (5), otherwise (6).

$$X_{i,j}(i+1) = best(x_j) - OSC(i) \times ((ub_j - lb_j) \times \mu + lb_j), k_3 < 0.5 \quad (5)$$

$$X_{i,j}(i+1) = best(x_j) + OSC(i) \times ((ub_j - lb_j) \times \mu + lb_j), otherwise \quad (6)$$

2.2 Auto-encoder

2.2.1 Sparse Auto-encoder (SAE)

SAE is built by stacking several sparse auto-encoders, where each sparse auto-encoder consists of an encoder and a decoder. The encoder can convert the input into a feature representation, while the decoder can reconstruct the input. Suppose the training set is $\{x^i\}_{i=1}^K$ where K is the number of samples. The feature representation h^i and reconstruction $\{y^i\}_{i=1}^K$ can be expressed as (7) and (8).

$$h^i = f(W_E x^i + b_E), i = 1, 2, \dots, k \quad (7)$$

$$y^i = f(W_D h^i + b_D), i = 1, 2, \dots, k \quad (8)$$

Where $f(\cdot)$ is the activation function, W_E and W_D are the weight matrix of the encoder and decoder respectively, b_E and b_D are the bias vectors.

For each sparse auto-encoder, the cost function is given in (9).

$$J_{sae} = \frac{1}{2k} \sum_{i=1}^k \|y^i - x^i\|^2 + \lambda (\sum_{i,j} W_{E,i,j}^2 + \sum_{i,j} W_{D,i,j}^2) + \beta \sum_k \rho \log \frac{k\rho}{\sum_{i=1}^k h_n^i} + (1-\rho) \log \frac{k(1-\rho)}{k - \sum_{i=1}^k h_n^i} \quad (9)$$

Where x^i is the input of the encoder, y^i is reconstruction, λ is the coefficient specified by the user, β is the coefficient of the sparse penalty term, and ρ is the sparse factor.

2.2.2 Denoising Auto-encoder (DAE)

The DAE is constructed by stacking several denoising auto-encoders that learn feature representations and reconstruct data in the same way as a SAE. Unlike SAE, during training, data with noises is fed into the DAE. The noise input and cost function of each denoising auto-encoder can be expressed as (10) and (11), and the meanings of parameters are the same as those of SAE.

$$\hat{x} = x + noise(x) \quad (1)$$

$$J_{dae} = \frac{1}{2k} \sum_{i=1}^k \|y^i - x^i\|^2 + \lambda (\sum_{i,j} W_{E,i,j}^2 + \sum_{i,j} W_{D,i,j}^2) \quad (2)$$

2.2.3 Compression Auto-encoder (CAE)

CAE is constructed by stacking several compression auto-encoders which can learn more robust feature

representation by adding a compression penalty term to its cost function. For each compression auto-encoder, the cost function is given in (12) and (13), and the meanings of parameters are the same as those of SAE.

$$J_{cae} = \frac{1}{2k} \sum_{i=1}^k \|y^i - x^i\|^2 + \delta \sum_{i=1}^k \|J_f(x^i)\|_F^2 \quad (12)$$

$$J_f(x^i) = \begin{bmatrix} \frac{\partial h_1^i}{\partial x_1^i} & \dots & \frac{\partial h_1^i}{\partial x_n^i} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m^i}{\partial x_1^i} & \dots & \frac{\partial h_m^i}{\partial x_n^i} \end{bmatrix} \quad (13)$$

2.3 Dynamic Hypersphere Algorithm

The dynamic hypersphere algorithm[29] refers to the use of perceptron to construct a dynamic feature space for each type of health training data separately, and to construct a corresponding hypersphere on the feature space of each type of data for their aggregation. By continuously reducing the error of the classification results, and updating the parameters, the hypersphere can be updated dynamically so that as many similar points as possible are constrained in the smallest possible sphere. Each class of data corresponds to a hypersphere. The dynamic hypersphere algorithm can not only take advantage of the perceptron on non-linear problems, but also perform well on unbalanced data by training and optimizing each type of data separately. Suppose there are M types of data, and each type of data has $N_i (i = 1, 2, \dots, m)$ samples. The initial value of C_i (the center of the sphere i) can be obtained from (14) and d_i (the distance from the sample to the center of the sphere i) can be expressed as (15).

$$C_i = \frac{1}{N_i} \sum_{j=1}^{N_i} U_{i,j} \quad (14)$$

$$d_i = \sum_{j=1}^{N_i} \|U_{i,j} - C_i\| \quad (15)$$

Where $U_{i,j}$ represents the feature representation of j neurons in the last layer of the encoder for the i_{th} class of training data, in other words, $U_{i,j}$ is a reduced dimensional representation of the input data of class i .

The loss function $L(\theta)$ is defined as (16).

$$L(\theta) = p_1 (\sum_{i=1}^m D_i + D_{-i} + \text{ReLU}(-R_i)^2) + p_2 \sum_{i=1}^m D_{i,j} + p_3 \sum_{i=1}^m d_{i,j} \quad (16)$$

Where p_1 is the penalty coefficient of each sphere, p_2 is the penalty coefficient between spheres, and p_3 is the accelerated convergence coefficient that plays a role in controlling the degree of punishment for sample segmentation. And R_i is the radius of sphere i , which should be greater than 0.

D_i (17) represent the total distance from the i_{th} sample to the center i , D_{-i} (18) represents the total distance between the data that does not belongs to class i and the

center i , and $D_{i,j}$ (19) is the separation distance between sphere i and sphere j .

$$D_i = \left(\sum_{j=1}^{N_i} \text{Relu}(\|U_{i,j} - C_i\| - R_i) \right)^2 \quad (17)$$

$$D_{-i} = \left(\sum_{j=1}^{N_i} \text{Relu}(R_i - \|U_{-i,j} - C_i\|) \right)^2 \quad (18)$$

$$D_{i,j} = \left(\sum_{i=1}^m \sum_{j=i+1}^m ((R_i - R_j) - \|C_i - C_j\|) \right)^2 \quad (19)$$

Calculate the distance from the new sample point to the center of each hypersphere, and take the class in which the closest hypersphere is located as the class of the new sample point.

2.4 Layer-wise Relevance Propagation (LRP)

As an anomaly interpretation technique, LRP can provide correlation between input signals and diagnosis results. The greater the contribution of input layer neurons to model diagnosis, the higher the correlation score obtained during back propagation. By visualizing the correlation scores, the input neurons that contribute significantly to the output results of the model can be highlighted. The transmission mechanism of LRP is as follows:

It is known that the correlation R_j^{l+1} of neuron j at layer $l+1$ can be decomposed to all neurons at layer l . The greater the contribution of neuron i at layer l to neuron j at layer $l+1$ at the stage of fault diagnosis, the higher the correlation score can be divided. And R_j^{l+1} can be expressed as (20).

$$R_j^{l+1} = \sum_{i \in l} R_{i \leftarrow j}^{(l,l+1)} \quad (20)$$

After the correlation of all neurons in layer $l+1$ is decomposed, the correlation R_j^l can be obtained by summation of all correlations obtained by neuron i in layer l , and its mathematical expression is given in (21).

$$R_j^l = \sum_{j \in (l+1)} R_{i \leftarrow j}^{(l,l+1)} \quad (21)$$

Further, the correlation coefficient $R_{i \leftarrow j}^{(l,l+1)}$ can be obtained through ε -rule, and its specific mathematical expression is given in (22).

$$R_{i \leftarrow j}^{(l,l+1)} = \frac{z_{ij}}{z_j + \varepsilon \text{sign}(z_j)} R_j^{l+1} \quad (22)$$

$R_{i \leftarrow j}^{(l,l+1)}$ can be understood as the contribution of layer l neuron i to layer $l+1$ neuron j , where z_{ij} is the weighted activation of layer $l+1$ neuron j by neuron i of layer l , and z_j is the weighted activation of layer $l+1$ neuron j by all neurons of layer l .

3. The proposed ALEDA method for intelligent fault diagnosis

To address the difficulties faced by deep auto-encoder models in fault diagnosis, we proposed an ensemble deep

auto-encoder model with AOA and LRP, named ALEDA, which integrates DAE, CAE and SAE. DAE is used to learn useful information from input signals with noises; CAE is used to learn more robust feature representations. SAE is used to reduce the risk of over-fitting and a dropout layer is append to each hidden layer to enhance the ability. The training process of ALEDA model mainly includes four parts: optimizing the base encoder models to the best performance, determining encoder components of the ensemble model, obtaining the classification results of the ensemble model, and interpreting diagnosis results.

3.1 Optimizing the base encoder models

To maximize the diagnostic accuracy of the ensemble model in the face of a new operating environment, it is necessary to adjust the diagnostic accuracy of each encoder component to the maximum. The process can be divided into the following steps:

- 1) Determine the parameters to be optimized and the optimization algorithm to be used.

Since the learning rate affects the convergence of the model, and the number of nodes in the hidden layer directly affects the structure of the model, we use them as parameters to be optimized. To verify the effectiveness of the AOA algorithm in passing, we choose the AOA algorithm to automatically adjust the parameters of the model.

- 2) Determine the optimization range of the parameters.

A suitable search range will speed up the algorithm optimization. In this paper, empirical formula of neural network nodes of hidden layer (23) and pyramid geometric rules (24) and (25) are used to limit the optimization range of nodes of the hidden layer.

$$h_{\max}(k) = \sqrt{0.55h_{\max}^2(k-1) + 3.31h_{\max}(k-1) + 0.35} + 0.51, k=1,2 \quad (23)$$

$$h_{\min}(1) = h_{\text{out}} \left(\frac{h_{\text{input}}}{h_{\text{out}}} \right)^{2/3} \quad (24)$$

$$h_{\min}(2) = h_{\text{out}} \left(\frac{h_{\text{input}}}{h_{\text{out}}} \right)^{1/3} \quad (25)$$

where $h_{\max}(k)$ represents the maximum number of nodes of the k_{th} hidden layer, $h_{\min}(1)$ and $h_{\min}(2)$ represent the minimum number of nodes of 1_{th} and 2_{th} hidden layers respectively, h_{input} and h_{out} are the number of nodes in the input layer and output layer respectively.

- 3) Determine the objective function of the optimization algorithm.

Since the goal of the optimization algorithm is to make the value of the loss function of the classifier as small as possible, it's necessary to firstly determine the classifier.

The hypersphere classifier can transform the same class of data into corresponding hyperspheres by training and optimizing each class of data separately, which alleviates the impact of data imbalance problems caused by changes in bearing operating conditions. Thus, the loss function of the hypersphere classifier (16) is used as the objective function of AOA. In addition, to take advantage of the encoder model while retaining the nonlinear advantage of the dynamic hypersphere algorithm, the encoder model is used instead of the original single-layer perceptron. The pseudo-code of parameter optimization algorithm for encoder network is given in Table 1.

3.2. Determining encoder components of the ensemble model

The next step is to design a strategy to combine the three encoders into an ensemble model so as to take full advantage of the three encoders to cope with the changing work environment. The strategies adopted in this paper are as follows:

Three auto-encoders E_i are selected to form the final ensemble model. First, $N_i (i = 1, 2, 3)$ samples were randomly selected from the training set for dividing a new training set and the validation set by cross-validation. Next, SAE, DAE and CAE are used to learn the characteristics of the training set, and the hypersphere classifier is used to classify the training set, and the weights of the three basic encoders are updated according to (26). With each fold of cross-validation, the weight of each encoder changes dynamically. After the cross-validation, the encoder model with the highest weight is selected to compose the ensemble model, and it is trained with sample N_i for subsequent testing. The specific process is shown in Figure 1. The updating formula of weight w_j is given in (26), where the initial value is given in (27).

$$w_{j+1} = \frac{w_j \square e^{\rho \square acc}}{\sum w_j \square e^{\rho \square acc}}, j = 0, 1, 2 \quad (26)$$

$$w_0 = \{1/3, 1/3, 1/3\} \quad (27)$$

where ρ is used to control the change degree of w_j , acc is the verified accuracy.

3.3. Obtaining the classification results of the ensemble model

To get the diagnostic results of the ensemble model, the classification results of the three base encoder models need to be integrated. In this paper, Naive Bayes is used to further judge the classification results of the three classifiers to determine the final classification results. The specific process is given in Figure 2.

Table 1: The pseudo-code of the algorithm optimizing encoder parameters

Algorithm 1 Optimizing encoder parameters

Result: get the best parameters of encoders

```

Initialize the encoder's parameters
randomly using (21) – (23)
Determine the number of particles ( $p_n$ ) and maximum
iteration ( $Max\_i$ ) of AOA algorithm

for  $i$  to 3 do:
Train the encoder network with initial parameters and
training set data;
Calculate the Fitness Function of initial parameters;
    for  $j$  to  $Max\_i$  do
        for  $k$  to  $p_n$  do
            update  $OPC$  and  $OSC$  using (1) and (2);
            for  $m$  to 4 do
                Randomly generated  $k_1, k_2, k_3$ ;
                if  $k_1 < OPC$  then
                    if  $k_2 > 0.5$  then
                        update the  $m_{th}$ 
parameter using (3);
                    else:
                        update the  $m_{th}$  parameter
using (4);
                    else:
                        if  $k_3 > 0.5$  then
                            update the  $m_{th}$ 
parameter using (6);
                        else:
                            update the  $m_{th}$  parameter
using (5);
                        end
                    end
                end
            end
        end
    end

```

It is assumed that there are k health conditions, and the total number of samples in the training set is N , and the sample number of each health condition is N_k . Firstly, the prior probability (28) of each type of sample is calculated and the Laplacian correction is made to it.

$$p_{pro}(j) = \frac{N_j + 1}{N + k}, j = 1, 2, \dots, k \quad (28)$$

Next, after the training of each classifier is completed, their confusion matrix (Table 2) is calculated as the conditional probability (29).

$$C_i(s_1, s_2) = \frac{N(s_1, s_2)}{N_{s_1} + 0.01}, i = 1, 2, 3 \quad (29)$$

Where N_{s_1} represents the number of samples actually labeled as s_1 in the training set; and $N(s_1, s_2)$ represents the number of samples actually labeled as s_1 but classified into s_2 by the classifier.

Then, the posterior probability (30) of each health conditions of each classifier is obtained. For each test sample, the classification result of the classifier with the largest posterior probability is its final classification result.

$$P_{pos}(j) = p_{pro}(j)C_i(j, s_j) \quad (30)$$

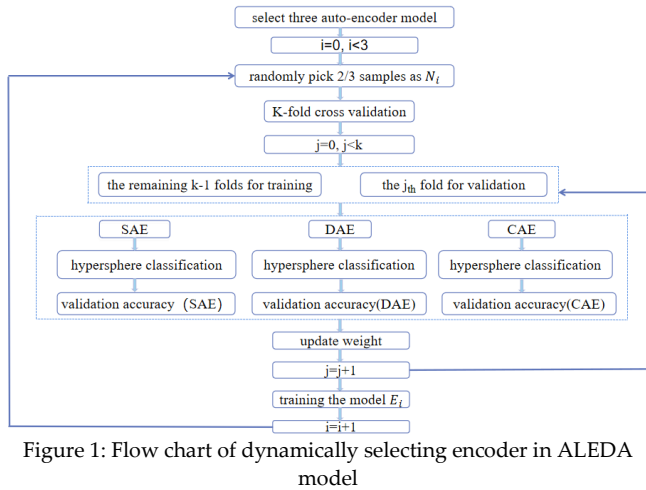


Figure 1: Flow chart of dynamically selecting encoder in ALEDA model

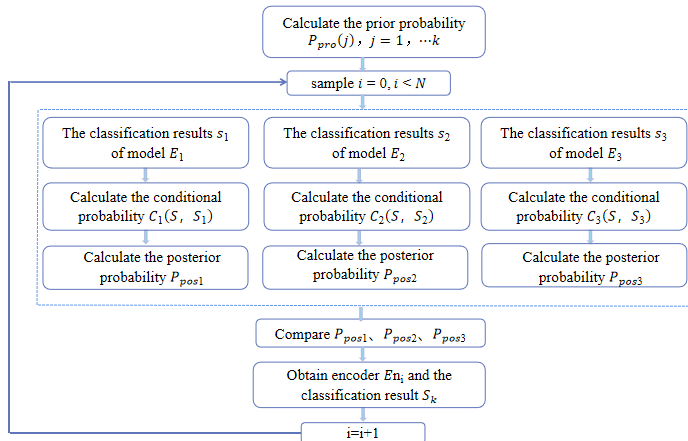


Figure 2: Flow chart of obtaining the classification results of the ensemble model

Table 2: The confusion matrix

Confusion Matrix		Predicted Value	
		Positive	Negative
Actual Value	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

3.4. Interpreting diagnosis results

To interpret the diagnostic results, the correlation between the final combined classification results of each test sample and the input features needs to be obtained. First, it is necessary to know how much each neuron contributes to the results of each encoder model during the

feature learning process of each hidden layer. Then, the correlation score between the classification results and the input features of each encoder model can be obtained. Next, the classifiers used in the previous stage are analyzed for each sample and the input layer correlation scores are recombined for each test sample. At last, the relationship between the diagnostic results of the ensemble model and the input features can be obtained. The specific process is given in Figure 3.

To more conveniently observe the prominent features of each test sample, the 100 neurons with the highest correlation score are visualized. Assuming a total of N test samples and M neurons, R_{mean} is defined as the mean value of the correlation score of neurons in the input layer, and its mathematical expression is given in (31).

$$R_{mean} = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m R_{i,j}^l \quad (31)$$

By counting the number of samples with correlation scores greater than R_{mean} in neuron $j(j = 1, 2, \dots, m)$ at the input layer, k neurons with the greatest contribution to the final classification results of each sample can be obtained.

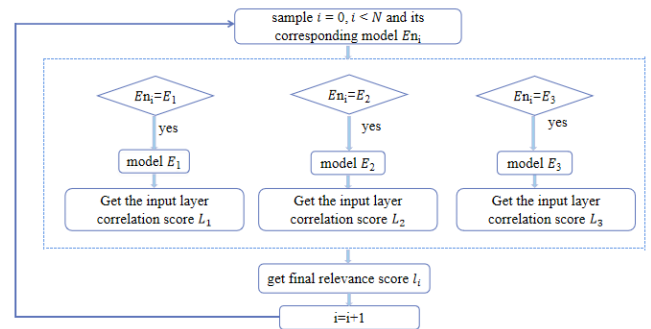


Figure 3: The process of getting the final relevance score

3.5. Algorithm pseudo-code

Table 3: Pseudo-code of ALEDA algorithm

Algorithm 2 ALEDA training

Result: final classification result of samples
 Initialize the encoder's parameter's randomly;
 Select the best parameter of three encoder by AOA;
for i in range (3) **do**
 Select 2/3 data randomly for training;
 Initialize weight = [1/3, 1/3, 1/3]
 Cross validation:
 Update weight;
 Select the encoder corresponding to the maximum value in the weight;
 Train three encoders and classification;

Calculate the posteriori probability for each category of failure in each encoder;

For each sample, the classification result of the classifier with the highest posteriori probability is selected as the final classification result;

4. Experimental verification

4.1. Data Preprocessing

Since the discriminant information in the time domain signal is not easy to be recognized, and the frequency domain signal in each health condition has its own different statistical characteristic parameters, the original vibration signal is converted into frequency domain signal by fast Fourier transform for further analysis and judgment. Since the frequency domain coefficients of the original data are symmetric after FFT transformation, half of the frequency domain data is used as the input of the model.

4.2. Experimental Design

It is well-known that bearing usually works in noisy working environment. Therefore, due to the influence of external noise, the quality of the collected data will be reduced, which directly affects the diagnostic effect of the model[7]. It is true that an excellent model should have good anti-noise ability. To evaluate the noise immunity performance of the model, the model was run separately in an additive White Gaussian noise (WGN) environment with different signal-to-noise ratios added. Besides, to reduce the influence of accidental factors on the experimental results, the average of five tests was taken as the result for each experiment.

We added WGN with SNR= 10dB, 5dB, -5dB and -10dB to the samples respectively to observe the influence of noise with different intensity on the sample (Figure 4). It can be observed that the frequency domain components of the original data are covered by noise after adding noise, especially when SNR = -10dB, the frequency component of the original signal is almost completely submerged and difficult to identify[7]. SNR is defined as (32).

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \quad (32)$$

where P_{signal} is the signal power and P_{noise} is the noise power.

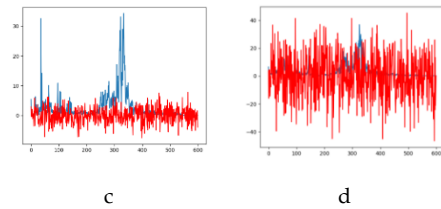
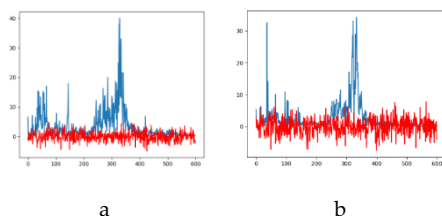


Figure 4: Power spectrum after adding noise (a) add 10dB noise (b) add 5dB noise (c) add -5dB noise (d) add -10dB noise

4.3 Evaluation indicators

Since multiple fault classes were considered in this paper, and the detection of each fault should be equally important, accuracy was selected as the main evaluation index. In addition, there may be data imbalance in the fault data, so precision and recall are used as evaluation indicators. The precision was used to judge the false positives of the model, and the recall rate was used to judge the false negatives of the model. Moreover, precision and recall restrict and influence each other, and F1 score takes both of them into consideration, so it was also taken as one of evaluation indicators. Their definitions are given in (33) – (36).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (33)$$

$$Precision = \frac{TP}{TP + FP} \quad (34)$$

$$Recall = \frac{TP}{TP + FN} \quad (35)$$

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \quad (36)$$

4.4. Validation on the Case Western Reserve University (CWRU) dataset

4.4.1 Data description

The motor bearing vibration data set of CWRU, as one of the widely used data sets, can be divided into ball fault, inner race fault and outer race fault according to different fault locations. Each fault location can be further divided into three categories: 7mils, 14mils and 21mils according to its severity. Thus, including healthy data, the data set can be divided into ten categories. The division of data sets is given in Table A1 in the Appendix.

As the sampling frequency of the data is 12KHz, and the motor speed changes between 1797 RPM and 1730 RPM, it can be calculated that the number n (37) of data points collected in each complete rotation of the rotating shaft is between 400 and 416. Therefore, to capture the impact of bearing failure at least once in each sample, the length of each sample is set to 1200 data points. For each category, it was divided into 400 samples based on the total number of data points. 80% of the samples were randomly selected as the training set and the rest as the test set. The original vibration signals are given in Figure A1 in the Appendix.

$$n = 60 f_s / \omega \quad (37)$$

where f_s is the sampling frequency and ω is the speed.

4.4.2. Model analysis

To verify that our strategy is effective in dealing with changing operating conditions, a series of comparative experiments were conducted. First, to verify the effectiveness of the AOA algorithm and the integration algorithm, the ALEDA model is compared with the manual parameter tuning model LEDA and the individual encoder models DAE, SAE, and CAE. Second, to verify the effectiveness of the strategy of dynamically selecting encoders based on weights, ALEDA was compared with the ensemble algorithm LEDA_Fix, where three encoders are fixed. Note that the classifiers used for comparison are all hypersphere classifiers.

the ensemble model has better performance in noisy environment. In summary, the parameter adjustment algorithm AOA and the integrated algorithm are beneficial to improve the efficiency of parameter adjustment and improve the classification accuracy of the fault diagnosis model.

To understand the training process of our ensemble model when determining the composition of encoders, the weight changes process of the three encoders during the first experiment in a noise-free environment were recorded and analyzed (Figure 6). When the first encoder was selected (Figure 6 (a)), the initial weight of the three models were set to 33.33% and then the weights of the three encoders were updated according to Equation 26 after the first folding of cross-verification. The weight of SAE was increased to 33.65% due to its highest validation accuracy while the weight of CAE was reduced to 33.08% due to its lower validation accuracy than the other two models. In the second fold, the validation accuracy of CAE was still the lowest among the three, so its weight was again reduced to 32.96%, a decrease of 0.12%. And DAE and SAE are validated with the same accuracy, so the weight of each is increased by 0.06%.

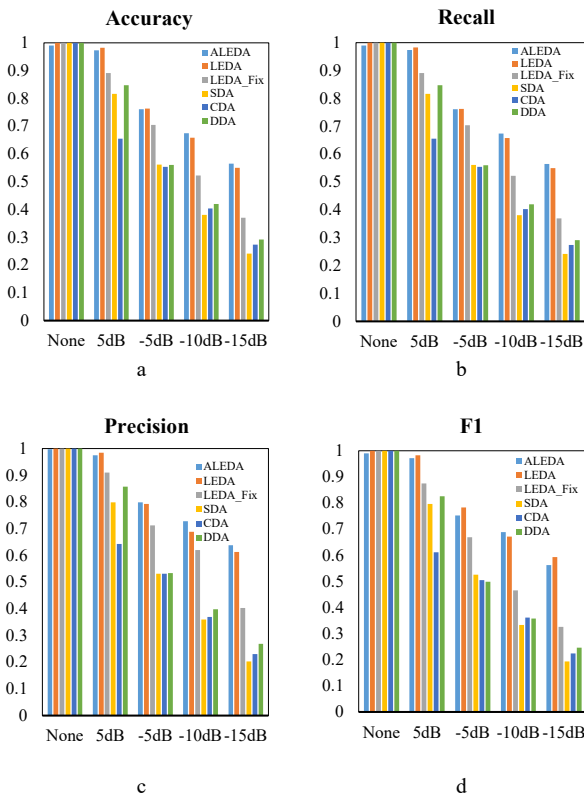


Figure 5: Comparison chart of evaluation indicators

By observing the comparison of various evaluation indexes in Figure 5, it can be found that under various noise environments, the classification accuracy of ALEDA is almost equal to LEDA, and even slightly better than LEDA in some cases. This shows that the network parameters automatically selected by the AOA algorithm and the parameters selected by manual repeated experiments have the same effect in the fault diagnosis model. Secondly, by comparing LEDA and LEDA_Fix, it can be found that LEDA shows higher classification accuracy in all kinds of noisy environments. And the advantage of LEDA gradually expands with the increase of noise. In addition, by comparing with individual encoder model DAE, SAE and CAE, it can be found that

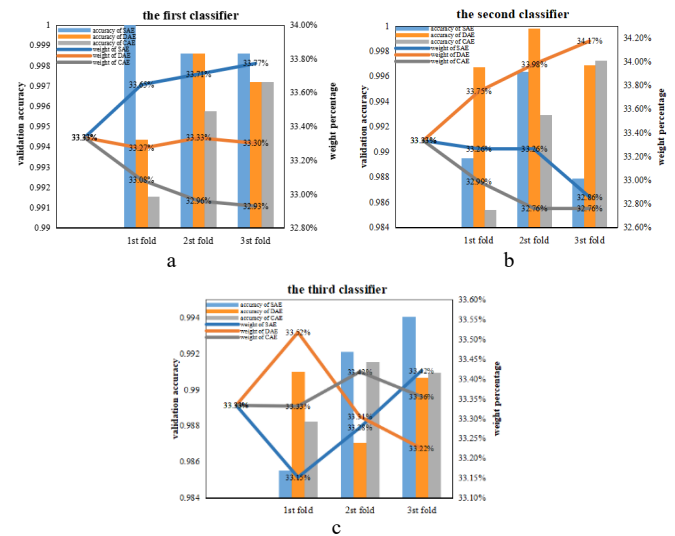


Figure 6: The validation accuracy and weight changes of each classifier during cross-validation under noise-free environment

In the third fold, the validation accuracy of SAE was still the highest, so its weight was further increased by 0.06%. At the same time, DAE and CAE were reduced by 0.03% respectively because of the same accuracy. In terms of overall performance, after the completion of cross-validation, SAE has the highest weight of 33.77%, which is better than the other two models, so it is determined as the final model. Figure 6(b) and Figure 6(c) show the changing process of the weight and verification accuracy when selecting the second and third encoders, respectively. Since the selection idea is the same as the above process, it will not be repeated here. At the end of training, an ensemble model consisting of SAE, DAE, and SAE can be obtained.

In addition, to intuitively compare the learning capability of feature representation between the ensemble model and the individual encoder model, the principal components of the ensemble model and each encoder model in a 5dB SNR environment were extracted using TSNE, and visualized as 2Ds and 3D plots. Meanwhile, the confusion matrix of each model is visualized. All these plots are placed in Figure 7, where each column is the same model.

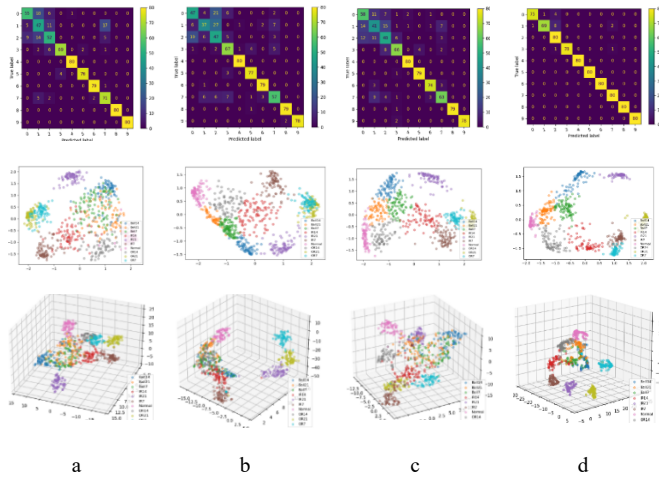


Figure 7: 2D and 3D feature visualization and confusion matrix visualization in 5dB SNR environment of (a) SAE, (b) DAE, (c) CAE and (d) ALEDA

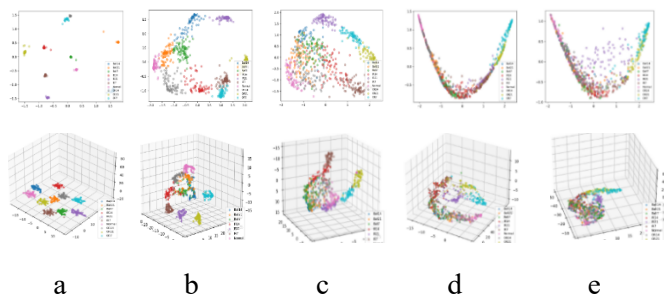


Figure 8: 2D and 3D feature visualization (a) noiseless of (b) 5dB, (c) 0dB, (d)-5dB and (e) -10dB.

For the SAE model, only the outer race fault samples OR7 and OR21 with fault depths of 7mils and 21mils and the inner race fault sample IR21 with a fault depth of 21mils were significantly separated from the other samples. All other types of samples have different degrees of overlap, especially the ball fault samples with failure depths of 7mils, 14mils and 21mils, which have a larger amount of overlap. Therefore, only OR7, OR21 and IR21 were completely correctly classified, while a large number of samples in other categories were incorrectly classified. For DAE model, IR21 samples were completely separated from other samples, and there was a small amount of overlap between IR7 and IR14, IR7 and OR21, OR7 and OR21, and normal and B21. There was a large amount of overlap between the samples of other categories, so only IR21 was completely correctly classified, and some samples of other categories were incorrectly classified. Similarly, for CAE models, only IR21 samples were

correctly classified. For ensemble model ALEDA, there was still a little overlap between the ball fault samples of the three fault depths, so a small number of samples were misclassified. However, compared with the three separate models, there was a significant improvement. Besides, there was a small amount of overlap between IR14 and B7, and a small amount of overlap between IR14 and OR14. Therefore, one IR14 sample was wrongly classified into B7, and one IR14 sample was misclassified into OR14. In general, ALEDA is superior to SAE, DAE and CAE in learning feature representation under 5dB SNR environment.

In addition, to observe the key features extracted by ALEDA under different noise levels, the principal components of the test samples in 5dB, 0dB, -5dB, -10dB and noise-free environment were extracted using TSNE, and visualized as 2Ds and 3D maps (Figure 8). It was found that in the noise-free environment, the test samples with different health status were separated obviously, and the different types of fault samples began to overlap gradually with the increase of noise.

4.4.3 Analysis of diagnostic results

After obtaining the classification results of the ensemble model, the correlation scores of the input features can be obtained through the propagation mechanism of LRP. Correlation scores of the health test samples and the test samples with a fault depth of 21mils were visualized as heat maps where the higher the correlation scores, the higher the color saturation.

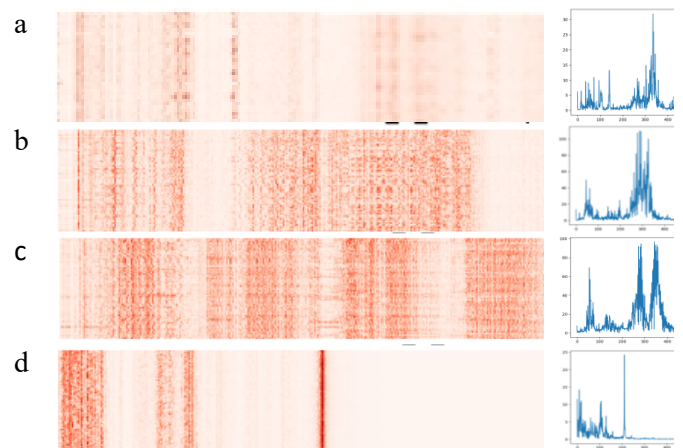


Figure 9: Visualization of LRP heat map and power spectrum visualization for (a) ball fault, (b) inner race fault, (c) outer race fault and (d) normal

According to the heat map, the contribution of input features of each test sample to the diagnosis results can be obtained. And we can see that each type of fault has its unique salient characteristics. To find the relationship between these high-scoring input features and their corresponding fault categories, a few samples were drawn from each type of the testing data and their power spectrum were visualized. The heat and spectrum plots

are placed in Figure 9, where each row is the same fault category.

parameters were then directly called in other four experiments to improve the running efficiency of the algorithm.

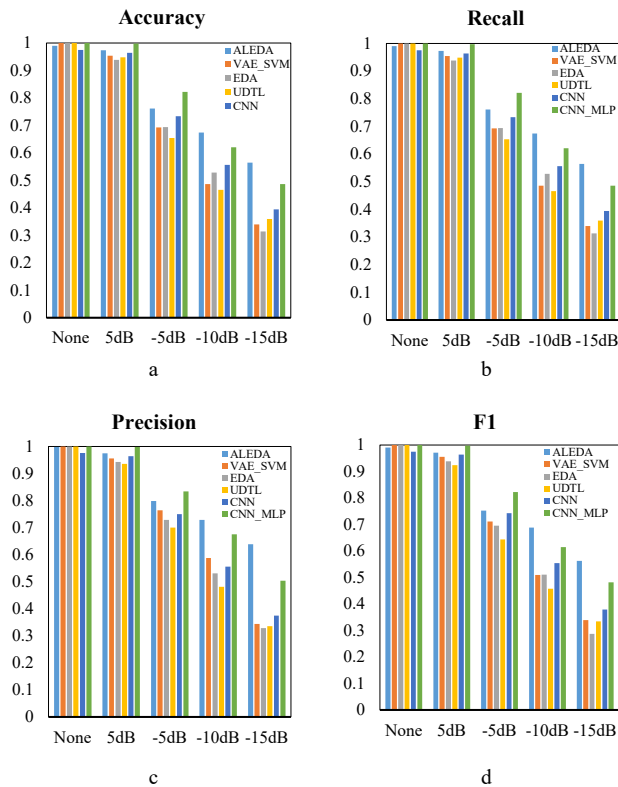


Figure 10: Comparison chart of evaluation indicators

In the traditional power spectrum analysis, when the peak value in the vibration spectrum is not equal to the multiple of the speed frequency, it means that the fault may occur. Further, if there are harmonics and side bands at the same time, the fault's more likely to occur. In addition, the fault frequency varies due to different fault locations. That is, each failure happens on a particular frequency component. By comparing heat map and power spectrum, it can be seen that features with higher scores in the LRP heat map usually correspond to frequencies with larger amplitudes in the power spectrum map. Besides, the regions of high score aggregation in the heat map roughly correspond to the regions with dense side bands and harmonics in the power spectrum map. Therefore, we infer that features with high scores in the heat map are related to abnormal frequencies such as side bands and harmonics in the power spectrum map. Our model learns and classifies all kinds of fault data through the data points that are different from normal frequency.

4.4.4 Comparison with other intelligent diagnosis methods

Compared with other intelligent fault diagnosis methods VAE_SVM (VAE for dimensional-reduction of data) [31], CNN_MLP [32], CNN, EDA(weighted integration) [33], UDTL (based on transfer learning)[34] . Five experiments were carried out for each algorithm in each noisy environment to get the average value where the parameters of the algorithm in this paper were tuned by the AOA algorithm in the first run, and the saved

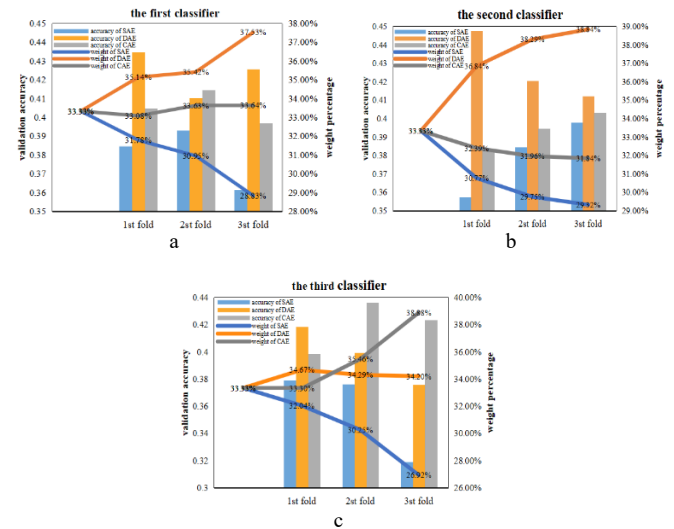


Figure 11: The validation accuracy and weight changes of each classifier during cross-validation under -10dB SNR environment

As can be seen from Figure 10, each model can accurately classify various faults in a noise-free environment, and both false negatives and false positives of each model perform well. With the increase of the noise in the data, the indicators of each model began to decline, but ALEDA performed better than VAE_SVM, EDA, UDTL and CNN. However, in 5dB and -5dB environments with little noise influence, all indexes of ALEDA are lower than CNN_MLP, which might be explained that CNN can extract deeper features for further classification by MLP with its powerful feature extraction ability. With the increase of noise, ALEDA gradually shows better classification ability than CNN_MLP.

To explore the reason why ALEDA gradually outperforms CNN_MLP under high noise, the weight changes process of the three encoders during the first experiment in -10dB were recorded and analyzed (Figure 11). Compared with Figure 6, it can be found that SAE accounted for a large proportion in the ensemble model in the noiseless environment, while under -10dB noise environment, the DAE model exhibits better validation accuracy and stability during the selection of the first encoder (Figure 11(a)) and the second encoder (Figure 11(b)). When choosing the third encoder model (Figure 11(c)), the CAE model outperformed. It is obvious that the combined form of the DAE and CAE models further enhances the noise immunity of the ensemble model ALEDA.

4.5. Dataset of the American Association for Mechanical Failure Prevention Technology (MFPT)

4.5.1. Data description

MFPT bearing fault dataset can be divided into inner race fault and outer race fault according to different

bearing fault locations, and each fault location can be further divided into 7 categories. Therefore, including healthy data, it is divided into 15 categories. Since the sampling rate of this data is 48,828sps and the input axis rate is 25Hz, about 1,953 data points can be collected every time when the axis is fully rotated. Therefore, to capture the impact of a bearing fault on each sample, the length of each sample was set to 2000 data points. The division of data sets and their load are given in Table A2 in the Appendix, and the original vibration signals are given in Figure A2 in the Appendix. Finally, 80% of the samples were randomly selected as the training set and the rest as the test set. It is noteworthy that the number of healthy samples in the MFPT dataset is larger than the other faulty samples.

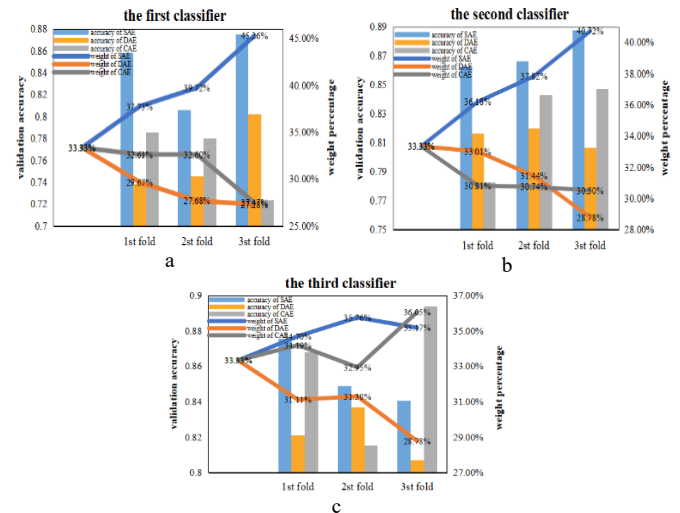


Figure 13: The validation accuracy and weight changes of each classifier during cross-validation under -10dB SNR environment.

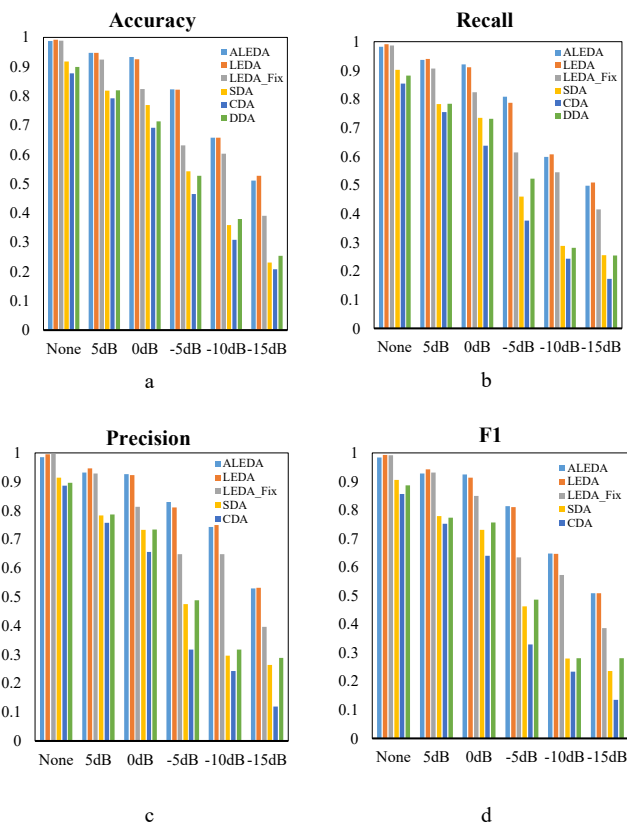


Figure 12: Comparison of accuracy (a), Recall (b), Precision (c) and F1-score (d) between the six models

4.5.2. Model analysis

Like Section 4.4.2, ALEDA, LEDA, LEDA_Fix and DAE, SAE and CAE models of individual encoder were compared in various noise environments. As can be seen from Figure 12, the performance of ALEDA is like LEDA in various noise environments. And compared with LEDA_Fix, ALEDA and LEDA showed higher classification accuracy in various noise environments. ALEDA also achieved obvious advantages over individual encoder models SAE, DAE and CAE. Therefore, it can be concluded that by using the parameter tuning algorithms AOA, the ensemble algorithm and the hypersphere classifier the diagnostic accuracy is improved regardless of the operating environments.

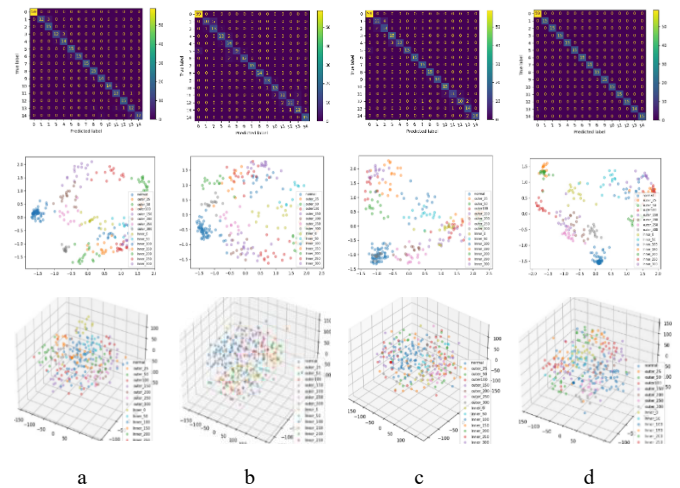


Figure 14: 2D and 3D feature visualization and confusion matrix under noiseless environment for (a) SAE, (b) DAE, (c) CAE and (d) ALEDA

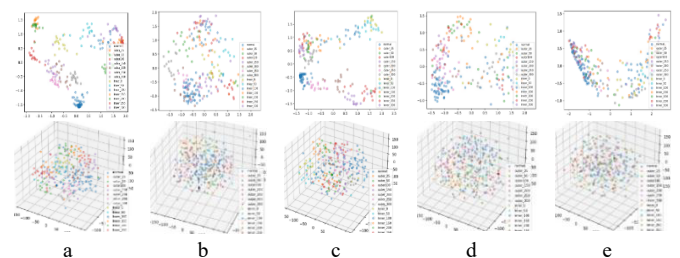


Figure 15: 2D and 3D feature visualization for scenarios of (a) noiseless, with (b) 5dB, with (c) 0dB, with (d) -5dB and with (e) -10dB noise.

T_SNE was used to extract the main components of the ensemble model and each encoder model in noiseless environment, and the two-dimensional visualization and three-dimensional visualization were carried out respectively. Meanwhile, the confusion matrix of each model was visualized. All these plots are placed in Figure 14, where each column describes the same model. By observing Figure 14, it can be found that in the three basic models, most of the samples of various health conditions cannot be completely separated, and there was sample overlap. While by ALEDA, most types of samples had

clear boundaries that can be distinguished from other types of samples. In general, ALEDA is superior to SAE, DAE and CAE in learning feature representation. Figure 15 showed the main components of ALEDA at different noise levels.

4.5.3 Analysis of diagnostic results

Correlation scores and the power spectrum (Figure 16) of the fault test sample at 300 loads and the health sample were visualized. Due to the limited amount of data per category in the MFPT datasets, the heat map has a narrow width. Although the amount of data is small, each type of data shows its unique characteristics. By comparing the frequency spectrum diagram and the heat map, the location of the fault frequency component that affects each kind of data can be obtained, and fault diagnosis can be achieved by learning these key features.

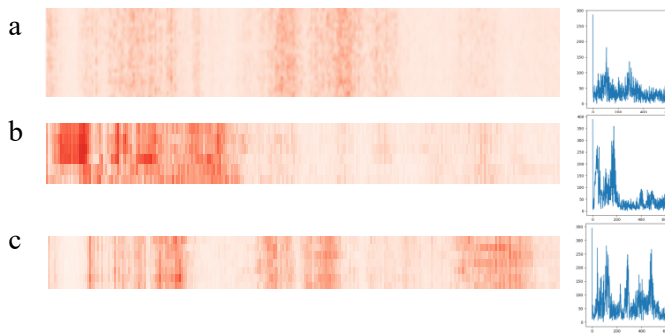


Figure 16: Visualization of LRP heat map and power spectrum visualization for (a) normal, (b) inner race fault and (c) outer race fault

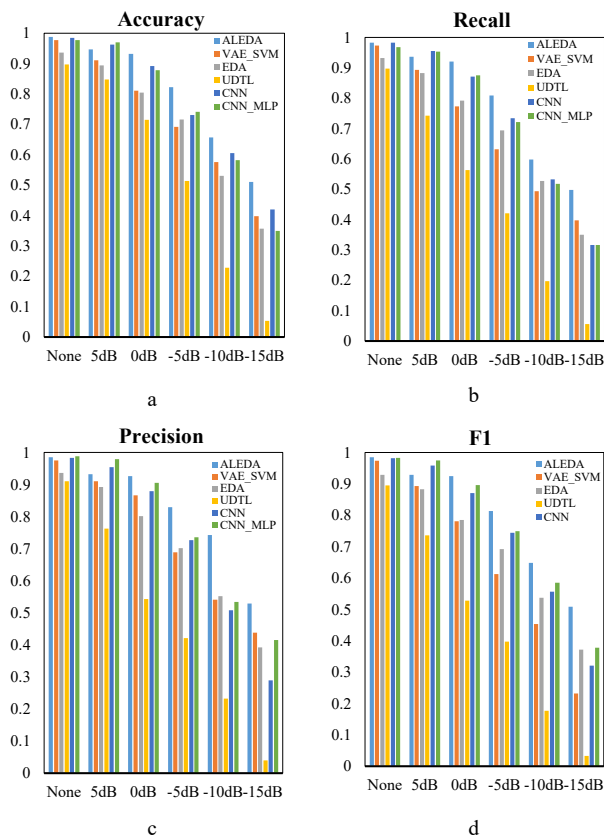


Figure 17: Comparison of accuracy (a), Recall (b), Precision (c) and F1-score (d) between the six models

4.5.4. Comparison with other intelligent diagnosis methods

The model ALEDA and several other fault diagnosis methods VAE_SVM, CNN_MLP, CNN, EDA and UDTL were respectively run in each noise environment and compared. As can be seen from Figure 17, the performance of ALEDA is similar to CNN_MLP and CNN in a noise-free environment. However, in a low-noise environment, say 5dB, both models based on convolutional neural network were slightly better than ALEDA. With the increase of noise, ALEDA gradually showed better classification ability. We've discussed the reasons in Section 4.4.4. In addition, ALEDA was always superior to VAE_SVM, EDA and UDTL in various noise environments.

4.6. Dataset of Jiang Nan University (JNU)

4.6.1. Data description

The JNU data set contains inner race fault, outer race fault, ball fault and normal data at speeds of 600, 800, and 1000 [35]. As the fault frequency of the same fault location changes with the change of rotation speed, the data set is divided into 12 types according to the fault location and rotation speed, and the division of the dataset is given in Table A3 in the Appendix. The original vibration signals are given in Figure A3 in the Appendix.

4.6.2. Model analysis

We compared ALEDA, LEDA, LEDA_Fix and DEA, SAE, CAE models in different noise environments (Figure 18). Overall, ALEDA showed better performance at all noise levels, proving that it has some advantages under different operating conditions.

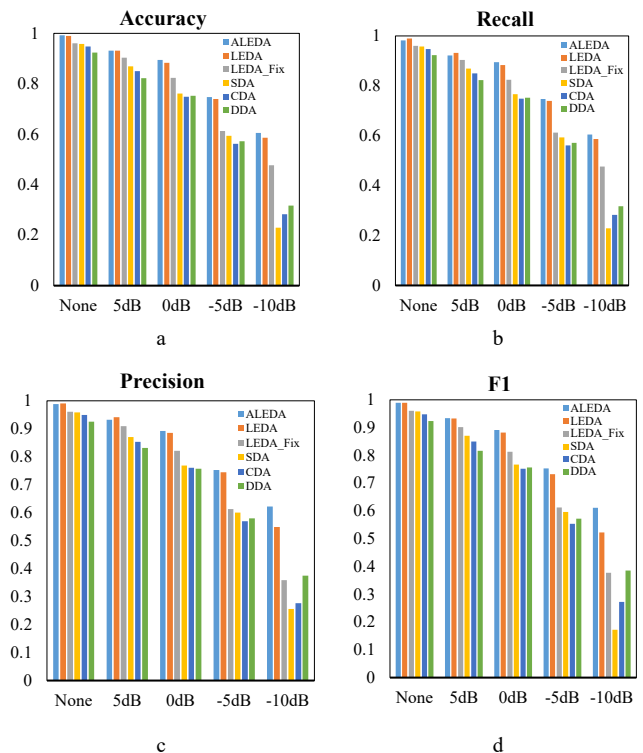


Figure 18: Comparison of accuracy (a), Recall (b), Precision (c) and F1-score (d) between the six models

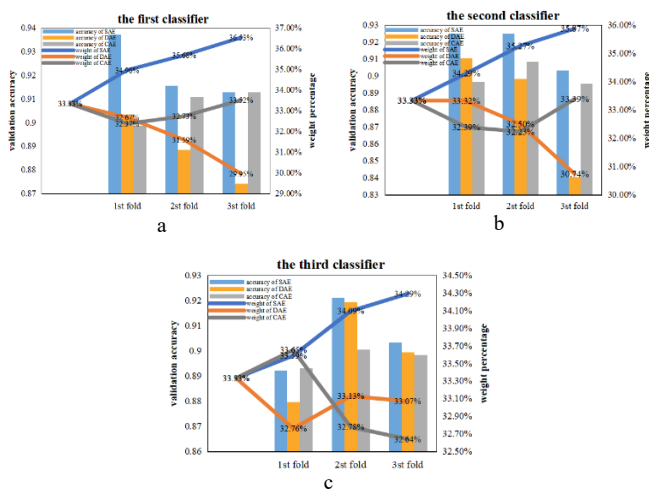


Figure 19: The validation accuracy and weight changes of each classifier during cross-validation under noiseless environment

Figure 19 shows the weight changes process of the three encoders during training in the first experiment without noise. When the first encoder was selected (Figure 19 (a)), the initial weight of the three models were set to 33.33%. And in the triple fold of cross-validation, the validation accuracy of SAE has always been the highest, so its weight has been increasing. And at the end of the cross-validation, the weight of SAE (36.53%) was higher than that of the other two models, so it was identified as the final model. In the process of selecting the remaining two encoders, as shown in Figure 19(b) and Figure 19(c), SAE is still the one with the largest weight, so the composition of our final integration model is SAE, SAE and SAE.

Figure 20 show the main components and confusion matrix of ALEDA, SAE, DAE and CAE at 5dB SNR. For the basic models SAE, DAE and CAE, at 5dB SNR, the normal samples at 600 RPM were almost completely separated from other samples, and there was a large amount of overlap between the samples of other categories. As for the ensemble model ALEDA, although there is still a small amount of overlap between various samples, it has been greatly improved compared with the basic model.

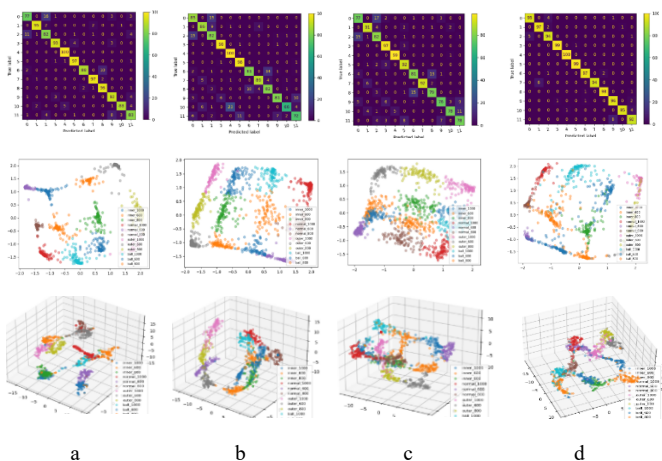


Figure 20: 2D and 3D feature visualization and confusion matrix under 5dB SNR environment for (a) SAE, (b) DAE, (c) CAE and (d) ALEDA

Figure 21 shows the main components of ALEDA at different noise levels. In the noise-free environment, the test samples with different health conditions were separated obviously. With the increase of noise, different types of fault samples began to overlap gradually.

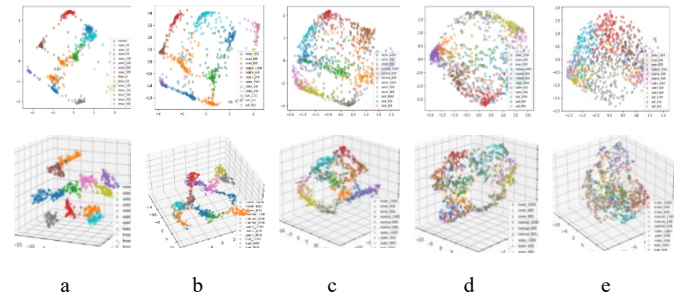


Figure 21: 2D and 3D feature visualization of scenarios of (a) noiseless, with (b) 5dB, (c) 0dB, (d) -5dB and (e) -10dB noise

4.6.3. Analysis of diagnostic results

Correlation scores of test samples at 1000 RPM were visualized as heat maps. Since the color difference between different features in the heat maps is not very large, the 100 features with the highest scores were visualized. In addition, several samples were randomly selected from each type of fault and their power spectrum were visualized. All the images are placed in Figure 22. By observing the power spectrum and heat maps, it can be found that each type of fault has unique prominent features, and the features with high scores in the heat map can roughly correspond to the harmonic components in the power spectrum map. Therefore, it can be concluded that the features of high color saturation in the heat map are equivalent to the frequency of abnormal faults in the power spectrum map, and our model classifies the fault data by learning these features.

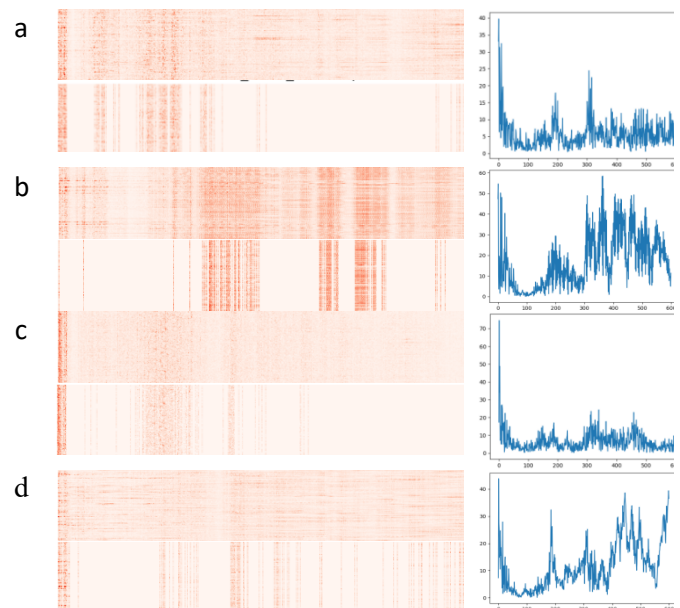


Figure 22: Visualization of LRP heat map and the top 100 input features with the highest scores and their power spectrums (a) inner race fault (b) outer race fault (c) normal (d) ball fault

4.6.4. Comparison with intelligent diagnosis methods

In this experiment, ALEDA and several other fault diagnosis methods, VAE_SVM, CNN_MLP, CNN, EDA and UDTL, were respectively run in SNR=5dB, 0dB, -5dB, -10dB and noiseless environment. And all running results were recorded in Figure 23. Through comparison, it is found that the performance of ALEDA was not much different from other models in the environment with little noise influence. With the reduction of SNR, ALEDA was gradually superior to other models. See section 4.4 for the reason.

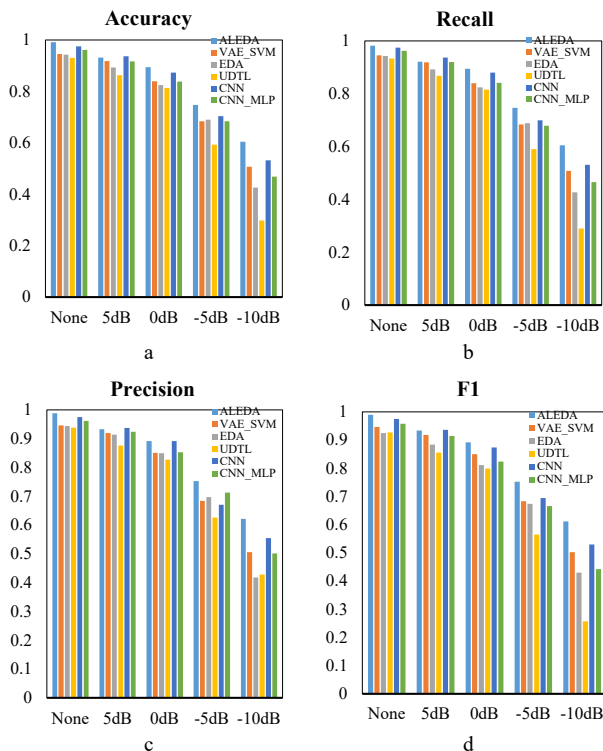


Figure 23: Comparison of accuracy (a), Recall (b), Precision (c) and F1-score (d) between the six models

5. Conclusions

In this paper, a new ensemble interpretable deep auto-encoder (ALEDA) method is proposed for intelligent fault diagnosis of rolling bearings. This method addresses the problem of diagnostic accuracy degradation of fault diagnosis models under variable operating conditions from three perspectives: adaptive adjustment of model parameters, targeted training of different types of fault data, and adaptive construction of integrated models. Power spectrum analysis and Layer-wise Relevance Propagation algorithm are combined to interpret the diagnostic results made by the model based on frequency domain data, which not only improves the reliability of the diagnostic results, but also provides enlightenment for interpreting the results. In addition, it is worth mentioning that this method verifies the effectiveness of the arithmetic optimization algorithm in the fault diagnosis model to a certain extent.

ALEDA was validated on three public datasets, including the CWRU, MFPT, and JNU. By comparing the

ALEDA model with the manual parameter tuning model, the effectiveness of the arithmetic optimization in the fault diagnosis model based on deep auto-encoder was verified. In addition, by visualizing and comparing the main features extracted by ALEDA and the individual encoder models, it was found that ALEDA can learn more critical features. Moreover, the impact of noise on the diagnostic accuracy of the ALEDA model was explored by visualizing the features extracted by ALEDA under different noises. What's more, ALEDA was compared with other advanced intelligent fault diagnosis methods at different noise levels. Various experimental results show that ALEDA is flexible enough to respond to changes in operating conditions and has excellent capabilities of diagnosis and generalization.

At present, ALEDA is only the ensemble of three basic encoder models. In future work, we will replace the basic model with the encoder variant model to explore whether the model's robustness and diagnostic accuracy can be further improved. Additionally, the data processing of the hypersphere classifier in the overlapping regions will be a meaningful study.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

Funding was received from the National Natural Science Foundation of China (61871061), which is gratefully acknowledged.

References

- [1] X. Wang, Y. Zi, Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis", *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 285-304, 2011, doi:10.1016/j.ymssp.2010.03.010.
- [2] Z. Wang, L. Jia, Y. Qin, "Adaptive diagnosis for rotating machineries using information geometrical kernel-ELM based on VMD-SVD", *Entropy*, vol. 20, no. 1, p. 73, 2018, doi:10.3390/e20010073.
- [3] J. Xie, Z. Li, Z. Zhou, S. Liu, "A novel bearing fault classification method based on XGBoost: The fusion of deep learning-based features and empirical features", *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-9, 2020, doi:10.1109/TIM.2020.3042315.
- [4] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification", *Measurement*, vol. 89, pp. 171-178, 2016, doi:10.1016/j.measurement.2016.04.007.
- [5] C. Shen, Y. Qi, J. Wang, G. Cai, Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder", *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170-184, 2018, doi:10.1016/j.engappai.2018.09.010.
- [6] F. Xu, X. Shu, X. Zhang, B. Fan, "Automatic diagnosis of microgrid networks' power device faults based on stacked denoising autoencoders and adaptive affinity propagation clustering", *Complexity*, vol. 2020, 2020, doi:10.1155/2020/8509142.
- [7] Y. Zhang, X. Li, L. Gao, W. Chen, P. Li, "Ensemble deep contractive auto-encoders for intelligent fault diagnosis of machines under noisy environment", *Knowledge-Based Systems*, vol. 196, p. 105764,

- 2020, doi:10.1016/j.knosys.2020.105764.
- [8] S. Haidong, J. Hongkai, Z. Ke, W. Dongdong, L. Xingqiu, "A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings", *Mechanical Systems and Signal Processing*, vol. 110, pp. 193-209, 2018, doi:10.1016/j.ymssp.2018.03.011.
- [9] H. Shao, H. Jiang, Y. Lin, X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders", *Mechanical Systems and Signal Processing*, vol. 102, pp. 278-297, 2018, doi:10.1016/j.ymssp.2017.09.026.
- [10] H. Shao, H. Jiang, F. Wang, H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis", *Knowledge-Based Systems*, vol. 119, pp. 200-220, 2017, doi:10.1016/j.knosys.2016.12.012.
- [11] Y. Zhang, X. Li, L. Gao, W. Chen, P. Li, "Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method", *Measurement*, vol. 151, pp. 107232, 2020, doi:10.1016/j.measurement.2019.107232.
- [12] W. Deng, R. Yao, H. Zhao, X. Yang, G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm", *Soft Computing*, vol. 23, no. 7, pp. 2445-2462, 2019, doi:10.1007/s00500-017-2940-9.
- [13] H. Chen, D. L. Fan, L. Fang, W. Huang, J. Huang, C. Cao, L. Yang, Y. He, L. Zeng, "Particle swarm optimization algorithm with mutation operator for particle filter noise reduction in mechanical fault diagnosis", *International journal of pattern recognition and artificial intelligence*, vol. 34, no. 10, pp. 2058012, 2020, doi:10.1142/S0218001420580124.
- [14] D. Lee, J. Ahn, B. Koh, "Fault detection of bearing systems through EEMD and optimization algorithm", *Sensors*, vol. 17, no. 11, p.2477,2017, doi:10.3390/s17112477.
- [15] C. Lee, T. Le, "An Enhanced Binary Particle Swarm Optimization for Optimal Feature Selection in Bearing Fault Diagnosis of Electrical Machines", *IEEE Access*, vol. 9, pp. 102671-102686, 2021, doi:10.1109/ACCESS.2021.3098024.
- [16] W. Zhang, G. Han, J. Wang, Y. Liu, "A BP neural network prediction model based on dynamic cuckoo search optimization algorithm for industrial equipment fault prediction", *IEEE Access*, vol. 7, pp. 11736-11746, 2019, doi: 10.1109/ACCESS.2019.2892729.
- [17] H. Qu, Z. Qiu, X. Tang, M. Xiang, P. Wang, "Incorporating unsupervised learning into intrusion detection for wireless sensor networks with structural co-evolvability", *Applied Soft Computing*, vol. 71, pp. 939-951, 2018, doi:10.1016/j.asoc.2018.07.044.
- [18] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144, 2016, doi:10.1145/2939672.2939778.
- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, A. Kaiser, I. Polosukhin, "Attention is all you need", *Advances in neural information processing systems*, vol. 30,2017.
- [20] X. Li, W. Zhang, Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism", *Signal processing*, vol. 161, pp. 136-154, 2019, doi:10.1016/j.sigpro.2019.03.019.
- [21] Y. Yang, V. Tresp, M. Wunderle, P. A. Fasching, "Explaining therapy predictions with layer-wise relevance propagation in neural networks", *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 152-162, 2018, doi:10.1109/ICHI.2018.00025.
- [22] B. Zhao, C. Cheng, G. Tu, Z. Peng, Q. He, G. Meng, "An interpretable denoising layer for neural networks based on reproducing kernel Hilbert space and its application in machine fault diagnosis", *Chinese Journal of Mechanical Engineering*, vol. 34, no. 1, pp. 1-11, 2021, doi:10.1186/s10033-021-00564-5.
- [23] J. Grezma, J. Zhang, P. Wang, K. A. Loparo, R. X. Gao, "Interpretable convolutional neural network through layer-wise relevance propagation for machine fault diagnosis", *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3172-3181, 2019, doi:10.1109/JSEN.2019.2958787.
- [24] A. Binder, S. Bach, G. Montavon, K. Müller, W. Samek, "Layer-wise relevance propagation for deep neural network architectures", pp. 913-922, 2016, doi:10.1007/978-981-10-0557-2_87.
- [25] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. Müller, W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation", *PLOS One*, vol. 10, no. 7, pp. e130140, 2015, doi:10.1371/journal.pone.0130140.
- [26] A. Binder, G. Montavon, S. Lapuschkin, K. Müller, W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers," *International Conference on Artificial Neural Networks*, pp. 63-71, 2016, doi:10.1007/978-3-319-44781-0_8.
- [27] A. Rios, V. Gala, S. McKeever, "Explaining Deep Learning Models for Structured Data using Layer-Wise Relevance Propagation", *arXiv preprint arXiv: 2011.13429*, 2020, doi:10.48550/arXiv.2011.13429.
- [28] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, A. H. Gandomi, "The arithmetic optimization algorithm", *Computer methods in applied mechanics and engineering*, vol. 376, pp. 113609, 2021, doi:10.1016/j.cma.2020.113609.
- [29] M. Du, Q. Yu, L. Ruisen, "Hypersphere Algorithm for Classification on Dynamic Feature Space", *CEA*, vol. 56, no. 22, pp. 6, 2020, doi:10.3778/j.issn.1002-8331.1908-0352.
- [30] J. Zheng, H. Qu, Z. Li, L. Li, X. Tang, "An irrelevant attributes resistant approach to anomaly detection in high-dimensional space using a deep hypersphere structure", *Applied Soft Computing*, vol. 116, pp. 108301, 2022, doi:10.1016/j.asoc.2021.108301.
- [31] J. An, S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability", *Special Lecture on IE*, vol. 2, no. 1, pp. 1-18, 2015.
- [32] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, "Unsupervised deep transfer learning for intelligent fault diagnosis: An open source and comparative study", *arXiv preprint arXiv:1912.12528*, 2019, doi:10.48550/arXiv.1912.12528.
- [33] Y. Zhang, X. Li, L. Gao, W. Chen, P. Li, "Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method", *Measurement*, vol. 151, pp.107232, 2020, doi:10.1016/j.measurement.2019.107232.
- [34] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: a survey and comparative study", *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 3525828, pp. 1-28, 2021, doi:10.1109/TIM.2021.3116309.
- [35] K. Li, X. Ping, H. Wang, P. Chen, Y. Cao, "Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis", *Sensors*, vol. 13, no. 6, pp. 8013-8041, 2013, doi:10.3390/s130608013.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Kaibi Zhang has done her bachelor's degree from Shandong University of Science and Technology in 1996. She has done her master's degree from Chongqing University of Posts and Telecommunications in 2009.

She has published 19 papers, of which 4 are indexed by EI, 5 are in core journals, and 2 textbooks are edited.



Yanyan Wang has done her bachelor's degree from ABC institution in 2015. She is currently studying for a master's degree at Chongqing University of Posts and Telecommunications. Her current research interests include deep learning and anomaly detection



Hongchun Qu has done his bachelor's degree from Anhui University of Science and Technology in 2001. He has done his master's degree from Chongqing University in 2005. He has completed his PhD degree in Computer Software and Theory at Chongqing University in 2009.

He has published more than 40 peer reviewed journal papers in high-level international journals such as OIKOS, COMPAG, IEEE TFS, ISA TRANS.

Appendix

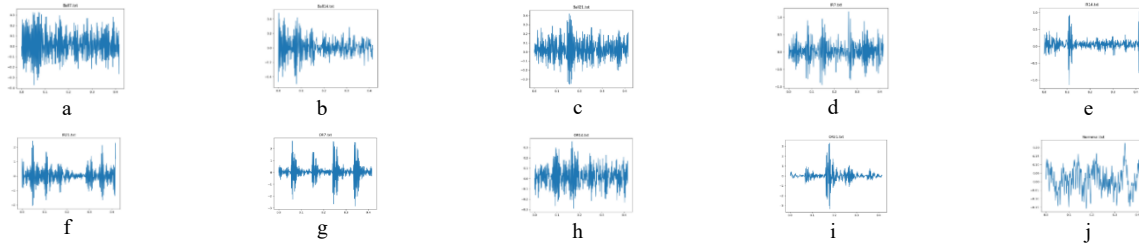


Figure A1: Original vibration signal of CWRU dataset of (a)B7, (b)B14, (c)B21, (d)IR7, (e)IR14, (f)IR21, (g)OR7, (h)OR14, (i)OR21 and (j)Normal

TableA1: Description of health condition and samples of CWRU dataset

Health condition	Motor Load (HP)	Fault Diameter	Sample number	Sample length
Inner Race fault	0,1,2,3	7mils	400	1200
Inner Race fault	0,1,2,3	14mils	400	1200
Inner Race fault	0,1,2,3	21mils	400	1200
Ball fault	0,1,2,3	7mils	400	1200
Ball fault	0,1,2,3	14mils	400	1200
Ball fault	0,1,2,3	21mils	400	1200
Outer Race fault	0,1,2,3	7mils	400	1200
Outer Race fault	0,1,2,3	14mils	400	1200
Outer Race fault	0,1,2,3	21mils	400	1200
Normal	0,1,2,3	-	400	1200

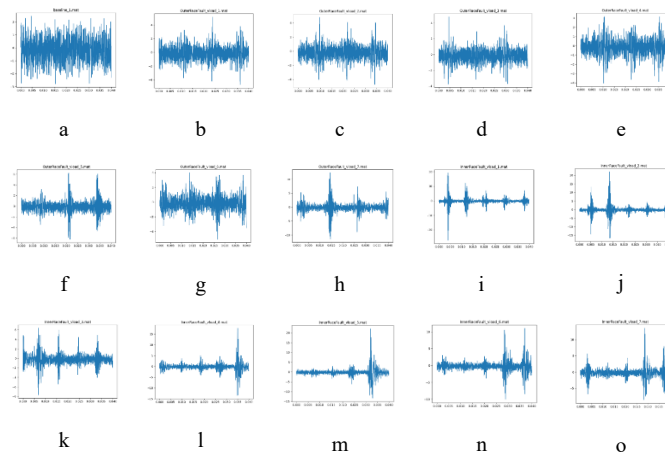


Figure A2: Original vibration signal of the 15 categories of MFPT dataset: (a) Normal (b) OR_25 (c) OR_50 (d) OR_100 (e) OR_150 (f) OR_200 (g) OR_250 (h) OR_300 (i) IR_0 (j) IR_50 (k) IR_100 (l) IR_150 (m) IR_200 (n) IR_250 (o) IR_300

Table A2: Description of health condition and samples of MFPT dataset.

Health condition	Load (LBS)	Sample number	Sample length
Inner Race fault	0	73	2000
Inner Race fault	50	73	2000
Inner Race fault	100	73	2000
Inner Race fault	150	73	2000
Inner Race fault	200	73	2000
Inner Race fault	250	73	2000
Inner Race fault	300	73	2000
Outer Race fault	25	73	2000
Outer Race fault	50	73	2000
Outer Race fault	100	73	2000
Outer Race fault	150	73	2000
Outer Race fault	200	73	2000
Outer Race fault	250	73	2000
Outer Race fault	300	73	2000
Normal	270	293	2000

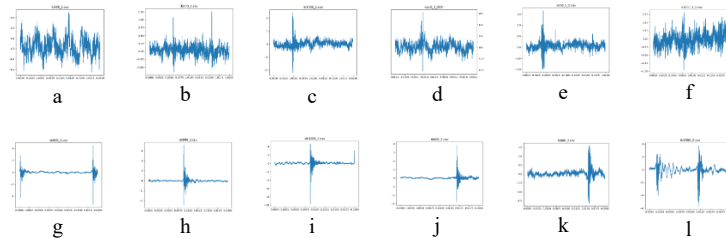


Figure A3: Original vibration signal of the 12 categories of JNU dataset, i.e., (a) IR_600, (b) IR_800, (c) IR_1000, (d) Normal_600, (e) Normal_800, (f) Normal_1000, (g) OR_600, (h) OR_800, (i) OR_1000, (j) B_600, (k) B_800 and (l) B_1000

Table A3: Description of health condition and samples of JNU dataset

Health condition	Motor speed (RPM)	Sample number	Sample length
Inner Race fault	600	500	1200
Inner Race fault	800	500	1200
Inner Race fault	1000	500	1200
Ball fault	600	500	1200
Ball fault	800	500	1200
Ball fault	1000	500	1200
Outer Race fault	600	500	1200
Outer Race fault	800	500	1200
Outer Race fault	1000	500	1200
Normal	600	500	1200
Normal	800	500	1200
Normal	1000	500	1200

Call Admission Control for Real-Time Applications of TVWS Wireless Access from HAP

Habib M. Hussien ^{1*}, Sultan F. Meko ², Konstantinos Katzis ³, Luzango P. Mfupe ⁴, Ephrem T. Bekele ¹

¹ Addis Ababa Institute of Technology (AAiT), School of Electrical and Computer Engineering, Addis Ababa, Ethiopia

² Addis Ababa Science and Technology University (AASTU), Department of Electrical and Computer Engineering, Addis Ababa, Ethiopia

³ European University of Cyprus, Department of Computer Science and Engineering, School of Sciences, Nicosia, Cyprus

⁴ CSIR, NGEI, Pretoria, South Africa

* Corresponding author: Habib M. Hussien, E-mail: habib.mohammed@aait.edu.et

Corresponding author ORCID: <https://orcid.org/0000-0002-9731-537X>

ABSTRACT: The rapid change in link capacity and user count induced by platform mobility in communication systems based on high altitude platform stations (HAPS) exploiting TV White Space (TVWS) spectrum may result in a high rate of handover failure and reduced resource utilization. In addition, in High Altitude Platform (HAP) wireless networks exploiting TV White Space (TVWS) spectrums, radio resources are frequently shared across numerous customers. When the number of users accepted into a network exceeds the network's capacity, network congestion occurs, resulting in a decrease in Quality of Service (QoS) or user displeasure. To address these issues, Call admission control (CAC) may be used. This article proposed a novel call admission control scheme using deadline, channel, and tolerance aware scheduling (DCTS) algorithm to solve the challenge of scheduling real-time flows in wireless networks while maintaining tight latency guarantees. The DCTS system ensures that the average packet drop due to deadline violation converges to the preset packet loss tolerance for a given deadline requirement, packet loss tolerance, and arrival rate. Our approach covers how to handle multiple real-time packet flows at the same time with a high risk of packet losses due to latency violations without surpassing a set threshold. It also discusses how real-time application scheduling in a wireless context must account for the complicated relationship between packet deadlines, channel circumstances, and flow tolerance, as well as how to propose such scheduling policies. We also test our proposed algorithm's performance for various arrival, channel state, deadline, and threshold scenarios. The convergence of packet drops near the threshold was demonstrated analytically. In wireless networks, CAC is a critical component in ensuring guaranteed quality of service. For real-time wireless applications that employ the DCTS scheduler, we present a threshold-based CAC method. We use the assumption that all flows belong to the same traffic class for determining the admission criteria. Our goal is to create a CAC algorithm that ensures that packet loss due to deadline violations is kept to a minimum for all allowed users. As a result, our CAC is based on a set of criteria that includes the maximum packet deadline, loss tolerance, and pace of newly received calls, as well as the accepted flows' minimum flow priority. The admission controller threshold is compared to the properties of freshly arrived flow in our CAC method. We compare our scheme's performance to that of the CAC of Violation Fair Exponential Rule (VFEXP) algorithm and the Modified Largest Weighted Delay First (MLWDF) methods.

KEYWORDS: CAC scheme; Flow Priority; DCTS; High Altitude Platforms, QoS; VFEXP; MLWDF, TVWS.

1. Introduction

Broadband wireless technology has progressed to keep up with the ever-increasing need for high-bandwidth data and applications. Due to the exponential growth in the number of users, there is a tremendous need for high-

quality wireless communication services. As a result of this demand, more novel communication infrastructures are being developed. To provide communication services to users, terrestrial and satellite networks have been built. The only creative option to deliver cellular

communications services using High Altitude Platform (HAP) is to overcome the drawbacks of both terrestrial and satellite systems [1]. HAP might be an aircraft or an airship, and it could be a lighter or heavier-than-air balloon. HAP functions in the stratospheric layer at a height of 17–22km [2-7] by keeping platform mechanisms, manned/unmanned aircraft, or airships flying in a close circle. Low propagation delay, big coverage area, line of sight, low cost, broadband capacity, broadband and broadcast services, and quick deployment are the most essential advantages of HAP [4]. Because of its efficient capacity, HAPs may give service to the shadow region and edge of the cell placed user, making them complementary to terrestrial systems [5-7]. Several concerns, including as resource allocation, scheduling, and packet flow management, in HAP wireless system using TVWS spectrum must be explored in this respect [8 - 13]. Wireless networks, like wired networks, have a range of QoS requirements. Delay, throughput, fairness, and packet loss threshold are all frequent QoS metrics. Because radio spectrum is a limited resource, customers are only allowed to utilize the service when the requisite resource is available. A freshly arrived user will be barred if this does not happen. As a result, call admission policy determines whether or not users are admitted. The goal of Call Admission Control (CAC) is to reduce the likelihood of users' calls being blocked while also maximizing the utilization of available resources. CAC aids in the regulation of network traffic in real-time applications, reducing congestion in wireless network traffic. CAC is also used in HAP wireless networks exploiting spectrum, to provide a particular degree of QoS. Consider a real-time service with a deadline, where the number of admitted users influences the proportion of packets dropped owing to a deadline violation. When an excessive number of users are permitted to a single HAP link wireless network, the percentage of packets that are dropped skyrockets. This, in turn, leads to client unhappiness, which lowers the service provider's income. The factors that may be controlled in most CAC algorithms include bandwidth, total number of simultaneous calls, total number of packets transferred per unit time, and total number of packets discarded owing to network traffic congestion. A new call may be prohibited from accessing the network until at least one of the allowed users ends if a predetermined limit of one of these parameters is exceeded. Alternatively, a parameter for CAC can be used to create a moderate decrease of QoS. Packet losses due to a deadline violation, for example, may be inevitable; yet, real-time applications can tolerate some loss tactfully. As a result, the threshold-based CAC algorithm may ensure that packet loss due to deadline violations is kept to a minimum for each real-time flow. CAC's capabilities go beyond just accepting or rejecting incoming calls depending on available resources. The allowed mobiles

can also travel from one cell to another; if the needed QoS is met, a handoff call can be accepted by an adjacent cell; otherwise, the user is discarded owing to handoff failure. As a result, the percentage of packets missed owing to handoff is an essential QoS statistic. A reservation-based CAC has been proposed in [14] to decrease handoff call dropping. Some specific channels are set aside for handoff calls under this arrangement. As a result, a handoff call is permitted if a reserved channel or free channel is available for new calls. As a result, when the number of open channels exceeds the number of reserved channels, a new call is accepted. CAC has been widely investigated in wired networks with the goal of congestion control and QoS provisioning. Because of time and location dependent channel circumstances, handoff requirements owing to mobility, and restricted bandwidth or slots, these algorithms cannot be employed directly for wireless counterpart. In addition, additional QoS requirements are available for wireless multimedia applications such as audio, video, and data. As a result, users must be supported by the CAC for wireless apps depending on QoS requirements. Literatures [15-18] presents a survey analysis of relevant wireless CAC. The scheduler is a crucial component of the CAC scheme's architecture. In this research, we suggested a scheduling method that takes three crucial factors into account when determining scheduling circumstances: channel conditions, flow tolerances, and packet deadlines in queues. The authors of [19] presented a modification to the Largest Weighted Delay First (MLWDF) scheduling discipline that accounts for wireless channels' time changing properties. The authors of [20] presented a comparison of different CAC scheme including Violation Fair Exponential Rule (VFEXP). As can be shown in the simulation, our suggested algorithm performs considerably better than the VFEXP and MLWDF schemes in terms of both percentage of satisfied consumers and percentage of call blocking. The rest of this paper is laid out as follows. We outline our system model and explore our threshold-based CAC policy in Section II. The Deadline, Channel, and Tolerance aware scheduling (DCTS) method is used to build an admission criterion. In Section III, we look into CAC and go over the suggested method in detail. In section IV, we use Matlab simulations to test the performance of our method and analyze the findings. We analyze the proportion of satisfied users and the percentage of call blocking using various system thresholds. We also compare our algorithm's performance to that of other methods like MLWDF and VFEXP. Finally, Section V brings the paper conclusion.

2. System Model and Description

In this scenario, a HAP with a TVWS base station antenna payload at a 20-kilometer height assigned to the TVWS base station on the ground is considered, with the

TVWS base station covering a coverage radius of 33km-100km. We examine a single-link HAP wireless system using TVWS spectrums operating in slotted time t where $t = \{0, 1, 2, \dots\}$. Consider a downlink situation in which users come and depart at different times. Each incoming call is assigned to a service if it meets the CAC controller's admission criteria. The channel holding time is exponentially distributed with a mean of $1/h$, while the call arrival follows a Poisson process with rate, λ_i . Call durations are independent and distributed exponentially with a mean of $1/\mu$. Once the call is accepted into the system, it will be handled in the DCTS scheduling scheme's priority sequence. During each time slot, the scheduler permits just one flow for packet transmission. If no slot is available at that moment, the call will be placed in a queue and will have to wait until a slot becomes available. After a small number of flows have been allowed to the service, every new call is approved based on system admission criteria, as illustrated in Fig.1. If a new call meets the admission criteria, it will be accepted into the system. The admitted flows are serviced in accordance with DCTS policy. Because accepted packets must wait for a period of time before being scheduled, the system prioritizes calls with a longer deadline and higher packet loss tolerance, i.e., call admission criteria is proportional to flow tolerance and deadline. This scheme's QoS performance is measured by the proportion of satisfied users and call blocking. Following that, we'll go into CAC analysis, which includes admission criteria, percentages of admitted users, and percentages of call blocking.

2.1. Deadline, Channel and Tolerance aware Scheduling (DCTS)

When making scheduling decisions, this method considers three factors: channel conditions, flow tolerance, and packet deadlines in queues. At first, this policy provides weight to each packet in a queue based on arrival time and packet delay, as well as the value of flow priority. The weights are assigned in decreasing order. The heavier weights going to packets that have a longer wait period. Furthermore, the total weight of each queue is computed by adding the weights of packets in the same queue that the link channel can support during slot t , i.e., only the weights of packets that the channel can contribute to the total weight of the respective queue among the number of

packets in the queue. We utilize the decreasing function f to implement DCTS, which is defined as

$$w_{i,k}(t) = f(x_i(t), \tau_{i,k}(t)) \quad (1)$$

where, $w_{i,k}(t)$ denotes the weight allocated to the k^{th} packet of the i^{th} queue. The parameters, $x_i(t)$ reflect the current priority of flow i and $\tau_{i,k}(t)$ represent the time left until a packet, $p_{i,k}$ at slot t expires. We suppose that f decreases monotonously with $x_i(t)$ and $\tau_{i,k}(t)$. e.g. $f(x_i(t), \tau_{i,k}(t)) = \exp\{-x_i(t)\tau_{i,k}(t)\}$. Equation (1) shows that the packet with the shortest time to expiration receives the most weight, while the package with the shortest time to expiry receives the least weight. This increases the likelihood of scheduling packets with a shorter expiration date. Each packet's weight is determined solely on the basis of its deadline. Following that, DCTS computes weight $w_i(t)$ for each flow, taking into account channel effects as follows:

$$w_i(t) = \sum_{k=h_i(t)+\varphi_i(t)+1}^{k=h_i(t)+1} f(x_i(t), \tau_{i,k}(t)) \quad (2)$$

where, $h_i(t)$ signifies the highest sequence number of the packet that was either transmitted or discarded until the start of t , i.e., $h_i(t)$ indicates the index of the head-of-line (HoL) packet of the i^{th} flow in slot t . Here, we assumed that $\tau_{i,k}(t) = \infty$ if $a_{i,k} > t$; otherwise, at the upper limit of the summation in (2), $\varphi_i(t)$ must be substituted by the minimum of $\varphi_i(t)$ and the number of packets existing in flow i 's queue. In the following stage, scheduling choices are determined using the weights $w(t) = \{w_1(t), \dots, w_N(t)\}$. The chance of selecting a flow is a function of w in DCTS, which is a randomized scheduling scheme $w(t)$. DCTS, in particular, makes the following scheduling decisions:

$$DCTS(t) = i \text{ with probability } \frac{e^{Lw_i(t)}}{1 + \sum_{j=1}^N e^{Lw_j(t)}} \quad (3)$$

L is a constant in this equation. In general, L is selected to be big such that scheduling a flow with the highest weight has a high likelihood. It's worth noting that even if flow i has no waiting packets, $w_i(t) = 0$, it has a positive chance of being scheduled. If this occurs, the system will just remain inactive. Now we will go over how DCTS updates flow priority, $x_i(t)$ for $i = 1, \dots, N$. The priority updates are based on the same principle as the gradient decent method. Assume that $l_i(x(t))$ represents the steady-state packet loss for priority assignments, with $x(t) = [x_1(t), \dots, x_N(t)]$. If $l_i(x(t)) > l_i a_i$, then the priority for flow i should go up (and $x_i(t)$ should go down). If $l_i(x(t)) < l_i a_i$, on the other hand, the priority for flow i should be reduced (x_i should be increased), providing other flows a greater opportunity to attain the target throughput. This leads to the gradient decent technique below.

$$x_i(t+1) = \max\{x_i(t) + \beta(t)[\tilde{l}_{i a_i} - l_i(x(t))], 0\} \quad (4)$$

where $\beta(t)$ is the step-size satisfying

$$\beta(t) > 0, \sum_{t=1}^{\infty} \beta(t) = \infty, \sum_{t=1}^{\infty} \beta^2(t) < \infty \quad (5)$$

Unfortunately, the preceding procedure cannot be implemented without knowing $l_i(x(t))$. In many circumstances, obtaining $l_i(x(t))$ necessitates obtaining the steady state distribution of the underlying Markov chain, which is computationally expensive. We propose a two-time scale based stochastic approximation technique [21] to solve the aforementioned problem. In this method, the steady state packet loss is calculated as a temporal average of the observed packet drop on a quicker timeline, while the priorities are updated on a slower period. Because of the time scale split, the priorities appear to be quasi static for the quicker time scale operation. We specifically update the following priorities: Let $\xi_i(t)$ signify the number of packets discarded owing to a deadline violation among flow i 's first t arrivals under Δ .

$$l_i(t+1) = l_i(t) + \alpha(t)[\xi_i^{DCTS}(t) - l_i(t)] \quad (6)$$

$$x_i(t+1) = \max\{x_i(t) + \beta(t)[\tilde{l}_{iai} - l_i(t)], 0\} \quad (7)$$

To obtain the time scale separation, we use $\alpha(t)$ and $\beta(t)$ in (8) as follows:

$$\alpha(t) > 0, \beta(t) > 0, \sum_{t=1}^{\infty} \alpha(t) = \sum_{t=1}^{\infty} \beta(t) = \infty, \sum_{t=1}^{\infty} (\alpha^2(t) + \beta^2(t)) < \infty \text{ and } \frac{\alpha(t)}{\beta(t)} \rightarrow 0. \quad (8)$$

where, $\alpha(t) = \frac{1}{t}$ and $\beta(t) = \frac{1}{1+t \log_{10}(t)}$. Also note that (6) computes the temporal average of the packet loss for $\alpha(t)$. We consider all flows to have identical QoS restrictions, i.e., the same traffic class, when determining the admission criterion. Our goal is to create a CAC algorithm that ensures that packet loss due to deadline violations is kept to a minimum for all allowed users. As a result, our CAC is based on a set of criteria that includes the maximum packet deadline, loss tolerance, and pace of newly received calls, as well as the accepted flows' minimum flow priority. Our scheme is a single-threshold based CAC, but it may be expanded to a multiple-threshold based CAC that can handle a variety of traffic types. We compare our scheme's performance to that of the CAC of VFEXP and the MLWDF methods.

3. Admission Control Analysis for Real-Time Applications

We show how our CAC scheme dynamically accepts additional calls. When a new call comes in, a CAC procedure is started to see if the new MS should be accepted. Our threshold-based CAC employs the following parameters for newly arriving flow j , new call arrival rate α_j , new call packet deadline D_j , new call flow tolerance l_j , and flow priority of already admitted MSs. As a result, when a new MS applies for admission to a system, the CAC scheme looks at the four requirements listed above. Let $\sum_j(t)$ be the admission criterion for flow j , with j being the number of newly admitted users at time slot t . $\sum_j(t)$ is calculated as follows:

$$\sum_j(t) = \frac{x_i \tilde{D}_{j_{min}}}{\alpha_j} \quad (9)$$

The minimum of the flow priority of already accepted MSs is determined as $x_{min}(t)$.

$$x_{min}(t) = \min_i \{X_i(t)\} \quad (10)$$

Furthermore, the service provider establishes the CAC threshold \sum_{th} . As a result, the new MS's admission requirements are as follows: if $\sum_j(t) > \sum_{th}$, the new MS is accepted to the system; otherwise, the new MS is blocked. The two QoS measurement metrics in this system are user happiness (% of satisfied users) and percentage of call blocking. If the proportion of satisfied users rises, it means that the scheduler is effectively serving admitted consumers for a particular threshold. However, there is a symbiotic relationship between the number of satisfied users and the percentage of blocked calls. Increased call blocking suggests that fewer users are accepted to the system, resulting in reduced network congestion. As a result, the threshold $\sum_j(t)$ is determined by weighing the trade-off between the two parameters. This policy combines the CAC with the scheduling of accepted flows, as seen in Fig. 1. For better demonstration, we've included some simple scheduling tasks here. The CAC requires the minimal flow priorities of admitted users for each timeslot, hence the system follows the procedures below.

Step 1: DCTS gives a weight to each packet of accepted flows for a fixed t . Packet $p_{i,k}$ is given a weight of

$$w_{i,k}(t) = \exp\{-x_i(t)\tau_{i,k}(t)\}, \quad (11)$$

where $x_i(t)$ denotes the current priority of allowed flow i and $\tau_{i,k}(t)$ denotes the time until a packet $p_{i,k}$ in slot t expires.

Step 2: For each accepted flow, the CAC system computes weight $w_i(t)$ as follows:

$$w_{i,t} = \sum_{k=h_i(t)+1}^{h_i(t)+\phi_i(t)+1} \exp\{-x_i(t)\tau_{i,k}(t)\}, \quad (12)$$

where $h_i(t)$ is the maximum sequence number of the packet that was either transmitted or discarded until the start of t , i.e., $h_i(t)$ is the index of the i^{th} flow's head-of-line (HoL) packet in slot t .

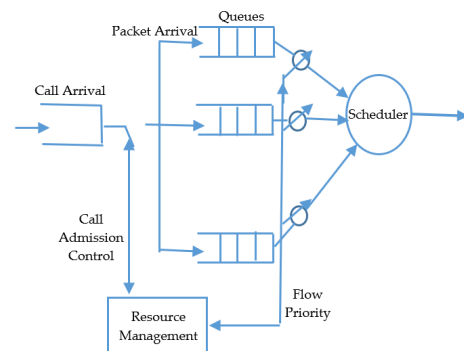


Figure 1: Call admission control based on a threshold for real-time applications

Step 3: Because DCTS is a randomized scheduling scheme, it selects flow i based on the probability F_i as:

$$F_i(t) = \frac{e^{LW_i(t)}}{1 + \sum_{j=1}^N e^{LW_j(t)}} \quad (13)$$

where L is a constant factor used to average the ratio of the maximum weight of flow i to the total weight of all flows using exponential averaging.

Step 4: CAC computes the $x_{min}(t)$ minimum of flow priority among all admitted flows, as specified in (2).

Step 5: CAC determines the decision criteria for the newly received flow j , as specified in (1).

Step 6: The CAC threshold Σ_{th} is compared to the decision criterion $\Sigma_j(t)$. Flow j is accepted to the queue if $\Sigma_j(t) > \Sigma_{th}$; else, it is denied.

Step 7: DCTS updates the $l_i(t)$ time average percentage of packet losses as follows:

$$l_i(t + 1) = l_i(t) + \beta_2(t)[\xi_i^{DCTS}(t) - l_i(t)], \quad (14)$$

Where $l_i(t)$ is equal to

$$l_i(t) = \frac{1}{t} \sum_{t=1}^t \xi_i^{DCTS}(t), \quad (15)$$

Step 8: At the conclusion of each frame, DCTS updates each user's flow priority as follows: where β_1 and β_2 are step-size in scheduling. The time average packet $l_i(n)$ for an n th frame is calculated as

$$x_i(n + 1) = \max\{x_i(n) + \beta_1(n)[\tilde{l}_i \alpha_i - l_i(n)], 0\} \quad (16)$$

The procedures outlined above can be used to calculate the proportion of satisfied users and the percentage of blocked calls. The fraction of admitted users among the number of arrived flows is used to calculate the percentage of user satisfaction. It is calculated for a certain time period or interval. We presume that after the call is accepted, the DSCT scheduler will handle each call according to their specific QoS requirements.

4. Simulation Results and Discussion

We compare our threshold-based CAC to CAC schemes based on MLWDF and VFEXP scheduling schemes to show how well it performs in terms of call blocking and percentage of satisfied users. Table 1 lists the parameters that were utilized in the simulation.

Table 1: List of parameters used in the simulation

Parameters	Values
Channel bandwidth	8MHz
Initial Number of flows	4
Maximum packet deadline of each flow	5,5,6,6 time slots
Arrival rate	10 packet per slot
Frame window size	10
L	2
Σ_{th}	Varies from 0.01 to 0.1

\tilde{l}	0.2
Channel holding time	0.1

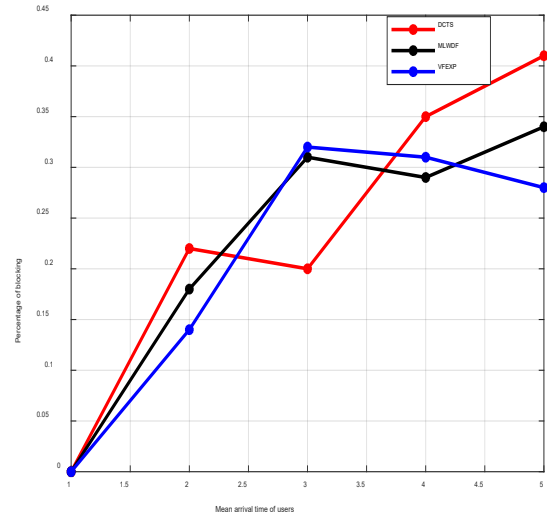


Figure 2: Percentage of Blocking for $\Sigma_{th} = 0.001$ and $\tilde{l} = 0.2$.

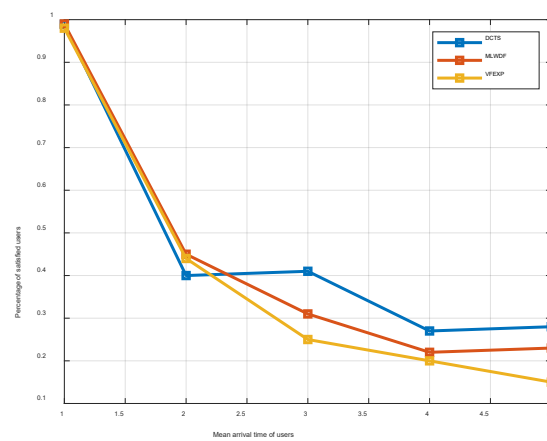


Figure 3: Percentage of satisfied users for $\Sigma_{th} = 0.001$ and $\tilde{l} = 0.2$.

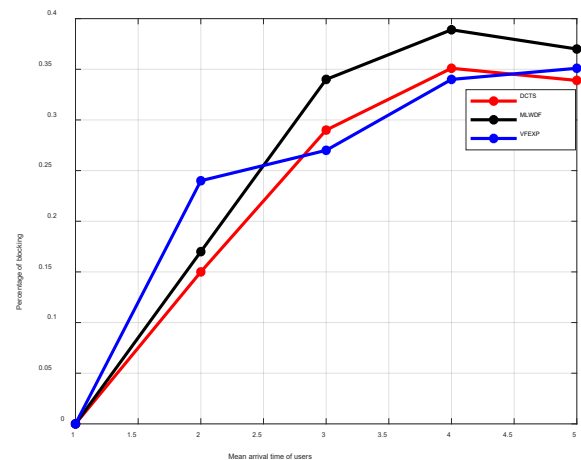


Figure 4: Percentage of Blocking for $\Sigma_{th} = 0.01$ and $\tilde{l} = 0.2$.

The proportion of call blocking for threshold-based CAC under DCTS, MIWDF, and VFEXP schedulers is shown in Fig.2. The flow tolerance and system threshold are 0.001 and 0.2, respectively. The percentage of calls that are

blocked is calculated for various mean arrival times. As demonstrated in the graph, our scheme's percentage of call blocking during the bulk of arrival time is lower than CAC under the MLWDF and VFEXP schedulers.

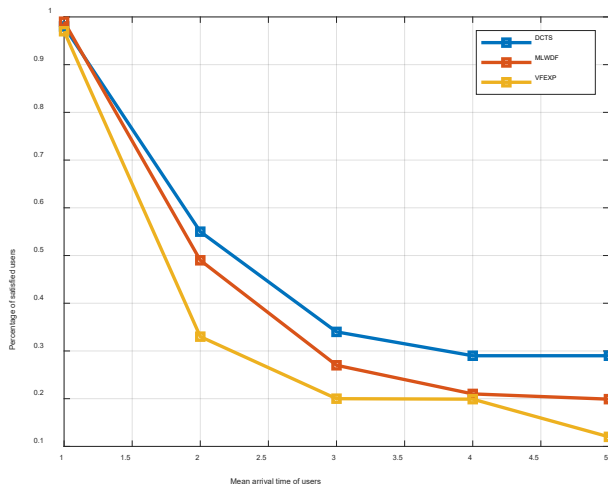


Figure 5: Percentage of satisfied users for $\Sigma_{th} = 0.01$ and $\tilde{l} = 0.2$.

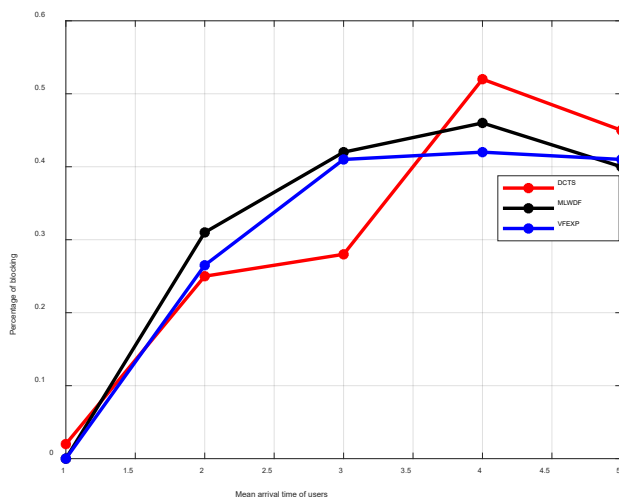


Figure 6: Percentage of Blocking for $\Sigma_{th} = 0.05$ and $\tilde{l} = 0.2$.

Figure 3 depicts the appropriate proportion of satisfied users. It also demonstrates that our scheme outperforms the other two in terms of mean arrival times of 1, 3, 4, and 5. The number of coming calls grows as the mean arrival time increases. As a result, the number of admitted users drops while the percentage of blocked calls rises. Figures 2 through 9 illustrate this point. The graph shows how changing the system threshold value impacts the call blocking percentage. A lower barrier allows for the admission of a greater number of flows. As a consequence, performance in terms of call blocking percentages and satisfied users improves. However, as

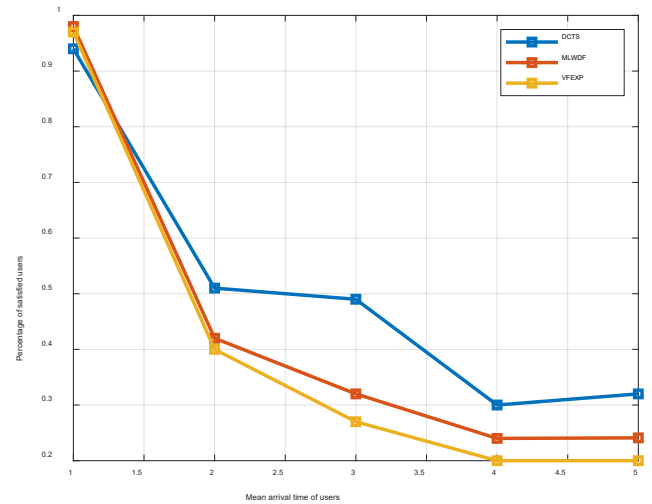


Figure 7: Percentage of satisfied users for $\Sigma_{th} = 0.05$ and $\tilde{l} = 0.2$.

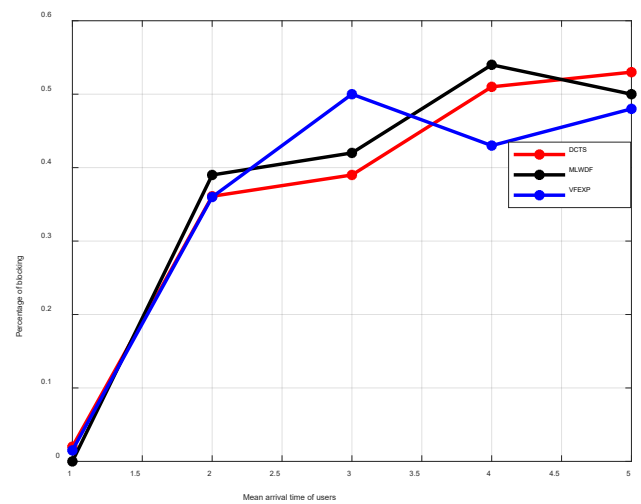


Figure 8: Percentage of Blocking for $\Sigma_{th} = 0.08$ and $\tilde{l} = 0.2$.

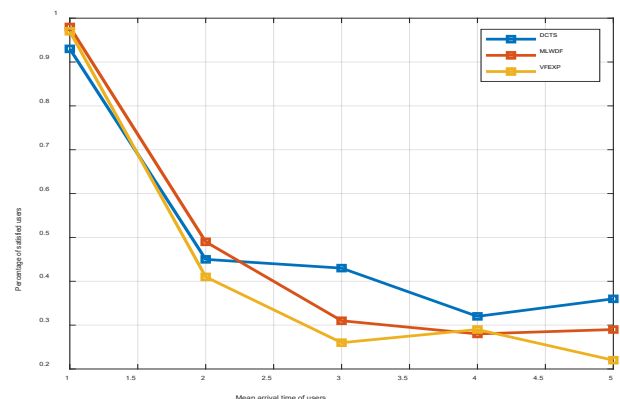


Figure 9: Percentage of satisfied users for $\Sigma_{th} = 0.08$ and $\tilde{l} = 0.2$.

the number of allowed users grows, the scheduler becomes overburdened, causing more packets to be dropped due to deadline breaches. Figures 3, 5, 7, and 9 have Σ_{th} values of 0.001, 0.01, 0.05, and 0.08, respectively. The equivalent percentages of satisfied consumers for our strategy with a mean arrival time of 1 are 0.97, 0.96, 0.92,

and 0.91, respectively. Even though the proportion of satisfied consumers has decreased, it may not be constant across various graphs (for mean arrival time of 2, 3, 4 and 5). When comparing the performance of the other two CAC schemes, the MLWDF-based CAC scheme outperforms the VFEXP-based CAC method.

5. Conclusion

We looked at the challenge of building a CAC scheme that supports our suggested DCTS scheduling scheme, which helps to reduce network congestion, in this research. We propose a novel call admission criterion based on system thresholds, with the goal of reducing packet losses in allowed flows. In terms of both percentage of call blocking and percentage of pleased consumers, our suggested CAC algorithm outperforms the CAC of the MLWDF and VF-EXP schemes, according to our simulation findings.

References

- [1] Z. Yang, A. Mohammed, T. Hult, & D. Grace, "Downlink coexistence performance assessment and techniques for WiMAX services from high altitude platform and terrestrial deployments," *EURASIP Journal on Wireless Communications and Networking*, 2008, DOI:10.1155/2008/291450.
- [2] H. M. Hussien, K. Katzis and L. P. Mfupe, "Dynamic Spectrum Allocation for TVWS Wireless Access from High Altitude Platform," *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 2021, pp. 1-6, DOI: 10.1109/ICECET52533.2021.9698667.
- [3] H. M. Hussien, K. Katzis and L. P. Mfupe, "Intelligent Power Allocation for Cognitive HAP Wireless Networks Using TVWS Spectrum," *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 2021, pp. 1-6, DOI: 10.1109/ICECET52533.2021.9698778.
- [4] H. M. Hussien, K. Katzis, L. P. Mfupe and E. T., "A Novel Resource Allocation for HAP Wireless Networks Exploiting TVWS Spectrum," *2021 IEEE AFRICON*, 2021, pp. 1-6, DOI: 10.1109/AFRICON51333.2021.9570928.
- [5] S.H. Alsamhi, N.S. Rajput, "Implementation of call admission control technique in HAP for enhanced QoS in wireless network deployment," *Telecommun Syst* 63, 141–151 (2016), doi:10.1007/s11235-015-0108-4
- [6] K. Katzis, L. Mfupe and H. M. Hussien, "Opportunities and Challenges of Bridging the Digital Divide using 5G enabled High Altitude Platforms and TVWS spectrum," *2020 IEEE Eighth International Conference on Communications and Networking (ComNet)*, 2020, pp. 1-7, DOI: 10.1109/ComNet47917.2020.9306090
- [7] H. M. Hussien et al., "Bridging the Urban-Rural Broadband Connectivity Gap Using 5G Enabled HAPs Communication Exploiting TVWS spectrum," *J. Engg. Res. & Sci*, 2022 (in press).
- [8] J. G. Andrews, A. Ghosh, R. Muhamed, "Fundamentals of WiMAX: Understanding Broadband Wireless Networking", *Prentice Hall*, 2007.
- [9] S. Ahson Mohammad Ilyas, "WiMAX Technologies, Performance Analysis, and QoS," *CRC Press*, 2008.
- [10] R. Krishna Rao and G.Radhamani, "WiMAX A Wireless Technology Revolution," *Auerbach Publications*, 2008.
- [11] M. D. Katz and F. H.P. Fitzek, "WiMAX Evolution Emerging Technologies and Applications" *John Wiley and Sons*, 2009.
- [12] Z. Abate, "WiMAX RF Systems Engineering," *Artech house*, 2009.
- [13] H. Wang, L. P. Kondi, A. Luthra, and S. Ci, "4G Wireless Video Communications," *John Wiley and Sons*, 2009.
- [14] J. Ni, D. H. K. Tsang, S. Tatikonda and B. Bensaou, "Threshold and reservation based call admission control policies for multiservice resource-sharing systems," *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, 2005, pp. 773-783 vol. 2, DOI: 10.1109/INFCOM.2005.1498309.
- [15] Z. Quan and J.M. Chung, "Admission control for probabilistic services with earliest deadline first scheduling In Local and Metropolitan Area Networks," *The 13th IEEE Workshop on*, pages 211 – 215, April 2004, DOI: 10.1109/LANMAN.2004.1338433
- [16] M. H. Ahmed, "Call admission control in wireless networks: A comprehensive survey," in *IEEE Communications Surveys & Tutorials*, vol. 7, no. 1, pp. 49-68, First Qtr. 2005, DOI: 10.1109/COMST.2005.
- [17] S.S Tzeng, "Call admission control policies in cellular wireless networks with spectrum renting," *Computer Communications*, 32(18), December 2009, DOI: 10.1016/j.comcom.2009.07.017
- [18] S. Ahson Mohammad Ilyas, "WiMAX Technologies, Performance Analysis, and QoS," *CRC Press*, 2008.
- [19] M. Andrews et al. "CDMA data QoS scheduling on the forward link with variable channel conditions", *Bell Laboratories Technical Report*, April, 2000
- [20] M.F. KHALED et al. "Channel-Aware Earliest Deadline Due Fair Scheduling for Wireless Multimedia Networks," *Wireless Personal Communications* (2006) 38: DOI: 233–252, 10.1007/s11277-006-9013-1
- [21] S.B. Vivek S., "Stochastic Approximation: a dynamic systems viewpoint," *Cambridge University Press*, 2008.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



HABIB MOHAMMED HUSSEIN received his bachelor degree from Adama Science and Technology University, Ethiopia in 2008 and his MSc degree from Tianjin University, Tianjin,

Engineering (Signal and Information Processing Technology). Now, he is a PhD Doctoral candidate at Addis Ababa University, Addis Ababa Institute of Technology (AAiT), Ethiopia. His current research interest includes high altitude platform, handoff schemes, TV white space technology (TVWS), radio resource allocation and optimization, call admission control, heterogenous network coexistence issues and 5G IoT.



SULTAN FEISSO MEKO received the B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from Adama University, Addis Ababa University, Ethiopia and IIT-Bombay, India in 1999, 2004 and 2014 respectively.

He served as Lecture at Maichew Technical College, Tigray, Ethiopia from 1999 to 2005, and as Lecturer and Assistance Professor at Adama Science and Technology University, Oromia, Ethiopia from 2006 to 2015. Since 2016, he has been with Addis Ababa Science and Technology University, Addis Ababa, where he is currently Assistance professor with the Department of Electrical and Computer Engineering. His current research interests include optimization of stochastic systems, resource allocation, scheduling and queuing model in wired/wireless networks.



KONSTANTINOS KATZIS received his BEng degree in Computer Systems Engineering and his MSc degree in Radio Systems Engineering from the University of Hull in 2000 and 2001 respectively. degree in Computer Systems

In 2006, Dr Katzis received his PhD degree in Electronics from University of York (UK). His current research interests include dynamic spectrum access and cognitive radio, architectures for 5G and beyond and wireless communications from aerial platforms. He currently holds the position of the secretary in IEEE1900.6 standard and he is also MC member of COST action CA15014 (IRACON). Recently, he has been awarded with the Fulbright Visiting Scholar fund for his proposal "Requirement Analysis of 5G Networks Supporting IoT-Health Applications" in collaboration with the Information Technology Laboratory (ITL) of the National Institute of Standards & Technology (NIST) in Washington DC (USA). Dr Konstantinos Katzis is an Associate Professor at European University Cyprus.



LUZANGO PANGANI MFUPE earned his Bachelor's, Master's and PhD degrees in Electrical Engineering at Tshwane University of Technology in Pretoria, South Africa.

He is an experienced principal research scientist, Innovator and Entrepreneur with demonstrated technical leadership in large & cutting-edge technology projects. Skilled in Dynamic Spectrum Management (DSM), Television White Spaces (TVWS), Defense & Civilian Spectrum Sharing Models, Formulation of Spectrum Regulations & Policies, Development of Geo-location Spectrum Databases, Advanced Wireless Networks Design and Modeling, Satellite Communications, Airborne Wireless Networks (AWNs), Artificial Intelligence/Machine Learning for Future Wireless Communications such as 5G, Computer Simulations, Lecturing, Supervision & SMMEs support.



EPHREM TESHALE BEKELE received the BSc degree in Electrical Engineering from Bahir Dar University, Bahir Dar, Ethiopia, in 2007, the MSc degree in Telecommunications Engineering in 2011,

and the PhD in Information and Communications Technology in 2015 both from the University of Trento, Trento, Italy. He has worked as an Assistant Lecturer in Bahir Dar University, from 2007 to 2009. He is currently an Assistant Professor at the Addis Ababa Institute of Technology (AAiT) in Addis Ababa University, Ethiopia, and a Associate Faculty Member of the ELEDIA Research Center. He is also a visiting lecturer at Bahir Dar University, Bahir Dar, Ethiopia, and a member of the Applied Electromagnetic Research Group in AAiT. His main research interests are Electromagnetic Nondestructive Testing, Technology 'Domestication', and RF Regulation.

Buckling Analysis of a Three-Dimensional Rectangular Plates Material Based on Exact Trigonometric Plate Theory

Onyeka Festus Chukwudi ^{1,3}, Okeke Thompson Edozie ^{2,*}, Nwa-David Chidobere ³

¹Department of Civil Engineering, Edo State University Uzairue, Edo State, 312102, Nigeria.

²Department of Civil Engineering, University of Nigeria, Nsukka, Enugu State, 410101, Nigeria.

³Department of Civil Engineering, Michael Okpara University of Agriculture, Umudike, Abia State, 440109, Nigeria.

*Corresponding author: Okeke, Thompson Edozie, edozie.okeke@unn.edu.ng

Corresponding author ORCID: <https://orcid.org/0000-0002-2668-9753>

ABSTRACT: In this study, exact trigonometric displacement function was used to solve the buckling problem of a three-dimensional (3-D) rectangular plate that is clamped at the first-three edges and the other remaining edge simply supported (CCCS) under uniaxial compressive load. Employing 3-D constitutive relations which consist of entire components, the functional for total potential energy was obtained. After that, the rotation and deflection at x-axis and y-axis were formulated from the established compatibility equations to get an exact trigonometric deflection function. The characteristics equation was obtained by differentiating energy equation with respect to deflect to obtain the relations between deflection and rotation. The equation of the total potential energy is minimized with respect to the deflection coefficient after incorporating the deflection and rotation function, the critical buckling load formula was established. The solution for the buckling problem gotten shown that the structure of the plate is safe when the plate thickness is increased as the outcome of the study showed that the critical buckling load increased as the span- thickness ratio increased. The overall difference in form of percent between the present work and previous studies recorded is 5.4%. This shows that at about 95% certainty, the present work is perfect. The comparison of this study with the results of previous similar studies revealed the uniformity 3-D plate theory and the variations of CPT and RPT theories in the exact buckling analysis of a rectangular plate. However, this approach which includes all the six stress elements of the plate material in the analysis produced an exact deflection function unlike the previous studies which used assumed functions. Furthermore, the theoretical analysis of this study demonstrates a novel approach to solve the buckling problem rectangular plate which is capable of analyzing rectangular plates of any thickness configuration.

KEYWORDS: CCCS rectangular thick plate, exact trigonometric deflection function, three-dimensional plate theory, critical buckling load under uniaxial compression

1. Introduction

Plates are three dimensional elements that comprises of two large plane perpendicular surfaces detached by a thickness [1, 2] and its application is widely seen in several engineering fields such as aeronautics, astronautic engineering, structural industries and marine structures. In recent years, the use of plates has greatly increased as a result of some benefits such as its versatility, affordability lightweight and its ability to resist heavy loads [3].

It is essential to examine the geometry of the plates and their supports, the behavior of the material used, the

type of loads and their method of application; due to its complexity [4]. Plates can be rectangular, square or circular in shape. Based on the materials of construction, plates can be anisotropic, isotropic and orthotropic. The plates also have free edge, simply supported and fixed supported edge conditions [5]. Plates can also be classified as thin, thick and moderately thick plates based on their thickness [6, 7]. Based on span-to-depth ratio, [7] considered $a/t \leq 10$ as thick plate, $15 \leq a/t \leq 40$ as moderately thick plate and $a/t \geq 45$ as thin plate. The importance of thick plate and its wide application in different aspects of engineering has necessitated more researches so as to maximize its application potential.

The bending, vibration and the buckling of plates are research areas for analyzing thick plates [8]. The phenomenon where a critical load value without further load application makes a material under the influence of in-plane loads to move from a stable to unstable equilibrium state is called buckling [9]. Buckling of plates can be classified as elastic and inelastic. For elastic buckling analysis, when the plate is stressed beyond the elastic limit before the beginning of buckling, it is said to be an inelastic or plastic buckling problem [10]. When the in-plane compressive load is increased beyond their critical values, the outcome is totally plate failure. Many scholars have employed different methods to solve the elastic buckling problems of thick plates. Generally, these methods include the equilibrium (Euler) methods, numerical methods, and energy methods [7].

Equilibrium methods such as the Fourier series method, the methods of integral transformation, Navier's double trigonometric series method, and the separation of variables method; Subject to the restraints of the plate edges and the loading boundary conditions, yields closed form solutions to the governing partial differential equations of equilibrium of the elastic buckling problem of plates within the domain of the plate [11]. However, the limitation of this method is that their application to plates with fixed edges, free edges and mixed support condition results in analytical difficulties. Numerical methods such as finite difference methods, finite strip methods, the finite element methods, and boundary element methods; yield approximate numerical solutions of the plate problem. Energy methods such as Rayleigh-Ritz method, Ritz variational method, and Kantorovich variation method, are methods of investigating the elastic buckling behavior of plates based on the application of the minimum principle to the total potential energy of the plate buckling problem. To obtain the characteristic buckling equation whose roots yields the buckling loads, energy methods minimize the total energy function of the plate's elastic buckling problem as regards to the parameters of the shape function.

A series of theories have been developed and employed to investigate the stability of plates. Such theories include the classical plate theory (CPT) and the refined plate theory (RPT). The CPT is a plate theory that ignores shear deformation that is lying across it, which is mostly used for analyzing thin plates [12]. The CPT overestimates buckling loads, natural frequencies and underestimates deflections for moderately thick or thick plate. It is inadequate for thick plate's analysis as a result of neglected transverse shear strains.

The refined plate theory (RPT) was formulated to overcome the limitations of the CPT. The RPT proposed by authors in [13] and [14] addressed vertical shear deformation effect in plate analysis. The RPT include;

First order shear deformation theory (FSDT), Second order shear deformation theory (SSDT). To produce accurate results, the FSDT and SSDT require a shear correction factor. It is difficult to obtain exact values because the shear correction factor depends on parameters such as support case, loading type and geometry. The High-order shear deformation theory (HSDT) was developed in order to consider the shear deformation effect without the factor of shear correction in plate analysis and to achieve a real difference between transverse shear stress at the plate surfaces [15]. The negligence of normal stress and strain along the thickness axis of the plate, makes the refined plate theory inconsistent, hence it can be addressed as a 2-D theory or an incomplete 3-D theory [16].

Employing a modified Stowell's technique, the authors in [16] investigated the buckling of an isotropic rectangular plate that is uniaxially loaded with a free edge that is not loaded. The authors considered plates with edge conditions of SCFC, CSFS, SSFS, CCFC, CSFC and SCFS. Applying polynomial functions from Taylor-Maclaurin series for each plate, the shape function was first determined. For different values of aspect ratio and moduli ratio, the authors presented the coefficients of buckling of each plate. Although their findings agreed reasonably with previous works, their study did not address a thick plate nor was a CCCS edge condition considered. The authors also did not employ the trigonometric functions.

To investigate the buckling behavior of a rectangular thick plate subjected to uniaxial in-plane compressive loading, the authors in [17] applied polynomial displacement functions. The direct governing equation which was obtained from orthogonal polynomial displacement functions, was solved to obtain the equation used to obtain the parameters for the determining buckling load. Although their results so similar when compared with previous studies as they did not involve CCCS edge condition and did not take trigonometric function into account.

The authors in [18] employed a split-deflection method and applied to the buckling analysis of thin rectangular plates under vibration. They also applied trigonometric to obtain the governing equation of the plate's equilibrium forces which was solved to get deflection function for CCCS rectangular plate analysis; their study did not address a thick plate as their assumption is limited to CPT which will not yield a good result when the plate is relatively thick.

The shape function derived from polynomial series was applied in the work of authors in [19] to determine the buckling behavior of the isotropic rectangular plate using the Ritz method. The authors substituted the shape function into the potential energy function, which was

later minimized and the critical buckling load was obtained. The authors did not take into account a three dimensional CCCS thick plate. There was also no consideration for trigonometric function to ensure the exact solution rather than other shape function whose solution cannot be reliable for analysis of thick and moderately thick plate [20].

The third-order energy functional was used by authors in [21] to investigate the buckling of CCSS and CCCS isotropic rectangular plates. At every point in the plate within the plate domain and adding the product of the stress and strain, the authors formulated third order strain energy. To get the total energy functions of third order, strain energy was also summed with the external load. The method of direct variation was adopted to actualize this. Although the results of their study showed some degree of accuracy when compared to previous works, they did not consider the application of trigonometric displacement function.

Variation energy method was used by authors in [22] to analyze the buckling of a 3-D CCCC thick plate applying both polynomial and trigonometric function. Although the three-dimensional plate theory and the uniaxial compressive load were applied, the authors failed to consider a thick plate with the CCCS boundary condition. The buckling of isotropic stiffened CCCC plate was investigated by the authors in [23]. The authors employed the use of the work principle approach and varied the aspect ratios, the properties of stiffness and the number of stiffeners, to analyze the critical buckling of the plates. The shape function was theoretically derived using Taylor Maclaurin's displacement function. The authors in [23] did not apply trigonometric displacement function and 3-D plate theory. The CCCS plate was not addressed in both studies. Since a thick plate is a typical three-dimensional structural element and not much work has been done applying 3-D plate theory, this present study is necessary.

Considering the approach of exact displacement function, the authors in [24, 25 and 26] obtained an analytic solution for the elastic buckling analysis of plates. The authors derived the displacement function from the compatibility equation to get a close form solution for 3-D stability analysis of thick plate under uniformly distributed load. The authors in [25] did not apply trigonometric function which predict the exact buckling load and both authors [24, 25 and 26] did not address isotropic thick plates with three clamped edges and one simply supported edge (CCCS).

The physical interpretation of CCCS plate is that, the three clamped edges are supported by a column and the other supported a beam, depicting the relevance of the present study. This is because the boundary condition depends on the type of beam/column support in the plate,

thus when CCCS initial condition occurs in the a plate material, analyzing it as any other type of plate as mentioned in the literature will not account for stresses induced. This is because, stresses are induced due to the applied load (in-plane load) in this case, and hence, non-negligible error results.

This work filled the gap as they applied the variation energy method with a trigonometric displacement function to get an exact three-dimensional stability rectangular plate under uniaxial compressive load. The focus of this study is to determine the calculating formula for the critical buckling load of thick plate that was clamped at the three edges and other edge simply supported (CCCS). Furthermore, the aspect ratio effect of the critical buckling load of the plate was evaluated to show its capacity to analyze different categories of plate, ranging from thin, thick and moderately thick plate.

2. Methodology and Theoretical Analysis

The theoretical analysis of this study lies in the buckling behavior of a three-dimensional rectangular plate's material based on exact trigonometric plate theory. This study demonstrates a novel approach to solve the buckling problem of a three-dimensional CCCS rectangular plate subjected to compressive load. In this work, and displacement in x , y and z axis; u , v and w respectively are applied and presented in Equation (1), (2) and (3) (see [22]):

$$u = z\theta_x \quad (1)$$

$$v = z\theta_y \quad (2)$$

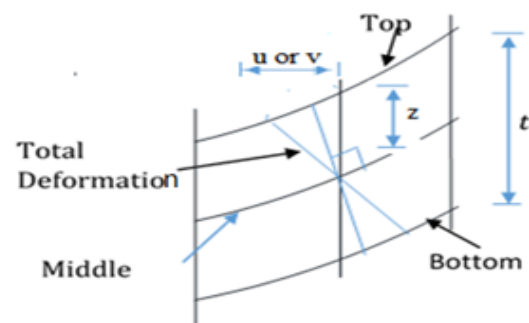


Figure 1: displacement of x - z (or y - z)

It can be seen in figure 1, that the six strains and stress elements required for the analysis were determined in line with the work of authors in [22].

2.1. Formulation of Total Potential Energy

The energy $[\Pi]$ equation were obtained in line with the authors in [23] and presented as:

$$\begin{aligned}
 \Pi &= \frac{D^*ab}{2a^2} \int_0^1 \int_0^1 \left[(1-\mu) \left(\frac{\partial \theta_x}{\partial R} \right)^2 + \frac{1}{\beta} \frac{\partial \theta_x}{\partial R} \cdot \frac{\partial \theta_y}{\partial Q} \right. \\
 &+ \frac{(1-\mu)}{\beta^2} \left(\frac{\partial \theta_y}{\partial Q} \right)^2 + \frac{(1-2\mu)}{2\beta^2} \left(\frac{\partial \theta_x}{\partial Q} \right)^2 + \frac{(1-2\mu)}{2} \left(\frac{\partial \theta_y}{\partial R} \right)^2 \\
 &+ \frac{6(1-2\mu)}{t^2} \left(a^2 \theta_x^2 + a^2 \theta_y^2 + \left(\frac{\partial w}{\partial R} \right)^2 + \frac{1}{\beta^2} \left(\frac{\partial w}{\partial Q} \right)^2 \right. \\
 &+ 2a \cdot \theta_x \frac{\partial w}{\partial R} + \frac{2a \cdot \theta_y}{\beta} \frac{\partial w}{\partial Q} \left. \right) + \frac{(1-\mu)a^2}{t^4} \left(\frac{\partial w}{\partial S} \right)^2 \\
 &\left. - \frac{N_x}{D^*} \cdot \left(\frac{\partial w}{\partial R} \right)^2 \right] dR dQ \quad (3)
 \end{aligned}$$

Where; N_x is the uniform applied uniaxial compression load of the plate.

2.2. Compatibility Equations

According to the work of authors in [24], the compatibility equation was obtained by differentiation of total potential energy with respect to θ_x and θ_y to establish the relationship between the rotation and deflection:

$$\begin{aligned}
 \frac{\partial \Pi}{\partial \theta_x} &= \frac{D^*ab}{2a^2} \int_0^1 \int_0^1 \left[(1-\mu) \frac{\partial^2 \theta_x}{\partial R^2} + \frac{1}{2\beta} \cdot \frac{\partial^2 \theta_y}{\partial R \partial Q} \right. \\
 &+ \frac{(1-2\mu)}{2\beta^2} \frac{\partial^2 \theta_x}{\partial Q^2} \\
 &\left. + \frac{6(1-2\mu)}{t^2} \left(a^2 \theta_x + a \cdot \frac{\partial w}{\partial R} \right) \right] dR dQ \\
 &= 0 \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial \Pi}{\partial \theta_y} &= \frac{D^*ab}{2a^2} \int_0^1 \int_0^1 \left[\frac{1}{2\beta} \cdot \frac{\partial^2 \theta_x}{\partial R \partial Q} + \frac{(1-\mu)}{\beta^2} \frac{\partial^2 \theta_y}{\partial Q^2} \right. \\
 &+ \frac{(1-2\mu)}{2} \frac{\partial^2 \theta_y}{\partial R^2} \\
 &\left. + \frac{6(1-2\mu)}{t^2} \left(a^2 \theta_y + \frac{a \cdot \partial w}{\beta \partial Q} \right) \right] dR dQ \\
 &= 0 \quad (5)
 \end{aligned}$$

Given that [2];

$$\theta_x = \gamma_{xz} - \frac{1}{a} \frac{\partial w}{\partial R} = \frac{c}{a} \frac{\partial h}{\partial R} \quad (6)$$

$$\theta_y = \gamma_{yz} - \frac{1}{a\beta} \frac{\partial w}{\partial Q} = \frac{c}{a\beta} \frac{\partial h}{\partial Q} \quad (7)$$

Hence, for zero integrands, the true solution was gotten by simplifying and factorizing the outcome of the compatibility equation to give the algebraic solution in Equation (8) which is the relation of known and unknown variable to get the constant quantity c .

$$\frac{6(1-2\mu)(1+c)}{t^2} = -\frac{c(1-\mu)}{a^2} \left(\frac{\partial^2 w}{\partial R^2} + \frac{1}{\beta^2} \frac{\partial^2 w}{\partial Q^2} \right) \quad (8)$$

2.3. Governing Equations

The governing differential equation was obtained after the total potential energy was differentiated with respect to deflection (w) and its solution gives the exact deflection function:

$$\begin{aligned}
 \frac{\partial \Pi}{\partial w} &= \frac{D^*}{2a^2} \int_0^1 \int_0^1 \left[\frac{12(1-2\mu)}{t^2} \left(\frac{\partial^2 w}{\partial R^2} + \frac{1}{\beta^2} \frac{\partial^2 w}{\partial Q^2} + a \cdot \frac{\partial \theta_x}{\partial R} \right. \right. \\
 &+ \left. \frac{a}{\beta} \frac{\partial \theta_y}{\partial Q} \right) + 2 \frac{(1-\mu)a^2}{t^4} \cdot \frac{\partial^2 w}{\partial S^2} \\
 &\left. - 2 \frac{N_x}{D^*} \cdot \frac{\partial^2 w}{\partial R^2} \right] dR dQ = 0 \quad (9)
 \end{aligned}$$

By factorizing and simplifying the outcome, gives:

$$\begin{aligned}
 \left(\frac{\partial^4 w_1}{\partial R^4} + \frac{2}{\beta^2} \cdot \frac{\partial^4 w_1}{\partial R^2 \partial Q^2} + \frac{1}{\beta^4} \cdot \frac{\partial^4 w_1}{\partial Q^4} - \frac{N_{x1} a^4}{g D^*} \cdot \frac{\partial^2 w_1}{\partial R^2} \right) w_S \\
 + \frac{w_1}{g} \left(\frac{(1-\mu)a^4}{t^4} \cdot \frac{\partial^2 w_S}{\partial S^2} \right) = 0 \quad (10)
 \end{aligned}$$

The true solution becomes:

$$\frac{\partial^4 w_1}{\partial R^4} + \frac{2}{\beta^2} \cdot \frac{\partial^4 w_1}{\partial R^2 \partial Q^2} + \frac{1}{\beta^4} \cdot \frac{\partial^4 w_1}{\partial Q^4} - \frac{N_{x1} a^4}{g D^*} \cdot \frac{\partial^2 w_1}{\partial R^2} = 0 \quad (11)$$

$$\frac{(1-\mu)a^4}{t^4} \cdot \frac{\partial^2 w_S}{\partial S^2} = 0 \quad (12)$$

The solution of Equation (11) to get an exact deflection and slope of the plate.

$$w = \Delta_0 [1 \ R \ R^2 R^3 R^4] \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \cdot [1 \ Q \ Q^2 Q^3 Q^4] \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad (13)$$

Similarly:

$$\begin{aligned}
 \theta_x &= \frac{c}{a} \cdot \Delta_0 [1 \ 2R \ 3R^2 \ 4R^3] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \cdot [1 \ Q \ Q^2 Q^3 Q^4] \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \\
 &= \frac{A_{2R}}{a} \cdot \frac{\partial h}{\partial R} \quad (14)
 \end{aligned}$$

And;

$$\begin{aligned}
 \theta_y &= \frac{c}{a\beta} \cdot \Delta_0 [1 \ R \ R^2 R^3 R^4] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \cdot [1 \ 2Q \ 3Q^2 \ 4Q^3] \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \\
 &= \frac{A_{2Q}}{a\beta} \cdot \frac{\partial h}{\partial Q} \quad (15)
 \end{aligned}$$

Where:

$$A_{2R} = c \cdot \Delta_0 \cdot \frac{\partial h}{\partial R} \cdot A_Q \quad (16)$$

$$A_{2Q} = c \cdot \Delta_0 \cdot \frac{\partial h}{\partial Q} \cdot A_R \quad (17)$$

The constants; c , Δ_0 , A_R and A_Q , thus, putting Equations (13), (14) and (15) into (3), simplifying and gives:

$$\begin{aligned}
 \Pi &= \frac{D^*ab}{2a^4} \left[(1-\mu) A_{2R}^2 k_{RR} \right. \\
 &+ \frac{1}{\beta^2} \left[A_{2R} \cdot A_{2Q} + \frac{(1-2\mu) A_{2R}^2}{2} \right. \\
 &+ \left. \frac{(1-2\mu) A_{2Q}^2}{2} \right] k_{RQ} + \frac{(1-\mu) A_{2Q}^2}{\beta^4} k_{QQ} \\
 &+ 6(1 \\
 &- 2\mu) \left(\frac{a}{t} \right)^2 \left([A_{2R}^2 + A_1^2 + 2A_1 A_{2R}] \cdot k_R \right. \\
 &+ \left. \frac{1}{\beta^2} \cdot [A_{2Q}^2 + A_1^2 + 2A_1 A_{2Q}] \cdot k_Q \right) \\
 &\left. - \frac{N_x a^2 A_1^2}{D^*} \cdot k_R \right] \quad (18)
 \end{aligned}$$

where:

$$k_{RR} = \int_0^1 \int_0^1 \left(\frac{\partial^2 h}{\partial R^2}\right)^2 dRdQ; k_R = \int_0^1 \int_0^1 \left(\frac{\partial h}{\partial R}\right)^2 dRdQ \quad (19a)$$

$$k_{QQ} = \int_0^1 \int_0^1 \left(\frac{\partial^2 h}{\partial Q^2}\right)^2 dRdQ; k_Q = \int_0^1 \int_0^1 \left(\frac{\partial h}{\partial Q}\right)^2 dRdQ \quad (19b)$$

$$k_{RQ} = \int_0^1 \int_0^1 \left(\frac{\partial^2 h}{\partial R\partial Q}\right)^2 dRdQ \quad (19c)$$

Minimizing Equation (18) with respect to A_{2R} gives:

$$(1 - \mu)A_{2R}k_{RR} + \frac{1}{2\beta^2}[A_{2Q} + A_{2R}(1 - 2\mu)]k_{RQ} + 6(1 - 2\mu)\left(\frac{a}{t}\right)^2 [A_{2R} + A_1] \cdot k_R = 0 \quad (20)$$

Minimizing Equation (18) with respect to A_{2Q} gives:

$$\frac{(1 - \mu)A_{2Q}}{\beta^4}k_{QQ} + \frac{1}{2\beta^2}[A_{2R} + A_{2Q}(1 - 2\mu)]k_{RQ} + \frac{6}{\beta^2}(1 - 2\mu)\left(\frac{a}{t}\right)^2 ([A_{2Q} + A_1] \cdot k_Q) = 0 \quad (21)$$

Rewriting Equations (20) and (21) gives:

$$\left[(1 - \mu)k_{RR} + \frac{1}{2\beta^2}(1 - 2\mu)k_{RQ} + 6(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_R\right]A_{2R} + \left[\frac{1}{2\beta^2}k_{RQ}\right]A_{2Q} = \left[-6(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_R\right]A_1 \quad (22)$$

$$\left[\frac{1}{2\beta^2}k_{RQ}\right]A_{2R} + \left[\frac{(1 - \mu)}{\beta^4}k_{QQ} + \frac{1}{2\beta^2}(1 - 2\mu)k_{RQ} + \frac{6}{\beta^2}(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_Q\right]A_{2Q} = \left[-\frac{6}{\beta^2}(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_Q\right]A_1 \quad (23)$$

Solving Equations (22) and (23) simultaneously gives:

$$A_{2R} = G_2 A_1 \quad (24)$$

$$A_{2Q} = G_3 A_1 \quad (25)$$

Let:

$$G_2 = \frac{(c_{12}c_{23} - c_{13}c_{22})}{(c_{12}c_{12} - c_{11}c_{22})} \quad (26)$$

$$G_3 = \frac{(c_{12}c_{13} - c_{11}c_{23})}{(c_{12}c_{12} - c_{11}c_{22})} \quad (27)$$

$$c_{11} = (1 - \mu)k_{RR} + \frac{1}{2\beta^2}(1 - 2\mu)k_{RQ} + 6(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_R \quad (28)$$

$$c_{22} = \frac{(1 - \mu)}{\beta^4}k_{QQ} + \frac{1}{2\beta^2}(1 - 2\mu)k_{RQ} + \frac{6}{\beta^2}(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_Q \quad (29)$$

$$c_{12} = c_{21} = \frac{1}{2\beta^2}k_{RQ}; c_{13} = -6(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_R; c_{23} = c_{32} = -\frac{6}{\beta^2}(1 - 2\mu)\left(\frac{a}{t}\right)^2 k_Q \quad (30)$$

Minimizing Equation (18) with respect to A_1 and simplifying the outcome gives:

$$6(1 - 2\mu)\left(\frac{a}{t}\right)^2 \left([1 + G_2] \cdot k_R + \frac{1}{\beta^2} \cdot [1 + G_3] \cdot k_Q\right) - \frac{N_x a^2}{D^*} \cdot k_R = 0 \quad (31)$$

Rearranging Equation (31) and simplify to give:

$$\frac{N_x a^2}{D^*} = 6(1 - 2\mu)\left(\frac{a}{t}\right)^2 \left([1 + G_2] + \frac{1}{\beta^2} \cdot [1 + G_3] \cdot \frac{k_Q}{k_R}\right) \quad (32)$$

2.4. Formulation of Total Potential Energy

A numerical analysis is performed on the rectangular thick plate that was clamped in the first three edges and remaining edge simply supported under uniaxial compressive load as presented in Figure 2.

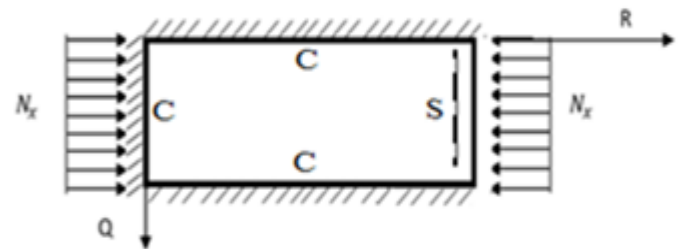


Figure 2: A section of CCCS rectangular plate

The plate at figure 2 has the following initial conditions:

$$\text{At } R = Q = 0; \text{ deflection } (w) = 0 \quad (33)$$

$$\text{At } R = Q = 0; \text{ slope } \left(\frac{dw}{dR} \text{ and } \frac{dw}{dQ}\right) = 0 \quad (34)$$

$$\text{At } R = Q = 1; \text{ deflection } (w) = 0 \quad (35)$$

$$\text{At } R = Q = 1, \text{ bending moment } \left(\frac{d^2 w}{dR^2}\right) = 0; Q = 1, \text{ slope } \left(\frac{dw}{dQ}\right) = 0 \quad (36)$$

Putting Equation (33) to (36) into the derived function of deflection gives:

$$2\text{Cos } g_1 + g_1 \text{Sin } g_1 - 2; 2\text{Cos } g_1 + g_1 \text{Sin } g_1 - 2 = 0 \quad (37)$$

The value of g_1 that satisfies Equation (37) is:

$$g_1 = 2m\pi \text{ [where } m = 1, 2, 3 \dots \text{]} \quad (38)$$

Substituting Equation (38) into the derivatives of w and satisfying the boundary conditions of Equation (33) to (36) gives the following constants:

$$a_0 = g_1 a_3; a_1 = -g_1 a_3; b_1 = b_3 = 0; b_0 = -b_2 = 0 \quad (39)$$

Substituting the constants of Equation (38) and (39) into Equation (10) and simplifying the outcome gives:

$$w = a_3 (g_1 - g_1 R - g_1 \text{Cos } g_1 R + \text{Sin } g_1 R) \times b_2 (\text{Cos } 2m\pi Q - 1) \quad (40)$$

This gives;

$$w = a_3 \times b_2 (g_1 - g_1 R - g_1 \text{Cos } g_1 R + \text{Sin } g_1 R) \cdot (\text{Cos } 2\pi Q - 1) \quad (41)$$

Let,

$$A_1 = a_3 \times b_2 \quad (42)$$

and;

$$h = (g_1 - g_1 R - g_1 \text{Cos } g_1 R + \text{Sin } g_1 R) \cdot (\text{Cos } 2\pi Q - 1) \quad (43)$$

Thus;

$$w = A_1 (g_1 - g_1 R - g_1 \text{Cos } g_1 R + \text{Sin } g_1 R) \cdot (\text{Cos } 2\pi Q - 1) \quad (44)$$

Where;

w and h are deflection and shape function of the plate respectively.

Using the particular trigonometric displacement function presented in Equation (44), the numerical of solution of the stiffness coefficients for deflection of rectangular thick plate analysis subjected to the CCCS boundary condition was obtained and presented.

Let; trigonometric stiffness coefficient values of CCCS plate $k_{RR} = 6173.94$; $k_{RQ} = 4061.11$; $k_{QQ} = 14851.88$; $k_R = 308.61$; $k_Q = 376.2$

Where; the Poisson's ratio of the plate be 0.3.

3. Results and Discussions

The numerical result of the analysis was obtained by employing trigonometric function at a different aspect ratio to get the result of non-dimensional critical buckling load on the three-dimensional CCCS rectangular thick plate presented in Figure 3 through 12. It is observed that the value of the critical buckling load N_x decreases as the value of the aspect ratio (1.0, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 2.0) increases at varying stiffness as shown in Figure 3 through 12. This indicates that the chance of failure in a plate structure increases with an increase in plate width. This value begins to vary as the aspect ratio increases. This is quite expected because the increase in the plate's aspect ratio decreases the capacity of the plate to resist buckling.

The span to thickness ratio of the plate varies between 4, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000 and 1500 at the varying breadth of the plate. More so, it is seen in the figures that the results the values of critical buckling load increase as the span- thickness ratio increases. This means that the failure in a plate structure is bound to occur as the in-plane load on the plate increases and gets to the critical buckling. The present theory predicts the buckling load of 64.3, 91.7, 102.5, 109.4, 112.1, 114.1, and 114.8 for a square plate in the span to thickness ratio of 5, 7, 15, 20, 30, and 40. Meanwhile, the value of the buckling load predicted for the same plate beyond the span to thickness ratio of 40 gives a constant value of about 115.4 when corrected to two decimal places. This shows that the result of the critical buckling load of thin and moderately thick plate using the 3-D theory is the same for the stability analysis of rectangular plate under the CCCS boundary condition. But looking at a graphical representation of these results in Figures 3 to 12 shows that at span-thickness ratio of 20 and above the graph appears straight along the horizontal axis. This confirms the study of the authors in [8] which showed that thin and moderately thick plates with the span to thickness ratio are from 20 and above.

In the validation of the result of the present study, a comparative analysis is performed to show the degree of divergence between the result of the present study with those of classical plate theory (CPT) and refined plate theory (RPT) as presented in Table 1&2 and Figure 13. The average percentage difference between the present study

and those of authors in [18], [20], [21] and [27] is about 6.33%, 6.37%, 6.47 and 2.59% respectively. Moreover, the result of the result of percentage difference being lower is quite expected because 3-D theory predict more close-form answer to the problem and prudent to use compared to the CPT and RPT.

The overall difference in percentage between the present study and those of 3-D elasticity theory in [27] is about 2.59%, while the average total percentage difference between the values from the present study and those of the 2-D RPT and CPT are 6.35 and 6.47 respectively. Here from, the overall difference in terms of percentage between the present study and those of the 3-D elasticity theory in [27] being lower than 2.6% show high level of insignificance in comparison, therefore debits that present theory gives an exact solution and can be reliable in the 3-D stability analysis of all types of a rectangular plate under such support configuration. From Figures 6, 7 and Table 4, the present study showed good agreement with previous studies but varied widely when considered as a thick plate at span to thickness ratios of 4 and 7. Meanwhile, the average difference in terms of a percentage between the present work and those of CPT in [18] and [20] and 2-D RPT in [21] is higher than 6.4%. It is noticed from the analysis that the present theory converges faster with exact elasticity theory than the CPT and RPT, which proves exactness and dependable nature of the derived relations. Thus, it was discovered that the present study become closer to those obtained using the CPT in [24] and [25] as the span-thickness ratio increases, and closest to those using RPT in [20] for a square plate at span to thickness ratio of 7. It should be noted that the present result when compared with exact elasticity theorem according to the authors in [27] gave a small percentage difference of 2.5%; the value which appears to be almost the same with exact elasticity theorem proves the exactness of the present theory. The result of present work being slightly higher than those of authors in [26] as seen in the Figure 12 does not invalidate the result as it's quite expected because the present theory which applied trigonometric theory predict more exact answer to the problems and prudent to use compared to other shape functions in the exact stability analysis of the plate. Meanwhile, the result of the present study is obviously higher than those of authors in [18], [20] and [21] because the authors in [18], [20] and [21] did not consider a typical 3-D analogy rather 2-D which was gotten by making suitable assumption to the kinematics of the relation or state of stress through the thickness axis of the plate which make their result inexact.

The result of comparative study confirmed that the 2-D RPT is only an approximate relation while the 3-D trigonometric function established in this study gives an exact solution for stability analysis of plates under the same boundary condition compared to polynomial [27,

28]. However, the overall difference in terms of percent between the present work and past works in comparison is about 5%. This means that at about 95% certainty, the values from the present study are the same as those from past scholars.

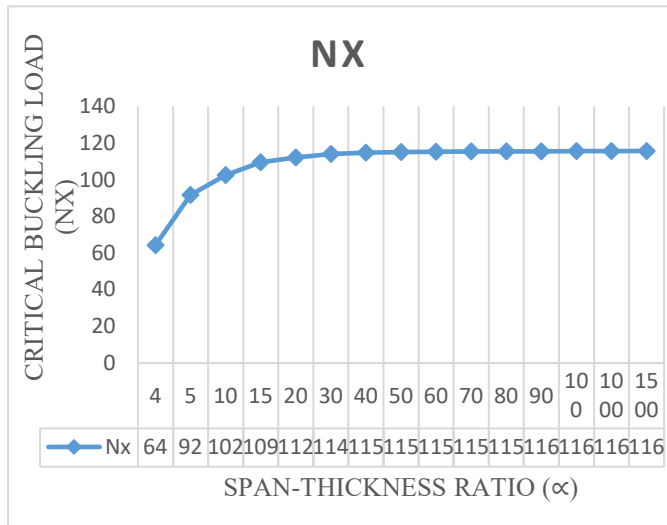


Figure 3: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a square rectangular plate.

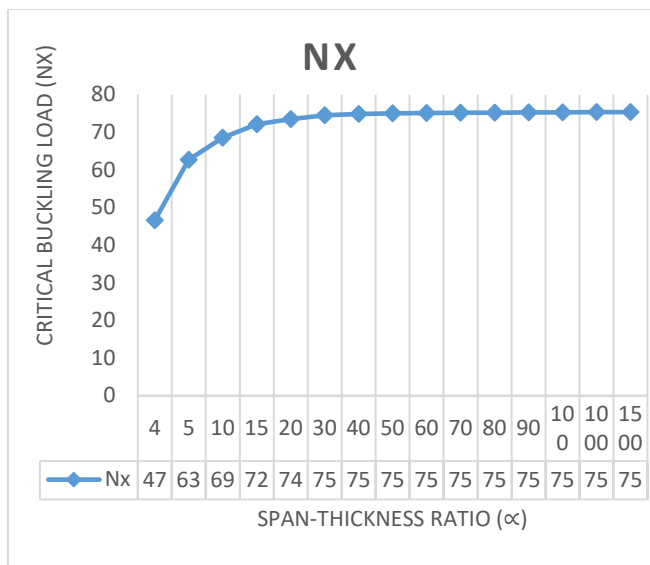


Figure 4: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 1.5

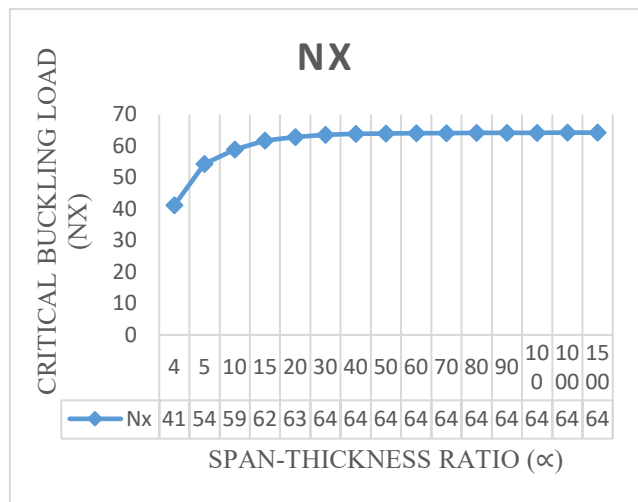


Figure 5: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 2.0

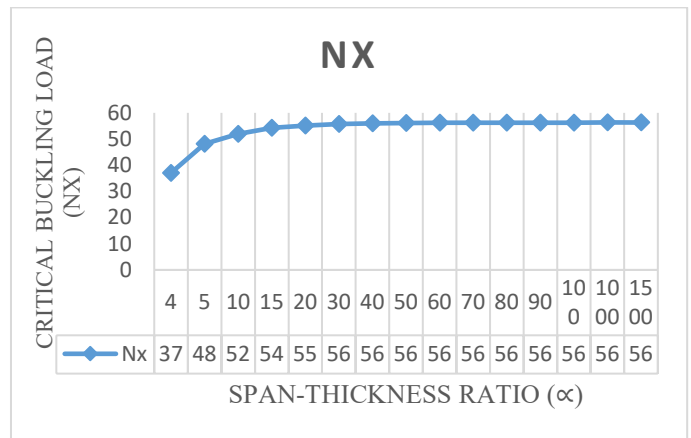


Figure 6: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 2.5

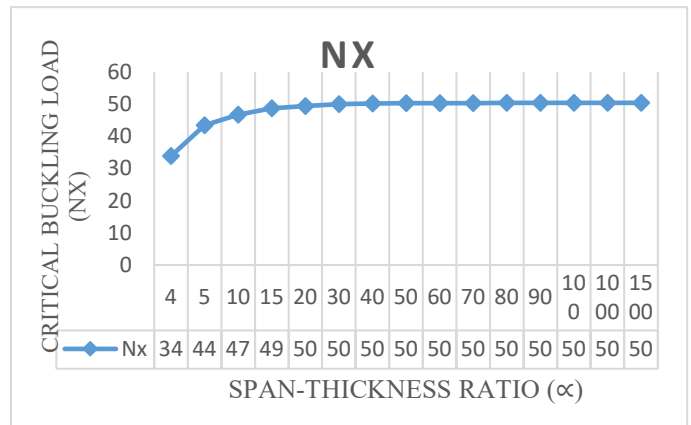


Figure 7: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 3.0

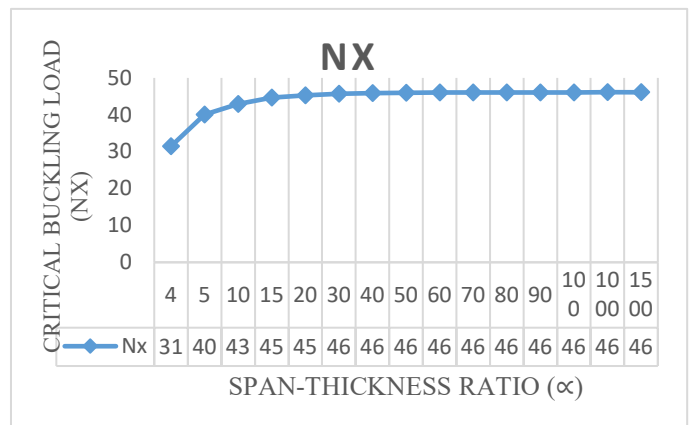


Figure 8: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 3.5

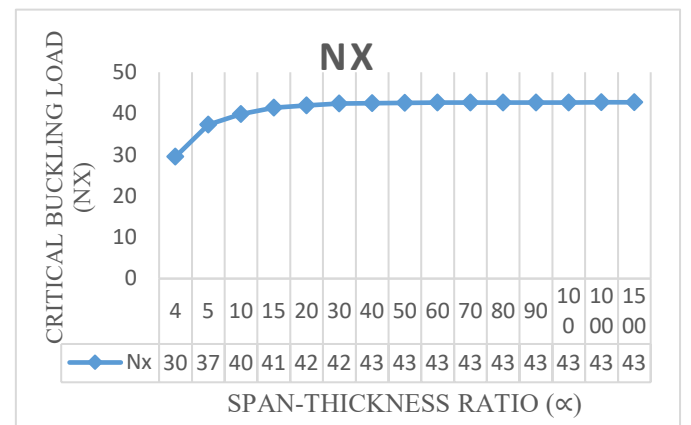


Figure 9: Graph of Critical buckling load (Nx) versus span-thickness ratio (α) of a rectangular plate at aspect ratio of 4.0

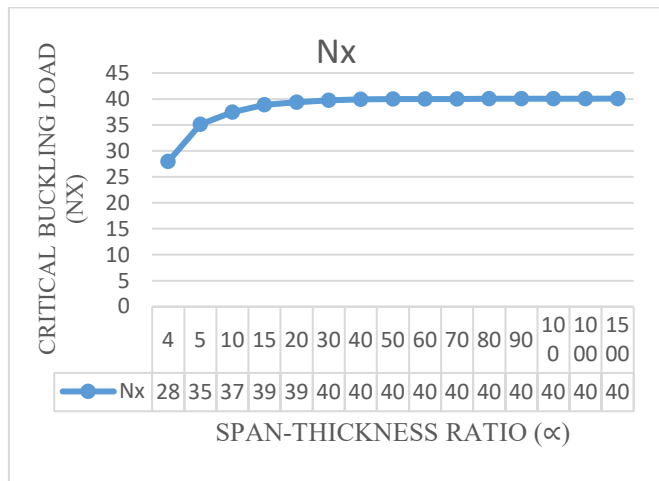


Figure 10: Graph of Critical buckling load (Nx) versus span-thickness (α) of a rectangular plate at aspect ratio of 4.5

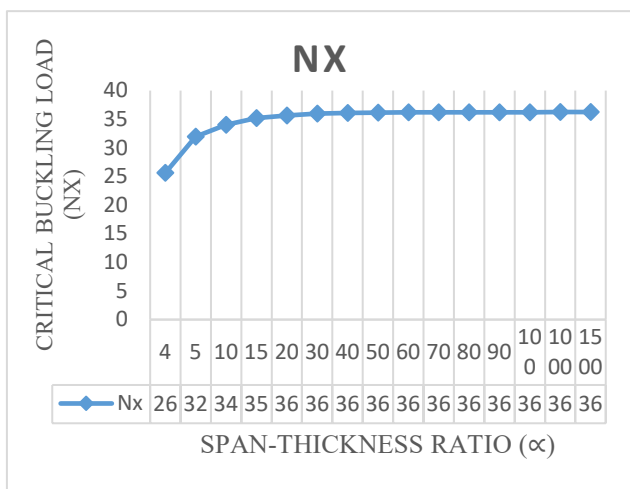


Figure 11: Graph of Critical buckling load (Nx) versus span-thickness (α) of a rectangular plate at aspect ratio of 5.0

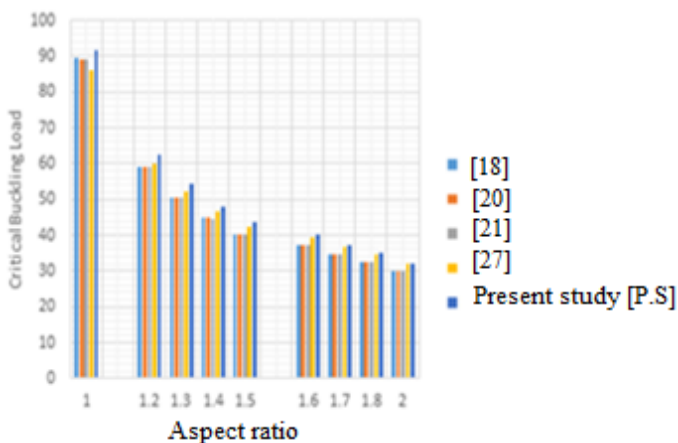


Figure 12: Graph of critical buckling load against aspect ratio (β) on comparative analysis of results from previous and present study at span-thickness ratio (α) of 7

Table 1: Comparison of critical buckling load of plate between the present study [P.S] and previous studies on CCCS plate at span to thickness ratio of 7

Critical Buckling Load Coefficients					
b/a	[18]	[20]	[21]	[27]	[P.S]
1.0	89.354	89.337	89.227	86.092	91.658
1.2	59.060	59.047	58.976	59.899	62.695
1.3	50.735	50.722	50.663	52.281	54.289

1.4	44.794	44.777	44.731	46.706	48.143
1.5	40.434	40.415	40.375	42.529	43.542
1.6	37.148	37.132	37.095	39.332	40.022
1.7	34.620	34.605	34.571	36.840	37.279
1.8	32.637	32.623	32.592	34.865	35.105
2.0	29.777	29.764	29.736	31.980	31.931

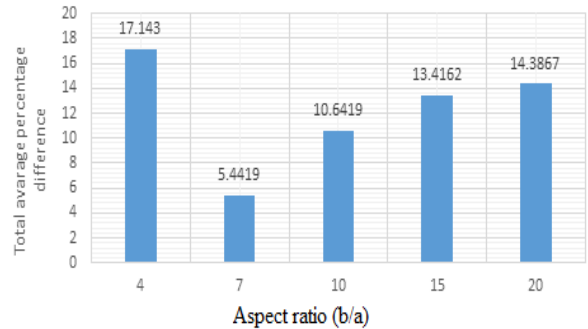


Figure 13: Graph of total average percentage difference against aspect ratio $\beta = b/a$

Table 2: Percentage difference analysis between the present study [P.S] and previous studies on CCCS plate at span to thickness ratio of 7

b/a	[18]&P.S	[20]&P.S	[21]&P.S	[27]&P.S
1.0	2.514	2.532	2.653	6.073
1.2	5.798	5.819	5.932	4.459
1.3	6.547	6.570	6.680	3.699
1.4	6.955	6.993	7.087	2.986
1.5	7.141	7.180	7.272	2.326
1.6	7.182	7.223	7.313	1.723
1.7	7.133	7.173	7.263	1.177
1.8	7.030	7.072	7.159	0.685
2.0	6.747	6.788	6.876	0.152
Av. % Diff	6.33	6.37	6.47	2.59
Tot al % Diff	5.4			

References

- [1]. F. C. Onyeka, O. M. Ibearugbulem, "Load Analysis and Bending Solutions of Rectangular Thick Plate," *International Journal of Emerging Technologies*, vol. 11, no. 3, pp. 1030-1110, 2020.
- [2]. F. C. Onyeka, T. E. Okeke, "Analysis of Critical Imposed Load of Plate Using Variational Calculus," *Journal of Advances in Science and Engineering*, vol. 4, no. 1, pp. 13-23, 2021, doi:10.37121/jase.v4i1.125.
- [3]. F. C. Onyeka, D. Osegbowa, E. E. Arinze, "Application of a New Refined Shear Deformation Theory for the Analysis of Thick Rectangular Plates," *Nigerian Research Journal of Engineering and Environmental Sciences*, vol. 5, no. 2, 901-917, 2020.
- [4]. F. C. Onyeka, "Critical Lateral Load Analysis of Rectangular Plate Considering Shear Deformation Effect," *Global Journal of Civil Engineering*, vol. 1, pp. 16-27, 2020, doi:10.37516/global.j.civ.eng.2020.0121.
- [5]. K. Shwetha, V. Subrahmanya, P. Bhat, "Comparison between thin plate and thick plate from navier solution using matlab software," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 6, pp. 2675 - 2680, 2018.

- [6]. F. C. Onyeka, E. T. Okeke, J. Wasiu, "Strain-Displacement expressions and their effect on the deflection and strength of plate," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 5, pp. 401-413, 2020, doi:10.25046/AJ050551.
- [7]. S. P. Timoshenko, J. M. Gere, W. Prager, *Theory of Elastic Stability*, Second Edition. In *Journal of Applied Mechanics* (2nd ed., Vol. 29, Issue 1). McGraw-Hill Books Company, 1962, doi:10.1115/1.3636481.
- [8]. F. C. Onyeka, E. T. Okeke, "Analytical Solution of Thick Rectangular Plate with Clamped and Free Support Boundary Condition using Polynomial Shear Deformation Theory," *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 1, pp. 1427-1439, 2021, doi:10.25046/aj0601162.
- [9]. J. N. Reddy, *Classical Theory of Plates*, In *Theory and Analysis of Elastic Plates and Shells*, CRC Press, 2006, doi:10.1201/9780849384165-7.
- [10]. D. O. Onwuka, S. E. Iwuoha, "Elastic Instability Analysis Of Biaxially Compressed Flat Rectangular Isotropic All-Round Clamped (CCCC) plates," *MOJ Civil Eng*; vol. 2, no. 2, pp. 52-56, 2017, doi: 10.15406/mojce.2017.02.00027.
- [11]. C. C. Ike, "Kantorovich-Euler Lagrange-Galerkin's Method for Bending Analysis of thin plates," *Nigerian Journal of Technology (NIJOTECH)*, vol. 36, no. 2, pp. 351 - 360, 2017, doi: 10.4314/njt.v36i2.5.
- [12]. G. Kirchoff, "Über das Gleichgewicht und die Bewegung einer elastischen Scheibe. J. Reine Angew," *Math. (Crelle's J.)*, vol. 40, pp. 51-88, 1850.
- [13]. E. Reissner, "The Effect of Transverse Shear deformation on the Bending of Elastic Plates," *J. Appl. Mech. Trans. ASME*, vol. 12, no. 2, pp. A69-A77, 1945, doi: 10.1115/1.4009435.
- [14]. R. D. Mindlin, "Influence of Rotatory Inertia and Shear on Flexural Motions of Isotropic, Elastic Plates," *J. Appl. Mech. Trans. ASME*, vol. 18, no. 1, pp. 31-38, 1951, doi: 10.1115/1.4010217.
- [15]. F. C. Onyeka, B. O. Mama, "Analytical Study of Bending Characteristics of an Elastic Rectangular Plate using Direct Variational Energy Approach with Trigonometric Function," *Emerging Science Journal*, vol. 5, no. 6, pp. 916-928, 2021, doi:10.28991/esj-2021-01320.
- [16]. U. G. Eziefula, "Analysis of inelastic buckling of rectangular plates with a free edge using polynomial deflection functions," *International Review of Applied Sciences and Engineering*, vol. 11, no. 1, pp. 15-21, 2020, doi:10.1556/1848.2020.00003.
- [17]. I. Shufrin, M. Eisenberger, "Stability and vibration of shear deformable plates - First order and higher order analyses," *International Journal of Solids and Structures*, vol. 42, no. (3-4), pp. 1225-1251, 2005, doi:10.1016/j.jisolsr.2004.06.067.
- [18]. O. M. Ibearugbulem, J. C. Ezeh, U. C. Nwachukwu, "Application of Split - Deflection Method in Buckling Analysis of CCSS and CCCS Thin Rectangular Isotropic Plates under Vibration," *International Journal of Innovative Research and Advanced Studies (IJIRAS)*, vol. 6, no. 2, pp. 7-12, 2019.
- [19]. O. M. Ibearugbulem, "Using the product of two mutually perpendicular truncated polynomial series as shape function for rectangular plate analysis. *International Journal of Emerging Technologies and Engineering (IJETE)*, 2014 ICRTIET Conference proceeding.
- [20]. E. Ventsel, T. Krauthammer, *Thin plates and shells: theory, analysis and applications*, 2001, New York: Marcel Dekker.
- [21]. S. Uzoukwu, O. M. Ibearugbulem, C. E. Okere, J. I. Arimanwa, "Stability Analysis of Rectangular CCSS and CCCS Isotropic Plates using 3rd Order Energy Functional," *Global Scientific Journals*, vol. 9, no. 1, pp. 637-649, 2021.
- [22]. F. C. Onyeka, F. O. Okafor, H. N. Onah, "Buckling Solution of a Three-Dimensional Clamped Rectangular Thick Plate Using Direct Variational Method," *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, vol. 18, no. 3, pp. 10-22., 2021, doi: 10.9790/1684-803031022.
- [23]. V. T. Ibeabuchi, O. M. Ibearugbulem, C. Ezeah, O. O. Ugwu, "Elastic Buckling Analysis of Uniaxially Compressed CCCC Stiffened Isotropic Plates," *Int. J. of Applied Mechanics and Engineering*, vol. 25, no. 4, pp.84-95, 2020, doi: 10.2478/ijame-2020-0051.
- [24]. F. C. Onyeka, B. O. Mama, J. Wasiu, "An Analytical 3-D Modeling Technique of Non-Linear Buckling Behavior of an Axially Compressed Rectangular Plate. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, vol. 6, no. 1, pp. 91-101, 2022, doi: 10.47001/IRJIET/2022.601017
- [25]. F. C. Onyeka, B. O. Mama, C. D. Nwa-David, "Analytical Modelling of a Three-Dimensional (3D) Rectangular Plate Using the Exact Solution Approach," *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, vol. 11, no. 1, pp. 76-88, 2022, doi: 10.9790/1684-1901017688.
- [26]. A. Moslemi, B. N. Navayi, J. A. Vaseghi, "3-D Elasticity Buckling Solution for Simply Supported Thick Rectangular Plates using Displacement Potential Functions," *Applied Mathematical Modelling*, vol. 40, no. 11-12, pp. 5717-5730. 2016, doi:10.1016/j.apm.2015.12.034.
- [27]. F. C. Onyeka, T. E. Okeke, C. D. Nwa-David, "Stability Analysis of a Three-Dimensional Thick Rectangular Isotropic Plates with Arbitrary Clamped and Simply Supported Boundary Conditions," *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, vol. 19, no. 1, Ser. IV, pp. 01-09, 2022, doi: 10.9790/1684-1901040109.
- [28]. F. C. Onyeka, "Direct Analysis of Critical Lateral Load in a Thick Rectangular Plate using Refined Plate Theory," *International Journal of Civil Engineering and Technology*, vol. 10, no. 5, pp. 492-505, 2019.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Engr. Dr. F. C. Onyeka has done his bachelor's degree (B.Eng) in Civil Engineering from Anambra State University, Uli, Nigeria in 2006. He has done his master's degree (M.Eng) and doctorate degree (Ph.D) in Structural Engineering from University of Nigeria Nsukka in 2010 and 2018 respectively. His research area includes; Structural

Engineering Mechanics, Plates and Shell theory and Theory of Elasticity, Variation Calculus and Stability of structures.

He is a University lecturer and a renowned researcher in the area of Structural Engineering Mechanics. He has about 52 publications which includes journals and conferences both locally and internationally with awards in some research breakthrough.

He is a member of professional bodies which includes; corporate member of Nigeria Society of Engineers (NSE) and member Nigeria Institute of Civil Engineers (NICE), a registered engineer in Council for regulation of Engineering in Nigeria (COREN).



Engr. Okeke, Edozie Thompson obtained bachelor's degree (B.Eng) in civil engineering from Enugu state University of Science and Technology in 2006. He got master's degree (M.Eng) in structural engineering at University of Nigeria Nsukka, in 2014. He currently undergoing Ph.D program at the University of Nigeria Nsukka. He is a University lecturer and a renowned

researcher in the area of Structural Engineering Mechanics with about 10 publication which includes journals and conferences both locally and internationally. He is a member of professional bodies which includes; corporate member of Nigeria Society of Engineers (NSE) and member Nigeria Institute of Civil Engineers (NICE), a registered engineer in Council for regulation of Engineering in Nigeria (COREN).



Engr. Nwa-David, Chidobere David obtained B. Eng (First class honors' in Civil Engineering) from Michael Okpara University of Agriculture, Umudike, M. Eng (Structural Engineering) from Department of Civil Engineering, Federal University of Technology, Owerri. He is a University lecturer and a renowned researcher and his area research interests

cut across all areas of civil engineering with specialized focus on structural engineering particularly in dynamics of structures, elasticity theory of plates, structural mechanics, concrete materials, variational calculus, sustainable structural systems, sustainable construction materials and studies with soft computing techniques. He is a Registered Engineer, Council for the Regulation of Engineering in Nigerian (COREN). He has participated in conferences where he presented papers. He has several scholarly articles published in local and international journals.

A Review- Modelling Approach and Numerical Analysis of Additive Manufacturing

Vaishnavi Kohale*, Samidha Jawade, Ganesh Kakandikar

School of Mechanical Engineering, Dr. Vishwanath Karad MIT- World Peace University, Pune, Maharashtra, India

*Corresponding author: Vaishnavi Anil Kohale, Pune, 8623917417, vaishnavikohale1998@gmail.com

ABSTRACT: Additive manufacturing creates 3-dimensional objects by depositing materials layer by layer. Different applications of additive manufacturing were examined to determine future growth possibilities. The current research seeks to discover existing additive manufacturing techniques based on the process mechanisms, evaluate modelling approaches based on modelling methodologies, and identify required studies. A significant number of numerical simulations are conducted to evaluate the thermal FE structure in terms of solid and powder material thermo - physical properties and permissible boundary conditions. The transient heat conduction is investigated using thermal analysis with a moving heat source.

KEYWORDS: Additive manufacturing, Modelling approach, Numerical analysis, Finite Element Analysis, Thermal analysis

1. Introduction

Nowadays, industrial 3D printing is a slow and costly trial and error process. On the other hand, additive manufacturing is a group of manufacturing technologies that use a solid digital model to produce an item additively. The interaction between solid models and additive manufacturing, as well as material deposition, are entirely computer-controlled. The method, which is based on CAD models, can manufacture reliable yet complicated parts. Moreover, it pushes towards a tool-less manufacturing environment, which means better quality and efficiency in many circumstances. 3D printing, freeform fabrication, and additive fabrication terms are used for representation of additive manufacturing. Low-Vol. components with complicated forms and geometric characteristics have traditionally been produced using traditional additive manufacturing methods. However, SLM technology allows for creating geometries with complex features that are impossible to achieve using conventional production processes such as casting, powder metallurgy, forging, and extrusion.

1.1. Additive Manufacturing

The process of joining materials to make objects from three-dimensional model data, usually layer upon layer, instead of subtractive manufacturing methodologies i.e

traditional machining is defined as Additive manufacturing (AM) [1]. The AM manufactures component layer-by-layer deposition using a laser. Using computer-aided design (CAD), additive manufacturing has the freedom to create objects with complex geometric shapes. Each process will differ depending on the material and machine technology used. Many additive manufacturing processes are available. The processes differ in the manner in which layers are deposited to form a component in the working principle and the materials used. The additive manufacturing method emerged as the primary tool in rapid prototyping. Continuous filament of a material is used in Fused deposition modelling (FDM) [2]. In DED method for Inconel 718 material the unique Mechanical properties and energy storage capacity are improved with the application in nuclear energy, Oil & gas. For powder bed fusion manufacturing in 316L stainless steel material the improved properties are better radiation tolerance and lower IASCC susceptibility for application of nuclear energy, Oil & gas [3]. Because materials for AM systems are limited, research and development are continuing to broaden materials and the application of present metal AM processes to a wider spectrum of materials. Polymers, ceramics, and metals are among the materials that can be used in AM technology [4]. Researchers and companies are becoming

more interested in metallic materials among these materials.

1.2. Applications

- **Aerospace** – Laser-sintering meets commercial and military aircraft demands for air ducts, fittings, and mountings that carry special aeronautical equipment.
- **Manufacturing** – Laser sintering is a cost-effective way to service specialized markets with low Vol.s. As economies of scale do not affect laser sintering, manufacturers may focus about batch size optimization instead.
- **Medical** – Medical gadgets are high-value, complicated goods. They must precisely fulfil the needs of their customers. These criteria must be met not just because of the operator's personal preferences but also because of regionally differing legal requirements or conventions. As a result, there are several varieties and, as a result, tiny amounts of the variants available.
- **Prototyping** – Laser-sintering allows the creation of both design and functional prototypes. As a result, functional testing may begin immediately and with greater flexibility. Simultaneously, these prototypes may be utilized to assess consumer acceptability.
- **Tooling** – The natural method removes the need for tool-path generation as well as various machining techniques like EDM. Tool inserts can be made overnight or in a matter of hours. Furthermore, design flexibility may be used to improve tool performance, for as by including conformal cooling channels within the tool.

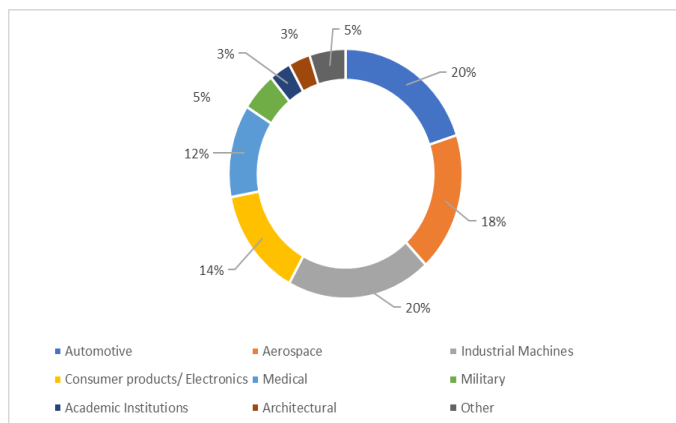


Figure 1: Industrial Adoption of additive manufacturing (AM) [3].

2. Modelling approach and Numerical analysis

An abstract description of a process that creates a relationship between input and output values is referred to as a model. Models that seek to predict the performance of the real system simulate it. Modeling involves making a 3D model data. Methods used for modelling and numerical analysis are:

2.1. Process modelling

The majority of technology process models are classified into one of three categories. Three models are possible: white box or simulation models - are built utilizing physical relations and engineering expertise to describe the process at the needed level, black box models - To describe a process, use data or knowledge collected from studies, grey box models - while still reasonably easy to solve, also include details than an equivalent black box model [5]. The interaction amongst the laser beam and even the powder material is perhaps the most typically modelled element in laser-based powder bed technology. Models at the powder scale are frequently used to determine the required laser input energy and can also be used to explore specific phenomena such as melt pool temperature histories or microstructures [5]. The figure 2 shows overview of approach.

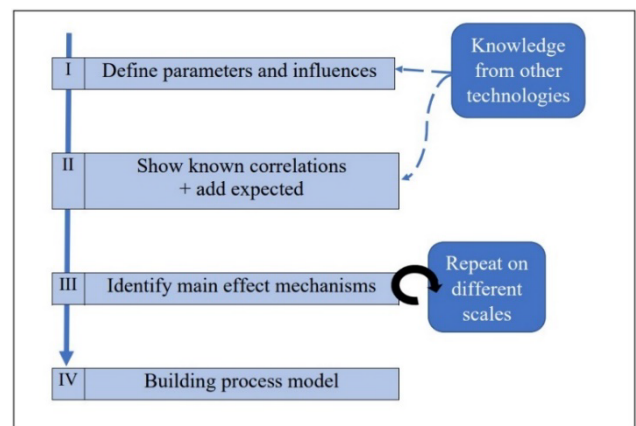


Figure 2: Overview of approach [5].

There are two primary additive manufacturing technologies: powder-bed and directed energy deposition. Material deposition is confined and happens concurrently with laser heat deposition in direct energy technology, whether powder fed or wire fed. Computational Fluid Dynamics (CFD) is used for non-linear partial differential equations for macroscale simulation of the solidification process [6]. In direct energy deposition, a continuum thermo-mechanical modelling tool is used to simulate the melt pool shape. Melt pool calculates the process's stability based on the powder layer thickness, scanning velocity, and optical and thermal characteristics of the material. Methods used are coupled radiation, heat transfer, consolidation kinetics, conventional radiation transfer equation [7]. For simulation of melt pool geometry finite element method, Discrete Element Method and fluid flow simulation and a 3-D transient heat are approached. During processing, finite-element analysis software i.e Finite-Element Analysis (FEA) is utilized to solve the coupled governing equations for solid deformation and heat transfer [6]. The proposed integrated process structure properties performance modelling and simulation methodology is presented in fig 3 [6].

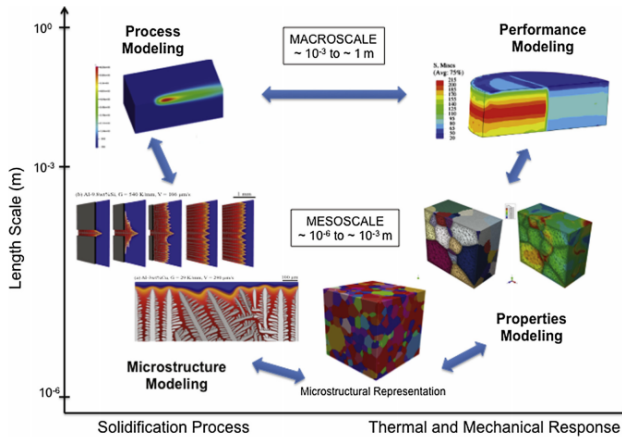


Figure 3: The proposed integrated process-structure-properties-performance modelling and simulation methodology, as well as the related length scales, are presented [6].

The temperature distribution in a single powder process is defined using a finite element-based heat transfer model. The geometrical characteristics of a single-track configuration have been quantitatively studied [8]. The created simulation model determines how process factors affect geometric features. Investigators have recently been drawn to wire-arc manufacturing processes (WAAM) in order to build metal parts with a higher deposition rate. A framework for component certification based on international standards have been developed using a computationally efficient mathematical model metho for the metal-AM process [9]. Manufacturing companies have embraced directed energy deposition technologies because of their capacity to build significantly bigger mechanical and structural components directly from a CAD model. 3D CAD model generation, slicing 3D CAD model to acquire set of 2D geometries, producing tool path for each one of these 2D geometries, selecting deposition parameters for every layer, selecting welding parameters such as travel speed, voltage, feed rate, and so on are all part of the process planning [10]. Finally, using the WAAM process and the supplied parameters and tool route, the products can be deposited. To build 2D geometries from a 3D model, there are two methods: unidirectional slicing and multidirectional slicing. Due to its ability to produce big and slightly less difficult geometric components, WAAM methods are conquering the industrial industries. The WAAM process is highly recommended for low and medium complicated geometry component manufacture. Mesh-based approaches and mesh-free methods are the two types of numerical modelling methods [9]. To avoid wasting time and money on hit and trial testing, the use of FEM has been extended to the models of WAAM processes and their parametric optimization. At the macro-scale level, WAAM processes use a multi-physics continuum modelling method. the applicability of its versions on an industrial scale, including the multi-physics continuum modelling method at the macroscale, i.e., a thermo-mechanical model to estimate residual

stress and distortion in WAAM-produced components. The numerical modelling aids in the understanding of the physics participating in the WAAM process, allowing it to be improved and changed in the field. The thermo-mechanical model's fundamental bifurcations are the thermal and mechanical models [9]. It is further characterized as coupled and decoupled, poorly coupled or uncoupled model based on the interconnection amongst two models.

2.2. Microstructure evolution modelling

Quantitative predictions of microstructures that are additively manufactured, and therefore their performance and properties, will necessitate collaborative work in solidification modelling at many lengths and time. For modelling microstructure evolution, the Potts Kinetic Monte Carlo technique in the simulation of melt pool geometry. EBSD data and the open-source tool DREAM3D is used to generate the microstructure representation [6].

2.3. Performance modelling

For representation of accurate physical quantities recording and tracking the evolution of metallic materials, physically-based macro-scale continuum models are created. There might be a variety of physical factors at issue, any of which could be contributing to this variation in behaviour. Finite-Element Analysis software (FEA) like Diablo and ABAQUS, are used to solve the governing equation that represents the physical mechanisms in performance modelling [6].

2.4. Topology optimization

Different types of tools are used for topology optimization. A recently used tool for the detection of shapes is the PLATO tool. However, the development of optimization software has focused chiefly on geometric optimization which can be simulated in ANSYS. The software will include numerous constraints meant for processing parameter to achieve optimal material performance as well as optimized topology in future [6].

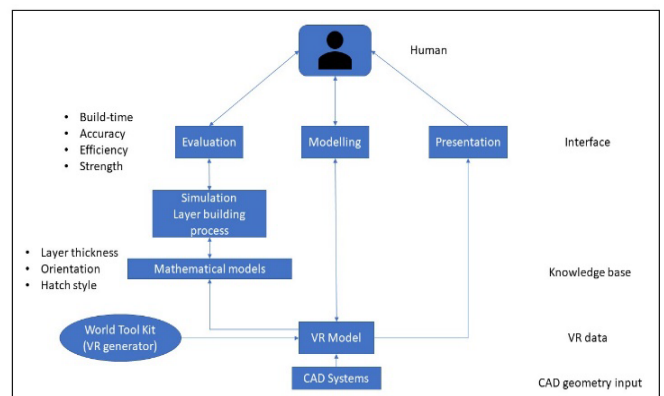


Figure 4: Virtual system of rapid prototyping [11].

The optimization of rapid prototyping is estimated using various processes such as virtual prototyping and virtual reality, virtual simulation, virtual fabrication, modelling module, simulation module, evaluation module. Fig shows the virtual system representation of rapid prototyping. In addition, process parameters such as nuisance, constant and control parameters, hatch space, orientation, layer thickness, overcure depth, build time, hatch style is considered [11].

2.5. Multiphysical modelling

For simulation of selective laser sintering process as shown in fig 3 for single layer of particles, a multiphysical modelling technique is used. Mechanical and thermal interactions of particle-to-wall and particle-to-particle are studied using a discrete element method by Beer-Lambert Law. Powdered particles to be sintered are represented by individual spheres of various sizes in a discrete element model. The technique used to solve and analyze the Runge-Kutta scheme. The process parameters used for simulation are Particle Diameter, Domain Cross-Sectional dimensions, Number of Particles, Initial Powder Temperature, Scan Speed, Laser Power, Particle Density, Laser Beam Diameter, Melting Temperature, Vaporization Temperature, Thermal Conductivity, Heat Capacity [12].

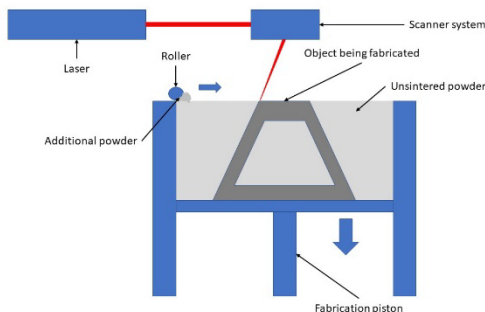


Figure 5: Schematic of a SLS set-up [12].

The surface-tracking approach used in this class of CFD models allows for explicit depiction of individual powder particles, which is particularly useful for the L-PBF process. In this case, unlike earlier models, the material properties of the powder layer will be similar to that of the bulk stuff. In their CFD simulation, they apply the Lattice Boltzmann Methodology (LBM). ALE3D is used to create a FEM-based CFD model for the L-PBF process of 316-L stainless steel (Arbitrary Lagrangian Eulerian code) [13]. CFD model based on Finite Vol. Method FVM for IN718's and aluminium L-PBF process is also developed. The influence of linear energy density here on creation of porosity in IN718 during L-PBF using the FVM model and Marangoni effect with recoil is also implemented [13]. There's also a subset of multiphysics models that don't use a surface-tracking method and then use a Lagrangian explanation of such melt pool surface

layer. These models can forecast the creation of porosity and trace the metallurgical evolution throughout the process. Recent multi-physics models, on the other hand, have a number of drawbacks.

2.6. Multi-scale modelling approach

In the L-PBF process the Build-up model, single laser track model, multi-scale approach, single layer model is used to study the various process parameters, melt pool dimension, temperatures through layer solidification applied heat input load parameters [14]. In addition, the method is used to calculate the process related to thermally induced distortions.

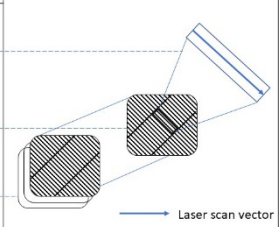
Dimension	Type of Simulation	Illustration
nm to μm	Single laser track simulation	
μm to mm	Single layer simulation	
mm to cm	Build-up simulation	

Figure 6: Simulation approaches on different scales [14].

To accomplish quick distortion prediction for MAM parts, a multi-scale simulations framework was created. Three models at various scales are included in the framework. First, a micro-scale heat source modelling is used to calibrate the heat input. The heat input is then used to determine the intrinsic stresses using a meso-scale hatching model. The inherent strains are then used to a macro-scale layer model to anticipate the part's distortion. Multi-scaling methods- AG (agglomerated heat source), MS (multi-step method), FH, SFH (flash heating and sequential flash heating), AMR (adaptive mesh refinement) [13]. Flash heating (FH) is a multi-scaling approach for parts that is focused on the layer-lumping principle. A section is split into a particular number of pieces is along build direction in this manner. These chunks are actually a collection of actual layers that have been fused together and are referred to as meta-layers [13]. The meta-layer size is a model input variable that can be changed.

2.7. Hole-drilling method

In the L-PBF process, the Integral Method determines the residual stresses in the sample thickness. In this process the criteria used to choose the dimensions and geometries of specimens are stated as follows: Using traditional and easily measurable geometry, closed and open sections, thin-walled sections, flat surfaces, alignment along the normal or in the building direction, surfaces big enough to contain the strain gauges for the testing without suffering from edge effects induced by sharp edges [15]. Then, MATLAB estimates the residual stresses measurements. The advancement of the

specimen's stress profile is shown as a function of specimen geometry, in terms of open/closed sections and orientations.



Figure 7: MTS3000 RESTAN (Residual Stress Analyzer) system [15].

2.8. X-ray (EDX) mapping

The SLM technique includes scanning a powder bed with a laser beam to manufacture layer by layer. The microstructure and nano-mechanical properties of the material is studied. Then, the changes in scan speed influence the development of fusion lines and single tracks. Energy dispersive X-ray (EDX) mapping is used to compare the SLM material's chemical composition distribution [16]. To describe the mechanical characteristics of SLM-processed materials and assess the impact of process-induced defects heat treatment method is used. The material is processed under the SLM machine and AutoFab software. With increasing laser scan speed, the diameters of SLM-formed lines and tracks decreased linearly. Furthermore, abnormalities are identified at high scan rates [16].

2.9. CFD simulation

The high-speed imaging and CFD simulation are used for systematic parametric study to investigate the influence of laser scanning speed and powder layer thickness on porosity development and correlate porosity development with top sample surface and melt pool and flow behaviour. High-speed imagery and computational fluid dynamics (CFD) calculations were used to investigate the interaction between the powder particles and laser beam. C++ open-source the CFD toolbox called Open Field Operation and Manipulation, a simulation of the interaction between the laser heat source and the powder material is carried out [17].

2.10. Finite element modelling

To be more precise, statistical analysis and machine learning both require a large amount of data. Machine

learning process parameters - Artificial neural network, Genetic Algorithm, Ensemble, Support Vector Regression. The statistical analysis method, that is Taguchi, ANOVA and regression modelling, two-level factorial design of experiment by Minitab software [18]. The mechanical performance of selective laser melting manufactured components is essential. The yield stress is the key characteristic, and it the main factor for the SLM manufactured component. Here the processing parameters such as laser power, scanning speed, and hatch space of an SLM process can be investigated [18].

During the Additive Manufacturing (AM) of metal parts, part distortion is a major concern. Finite element Modelling is used by AM for powder bed fusion manufacturing. The inherent strain technique is a quick and accurate approach for predicting residual stresses and deformation. The inherent strain method's origins may be traced back to Computational Welding Mechanics, and it's been widely adopted. It comprises of an FE quasi-static evaluation with user-defined inherent strains causing deformation [19]. To simulate the deformation of a twin-cantilever beam with different scanning tactics, a 3D simple mechanical model is created by using commercial software ABAQUS. The modelling methodology, namely layer lumping, allows for the usage of a rather coarse FE mesh. Two methods are applied for calculation of deformation purpose: Reduced order methods and Empirical methods [19]. An empirical methodology can be specified for determining characteristic inherent strains for just a specific scanning strategy, which are used as input data for linear-elastic analysis to derive distortion and residual stress fields generated by LPBF processes.

2.11. Mathematical modelling

In the RAM process the Deposition Principle and Establishment of Temperature Field Equation is required for the numerical simulation. The Gaussian distribution is analogous to the spatial accumulation distribution of mass flow [20]. By integrating the external mass supply element to the mass conservation equation, the mass growth process can be accomplished. The temperature field of a three-dimensional model must be generated using the Fourier law of conduction of heat and the principle of conservation of energy, and heat transfer problem's governing equation must be established. The single-channel melting layer model using ANSYS is made and meshed. The temperature change of the molten layer under diverse currents was studied using a RAM simulation model, and indeed the present parameter range of the molten wire is preliminarily determined [20]. The actual physical events that occur during the AM process must be reduced to make the process available for numerical simulation. A Newtonian fluid present in melt pool is laminar and incompressible. Powder size follows a

sphere-shaped Gaussian distribution. The flow at the solid-liquid interface is called mushy zones, regarded as a porous medium's isotropic permeability. A Boussinesq approximation is used to evaluate the density fluctuation of the molten pool in the momentum equation for the buoyancy term.

- a. Gaussian packing
- b. Laser source and laser absorptivity
- c. Governing equations
- d. Boundary conditions
- e. Material properties and numerical simulation process

2.12. Heat transfer analysis

In the numerical modelling for the heat transfer analysis of AM operations using powder-bed method FE framework is utilized. To cope with the sintering process, that converts the metal powder into a new solid part, the formulation is reinforced with an appropriate FE activation approach. a good balance of computational effort and accuracy is achieved by simplified hatch-by-hatch patterns.

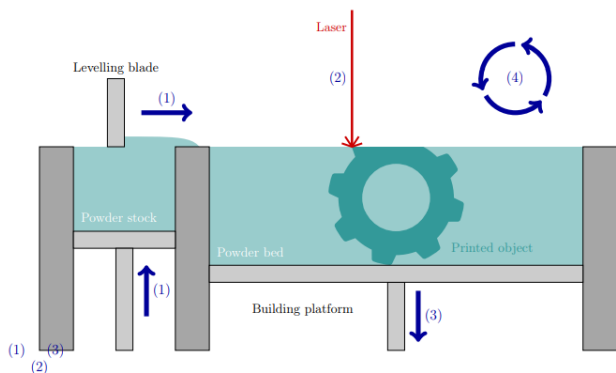


Figure 8: Printing process by DMLS. (1) Levelling blade is used to spread over new layer of powder over the building platform. (2) the area of powder that corresponds to the object's present cross section melts by laser source. (3) The building platform is lowered to take in a new layer. (4) 1 and 3 steps are performed repeatedly until the whole model is formed. (5) During post-processing, unfused powder which is loose is eliminated [21].

The governing equation for describing the temperature evolution of the AM process during the printing and cooling stages is the equation of balance of energy,

$$H = -\nabla \cdot q + 'r, \text{ in } \Omega, t > 0 \dots \dots \dots [21]$$

The temperature and the latent heat rate released or absorbed during the phase-change process are used to calculate the enthalpy rate per unit Vol.

- i. Boundary conditions.

The boundary conditions in the partition of the margins where,

$\partial\Omega_c$ = surface contact with the printing platform,

$\partial\Omega_p$ = surface contact with the powder-bed,

$\partial\Omega_e$ = external surface in contact with the surrounding environment [11]

- Heat conduction through the building platform

$$T = T_c, \text{ for } \partial\Omega_c$$

where T_c is the temperature of the building platform.

- Heat conduction in powder bed.

The density and the specific heat are straightforwardly computed as

$$\rho_p = \rho_{solid}(1 - \varphi) \dots \dots [21]$$

$$C_p = C_{solid}$$

for low porosity powders,

$$k_p = (6.3 + 22\sqrt{0.09k_{solid} - 0.016}) \frac{k_{solid}(1-\varphi)}{\left(\frac{k_{solid}}{k_{gas}}\right) - (10^{0.523-0.594\varphi}) - 1}$$

When the powder bed is excluded from the computational domain,

$$q_p(T) = h_p(T - T_p), \text{ on } \partial\Omega_p,$$

The numerical model provides for power input and absorption, heat dissipation across boundaries via conduction, convection, and radiation and temperature dependence of material properties.

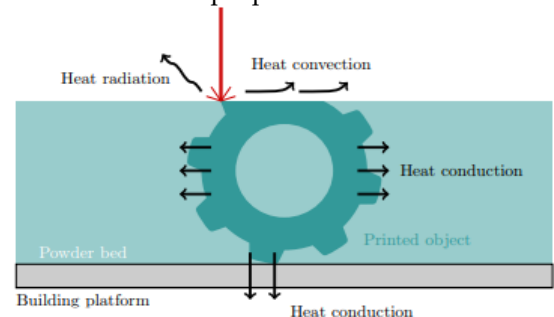


Figure 9: detailed image of figure gathering the boundary conditions of the problem: (1) Conduction of heat via building platform (2) Conduction of heat via powder bed (3) Conduction of heat and radiation of heat via free surface [21].

- Heat convection through the surrounding environment.

$$q_{conv}(T) = h_{conv}(T - T_e), \text{ on } \partial\Omega_e$$

- Heat radiation.

$$q_{rad}(T) = \sigma (T - T_e), \text{ on } \partial\Omega_e$$

For heat radiation and convection,

$$q_{loss}(T) = h_{loss}(T - T_e), \text{ on } \partial\Omega_e$$

2.13. Thermo-mechanical model

The SLM manufacturing process is numerically analysed utilizing a 3D thermo-mechanical modelling method based on finite element coupling. During the SLM process,

this method allows for the assessment of both the mechanical stress distribution and the transient temperature.

- Heat transfer modelling

The first law of thermodynamics is used to develop the differential equation controlling transient heat conduction inside a continuous medium of arbitrary Vol. [22].

The heat equation gives the temperature with respect to time,

$$k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + q = \rho c_p \frac{\partial T}{\partial t}$$

where,

k = thermal conductivity coefficient,

ρ = mass density,

c_p = specific heat,

q = power generated per Vol. in the workpiece.

A hemispherical shape heat source model's power density distribution is represented by,

$$\dot{q} = \frac{2^{5/2} \beta P}{\pi^{3/2} r_0^3} \exp \left\{ -2 \frac{x^2 + y^2 + z^2}{r_0^2} \right\}$$

where,

P = power of the laser source,

β = absorptivity of the laser beam

r_0 = radius of the laser beam.

- Mechanical modelling

The decoupled or poorly coupled approach is commonly employed for AM process modelling since it requires less computational effort and time than the fully coupled method. The mechanical or structure analysis stage, however, dominates the analysis time in decoupled modelling. Since the thermal gradient near the thermal source is extremely strong, the mesh size must be adequate to capture the high graded residual stress and distortion of the diffusion layer and many layers beneath it: in the heat affected zone (HAZ) [23]. As a result, when a fine mesh is used in the model, the computational time increases. To reduce the computational time of DMD, a Finite Element-based mesh coarsening technique is developed in both thermal and mechanical analyses. When compared to a traditional analysis process that used just fine meshes with no coarsening, the computing time for the coarsening method is reduced by around three times, and the findings are comparable. The ABAQUS solution map technique is used for meshing purpose. In two steps, the adaptable meshing technique is used: fine

mapping and coarsening steps which utilizes the concept of coherent strain. The model can be separated into three separate scales (micro, meso, and macro), the mechanical layer equivalent can be achieved [23]. With a true thermal heat model structure and thermal boundary conditions, a pure thermal analysis is done on a very small structure that is connected to a substrate that replicates the lower layers from the real scale. Many AM factors that have a huge consequence on the mechanical properties of the parts produced demand more inventive FE-based modelling framework.

- Quasi-static mechanical analysis

It is carried out to obtain mechanical response of the workpiece.

The balance of linear momentum in any point of the body

$$\text{div}(\sigma) + b = 0$$

where,

σ = stress tensor

b = body forces,

The sum of the following terms yields the total strain increment:

$$\Delta \epsilon^{total} = \Delta \epsilon^e + \Delta \epsilon^p + \Delta \epsilon^{th}$$

where,

$\Delta \epsilon^e$ = elastic strain increment,

$\Delta \epsilon^p$ = plastic strain increment,

$\Delta \epsilon^{th}$ = thermal strain increment.

the resulting stress from the elastic strain.

$$\sigma^e = C : \epsilon^e$$

where,

C = fourth-order material stiffness tensor (Elastic moduli).

For the associated flow rule the plastic strain increment is,

$$\Delta \epsilon^p = \lambda \frac{\partial f}{\partial \sigma}$$

where,

λ = plastic multiplier calculated through the consistency condition

The total thermal strain is calculated,

$$\epsilon^{th} = \left(\alpha_T (T - T_{ref}) - \alpha_{ini} (T_{ini} - T_{ref}) \right) I$$

where,

α_T and α_{ini} = Vol. tric thermal expansion coefficients evaluated

T = current temperature

T_{ini} = initial temperature respectively

T_{ref} = reference temperature for the thermal expansion coefficients

I = second-order identity tensor

Numerical simulation can efficiently determine thermal development, molten pool shape, residual stress, and deformation. The model combines heat transfer and fluid dynamics to track the boundaries of melt pools.

The precise calculation of nodal temperature data, dispersion of residual stresses & deformation of components manufactured by AM, as well as large computational efforts, are the main issues in AM modelling. Innovative solutions are proposed and applied to address these challenges in metal alloy AM techniques, as well as modelling methodologies to increase the process' efficiency and accuracy. residual stress and deformation evaluation of Direct metal deposition [23]. As the deposition progresses and multiple layers are created, the thermal expansion of layers under the heat source causes compressive pressures to be applied to the lower layers, resulting in compressive plastic strains. Because of the nature of the operation, stress residual induction is unavoidable in AM procedures. To analyse material deposition onto the substrate and the generation of residual stresses & geometric distortions in AM parts, the aforementioned models often use simplifying assumptions or employ particular methodologies.

The modelling of the AM process may be separated into two categories: (1) doing a pure thermal or thermal transfer analysis in order to acquire the nodal temperatures in the FE based model, and (2) building a structural configuration in order to evaluate mechanical behavior of the FE based model with applied nodal temperature gradients, and lastly, obtaining the deformation and residual stresses of the produced component [23]. The approach is known as coupled thermomechanical analysis when the thermal study is followed by a structural assessment for each increment. The approach is known as coupled thermomechanical analysis when the thermal study is followed by a structural assessment for each increment.

The heat transfer analysis is built on the idea of the body's energy conservation. The residual stresses and distortions are estimated by applying boundary conditions to the mechanical FE model. Properties of the material should be considered temperature dependent in both the thermal or mechanical analyses to provide a

more realistic representation of the process. The application of incremental material to the substrate necessitates the use of numerical analytical tools. The most well-known strategies are (a) silent elements activation, (b) inactive parts activation, and (c) hybrid elements activation [23].

2.14. Thermal modelling

The precision of the temperature history data produced through thermal analysis in the FE based decoupling approach for a model is critical for precise estimate of a residual stress distribution & deformation of an AM part. The thermal study of AM processes has been the subject of numerous studies. Various features and improvements in the modelling of thermal analysis with AM processes is given. The heat flux model, which represents the thermal source, is an important part of the thermal analysis. The laser/electron beam characteristics, such as power, speed, orientation, shape, and efficiency, must be included in the thermal source model. 3D super Gaussian, 3D Gaussian and 3D inverse Gaussian beams are three different forms of thermal body heat flux distribution [23]. Degrading the overall heat transfer phenomenon within the melt pool is one of the simple suppositions in the thermal evaluation of the AM process, which can lead to an overestimation of a nodal temperature history and, as a result, an overestimation in residual stresses. During an AM process, the shape and parameters of the melting pool have always been a critical component in determining nodal temperature, as well as deformation and residual stress measurement [23].

2.15. Interactive CAD modelling

The production of 3D patterns that may be tailored to the CAD model using generative algorithms in parametric modelling. GD has spread to a variety of industries, including architecture, jewellery, and industrial design. GD's parametric modelling enables for the automated production of any project piece based on parameters. This means that certain algorithm-generated rule sets control the development and change of pieces inside a project [24]. Elements are drawn automatically based on user-defined algorithms, and parameters inside the algorithm can be changed. Using generative algorithms, it is possible to identify answers to difficulties that can occur with traditional CAD systems. In inclusion to reducing the structure, generative algorithms are used to simulate non-structural aspects. For the production of 3D patterns and Voronoi tessellations, two algorithms have been implemented [24]. The first allows to create patterns on complex surfaces, while the second helps in creating a Voronoi tessellation. The new methodologies enable for the modelling and modification of non-structural components, allowing for an interactive aesthetic assessment of the geometries formed [24]. Only the

structural parts are subjected to FEM analysis. To overcome any non-convergence issues caused by high displacements during flexion and extension movements, explicit dynamic simulations are conducted [24].

3. Conclusion

A survey and evaluation of modelling techniques is presented in this work. The classification of AM methods is based on their working principle rather than the materials utilized. Following that, modelling processes in the field of additive manufacturing were presented and classified, not only based on the operating principle but also on the simulated process characteristic and the modelling technique. The most commonly used way for handling the challenge of dimensional accuracy is through empirical models and statistical approaches (ANOVA etc.). Mechanical characteristics and dimensional stability are often modelled using numerical heat transfer models, which focus on the melt pool and material phase transition, whereas build time has been studied both analytically and numerically. New features in the CAD modelling industry that can help solve challenges related to modelling in Additive Manufacturing.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Pinar Zorlutuna, Nasim Annabi, Gulden Camci-Unal, Mehdi Nikkha, "Microfabricated Biomaterials for Engineering 3D Tissues", *Advanced materials*, Vol. 24, Issue 14, pp. 1782-1804, 2012, doi:<https://doi.org/10.1002/adma.201104631>.
- [2] Panagiotis Stavropoulos, Panagis Foteinopoulos, "Modelling of additive manufacturing processes: a review and classification", *Manufacturing Rev*, Vol. 5, Issue 2, pp. 1- 26, 2018, doi:<https://doi.org/10.1051/mfreview/2017014>.
- [3] Cheng Sun, Yun Wang, Michael D. Mc Murtrey, Nathan D. Jerred, Frank Liou, Ju Li, "Additive manufacturing for energy: A review", *Applied Energy*, Vol. 282, pp. 1- 18, 2021, doi:<https://doi.org/10.1016/j.apenergy.2020.116041>.
- [4] Ana Vafadar, Ferdinando Guzzomi, Alexander Rassau, Kevin Hayward, "Advances in Metal Additive Manufacturing: A Review of Common Processes, Industrial Applications, and Current Challenges", *Appl. Sciences*, Vol. 11, pp. 1- 26, 2021, doi:<https://doi.org/10.3390/app11031213>.
- [5] Sarah Müller, Engelbert Westkämper, "Modelling of Production Processes: Theoretical Approach to Additive Manufacturing", *Procedia CIRP*, Vol. 72, pp. 1524-1529, 2018, doi:<https://doi.org/10.1016/j.procir.2018.03.010>.
- [6] M.M. Francois, Wayne King et al., "Modelling of additive manufacturing processes for metals: Challenges and opportunities", *Current Opinion in Solid State and Materials Science*, Vol. 21, Issue 4, pp. 198-206, 2017, doi:<https://doi.org/10.1016/j.cossms.2016.12.001>.
- [7] N.T. Aboulkhair et al., "On the formation of AlSi10Mg single tracks and layers in selective laser melting: Microstructure and nano-mechanical properties", *Journal of Materials Processing Technology*, Vol. 230, pp. 88-98, 2016, doi:<https://doi.org/10.1016/j.jmatprotec.2015.11.016>.
- [8] A.V. Gusarov, I. Smurov, "Modeling the interaction of laser radiation with powder bed at selective laser melting", *Physics Procedia*, Vol. 5, pp. 381-394, 2010, doi:<https://doi.org/10.1016/j.phpro.2010.08.065>.
- [9] Shekhar Srivastava, Rajiv Kumar Garg, Vishal S. Sharma, Anish Sachdeva, "Measurement and Mitigation of Residual Stress in Wire Arc Additive Manufacturing: A Review of Macro Scale Continuum Modelling Approach", *Archives of Computational Methods in Engineering*, pp. 3491-3515, 2020, doi:<https://doi.org/10.1007/s11831-020-09511-4>.
- [10] K.E.K. Vimal, M. Naveen Srinivas, Sonu Raja, "Wire arc additive manufacturing of aluminium alloys: A review", *Materials Today: Proceedings*, pp. 1139-1145, 2020, doi:<https://doi.org/10.1016/j.matpr.2020.09.153>.
- [11] S.H. Choi, S. Samadevam, University of Hong Kong, "Modelling and optimization of rapid prototyping", *Computers in Industry*, Vol. 47, Issue 1, pp. 39-53, 2002, doi:[https://doi.org/10.1016/S0166-3615\(01\)00140-3](https://doi.org/10.1016/S0166-3615(01)00140-3).
- [12] Rishi Ganeriwala, Tarek I. Zohdi, "Multiphysics modeling and simulation of selective laser sintering manufacturing processes", *Procedia CIRP*, Vol. 14, pp. 299-304, 2014, doi:<https://doi.org/10.1016/j.procir.2014.03.015>.
- [13] Mohamad Bayat, Wen Dong, Jesper Thorborg, Albert C. To, Jesper H. Hattel, "A review of multi-scale and multi-physics simulations of metal additive manufacturing processes with focus on modeling strategies", *Additive Manufacturing*, Vol. 47, pp. 1-25, 2021, doi:<https://doi.org/10.1016/j.addma.2021.102278>.
- [14] Chritian Seidal, Michael F. Zaeh, "Multi-scale modelling approach for contributing to reduced distortion in parts made by laser-based powder bed fusion", *Procedia CIRP*, Vol. 67, pp. 197-202, 2018, doi:<https://doi.org/10.1016/j.procir.2017.12.199>.
- [15] Alessandro Salmi et al., "On the effect of part orientation on stress distribution in AlSi10Mg specimens fabricated by laser powder bed fusion (L-PBF)", *Procedia CIRP*, Vol. 67, pp.191-196, 2018, doi:<https://doi.org/10.1016/j.procir.2017.12.198>.
- [16] Kurian Antony, N. Arivazhagan, K. Senthilkumaran, "Numerical and experimental investigations on laser melting of stainless steel 316L metal powders", *Journal of Manufacturing Processes*, Vol. 16, Issue 3, pp. 345-355, 2014, doi:<https://doi.org/10.1016/j.jmapro.2014.04.001>.
- [17] Chunlei Qiu, Chinnapat Panwisawa, "On the role of melt flow into the surface structure and porosity development during selective laser melting", *Acta Materialia*, Vol. 96, Pages 72-79, 2015, doi:<https://doi.org/10.1016/j.actamat.2015.06.004>.
- [18] Ivanna Baturynska et al., "Optimization of process parameters for powder bed fusion additive manufacturing by combination of machine learning and finite element method: A conceptual framework", *Procedia CIRP*, Vol. 67, pp. 227-232, 2018, doi:<https://doi.org/10.1016/j.procir.2017.12.204>.
- [19] Iñaki Setien, Michele Chiumenti, Sjoerd van der Veen, Maria San Sebastian, Fermín Garcíandía, Alberto Echeverría, "Empirical methodology to determine inherent strains in additive manufacturing", *Computers and Mathematics with Applications*, Vol. 78, Issue 7, pp. 2282-2295, 2019, doi:<https://doi.org/10.1016/j.camwa.2018.05.015>.
- [20] Suli Li, Kaiyue Ma, Chao Xu, Laixia Yang, Bingheng Lu, "Numerical Analysis and Experimental Verification of Resistance Additive Manufacturing", *Crystals*, Vol. 12, Issue 2, pp. 1-32, 2022, doi:<https://doi.org/10.3390/cryst12020193>.
- [21] Michele Chiumenti, Eric Neiva, et al., "Numerical modelling and experimental validation in selective laser melting" *Additive Manufacturing*, Vol. 18, pp. 171-185, 2017, doi:<https://doi.org/10.1016/j.addma.2017.09.002>.
- [22] Bruno M. Marques et al., "Numerical Analysis of Residual Stresses in Parts Produced by Selective Laser Melting Process", *Procedia Manufacturing*, Vol. 47, pp. 1170-1177, 2020, doi:<https://doi.org/10.1016/j.promfg.2020.04.167>.
- [23] Farshid Hajjalizadeh, Ayhan Ince, "Short review on modeling approaches for metal additive manufacturing process", *Material design and processing communications*, Vol. 2, Issue 2, pp. 1-7, 2020, doi:<https://doi.org/10.1002/mdp2.56>.

- [24] V. Ricottal, R. I. Campbell, T. Ingrassia, V. Nigrelli, "A new design approach for customised medical devices realized by additive manufacturing", *International Journal on Interactive Design and Manufacturing*, Vol. 14, pp. 1171–1178, 2020, doi:<https://doi.org/10.1007/s12008-020-00705-5>.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



VAISHNAVI KOHALE has completed her bachelor's degree in Mechanical Engineering from Sinhgad Academy of Engineering, Kondhwa, Maharashtra. Now currently pursuing her Master's degree in CAD/CAM/CAE from MIT World Peace University from, Pune, Maharashtra. Her current research includes additive manufacturing.



Prof. SAMIDHA JAWADE got her master's degree in process metallurgy and is currently pursuing her Ph.D. in mechanical engineering from MIT World Peace University in Pune, Maharashtra. She is currently employed as an Assistant Professor at the MIT World Peace University of Dr. Vishwanath Karad. Composite materials and additive manufacturing are two of her current study topics.



Dr. GANESH M. KAKANDIKAR is Professor Associate Head of School of Mechanical Engineering at Dr. Vishwanath Karad MIT World Peace University, Pune. He has completed Ph.D. in mechanical engineering from Swami Ramanand Teerth Marathwada University, Nanded in 2014.

He has 21 years of experience in teaching, research and administration. He has authored 03 books published internationally by Lambert Academic and CRC Press, Taylor and Francis Group and contributed book chapters published by Wiley, Springer, Elsevier. 65+ publications are at his credit in National/International Journals and conferences. He is reviewer of many reputed journals.

Micro Forming and its Application: A Critical Review

Neha Tiwari *, Ganesh Kakandikar, Omkar Kulkarni

School of Mechanical Engineering, Dr. Vishwanath Karad MIT-World Peace University, Pune, Maharashtra India.

*Corresponding author: Neha Tiwari, Jalgaon, 7841812083, tssnehatiwari173@gmail.com

ABSTRACT: In terms of manufacturing methods/processes, micro-manufacturing has received a lot of attention around the world. Micro-forming is one of the most widely used micro-manufacturing techniques. The micro forming is based on the properties of materials based on the process of shaping parts and object by mechanical deformation. Many efforts had been focused on micro-forming, in particular the deep drawing process, because of the method's ability to produce an extensive variety of products, particularly in its conventional macro-process. This method is used to create the majority of everyday items. Although efforts were made to develop micro-forming for industrial use, the technique was deemed to be insufficiently advanced. Much development effort was required, in particular, to design a completely computerized high-extent production micro-forming machine that is dependable and ready to perform always in terms of procedures, material handling, and tooling to assure effective micro-product production. Micro forming, which is discussed in this work, is also one of the often-used micro-forming methods in deforming procedures. Finally, in addition to continuing to improve micro forming, this study aims to investigate the essential methods of the Limit dome height test, Nakajima test, M K Model test, and deep drawing processes and their major concerns in a systematic manner.

KEYWORDS: Deep drawing, Forming Limit Diagram (FLD), Limit dome height test (LDH), Micro-forming, Nakajima test

1. Introduction

Metal forming, in addition to being a substantial industry in and of itself, is a pillar of modern manufacturing. Every year, hundreds of millions of tons of metal pass through metal forming processes all over the world. Metal forming accounts for 15-20% of the GDP of most industrialized countries. It also serves a social purpose by giving employment to millions of people. Bulk manufacturing of semi-finished and finished goods is a metal forming industry, which is one of the reasons why large-scale research and development advantages are feasible because even minor per-ton savings mount as much as sums.

Surface strain measurements were also used to determine the Forming Limit Diagram. For surface strain measurement, a circular grid on foil is required [1]. The forming limit diagram (FLD) illustrates metal formability in low-cost metal forming [2]. FLDs can be created by hemispheric punch-stretch tests and Marciniak cup testing. These experimental procedures required a significant amount of effort and time. To improve the

efficiency of FLDs, some analytical methods are applied. Analytical methods, however, are unable to predict failure. A forming limit diagram (FLD) is a graph that shows the major strains (ϵ_1) at the onset of localized necking for all values of the minor strain (ϵ_2) [3].

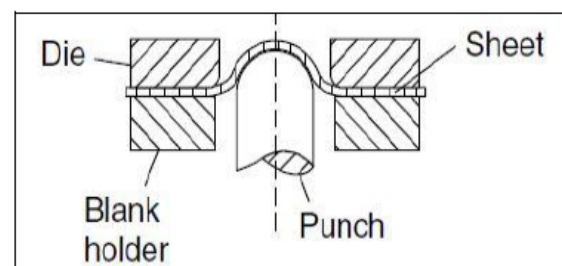


Figure 1: Sheet metal forming [4]

Plastic deformation is used to create product shapes in the metal forming process. As a result, for enhancing methods studying the plastic flow properties of alloys and metals is critical. The power and constraints of plastic deformation in forming also affect the factor properties. Raw materials are supplied for other techniques by many forming techniques, which manufacture finished or semi-

finished items. Steel facilities, for example, manufacture sheet metal that is utilized in automobiles and their parts. A variety of manufacturers used sheet metal material to create a wide range of residential and commercial products. Similarly, re-rolling mills use billets produced by steel manufacturers to roll into components such as angles, channels, and bars.

The formability analysis will be discussed, as well as its advantages. As well as the variables that affect sheet metal formability. Blank temperature, blank shape, punch velocity, Punch nose radius, blank holding force (BHF), die arc radius and spring back are some of the parameters.

1.1 Formability

The formability of the material is defined as the given ability of a sheet metal workpiece to go through plastic deformation to a provided shape without damage. The sheet metal forming methods are deep drawing and bending, flaws must be studied separately. The mechanics of the forming process determine the difference between these types of stamping techniques. The normal defects of the produced parts are presented for deep drawing.

1.1.1. Types of Sheet metal forming process

1.1.1.1. Limit dome height test (LDH)

The LDH method is the combination of both forming limit diagrams and simulated tests. LDH is the out-of-plane deformation method and is performed in different strain paths like uniaxial, plane strain, and biaxial [5]. Ghosh aimed to describe the height of the component as a function of the lowest strains that rectangular Nakajima specimens stretched on a hemispherical punch till fracture is based on Drewes' observation [6]. A diagram is presented below, which was created by sketching an arc around the result of experimental points recorded with samples of various widths. The approach was later adapted by foreign researchers as the strip stretch test, and by USA research scholars as to the Limiting Height Test.

A formability index defined by LDH is the height of the part after planar strain. As compared to the heights achieved in other stages of tension, this is the lowest. Material property is the width of the specimen about plane strain. Brings lots of advantages, the method is rarely used in industry because of its wide range of LDH values and the considerable quantity of testing required.

1.1.1.2. Deep Drawing process

A sheet metal forming technique is known as a deep drawing that involves mechanically dragging a sheet metal blank radially into a forming die [2]. As a result, it's a shape transition with material retention. When the

sketched part depth exceeds its diameter, the procedure is known as deep drawing. This can be performed by redrawing the part with a series of dies. Deep drawing is a method for making cup-shaped objects, pressure vessels, and gas cylinders, among many other things [8]. The deep drawing test is also known as the micro deep drawing test (MDD). The MDD toolset was controlled by PLC modules in the control and data logging box, which have driven by the press machine [9].

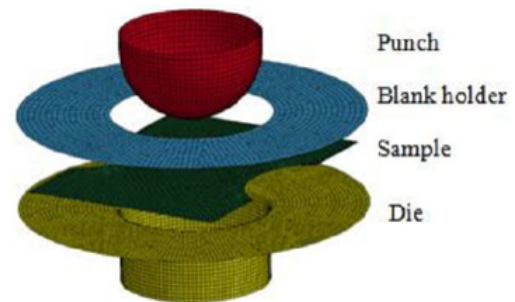


Figure 2: Limit Dome Height Test [7]

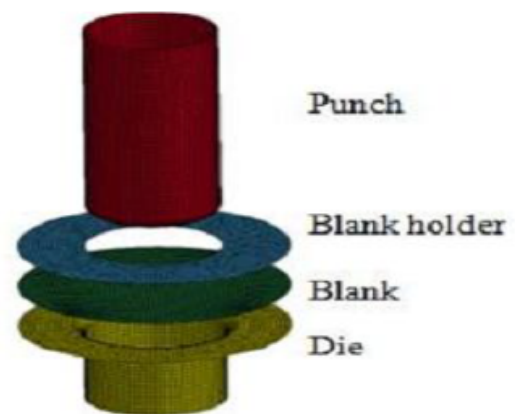


Figure 3: Deep Drawing Test [7]

1.1.1.3. Nakajima test

These tests involve a hemispherical punch to deform sheet metal blanks of various shapes until a fracture occurs [10]. The means growth using a circular die and hemispherical punch to create rectangular specimens of various sizes. The positive and negative domains of the FLD can be obtained by altering the specimen width and lubricant. The test's benefits include the ease of use of the equipment, the square shape of the samples, and the ability to cover the whole Forming Limit Diagram (FLD) [11]. The possibility of wrinkling and measurement error caused by the punch's shape are disadvantages. The ISO 12004 standard 'Metallic materials' uses this procedure as a standard method. The formation of limit curves is determined.

1.1.1.4. M K Model

The M K Model theory is invented by professor Marciniak and Kuczynski in 1967. A model to investigate instability based on the assumption that there was a pre-

existing flaw in the plane of metal sheets before deformation was suggested by Marciniak and Kuczynski. According to the hypothesis of Marciniak, sheet metal has formed manufacturing geometrical thickness variation and structural imperfection composition and gap. The defects gradually involve plastic forming of the sheet metal that is fully focused on having necking of sheet metal. The initial inhomogeneity factor is assumed in the direction perpendicular to the maximum principal stress. Hutchinson invented the M K model to detect FLD, by considering the arbitrary angle between minor principal stress and the imperfection. The M K model's limit strains are sensitive to geometrical imperfections, and this magnitude is far too large for limit strain prediction to be useful. At a high strain ratio developed limit strains are overestimated. The M K model was created to solve this problem by including surface roughness and void growth in the imperfection hypothesis [6].

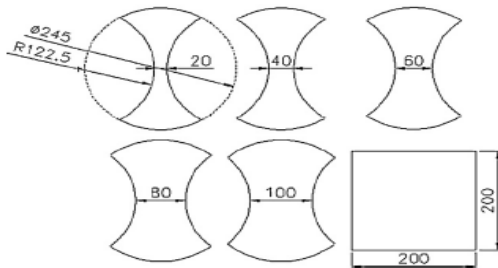


Figure 4: Nakajima Test [11]

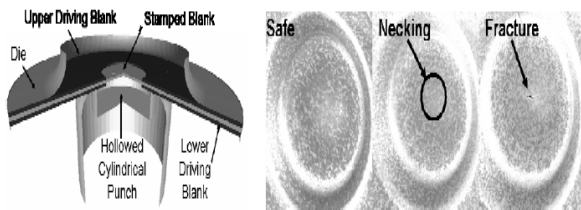


Figure 5: Experimental set up (left side) & 3 specimen [6]

By using Nakajima set up biaxial stretching cannot obtain therefore Marciniak setup has been designed. The main difference between them is the shape of the punch. The Nakajima test uses spherical punch while the Marciniak test uses flat shape punch. Marciniak also developed a specific design: one or more additional sheet metal rings are used to prevent the punch from making contact with the blank. This test is difficult in practice because a hot blank from the furnace must be introduced between two cold carrier blanks in a very short amount of time. This is why a finite element analysis was performed to examine the strain distribution and optimize the driving blanks such as material and central hole diameter [6].

2. Effect of parameters

The smooth and defect-free operation of sheet metal forming is dependent on the selection of various process

parameters. Studying forming in sheet metals is made easier with the use of the circle grid analysis and forming limit diagram (FLD). The above-mentioned process parameters, as well as their combinations, have been discussed in terms of experimental and analytical approaches in this paper [12].

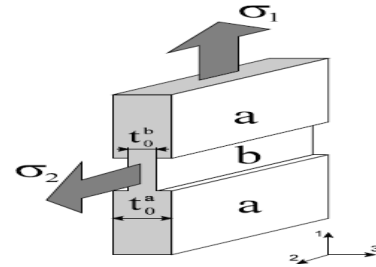


Figure 6: Geometrical model of M K Model [6]

2.1. Punch nose radius

The impact of the punch nose radius on deep drawing. Six different punches with different nose radii were utilized to create a cylindrical with a 44mm outer diameter (OD), a 28mm height, and a 0.5mm sheet thickness of mild steel with 0.15% carbon. The results reveal that it costs substantially more to create pieces with large nose radii than it does to form parts with small punch nose radii. Due to the significant stretching of the metal over the punch head, the hemispherical punch straightens the most [12].

2.2. Blank Temperature

On the deep drawing properties of the aluminum 7075 sheet, the effect of 3 main deep drawing process parameters: punch velocity, blank temperature, and die arc radius was studied [12].

2.3. Blank Holding Force

The behavior of wrinkling and the limiting drawing ratio with sheet thickness was studied by varying the blank holding force. The amount of blank holding power required to eliminate wrinkling grew significantly as the sheet thickness dropped, according to the authors. When the sheet thickness was extremely thin, the coefficient of friction had a significant impact on the blank holding force. As sheet thickness fell, the limiting drawing ratio reduced rapidly below 0.04mm thickness. When the thickness of the sheet was extremely thin, the coefficient of friction had a significant impact on the limiting drawing ratio [12].

2.4. The initial blank shape

The effect of the original blank shape and holding force on the final part quality, such as wrinkling and breakage. From the point of view of damage and fracture, the oval blank shape was discovered to have the worst formability in the initial testing. The three blank shapes oval, oblong,

and rectangle were evaluated, and a mute conclusion was formed [12].

2.5. Spring back

The spring-back tendency of material during sheet metal forming has been researched, with positive results for decreasing defects such as wrinkling, earing, damage, crack, and fracture. According to the researchers, spring-back occurs at the V-region and die-lip of the die model [12].

3. Micro Forming

Miniaturization of products is a current trend in many fields of engineering, enabling product functionality to be performed in the micro-geometry to minimize volume, weight, contamination, and cost [13]. It's a micro-fabrication process for micro components that makes use of metal's superior mechanical and functional qualities, and it's a significant forming technology that's has been reduced in size to the micron size range, making it perfect for manufacturing. According to Yole Développement, the microelectromechanical systems (MEMS) industry in micro-manufacturing would grow at a CAGR of 7.5% for medical applications, 43.6% for telecom applications, and 11.0% for industrial applications from 2018 to 2024 [14]. Due to increasing miniaturization in a variety of fields, such as medical equipment, precision equipment, communication equipment, micro-electromechanical systems, and microfluidic systems, demand for small-sized products has grown significantly. Over the last 20 years, research into metal micro-forming technologies has improved. Embossing, micro-sheet metal forming, Micro-bulk forming, micro-hydroforming, surface engineering for die and tooling, laser-assisted micro-forming, and micro-sintering are some of the research and development activities included in these studies [13]

4. Limitations of Micro-forming

It's critical to understand the process's limiting parameters in industrial micro forming applications. The limiting issues could be related to the micro forming's special characteristics. There is a lot of study being done on how to optimize the micro forming process [15].

Size Effect- It is known that size effects play a key role in micro forming processes, and as a result of size effects, many unexpected changes occur while handling micro-scale parts/work-pieces. All of the sources or causes of size effects have offered a distinct idea of their appearance, and scaling law explained the phenomena in which all of the characteristics (size and weight) change in a fixed relationship with the change in size effects. The conclusion is that because size effects characterize the micro-world, a study of size effects in micro-forming is required to analyze all deformation behaviors, including maternal

behaviour, stress distribution, and wrinkle creation on the material surface [16].

The size effect had an impact on flow, and deformation behaviour's needed to be looked into. Experimental studies on the influence of a few significant characteristics such as geometry and grain sizes on the degree of flow-induced defects in the micro forming process of pure copper were conducted in this research to see how size effects affect flow-induced defects. The microstructure and flow pattern in microscale extrusion of components with intricate shapes are investigated, as well as flow-induced flaws in micro forming of a planned part [17].

Grain Size Effect- Micro-forming processes have shown that insignificant elements in bulk materials, such as grain-size effects, should not be ignored. In the micro-forming world, studies have demonstrated that material behavior differs dramatically as the scale is increased. As a result, in any microform-related process, particularly micro-stamping, studies and analyses of each material used in the process must be considered essential to understand their behavior and associate their behavior with the process [18].

A material with a larger grain size has less strength than one with smaller grain size, according to the equation. This impact is entirely dependent on the material's average grain size and is most significant in macro-scale materials. Despite their thinness, the examined strips can still be classified as macro-scale material due to their relatively substantial width and length [18].

Flow Defects- In the micro-extrusion process, irregular material flow causes flow-induced defects, which is similar to the scenario in the macro-forming process. The flow line, on the other hand, becomes more apparent as the specimen size increases [17].

Geometry Defects- The effect of geometry size on the formation of folding defects is greater than the effect of grain size. No matter what grain size is used, the folding defect in deformed objects significantly improves when the geometry size is reduced [17].

5. Research Work

In this paper Amin Teyfourli studied various micro forming processes. Micro forming is one of the frequently used micro-manufacturing methods in deforming processes. This study aims to analyze the key methods of stamping, forging, bending, and deep drawing processes, as well as their major challenges, in addition to continuing to improve micro forming [19]. By miniaturizing or downscaling, conventional and non-traditional techniques have been commonly used to produce micro-products. Subtractive, joining, hybrid, and deforming processes are the four types of general manufacturing processes. The capacity to produce parts with feature sizes

of less than $100\mu\text{m}$ is the most massive benefit of micro-manufacturing. High-volume production should be a primary consideration when developing micro-manufacturing in micro-components. More study is needed to improve micro-forming methods, especially deep drawing and bending processes, as well as more basic research on the interaction of handling tools and materials.

In this work V. Chowthies said that further research is needed to improve some micro forming methods especially deep drawing and bending processes, as well as more fundamental research into the interaction of handling tools and materials [20]. Micro-manufacturing will become a stronger link between macro-manufacturing and Nano-manufacturing in the future due to developments in machining and handling technology.

In this research paper Tsung-Chia Chen presents an analysis of the deep drawing technique for stainless steel micro-channel arrays is presented. The fuel flow field, electric connectivity, and fuel sealing are all performed by the micro-channel of bipolar plates. The updated degenerated shell finite element analysis, Lagrangian formulation, and the r-minimum rule investigates the major link between punch load and stroke, strain, and stress distributions, thickness fluctuations, and depth variations of individual micro-channel sections. The depth and width of a single micro-channel are 0.5 mm, 0.75 mm, respectively, and a micro-channel array is effectively formed [21]. Fractures were most usually seen in the micro-channel bottom's fillet corner. According to the results of the experiments, greater consideration should be given to the punch and die fillet dimension design. A bigger die fillet can contribute to improved formability and a lower punch load.

In this paper Fanghui Jia tested MDD experimentally and numerically to better understand the microscale deformation behavior of a 2-layer Al-Cu composite. The composite material was rolled to a thickness of 50 μm [22]. The annealing of material is then carried out at 400°C . MDD experiments were used to evaluate the draw-ability of the annealed composite. For simulation of the composite material through the micro deep drawing process, FE models with Voronoi tessellations were developed. To account for the grain heterogeneity, several mechanical characteristics have been assigned to each Voronoi tessellation based on experimental data. The results of the simulation are very similar to the experiment results. With the MMD technique, micro components can be manufactured, and the resulting cup has no fractures and fewer wrinkles. To improve the accuracy of FE simulation, grain sizes of Al and Cu can be immersed into the Voronoi model. The result of the FEA analysis with the Voronoi diagram is closer to the experimental result in terms of drawability (cup mouth wall thickness, flange

height, and cup mouth radius), wrinkles, and drawing force [22].

In this research work N. Srinivasan analyzed the integration of growths from the macro to the micro-scale using progressive micro forming techniques, focusing on the size effect deformation behavior in materials, dimensional accuracy, deformation load, failure analysis, and roughness of micro components, all of which are related to micro-scale forming processes such as extrusion, forging, and incremental [23]. The research challenges of micro mass production related to industrial complexity, as well as the micro forming process that requires extensive attention, are discussed. For medical applications such as prosthetic teeth, the study should be extended to include other metals such as magnesium and titanium, and the effect of size on material behavior will be profoundly studied.

In this work A Dhal study shows that refined microstructures have good micro forming capabilities in a variety of deformation domains (0.7 mm to 0.1 mm) [24]. The micro forming capability of micro-scale sheet metal forming is very sensitive to plastic instability. As a result, a finer microstructure with good strain hardening properties is desired. Cry rolling causes a significant amount of dislocation, which must be relieved with annealing. Microscale sheet metal forming requires a refined microstructure with dislocation-free, equiaxed grains.

In this research paper Dr. R. Sridhar focused on recent research and findings in the field of deep drawing. Deep-drawing processes are used to create a product that is light in weight, strong, low in density, and corrosion-resistant. These criteria will make the product more subject to wrinkling and other types of failure. Deep drawing is affected by several parameters such as blank-holder pressure, die radius, punch radius, coefficient of friction, and material characteristics [25]. To develop a defect-free product, a full understanding of the complete process is essential. The researcher aims to gather the latest advancements and research in the field of micro deep drawing. The main process factors and their impacts on micro deep drawing have been critically studied. Material thickness, coefficient of friction, thermal characteristics, stress and strain concentration, and wrinkling are shown to be important process parameters. Micro deep drawing of thin foil materials is found to be profitable for vehicle and electrical applications.

In this paper Sridhar Ramasamy studied the formability of pure copper foils, as well as the thickness distribution in micro deep drawing. The limit drawing ratio is obtained by selecting pure copper C1100 foil thicknesses of 100, 200, and 300 μm [26]. A micro deep drawing tool assembly is created to determine the forming depth in microscale, as well as the force, required

informed items. In both the tensile and Ericson cupping tests, the results show that 300 m foil has good formability. The cup shoulder corner and the upper section of the sidewall are where there is an excessive loss in thickness. In the transverse direction, the thinning is most visible at the cup's sidewall. The research shows that with the right forming settings, high-quality copper micro cups may be made quickly. A fundamental understanding of the micro-deep drawing technique is established. Furthermore, it has been discovered that scaling effects must be considered not just inside the process but also in all other aspects of the forming process. Various copper foil thicknesses are based on micro drawing tests.

In this work the authors study project work is related to micro-farming research on a 90-micron ultra-fine brass sheet is used to mark FLD according to the ASTM E2218-02 limit dome height test. The experimental apparatus was designed and invented to perform micro-form standard specimens using a 15 mm hemispherical shape of the punch [13]. The failure limit curve for FLD shown with experimental and empirical methodologies shows agreement. Many industries, including medical instrumentation, automotive, mobile manufacturing, and aerospace, have embraced miniaturization. In the forming process, appropriate strain distribution promotes higher-quality products. To plot major and minor strains as well as form limit curves, the forming limit diagram (FLD) was contrived. FLD, in other words, exhibits localized necking, wrinkling, and safe zones with varied strain routes, and it proves to be a useful tool for determining the material's rupture properties. Empirical (analytical), numerical, and experimental ways to plot FLD are available.

In this research the authors studied a new micro deep drawing approach for making micro-cups that combines a floating ring which acts as the principal rigid die, while a rubber pad acts as the primary flexible die [27]. The floating ring's goal is to defend tiny wrinkles from forming at the flange, while the flexible die's job is to finish the forming stroke. Simulations and tests are used to investigate the effects of drawing ratio, initial sheet thickness, rubber height, and punch corner radius. In addition, three size scales are used to test the feasibility of applying the suggested technique to various process dimensions.

In this research the authors experimented with the forming limit curves, fine alpha brass sheets of 40 m and 90 m with different orientations to rolling directions (0°, 45°, 90°) were subjected to an experimental test [28]. According to ASTM-2218-14, a hemispheric punching test for micro forming is done. A standard test was done using a sample of Single-Axial, Intermediate Single-Axial, and Two-Axial, Intermediate Two-Axial, and Plane Strain pathways to determine limit strains. The micro-forming process is modeled using explicit dynamic numerical analysis on the Abaqus platform and forming limit curves

are displayed. Experimental data have been used to validate numerical results.

6. Conclusions

In this paper, High-volume production should be the main consideration when developing micro-manufacturing in micro-components. Currently, the majority of sheet metal research is focused on establishing the failure limit diagram using finite element methods, experimental methods, numerical analysis, and sheet metal material management. As described in this paper, additional research is needed to improve micro-forming methods, particularly deep drawing and bending processes, as well as more basic research into the interaction of handling tools and materials.

This paper explains the micro forming and various micro forming processes. This paper also discussed the effect of process parameters are punch nose radius, Blank temperature, spring back, and blank holding force.

References

- [1] Gyan Patel and Ganesh Kakandikar, "Investigations on effect of thickness and rolling direction of thin metal foil on forming limit curves in microforming process," in *Modern Manufacturing Processes*, (Elsevier, 2020), 145–155, doi:10.1016/b978-0-12-819496-6.00007-5.
- [2] U. V Mangudkar, S. P. Mankani, "COMPARATIVE STUDIES ON FORMABILITY ANALYSIS IN METAL FORMING," *IJRET: International Journal of Research in Engineering and Technology*. .
- [3] V. R. Shinge, U. A. Dabade, "Experimental Investigation on Forming Limit Diagram of Mild Carbon Steel Sheet," *Procedia Manufacturing*, vol. 20, pp. 141–146, 2018, doi:10.1016/j.promfg.2018.02.020.
- [4] M. S. Kulkarni, S. Y. Gajjal, "Review of Sheet Metal Forming Analysis," *SSRG International Journal of Mechanical Engineering*, vol. 2, no. 1, . 2015.
- [5] J. Sahu, S. Mishra, "Limit dome height test of very thin brass sheet considering the scaling effect," *Journal of Physics: Conference Series*, vol. 734, no. 3, 2016, doi:10.1088/1742-6596/734/3/032114.
- [6] D. Banabic et al., *Sheet metal forming processes: Constitutive modelling and numerical simulation* (Springer Berlin Heidelberg, 2010).
- [7] J. M. K. Rao, "Numerical Simulation for Predicting Failure in Deep Drawing Process Using Forming Limit Diagram (Fld)," no. 4, pp. 11–15, 2015.
- [8] J. Pavan Kumar et al., "Formability of sheet metals - A review," *IOP Conference Series: Materials Science and Engineering*, vol. 455, no. 1, 2018, doi:10.1088/1757-899X/455/1/012081.
- [9] J. Zhao et al., "Experimental Investigation on Micro Deep Drawing of Stainless Steel Foils with Different Microstructural Characteristics," *Chinese Journal of Mechanical Engineering (English Edition)*, vol. 34, no. 1, 2021, doi:10.1186/s10033-021-00556-5.
- [10] N. Ayachi et al., "Development of a nakazima test suitable for determining the formability of ultra-thin copper sheets," *Metals*, vol. 10, no. 9, pp. 1–18, 2020, doi:10.3390/met10091163.
- [11] S. B. Kim et al., "Forming limit diagram of auto-body steel sheets for high-speed sheet metal forming," *Journal of Materials Processing Technology*, vol. 211, no. 5, pp. 851–862, 2011, doi:10.1016/j.jmatprotec.2010.01.006.
- [12] S. Dwivedi, A. Kumar Bhaisare, A. Singh, "A REVIEW-EFFECT OF PARAMETERS AND ITS ANALYSIS ON FORMABILITY." .
- [13] A. Mashalkar, G. Kakandikar, V. Nandedkar, "Micro-forming analysis of ultra-thin brass foil," *Materials and Manufacturing Processes*, vol. 34, no. 13, pp. 1509–1515, 2019, doi:10.1080/10426914.2019.1655158.

- [14] Ken Ichi Manabe, "Metal micro-forming," *Metals*, vol. 10, no. 6, . MDPI AG, : 1–3, 2020.
- [15] M. Oraon, V. Sharma, "Sheet Metal Micro Forming: Future Research Potentials." 2010.
- [16] M. Singh et al., "International Review of Mechanical Engineering (IREME) Contents: REPRINT Size effects in Micro-forming: A review by Mandeep Singh, Dongbin Wei 70 (continued on inside back cover) International Review of Mechanical Engineering (IREME) Editor-in-Chief: Edit," *International Review of Mechanical Engineering*, 2018.
- [17] J. L. Wang, M. W. Fu, J. Q. Ran, "Analysis of size effect on flow-induced defect in micro-scaled forming process," *International Journal of Advanced Manufacturing Technology*, vol. 73, no. 9–12, pp. 1475–1484, 2014, doi:10.1007/s00170-014-5947-8.
- [18] M. A. Musa, A. R. Razali, N. I. Kasim, "Grain and feature size effect on material behavior for micro-sheet-forming," *Applied Mechanics and Materials*, vol. 680, pp. 77–80, 2014, doi:10.4028/www.scientific.net/AMM.680.77.
- [19] A. Teyfour et al., "A Review on Micro Formings," *Modern Applied Science*, vol. 9, no. 9, 2015, doi:10.5539/mas.v9n9p230.
- [20] V. Chowthies, A. Prof, "IJRME-International Journal of Research in Mechanical Engineering A REVIEW ON MICRO-MANUFACTURING, MICRO-FORMING AND THEIR KEY ISSUES," .
- [21] T. C. Chen, J. C. Lin, R. M. Lee, "Analysis of deep drawing process for stainless steel micro-channel array," *Materials*, vol. 10, no. 4, 2017, doi:10.3390/ma10040423.
- [22] F. Jia et al., "Experimental and numerical study on micro deep drawing with aluminium-copper composite material," *Procedia Engineering*, vol. 207, pp. 1051–1056, 2017, doi:10.1016/j.proeng.2017.10.1129.
- [23] N. Srinivasan, D. Rajenthirakumar, R. Sridhar, "Micro Forming-Key Issues and Research Opportunities," *Advances in Natural and Applied Sciences*, vol. 11, no. 4, pp. 557–579, 1998.
- [24] A. Dhal, S. K. Panigrahi, M. S. Shunmugam, "Investigation into the micro deep drawing capabilities of a specially engineered refined aluminium alloy," *MATEC Web of Conferences*, vol. 190, 2018, doi:10.1051/mateconf/201819010001.
- [25] R. Sridhar, S. ShankarC, "Review On Micro Deep Drawing Process For Thin Foil Materials," *International Journal of Innovations in Engineering and Technology (IJJET)*, vol. 11, 2018, doi:10.21172/ijjet.113.03.
- [26] S. Ramasamy et al., "EXPERIMENTAL EVALUATION OF SIZE EFFECTS IN MICRO DEEP DRAWING PROCESS OF THIN FOIL MATERIALS." .
- [27] I. Irthia, Z. Mahmood, "Effect of process parameters on micro flexible deep drawing of stainless steel 304 cups utilizing floating ring: Simulation and experiments," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 235, no. 1–2, pp. 134–143, 2021, doi:10.1177/0954405420949199.
- [28] G. Patel, K. Ganesh M, O. Kulkarni, "Experimental and numerical investigations on forming limit curves in micro forming," *Advances in Materials and Processing Technologies*, pp. 1–12, 2020, doi:10.1080/2374068X.2020.1793268.



Neha Tiwari has completed her bachelor's degree in mechanical engineering from G.H.Raisoni college of engineering, Jalgaon, Maharashtra. Now currently pursuing her master's degree in computer Aided Design, Manufacturing and Analysis from MIT World Peace University from, Pune, Maharashtra.



Ganesh M. Kakandikar is Professor Associate Head of School of Mechanical Engineering at Dr. Vishwanath Karad MIT World Peace University, Pune. He has completed a Ph.D. in mechanical engineering from Swami Ramanand Teerth Marathwada University, Nanded in 2014. He has 21 years of experience in teaching, research, and administration. He has authored 03 books published internationally by Lambert Academic and CRC Press, Taylor, and Francis Group and contributed book chapters published by Wiley, Springer, Elsevier. 65+ publications are at his credit in National/International Journals and conferences. He is the reviewer of many reputed journals



Omkar Kulkarni has received his master's in computer-aided design, manufacture, and engineering, currently pursuing his Ph.D. degree in mechanical engineering from MIT World Peace University from, Pune, Maharashtra. Now he works as Assistant Professor in Dr. Vishwanath Karad MIT World Peace University. His current research includes applications of nature-inspired algorithm in mechanical applications and micro forming of sheet metal

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Detailed Overview on POLYMATH Software for Chemical Engineering Analysis

Abdulhalim Musa Abubakar ^{*1}, Bello Iliyasu ², Zakiyyu Muhammad Sarkinbaka ³

¹ Department of Chemical Engineering, Modibbo Adama University (MAU), P.M.B 2076, Yola, Adamawa State-Nigeria

² Department of Mathematics, Modibbo Adama University (MAU), P.M.B 2076, Yola, Adamawa State-Nigeria

³ Department of Chemical Engineering, Federal University Wukari, Faculty of Engineering, P.M.B 1020, Wukari, Taraba State, Nigeria

*Corresponding author: Abdulhalim Musa Abubakar, +2347050244277 & abdulhalim@mautech.edu.ng

ABSTRACT: It is pertinent to highlight areas POLYMATH software is useful for chemical engineering analysis. Its applications had been demonstrated in this paper using 10 Problem Set, in areas that includes transport phenomena, heat transfer, reaction, and bioreaction kinetics to solve differential equations, nonlinear equations, simultaneous linear equations, graphical representation and regression problems arising in these fields using Licensed POLYMATH Software Version 6.10.261. POLYMATH is mostly used by students, teachers and researchers for educational purposes, but however limited in application by engineers in process industries, as more sophisticated softwares are preferred. Incorporating many solution approaches there in the literature for solving mathematical method problems, especially the newly proposed novel Extended Runge-Kutta Method of Order 6 into POLYMATH and the enabling of log-log plots is hereby recommended.

KEYWORDS: POLYMATH, Problem Set, Chemical Engineering Analysis, Runge-Kutta-Fehlberg, ODE

1. Introduction

Tiring engineering task, including calculations and analysis, today are carried out using computerized systems or specific softwares designed for specific type or large variety of functions. Computers are therefore, powerful tools employed to collect, store, process and analyze data for simulation, explaining mathematical models and analysis of basic scientific phenomena [1]. Chemical engineers employ the services of softwares such as POLYMATH, ASIM, MathCad, AQUASIM, MAPLE, Mathematica, FORTRAN, Excel, MATLAB and ASPEN among others for purposeful analysis. So, the act of finding the root causes of observed phenomenon utilizing both statistical, graphical and numerical analysis softwares is termed chemical engineering analysis. Chemical engineering analysis is thus broadly applied in chemical reaction engineering (CRE), transport

phenomena (e.g. fluid mechanics, heat and mass transfer), process dynamics and control, thermodynamics, phase equilibria and distillation and biochemical engineering among others [2]. In CRE, problems solved with POLYMATH are co-current and countercurrent heat exchange, nonlinear regression of rate equations, nonlinear algebraic balances on continuous stirred tank reactors (CSTRs), growth kinetics in bioreactors and membrane reactors, multiple reactions with heat effects, to mention a few [3].

POLYMATH is an educational software recommended for mathematicians, engineers [4] and financial analysts, created by Shacham and Cutlip in January 1, 2003. The challenge of using FORTRAN and C++ is that, a very lengthy time is spent developing and testing codes [3] in addition to hours spent analyzing the result based on different possible scenarios. Unlike

ASPEN and COMSOL where the equations behind the computations are hidden, POLYMATH allow the user to provide their mathematical model where necessary [5] and are solved immediately. Errors in the typed mathematical models, entry of wrong initial estimates or the selection of an inappropriate solution algorithm would prevent the software package from given the correct solutions to problems entered [6, 7]. A table of results containing initial estimates, problem statement and calculated unknown parameters are presented by POLYMATH together with graphs relating some of the parameters in the expression [8]. A hardcopy of this result can sometimes be printed by pressing the function key (F3) on the keyboard [9].

Famously, a solution to ten (10) chemical engineering numerical problems that cut across different aspects of chemical engineering, titled (i) molar volume and compressibility factor from Van Der Waal's equation, (ii) steady state material balances on a separation train, (iii) vapor pressure data representation by polynomials and equations, (iv) reaction equilibrium by multiple gas phase reactions, (v) terminal velocity of falling particles, (vi) unsteady state heat exchange in a series of agitated tanks, (vii) diffusion with chemical reaction in a one dimensional slab, (viii) binary batch distillation, (ix) reversible, exothermic, gas phase reaction in a catalytic reactor, and (x) dynamics of a heated tank with PI temperature control, had been solved using six numerical software packages including Excel, MAPLE, MATHCAD, MATLAB, MATHEMATICA, and POLYMATH [4] for undergraduate education [1]. Statistical analysis of adsorption isotherm models and estimation of growth kinetic parameters can be done using POLYMATH [10, 11]. It is a user friendly software for numerical computation with short learning time, requiring minimal user intervention in technical details of the solution process and had been incorporated into several text, such as Kyle's textbook of "Chemical and Process Thermodynamics" and Fogler's textbook of "Elements of Chemical Reaction Engineering" among others [1]. An android application that can run on a minimum of Android 2.2 version called PolyMath Lite had been developed by Cutlip in 2016 [12].

2. Features of POLYMATH Software Package

POLYMATH is a simple software package developed to help students, academicians and engineers

solve numerical analysis problems. In a book written by [2], 170 chemical and biochemical engineering problems requiring numerical solutions had been highlighted. The package consists of a calculator, units converter, records of scientific constants and numerical method icons. Numerical methods in chemical engineering deals broadly with subjects like (i) quantum mechanical calculations of atoms and molecules, (ii) numerical treatment of chemical reaction kinetics, (iii) transport processes, (iv) mathematical description of unit operations, and (v) stationary and instationary simulation and optimization of chemical plants among others [13]. Thus, POLYMATH solves four type of numerical problems arising in those subjects, namely, system of nonlinear equations, system of ordinary differential equations (ODEs), data regression and system of linear equations.

POLYMATH can solve up to 300 set of nonlinear equations with 300 explicit equations using four types of solution approaches, including 'safenewt', 'fastnewt', 'safebroydn' and constrained nonlinear equations [5]. Same number of ODEs can be solved by POLYMATH by either Runge-Kutta-Fehlberg (RKF45), RKF56, Burlirsch-Stoer (BS), STIFF or STIFFBS method. The data regression window is divided into regression, analysis and graph tabs. Multiple linear, nonlinear and linear and polynomial regression with room for finding polynomial derivative and integration are found under the regression tab which is capable of regressing 200 variables with up to 1000 data points for each. Under the analysis tab, numerical differentiation, integration and interpolation can be done using 3 solution methods each. Graphs are plotted entering the 'x' and 'y' data in cell box sheet provided by POLYMATH under the graph tab. Lastly, POLYMATH can accommodate 264 simultaneous linear equations as stated by [5]. Table 1 shows selected chemical engineering problems that can be solved using the four numerical methods discussed above:

Table 1: Chemical Engineering Problem Set with POLYMATH Solutions [14–17]

No .	Mathematical Method	Problem Statement	Equation, Nomenclature & POLYMATH Result
1.	Nonlinear with Constants	Given the equation of saturation concentration of dissolved oxygen in freshwater, $o_{sf} = 14.621$ mg/L, find the absolute temperature, T_a , taking initial guess of 273.15K.	$\ln o_{sf} = -139.34411 + \frac{1.575701 \times 10^5}{T_a} - \frac{6.642308 \times 10^7}{T_a^2} + \frac{1.243800 \times 10^{10}}{T_a^3} - \frac{8.621949 \times 10^{11}}{T_a^4}$ <p>where, o_{sf} = the saturation concentration of dissolved oxygen in fresh-water at 1 atm (mg/L) and T_a = absolute temperature (K). From Problem Set 1, it could be concluded that o_{sf} ranges from 14.621 mg/L at 0°C to 6.413 mg/L at 40°C.</p>
2.	Nonlinear with Constants	Solve the Manning equation for a rectangular open channel for H given $Q = 5$, $S = 0.0002$, $B = 20$, and $n = 0.03$ taking initial guess ≥ 0 . Prove that the scheme converges for all initial guesses of $H \geq 0$.	$Q = \frac{\sqrt{S} (BH)^{5/3}}{n(B + 2H)^{2/3}}$ <p>where, Q = flow (m³/s), S = slope (m/m), H = depth (m) and n = the Manning roughness coefficient. Results of Problem Set 2 is $H = 0.7022933$m for $H \geq 0$.</p>
3.	Nonlinear with Constants	Von Karman equation can be used to predict the fanning friction factor, f , for pipes. For the range and initial value of $f = 0.001(0.0001)0.01$, estimate f by numerical analysis taking typical Reynold's number, Re for turbulent flow from 10,000 to 500,000	$\frac{1}{\sqrt{f}} = 4 \log(Re\sqrt{f}) - 0.4$ <p>where, f = fanning friction factor and Re = Reynold's number. POLYMATH gives $f = 0.007727$ when $Re = 10,000$ and $f = 0.003291$ when $Re = 500,000$.</p>
4.	Nonlinear with Explicit Function	Reynolds' dimensionless expression, Colebrook equation and pressure drop (ΔP) can be solved	$Re = \frac{\rho V D}{\mu}$ $\frac{1}{\sqrt{f}} = -2.0 \log \left(\frac{\epsilon}{3.7D} + \frac{2.51}{Re\sqrt{f}} \right)$ $\Delta P = \frac{f L \rho V^2}{2D}$

		<p>simultaneously for f and ΔP (Pa). Take $L = 0.2\text{m}$, $\rho = 1.23 \text{ kg/m}^3$, $\mu = 1.79 \times 10^{-5} \text{ N.s/m}^2$, $D = 0.005\text{m}$, $V = 40 \text{ m/s}$ and $\varepsilon = 1.5 \times 10^{-6}\text{m}$ using the Blasius formula, $f = 0.316/Re^{0.25}$ as good initial guess of f for a smooth tubing.</p>	<p>where, f = friction factor, L = length of pipe, ρ = density, V = velocity, μ = dynamic viscosity, D = diameter, ε = roughness, and ΔP = pressure drop. The results window presents $f = 0.02844$ for $\Delta P = 1119.459 \text{ Pa}$.</p>
5.	Simultaneous Nonlinear	<p>Fig. 2a. displays heat transfer in the Thermos, q_1, q_2 and q_3 in the inner compartment, middle layer and outside shell which are by radiation, convection in a small space and natural convection respectively. Find the temperatures, T_1 and T_2 at steady state when $T_0 = 450^\circ\text{C}$ and $T_3 = 25^\circ\text{C}$, considering that $q_1 = q_2 = q_3$ from each region are equal.</p>	$q_1 = 10^{-9}[(T_0 + 273)^4 - (T_1 - 273)^4]$ $q_2 = 4(T_1 - T_2)$ $q_3 = 1.3(T_2 - T_3)^{4/3}$ <p>where, q_1 = heat flux from inner compartment to the next layer, q_2 = heat flux between middle layer and outside shell, q_3 = heat flux from the outside shell, and T = temperature in those regions ($^\circ\text{C}$). Constrained nonlinear POLYMATH solution approach gives, $q = 273.0047$, $T_1 = 148.417^\circ\text{C}$ and $T_2 = 80.16582^\circ\text{C}$.</p>
6.	Differential Equation	<p>A half full storage tank in which liquid is supplied at a sinusoidal rate, $3Q\sin^2(t)$, and withdrawn at constant flowrate to meet demand is shown in Fig. 2b. Use the continuity equation to solve for y which is initially zero, taking area, $A = 1200 \text{ m}^2$, $Q = 500$</p>	$\frac{dy}{dt} = 3\frac{Q}{A}\sin^2(t) - \frac{Q}{A}$ <p>where, Q = liquid flowrate (m^3/day), A = area (m^2), y = liquid depth (m) and t = time (days). The algorithm of Runge-Kutta-Fehlberg (RKF45) method gives a depth, $y = 1.798038\text{m}$ at $t = 10$ days. Though Excel solution of this problem is a tedious step, results are however the same.</p>

		m ³ /day, at time from t = 0 to 10 days, selecting a suitable step size.											
7.	Simultaneous Nonlinear Differential Equations	Given the Monod model for bioreaction kinetics, find the substrate concentration, s , and the biomass concentration, x for values of $y = 0.05$, $k_s = 20$, $k = 5$, and $b = 0.01$, starting with $s_0 = 1000$ and $x_0 = 100$ at time ranging from t = 0 to 1.	$\frac{ds}{dt} = -\frac{ksx}{k_s + S}$ $\frac{dx}{dt} = y \frac{ksx}{k_s + S} - bx$ <p>where, s = growth limiting substrate concentration (ML⁻³), x = biomass concentration (ML⁻³), k = maximum specific uptake rate of the substrate (T⁻¹), k_s = half saturation constant for growth (ML⁻³), y = yield coefficient (MM⁻¹), and b = decay coefficient (T⁻¹). Both RKF45 and RKF56 solution approaches gives $s = 452.5393$ and $x = 126.2464$.</p>										
8.	Multiple linear Regression	Fit the data of Figure 3 to Antoine equation and estimate the values of the constant parameters, A, B & C. Take 10, 1000 and 100 as initial values of A, B and C respectively. The expected fitted data curve is also displayed in column 3 of the table in Figure 3.	$\log P = A - \frac{B}{T + C}$ <p>where, P = vapor pressure (mmHg) and T = temperature (°C). Hint: Rewrite the above equation in form of $y = a_0 + a_1x_1 + a_2x_2$; a multiple linear regression format, i.e.</p> $\log P = A + \frac{AC - B}{T} - \frac{C \log P}{T}$ <p>and compare the two for constants, a_0, a_1 and a_2 which in turn gives A, B and C. POLYMATH gives A = 17.49696, B = 4515.673 and C = 241.6434 with coefficient of variation, R² = 0.99998</p>										
9.	Linear Regression	Find the constants, 'a' and 'b' in the given formula relating thermal conductivity of iron (k) with temperature (T) by regression with POLYMATH. Hint: Linearize the given function.	$Tk^a = b$ <p>where, T = temperature (K) and k = thermal conductivity (W/cm-K)</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>T/K</th> <th>200</th> <th>600</th> <th>1000</th> <th>1400</th> </tr> </thead> <tbody> <tr> <td>k(W/cm-K)</td> <td>1.0</td> <td>0.4</td> <td>0.3</td> <td>0.25</td> </tr> </tbody> </table> <p>Using the above data, it was found that, $a = 1.3739$, $b = 191.97$ and R² = 0.9895.</p>	T/K	200	600	1000	1400	k(W/cm-K)	1.0	0.4	0.3	0.25
T/K	200	600	1000	1400									
k(W/cm-K)	1.0	0.4	0.3	0.25									
10.	Numerical Interpolation, Differentiation	Day and night variation of temperature of a	Measured Data:										

n & Integration.	furnace wall with time is given in the table of the side column. Interpolate for T at t = 12hrs. Differentiate, $\frac{dT}{dt}$ at t = 15 hrs. Carryout numerical integration to determine, $\int_2^{24} f(t)$.	t/(hr)	2	3	5	8	10	15	18	22	24
		T/(°C)	86.5	97.7	104.0	101.7	92.5	62.3	55.0	67.5	80.0

POLYMATH gives T = 80.086, 79.89355 and 80.42°C selecting RATINT, SPLINT and Linear solution type respectively, interpolated at t = 12 hrs. At t = 15 hrs, $\frac{dT}{dt} = -4.5996$ °C/hr by STANDA and SPLINE solution method and -4.8155 °C/hr by LAGRANGE's numerical differentiation method. GAULEG and SPLINE integration gives approximately 1745 while SIMPSON numerical integration technique gives 1742.8

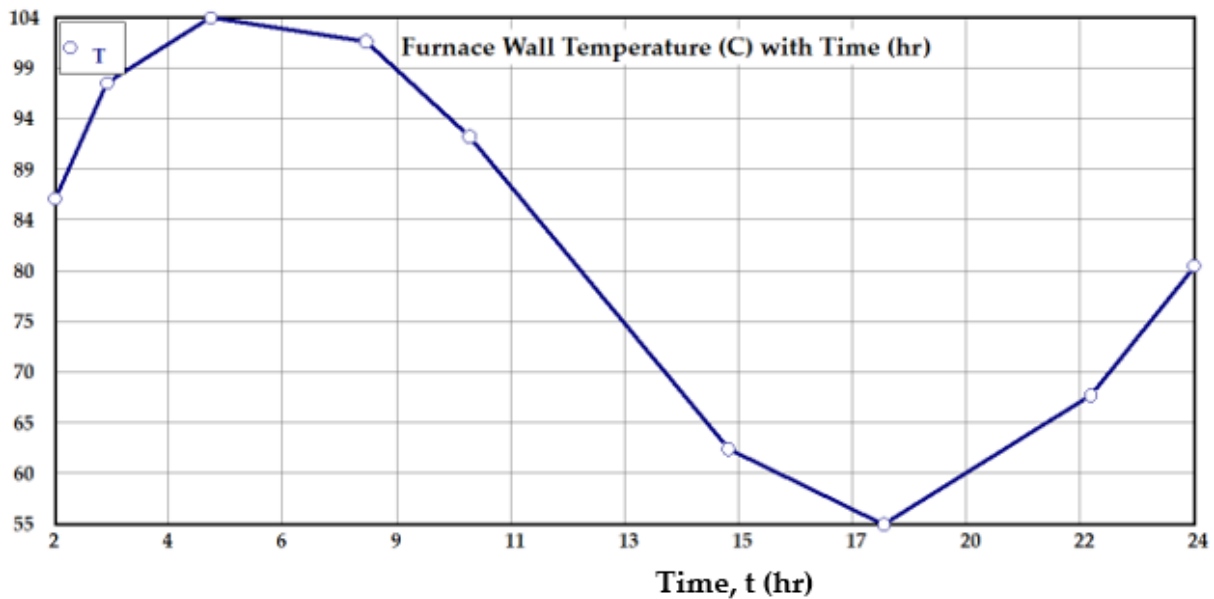


Figure 1: Plot of T(°C) versus t(hrs) for Problem Set 10

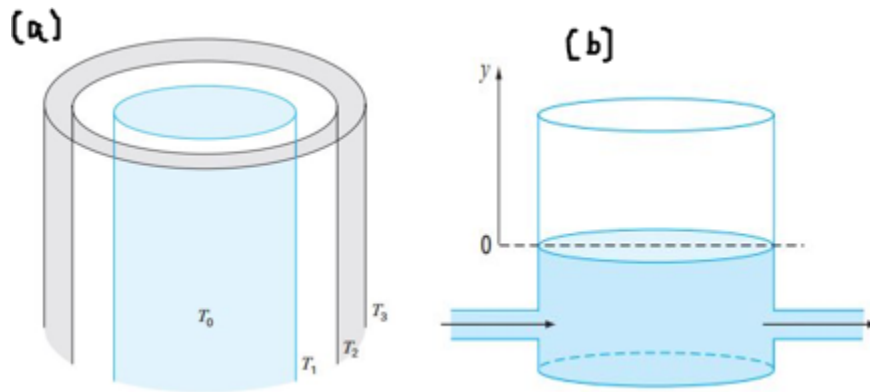


Figure 2: (a) Heat Transfer in Thermos and (b) Storage Tank Liquid Flow [15]

3. Nonlinear Equations

Chemical engineering problems can possibly result in a single or sets of nonlinear equations solvable by numerical methods, including Secant Method, Regula-Falsi, Iterative Method, Bisection, and Newton's Method. Equations of state (EOS) in Thermodynamic, flow equations in Transport Phenomena (e.g. Manning, Von Karman and Colebrook equations as defined in Problem Set 2-4 of Table 1), equilibrium equation in Separation Processes, and Heat Transfer problems seen in Problem Set 5, to mention a few are solved using any of these methods. In [18], Newton's algorithm had been applied to formulate a programming codes capable of solving cubic EOS. Among these EOS, [6] demonstrated how Redlich-kwong equation of state is solved using successive substitution method, a method similar to how Problem Set 5 in this work was solved using POLYMATH. In [19], the modified Raoult's law had been solved using simultaneously with Wilson equation to determine the composition of a non-ideal binary mixture.

In POLYMATH, nonlinear equations are solved with basically two methods: (a) Newton's, and (b) Broyden's method. The "safenewt" algorithm uses the Newton's technique with line search to guarantee reduction of the norm of the function values in all iteration [20]. Broyden is a quasi-Newton method (or the generalization of the Secant method to nonlinear systems) that is superlinearly convergent for finding roots in k variables, as it reduces the amount of calculation at each iteration without significantly degrading the speed of convergence [21]. Therefore, both 'safenewt' and 'safebroyd' includes function minimization along the Newton and Broyden directions.

4. Differential Equations

POLYMATH would request the typing of a function to be solved by numerical methods or user-defined functions to be entered for regression purposes. Functions generate one unique output for each input [8] which are termed as arguments, often enclosed in parenthesis [9]. Some of the functions recognized by POLYMATH are 'sin', 'sinh', 'cos', 'cosh', 'tan', 'tanh', 'ln', 'log', 'arcsin', 'arccos', 'arctan', 'exp', 'frac', 'int', 'abs',

'sign', and 'round'. For example, absolute value of say 'x' can be written as "abs(x)". Also, factorial of an integer, say '5' is written as "5!" in POLYMATH. Differential functions which are implicit type of functions as well as explicit functions must be specified in POLYMATH.

Differential equations are classified into ordinary or partial differential equations and linear or nonlinear equations [22]. A mathematical equation that relates some functions with their derivatives are known as differential equations. Practically, the functions represent physical quantities, the derivatives stand for their rates of change, and the equation defines a relationship between the two [8]. Modeling of the variation of a physical quantity, such as temperature, pressure, displacement, velocity, stress, strain, current, voltage, or concentration of a pollutant, with the change of time or location, or both would result in differential equations [23]. Differential equation govern almost all physical processes of interest to chemical engineers, namely [24]; description of thermodynamic laws, description of flow processes (e.g. using Darcy's law, Bernoulli's equation, Navier-Stokes and mass conservation equations), microbial growth kinetic constant estimation, representation of reaction rates in CRE, Fick's law of diffusion in mass transfer and many more. Elsewhere, [25] highlighted some of the applications of ODEs in reaction engineering. Partial differential equations (PDEs) on the other hand, finds applications in CRE to model complex reactors and reacting systems [3]. Important examples of PDEs are, Tricomi equation, Laplace equation, Helmholtz equation, Poisson-Boltzmann equation, Lorenz equations, Cauchy-Riemann equations, Poisson's equation, Plate bending, Schrödinger's equation, Geodesic equation, Navier-Stokes, Wave equation (1D-3D) and Fourier equation [22, 23]. Softwares that can deal with PDEs are COMSOL (formerly FEMLAB) and MATLAB, especially for reactions taking place in tubular reactors [3]. ODEs in Problem Set 6 & 7 where solved by POLYMATH which is also utilizing the six parameters (k_1, k_2, k_3, k_4, k_5 and k_6) of the RKF method in equations (1) [26, 27].

$$k_1 = h f(t_k, y_k)$$

$$k_2 = h f\left(t_k + \frac{1}{4}h, y_k + \frac{1}{4}k_1\right)$$

$$\begin{aligned}
 k_3 &= h f\left(t_k + \frac{3}{8}h, y_k + \frac{3}{32}k_1 + \frac{9}{32}k_2\right) \\
 k_4 &= h f\left(t_k + \frac{12}{13}h, y_k + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right) \\
 k_5 &= h f\left(t_k + h, y_k + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right) \\
 k_6 &= h f\left(t_k + \frac{1}{2}h, y_k - \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)
 \end{aligned} \tag{1}$$

An approximate solution, y_{k+1} , to given initial value problems (IVPs) can then be executed using the RKF45 approximation as in equation (2):

$$y_{k+1} = y_k + \frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4101}k_4 - \frac{1}{5}k_5 \tag{2}$$

An improved value of the solution, z_{k+1} , is determined using another approximation. Check equation (3):

$$z_{k+1} = y_k + \frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56460}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6 \tag{3}$$

Note that the approximations in (2) and (3) of the RKF45 method [26] ignores, k_2 . This makes equation (2) of order 4 and (3), an order 5 equation; basically, making use of the classical Runge-Kutta methods of order 4 and 5 together, therefore defining RKF45 algorithm holistically. The selection of appropriate step sizes, h , guarantees the accuracy of most IVPs. The advantage of the RKF45 algorithm is that, it ensures that the right, h is used on its own and not allowing the user to specify any step size. Problem Set 6 had been coded to give, y_{k+1} and z_{k+1} using a step size of $h = 0.1$ for the RKF45 technique using C++ as seen in the Appendix. RKF56 on the other hand is termed the 6th order Runge-Kutta-Fehlberg method and is presented by [28]. The Runge-Kutta-Fehlberg method is generally thought of as one of the best methods existing for solving non-stiff systems of equations [27].

Starting from order 1 to 6, differential equations are solved using (i) Euler's method of order 1, (ii) Heun's method (or modified Euler or leap-frog formula) of order 2, (iii) Bogacki-Shampine method of order 3, (iv) Runge-Kutta method of order 4, (iv) R-K-Fehlberg method of order 4 with an error estimator of order 5 (RKF45), and (v) RKF method of order 5 with an error estimator of order 6 (RKF56). Beyond this, not long ago, [29] proposed a novel Extended RK 6th order method to bring the

number of the solution approaches to six. A differential equation of the form, $y' = f(t, y)$ is said to be stiff if its exact solution $y(t)$ includes a term that decays exponentially to zero as t increases (i.e. e^{-ct} , where c is a large positive constant), but whose derivatives are much greater in magnitude than the term itself [30]. A stiff differential equation is further characterized as being numerically unstable unless the step size is extremely small [30]. Stiff and Burlirsch-Stoer (BS) are alternative solution techniques similar to Runge-Kutta.

5. Data Regression

The most extensively used statistical technique is regression analysis, which involves detecting, evaluating, and analyzing the connection between the dependent and independent variables [31]. As mentioned earlier, the regression window in POLYMATH is divided into 3 tabs, namely, 'Regression', 'Analysis' and the 'Graph' tab. User-defined linear and multiple linear regression was carried out in Problem Set 8 and 9 to determine unknown constants, however, using the nonlinear function definition provided by POLYMATH. Regressed graph of the calculated and experimental data of Problem 8 is shown in Figure 3.

In chemical engineering, regression tools are widely used to determine unknown constant parameters in Arrhenius equation, Antoine equation, microbial growth models, EOS parameters, rate equations, biogas kinetic models and many other proposed models in the literature. KaleidaGraph program (Synergy software), PAST, SigmaPlot, SAS, EViews, GraphPad Prism, SPSS, Matrixer, SHAZAM, Statistica, ORIGINPro, Datafit, JASP, PSPP, Xlstat Excel Add-in, Excel Solver, Jamovi, Minitab, Stata, KyPlot, Scilab, nonlinear curve fitting toolbox of MATLAB and NCSS are competitive softwares for regression analysis apart from POLYMATH [32, 33].

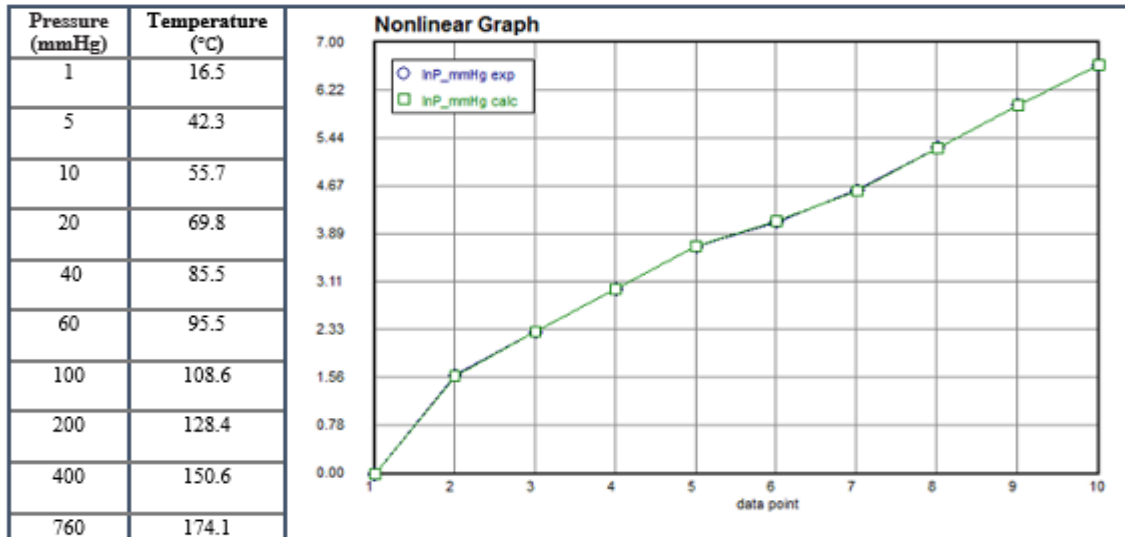


Figure 3: Experimental Vapor Pressure for N-Decane with Regressed Curve [8]

We have demonstrated how to use POLYMATH to carryout numerical differentiation, interpolation, integration as well as data plots for Problem Set 10. Interpolation, formerly can be described as a method of estimating missing data values that are in-between a given set of field data [34]. Typical example is approximating of function values (e.g. temperature) in heat transfer and/or thermodynamics at points where there is no data reading and when computing derivatives from empirical data values.

In chemical engineering, regression tools are widely used to determine unknown constant parameters in Arrhenius equation, Antoine equation, microbial growth models, EOS parameters, rate equations, biogas kinetic models and many other proposed models in the literature. KaleidaGraph program (Synergy software), PAST, SigmaPlot, SAS, EViews, GraphPad Prism, SPSS, Matrixer, SHAZAM, Statistica, ORIGINPro, Datafit, JASP, PSPP, Xlstat Excel Add-in, Excel Solver, Jamovi, Minitab, Stata, KyPlot, Scilab, nonlinear curve fitting toolbox of MATLAB and NCSS are competitive softwares for regression analysis apart from POLYMATH [32, 33]. We have demonstrated how to use POLYMATH to carryout numerical differentiation, interpolation, integration as well as data plots for Problem Set 10. Interpolation, formerly can be described as a method of estimating missing data values that are in-between a given set of field data [34]. Typical example is

approximating of function values (e.g. temperature) in heat transfer and/or thermodynamics at points where there is no data reading and when computing derivatives from empirical data values.

To show how one quantity vary with another related quantity, graphs are readily used to convey information to readers than if same information are explained in words [35]. Graphs can also be plotted in POLYMATH just like Excel as depicted in Figure 1. However, POLYMATH cannot be used to plot a log-log scale. Log-log graphs is a scatterplot in which both axes have a logarithmic scale [36]. A log-log plot displays the relationship, such as one in Problem Set 9 ($Tk^a = b$) as a straight-line equivalent to " $\ln T = \ln b - a \ln k$ ". The Bode plot, a graphical depiction of the frequency responses of a system in control and electrical engineering is a log-log plot [37]. The idea behind this type of plot according to [36] is to use the axes so we can more easily see details for small and large values of the dependent variable, and to show % change or multiplicative factors. Log-log scales are plotted very easily in Excel for an existing scatter plot by clicking on the "Logarithmic Scale" check box under Format Axis after right-clicking on the values along the x-axis. This is then repeated for the y-axis.

6. Linear Equations

Not one, but sets of simultaneous linear equations are solved using numerical analysis softwares. Methods

of solution are usually, Cramer's rule, substitution method, elimination method [35], graphical, and adjoint method which requires painstaking steps if the number of equations is more than 3. The fastest tool of solving a system of simultaneous linear equations is POLYMATH. In chemical engineering, simultaneous equations are used to balance chemical equations, finding unknowns in material balances, constrained optimization problems of differential equations after applying Laplace transform, solving certain functional equations among others. In real life, a smartphone user can use system of linear equations to select, best cell phone data plan knowing the charge rate of different companies per minute [38]. Also, to determine if one will make more money at one job or another, taking multiple variables (e.g. salary, benefits and commissions) into account [38]. POLYMATH cannot solve sets of linear inequality constraint equations used in cost optimization in chemical plants [39].

7. Conclusion

POLYMATH provisions for solutions to numerical problems makes it a useful chemical engineering analysis tool. An Android application called PolyMathLite had been created to further go into user's smart phones. The software package is limited to not possessing the features for solving PDEs in engineering. Unlike Excel, Tora, Lingo and sometimes C++ programs, it cannot solve linear programming problems, involving cost minimization or maximization in chemical plants. POLYMATH has the advantage of solving ODEs by selecting the best step size necessary while the use of programming codes would take much time and lengthy iterations before arriving at a solution and sometimes crash a system if 'h' is very negligible.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors acknowledge the support given by Federal University Wukari as well as the University of Maiduguri.

References

- [1] A. Brenner, M. Shacham, M. B. Cutlip, "Applications of mathematical software packages for modelling and simulations in environmental engineering education," *Environmental Modelling & Software*, vol. 20, pp. 1307-1313, 2005, doi:10.1016/j.envsoft.2004.09.007.
- [2] M. B. Cutlip, M. Shacham, *Problem solving in chemical and biochemical engineering with POLYMATH, Excel, and MATLAB*, 2nd ed. (Prentice Hall International Series, 2008).
- [3] H. S. Fogler, M. B. Cutlip, "Chemical Reaction Engineering (CRE) education: From the era of slide rule to the digital age." www.fdcocuments/in . (accessed: 30-Sep-2021).
- [4] M. B. Cutlip, M. Shacham, "Polymath solutions to the chemical engineering demonstration problem set." .
- [5] R. Hesketh, "Application of numerical problem solving in chemical engineering coursework," *ASEE 2017 Chemical Engineering Summer School*, pp. 1-6, 2017.
- [6] M. Shacham, N. Brauner, "Solving a system of nonlinear algebraic equations: You only get error messages-What to do next?," *Chemical Engineering Education*, vol. 15, no. 2, pp. 75-82, 2017.
- [7] M. Shacham et al., "Can I trust this software package? An exercise in validation of computational results," *Chemical Engineering Education*, vol. 42, no. 1, pp. 53-59, 2008.
- [8] Hashim Khan, "Polymath tutorial for Chemical Engineers." : 1-4, 2005.
- [9] M. Shacham, M. B. Cutlip, "User-Friendly numerical analysis programs." <http://www.polymath-software.com> . (accessed: 27-Sep-2021).
- [10] G. K. Rajahmundry et al., *Statistical analysis of adsorption isotherm models and its appropriate selection*, vol. 276, (2021).
- [11] A. M. Abubakar, Z. Soltanifar, M. M. Kida, *Sensitivity analysis of kinetic growth model data: Monod equation* (Revista Rebibio, 2022).
- [12] M. B. Cutlip, "Enabling extensive numerical problem solving on smartphones and tablets," *ASEE International Forum*, pp. 1-10, 2014.
- [13] F. J. Keil, "Application of numerical methods in chemical process engineering," *Scientific Computing in Chemical Engineering*, pp. 1-19, 1996, doi:10.1007/978-3-642-80149-5_11.
- [14] V. J. Law, *Numerical methods for Chemical Engineers using Excel, VBA, and MATLAB* (New York: CRC Press, 2013).
- [15] S. C. Chapra, R. P. Canale, *Numerical methods for engineers*, 6th ed. (New York, United States: McGraw Hill Companies, 2020).
- [16] CheGuide, "Curve fitting data to an Antoine equation-A guide for Chemical Engineers working in process industry," in *Chemical Engineer's Guide (CheGuide)*, (2015).
- [17] UMD, "Regression practice problems," in *Practical Numerical Methods*, (University of Minnesota Duluth, 2021), 30-36.
- [18] A. M. Abubakar, A. A. Mustapha, "Newton's method cubic equation of state C++ source code for iterative volume computation," *SSRG International Journal of Recent Engineering Science*, vol. 8, no. 3, pp. 12-22, 2021, doi:10.14445/23497157/IJRES-V8I3P103.
- [19] J. G. Eberhart, "Solving nonlinear simultaneous equations by the method of successive substitution: Applications to equations of state," *Journal of Chemistry Education*, vol. 71, no. 12, pp. 1038-1040, 1994, doi:https://doi.org/10.1021/ed071p1038.
- [20] W H Press et al., "Attempting to solve the systems of equations," in *Numerical Recipes*, 2nd ed. , (Cambridge, United States: Cambridge University Press, 1992), 40.

- [21] A. Ramli, M. L. Abdullah, M. Mamat, *Broyden's method for solving Fuzzy nonlinear equations* (Hindawi, 2010).
- [22] Cheng Yung Ming, "Solution of differential equations with applications to engineering problems," (IntechOpen, 2017), 33, doi:<http://dx.doi.org/10.5772/67539>.
- [23] B. Sumithra, "Engineering applications of differential equations," *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 6, no. 7, pp. 110–114, 2017.
- [24] P Rees, O Kryvchenkova, and P D Ledger, "Engineering Analysis 2: Ordinary Differential Equations." College of Engineering, Swansea University, United Kingdom, : 1–37, 2017.
- [25] C. B. Mohod, "Application of Ordinary Differential Equations (ODEs) in reaction engineering: Review," *International Journal of Chemical and Molecular Engineering*, vol. 2, no. 1, pp. 1–9, 2016.
- [26] John H Mathews and Kurtis K Fink, "Runge-Kutta Methods," in *Numerical Methods Using Matlab*, 4th ed. , (New Jersey: Prentice-Hall Inc, 2004), 497–499.
- [27] C. F. Mayo, "Implementation of the Runge-Kutta-Fehlberg method for solution of ordinary differential equations on a parallel processor," (Naval Postgraduate School (NPS)-Dudley Knox Library, 1987).
- [28] L. Derr, C. Outlaw, D. Sarafyan, "A new method for derivation of continuous Runge-Kutta formulas," *Computers Math. Application*, vol. 26, no. 3, pp. 7–13, 1993.
- [29] W. Suryaningrat, R. Ashgi, S. Purwani, "Order Runge-Kutta with extended formulation for solving ordinary differential equations," *International Journal of Global Operations Research (IJGOR)*, vol. 1, no. 4, pp. 160–167, 2020.
- [30] Jim Lambers, "Stiff differential equations: Lecture 9 notes," in *MAT 461/561*, (2010), 1–4.
- [31] J. O. Rawlings, S. G. Pantula, D. A. Dickey, *Applied regression analysis: A research tool*, 2nd ed. (Carolina, USA: Springer, 1998).
- [32] J. Tellinghuisen, "Using least squares to solve systems of equations." Tennessee, USA, 2015.
- [33] P. Rani et al., *Experimental and kinetic studies on co-digestion of agrifood and sewage sludge for biogas production*, vol. 16, no. 1, (Taylor and Francis Group, 2022).
- [34] C. R. Chikwendu, H. K. Oduwole, S. I. Okoro, *An application of spline and piecewise interpolation to heat transfer (cubic case)*, vol. 5, no. 6, (IISTE Knowledge Sharing Partners, 2015).
- [35] J. Bird, *Basic engineering mathematics*, 5th ed. (United Kingdom: Elsevier Ltd, 2010).
- [36] M. Bourne, "Graphs on logarithmic and semi-logarithmic axes." <https://www.intmath.com/exponential-logarithmic-functions/7-graphs-log-semilog.php> . (accessed: 06-Oct-2021).
- [37] Y. Bavafa-Toosi, "Introduction to linear control systems," *SIAM Review*, vol. 61, no. 1, pp. 209–225, 2019, doi:<https://doi.org/10.1137/19N974750>.
- [38] M. H. Snyder, "10 ways simultaneous equations can be used in everyday life." www.sciencing.com/10-can-used-everyday-life-8710568.html . (accessed: 04-Oct-2021).
- [39] A. M. Abubakar et al., "Simplex C++ syntax for solving chemical engineering cost optimization problems," *Research Inventy: International Journal of Engineering And Science*, vol. 11, no. 7, pp. 39–47, 2021, doi:10.5281/zenodo.5146856.

Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



ABDULHALIM MUSA ABUBAKAR bagged his B.Eng. (2018) Chemical Engineering from University of Maiduguri-Nigeria.

He works as research assistant with Modibbo Adama University-Nigeria. He has carried out researches in the areas of Linear Optimization, Petroleum and Reservoir Engineering. He is also interested in programming application in chemical engineering. He is a member of International Association of Engineers (IAENG), Graduate Member of the Nigerian Society of Engineers (NSE) and an affiliate member of Onshore Offshore Oil and Gas Professionals (O3GP). He had several other qualifications in areas of soft skill, data science with POLYMATH, ORIGIN software, SPSS and Excel, C++ Programming, Project Management, Webb Development, Information Security and had undergone a Professional Course in Safety Management and the use of AutoCAD 2D. He had 6 publications, two of which is related to the application of C++.



BELLO ILIYASU has bagged a Bachelor of Science (B.Sc.) Degree in Mathematics at a reputed Nigerian University. He had studied at Federal College of Education, Yola, Adamawa State.

He works as research fellow at Modibbo Adama University (MAU) being an experienced researcher. He is currently working on various research papers.



ZAKIYU MUHAMMAD SARKINBAKA is a research assistant at Federal University Wukari (FUWukari), Taraba-Nigeria. He is a graduate of Chemical Engineering from the University of Maiduguri.

He is an experienced high school teacher before his current job at FUWukari. He enrolled for a Masters Program in Chemical Engineering in 2019 in the same university. He had published three articles in reputed journals. Two among his publications are titled: "Neural Network Based Performance Evaluation of a Waterflooded Oil Reservoir" and "Multivariate Optimization of a Jacketed Heating System: A Generic Algorithm Approach".

Copyright: This article is an open access article distributed under the terms and conditions of the

Appendix

Figure 4 is a Dev C++ 5.11 coding for the Runge-Kutta-Fehlberg (RKF45) Method of order 4 and 5.

```

RKF45 Method.cpp  Untitled1.cpp
1  /* RUNGE-KUTTA-FEHLBERG (RKF45) METHOD */
2  /* PROBLEM SET 6 */
3  #include <iostream>
4  #include <math.h>
5  #include <iomanip>    // For setw, setprecision & setfill
6  using namespace std;
7
8  #define f(t,y) (3*Q/A)*pow(sin(t),2)-0.4167 // Defining ODE to be solved "Problem Set 6"
9  int main()
10 {
11     float t0, y0, tn, h, k1, k2, k3, k4, k5, k6, y_4, y_5;
12     int i, n;
13     int Q=500, A=1200;    // Defining Flowrate (Q) and Area (A)
14     cout<<"RUNGE-KUTTA-FEHLBERG (RKF45) METHOD"<<endl<<endl;
15     cout<<"Enter Initial Condition: t = ";
16     cin>>t0;
17     cout<<setw(29)<<"y = ";
18     cin>>y0;
19     cout<<"\n"<<"Calculation point, tn = ";
20     cin>>tn;    // Stopping value of t
21     cout<<"Suitable Step Size, h = ";
22     cin>>h;
23     cout<<endl;
24     n=(tn-t0)/h;    // Number of Iterations, n
25
26     cout<<setw(3)<<"i"<<setw(10)<<"t(i)"<<setw(11)<<"k1"<<setw(10)<<"k2"<<setw(13)<<"k3"
27     <<setw(15)<<"k4"<<setw(16)<<"k5"<<setw(17)<<"k6"<<setw(18)<<"y(i)4"<<setw(20)<<"y(i)5"<<endl;
28
29     for (i=1; i<=n; i++)
30     {
31         k1=h*(f(t0,y0));    // RKF45 parameters: k1, k2, k3, k4, k5 & k6
32         k2=h*(f((t0+h/4),(y0+k1/4)));
33         k3=h*(f((t0+3*h/8),(y0+3*k1/32+9*k2/32)));
34         k4=h*(f((t0+12*h/13),(y0+1932*k1/2197-7200*k2/2197+7296*k3/2197)));
35         k5=h*(f((t0+h),(y0+439*k1/216-8*k2+3680*k3/513-845*k4/4104)));
36         k6=h*(f((t0+h/2),(y0-8*k1/27+2*k2-3544*k3/2565+1859*k4/4104-11*k5/40)));
37         y_4=y0+(25*k1/216)+(1408*k3/2565)+(2197*k4/4104)-k5/5;    // RKF45 METHOD of ORDER 4
38         y0=y_4;
39
40         y_5=y0+(16*k1/135)+(6656*k3/12825)+(28561*k4/56460)-(9*k5/50)+(2*k6/55); // RKF45 METHOD of ORDER 5
41         tn=t0+h;    // Time interval equation
42         t0=tn;
43         cout<<setw(3)<<i<<setw(10)<<t0<<setw(11)<<k1<<setw(10)<<k2<<setw(13)<<k3<<setw(15)
44         <<k4<<setw(16)<<k5<<setw(17)<<k6<<setw(18)<<y0<<setw(20)<<y_5<<endl;
45     }
46     cout<<"\n"<<"Value of y RKF45 of ORDER 4 after "<<n<<" iterations, at t = "<<tn<<" is "<<y_4<<endl;
47     cout<<"\n"<<"Value of y RKF45 of ORDER 5 after "<<n<<" iterations, at t = "<<tn<<" is "<<y_5;
48     return 0;
    }
    
```

Figure 4: Problem Set 6 Coded C++ Solution

As explained earlier the RKF45 method as used in POLYMATH selects the appropriate step size on its own for the solution of a given problem. Entering a very

minute step size in C++ execution of Problem Set 6 will display a lengthy table of calculation before given the result, simply because, the smaller the value of h, the

higher the number of iterations. This is the reason why, the execution of the code in Figure 4, for a step size of $h = 0.1$ gave $y = 0.604765m$ and $0.58828m$ (for RKF45 of order 4 & 5 respectively) in Figure 5, which is not the same with the value obtained by POLYMATH and Excel, which are $y = 1.798038m$. The step size used by POLYMATH for Problem Set 6 is $h = 0.000001$. Despite that, the C++ result cannot be discarded because it falls between the minimal and maximal value given by POLYMATH in Figure 6.

Note that varying the step size, h , during C++ execution of the problem will definitely affect the result. A step size of $h = 0.001, 0.01$ and 0.5 gives 0.604776 & $0.604656, 0.604736$ & 0.603485 and 0.604765 & $0.455666m$ for RKF45 of order 4 and 5 respectively. Therefore, in chemical engineering where time will not permit the use of C++ codes due to the 'h' myths, POLYMATH have an edge in this case, over C++ and other coding software programs for solving ODEs by Runge-Kutta-Fehlberg method.

```

RUNGE-KUTTA-FEHLBERG (RKF45) METHOD
Enter Initial Condition: t = 0
                      y = 0
Calculation point, tn = 10
Suitable Step Size, h = 0.1
    
```

i	t(i)	k1	k2	k3	k4	k5	k6	y(i)4	y(i)5
1	0.1	-0.04167	-0.0416075	-0.0415294	-0.0408203	-0.0406733	-0.0414202	-0.0413373	-0.0826637
2	0.2	-0.0406733	-0.0401156	-0.0397913	-0.0380171	-0.037723	-0.0394368	-0.0806946	-0.120042
3	0.3	-0.037723	-0.0366924	-0.0361346	-0.0333662	-0.0329368	-0.0335491	-0.116171	-0.151638
4	0.4	-0.0329368	-0.0314742	-0.0307053	-0.027053	-0.0265053	-0.0299121	-0.146019	-0.17586
5	0.5	-0.0265053	-0.0246692	-0.0237198	-0.0193292	-0.0186851	-0.0227505	-0.168718	-0.191411
6	0.6	-0.0186851	-0.0165486	-0.0154566	-0.0105027	-0.00978789	-0.0143498	-0.18303	-0.197339
7	0.7	-0.00978789	-0.00743612	-0.00624508	-0.000925358	-0.000168352	-0.00504494	-0.188052	-0.193075
8	0.8	-0.000168352	-0.00230486	0.00354751	0.00902093	0.00978998	0.00479315	-0.183253	-0.178457
9	0.9	0.00978998	0.012286	0.0135308	0.0189397	0.0196901	0.0147722	-0.168492	-0.153735
10	1	0.0196901	0.0221095	0.0233067	0.0284355	0.0291374	0.0244945	-0.144024	-0.119564
11	1.1	0.0291374	0.0313836	0.0324855	0.0371297	0.0377551	0.0335723	-0.110494	-0.0769738
12	1.2	0.0377551	0.0397387	0.0407014	0.0446758	0.0451997	0.0416438	-0.0689058	-0.0273295
13	1.3	0.0451997	0.0468416	0.0476267	0.0507729	0.0511744	0.0483872	-0.0205853	0.0277215
14	1.4	0.0511744	0.0524091	0.0529853	0.0551779	0.0554411	0.0535336	0.032873	0.0863165
15	1.5	0.0554411	0.0562194	0.0565637	0.0577152	0.0578296	0.0568779	0.08967	0.146452
16	1.6	0.0578296	0.0581204	0.0582192	0.0582837	0.0582447	0.0582868	0.147874	0.206061
17	1.7	0.0582447	0.0580365	0.0578857	0.0568607	0.0566699	0.057704	0.205495	0.263102
18	1.8	0.0566699	0.0559709	0.0555766	0.053503	0.0531679	0.0551528	0.26057	0.315631
19	1.9	0.0531679	0.052006	0.051384	0.0483443	0.0478784	0.050735	0.311235	0.361886
20	2	0.0478784	0.0462999	0.0454749	0.0415904	0.0410122	0.0446266	0.355801	0.400356
21	2.1	0.0410122	0.0390801	0.0380849	0.0335105	0.032843	0.0370712	0.392824	0.429838
22	2.2	0.032843	0.0306344	0.0295088	0.0244268	0.0236966	0.02837	0.42116	0.44949
23	2.3	0.0236966	0.0212994	0.0200883	0.0147013	0.0139076	0.0188698	0.440013	0.458861
24	2.4	0.0139076	0.0114474	0.010199	0.00472181	0.00395507	0.00894944	0.448961	0.457908
25	2.5	0.00395507	0.00147115	0.000235257	-0.00511382	-0.00585309	-0.0009956	0.447981	0.447002
26	2.6	-0.00585309	-0.00823168	-0.0094058	-0.0144135	-0.0150958	-0.0105689	0.437444	0.42691
27	2.7	-0.0150958	-0.0172743	-0.0183398	-0.0228065	-0.0234046	-0.0193887	0.418101	0.398765
28	2.8	-0.0234046	-0.0252961	-0.0262105	-0.0299581	-0.0304483	-0.0271035	0.391057	0.36402

29	2.9	-0.0304483-0.0319773	-0.0327043	-0.0355833	-0.0359459	-0.0334056	0.357721	0.324394
30	3	-0.0359459-0.0370516	-0.0375621	-0.0394579	-0.0396785	-0.0380439	0.319754	0.281798
31	3.1	-0.0396785-0.0403167	-0.0405904	-0.0414273	-0.0414971	-0.0408334	0.279003	0.238262
32	3.2	-0.0414971-0.0416425	-0.0416683	-0.041413	-0.0413293	-0.0416629	0.237423	0.195854
33	3.3	-0.0413293-0.0409759	-0.040753	-0.0394157	-0.0391817	-0.0404994	0.197005	0.156597
34	3.4	-0.0391817-0.0383438	-0.0378809	-0.0355148	-0.0351399	-0.0373892	0.159692	0.122388
35	3.5	-0.0351399-0.0338508	-0.0331666	-0.0298661	-0.0293652	-0.0324563	0.127303	0.09423
36	3.6	-0.0293652-0.0276763	-0.0267978	-0.0226945	-0.0220877	-0.0258974	0.101463	0.0756287
37	3.7	-0.0220877-0.0200663	-0.0190287	-0.0142862	-0.0135975	-0.017974	0.0835328	0.0656065
38	3.8	-0.0135975-0.0113242	-0.0101689	-0.00497615	-0.00423314	-0.0090019	0.0745598	0.0655881
39	3.9	-0.00423314-0.00179864	-0.000571533	-0.00486433	-0.00563207	0.000661156	0.0752337	0.0759064
40	4	0.00563207-0.0013077	0.00938067	0.014843	0.0156048	0.0106299	0.0858599	0.096482
41	4.1	0.0156048-0.0180681	0.019291	0.024562	0.0252876	0.020507	0.106347	0.126827
42	4.2	0.0252876-0.0276173	0.0287644	0.0336339	0.0342943	0.0298987	0.136209	0.166063
43	4.3	0.0342943-0.0363975	0.037423	0.041697	0.0422659	0.0384305	0.17459	0.212959
44	4.4	0.0422659-0.0440587	0.0449219	0.0484298	0.0488845	0.0457622	0.22029	0.265976
45	4.5	0.0488845-0.0502956	0.0509619	0.053564	0.0538864	0.0516017	0.271819	0.323334
46	4.6	0.0538864-0.0548594	0.0553024	0.0568949	0.0570721	0.055716	0.327456	0.383078
47	4.7	0.0570721-0.0575682	0.0577702	0.0582897	0.0583146	0.0579412	0.385315	0.443158
48	4.8	0.0583146-0.0583141	0.058267	0.0576927	0.0575644	0.0581886	0.44342	0.50151
49	4.9	0.0575644-0.0570673	0.0567729	0.0551278	0.0548514	0.0564483	0.499789	0.556142
50	5	0.0548514-0.0538775	0.0533476	0.0506972	0.0502837	0.0527897	0.552504	0.605206
51	5.1	0.0502837-0.0488718	0.0481276	0.0445776	0.0440435	0.0473585	0.599798	0.647079
52	5.2	0.0440435-0.0422499	0.041321	0.0370129	0.0363794	0.0403715	0.640116	0.680424
53	5.3	0.0363794-0.0342757	0.0331992	0.0283047	0.0275972	0.0321071	0.672183	0.704243
54	5.4	0.0275972-0.0252671	0.0240858	0.0188002	0.0180468	0.0228947	0.695054	0.717919
55	5.5	0.0180468-0.0155833	0.0143444	0.00887822	0.008109	0.0131017	0.708148	0.721239
56	5.6	0.008109-0.00561033	0.00436318	-0.00106556	-0.00181995	0.00311846	0.711275	0.714402
57	5.7	-0.00181995-0.00425421	-0.00545989	-0.0106348	-0.0113443	-0.006657	0.704643	0.698014
58	5.8	-0.0113443-0.0136171	-0.0147332	-0.0194479	-0.0200842	-0.015835	0.688848	0.673059
59	5.9	-0.0200842-0.0221049	-0.0230871	-0.0271537	-0.0276914	-0.0240496	0.664852	0.640864
60	6	-0.0276914-0.0293795	-0.0301884	-0.0334448	-0.0338625	-0.0309733	0.633945	0.603046
61	6.1	-0.0338625-0.0351507	-0.0357542	-0.038705	-0.0383516	-0.0363302	0.597689	0.561443
62	6.2	-0.0383516-0.0391884	-0.0395625	-0.0408463	-0.0409796	-0.0399065	0.557863	0.518047
63	6.3	-0.0409796-0.0413218	-0.0414614	-0.0416617	-0.0416617	-0.0415599	0.516386	0.47492
64	6.4	-0.0416617-0.0414953	-0.0413753	-0.040484	-0.0403117	-0.0412243	0.475244	0.434113
65	6.5	-0.0403117-0.0396724	-0.0393077	-0.0373603	-0.0370425	-0.0389131	0.43641	0.397585

66	6.6	-0.0370425-0.0359358	-0.0353408	-0.0324151	-0.0319644	-0.0347185	0.401763	0.367125
67	6.7	-0.0359358-0.0304345	-0.029633	-0.0258455	-0.02528	-0.0288077	0.373917	0.344278
68	6.8	-0.02528-0.0233878	-0.0224117	-0.0179134	-0.0172556	-0.0214163	0.35165	0.330288
69	6.9	-0.0172556-0.0150766	-0.0139648	-0.00893507	-0.00821114	-0.0128391	0.338846	0.326045
70	7	-0.00821114-0.00583221	-0.00462912	-0.000731546	-0.00149271	-0.00341788	0.335448	0.33205
71	7.1	0.00149271-0.00397675	0.00522322	0.0107011	0.0114691	0.00647167	0.341923	0.348395
72	7.2	0.0114691-0.0139593	0.0151994	0.0205761	0.0213204	0.0164353	0.358345	0.374761
73	7.3	0.0213204-0.0237174	0.0249017	0.0299629	0.0306538	0.0260758	0.384391	0.410429
74	7.4	0.0306538-0.032862	0.0339434	0.0384873	0.0390973	0.0350088	0.419355	0.454309
75	7.5	0.0390973-0.0410286	0.041964	0.0458094	0.0463141	0.0428783	0.462176	0.504984
76	7.6	0.0410286-0.0478917	0.0486436	0.0516373	0.0520166	0.0493704	0.511478	0.560766
77	7.7	0.0478917-0.0531775	0.0537161	0.0573787	0.0559775	0.0542264	0.565628	0.619763
78	7.79999	0.0559775-0.0566755	0.0569792	0.0579501	0.0580388	0.0572526	0.622799	0.679954
79	7.89999	0.0566755-0.058246	0.0583028	0.0581832	0.0581184	0.0583284	0.681044	0.739273
80	7.99999	0.058246-0.0578266	0.0576342	0.0564289	0.0562131	0.057411	0.738373	0.795687
81	8.09999	0.0578266-0.0554338	0.0549999	0.052757	0.0523989	0.0545369	0.792833	0.847279
82	8.2	0.0523989-0.0511633	0.0505051	0.047314	0.0468277	0.0498206	0.842585	0.892323
83	8.3	0.0468277-0.045185	0.0443288	0.0403167	0.0397218	0.0434502	0.885976	0.929357
84	8.4	0.0397218-0.0377375	0.0367174	0.0320443	0.0313643	0.0356797	0.92161	0.957236
85	8.5	0.0377375-0.0291176	0.0279743	0.0228264	0.0220885	0.0268188	0.948398	0.97518
86	8.6	0.0291176-0.0196689	0.018448	0.0130306	0.0122643	0.0172208	0.965604	0.982807
87	8.7	0.0196689-0.00976816	0.00851828	0.0030474	0.00228312	0.00726838	0.972875	0.980144
88	8.8	0.00976816-0.000189899	-0.0014189	-0.00672521	-0.00745693	-0.00264172	0.970251	0.967629
89	8.9	0.000189899-0.0008083	-0.0109674	-0.0158976	-0.0165676	-0.0121144	0.958171	0.946095
90	9	-0.0158976-0.0187036	-0.0197466	-0.0241041	-0.0246857	-0.020772	0.937447	0.91673
91	9.1	-0.0246857-0.0265211	-0.0274065	-0.0310176	-0.0314875	-0.0282695	0.909239	0.881039
92	9.2	-0.0310176-0.0329492	-0.0336416	-0.0363624	-0.036702	-0.0343078	0.875002	0.840775
93	9.3	-0.0329492-0.0377316	-0.0382035	-0.0399254	-0.0401211	-0.0386462	0.836434	0.797877
94	9.4	-0.0401211-0.0406777	-0.0409102	-0.0415646	-0.0416086	-0.0411119	0.795404	0.754386
95	9.5	-0.0406777-0.04167	-0.04167	-0.0412147	-0.0411052	-0.0416064	0.753881	0.712369
96	9.6	-0.04167-0.0406689	-0.0404047	-0.0388895	-0.038631	-0.0401101	0.713852	0.673833
97	9.7	-0.0406689-0.0377144	-0.0372127	-0.0346819	-0.0342846	-0.0366827	0.677244	0.640646
98	9.8	-0.0377144-0.0329242	-0.0322051	-0.0287595	-0.0282392	-0.0314607	0.64585	0.614463
99	9.9	-0.0329242-0.0264893	-0.0255814	-0.0213584	-0.020736	-0.0246524	0.621252	0.596661
100	10	-0.0264893-0.0186663	-0.0176058	-0.0127738	-0.0120739	-0.0165291	0.604765	0.58828

value of y RKF45 of ORDER 4 after 100 iterations, at t = 10 is 0.604765

value of y RKF45 of ORDER 5 after 100 iterations, at t = 10 is 0.58828

Figure 5: RKF45 C++ Output Results of Problem Set 6

The continuity equation in Problem Set 6 is a single first order differential equation. Column 2 of Table 2 gives the solution of the Monod Model Kinetic equation, often used to explain growth and decay rate of microbes in bioreaction engineering. Problem 7, though solved by RKF56 method in POLYMATH, differs from Problem 6, being a multiple first order differential equation solved simultaneously. Table 2: POLYMATH Result of Problem Set 6 and 7

Problem Set 6

d(e) x= ini-
finl i → RKF45

Differential Equations: 1 | Auxiliary Equations: 2 | Ready for solution

```
# CONTINUITY EQUATION OF LIQUID FLOW
d(y) / d(t) = 3*(Q/A)*sin(t)*sin(t)-(Q/A)

# Explicit Equation
Q = 500      # Flowrate (m3/day)
A = 1200     # Area (m2)

# Time Range (0 to 10 days)
t(0) = 0     # Initial Time (days)
t(f) = 10    # Final Time (days)

# Initial Condition
y(0) = 0     # Depth of the liquid (m)
```

POLYMATH Report
Ordinary Differential Equations

Calculated values of DEQ variables

Variable	Initial value	Minimal value	Maximal value	Final value
1 A	1200.	1200.	1200.	1200.
2 Q	500.	500.	500.	500.
3 t	0	0	10.	10.
4 y	0	-0.1645208	2.12963	1.798038

Differential equations

1 $d(y)/d(t) = 3*(Q/A)*\sin(t)*\sin(t)-(Q/A)$

Explicit equations

1 Q = 500
Flowrate (m3/day)

2 A = 1200
Area (m2)

Problem Set 7

d(e) x= ini-
finl i → RKF56

Differential Equations: 2 | Auxiliary Equations: 4 | Ready for solution

```
# MONOD MODEL: BIOREACTION KINETICS
d(s) / d(t) = -(k*s*x)/(ks+s)
d(x) / d(t) = y*(k*s*x)/(ks+s)-b*x

# Explicit Equation
k = 5      # Maximum specific uptake rate of the substrate
b = 0.01   # Decay coefficient
y = 0.05  # Yield coefficient
ks = 20    # Half saturation constant for growth

# Initial Conditions
s(0) = 1000 # Growth limiting substrate concentration
x(0) = 100  # Biomass concentration
t(0) = 0    # Initial time
t(f) = 1    # Final time
```

POLYMATH Report
Ordinary Differential Equations

Calculated values of DEQ variables

Variable	Initial value	Minimal value	Maximal value	Final value
1 b	0.01	0.01	0.01	0.01
2 k	5.	5.	5.	5.
3 ks	20.	20.	20.	20.
4 s	1000.	452.5393	1000.	452.5393
5 t	0	0	1.	1.
6 x	100.	100.	126.2464	126.2464
7 y	0.05	0.05	0.05	0.05

Differential equations

1 $d(s)/d(t) = -(k*s*x)/(ks+s)$

2 $d(x)/d(t) = y*(k*s*x)/(ks+s)-b*x$

Experimental Methodology to Find the Center of Gravity of a Solid

Joohoon Je¹, Eunsung Jekal^{*2}

¹ Yaksa middle school, 10, Jongga 4-gil, Jung-gu, Ulsan, Republic of Korea

²Jekal's Laboratory, 471, Munsu-ro, Nam-gu, Ulsan, Republic of Korea

*Corresponding author: Eunsung Jekal, esjekal.jekalslab@gmail.com

ABSTRACT: The center of gravity of a three-dimensional object found through an experimental method can be made easier and faster than calculating the movement manually to make it look more natural in graphic computer images. In addition, in various sports such as skating, the score can be increased by appropriately moving the position of the center of gravity. Lastly, it is expected that it can be used even when the performance is high in the manufacturing process to increase the stability and speed of various means of transportation (eg, automobiles, airplanes, etc.).

KEYWORDS Center of gravity, three-dimensional object

1. Introduction

Tools or techniques using the center of gravity are deeply embedded in our daily lives [1, 2, 3]. For example, when creating animations, a motion generation technique that finds and moves the center of gravity of a character in order to provide a natural motion is considered an important task in the field of computer graphics [4, 5, 6, 7]. In addition, the traveling speed and fuel efficiency vary greatly depending on the load and the position of the center of gravity of the ship or airplane. In addition to this, as shown in Figure 1, there is also a simple toy that is balanced by the center of gravity when the beak is hung on the pedestal. In the case of this toy, it looks simple, but if one part breaks or other materials stick, it loses its center and collapses.



Figure 1: balancing bird model.

Therefore, this paper focused on the information of the center of gravity. The center of gravity information has enough advantages to be used as a parameter to determine the shape and motion of an object. First of all, the center of gravity can be expressed only in very small dimensions

[8, 9, 10, 11]. Only the position values of the x , y , and z axes are required to indicate the position of the center of gravity in the coordinate system. In addition, many motions can be predicted even by changing the position of the center of gravity. For example, in the case of a walking motion, assuming that the y -axis is perpendicular to the ground, the position values in the x and z planes change, and the center of gravity vibrates with a very small width in the y -axis direction. The running motion is similar to the walking motion, but the speed in the x and z planes is different. The change in y value increases in jumping and squatting motions. In this paper, in order to create a more natural result in addition to the center of gravity information, motion history information is additionally utilized to increase the accuracy of pose prediction through the center of gravity. In addition, it can be used to increase the score by adjusting the position of the center of gravity in sports, or to increase speed and stability by adjusting the center of gravity of transportation means such as ships and airplanes [12, 13, 14].

2. Task description and data construction

The center of gravity is the center point of an object's weight, and the center of mass is calculated based on the distribution of mass only. The difference is that the sum of the torques representing the action of the applied force to make it happen is zero. However, in special circumstances (such as the surface of the Earth) where gravity (mg) is constant, the center of gravity and the center of mass are the same. In everyday life, the center of gravity can be seen as equal to the balance point of an object. More specifically, if an object maintains its balance when a certain point is placed on a pointed pedestal, it can be said to be the center of gravity. Therefore, the center of gravity of a symmetrical object (sphere, regular polyhedron, etc.) with constant density

coincides with the geometric center of the object.

2.1. Center of gravity of the plane

Center of gravity of the plane triangle The center of gravity of a triangle is the intersection of the medians, which are lines connecting one vertex of a triangle and the midpoint of its opposite side as shown in Figure 2.

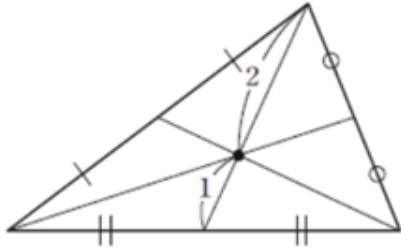


Figure 2: The center of gravity of a triangle.

For example, if the coordinates of the center of gravity G of a triangle ABC with three points as vertices are $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$, the center of the side BC is $M(x', y')$, $x' = \frac{x_2+x_3}{2}$, $y' = \frac{y_2+y_3}{2}$, then the center of gravity G(x,y) is a point that divides the segment AM by 2:1. Mathematical coordinate can be obtained as equation(1) and (2).

$$x = \frac{2 \times x' + 1 \times x_1}{2 + 1} = \frac{2 \times \frac{x_2+x_3}{2} + 1 \times x_1}{2 + 1} = \frac{x_1 + x_2 + x_3}{3} \tag{1}$$

$$y = \frac{2 \times y' + 1 \times y_1}{2 + 1} = \frac{2 \times \frac{y_2+y_3}{2} + 1 \times y_1}{2 + 1} = \frac{y_1 + y_2 + y_3}{3} \tag{2}$$

Therefore, the coordinates of the center of gravity G is expressed as $G(\frac{x_1+x_2+x_3}{3}, \frac{y_1+y_2+y_3}{3})$.

Center of gravity of the plane n-polygon First, the center of gravity of a rectangle is a point obtained by dividing the rectangle into two triangles along the diagonal of the rectangle, finding the center of gravity of the triangles, and dividing the line segment made by connecting the points according to the ratio of the area of the triangle (for example, if the triangle Assuming that the area of is 1 and 3, respectively, the 3:1 point of the line segment made by connecting the center of gravity of the triangle is the center of gravity of the rectangle). The center of gravity of a pentagon is divided along a diagonal like a square to make three triangles, and then the center of gravity of the triangle made by connecting the centers of gravity of the triangles is the center of gravity of the pentagon. Shapes other than quadrilaterals and pentagons are divided into triangles and have centers of gravity, and if you repeat dividing the shapes again, a rectangle or a pentagon comes out.

2.2. Center of gravity of the solid

Since the z-coordinate value is added under the assumption that the triangle is three-dimensionally expressed in the three-dimensional space, the coordinates of the center of gravity G of the triangle ABC with the vertex $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$, $C(x_3, y_3, z_3)$ can be expressed as

$G(\frac{x_1+x_2+x_3}{3}, \frac{y_1+y_2+y_3}{3}, \frac{z_1+z_2+z_3}{3})$. However, in the case of various three-dimensional figures outside the triangle category, it is difficult to find them without using the center of mass using the integral taught in university physics. Therefore, in this paper, we try to find the center of gravity of a three-dimensional figure in a relatively simple and intuitive way.

A point of mass placed at the vertex of a tetrahedron and the center of the inscribed sphere through mathematical proof A tetrahedron can be considered as an analogy of space for a triangle in a plane, and just as a triangle has an inscribed circle, a tetrahedron has an inscribed sphere.

Theorem Let I be the center of the inscribed sphere of the tetrahedron, and let J be the intersection of the straight line DI and the plane ABC. Then the following equation holds:

$$IJ = \frac{S_{JBC} \cdot IA}{S_{JBC}} = \frac{S_{JAC} \cdot IB}{S_{JAC}} = \frac{S_{JAB} \cdot IC}{S_{JAB}} \tag{3}$$

Proof. Construct the segments AA_1 and BB_1 from the vertices A and B of the triangle ABC through the point J. First, let's express CJ using JA and JB. To do this, construct a parallelogram CMJN by drawing a straight line parallel to JA and JB through point C. Then $CJ = CM = CN$.

At this time, since the triangles CA_1M and B_1AJ are similar, it becomes $\frac{CM}{BJ} = \frac{CA_1}{BA_1} = \frac{CN}{JA} = \frac{CB_1}{AB_1}$. However, the ratio of CA_1 and BA_1 is equal to the ratio of the areas of triangles CAA_1 and BAA_1 . In this case, since the areas of triangles CJA_1 and BJA_1 are the same, $\frac{CA_1}{BA_1} = \frac{S_{JAC}}{S_{JAB}}$. For the same reason, $\frac{CB_1}{AB_1} = \frac{S_{JBC}}{S_{JAB}}$.

In the end, $CM = \frac{S_{JAC}}{S_{JAB}} \cdot JB$ is obtained from the equations $\frac{CM}{BJ} = \frac{CA_1}{BA_1}$ and $\frac{CA_1}{BA_1} = \frac{S_{JAC}}{S_{JAB}}$, and it can be seen that $CN = \frac{S_{JBC}}{S_{JAB}} \cdot JA$ in a similar way. Since $CJ = CM = CN$, it is $CJ = \frac{S_{JAC}}{S_{JAB}} \cdot JB = \frac{S_{JBC}}{S_{JAB}} \cdot JA$, and it is $S_{JAB} \cdot CJ = S_{JAC} \cdot JB = S_{JBC} \cdot JA$.

Now, let's think $CJ = CI + IJ, JB = JI + IB, JA = JI + IA$, it would become $CJ = IJ + S_{JAC} \cdot JI, IB = S_{JBC} \cdot JI, IA = JI + S_{JAC} \cdot JI$. Considering that it is $JJ = -IJ$ on the right side of the equation, the equation is derived, considering that it is $CI = -IC$ on the left side. In theorem's equation, the area S of the triangles JAB, JBC, and JAC can be expressed using the areas $S_{JAB}, S_{JBC}, S_{JAC}$ of the sides DAB, DAC, and DBC of the tetrahedron. Consider the volumes of the tetrahedron DABJ, DBCH, and DACJ in the Figure 3. If their bases are called faces JAB and JBC JAC, respectively, the height of these tetrahedra is the same as h, so the volume of the tetrahedron is $\frac{S_{JAB} \cdot h}{3}, \frac{S_{JBC} \cdot h}{3}, \frac{S_{JAC} \cdot h}{3}$ respectively. On the other hand, if we take DAB, DAC, DBC as the base of these tetrahedrons, and consider the heights drawn from the point J to these surfaces, all of these heights will be the same (let the height be k), then the volume of the tetrahedron will be $\frac{S_{DAB} \cdot k}{3}, \frac{S_{DAC} \cdot k}{3}, \frac{S_{DBC} \cdot k}{3}$, respectively. Eventually we can get the equation $S_{JAB} = S_{DAB} \cdot \frac{k}{h}, S_{JBC} = S_{DBC} \cdot \frac{k}{h}, S_{JAC} = S_{DAC} \cdot \frac{k}{h}$ which becomes $\frac{S_{JAB} \cdot h}{3} = \frac{S_{DAB} \cdot k}{3}, \frac{S_{JBC} \cdot h}{3} = \frac{S_{DBC} \cdot k}{3}, \frac{S_{JAC} \cdot h}{3} = \frac{S_{DAC} \cdot k}{3}$. Now, by substituting the obtained equations into the equation of theorem and rearranging them, we can get equation(4).

$$IJ = \frac{S_{DBC} \cdot IA}{S_{DBC}} = \frac{S_{DAB} \cdot IC}{S_{DAB}} \tag{4}$$

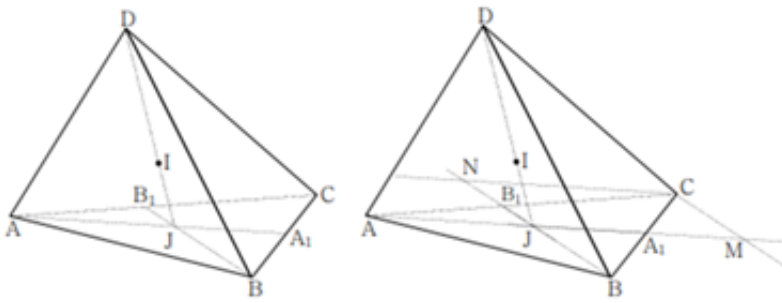


Figure 3: Triangle model for mathematical proof.



Figure 5: The dodecahedron model.

3. Experimental method

3.1. Regular polyhedron

3.1.1. tetrahedron

Hang the thread to the center of gravity of each side and leave it there until it stops shaking, then put a dot at the intersection of the hanging thread and the paper. After that, switch sides and repeat. At this time, the position of the paper and the multifaceted tooth should not be changed, only the hanging side should be changed. At this time, if all points coincide, the center of gravity is the midpoint of each side and the point of intersection of the lines connecting the extension line of the single thread to the point and the intersection of the surface.



Figure 4: The tetrahedron model.

3.1.2. dodecahedron

It is the same as the experiment method of the tetrahedron, but in the case of the dodecahedron, a string is hung at each vertex instead of the center of gravity of each face. This is because in the case of a tetrahedron, if a perpendicular is drawn from a vertex, it meets the center of gravity of the face, but in a regular dodecahedron, if a perpendicular is drawn from a vertex, it continues to the opposite vertex. Therefore, after connecting the starting point of the thread to one vertex and hanging it, find the place where the opposite vertices meet and make a dot on the paper fixed to the floor. At this time, if all points coincide, it can be seen that the intersection line connecting each vertex and the furthest vertex is the center of gravity.

3.2. Car model

First, the part to be leveled in the erected form is found and marked on the upper part of the car model. Then, after turning it 90 degrees to the side, find the part to be leveled in the same way on the side part of the car model. Finally, after turning forward 90 degrees, find the part that is leveled, connect the intersections of the points, and check whether the point coincides with the center of mass.



Figure 6: The car model.

Table 1: The group of the distance between each dots in the tetrahedron model.

Dot pairs	[1,2]	[1,3]	[1,4]	[2,3]	[2,4]	[3,4]
Distance (cm)	0	0.5	0.7	0.5	0.7	0.4

4. Results

As shown in Figure 7, considering that the distance distribution is gathered and distributed within a maximum of 4.5cm, it can be seen that the experiment was elaborated enough to ignore the error. In addition, the experimental results fit well with the assumptions made mathematically at the beginning.

In Figure 8, we represent each number of dots of top, back, right, left and front as 1,2,3,4 and 5.

According to $e^{-1/|\theta|}$, the attribution degree is 0.778, input e into the cloud generator, activate the comment "better" and "very good", and activate the degree of "very good" more than "better", then the performance evaluation result of this type of man-portable is "very good". Analysis and evaluation results can be obtained: protective index and self-measurement index cloud center of gravity vector deviation is large, indicating poor performance of the index, but the weight is low, the impact on the performance of the load is not significant, the improvement is not significant. The

Table 2: The group of the distance between each dots in the dodecahedron models.

Dot pairs	[1,2]	[1,3]	[1,4]	[1,5]	[1,6]	[1,7]	[1,8]	[1,9]	[1,10]
Distance (cm)	2.5	2.3	0.7	3.3	2.1	1.6	3.0	3.4	3.9
Dot pairs	[2,3]	[2,4]	[2,5]	[2,6]	[2,7]	[2,8]	[2,9]	[2,10]	[3,4]
Distance (cm)	1.6	2.5	1.3	1.9	0.8	0.8	1.2	1.6	1.9
Dot pairs	[3,5]	[3,6]	[3,7]	[3,8]	[3,9]	[3,10]	[4,5]	[4,6]	[4,7]
Distance (cm)	1.3	0.5	1.4	1.6	2.8	3.0	3.0	1.6	1.8
Dot pairs	[4,8]	[4,9]	[4,10]	[5,6]	[5,7]	[5,8]	[5,9]	[5,10]	[6,7]
Distance (cm)	3.1	3.1	4.1	1.8	1.9	1.7	1.9	2.1	1.7
Dot pairs	[6,8]	[6,9]	[6,10]	[7,8]	[7,9]	[7,10]	[8,9]	[8,10]	[9,10]
Distance (cm)	2.1	3.1	3.4	1.5	1.9	2.3	1.7	1.8	0.5

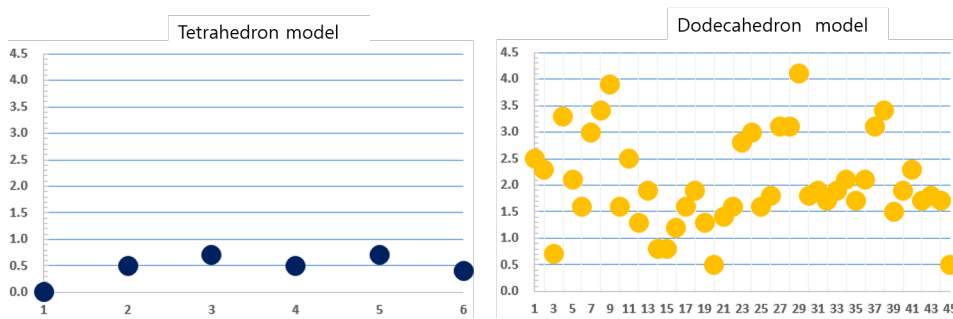


Figure 7: Distance distribution of the tetrahedron and the dodecahedron models.

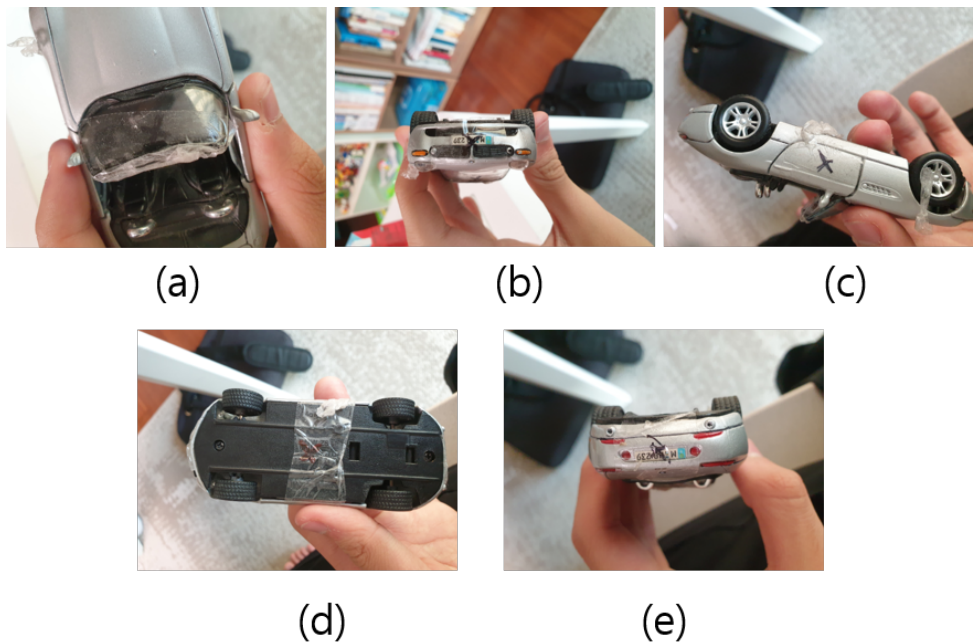


Figure 8: Expressed number of dots in the car model.

Table 3: The group of the distance between each dots in the car model.

Dot pairs	[1,2]	[1,3]	[1,4]	[1,5]	[2,3]	[2,4]	[2,5]	[3,4]	[3,5]	[4,5]
Distance (cm)	0.6	0.7	0.5	0.6	1.3	1.1	1.0	0.4	0.6	0.7

cloud center of gravity deviation of the three indicators of tactics, comfort, and manning effectiveness is similar to the system deviation, so the improvement of the performance of these three indicators can help this type of man-portable device improve its performance.

5. Conclusion

As mentioned in the introduction, these methods automatically calculate the motion of objects to make the motion in computer graphic images look more natural, making it easier and faster than manual work. In various sports such as skating, you can increase your score by moving the center of gravity appropriately. In addition, it can be used when the performance is high in the manufacturing process to increase the stability and speed of various means of transportation (eg, automobiles, airplanes, etc.).

6. Limitations

Our research presents the final design of a low-cost device with reliable and accurate performance. Though it would be interesting to compare our device with other similar ones, it is challenging to replicate the same device(s) and scenarios used by other researchers. Future research could involve a comparative study involving our device and other similar devices to assess their relative performance and usability and to gain further insights into the development of low-cost devices that can simulate the weight of virtual objects and changes in their center of gravity. Any electronic system communication can potentially have transmission delays. However, we did not calculate this delay because the time for communication was of a smaller magnitude than the time used for the system to be filled. Also, our system was faster than other reported state-of-the-art system.

In our experiments, the perception was based on a single sample from each participant. We did this because we observed in pilot studies that, after a few repetitions of the same trials, fatigue could kick in and could affect their responses. However, the results we had were consistent among themselves, showing errors were committed by a similar number of participants, and independent participants often chose similar answers. As such, to a large extent, our approach is valid, as the performance of the final complete prototype linked to the VR application shows.

References

- [1] H. Uitenbroek, "The accuracy of the center-of-gravity method for measuring velocity and magnetic field strength in the solar photosphere", *The Astrophysical Journal*, vol. 592, no. 2, p. 1225, 2003, doi:<https://doi.org/10.1086/375736>.
- [2] A. Üsküdar, Y. S. Türkan, Y. S. Özdemir, A. H. Öz, "Fuzzy ahp-center of gravity method helicopter selection and application", "2019 8th International Conference on Industrial Technology and Management (ICITM)", pp. 170–174, IEEE, 2019, doi:[10.1109/ICITM.2019.8710703](https://doi.org/10.1109/ICITM.2019.8710703).
- [3] B. Khorshidi, "A new method for finding the center of gravity of polygons", *Journal of Geometry*, vol. 96, no. 1-2, pp. 81–91, 2009, doi:<https://doi.org/10.1007/s00022-010-0027-1>.
- [4] M. Murray, A. Seireg, R. Scholz, "Center of gravity, center of pressure, and supportive forces during human activities", *Journal of applied physiology*, vol. 23, no. 6, pp. 831–838, 1967, doi:<https://doi.org/10.1152/jappl.1967.23.6.831>.
- [5] . Shermukhamedov, K. Baynazarov, "Graphic-analytical method for calculating the distribution of forces over the frame in the working process of the unloading", *Scientific-technical journal*, vol. 4, no. 2, pp. 79–86, 2021, doi:<https://uzjournals.edu.uz/ferpi/vol4/iss2/11>.
- [6] P. Wu, L. Zhou, H. Chen, H. Zhou, "An improved fuzzy risk analysis by using a new similarity measure with center of gravity and area of trapezoidal fuzzy numbers", *Soft Computing*, vol. 24, no. 6, pp. 3923–3936, 2020, doi:<https://doi.org/10.1007/s00500-019-04160-7>.
- [7] Y. Tsukamoto, S. Funatani, "Application of feature matching trajectory detection algorithm for particle streak velocimetry", *Journal of Visualization*, vol. 23, no. 6, pp. 971–979, 2020, doi:<https://doi.org/10.1007/s12650-020-00677-4>.
- [8] B. Wiecek, M. Kukla, Ł. Warguła, "The symmetric nature of the position distribution of the human body center of gravity during propelling manual wheelchairs with innovative propulsion systems", *Symmetry*, vol. 13, no. 1, p. 154, 2021, doi:<https://doi.org/10.3390/sym13010154>.
- [9] S. Zagorski, D. Andreatta, G. Heydinger, "Development of a passenger vehicle seat center-of-gravity measuring device", Tech. rep., SAE Technical Paper, 2020, doi:<https://doi.org/10.4271/2020-01-1061>.
- [10] P. Han, J. Zhang, L. Yang, "Research on active gravity center fault tolerance control of fuel system", "Advances in Guidance, Navigation and Control", pp. 5157–5164, Springer, 2022, doi:https://doi.org/10.1007/978-981-15-8155-7_425.
- [11] R. Holubek, M. Vagaš, "Center of gravity coordinates estimation based on an overall brightness average determined from the 3d vision system", *Applied Sciences*, vol. 12, no. 1, p. 286, 2022, doi:<https://doi.org/10.3390/app12010286>.
- [12] B. Firmanto, A. F. O. Gaffar, B. Suprpty, A. B. W. Putra, et al., "Multi-modal biometric system based on feature source compaction and the proposed vcg (virtual center of gravity) feature", "2021 International Seminar on Intelligent Technology and Its Applications (ISITIA)", pp. 95–100, IEEE, 2021, doi:[10.1109/ISITIA52817.2021.9502224](https://doi.org/10.1109/ISITIA52817.2021.9502224).
- [13] A. Brian, N. Getchell, L. True, A. De Meester, D. F. Stodden, "Reconceptualizing and operationalizing seefeldt's proficiency barrier: Applications and future directions", *Sports Medicine*, pp. 1–12, 2020, doi:<https://doi.org/10.1007/s40279-020-01332-6>.
- [14] A. Mohamed, "Gravity applications to groundwater storage variations of the Nile delta aquifer", *Journal of Applied Geophysics*, vol. 182, p. 104177, 2020, doi:<https://doi.org/10.1016/j.jappgeo.2020.104177>.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Disinfecting Omnidirectional Mobile Robot with Vision Capabilities

Waqas Qaisar *, Muhammad Tanveer Riaz, Abdul Basit, Yasir Naseem, Zohaib Nazir

Department of Mechanical, Mechatronics and Manufacturing Engineering, University of Engineering and Technology Lahore, Faisalabad Campus, Faisalabad, 38000, Pakistan

*Corresponding author: Waqas Qaisar, Email: m.waqasqaisar777@gmail.com

Corresponding author ORCID: 0000-0001-7599-0908

ABSTRACT: The disinfecting mobile robot with omnidirectional movement is a vision-capable robot that uses image processing to follow a dedicated path regardless of the change in direction required and uses Ultraviolet rays from the UV light tubes to disinfect everything in its path and disinfect the entire room. The basic premise of the project in this paper is the principle of designing a mobile robot that has high mobility and can move in every direction to follow a dedicated path that can be used to disinfect certain places that are not feasible for human beings. This proposed project is an omnidirectional mobile robot that will be designed such that it will use controllers, a camera for image processing to avoid any obstacle in its path, and actuators all communicating with one another to rotate the wheels of the robot individually to achieve the desired linear and rotatory motion to avoid any obstacles in the path of the robot and clean all the bacteria and germs in the room that might be harmful to humans. All of these components work through feedback which is being given through an encoder to achieve the desired output motion of the robot. The main benefit of this disinfecting mobile robot will be its increased mobility due to the combined effect of its rotatory and linear motion. This increased mobility combined with a set of ultraviolet light rays and a camera in the front which uses image processing to detect any object in its path and avoid it to disinfect an entire targeted area and allow it to access areas where conventional robots and humans can't go.

KEYWORDS: Omnidirectional, Disinfecting, Control System, Image Processing, Obstacle

1. Introduction

The basic premise of this project is the principle of designing an omnidirectional mobile robot that uses image processing to detect its path and uses Ultraviolet rays to disinfect everything in its dedicated path. As we know that a mobile robot is a typical moving robot which uses sensors or cameras along with other technology to identify its surroundings and move around its environment and is controlled using a specific software. Mobile robots normally use a combination of wheels, tracks, and legs along with artificial intelligence (AI) and they are becoming increasingly popular across the different major business sectors of the world [1]-[5]. This proposed project is a mobile robot that will be designed such that it will use controllers, sensors, and actuators all communicating with one another to rotate the wheels of this robot individually to achieve the desired rotation to avoid any obstacles in the path of the robot [6]-[9]. The movement of the robot is done using omnidirectional

caster wheels which are rotated using a motor [10]. But first, we shall discuss how we will use all these components together to form a closed-loop control system and achieve this end goal.

2. Methodology

The primary component that will initiate this process of rotation of the wheel at a certain angle is the camera of this robot which will act as the sensor for this control system. The camera will act as the switch of this system because when the camera sees nothing and no information is being given to the microprocessor then the robot will continue to move in a straight path. But, as soon as some visuals are picked up by the camera of some obstacle in the way then it will convey that information to the microprocessor which will then do its function to avoid that particular obstacle.

The next component of this system to be activated after receiving the input signal from the sensor will be the

microprocessor. The microprocessor which we are using in this system is a Raspberry Pi microprocessor and it will be programmed using python programming specifically for the type of actuators that will be attached to this system. The microprocessor in this system can be said to act as the messenger as it takes the signal from the input that is the camera and then uses that signal to the controller, for which we shall use an Arduino that will control the rotation of motors.

The image picked up by the camera module is processed by the microprocessor using python programming code embedded in it. The Raspberry Pi has a dedicated camera input port for recording HD video and high-resolution photos [11]-[13]. We will create tools that take photos and video and analyze them in real-time for the controller to understand the image signal using Python and specific libraries written for the Pi [14]. This is done by taking the input image in the form of RGB colors which are not in a readable format for the motor controller. The image data is a two-dimensional array of pixels, with each pixel representing a tuple of three values: the relative intensity of red, green, and blue colors in the range 0 to 255 [15]-[18]. This python code will access a specific pixel in the image by treating the image as an array and providing a tuple of the pixel's x and y positions [19]. This will convert the image signal from the camera and transfer it to the microcontroller in a form that it will understand to control the Servo and DC gear motors that are used for the motion of this omnidirectional robot.

The final component of this system is actuators. Two kinds of motors are being used in this mobile robot them being the Servo motor and a simple DC gear motor. The reason for these two input actuators is because the Servo motors act as the steering motor and the DC gear motor acts as the driving motor for this omnidirectional mobile robot [20]. The actuators are controlled with the help of the controller and the processing is done by the microprocessor. After receiving the signal from the controller the DC gear motor will provide the driving motion for this robot and the steering will be done by the Servo motor as it will change its angle and subsequently the angle of the wheel according to the input signal taken from the camera in the form an image which will be then processed by the raspberry pi microprocessor.

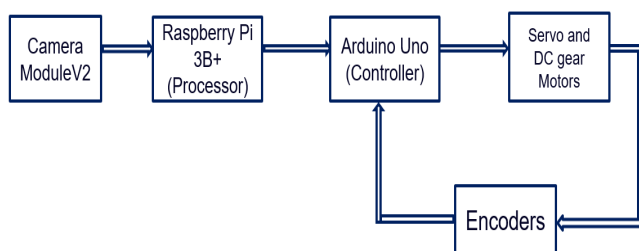


Figure 1: Block Diagram Representation of methodology

2.1. Disinfecting mechanism in the Omnidirectional Robot

The omnidirectional motion of the robot is to make it self-aware such that it follows a dedicated path and avoids any obstacle that comes in its way. The reason for this path following is to disinfect the entire area which the robot covers as it moves around an entire room using ultraviolet light. The top of the robot will include a socket that will be able to attach Ultraviolet tubes in a circular manner which will kill any germs and viruses in the path around the robot and ultimately make the whole area safe for human beings [21]-[24]. Hence the disinfecting mechanism of this robot is a very effective and practical one and also very easy to implement even on an industrial scale.

Now, there are three basic categories of Ultraviolet radiation which are used for disinfecting purposes in mobile robots. The three categories are as follows.

- UV- A
- UV-B
- UV-C

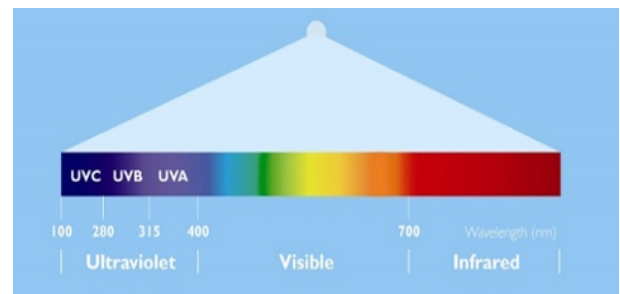


Figure 2: UV rays types

The type of Ultraviolet rays we are going to use for disinfecting purposes in our robot is the UV type C rays [25]-[28]. This is because the most effective and safe UV radiation for disinfecting is UV-C. The purpose of discussing all this is that a disinfecting mobile robot will move from one point to another and will stay at a particular distance from the wall and for some particular time it will exert UV radiation on the wall or a surface that we want to disinfect.

The purpose of discussing all of this is that that time the robot will take to disinfect the face independent of the RPM of the motor. Motor's job is simply to carry the robot from one place to another for disinfecting.

3. Literature Review

In [29], the authors explained the Exploring the Applicability of Robot-Assisted UV Disinfection in Radiology. This robot has used this Ultraviolet disinfection technology in their robot to make a room completely germ-free. The robot designed in this paper has a wheeled mechanism along with sensors that allows it to move across a room and disinfect the entire area eliminating the use of current disinfecting methods which

include using mops, and chlorine-based fluids by human cleaners.

This robot is very practical and our mobile robot aims to achieve a similar result but in a much better, and easier manner. Our robot uses a camera for path following and obstacle avoidance which increases the accuracy of our robot. Moreover, our robot uses very inexpensive parts like Caster wheels, and Arduino unlike the robot in this paper which uses mecanum wheels and an overall expensive and complicated mechanism to achieve the same result that our robot achieves with a less production cost.

In [30], the authors explained Motion Analysis and Control of Three-Wheeled Omnidirectional Mobile Robot. This is a three-wheel mobile robot with all three wheels being active omnidirectional wheels. The wheel is used are mecanum wheels which have very low torque and a very low speed as compared to what we require in our system. Hence mecanum wheels are a very expensive and relatively ineffective tool for an obstacle-avoiding omnidirectional robot. To avoid this, we are using caster wheels that are relatively inexpensive and do the job fairly well to avoid obstacles in the path of the robot and to follow a dedicated path.

In [31], the authors explained Structures of the Omnidirectional Robots with Swedish Wheels. Swedish Wheels can be Elliptical Mecanum double wheel or Mecanum wheel with rotary rollers. We are not using any kind of Special wheels as we are using omnidirectional caster wheels. Caster wheels can rotate and translate and these wheels yield the same result as mecanum and omnidirectional wheels and caster wheels are more practical in addition to being relatively inexpensive as compared to Swedish wheels.

In [32], the author focuses on a robot that is capable of moving around an entire room and disinfecting it from germs using ultraviolet radiation type C and does it similarly to our omnidirectional disinfecting mobile robot. But, the design of this robot limits its mobility and restricts the robot from moving between tight spaces due to the robot having a massive base, and a large turning radius. Our robot on the other hand has a relatively small base and three 360 degrees rotating wheels which enables it to rotate in very tight spaces. This makes our omnidirectional robot much more practical, especially for hospitals where there are many tight spaces between patient beds and robots such as proposed in this paper tend to see it very difficult to move in such spaces.

In [33], the authors explained a design of a three-wheel mobile robot with mecanum wheels. These wheels produce very high torque and can move in an omnidirectional manner. The working of these kinds of wheels is very complex and can be very hard to control

using simple DC motors. The robot in this paper is also using an ARM Cortex-M4 microcontroller which is incapable of being connected to a camera for image processing purposes. Our design on the other hand uses an Arduino and a Raspberry Pi which work in serial communication and are capable of connecting any kind of sensor or camera with our robot.

In [34], the authors explained the wheel mechanism in this mobile robot uses an AC servo motor with magnetic clutches for driving the wheel and steering shaft. The magnetic clutch mechanism is a very impractical approach for our robot due to its high initial cost and the overheating of the clutches.

The clutch and gear system can be very useful in a lot of scenarios but cannot be controlled using image processing techniques. To compensate for this deficiency of magnetic clutches, our design uses three motors (One for each wheel) which are controlled using an Arduino according to the commands given to it by the Raspberry Pi.

In [35], the authors explained mobile robot consists of three active wheels. We are not using any kind of Special wheels, we are using motors to control simple wheels which can provide the best result as compared to these mecanum wheels.

The design of caster wheels being controlled using motors for each wheel is a very practical one due to its inexpensive nature and ease with which the entire mechanism can be assembled and controlled.

In [36], the authors explained the system in this paper is very efficient in terms of what is said to achieve. But it only rotates the wheel at 180 degrees while our requirement is for a 360-degree rotation mechanism. If the load on this system increases then motors of higher power will be required. The components used in the design of this robot include a bearing, chain drive, DC motor, sprocket, Steering, and wheel.

Our design of an omnidirectional robot on the other hand is a relatively simple one and is also easier to implement and produces a rotation of a full 360 degree to help the robot avoid any obstacle in its way.

In [37], the authors explained the Design of 360-degree rotating car aided for parking. The robot in this paper uses a locking mechanism and rotates the entire robot by lifting it for parking purposes while our system will use a microprocessor to control the movement of motors and rotate the wheels to follow a dedicated path.

The robot in this paper has very little practicality as the lifting and rotating mechanism can be used for very little practical use in our everyday lives and it also cannot be implemented on a much larger scale. Our omnidirectional

path following robot, on the other hand, is a very practical one and has a wide array of uses in hospitals, and labs for disinfecting purposes and it can also be implemented on an industrial level.

In [38], the authors explained Intelligent disinfection robots assist medical institutions in controlling environmental surface disinfection. This robot uses a machine vision algorithm for the movement of the robot as a dedicated path is fed into the robot memory and it follows it and uses UV rays to disinfect the entire area from any germs and viruses [39].

Our robot on the other hand uses cameras, and image processing techniques to detect any object in its way and follow a dedicated path and continue to move without any interruption whereas the robot in this paper can be very inaccurate if an object comes in its way and it may stop the disinfecting process entirely due to a lack of sensors. Moreover, the material and circuitry required for this robot are very expensive and impractical to manufacture on a large scale whereas our robot uses normal electronics which are relatively inexpensive and allow for large-scale manufacturing.

4. Calculations and Analysis of three-wheel omnidirectional robot

4.1. RPM Calculation of motor

Motor that we are using for our project is TG-47E-LG having 753 rpm.



Figure 3: TG-47E-LG DC Gear Motor

Formulas Used

To calculate the Wheel travel with this motor we use

$$W_T = 2\pi r \tag{1}$$

where:

W_T = Wheel Travel

π = 3.14

r = radius of wheel

The radius of our specific wheel is 0.0325 m

Hence

$$W_T = 0.2041 \text{ m}$$

To calculate the desired velocity we use

$$s = v \times t \tag{2}$$

where:

s = distance traveled

v = desired velocity

t = time taken to travel distance

By rearranging we get:

$$v = s/t$$

For $s=2$ and $t=1$

$$v = 2 \text{ ms}^{-1}$$

For the desired RPM using our desired velocity we use:

$$\text{RPM} = (v \times W_T) \times 60 \tag{3}$$

Hence

$$\text{RPM} = 587$$

where 587 means that it will be almost 80% of 753 which is the rated speed of motor. Hence, 80% of 753 is 204, this means that we require 204 value of PWM in our Arduino to rotate the motor at our desired speed according to our wheel.

4.2. Kinematic Analysis

For kinematic analysis we are going to consider the omnidirectional wheel we are using for achieving the omnidirectional movement. The wheel we are using looks like:



Figure 4: Caster Wheel

The calculation of the kinematic of the wheel is done as:

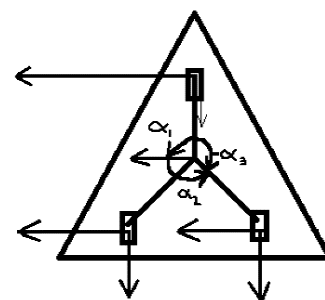


Figure 5: FBD of the three wheel robot

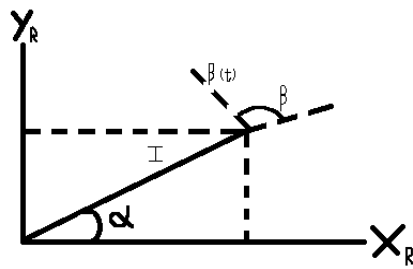


Figure 6: FBD of a single wheel

Our goal is to find the rolling and sliding constraints where:

θ = Wheel rotation

r = Wheel radius

Speed of wheel = $v = r \times \theta$

Translational Motion of Robot Motion is given by:

$$\zeta = [\dot{x} \ \dot{y} \ \theta]^T$$

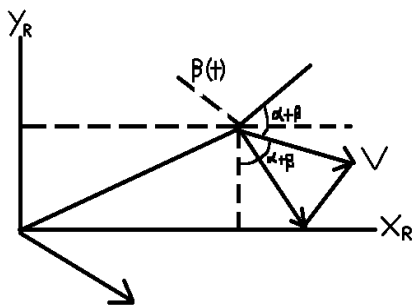


Figure 7: Translational motion of a single wheel

$$v_x = \sin(\alpha + \beta)$$

$$v_y = -\cos(\alpha + \beta)$$

$$[v_x \ v_y] = [\sin(\alpha + \beta) \ -\cos(\alpha + \beta)]$$

$$v_1 = [v_x \ v_y] \cdot \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}$$

$$v_1 = [\sin(\alpha + \beta) \ -\cos(\alpha + \beta)] \cdot \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}$$

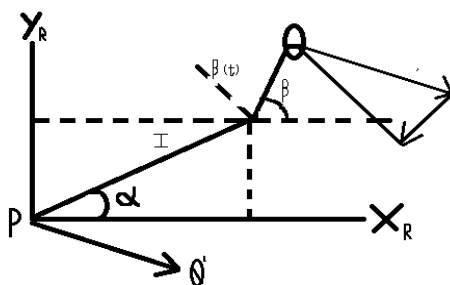


Figure 8: Rotational motion of a single wheel

$$\text{Rotational Motion} = -l \times \theta = -l \cdot \cos(\beta) \cdot \theta$$

By Combining:

$$v = r \times \theta = \sin(\alpha + \beta) \dot{x} - \cos(\alpha + \beta) \dot{y} - l \cdot \cos(\beta) \cdot \theta$$

We can write it as:

$$\sin(\alpha + \beta) \dot{x} - \cos(\alpha + \beta) \dot{y} - l \cdot \cos(\beta) \cdot \begin{bmatrix} \dot{x} \\ \dot{y} \\ \theta \end{bmatrix} - r \cdot \theta = 0$$

This gives us:

$$\sin(\alpha + \beta) \dot{x} - \cos(\alpha + \beta) \dot{y} - l \cdot \cos(\beta) \cdot \zeta - r \cdot \theta = 0$$

Which is the rolling constraint.

Similarly our Sliding constraint is

$$[\cos(\alpha + \beta) \cdot \sin(\alpha + \beta)d + l \cdot \sin\beta] \cdot R \cdot \zeta + d \cdot \beta$$

where

d = offset from the wheel axis

β = Variable steering

4.3. Dynamic Analysis:

The Dynamic analysis of this omnidirectional robot is going to be derived by using the Euler - Lagrange method:

$$L = K.E - P.E$$

where:

K.E = Kinetic Energy

P.E = Potential Energy

Our mobile robot will be derived by considering the surface of its movement to be planar, hence:

$$P.E = 0 \text{ J}$$

The Lagrange dynamic formulation is described as:

$$Y = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q}$$

Now m_w, m_l, m_p are the mass of wheel, mass of link and the mass of platform respectively.

I_w, I_l, I_{mp} are the inertial of wheel, link and platform respectively.

V_{wi}, V_{li} are the linear velocity of wheel and a link at t .

Ω_{wi}, Ω_{li} are the angular velocity of wheel and a link at t .

V_p, Ω_p are the linear and angular velocities of the platform.

$$K_{wi} = \frac{1}{2} m_w \cdot v_{wi}^T \cdot v_{wi} + \frac{1}{2} I_w \cdot \Omega_{wi}^T \cdot \Omega_{wi}$$

For Wheel:

$$K_{li} = \frac{1}{2} m_l \cdot v_{li}^T + \frac{1}{2} I_l \cdot \Omega_{li}^T \cdot \Omega_{li}$$

For link:

$$K_p = \frac{1}{2} m_p \cdot v_p^T \cdot v_p + \frac{1}{2} I_p \cdot \Omega_p^T \cdot \Omega_p$$

For Platform:

Total Kinetic Energy is:

$$K.E = \sum_{i=1}^3 X_{wi} + \sum_{i=1}^3 X_{li}$$

The forward dynamic equation of motion is a relation between the activated torque of wheel (as input) and angular velocities of wheel (as output). The activated variable of wheel is angular velocity \dot{q}_x

$$q = [\theta_x \ \theta_y \ \theta_z]$$

$$\dot{q} = [\dot{\theta}_x \ \dot{\theta}_y \ \dot{\theta}_z]$$

By using Lagrange:

$$T_x = \frac{d}{dt} \left(\frac{\partial x}{\partial \dot{q}_x} \right) - \frac{\partial x}{\partial q_x}$$

$$\frac{\partial L}{\partial q_x} = \begin{bmatrix} \frac{\partial x}{\partial \theta_y} \\ \frac{\partial x}{\partial \theta_y} \\ \frac{\partial x}{\partial \theta_z} \end{bmatrix}, \frac{\partial L}{\partial q} = \begin{bmatrix} \frac{\partial x}{\partial \theta_y} \\ \frac{\partial x}{\partial \theta_y} \\ \frac{\partial x}{\partial \theta_z} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This implies that:

$$T_x = M(q_s) \cdot \ddot{q} + G_x(q_s, \dot{q}_x)$$

$$T_x = [T_x \ T_y \ T_z]^T$$

$G = [q_s \ \dot{q}_x]^T$ is neglected as it is very small

$$T_s = M(q_s) \cdot \ddot{q}$$

Hence, we find the torque vector as function of angular wheel acceleration.

4.4. Robot Constraints Calculations

4.4.1. Rolling Constraints

$$[\sin(\alpha + \beta) - \cos(\alpha + \beta) - 1 \cdot \cos(\beta)] \cdot \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = r \cdot \Phi$$

This gives us

$$\sin(\alpha + \beta) \dot{x} - \cos(\alpha + \beta) \dot{y} - 1 \cdot \cos(\beta) = r \cdot \Phi$$

$$\sin(\alpha + \beta) \dot{x} = r \cdot \Phi + \cos(\beta) \cdot \Phi \tag{4}$$

Now.

$$\alpha = 120$$

$$\beta = 0$$

$$r = 3$$

$$\Phi = 50$$

Putting values in equation (4), we get:

$$\sin(120) \dot{x} = [3(50)] + [4(\cos(0))] \cdot 50$$

$$0.866 \dot{x} = 350$$

$$\dot{x} = 350 / 0.866$$

$$\dot{x} = 404 \text{ cms}^{-1}$$

As the wheels are same with similar sizes and dimensions so the constraints are also the same for each wheel.

4.4.2. Sliding Constraints:

$$[\cos(\alpha + \beta) + \sin(\alpha + \beta) d + l \sin(\beta) \beta] + d \cdot \beta = 0$$

After evaluation of our servo motor, we observe that our sliding constraints tend to move towards zero and at a certain point there are no sliding constraints in our wheel as there is no sliding after a certain point in the robot's movement.

Forward Dynamics

$$T = M(q_s) \cdot \ddot{q}$$

where $M(q_s)$ = Mass and initial parameters.

$$\ddot{q} = [\ddot{\theta}_x \ \ddot{\theta}_y \ \ddot{\theta}_z]$$

$$M = \begin{bmatrix} M1 & 0 & 0 \\ 0 & M2 & 0 \\ 0 & 0 & M3 \end{bmatrix}$$

$$T_s = [T_x \ T_y \ T_z]$$

Now, both T_y and T_z are zero, hence after solving we get:

$$\theta = 197 \text{ rad/s}^{-1}$$



Figure 9: Raspberry Pi

5. Serial Communication between Raspberry Pi and Arduino

5.1. Raspberry Pi

A Raspberry Pi board is a fully functional computer with all the bells and whistles of a full-fledged computer,

including dedicated memory, a processor, and a graphics card for output via HDMI. It even runs a specially designed version of the Linux operating system and is simple to install in most Linux software, and with little effort.

5.2. Arduino

Arduino is made up of three components. There is a hardware prototype platform, the Arduino language, as well as an IDE and libraries. The Arduino boards are not full-fledged computers, but rather microcontrollers. They don't run a full operating system; instead, they write code and execute it as their firmware interprets it. The Arduino board's primary function is to interface with devices and sensors, making it ideal for hardware projects in which you simply want things to respond to various sensor readings and manual input.

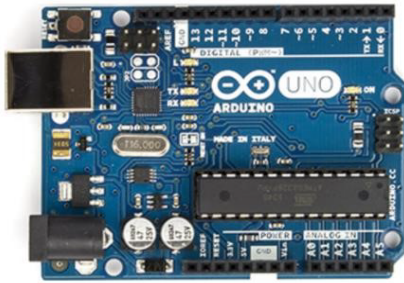


Figure 10: Arduino Microcontroller

5.3. Arduino and Raspberry Pi

We are using Arduino with Raspberry Pi because of many factors like we are using some Analog Sensors in our prototype and due to the unavailability of Analog to Digital Converters in Raspberry Pi we are using it and we can also get a much better real-time response in Arduino. It is very easy to get started with Arduino due to the easy programming language and libraries availability that's why we are controlling the actuators (DC and Servo Motors) and getting the real-time response from the sensors (Compass Sensors, Encoders, etc.) and using Raspberry Pi just for Path Tracking, Object Detection, and Distance Measurement.

5.4. Serial Communication between Arduino and Raspberry Pi

Now, due to the Raspberry Pi consisting only 26 GPIO Pins and no ADC channels, it cannot handle all of the interactions in projects like 3D printers. As a result, we require more output pins and additional functions, and in order to add more functions to PI, we must establish communication between PI and UNO. With this, we can use all of UNO's functions as if they were PI functions.

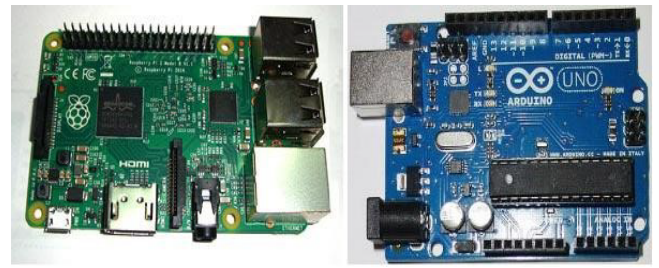


Figure 11: Side by side image an Arduino and Raspberry Pi

5.5. Working and Programming Explanation:

Arduino Uno Part

In order to do programming we connect the Arduino with our computer or laptop and then write the program in the Arduino IDE software, and then upload it to the UNO. Then, unplug the UNO from the PC. After programming, connect the UNO to the PI and connect an LED and a button to the UNO. The program now initializes the Arduino's Serial Communication. When we press the UNO's button, the UNO sends a few characters serially to the PI via the USB port. The PI's LED blinks to indicate the characters being sent.

Raspberry Pi Part:

Following that, we must write a program for PI (see Code section below) to receive the data sent by UNO. To do so, we must first understand the commands listed below. We'll import serial files from the library, and this function will allow us to send and receive data serially or via USB port.

Now we must specify the device port and bit rate in order for the PI to receive data from the UNO without errors. The following command enables serial communication at 9600 bits per second on the ACM0 port.

```
ser = serial.Serial('/dev/ttyACM0', 9600)
```

Go to the PI terminal and look for the port to which the arduino is connected and then enter.

```
ls /dev/tty*
```

On the Raspberry Pi, you will then see a list of all attached devices. Connect the Arduino Uno to the Raspberry Pi using a USB cable and repeat the command. The displayed list makes it simple to identify the UNO attached port.

The following command is used to create a forever loop, which means that the statements inside the loop will be executed indefinitely.

While 1:

We will display the characters on the PI screen after receiving the data serially.

```
print (ser.readline())
```

6. Image Processing

Image Processing is basically the use of computers to process digital images by using an algorithm or a programming code. It should be noted that every image is made up of several pixels whose pixel count is dependent on the resolution of the camera for a 640x480 resolution camera module there will be a total of 307,200 pixels make up the image in each frame for that particular camera. Each colored picture every frame is made up of three main colors them being Red, Green, and Blue. Now for a digital image representation:

- We need to know the position of each pixel in that image.
- We need to know the corresponding RGB values.
- This same process can be carried out for videos too as a video is just a series of images.
- All of this is carried out in our system using a raspberry pi board connected with a raspberry pi camera module V2.

Path planning is one of the most crucial things in any obstacle avoiding and path following mobile robot and this path planning is achieved through the process of image processing. For this, the first step is to enhance the image and to reduce noise to make the objects clear for the camera to detect and for this we are using histogram manipulation.

Segmentation in RGB vector space is simple. The goal of segmentation is to categorize each RGB pixel in a given image. Otsu's method is used in image processing and computer vision to automatically reduce a grey level image to a binary image or to perform histogram shape based image shareholding.

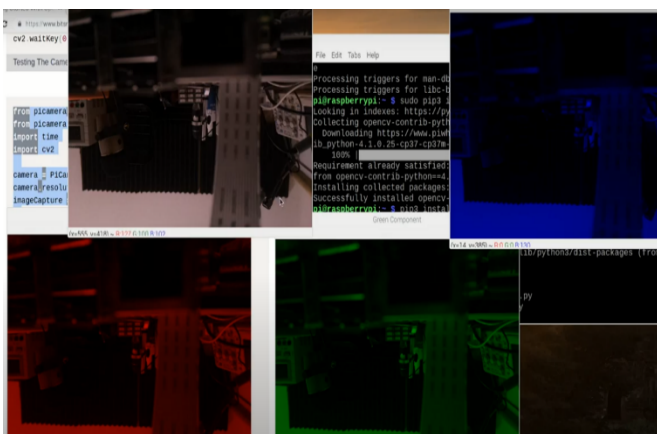


Figure 12: RGB Segmentation of Image

Then we write programming using python on a raspberry pi module in such a way that uses this exact method to clarify each pixel and help the camera detect objects. After that, we do camera calibration as it is a necessary step in 3D computer vision to extract information from 2D images from the camera and we are using a method called the model plane method.

This technique only requires the camera to observe a planar pattern made up of a grid pattern in various orientations. Corner points are extracted for each view in order to calculate the correspondence between the model plane and its image. This helps the camera to detect the position of objects and send that information to the controller which then gives the command to rotate the servo motors accordingly.

6.1. Object Detection

Object detection is a very important part of image processing when designing a vision-capable mobile robot. The reason is that the camera will detect any change in the reference frame and detect the presence of any object and then use the python coding to calculate the depth and distance of that object and then avoid it by giving the signal to the controller. As a result, object detection can be seen as a computer vision technique that enables us to recognize and locate objects in an image or video. Object detection can be used to count objects in a scene, determine and track their precise locations, and accurately label them using this type of identification and localization. Normally it is observed that the concepts of object detection and image recognition are frequently confused, so we must first distinguish between them in order to further learn about the working of image processing.

Image recognition is the process by which an image is labelled. The label "dog" is applied to a picture of a dog. A picture of two dogs is still labelled "dog." Object detection, on the other hand, draws a circle around each dog and labels it "dog." The model predicts where each object will be and what label will be assigned to it. Object detection, in this sense, provides more information about an image than recognition.

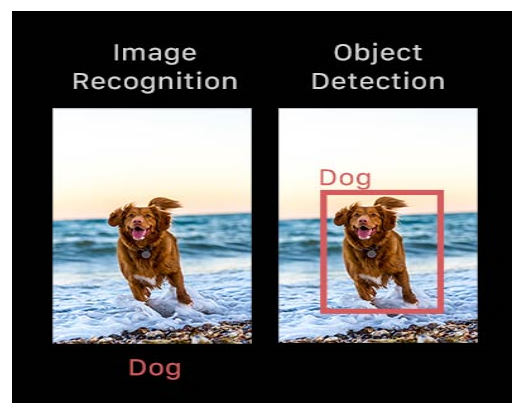


Figure 13: Difference between Object detection and image recognition

6.2. Object Detection Methodology

First, we take a reference image without the object to be focused and then take another image in the same frame with an object. Sample image is given as:

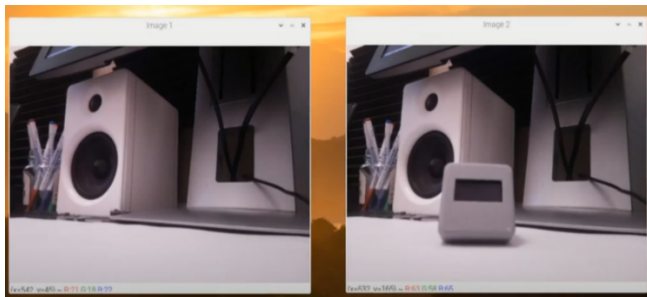


Figure 14: Object Segmentation

Secondly, this image is converted from colored to a greyscale image so that all the color characteristics are gone and now we only have two colors to work with them being black and white.

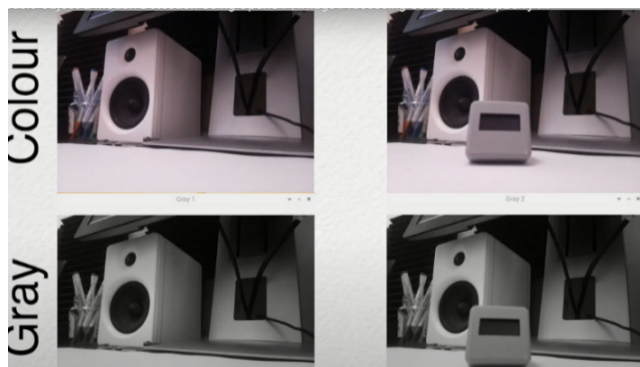


Figure 15: Black and White

Then we subtract the two images to differentiate the object to be detecting from the surrounding environment.



Figure 16: Object subtraction from background

After that this image is converted into a binary black and white image which acts to finally segment the object in focus from the surroundings as the objects gets a white color and the environment is represented in a black shade.



Figure 17: Binary object coloration

7. 3D Cad Model of Omnidirectional Robot

The 3D Cad Model of this omnidirectional vision capable mobile robot is carried out using solid works software and is rendered according to the physical hardware of the robot which is desired at the completion of this project. The cad design is shown as:

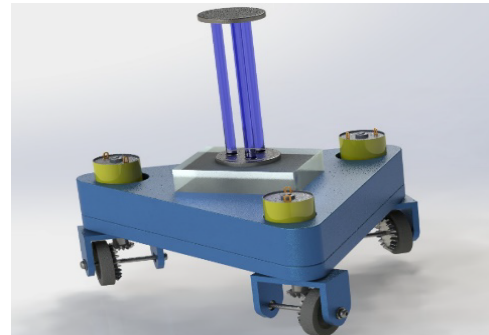


Figure 18: Isotropic view of Tri wheel omnidirectional robot

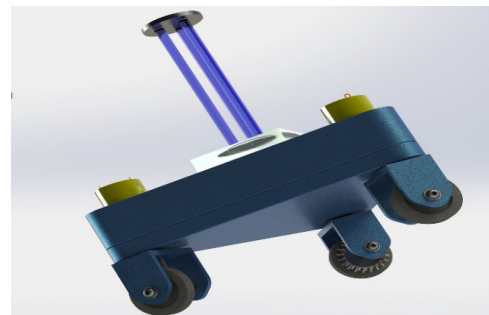


Figure 19: Cad design of the wheels being used

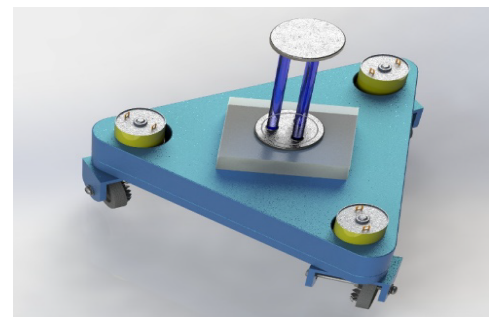


Figure 20: Another view of the 3D Cad Design



Figure 21: Base, Single wheel and DC gear motors in the robot

8. Mechanical Model

Using this 3D Cad model as a reference we have made an aluminum base frame with motors attached to it and also cut it according to the dimensions we used in our solid works' design and also bought and used all the other components which include 3 Servo Motors, 3 DC gear motors, crown and pinion, Arduino Uno, Raspberry Pi 3B+ and Raspberry Pi camera module V2. Some of the images of the physical components and hardware are given in figure 21.

This entire design is made according to the calculations and theoretical derivations defined in the previous headings. The kinematic and dynamic analysis of our robot allows us to check as to what kind of components are required to make this robot and what kind of parameters do we need to make our robot a working success.

9. Conclusion

The main outcome of this paper is to explain the workings of our omnidirectional mobile robot with vision capabilities and explain how we can achieve this omnidirectional movement and how this robot will be used to avoid any and every object in the path of the robot. After learning everything, it can be concluded that the basic things required to achieve the omnidirectional motion of caster wheels are two kinds of motors, them being Servo and DC gear and control these motors using encoders and use pulleys to transfer the rotation torques and then use image processing to detect the objects in front of the robot and tell the motor to rotate or stop accordingly so that it would follow a dedicated path and use the Ultraviolet rays to disinfect the entire area in its pathway and make it safe for humans to enter and do their work in a germ-free and disinfected area.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is completed in Mechatronics Engineering department of the university of engineering and technology Lahore (Faisalabad Campus).

References

- [1] H. Nacer, B. Mendil, "Motion analysis and control of three-wheeled omnidirectional mobile robot," *Journal of Control, Automation and Electrical Systems*, vol. 30 no. 2, pp. 194-213, 2019.
- [2] M.T. Riaz et al., "A wireless controlled intelligent healthcare system for diplegia patients", *Mathematical Biosciences and Engineering*, 19(1): 456-472, 2022. doi: 10.3934/mbe.2022022
- [3] H.A. Javaid et al., Classification of Hand Movements Using MYO Armband on an Embedded Platform. *Electronics*, 10, 1322, 2021. DOI: 10.3390/electronics10111322
- [4] M. T. Riaz et al., "The Intelligent Transportation Systems with Advanced Technology of Sensor and Network," 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 2021, pp. 1-6, doi: 10.1109/ICECube53880.2021.9628331.
- [5] M. T. Riaz et al., "Design and Experimental Validation of a Small-Scale Prototype Active Aerostatic Thrust Bearing," 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 2021, pp. 1-6, doi: 10.1109/ICECube53880.2021.9628270.
- [6] M. A. Akbar, M. Shafiq, T. Kamal, M. T. Riaz, and M. K. Shad, "An empirical Study Investigation of Task Allocation Process Barriers in the Context of Offshore Software Development Outsourcing: An Organization Size Based Analysis," *International Journal of Computing and Digital Systems*, 8(04), 343-350, 2019. DOI: <http://dx.doi.org/10.12785/ijcds/080403>
- [7] M.A. Akbar et al., "Multicriteria Decision Making Taxonomy of Cloud-Based Global Software Development Motivators", *IEEE Access*, 8, 185290-185310, 2020. DOI: 10.1109/ACCESS.2020.3030124
- [8] M. A. Akbar et al., "Identification and Prioritization of Cloud Based Global Software Development Best Practices," in *IEEE Access*, vol. 8, pp. 191242-191262, 2020, doi: 10.1109/ACCESS.2020.3031365.
- [9] M. A. Akbar et al., "Requirements Change Management Challenges of Global Software Development: An Empirical Investigation," in *IEEE Access*, vol. 8, pp. 203070-203085, 2020, doi: 10.1109/ACCESS.2020.3035829.
- [10] M. A. Akbar et al., "AZ-Model of software requirements change management in global software development," in 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 1-6, 2018. DOI: 10.1109/ICECUBE.2018.8610964
- [11] A. Ahmed et al., "Modeling and Simulation of Office Desk Illumination Using ZEMAX," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 1-6, 2019. DOI: 10.1109/ICECCE47252.2019.8940756
- [12] M. Taiyaba et al., "Secure V2X Environment using Blockchain Technology." In *Proceedings of the Evaluation and Assessment in Software Engineering*, pp. 469-474. 2020. DOI: <https://doi.org/10.1145/3383219.3383287>
- [13] M. Shafiq et al., "Towards successful global software development." In *Proceedings of the Evaluation and Assessment in Software Engineering*, 445-450. 2020. DOI: <https://doi.org/10.1145/3383219.3383283>
- [14] M. T. Riaz et al., "Investigation of Electrical Properties of Epoxy Resin Composite with the Surface Modification of SiO₂ Nanoparticles," 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 2021, pp. 1-5, doi: 10.1109/ICECube53880.2021.9628354.
- [15] M. T. Riaz et al., "Wireless Model for High Voltage Direct Current Measurement using Hall Sensor," 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), 2021, pp. 642-647, doi: 10.1109/IBCAST51254.2021.9393186.
- [16] M. T. Riaz, M. M. Afzal, S. M. Aaqib and H. Ali, "Analysis and Evaluating the Effect of Harmonic Distortion Levels in Industry," 2021 4th International Conference on Energy Conservation and Efficiency (ICECE), 2021, pp. 1-7, doi: 10.1109/ICECE51984.2021.9406283.
- [17] M.T. Riaz et al., "Design of a Free Energy Generator using Gravity Wheel & Dynamo," 2021 4th International Conference on Energy Conservation and Efficiency (ICECE), 2021, pp. 1-5, doi: 10.1109/ICECE51984.2021.9406299.
- [18] M. T. Riaz et al., "Steady State Analysis of HVDC Transmission System Based on MATLAB/SIMULINK," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 1-6, 2019. DOI: 10.1109/ICECCE47252.2019.8940745
- [19] M. T. Riaz et al., "Research on the Protection of Hybrid HVDC System," in 2018 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), 1-6, 2018. DOI: 10.1109/PGSRET.2018.8686007

- [20] M. T. Riaz et al., "Wireless Android Based Home Automation System," *Advances in Science, Technology and Engineering Systems Journal*, 2(1), 234–239, 2017. DOI: 10.25046/aj020128
- [21] H. A. Raza et al., "Analysis the effect of 500kv High-Voltage Power Transmission Line on the Output Efficiency of Solar-Panels," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 1–6, 2019. DOI: 10.1109/ICECCE47252.2019.8940803
- [22] L. Hanwu et al., "Regularity of Current Dispersion in Different Kinds of Grounding Electrode," in 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), 1–4, 2018. DOI: 10.1109/ICHVE.2018.8642240
- [23] M. Idrees et al., "Fuzzy Logic Based Calculation and Analysis of Health Index for Power Transformer Installed in Grid Stations," in 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE), 4, 1–6, 2019. DOI: 10.1109/RAEE.2019.8887016
- [24] U. Farooq et al., "A Reliable Approach to Protect and Control of Wind Solar Hybrid DC Microgrids," 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), Changsha, China, 348–353, 2019. doi: 10.1109/EI247390.2019.9062101.
- [25] A. Waleed et al., "Study on Hybrid Wind-Solar System for Energy Saving Analysis in Energy Sector," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2020, pp. 1–6, doi: 10.1109/iCoMET48670.2020.9073901.
- [26] M. R. Javed et al., "Protection Scheme Design for Star Connected PMSG Based Wind Farm against Line Faults," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 1–6, 2020. doi: 10.1109/iCoMET48670.2020.9074139.
- [27] A. Waleed et al., "Effectiveness and Comparison of Digital Substations Over Conventional Substations," *Adv. Sci. Technol. Eng. Syst. J.*, 4(4), 431–439, 2019. DOI: 10.25046/aj040452
- [28] M. R. Javed et al., "A Comparative Study of Maximum Power Point Tracking Techniques for Solar Systems," 2019 22nd International Multipoint Conference (INMIC), Islamabad, Pakistan, 1–6, 2019. doi: 10.1109/INMIC48123.2019.9022762.
- [29] C. McGinn, R. Scott, N. Donnelly, K. Roberts. M. Bogue, C. Kiernan, M., Beckett, "Exploring the applicability of robot-assisted uv disinfection in radiology," *Frontiers in Robotics and AI*, p. 193, 2021.
- [30] S. Z. Hassan et al., "Intelligent Control of Wind-Assisted PHEVs Smart Charging Station," *Energies*, vol. 12, no. 5, p. 909, 2019.
- [31] Tătar, M. Olimpiu, C. Cirebea, D. Măndru, "Structures of the omnidirectional robots with swedish wheels. In *Solid State Phenomena*," Trans Tech Publications Ltd, vol. 198, pp. 132–137, 2013.
- [32] E. Schahawi, W. Zingg, M. Vos, H. Humphreys, L. Lopez-Cerero, A. Fueszl, E. Presterl, "Ultraviolet disinfection robots to improve hospital cleaning: Real promise or just a gimmick," *Antimicrobial Resistance & Infection Control*, vol. 10 no. 1, pp. 1–3, 2021.
- [33] J. Moreno, E. Clotet, R. Lupiañez, M. Tresanchez, D. Martínez, T. Pallejà, J. Palacín, "Design, implementation and validation of the three-wheel holonomic motion system of the assistant personal robot (APR)," *Sensors*, vol. 16 no. 10, pp. 1658, 2016.
- [34] K. Miyashita, M. Wada, "Study on Self-Position Estimation and Control of Active Caster Type Omnidirectional Cart with Automatic/Manual Driving Modes," 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 1798–1803. IEEE, 2020.
- [35] E. Rubies, J. Palacín, "Design and FDM/FFF Implementation of a Compact Omnidirectional Wheel for a Mobile Robot and Assessment of ABS and PLA Printing Materials," *Multidisciplinary Digital Publishing Institute (MDPI)*, vol. 9, no. 2, p. 43, 2020.
- [36] V. Mallikarjuna et al., "Design and fabrication of 180 degree wheel rotation vehicle," *International Journal of Latest Engineering Research and Applications*, vol. 5, no. 11, 2018.
- [37] A. Thankachan, A. Raveendran, S. Kumar, P. Arun Gawtham, P. Roy, A. Vishnu, "Design of 360 degree rotating car aided for

parking," *International Journal of Latest Engineering Research and Applications (IJLERA)*, vol. 7, no. 4, 2018, doi:10.15680/IJIRSET.2018.0704112.

- [38] Y. Fan, Y. Hu, L. Jiang, Q. Liu, L. Xiong, J. Pan, Q. Zhang, "Intelligent disinfection robots assist medical institutions in controlling environmental surface disinfection". *Medicentral.net*, vol. 1, no. 1, pp. 19–23, 2021.
- [39] A. Waleed et al., "Solar (PV) Water Irrigation System with Wireless Control," in 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE), 4, 1–4, 2019. DOI: 10.1109/RAEE.2019.8886970

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Waqas Qaisar is doing B.Sc. in Mechatronics and Control Engineering from the University of Engineering and Technology Lahore (Faisalabad Campus), Pakistan. Currently, he is working on a project whose title is an Omni-directional mobile robot with vision capabilities. His current research interests include Control of Mechatronics System, Robotics, Power Electronics and Automation.



Muhammad Tanveer Riaz received the B.Sc. degree in electrical engineering from the University of Engineering and Technology Lahore, Pakistan, in 2015, and masters of engineering degree in electrical engineering from the Chongqing University, Chongqing, China, in 2018. He is serving as Lecturer in the department of mechatronics and control engineering at the University of Engineering & Technology Lahore, Pakistan since May 2019. His current research interests include Smart Integration of Renewable energy sources, Smart grid and micro grid security issues, HVDC grounding parameters analysis, Hybrid HVDC protection system, new electrical technologies and energy management.



Abdul Basit is doing his bachelor's degree in Mechatronics Engineering from the University of Engineering and Technology Lahore (Faisalabad Campus). Currently, he is working on a project whose title is an Omni-directional mobile robot with vision capabilities. His current research interests include Control of Mechatronics System, Robotics, computer programming, and web development. His previous engineering projects include working on home automation system, and Tesla coil.



Rana Muhammad Yasir Naseem will receive the B.Sc. degree in Mechatronics and Control engineering from the University of Engineering and Technology Lahore, Pakistan, in 2022. He is a student in the department of Mechatronics and control engineering at the University of Engineering & Technology Lahore. His current research interests include omnidirectional mobile robots with vision capability, e-bike, robotic arm, ambulance drone, and new electrical technologies and energy management.



Muhammad Zohaib Nazir will receive the B.Sc. degree in Mechatronics and Control engineering from the University of Engineering and Technology Lahore, Pakistan, in 2022. He is a student in the department of Mechatronics and control engineering at the University of Engineering & Technology Lahore. His current research interests include omnidirectional mobile robots with vision capability, e-bike, and robotic arm.

Cascaded Keypoint Detection and Description for Object Recognition

Abdulmalik Danlami Mohammed ^{*,1}, Ojerinde Oluwaseun Adeniyi ¹, Saliu Adam Muhammed ¹, Mohammed Abubakar Saddiq ², Ekundayo Ayobami ¹

¹Department of Computer Science, Federal University of Technology, Minna, Niger State, P.M.B.65, Nigeria

²Department of Electrical/Electronic Engineering, Federal University of Technology, Minna, Niger State, P.M.B.65, Nigeria

*Corresponding author: Abdulmalik Danlami Mohammed, +2349069148660, drmalik@futminna.edu.ng

Corresponding author ORCID: <https://orcid.org/0000-0002-0217-7411>

ABSTRACT: Keypoints detection and the computation of their descriptions are two critical steps required in performing local keypoints matching between pair of images for object recognition. The description of keypoints is crucial in many vision based applications including 3D reconstruction and camera calibration, structure from motion, image stitching, image retrieval and stereo images. This paper therefore, presents (1) a robust keypoints descriptor using a cascade of Upright FAST -Harris Filter and Binary Robust Independent Elementary Feature descriptor referred to as UFAHB and (2) a comprehensive performance evaluation of UFAHB descriptor and other state of the art descriptors using dataset extracted from images captured under different photometric and geometric transformations (scale change, image rotation and illumination variation). The experimental results obtained show that the integration of UFAH and BRIEF descriptor is robust and invariant to varying illumination and exhibited one of the fastest execution time under different imaging conditions.

KEYWORDS: Image keypoints, Feature detectors, Feature descriptors, Image retrieval, Image recognition, Image dataset

1. Introduction

The description of image keypoints is at the core of many computer vision applications since it simplifies the task of object recognition and object tracking. Some of the computer vision applications where keypoints description has been found useful include pose estimation, 3D reconstruction and camera calibration, structure from motion, image stitching, image retrieval and stereo images. The job of a descriptor is to describe the intensity distribution of neighbouring pixels around an interest points. As a result, the performance of many vision-based applications such as object recognition, image retrieval and 3D reconstruction can be enhanced with a stable and distinctive descriptor. The development of computer vision based application on mobile phones in time past, has been a challenging task due their low processing power. This has, however, led to a new research direction in image processing and computer vision on low memory devices such as the smart phones. The outcome of such research direction is the development of different methods for describing interest points from an image structure.

These new methods of decomposing the whole image structure into a subset of descriptors reduce computational burden that would otherwise make the process of development and deployment of many computer vision based application cumbersome on a low processing devices (e.g smartphones). In recent time, few works have been proposed to improve the computation of image keypoints and their description with the aim of achieving real time performance and invariance to image transformations such as scale change, image rotation, illumination variation, and image blurring. Some of these works include the oriented FAST and Rotated BRIEF proposed in [1]. Binary Robust Invariant Scale Keypoint presented in [2] and the Fast Retina Keypoints proposed in [3].

In order to achieve robust description of keypoints with minimal computation, we expanded the Upright FAST-Harris Filter proposed in [4] to include Binary Robust Independent Elementary Feature descriptor proposed in [5]. The expansion is a cascaded approach in which keypoints are first detected using the Upright

FAST-Harris filter follow by the computation keypoints descriptor around its' neighbourhood based on Binary Robust Independent Elementary Feature descriptor. Finally, we compare the performance of UFAHB against other state of the art descriptors using dataset extracted from images captured under different imaging conditions.

2. Related work

A wide range of keypoints detectors and their descriptors are proposed in the literature. For example, the Oriented Fast and Rotated Brief also referred to as ORB is a fast and robust local keypoints detector and descriptor proposed in [1]. The algorithm uses the FAST keypoints detector to detect corners in an image and subsequently employs the Harris edge filter to order the FAST keypoints. The orientation of the detected keypoints is computed using the intensity centroid while the keypoints are described using a rotated Binary Robust Independent Elementary keypoint. The Binary Robust Invariant Scale Key points referred to as BRISK is proposed in [2]. It is a scale invariant feature detector in which keypoints are localized in both scale and image plane using the modified version of FAST. In [2], the strongest keypoints are found in octaves by comparing 8 neighbouring scores in the same octave and 9 scores in each of the immediate neighbouring layers above and below. In BRISK, keypoints are described by computing a weighted Gaussian average over a selected pattern of points around the points of interest and thus achieves invariance to rotation. BRISK is however regarded as a 512 bit binary descriptor. Fast Retina keypoints (FREAK) is proposed in [3]. It is an improvement over the sampling pattern and binary comparison test approach between points of BRISK. The pattern of FREAK is motivated by the retina pattern of the eye. However, in contrast to BRISK, FREAK employs a cascade approach for comparing pairs of points and uses 128 bits as against the 512 bits obtained in BRISK to enhance the matching process. The Binary Robust Independent Elementary Feature (BRIEF) is one of the first binary descriptors and which was presented in [5]. The descriptor(BRIEF) works by building a bit vector from the result of comparing the intensity patterns of a smoothed image. Even though BRIEF does not measure keypoint orientation, it still can tolerate a small image rotation. BRIEF is computationally efficient and faster in comparison to BRISK and FREAK. The Scale-Invariant Feature Transform (SIFT) is a scale and rotation invariant feature detector and descriptor that is proposed in [6]. SIFT has a wide area of applications in object recognition, image stitching, stereo image, image tracking and 3D reconstruction. The generation of a set of image keypoints using SIFT method involves a stage-filtering approach that includes detection of scale-space extrema, key points localization and key points description. SIFT uses a 4×4 sub-region to divide the gradient location and 8 different

orientations set aside for the gradient angles. The dimension of SIFT descriptor is 128. Speeded Up Robust Feature also referred to as SURF is a keypoints descriptor that is motivated by SIFT. SURF is proposed in [7] – a robust keypoints detector and descriptor based on the Hessian matrix. It has a wide area of applications that include object recognition, camera calibration, image registration, 3D reconstruction and objet tracking. SURF is computationally efficient with a high degree of repeatability, robustness and distinctiveness compared to other detectors including SIFT. The Harris detector or Harris edge filter as proposed in [8] works by finding keypoints in image area in which the matrix of the second order derivatives has two large eigenvalues. In [9], Features from Accelerated Segment Test detector also known as FAST is proposed. FAST works by comparing the intensities value of a pixel with its circular neighborhood pixels. The Local Binary Pattern is a local descriptor that works by acquiring the intensity value of an image in a small neighborhood around a central pixel. The local binary pattern consists of a string of bits in which each pixel in the neighborhood is represented by one bit. These binary patterns are rarely used directly instead they are first quantized and transformed into a histogram. LBP was made famous by the work presented in [10].

3. Methodology

The block diagram of our cascaded Upright FAST Harris and BRIEF method is shown in Figure 1. In the diagram, keypoints from input images are detected using the U-FAH method and for every keypoint extracted, its descriptions around the neighborhood are computed using the BRIEF method

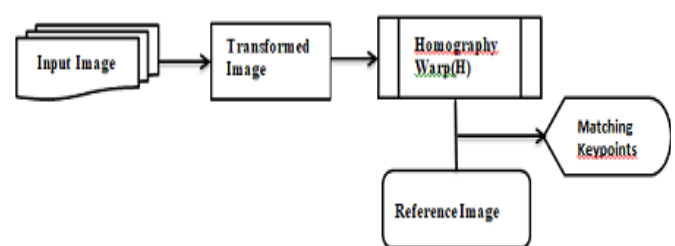


Figure 1: Schematic Diagram of keypoints matching between pair of images

As depicted in Figure1, keypoints and their descriptions are computed for both the transformed image and the reference images using U=FAHB. Subsequently, Homography warp is computed for the transformed image in order to align each of the transform images with the reference image. To keep the text clean and concise, we have a complete discussion of our proposed method in section 3.2 of this paper.

3.1. Dataset

In this work, dataset from real images that represent different types of scenes (see Figure 2) are extracted and

the recall and 1-precision criterion with regard to matching descriptor are used to evaluate the performance of U-FAHB against other state of the art descriptors.

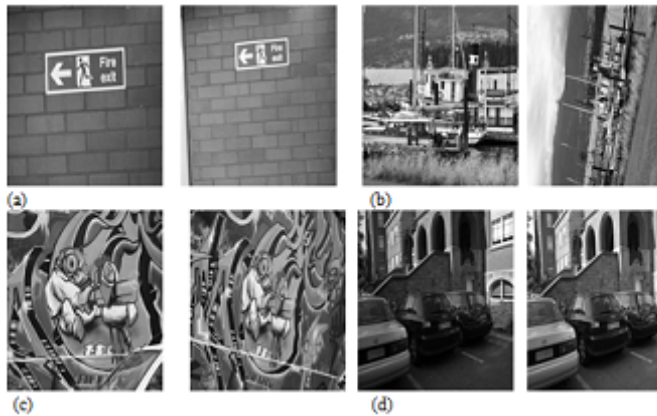


Figure 2: Image pairs representing different imaging conditions such as: (a) scale changes, (b) image rotation, (c) viewpoint changes and (d) illumination change. These image are standard images for evaluation as proposed in [11]

Figure 2(b) shows a pair of image rotations. The first image is referred to as the reference image, while the second image is obtained by rotating the camera optical axis. The angle of rotation in this case is 30 degrees; the average rotation angle in the dataset. Figure 2(c), shows a pair of image under varying view in which the first image referred to as reference, the second image is obtained by changing the camera position at 20 degree. The pair of image under illumination is shown in Figure 2(d). The illumination pair is obtained by a decreasing illumination

3.2. Upright FAST- Harris Filter with BRIEF

The modular approach employed to the design of Upright FAST-Harris Filter as proposed in [4] offers the benefit to combine UFAH with other descriptors. However, given the low computational resource of a mobile phone, it is important to combine UFAH with a computationally efficient descriptor. In this paper therefore, we consider the Binary Robust Independent Elementary Feature descriptor proposed in [5] due to its computational efficiency and speed. A complete discussion on UFAH can be found in [4]. Hence, in this paper, and for clarity, we restrict the discussion of our cascade approach to Binary Robust Independent Elementary Feature only.

3.2.1. Binary Robust Independent Elementary Feature

Binary Robust Independent Elementary Feature descriptor, referred to as BRIEF is a light-weight and simple to use descriptor of an image patch that is composed of a binary intensity test. The intensity test τ of a given smoothed image patch p is defined as follows:

$$\tau(p; x, y) = \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $p(x)$ is the intensity of the pixel within the smoothed patch p at point x . Here the outputs of the binary test are concatenated into a vector of n bits that is referred to as the descriptor. This vector of n bit string can be defined as:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (2)$$

While different types of test distributions are proposed in [5], we employed the Gaussian distribution around the center of the image patch for a better performance. From our test result, we observed that BRIEF descriptor with 512 length gave a better performance compared to the 256 employed in ORB. In order to reduce the noise associated with individual pixel when performing the binary test operation, a smoothing operation is applied to the image patch. Here, we use an integral image similar to the one used in [8] to perform the image smoothing operation.

3.3. Result of Matching Pair of Images using U-FAHB method

The dataset from images captured under scale change (see Figure 2a) is extracted by varying the camera zoom in the range 0 and 2.5 scale ratio. For all transformed image, the homography warp H that align each of the transformed images with the reference image is computed.

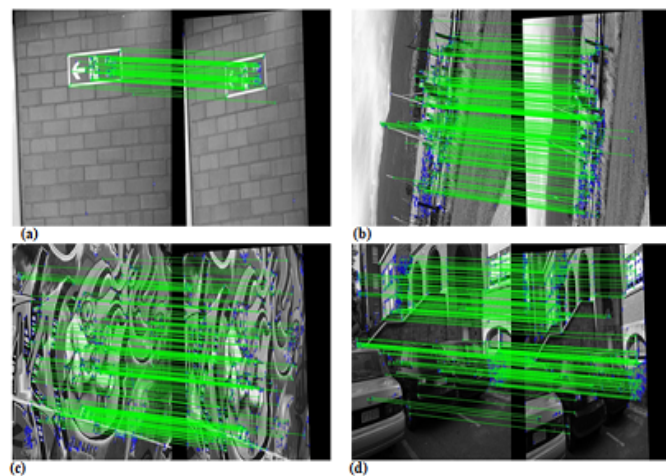


Figure 3: The result of matching descriptors from pair of image under (a) Scale change (b) Image rotation (c) View change and (d) Varying illumination

Figure 3(a) shows the result of matching the first image with the second image under scale change.

The dataset from images captured under rotation (see Figure 2b) is extracted by rotating the camera optical axis. In this experiment, the angle of rotation of the second image from the reference image is given as 30 degree representing the average rotation angle in this experiment. For each descriptor in the reference image, its nearest neighbor descriptor in the second image is computed and then cross checks their consistency in both directions to reduce false matches. The result of matching the first image with the second image observed under rotation is shown in Figure 3(b).

The dataset from images captured under view change (see Figure 2c) is extracted by changing camera position from a front-parallel view to more foreshortening. The view point angle of the second image from the reference image is given as 20 degrees. For each descriptor in the reference image, its nearest neighbor descriptor in the second image is computed and then cross checks their consistency in both directions to reduce false matches. Figure 3(c) shows the result of matching the first image with the second image under view change.

The dataset from the images captured under varying illumination (see Figure 2d) is extracted by changing the camera aperture. Figure 3(d) shows the result of matching the descriptors from the first image with their nearest neighbor descriptor in the second image observed under varying illumination.

4. Performance Evaluation of Keypoint Descriptors

The joint performance of the Upright FAST-Harris Filter and the BRIEF descriptor is compared with the state of the art descriptors using the recall and 1-precision metrics. Given a pair of images, feature points and their description are computed for the reference images as well as for the transformed images using the appropriate methods. For each keypoint in the reference image, a nearest neighbor in the transformed image is located followed by a consistency check in both directions to reduce the number of false matches. Subsequently, the number of positive matches and the false matches are counted and the results are plotted using the recall vs 1-Precision curve.

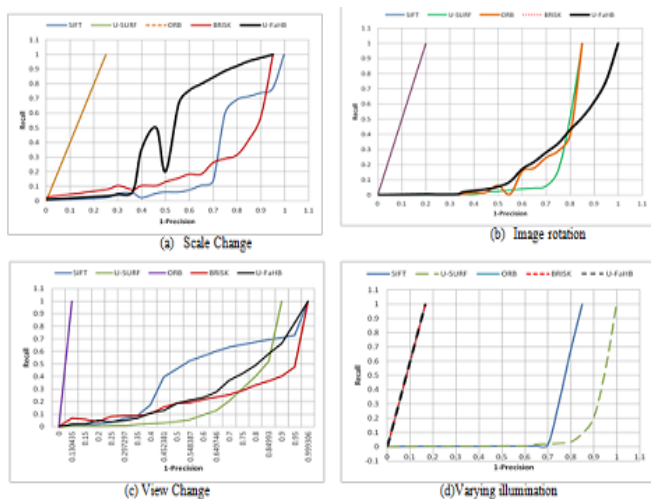


Figure 4: The Precision-recall curves for SIFT, U-SURF, ORB, BRISK and UFAHB descriptors using different dataset extracted from images observed under (a) Scale change (b) Image rotation (c) View change (d) Varying illumination

While recall in this context corresponds to the number of positively matched regions in relation to the number of corresponding regions obtained for a pair of image and is therefore expressed as:

$$recall = \frac{\text{number of positive matches}}{\text{number of corresponding regions}} \quad (3)$$

1-Precision on the other hand corresponds to the number of false matches in relation to the sum total of positive matches and false matches, which can be expressed as:

$$1 - precision = \frac{\text{no of false matches}}{\text{no of (+) matches} + \text{no of (-) matches}} \quad (4)$$

The recall vs 1-Precision curve for dataset from image observed under scale change is shown Figure 4(a). As can be observed from the graph, ORB and U-SURF have better performance on scale changes compare to SIFT, BRISK and UFAHB.

In Figure 4(b), the dataset extracted from image under rotation for all descriptors is plotted using a recall vs 1-Precision curve. The result shows both SIFT and BRISK to have similar performance and better score on image rotation than the remaining descriptors.

Figure 4(c) shows how well each descriptor has performed on a dataset extracted from pair of images observed under view point change. As observed from the graph in Figure 4(c), ORB descriptor has the highest score in terms of the number of correctly matched descriptors.

The recall and 1- precision curve obtained for all descriptors using a dataset from image observed under varying illumination is shown in Figure 4(d). Here, ORB, BRISK and UFAHB have the best performance for a small number of keypoints regions detected in images of decreasing illumination.

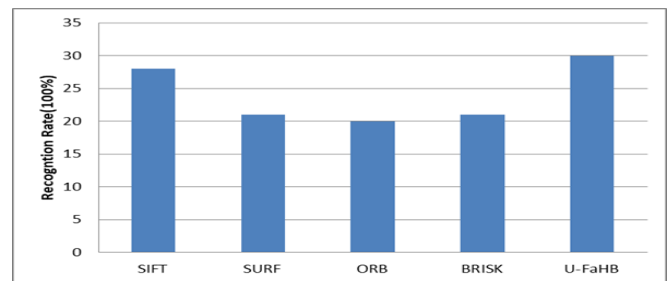


Figure 5: The recognition rate as obtained for all the algorithms (SIFT, SURF, ORB, BRISK and U-FaHB)

In Figure 5, the rate of recognition as observed by the different descriptors is shown. As can be deduced from graph, the Upright-FAST Harris combined with Binary Robust Independent Elementary Feature descriptor recorded the highest recognition rate as compare to the other descriptors.

Table 1: Description time in millisecond across all dataset

Descriptor	Average description time(ms)
SIFT	2.94643
U-SURF	2.52788
ORB	0.199153
BRISK	0.040877
UFAHB	0.086515

Table 1 and Figure 6 show the average time it takes for each descriptor to describe a feature region. From Table 1 it can be observed that BRISK has the fastest description time. This is followed by UFAHB, ORB, U-SURF and SIFT in that order.

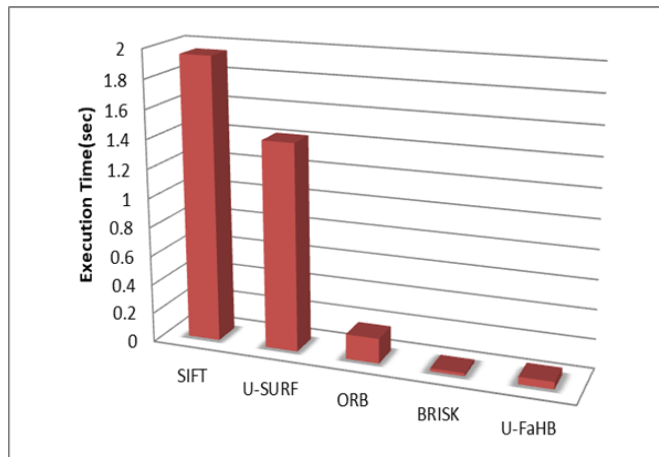


Figure 6: The Average execution time recorded for the different descriptors (SIFT, U-SURF, ORB, BRISK, U-FaHB)

5. Discussion and Conclusion

All descriptors are fairly evaluated using a different dataset of images that are exposed to different photometric and geometric transformations including scale change, rotation, viewpoint change and illumination change. Their performances are analyzed using the precision and recall curve. The description time, which is critical for real time performance is recorded for each descriptor using the same experimental setup.

The recall and 1-precision curve for a pair of images under scale change between 0 and 2.5 is evaluated. In this test (see figure 3a), ORB outperformed the other descriptors, with BRISK and UFAHB following closely in that order. This however, shows that descriptors based on the bit pattern performed extremely well in situation where the scale of an image varies. The recall and 1-precision curves obtained for a pair of images under rotation is evaluated to demonstrate the invariant nature of different descriptors (see figure 3b). The image rotation is 0-30 degrees representing the average rotation in the dataset. SIFT and BRISK outperformed other descriptors followed by UFAHB, ORB and U-SURF. This thus indicate that both SIFT and BRISK perform extremely well under rotation. The recall and 1-precision curve under viewpoint change was evaluated between two images whose viewpoint angles lie between 0 and 20 degrees. Looking at the curve in figure 3c, it is obvious that ORB descriptor is not distinctive even though it is able to match correctly a small number of keypoints correspondences. On the other hand SIFT is highly distinctive compared to other detectors under viewpoint changes, followed by UFAHB and BRISK. The robustness of each descriptor to illumination change was evaluated on a pair of images with decreasing brightness. In this test as shown in figure

3d, UFAHB, ORB and BRISK outperformed the other descriptors showing the robustness of bit pattern to illumination changes.

In order to evaluate the potential of individual descriptor for real time performance, the execution time for each descriptor was analyzed (see table 1). The results obtained show that BRISK is the fastest descriptor and closely followed by UFAHB. Even though, the description time recorded by UFAHB is slow compare to BRISK, it boasts of accurate recognition since all sample points are involved in the matching process. SIFT and U-SURF are the slowest of the descriptors due to their computational complexity.

In conclusion, descriptors based on binary patterns are faster to execute under different imaging conditions. Therefore, the combination of UFAH and BRIEF is promising especially for devices with low computational power.

Conflict of Interest

The authors declare no conflict of interest.

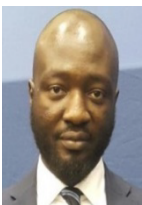
References

- [1] E. Rublee et al., "An efficient alternative to SIFT or SURF," *International Conference on Computer Vision* pp. 2564- 2571, 2011, doi: 10.1109/iccv.2011.6126544.
- [2] S. Leutenegger et al., "BRISK: Binary robust invariant scalable keypoints," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2548-2555, 2011,doi: 10.1109/iccv.2011.6126542.
- [3] A. Alah et al., "Fast retina keypoint," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510-517,2012, doi: 10.1109/cvpr.2012.6247715.
- [4] A. D. Mohammed, A. M. Saliu, I. M. Kolo, A. V. Ndako, S. M. Abdulhamid, A. B. Hassan and A. S. Mohammed, "Upright FAST-Harris Filter," *i-manager's Journal on Image Processing*, vol. 5, no. 3, pp. 14-20, 2018, doi: 10.26634/jip.5.3.15689.
- [5] M. Calonder, V. Lepetit, C. Strecha and P. Fua. , "BRIEF:Binary Robust independent elementary features," *European Conference on Computer Vision*, 2010, doi.org/10.1007/978-3-642-15561-1_56.
- [6] Lowe D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*,vol. 60 no. 2, pp. 91-110,2004, doi:10.1023/b:visi.0000029664.99615.94.
- [7] H. Bay et al., "Surf: Speeded up robust features," *European Conference on Computer Vision*, pp. 404-417, 2006, doi:10.1007/11744023_32.
- [8] Harris, M. Stephens, "A Combined Corner and Edge Detector," *Alvey vision conference* , pp. 147-151.,1988, doi:10.5244/c.2.23.
- [9] Rosten E., Porter R., Drummond T., "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105-119., 2010, doi: 10.1109/tpami.2008.275.
- [10] T. Ojala, T. Maenpana, D.Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," *Proceedings of the 12th IAPR International Conference on Computer Vision and Image Processing*, Vo 1, pp. 701-706.,1994, doi: 10.1109/icpr.1994.576366.
- [11] K. Mikolajczyk, C. Schmid., "A performance evaluation of local descriptors," *IEEE Transaction on Pattern Analysis and and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005, doi: 10.1109/tpami.2005.188.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



DR. ABDULMALIK DANLAMI MOHAMMED is a Lecturer in the Department of Computer Science at the School of Information and Communication Technology, Federal University of Technology, Minna. He received his PhD in Computer Science from The University of Manchester, United Kingdom, MSc in Computer Science from Belarussian National Technical University, Minsk, Belaruss and BSc in Computer Science from Saint Petersburg State Electro-Technical University, Saint Petersburg, Russia. He has published many academic papers in reputable International Journals, Conference Proceedings and Book chapters. Dr. Abdulmalik Mohammed is the Founder and CEO of Korasight Technovation hub. His research interest includes but not limited to Data Science, Feature Engineering for predictive models, Feature extraction and description for pattern recognition, the application of Machine learning and Deep learning techniques for Emerging Technologies such as the Internet of Things (IoT), Big Data, Computer Vision and Image processing. Dr. Abdulmalik Danlami Mohammed is a member of Nigeria Computer Society (NCS) and International Association of Engineers (IAENG).



DR. OLUWASEUN A. OJERINDE is a lecturer in the Department of Computer Science in the School of Information and Computer Technology in Federal University of Technology, Minna. He bagged his B.Sc. in Computer Technology at Babcock University in 2006. He received his M.Sc. in Mobile Communication System from Loughborough University in 2008. He also obtained his PhD in Mobile Communication System from Loughborough University in 2014. His research area is in Antenna, On-body systems, Multiple Input Multiple Output (MIMO) systems, spanning, Telecommunications, Networking and Radiation. He has worked on the effects of metallic objects on radiation for mobile devices. He is a member of CPN, IEEE and IET.

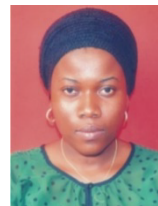


SALIU ADAM MUHAMMAD, received B. Tech. Mathematics/Computer Science from Federal University of Technology, Minna, Niger State- Nigeria, MSc. Computer Science from Abubakar Tafawa Balewa, Bauchi, Bauchi State Nigeria. He was a PhD. Student in Computer Science & Technology Department, School of Information Science and Electronic Engineering, Hunan University, Changsha, Hunan Province – PR. China, but could not

complete the programme owing to financial issues. His PhD programme is currently in view. He was a lecturer in the Department of Mathematics/Computer Science and currently a lecturer in the Department of Computer Science, School of Information & Communication Technology, Federal University of Technology, Minna, Niger State – Nigeria. He has authored and co-authored Thirteen papers in Journals (National & International). He has also participated in ten Conferences (all national) – with four papers in Book of Proceedings, three presentations and three without presentation.



DR. ABUBAKAR SADDIQ MOHAMMED has not only valuable experience in broadcasting, computing and networking engineering but many years of experience in lecturing and research. He holds a Doctor of Philosophy (Ph.D) (Micro & Nano Electronics) from Belarussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Republic of Belarus. He obtained an M.Eng. (Communication Engineering) and B.Eng. (Electrical, Computer & Electronics Engineering) both from Federal University of Technology, Minna, Nigeria. He is a member of Professional bodies among which are: The Council for



MRS EKUNDAYO AYOBAMI is a lecturer in the Department of Computer Science at the School of Information and Communication Technology of Federal University of Technology, Minna, Niger State, Nigeria. She received both B.Sc and M.Sc Computer Science from the University of Ilorin, Kwara State Nigeria. Her research interest includes but not limited to Data mining. Mrs Ekundayo Ayobami is a member of Nigeria Computer Society (NCS). She has published many academic papers in reputable journals and conferences.

Biodigester and Feedstock Type: Characteristic, Selection, and Global Biogas Production

Abdulhalim Musa Abubakar *

Department of Chemical Engineering, Faculty of Engineering, Modibbo Adama University (MAU), P.M.B 2076, Yola, Adamawa State, Nigeria

Corresponding author: +2347050244277 & abdulhalim@mautech.edu.ng

Corresponding author ORCID: 0000-0002-1304-3515

ABSTRACT: This work aims at providing factual details necessary for the utilization of diverse feedstock for anaerobic digestion (AD) to produce biogas using either conventional or non-conventional types of digesters. This is necessary as different substrates had peculiar merits or potentials of biogas production due to their unique characteristics. Selection of right feedstock is usually based on sustainability, quantity, output requirement, availability and metallic nutrient content apart from digester type which is affected by the weather condition of the location among other factors. Global biogas production is increasing annually, especially in areas of biogas utilization for electricity generation, heating and fuel for transportation.

KEYWORDS: Biogas, Anaerobic digestion, Feedstock type, Bioreactor, Chicken manure

1. Introduction

Biomass are organic matter of plant and animal products that can be broken down physically or chemically [1, 2]. The simple act of breaking down these materials into smaller pieces could be termed physical decomposition while the use of microorganism is a chemical biodegradation process. The biodegradation process is carried out with or without oxygen. Aerobic fermentation is referred to as oxygen-free decomposition while anaerobic is vice-versa. The composition of the biogas produced from the respective processes are almost similar [3]. Biogas can be synthesized using different biomass sources providing an oxygen-free environment in the presence of anaerobic microorganisms [4–6]. It can be called a ‘cell gas’, as it comes from “biogenic materials” [3, 7-8]. The process of converting organic waste to biogas is termed AD. It was first introduced in 1870 by Jean-Louis Mouras [9]. AD is an alluring, slow, versatile biotechnological commercial route to transform organic

waste or biodegradable material to useful resources by consortium of microorganisms living symbiotically [6-7, 9–13]. This biological process are carried out purposely to produce biogas and digestate [7, 14]. Biogas which is known for its composition of methane, carbon dioxide, oxygen, nitrogen, water, ammonia, hydrogen, siloxanes and hydrogen sulphide vary in this components composition if the substrate processed are not the same. In this case, the component that make up the biogas, methane is often targeted, because it is well pronounced in the characteristics of the gas. While the second co-product is used as fertilizer. It can be called biological fertilizer, an alternative of the conventional or chemical fertilizer. Merits and drawback lies in the use of the AD process. Major advantage is the contribution to conservation of non-renewable energy sources and biogas production, while long retention time and low heating value of the produced gas, constitutes a disadvantage [12, 15] as shown in Table 1.

Table 1: Advantages and Disadvantages of AD Process [8]

Advantages of AD Process	Disadvantages of AD Process
<ul style="list-style-type: none"> Operating costs for an anaerobic treatment plant are relatively very low when compared with aerobic treatment plant Low-energy consumption. Also, the system does not require external energy for its operation The flexibility of an anaerobic system allows the technology to be applied on either a small or a large scale Low sludge generation compared to aerobic systems due to a lower yield coefficient The excess sludge is well stabilized thereby resulting to limited environmental impact Low nutrient and chemical requirement Allows for efficient resource recovery, and conservation of non-renewable energy sources 	<ul style="list-style-type: none"> Long start-up: the slow growth rate causes a longer start-up period as compared to aerobic systems High buffer requirements for the pH control: The required pH for AD should be in the range of 6.5–8 High sensitivity of microorganisms: Methanogens are sensitive to pH and temperature Low pathogen and nutrients removal Process is more sensitive to the presence of toxic compounds and changes in temperature

It is well known that feedstock to anaerobic bioreactors are basically biodegradable organic wastes. Millions of tonnes of waste is generated annually across various countries of the world, in both developed and developing countries [3]. India alone produces cumulatively '44 million tons + 500 Mt' annually of waste [10]. These wastes can be subjected to various waste treatment methods to recycle or convert them to useful materials [16]. In Table 2, biogas yields of various feedstock are compared; where it is observed that different feedstock produces different throughput of biogas.

Table 2: Comparison of biogas yields and electricity produced from different potential substrates [10]

Type	Biogas production (m ³ /ton fresh matter)	Produced kilowatt-hours (kWh) per ton fresh matter
Cattle dung	55-68	122.5
Chicken litter/dung	126.0	257.3
Fat	826-1200	1687.4
Food waste (FW)	110	224.6
Fruit waste	74	151.6
Horse manure	56	114.3
Maize silage	200-220	409.6
	101.5	207.2
	11-25	23.5

Municipal solid waste (MSW)	47.0	96.0
Pig Slurry		
Sewage sludge		

During AD, feeding the biodigester continuously with water and organic waste is necessary to keep it running. Biogas automatically stops coming out of the outlet of the biodigester when the feedstock runs out [17]. Not all feedstock will require the addition of water. Where water is required to be added, this has to be done making sure the contents of the digester is not too diluted. Not adding the required amount, affects the working principle of the digester [17].

2. Conventional Biodigesters

There are three conventional reactor models used for biogas production, namely, fixed dome reactor model, floating drum digester model and the tube digester model [18]. Fixed-dome biogas digester, also called the hydraulic reactor model is a semi batch reactor originating from China in 1936, and it is the most commonly built [6, 19–21]. It has an underground pit or manhole lined with reinforced concrete slab or bricks in order to protect it from physical damage from excessive pressure and to save space [3, 21]. The filling hole or fermentation chamber is where organic waste resides for AD take place. It also have an inlet to add feed to the

digester, a non-movable or fixed dome gas holder or gas reservoir, constructed with granite, sharp sand, cement and iron rods at the upper part of the reactor for gas storage, a displacement or compensation tank where displaced slurry enter during gas production and an outlet pipe with a valve/faucet attached to the dome top where the emerging biogas flows to be collected in a storage tank also connected to the outlet pipe [3, 6, 21, 22]. Gas is produced under pressure. The volume of gas generated is proportional to the gas pressure and implies that, as the pressure increases, gas volume increases. The advantage of the Chinese digester is that, it is simple, occupies less area, having relatively low cost of construction, having a long lifespan of above 20 years and consist of no moving parts [3, 23]. The disadvantage of this type of digester model is that gas pressure is not stable (caused by absence of gas valve in the outlet) and is prone to cracks in the gas reservoir [22]. The amount of waste and water to use as feed and the local climate are some of the factors important to consider during its design [6]. Centre for Agriculture Mechanization and Rural Technology (CAMARTEC) model designed by GIZ for use in Tanzania, French model installed in Pakistan and Deenbandhu 2000 (a modification of the Janta model) developed in India are three variations of the fixed dome plant [24, 25].

The floating drum biogas plant was first designed in India and built with sand, cement materials and bricks [23]. It comprises of a cylindrical or dome-shaped underground digester and a moving gas holder floating over the fermentation slurry. This floating iron drum is placed upside down to hold the gas produced. The drum

moves up and down based on the volume of gas stored (up, when the gas increases and down, when it decreases) and a guiding frame is kept to prevent it from tilting [3]. The disadvantage of this design is that, it has a shorter life span compared to the Chinese model, it is prone to corrosion due to the high cost of steel drum and requires regular painting during its maintenance. The strength of the model is its simplicity in building and easy operation [22]. The balloon or tubular biodigester model is set up using a large and strong plastic bag placed on a dug trench for its safety and linked to a piece of drainpipe at either end [23]. These pipes are the outlet and exit pipes (which is at one top end) for discharge of slurry and removal of gas respectively. The top of the bag starts to inflate as biogas is produced and is piped away [17]. Gas pressure can be increased when some loads are placed on top of the bag. Balloon bioreactor merits are namely, easy to clean, simple construction and operation, easy to relocate, lightweight, easy to install and cheap manufacturing cost [26]. The biggest disadvantage associated with the model is its ability to get damaged easily and a short lifetime of 4 years [22]. It is also called the polyethylene tubular digester, prominently in use in South Africa, Vietnam, Cambodia, Colombia, Ethiopia, Bangladesh, Tanzania and countries of Latin America [21, 23, 25]. Unlike laboratory scale bioreactors where lots of research has been carried out on feedstock characteristics inside the digester (e.g. measure of cell and substrate concentration) and kinetic studies, the conventional reactors have not been so given attention. The distinguishing factors of the three digester types is shown in Table 3.

Table 3: Comparative table of the digesters [19, 21]

	Fixed Dome Plant	Floating Drum Plants	Balloon Plants
Lifespan	20 years or more	15 years maximum	2-5 years
Size	5-200 m ³	5-15 m ³	4-100 m ³
Investment costs	Low	High	Low
Cost of maintenance	Low	High	Low
Gas pressure	Between 60 and 130 mbar	Up to 20 mbar	Low gas pressure
Skilled required	High	High	Low
Methane emission	High	Medium	Low

Presently, there is the fourth kind of digester termed fiberglass digester consisting of a digesting part and storage part of which there is no any barrier or insulation between them. Advantages associated with its use is that it is movable, environmentally friendly, has a lightweight, has low investment cost, leaks are easily repaired and the ease of implementation and handling [22]. Also, membrane are suitable when managing inhibition triggered by ammonia accumulation because they have the capability of shielding microbes from inhibitors [27]. Membrane bioreactors or anaerobic membrane bioreactor (AnMBR) is a fascinating innovative technology for biogas production. Key setbacks related to an AnMBR system are its maintenance, installation, operating cost and membrane fouling [28]. AnMBR systems, just like conventional bioreactors operates at neutral pH.

2.1. Categories of Biodigesters

Based on method of operation, time and volume of feedstock available, biodigesters are classified into batch, semi-batch and continuous reactor. Batch reactors are in most cases fed with slurry to allow for digestion within a desired retention time and then discharged [5]. The continuous stirred tank reactor (CSTR) is one in which feedstock are continuously fed and products are continuously taken out. It is composed of an internal mixing system, based on wet process, operated in mesophilic conditions [29]. FW are commonly fed in CSTR after mixing adequately with water to prevent pumping and mixing difficulty as FW have high solid

content greater than 15% [30]. Batch reactors are characterized with handling diversified substrate and higher volume compared to continuous digester. The weakness of batch system is that the initial cost and cost of synthesizing the same amount of biogas is almost twice that of continuous digester [19].

Apart from those three, although base on them, other type of digesters are anaerobic contact reactor (ACR), anaerobic pond, internal circulation reactor (ICR), anaerobic sequencing batch reactor (ASBR), up-flow anaerobic solid-state reactor (UASS), anaerobic plug-flow reactor (APFR), anaerobic baffled reactor (ABR), up-flow anaerobic sludge bed reactor (UASB), anaerobic fluidized bed reactors (AFBR), up-flow anaerobic sludge fixed-film (UASFF) reactors, membrane anaerobic system (MAS), modified anaerobic baffled reactor (MABR), ultrasonic membrane anaerobic system (UMAS), expanded granular sludge beds (EGSB) and upflow anaerobic filtration (UAF) [31–34]. The UASB is currently in use in Morocco to treat recycled paper mill wastewater because it is a suitable environment for the survival of microorganism as it permits a dense sludge system [28, 33]. EGSB has been tested by [33] to digest palm oil mill effluent (POME) in Nigeria and was found to yield the highest methane gas at low retention time over other types. The plug flow system allows the treatment of high amount of waste per unit digester volume, requires little or no water and reduces the need for pretreatment [30]. All the listed digester type have the advantage of large-scale application [31]. Specifically, two types of digesters for FW is shown in Figure 1.



Figure 1: FW Biodigesters (a) Green Cone FW Digester and (b) Ecofys Plastic Bag Digester

Green Cone FW Digester created by a British company called Great Green Systems is however, producing small amount of biogas, but the Ecofys Plastic Bag Digester, a technology of Netherland is made up of recyclable waste. Ecofys digester is considered portable, cheap, very easy to install and could last for 8 years. In terms of size, biogas plants are divided into small size plants (household scale) capable of delivering 0.5m³/day of biogas and large sizes (industrial scale) that can generate 2500 m³/day [26]. According to a United Nations Development Programme (UNDP) report, it takes 5-8 pigs, 4 adults human and 1-2 cows to supply sufficient substrate daily for a single-household bioreactor [35]. Countries like Lesotho, Uganda, Zambia, Zimbabwe, Burundi, Botswana, Nigeria, Sudan and Swaziland are known for small-to-medium scale digesters while Rwanda and South Africa had several large-scale digesters [24, 35, 36]. Between 2007-2012, Non-Governmental Organisations (NGOs) have help develop 17000 digesters cumulatively, in Rwanda, Tanzania and Ethiopia [37]. The same author stated that the population of Africa is projected to reach 2 billion by 2050, but currently the number of household size plants is just 18.5 million compared to China's 200 million, showing a huge deficit and explains how far the continent is left behind as regards investing in the technology. Small sized and large digesters in Bangladesh in 2014 cumulatively generates 15 billion m³ of biogas, and plans to build additional 100,000 small scale biogas digesters in 2020 [21]. Several factors are considered when designing biogas plants, namely, plant safety, easy maintenance, weather condition, biogas output requirement, availability of feedstock, sustainability, easy operation among others. The priority of building a biogas plant is basically to minimize cost as it is a rural based technology where local materials are channeled to its construction [38].

3. Feedstock Classification for AD

Agricultural, municipal, and industrial wastes are the three primary kinds of organic waste [4, 32]. Livestock waste, harvest waste, grass and algae, energy crops, garden waste, and vegetable by-products are examples of agricultural wastes. Slaughterhouses, ranges, insect farms, and poultry houses all provide livestock wastes. Fish waste, insects and worms, poultry litter, keratin-rich waste, and manures are among them [39]. Insect farming technology produces biogas that is comparable to animal waste by cultivating silkworm and caterpillar excreta

[40]. Poultry litter is a type of lignocellulosic bedding made up of spilt feed, excrement, and feathers [41]. Wool, horns, chicken feathers, hooves, hair, claws, and nails produced by the fish, meat, and wool industries are examples of keratin-rich waste [27].

Animal waste is referred to as manure, which is made up of a combination of water, straw, excrement (feces and pee), and sand [29]. Cattle, goats, chickens, pigs, deer, horses, and other animals are the source of manures [42]. A kg of cow dung can generate 0.03-0.05m³ of biogas whereas 50,000 cattle could deliver around 20000 m³/day of biogas [25, 40, 41]. In 2006, Food and Agricultural Organisation (FAO)'s findings shows that there are up to 277 million heads of cattle in Africa [43]. Nigeria's biogas potential stood at 6.8 million m³/day from animal manure while in Zambia, animal dung potential for the gas is 1.473×10⁹m³ [41, 43-45]. From the 21 million cattle population in 2001, estimates of daily manure production in Nigeria ranges from 210-1260 million kg ([44] gave a specific amount = 227,500 tons daily) while its annual capacity is 76.7-450 million tonnes [46]. The aggregate is 542.5 million tons/yr of livestock waste with potential of 25.53 billion m³ of biogas, 169541.66 MW of energy and 88.19 million tonnes of biofertilizer [39, 41, 47]. Third largest cattle breeders in Africa is Tanzania followed by Ethiopia and Sudan; boosting of 40 million animals out of which 18.8 million are cattle [36]. Distribution of manures across states and regions of seven countries (namely, Germany, Austria, Australia, Norway, Canada, Ireland and the United Kingdom) and their potential for biogas production had been presented previously by [48]. Total weight of waste generated from a certain location or town per year at every animal slaughter house can be estimated based on Equation 1 [20],

$$M = [E \times A_m + N \times (A_b + A_r)] \times 365 \quad (1)$$

where, M = total amount of waste produced in that province (kg year⁻¹), E = total number of live animals, N = total number of animals slaughtered, A_m = amount of manure produced (kg day⁻¹) and A_b & A_r = amounts of blood and rumen (kg day⁻¹) produced at slaughter houses.

Silage grass, mushroom stick, and algae are all good agricultural AD substrates [42]. As illustrated in Figure 2, there are three varieties of microalgae for commercial biogas generation.

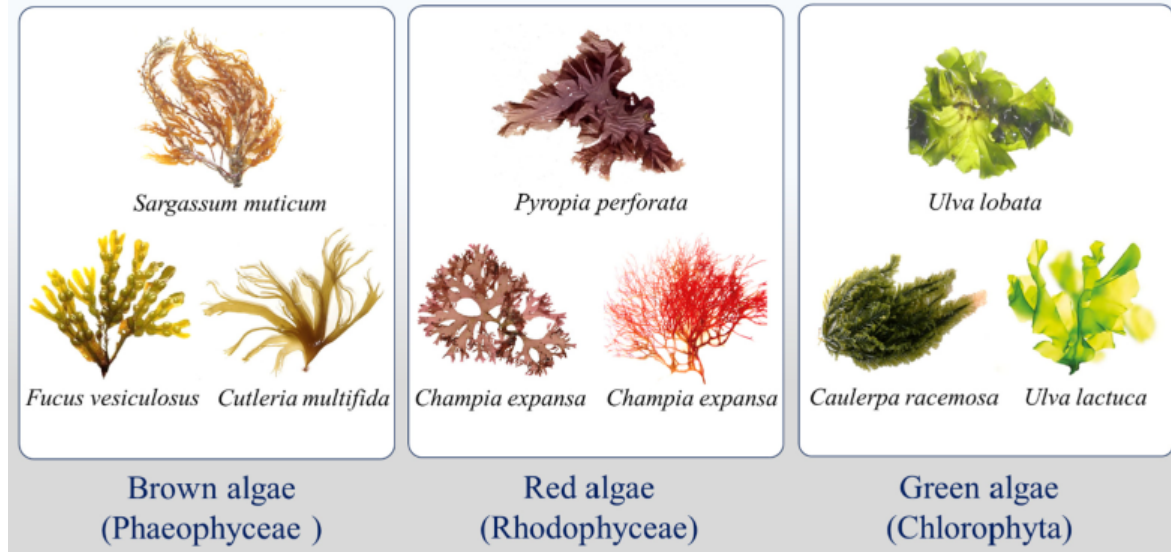


Figure 2: Examples of Different Types of Microalgae [49]

Red algae (*Rhodophyceae*) are the most common of these algae kinds [49]. Energy crops are grown specifically for the production of biofuels and biogas. Vinasse is a liquid leftover from sugarcane derivatives, sugar beet, and cassava, with a significant biogas synthesis potential [10, 13]. Leaves, garden clippings, plants, and cut grass are all examples of garden waste (GW) [50]. In many areas, harvest or agricultural wastes such as olive pomace, POME, stalk, straw, plant trimmings, and bark are a key source of AD feedstock [51–53]. Anaerobic bacteria have a hard time degrading crop straw to produce biogas [54]. Corn stover is made up of husks, stalks, leaves, and cobs that comes from corn grain production [55]. Wheat is one of the most widely grown crops on the planet, while silage corn is the most important crop for biodigester plants [46, 56]. The potential of crop residues for biogas generation in Zambia is $1.819 \times 10^9 \text{ m}^3$ according to [56].

Though there is no such thing as municipal liquid waste (MLW) in the literature, it is legitimate to separate municipal garbage into solid and liquid waste. FW, municipal waste water, landfill garbage, papers, green waste, urban sanitation, and aquatic biomass are only a few examples. These wastes come from a variety of places; possibly domestic, educational, medical, to mention a few [57]. FW is described as an uneaten, abandoned, or lost substrate of foods such as rice, noodles, nuts, pasta, eggs, fish, vegetables, fruits, meat, potato, and sweets during the stages of manufacturing, processing, distribution, and consumption [29, 40, 49–51].

Restaurants, canteens, markets, hotels, hostels, and household products all produce large volumes of FW [30]. Some of these wastes come from the listed buildings' and locales' kitchens [12]. In the kitchen, filthy water and various food remnants are produced. Vegetable residue, fruit peels, cooked food leftovers, and spices are just a few examples. Spices aren't good for the AD process. Red chili, black paper, cinnamon, coriander, garlic, turmeric, cardamom, and clove are some of the spices used [15]. Daily biogas output from FW in Benin metropolis of Nigeria of up to 28836.91 m^3 was generated from approximately 305.075 tonnes/day production rate based on [58].

Low biogas recovery is typical of municipal waste water [59]. Sewage sludge and gutter water are two examples [3]. Almost every sort of waste is accepted at the landfill, which is divided into organic and inorganic subtypes. MSW disposed off in a landfill by composting or open dumping is an example [4]. Leachate is generated in dump sites and produces a significant amount of biogas during AD [28]. School and printing presses both have paper. Cardboard, filter paper, waste paper, newspaper, and tissue paper are among them [60]. Estimates of Nigeria's potential of MSW was put at 17 million tonnes [61]. Aquatic biomass could include lignocellulosic biomass from aquatic weeds like water hyacinth (*Eichhornia crassipes*) and water primrose (*Ludwigia hyssopifolia*) [6, 54, 55].

Some of the industries that generate semi-solid and liquid organic waste for anaerobic digesters include the agricultural and food processing industries, fodder and brewery industries, wastewater treatment plants (WWTP), textile industries, fruit processing, sugar industry, and pharmaceutical industries [1, 2, 62]. Biomedical waste produces blood, which is also an important biogas source [63]. Sludge is a solid waste byproduct produced by WWTP [64]. Due to water scarcity, the number of WWTP has expanded in recent years, producing either settled primary sludge or waste activated sludge (WAS) from biological treatment [49, 57]. Four WWTP, namely, Abesan, Alausa, Iponri and Oke-Afa are too few for Lagos state (the most densely populated land area in Nigeria), being the largest emerging cities in the world with a small area of 3577 sq. km – that generates 1.4 billion litres of wastewater per day from a population of around 20 million [65]. In addition, this waste presents an enormous potential for biogas and biofertilizer production for the country.

Fruit processing plants generate two forms of trash [27]: (a) solid waste, which includes stones, skin, peels, and seeds, and (b) liquid waste, which includes washwaters and juice. Banana, watermelon, citrus, mango, pineapple, and other fruits are examples. Waste that can be digested into liquid effluent trash (e.g. wastewater, manure slurry, sewage sludge and agro-food effluents) and organic solid waste can be distinguished (e.g. agricultural, industrial and municipal waste) [19]. Substrates aren't picked at random. Always choose a substrate based on its long-term viability, energy efficiency, environmental impact, and economic value [11, 62].

3.1. Feedstock Characteristics

The composition of the feedstock should be the first deliberation while opting for organic matter that would give high yield of biogas. Handful of organic waste are difficult to break down in a digester, because they are indigestible (e.g. paper and impregnated wood), hard to digest, slow to digest, or contain inhibitors [1, 27]. Quantity and composition of feedstock can affect AD in the following way [66]: (a) dry matter content and viscosity of substrate causing stirring difficulty, (b) impurities affecting size and causing sedimentation, (c) digester's size and shape defined by the slurry volume, (d) feedstock physical and chemical compositions [67]

and (e) content of volatile solids (VS) and ammonia concentration. It is critical to dissect various feedstock and their composition; one which will significantly matter with regards to biogas output.

Animal dung contains parasites, viruses and bacteria (that keeps reproducing during AD) [17]. The most commonly used is cattle dung mixed with hot water (ratio = 1:1) [10]. Animal manure are low in C/N ratio as well as nitrogen content but rich in carbohydrate content [3, 9, 12]. Poultry residue are rich in nitrogen, and is therefore not recommended for efficient AD [3, 68]. Brown macroalgae are seaweed characterize with high polysaccharide content and negligible lignin content [49]. Other plant-based materials like vegetables, root crops, grains and fruits are rich in different polysaccharides [16]. The best pretreatment method for water hyacinth is 5%v/v H₂SO₄ with residence time (RT) of 1 hour [69]. Vinasse has low macro and micronutrients, deficient in carbon-to-nitrogen ratio and unpleasant smell [13]. Hence needs to be codigested with filter cake, straw and bagasse to solve the problem of its nutrient deficit. AD of agricultural waste is portrayed generally as having poor buffering capacity, low quality end products and potential variability [5].

Effluents wastewaters are sourced for AD basically from industries such as wastewater treatment plants. Waste activated sludge (WAS) from wastewater treatment plants (WWTPs) are rich in nitrogen, phosphorus, potassium, organic carbon, microbial biomass and exopolymeric substances; mainly proteins and carbohydrates [51, 70]. Apart from WWTP, wastewaters are sourced from industries discharging them. Municipal solid waste is typified by low chemical oxygen demand (COD) concentration, presence of toxic materials and high concentration of heavy metals [9, 52]. Domestic sewage are rich in nitrogen organisms [3]. Kitchen waste contains high nutritive and calorific value [71]. FW are non-homogenous in nature, has high water, VS and salinity content, low C/N ratio and is highly biodegradable [29, 49, 50]. They are either fat-rich, protein-rich or carbohydrate-rich materials. Fat-rich and protein-rich feedstock produces more methane than carbohydrate-rich feedstock [51].

3.2. Lignocellulosic Feedstock

It has been established that physical and chemical properties of the raw material are factors affecting the amount of biogas produced [70]. In [57], it was stated that characterization, elimination of contaminants, pretreatment, AD in optimum condition and utilization of energy crops enhanced the efficiency of an anaerobic digester. Cellulose, hemicellulose and lignin are lignocellulosic substrates [54, 59]. They are the main elements of the cell walls, require long retention time, have a high C/N ratio, high carbohydrate content and are recalcitrant towards AD process as they are highly nondegradable [4, 16, 60, 62, 69, 72]. High temperature and low retention time pretreatment of lignocellulose could effectively improve the porosity and delignification efficiency [5]. Cellulose and lignin are natural complex polymer found in wood [29]. Cellulose is a carbohydrate while lignin is non-carbohydrate. Examples of lignocellulosic matter are crop residue, corn stover, rice straw, corn stalk, wheat straw, water hyacinth, barley [16] etc.

Urea and acid pretreatment are the predominantly used pretreatment technologies for crop straw [54]. According to [55], corn stover typically comprises of 37.5 % cellulose, 22.4 % hemicellulose and 17.6 % lignin. Corn silage contain high C:N ratio and low ammonia [70]. Hard lignocellulosic structure and high C/N ratio of wheat straw of up to about 100 hinders production of biogas from wheat straw [12, 67]. Codigestion of rice straw is more desired because of its low nitrogen content, lignin percentage and high C/N ratio [5]. Just like tree plants, water hyacinth consisting of stem, leaf and root [69].

3.3. Chicken Manure as AD Feedstock

Chicken manure (CM) is an alkaline, semi-solid organic material that is made up of diverse composition of other organic materials, as well as being one of the most widely used feedstock for anaerobic production of biogas and biofertilizer [64, 73]. In [73], it was reported that daily chicken excretion ranges from 80-125g (wet)/chicken. Dry matter content or total solid (TS) content of CM is 20-25% of the excreta which is rich in nitrogen, with high amount of biodegradable fraction and VS content of 55-65% [65, 66, 73]. Percentage water content of more than 70% in CM is considered unattractive for utilization [74]. It also contain pathogens (methanogenic bacteria), high phosphorus, low C-N ratio, and high salinity level [71,

75]. [76] reported a CM with TS = 47.3%, pH = 8.1, VS = 68% and C-N ratio of 18. To prevent CM from decomposing, prior to AD, they are often kept at a very low temperature of -20°C [77]. Others have reported a higher temperature of 4°C. Dry fermentation has the merits of high biogas production rates, low water consumption and low cost [78].

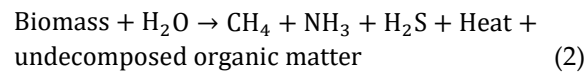
The agricultural sector where CM is derived, is the main source of total ammonia emission into the atmosphere [79, 80]. It has been stated clearly, that the high nitrogen content of CM makes it a suitable material for AD. However, nitrogen, together with sulphur inhibits the digestion process [81]. The level of nitrogen in CM is attributed to the conversion of uric acid and undigested proteins into total ammonia nitrogen (TAN) during digestion [82]. For a successful AD process, the inhibitory effect of ammonia (threshold value of 200 mg/L) as well as the low C-N ratio of CM must be overcome as it may cause volatile fatty acid (VFA) to accumulate and inhibit microbial activities [16, 76, 82, 83]. Air stripping is a pretreatment technique to get rid of ammonia from CM wastewater [74]. TAN inhibition can be reduced by feeding the system with feedstock containing low TS [65]. For a stable AD performance and a balanced nutrient, mono-digestion of CM is often frowned at. Anaerobic co-digestion of CM with other feedstock is mostly carried out by researchers as alternative method of solving the ammonia problem [84]. Example is co-digestion of FW, goat manure and CM and co-digestion of ethanol plant effluent with CM [64, 82] among others.

Amongst all agricultural activities, the poultry sector is one capable of generating huge amount of organic waste [85]. Livestock farmers, especially those handling poultry birds like geese, ducks, turkey, chicken, guinea fowl, quail, ostrich and pigeon in poultry houses disposes off the waste generated from these animals on the environment. Chicken waste are often applied on agricultural land as manure or compost as a traditional treatment approaches, dumped at landfill, or incinerated, contaminating the environment in the process [71, 86]. For instance, too much of nitrogen and phosphorus in CM results in eutrophication during landfill and composting [74]. In addition, chicken waste provides a breeding environment for flies and parasites, pathogen release, eutrophication of surface waters, threat to local air quality when used as fertilizer, pollution to soil, health risks and groundwater contamination [79, 86-88]. Its effect on soil

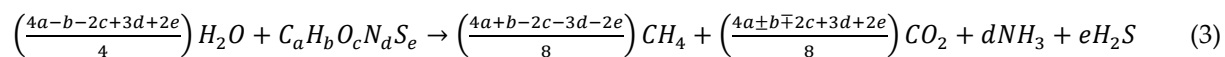
properties is classified into three, and includes physical, chemical, and biological effects. CM in solid or liquid phases used on irrigated soils risks increased soil and groundwater salinity, excessive nitrate leaching to ground water as chemical effect. Biologically, problems arising are; introduction of pathogens, deterioration of soil carbon, and decreasing populations of desirable microbes [71, 82, 88]. The soil physical property like structure/texture may also be affected [89]. Improper disposal of CM waste must be addressed to mitigate its effect on the ecosystem. CM should be pre-process or pretreated by thermo-chemical and/or physical processing technologies like torrefaction, ozone treatment, re-feeding to animals, composting, steam treatment, drying, ozone treatment, pyrolysis, esterification, gasification, co-gasification, fermentation or digestion, combustion and co-combustion [82, 90]. In [91], it was affirmed that, combustion can be a viable and dependable way to treat CM, principally when coupled with energy recovery.

3.4. Biogas Potential of Feedstock Constituents

Anaerobic conversion of organic material is defined by [57] in Equation 2.



Typically, biomass contains carbon, hydrogen, oxygen, nitrogen and Sulphur, depending on the feedstock in varying amount. The empirical formula is simply, $C_aH_bO_cN_dS_e$, where a, b, c, d and e are atomic numbers of the respective elements in the biomass. The degradable fraction of FWs mainly includes carbohydrates ($C_6H_{12}O_6$), proteins ($C_{13}H_{25}O_7N_3S$), and lipids ($C_{12}H_{24}O_6$) [31]. Lipids are found in meat processing byproducts, agro-industrial residues and fatty wastewater; carbohydrate are found in agricultural waste and in organic fraction of municipal solid waste; while proteins are found in waste from slaughterhouses and meat processing industry [64]. Biogas potential of feedstock constituents can be predicted by employing Buswell's empirical formula of Equation 3 [6, 15, 27, 29, 62].



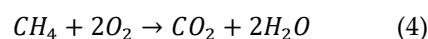
Equation 3 is the modified form of Equation 2. Equation 2 is mainly used to assess methane production

and to stabilize the digesting system, which can also be achieved using stoichiometry in Table 4 [67].

Table 4. Stoichiometry of biogas potential determination from various feedstock components [30, 62]

Feedstock	Methane Formation Stoichiometry	Methane Concentration (%)
Carbohydrate	$(C_6H_{10}O_5)_n + nH_2O \rightarrow 3nCH_4 + 3nCO_2$	50
Lipid	$C_{50}H_{90}O_6 + 24.5H_2O \rightarrow 34.75CH_4 + 15.25CO_2$	69.5
Protein	$C_{16}H_{24}O_5N_4 + 14.5H_2O \rightarrow 8.25CH_4 + 3.75CO_2 + 4NH_4^+ + 4HCO_3^-$	68.8

When the elemental composition is known, the theoretical methane production can be calculated using Equation 3. When both the elemental composition and the proportion of carbohydrates, proteins, and fats are not known, the theoretical methane yield can also be calculated from the COD of the feedstock using reaction 4 [27]:



From Equation 4, 2-kmols of O_2 (or 64 kg COD) are needed for the complete oxidation of 1-kmol of methane,

so 1 kg COD is equivalent to 1/64-kmol of methane or 0.35 m^3 CH_4 at standard temperature and pressure [27].

3.5. Anaerobic Codigestion (AcoD)

Multiple degradable waste may be mixed in the same digester at various combinations and fractions to increase biogas yield [1, 62]. This is called anaerobic codigestion (AcoD). AcoD comes with the merits shown in Table 5.

Table 5: Merits of Anaerobic Codigestion

S/No.	Advantage	References
-------	-----------	------------

1.	Reduction of investment costs and increase in energy efficiency	[64]
2.	Improve the proportion of macronutrient such as carbon, phosphorus, nitrogen and Sulphur	[67]
3.	Upgrade the feedstock for balanced nutrient	[92]
4.	Ensure reactor stability and improved performance	[93]
5.	Benefits of resolving demerits of using single substrate	[2]
6.	To take advantage of different waste streams	[94]
7.	Improvement of biogas production and methane yield	[2]

To optimize the AcoD process of biogas production technique, biodegradability, chemical composition, bioavailability, operational parameters (temperature, pH, loading rate etc), bioaccessibility, addition of nanoparticles, thermodynamic and kinetic model and characterization of substrates are crucial parameters to consider [2]. Enhancement of biogas output is divided into upstream, mainstream and downstream processes [85]. Upstream approaches including fungal pretreatment, enzymatic, microaeration, composting, and ensiling have been employed prior to AD to improve biogas yield and productivity [79]. Mainstream approaches include bioaugmentation, AcoD and integrated biogas production techniques while the downstream approaches focuses on biological removal of

CO₂, impurities and H₂S [94]. Table 6 and 7 is almost a demonstration of how AcoD of multiple feedstock will increase methane yield.

Table 6: Methane Yield from AcoD of MSW with Agricultural waste [4]

Substrate types	Methane yield (m ³ /kg VS)
MSW and activated sludge	0.376
MSW and sewage sludge	0.395
Hydroseparated MSW and sewage sludge	0.333
MSW and activated sludge	0.287
MSW and leachate	0.232
MSW and pig manure	0.377
MSW and cattle manure	0.443
MSW and rice straw	0.403

In several text, methane and biogas yield are used synonymously or interchangeably, especially in kinetic equations to estimate unknown parameters. The units of biogas and methane yield reported in Table 6 and 7 can be converted using equation given by [102] to other units. Either algae or lignocellulosic biomass that limits hydrolysis stage of AD can be codigested with animal manures rich in C/N ratio to avoid the resurgence of NH₃ [12, 40, 103]. To practicalize this in FW, mixing fat-rich materials with carbohydrate-rich materials (fast degradable and slowly degradable specie) are advantageous in microorganism enrichment, nutrition balance, increase in stability, reduction in the accumulation of inhibitors, high efficiency of biogas production and methane yield [2, 41].

Table 7. Biogas Yield of Chicken Manure and Other Substrates

Feedstock	Biogas yield (mL/g VS)	Reference
CM + algae	332	[95]
CM + cardboard waste	319.62	[96]
CM + energy crop residue	-	[97]
FW, goat, & CM	80-109	[81]
Oil refinery wastewater + CM	194.02	[98]
Chicken processing waste, seagrass and Miscanthus	400	[76]
Durian shell, chicken, dairy and pig manures	224.8	[100]
CM + poppy straw	-	[101]
Corn stover + CM	218.8	[71]

Replacing the feedstock should be applied with caution, though, it does not necessarily have negative impacts on biogas production. Frankly, when the substrate type changes, the microbial communities need to adapt to the new environment and conditions [104]. Bioaugmentation is a type of strategy of adding specific microbial cultures to a biological system to improve the operational function by manipulating the microbial consortium [105]. This is based on the belief that slow degradation is caused by the absence or low populations of microorganisms responsible for the particular degradation step [16]. Bioaugmentation has been investigated in lab-scale digesters [105]; however, full-scale applications are still limited.

4. Biogas Production Across the Globe

Further growth in global market value of biogas is foreseen to rise before 2025 as countries poor in developing biogas plants are investing in the technology [89, 106]. According to [56], there are around 35 million installed biogas plant across the globe. The development in biogas production across continents can be visualized.

4.1. Europe and North America

Europe has an energy target they hope to achieve; hence, majority of World's biogas plants are in Europe [10, 62, 107]. Europe can boost of 20,000 plants [108]. The continent also lead in electricity generation from biogas [29]. Out of the 20,000 plants mentioned, (17,662 plants - 88%) is exploited for electricity generation [109]. The capacity of the plants for biogas in 2017 was 10.9 million tonnes, currently around 200 billion m³ annually and projected to reach 18-20 billion m³ by 2030 [9, 46, 48]. In areas of research, commercial utilization and industrial uses, number of plants rises from 3,700 agricultural biogas plants in 2007 to (7.2 thousand) in 2017 in Germany, and considered highest, producing 6.7-10.9 million tonnes of biogas in 2017 due to federal government support [9, 11, 46, 49, 95]. All this progress is a result of the shift from energy crops used to substrate [7, 51]. One of these substrates is CM [105]. Italy use wheat as feedstock for biogas production while Belgium has a bio-plant capable of handling 58,000 tons of waste yearly [40, 52]. In Sweden, CSTR type biomass plants are widely used to process FW to biogas and subsequently into biomethane [28, 96]. Main feedstock is sewage and landfills, producing the largest amount of biomethane for

use in buses and cars [9, 11, 97]. Twelve thousand vehicles were estimated to have been fueled in 2007 with upgraded biogas the world over [25]. France main feedstock are agricultural, wastewater, FW and industrial beverages processed in about 309 plants [20]. Enlargement of biogas plants to 5-10 times their original volume is the common trend now in Austria [107]. Centralized biogas plants and farm biogas plants are the two classes of biogas plants in Denmark [110]. Apart from single substrate plants, Denmark are the country most-utilizing multiple substrates to co-digest feedstock for diverse function [1, 49]. Waste paper and the use of wheat straw obtained at Newcastle University, had been studied for production of biogas in the United Kingdom (UK) [53, 111]. Switzerland is characterized for installing large-scale biogas plants, accounting for 82%, largely from manure and other agricultural feedstock [50, 89]. Poland's substrate for large-scale plants are slurry, pork, maize silage and distillery effluent [112]. Figure 2.3 depicts two biogas plants in Poland that uses sewage and agricultural materials as feedstock. Efficacy of steam explosion physical pretreatment method is widely accepted at the moment in Czech Republic [20]. Norway developed the world biggest liquid biogas plant [67]. Scotland explored waste paper collected from School of Computing and Engineering at the University of West of Scotland (UWS) as feedstock [60].

North America (countries like US, Canada, Mexico and Brazil) had thousands of agricultural biogas plants [10]. United States has 2,200 biogas systems processing 70 million tons of organic matter and 0.2 billion kWh-1030 GWh of electricity annually [11, 42, 100]. Biogas had been utilized as bus fleet fuel for over 180, 000 buses in Brazil, the largest percentage of it coming from cattle manure, producing 584 billion m³/y of biogas [11, 101]. The Brazilian Association of Biogas and Methane is playing significant role in that direction [113]. Federal University of Fronteira Sul in Brazil was investigated by [77] looking at the potential of corn stalk for biogas synthesis.

Population of animal in Asia (countries including, India, China, Indonesia, Malaysia, Iran, Singapore, Japan, etc) cannot be compared with any region [20]. As regards biogas, the technology all started from China as mentioned earlier [19]. They can boost of 100,000 modern biogas plants generating 50 billion m³ of biogas yearly, 3000 MW of electricity being the biggest in Asia [11, 62, 104].



Figure 3: Biogas plants in Poland: (a) agricultural biogas plant (b) biogas plant in a sewage treatment plant [117]

In 2015, China had produced 787.4 million tons of crop straw, mostly wheat straw, amounting to 130 million tons in 2016 [45, 107]. More than 35,000 biogas plants by Indian Government's support had been built so far, plus 100,000 used for cooking in Indian households [4]. Indonesia, is currently the 4th most populous country in the world widely building fixed dome digesters for Small and Medium Enterprises (SMEs) for farmers utilizing water hyacinth, animal waste and palm oil [6, 18, 22, 54]. Indonesia is the largest producer of palm oil in the world in which POME is the major feedstock followed by Malaysia [114–116]. Fortunately, apart from POME, Malaysians make use of low cost animal waste as alternative AD feed [20]. There are 1040 tons/day of MSW and 152 tons/day of green waste that could be run in a 550 tons/day AD plant capacity in Isfahan, Iran [47].

Major feed to biogas industries in Singapore are sewage sludge, FW, animal manure and horticulture wastes [92]. Japan is the only country using thermophilic approach in Asia [4]; other Asian nations like Korea, Cambodia, Vietnam, Nepal and Bangladesh's adoption for domestic installations is rising. Pigs, water buffalo and cattle, numbering 30 million are the sources of livestock manure with potential for biogas production in Vietnam [118]. Nepal being in front in installation rate, boost of approximately 330,000 households [11]. Biogas production potential in Bangladesh is around 17000 million m^3 used mainly for cooking purposes [11, 20]. South American nations are hardly mentioned in the literature with regards to biogas synthesis. A pocket of

publications was recorded by [55] on the utilization of corn stover, a lignocellulosic waste in Santiago, Chile.

Several biogas plants are being in operation in Africa [4, 20]. To supply farm houses with energy, Ducellier and Isman build simple biogas machines in Algeria between 1930 and 1940, an act that signals the beginning or introduction of biogas technology in Africa [103]. Two notable large scale plants projects that are operational in Africa is the GOPDC-Ghana and PRESCO PLC-Nigeria [119]. Biogas potential in Mauritania is 520×10^6 m^3 /year and $258.7 (\pm 125.8) \times 10^6$ m^3 /year while in Egypt these plants are mostly buried underground [20, 119]. Ethiopia employ fixed dome bioreactor in about 4500 household utilizing kitchen waste [24, 120]. Data on biogas application in Somalia is scanty because only about 1% of communities in the 43rd largest country in the world consume the gaseous fuel according to [121]. Main feedstock for Sudan's 200 installed biogas facility is animal and agricultural waste according to a 2005 figure [36]. Summarily, East and North Africa could boost of 3.2 million m^3 /h of biogas production. In 2017, of the 700 biogas plants in South Africa, only 300 was reported to be in operation [56]. In 2021, [37] reported a lower figure (200 digesters) of working bioreactors in South Africa (a 66.7% further reduction).

Nigeria's biogas potential in 1999 was 1382×10^6 m^3 /yr – before then (1995), the pioneer plants build by the Sokoto Energy Research Center (SERC) in Zaria has the capacity of 10 m^3 of biogas; an 18 m^3 plant constructed by the Federal Institute of Industrial Research Oshodi

(FIIRO) at Ojokoro Ifelodun Piggery Farm, Lagos in 1996; Mayflower School Ikene in Ogun State; and at Usman Danfodio University Sokoto, most of which are still at research stage or non-operational [122–125]. Undoubtedly, enormous amount of solid waste (32 million tonnes) is generated in Nigeria, but real effort had not been made by government to build a biogas plant to benefit from any of its products; atleast to meet the demands of millions living without electricity in the nation [45, 124, 126]. One notable government interest was the setting up of a biogas plant at Karu Abattoirs, Nasarawa state by the Federal Capital Territory (FCT) Administration in Nigeria [124]. However, several studies had been done on a bench-scale biodigester in Polytechnics and Universities in Nigeria for research purposes only [117, 126]. Biogas plants in some academic institutions in Nigeria are Usman Danfodio University Biogas Plant, Obafemi Awolowo University Biogas Plant and patent University of Ibadan Prototype Plant [45, 122, 127]. Other University related research and installation of the plant are reported elsewhere by [19, 36, 59, 60, 75, 128]. The trading, usage and construction of biodigester came much earlier in Kenya (1948) compared to Nigeria when Tim Hutchinson built the 1st biodigester that uses coffee pulp in the country. The country is also in the frontline in the gas manufacture in Africa overseeing the construction of 11,529 plants through the Africa Biogas Partnership Programme (ABPP) between 2009-2013 [43, 129]. In 2013, the ABPP built 70,000 biogas reactors in six member countries of Uganda, Senegal, Kenya, Burkina Faso, Tanzania and Ethiopia [37]. Tanzania developed their first plant in 1950s, making them third in Africa base on literature consulted after Algeria and Kenya [36].

5. Conclusion

Between 2000-2005, Nigeria is the country with the highest deforestation rate in the world and ranks 8 in methane emission, as they consume 46 million tonnes of wood and 3.2 million tonnes of charcoal for cooking as reported by [61, 129, 130]. Biogas plant development for use as cooking gas will go a long way in reducing the over-dependence on firewood, thereby checkmating desert encroachment in sub-Saharan Africa. Because biogas can be produced from hundreds of biowaste materials, most of them have not been studied (kinetic model-wise) in order to optimize their production. In the same context, there are too many of these biogas models (e.g. Chen & Hashimoto, Logistic, Bi-logistic, modified

Gompertz, First-Order, modified First-Order, Richards, Biogas Production Kinetic (BPK), Transference function model, Cone, Transfert, Proposed model by [131] and Fitzhugh) to capture the multitude of feedstock in nature plus the variety of reactor types available and the dynamic operating conditions involved. Regression remains the most used chemical engineering analysis tool to explain these models used for a selected single or multiple feedstock [132].

Africa produces and utilizes less biogas compared to Europe and Asia as only few plants had been built for AD of feedstock in recent years. It is therefore recommended that Africa, increase the harnessing of the abundant waste generated annually from the continent to meet some of her challenges, especially in areas of electricity generation for its accelerated development. Since energy consumption has skyrocketed, the speculation, that global energy demand will increase by 50% in 2050 might be credible [42, 133] and surplus will be achieved to counter shortages if more countries key into biogas plant development. This work further aims at increasing research on feedstock utilization for the production of biogas as well as the application of kinetics to facilitate biogas plant design and optimization using different feedstock, most abundant in different countries. Biogas digesters are presumed to last for 100 years. Despite this merit, adoption and implementation of biogas projects in developing countries are hindered majorly due to resistance to change, inadequate research, training and expertise in the technology, lack of investment incentives and trade, insufficient funding and lack of policy, strategy and regulations.

References

- [1] O. Khayal, "Main types and applications of biogas plants," *Nile Valley University*, pp. 1–11, 2019, doi:10.13140/RG.2.2.32559.69287.
- [2] K. Hagos et al., "Anaerobic co-digestion process for biogas production: Progress, challenges and perspectives," *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 1485–1496, 2017, doi:http://dx.doi.org/10.1016/j.rser.2016.11.184.
- [3] I. A. Raja, S. Wazir, "Biogas production: The fundamental processes," *Universal Journal of Engineering Science*, vol. 5, no. 2, pp. 29–37, 2017, doi:10.13189/ujes.2017.050202.
- [4] M. R. Atelge et al., "Biogas production from organic waste: Recent progress and perspectives," *Waste and Biomass Valorization*, vol. 11, pp. 1–22, 2018, doi:10.1007/s12649-018-00546-0.
- [5] J. Kainthola, A. S. Kalamdhad, V. V. Goud, "A review on enhanced biogas production from anaerobic digestion of lignocellulosic biomass by different enhancement techniques," *Process Biochemistry*, vol. 84, pp. 81–90, 2019, doi:10.1016/j.procbio.2019.05.023.

- [6] A. M. Uche et al., "Design and construction of fixed dome digester for biogas production using cow dung and water hyacinth," *African Journal of Environmental Science and Technology*, vol. 14, no. 1, pp. 15–25, 2020, doi:10.5897/AJEST2019.2739.
- [7] I. Koniuszewska et al., "Intensification of biogas production using various technologies: A review," *International Journal of Energy Research*, vol. 44, no. 8, pp. 6240–6258, 2020, doi:10.1002/er.5338.
- [8] N. Sawyerr et al., "An overview of biogas production: Fundamentals, applications and future research," *International Journal of Energy Economics and Policy*, vol. 9, no. 2, pp. 105–116, 2019, doi:10.32479/ijeep.7375.
- [9] R. L. Granado et al., "Technology overview of biogas production in anaerobic digestion plants: A European evaluation of research and development," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 44–53, 2017, doi:10.1016/j.rser.2017.05.079.
- [10] B. Bharathiraja et al., "Biogas production – A review on composition, fuel properties, feed stock and principles of anaerobic digestion," *Renewable and Sustainable Energy Reviews*, vol. 90, pp. 570–582, 2018, doi:10.1016/j.rser.2018.03.093.
- [11] T. Chowdhury et al., "Latest advancements on livestock waste management and biogas production: Bangladesh's perspective," *Journal of Cleaner Production*, vol. 272, no. 122818, pp. 1–20, 2020, doi:https://doi.org/10.1016/j.jclepro.2020.122818.
- [12] S. A. Neshat et al., "Anaerobic co-digestion of animal manures and lignocellulosic residues as a potent approach for sustainable biogas production," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 308–322, 2017, doi:10.1016/j.rser.2017.05.137.
- [13] M. Parsaei et al., "A review of biogas production from sugarcane vinasse," *Biomass and Bioenergy*, vol. 122, pp. 117–125, 2019, doi:10.1016/j.biombioe.2019.01.034.
- [14] A. Wu et al., "A spreadsheet calculator for estimating biogas production and economic measures for UK-based farm-fed anaerobic digesters," *Bioresource Technology*, vol. 220, pp. 479–489, 2016, doi:10.1016/j.biortech.2016.08.103.
- [15] N. Sahu et al., "Evaluation of biogas production potential of kitchen waste in the presence of spices," *Waste Management*, vol. 70, pp. 236–246, 2017, doi:10.1016/j.wasman.2017.08.045.
- [16] M. Westerholm, A. Schnürer, *Microbial responses to different operating practices for biogas production systems* (Uppsala, Sweden: InTech Open, 2019).
- [17] Z. Lenkiewicz, M. Webster, "How to convert organic waste into biogas: A step-by-step guide." wasteaid.org.uk/toolkit . (accessed: 14-Aug-2021).
- [18] M. Khalil et al., "Waste to energy technology: The potential of sustainable biogas production from animal waste in Indonesia," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 323–331, 2019, doi:10.1016/j.rser.2019.02.011.
- [19] Y. Lahlou, *Design of a biogas pilot unit for Al Akhawayn University* (School of Science and Engineering, 2017).
- [20] M. M. Ali et al., "Mapping of biogas production potential from livestock manures and slaughterhouse waste: A case study for African countries," *Journal of Cleaner Production*, vol. 256, no. 120499, pp. 1–18, 2020, doi:https://doi.org/10.1016/j.jclepro.2020.120499.
- [21] Martina Pilloni and Tareq Abu Hamed, "Small-size biogas technology applications for rural areas in the context of developing countries," in *Anaerobic digestion in built environments*, ed Anna Sikora (IntechOpen, 2021), 24, doi:http://dx.doi.org/10.5772/intechopen.96857.
- [22] M. A. Fahriansyah, Sriharti, "Design of conventional mixer for biogas digester," *IOP Conference Series: Earth and Environmental Science*, pp. 1–8, 2019, doi:10.1088/1755-1315/277/1/012017.
- [23] IRENA, "Measuring small-scale biogas capacity and production." Abu Dhabi, United Arab Emirates, 2016.
- [24] E. W. Gabisa, S. H. Gheewala, "Potential, environmental, and socio-economic assessment of biogas production in Ethiopia: The case of Amhara regional state," *Biomass and Bioenergy*, vol. 122, pp. 446–456, 2019, doi:10.1016/j.biombioe.2019.02.003.
- [25] A. F. Akintade, "Biogas a viable source of energy: Case study, Nigeria," (Centria University of Applied Sciences, 2021).
- [26] O. Raymond, U. Okezie, "The significance of biogas plants in Nigeria's energy strategy," *Journal of Physical Sciences and Innovation*, vol. 3, pp. 11–17, 2011.
- [27] R. J. Patinvoth et al., "Innovative pretreatment strategies for biogas production," *Bioresource Technology*, vol. 224, pp. 13–24, 2016, doi:10.1016/j.biortech.2016.11.083.
- [28] S. M. A. Abuabdou et al., "A review of anaerobic membrane bioreactors (AnMBR) for the treatment of highly contaminated land fill leachate and biogas production: Effectiveness, limitations and future perspectives," *Journal of Cleaner Production*, vol. 255, no. 120215, pp. 1–12, 2020, doi:10.1016/j.jclepro.2020.120215.
- [29] M. C. Caruso et al., "Recent updates on the use of agro-food waste for biogas production," *Applied Sciences*, vol. 9, no. 6, pp. 1–29, 2019, doi:10.3390/app9061217.
- [30] M. Westerholm, T. Liu, A. Schnürer, "Comparative study of industrial-scale high-solid biogas production from food waste: Process operation and microbiology," *Bioresource Technology*, vol. 304, pp. 122–981, 2020, doi:10.1016/j.biortech.2020.122981.
- [31] S. Mirmohamadsadeghi et al., "Biogas production from food wastes: A review on recent developments and future perspectives," *Bioresource Technology Reports*, vol. 7, no. 100202, pp. 1–37, 2019, doi:10.1016/j.biteb.2019.100202.
- [32] M. A. Aziz et al., "Recent advances on palm oil mill effluent (POME) pretreatment and anaerobic reactor for sustainable biogas production," *Renewable and Sustainable Energy Reviews*, vol. 119, no. 109603, pp. 1–31, 2019, doi:https://doi.org/10.1016/j.rser.2019.109603.
- [33] E. I. Dhimain, S. C. Izah, "Potential of biogas production from palm oil mills effluent in Nigeria," *Sky Journal of Soil Science and Environmental Management (SJSSEM)*, vol. 3, no. 5, pp. 50–58, 2014.
- [34] M. Bakraoui et al., "Biogas production from recycled paper mill wastewater by UASB digester: Optimal and mesophilic conditions," *Biotechnology Reports*, vol. 25, pp. 1–8, 2020, doi:10.1016/j.btre.2019.e00402.
- [35] V. J. Brown, "Biogas: A bright idea for Africa," *Environmental Health Perspectives*, vol. 114, no. 5, pp. 300–303, 2006.
- [36] A. M. Mshandete, W. Parawira, "Biogas technology research in selected sub-Saharan African countries-A review," *African Journal of Biotechnology*, vol. 8, no. 2, pp. 116–125, 2009.
- [37] R. F. T. Tagne et al., "Technologies, challenges and perspectives of biogas production within an agricultural context: The case of China and Africa," *Environment, Development and Sustainability*, vol. 23, pp. 14799–14826, 2021, doi:https://doi.org/10.1007/s10668-021-01272-9.
- [38] A. I. Aigbodion et al., "Viability of biogas production from manure/biomass in Nigeria using fixed dome digester," *Universal Journal of Agricultural Research (UJAR)*, vol. 6, no. 1, pp. 1–8, 2018, doi:10.13189/ujar.2018/060101.
- [39] E. M. M. Esteves et al., "Life cycle assessment of manure biogas production: A review," *Journal of Cleaner Production*, vol. 219, pp.

- 411–423, 2019, doi:10.1016/j.jclepro.2019.02.091.
- [40] E. Małgorzata, J. Frankowski, "The biogas production potential from silkworm waste," *Waste Management*, vol. 79, pp. 564–570, 2018, doi:10.1016/j.wasman.2018.08.019.
- [41] K. Chaump et al., "Leaching and anaerobic digestion of poultry litter for biogas production and nutrient transformation," *Elsevier*, vol. 84, pp. 413–422, 2018, doi:10.1016/j.wasman.2018.11.024.
- [42] H. Wang et al., "Effect of addition of biogas slurry for anaerobic fermentation of deer manure on biogas production," *Energy*, vol. 165, pp. 411–418, 2018, doi:10.1016/j.energy.2018.09.196.
- [43] C. Mulinda, Q. Hu, K. Pan, "Dissemination and problems of African biogas technology," *Energy and Power Engineering*, vol. 5, pp. 506–512, 2013, doi:http://dx.doi.org/10.4236/epe.2013.58055.
- [44] M. N. Usman, M. A. Suleiman, M. I. Binni, *Anaerobic digestion of agricultural wastes: A potential remedy for energy shortfalls in Nigeria*, vol. 4, no. 1, (Scholarena, 2021).
- [45] M. B. Biodun, O. S. I. Fayomi, J. O. Okeniyi, "The possibility of biogas production in Nigeria from organic waste material: A review," *International Conference on Engineering for Sustainable World (ICESW 2020): Material Science and Engineering*, vol. 1107, no. 012166, pp. 1–9, 2020, doi:10.1088/1757-899X/1107/1/012166.
- [46] T. M. Simeon, "Techno-economic analysis of a model biogas plant for agricultural applications: A case study of the Concordia Farms Limited, Nonwa, Tai, Rivers State," (University of Nigeria, Nsukka, 2009).
- [47] S. M. M. N. Dehkordi et al., "Investigation of biogas production potential from mechanical separated municipal solid waste as an approach for developing countries (case study: Isfahan-Iran)," *Renewable and Sustainable Energy Reviews*, vol. 119, no. 109586, pp. 1–12, 2020, doi:10.1016/j.rser.2019.109586.
- [48] J. Liebetrau et al., *Potential and utilization of manure to generate biogas in seven countries* (IEA Bioenergy Task 37, 2021).
- [49] T. M. Thompson, B. R. Young, S. Baroutian, "Advances in the pretreatment of brown macroalgae for biogas production," *Fuel Processing Technology*, vol. 195, no. 106151, pp. 1–12, 2019, doi:10.1016/j.fuproc.2019.106151.
- [50] S. Achinas, G. J. W. Euverink, "Elevated biogas production from the anaerobic co-digestion of farmhouse waste: Insight into the process performance and kinetics," *Waste Management & Research*, vol. 37, no. 12, pp. 1240–1249, 2019, doi:10.1177/0734242X19873383.
- [51] C. P. C. Bong et al., "The characterisation and treatment of food waste for improvement of biogas production during anaerobic digestion– A review," *Journal of Cleaner Production*, vol. 172, pp. 1545–1558, 2017, doi:10.1016/j.jclepro.2017.10.199.
- [52] S. Tanigawa, "Fact Sheet | Biogas: Converting Waste to Energy Tags / Keywords." www-eesi-org-papers-view-fact-sheet-biogasconverting-waste-to-energy . (accessed: 14-Aug-2021).
- [53] N. I. H. Abdul Aziz, M. M. Hanafiah, M. Y. M. Ali, "Sustainable biogas production from agrowaste and effluents – A promising step for small-scale industry income," *Renewable Energy*, vol. 132, pp. 363–369, 2018, doi:10.1016/j.renene.2018.07.149.
- [54] Q. Yu et al., "A review of crop straw pretreatment methods for biogas production by anaerobic digestion in China," *Renewable and Sustainable Energy Reviews*, vol. 107, pp. 51–58, 2019, doi:10.1016/j.rser.2019.02.020.
- [55] V. Wyman et al., "Lignocellulosic waste valorisation strategy through enzyme and biogas production," *Bioresource Technology*, vol. 247, pp. 402–411, 2018, doi:10.1016/j.biortech.2017.09.055.
- [56] F. Kemausuor, M. S. Adaramola, J. Morken, "A review of commercial biogas systems and lessons for Africa," *Energies*, vol. 11, no. 2984, pp. 1–21, 2018, doi:10.3390/en11112984.
- [57] S. K. Srivastava, "Advancement in biogas production from the solid waste by optimizing the anaerobic digestion," *Waste Disposal & Sustainable Energy*, vol. 2, no. 2, pp. 85–103, 2020, doi:10.1007/s42768-020-00036-x.
- [58] E. P. Akhator, D. I. Igbinomwanhia, A. I. Obonor, "Potentials for commercial production of biogas from domestic food waste generated in Benin Metropolis, Nigeria," *Journal of Applied Sciences and Environmental Management (JASEM)*, vol. 20, no. 2, pp. 369–373, 2016, doi:http://dx.doi.org/10.4314/jasem.v20i2.19.
- [59] Z. Kong et al., "Large pilot-scale submerged anaerobic membrane bioreactor for the treatment of municipal wastewater and biogas production at 25°C," *Bioresource Technology*, vol. 319, pp. 1–12, 2021, doi:10.1016/j.biortech.2020.124123.
- [60] C. Rodriguez et al., "Mechanical pretreatment of waste paper for biogas production," *Waste Management*, vol. 68, pp. 157–164, 2017, doi:10.1016/j.wasman.2017.06.040.
- [61] F. O. Olanrewaju et al., "Bioenergy potential in Nigeria," *Chemical Engineering Transactions*, vol. 74, pp. 61–66, 2019.
- [62] G. Mancini et al., "Increased biogas production from wheat straw by chemical pretreatments," *Renewable Energy*, vol. 119, pp. 608–614, 2018, doi:10.1016/j.renene.2017.12.045.
- [63] A. M. A. Mohammed, F. M. A. Kabbashi, H. K. A. Hamad, "Production of biogas from biomedical waste (blood)," (Sudan University of Science and Technology, 2017).
- [64] D. Elalami et al., "Pretreatment and co-digestion of wastewater sludge for biogas production: Recent research advances and trends," *Renewable and Sustainable Energy Reviews*, vol. 114, no. 109287, pp. 1–23, 2019, doi:10.1016/j.rser.2019.109287.
- [65] O. Tomori, "Feasibility study of a large scale biogas plant in Lagos, Nigeria," (Murdoch University of Western Australia, 2012).
- [66] A. Nsair et al., "Operational parameters of biogas plants: A review and evaluation study," *Energies*, vol. 13, no. 15, pp. 1–27, 2020, doi:10.3390/en13153761.
- [67] S. Sarker et al., "A review of the role of critical parameters in the design and operation of biogas production plants," *Applied Sciences*, vol. 9, no. 9, pp. 1–38, 2019, doi:10.3390/app9091915.
- [68] S. K. Pramanik et al., "The anaerobic digestion process of biogas production from food waste: Prospects and constraints," *Bioresource Technology Reports*, vol. 8, pp. 1–38, 2019, doi:10.1016/j.biteb.2019.100310.
- [69] S. Sarto, R. Hildayati, I. Syaichurrozi, "Effect of chemical pretreatment using sulfuric acid on biogas production from water hyacinth and kinetics," *Renewable Energy*, vol. 132, pp. 335–350, 2019, doi:10.1016/j.renene.2018.07.121.
- [70] V. A. Mazur et al., "Agroecological prospects of using corn hybrids for biogas production," *Agronomy Research*, vol. 18, no. 1, pp. 177–182, 2020, doi:https://doi.org/10.15159/ar.20.016.
- [71] H. Gebretsadik, S. Mulaw, G. Gebregziabher, "Qualitative and quantitative feasibility of biogas production from kitchen waste," *American Journal of Energy Engineering*, vol. 6, no. 1, pp. 1–5, 2018, doi:10.11648/j.ajee.20180601.11.
- [72] H. T. T. Nong et al., "Development of sustainable approaches for converting the agro-weeds *Ludwigia hyssopifolia* to biogas production," *Biomass Conversion and Biorefinery*, pp. 1–9, 2020, doi:https://doi.org/10.1007/s13399-020-01083-4.
- [73] S. Wacławek et al., "Disintegration of wastewater activated sludge (WAS) for improved biogas production," *Energies*, vol. 12,

- no. 21, pp. 1–15, 2019, doi:10.3390/en12010021.
- [74] W. Cao et al., "Hydrogen production from supercritical water gasification of chicken manure," *International Journal of Hydrogen Energy*, pp. 1–10, 2016, doi:10.1016/j.ijhydene.2016.09.031.
- [75] A. Abraham et al., "Pretreatment strategies for enhanced biogas production from lignocellulosic biomass," *Bioresource Technology*, vol. 301, no. 122725, pp. 1–13, 2020, doi:10.1016/j.biortech.2019.122725.
- [76] S. Ali et al., "Evaluating the co-digestion effects on chicken manure and rotten potatoes in batch experiments," *International Journal of Biosciences (IJB)*, vol. 10, no. 6, pp. 150–159, 2017, doi:10.12692/ijb/10.6.150-159.
- [77] B. Venturin et al., "Effect of pretreatments on corn stalk chemical properties for biogas production purposes," *Bioresource Technology*, vol. 266, pp. 1–36, 2018, doi:10.1016/j.biortech.2018.06.069.
- [78] S. Mirmohamadsadeghi et al., "Pretreatment of lignocelluloses for enhanced biogas production: A review on influencing mechanisms and the importance of microbial diversity," *Renewable and Sustainable Energy Reviews*, vol. 135, no. 110173, pp. 1–19, 2021, doi:10.1016/j.rser.2020.110173.
- [79] M. Tabatabaei et al., "A comprehensive review on recent biological innovations to improve biogas production, Part 1: Upstream strategies," *Renewable Energy*, vol. 146, pp. 1204–1220, 2019, doi:https://doi.org/10.1016/j.renene.2019.07.037.
- [80] A. A. Rajput, C. Visvanathan, "Effect of thermal pretreatment on chemical composition, physical structure and biogas production kinetics of wheat straw," *Journal of Environmental Management*, vol. 221, pp. 45–52, 2018, doi:10.1016/j.jenvman.2018.05.011.
- [81] S. P. Lohani et al., "Anaerobic co-digestion of food waste, goat and chicken manure for sustainable biogas production," *International Journal of Energy Applications and Technologies*, vol. 7, no. 4, pp. 120–125, 2020, doi:10.31593/ijeat.748982.
- [82] K. Dalk, A. Ugurlu, "Biogas production from chicken manure at different organic loading rates in a mesophilic-thermophilic two stage anaerobic system," *Journal of Bioscience and Bioengineering*, vol. xx, no. xx, pp. 1–8, 2015, doi:10.1016/j.jbiosc.2015.01.021.
- [83] N. Duan et al., "Performance evaluation of mesophilic anaerobic digestion of chicken manure with algal digestate," *Energies*, vol. 11, no. 1829, pp. 1–11, 2018, doi:10.3390/en11071829.
- [84] M. A. K. B. D. T. T. Onay, "Enhanced biogas production from chicken manure via enzymatic pretreatment," *Journal of Material Cycles and Waste Management*, no. 0123456789, 2020, doi:10.1007/s10163-020-01039-w.
- [85] T. Keskin et al., "The determination of the trace element effects on basal medium by using the statistical optimization approach for biogas production from chicken manure," *Waste and Biomass Valorization*, vol. 0, no. 0, pp. 1–10, 2018, doi:10.1007/s12649-018-0273-2.
- [86] W. Fuchs et al., "Tackling ammonia inhibition for efficient biogas production from chicken manure: Status and technical trends in Europe and China," *Renewable and Sustainable Energy Reviews*, vol. 97, pp. 186–199, 2020, doi:10.1016/j.rser.2018.08.038.
- [87] N. A. Noori, Z. Z. Ismail, "Process optimization of biogas recovery from giant reed (*Arundo donax*) alternatively pretreated with acid and oxidant agent: Experimental and kinetic study," *Biomass Conversion and Biorefinery*, pp. 1–15, 2019, doi:https://doi.org/10.1007/s13399-019-00481-7.
- [88] Y. Li et al., "Biogas production from co-digestion of corn stover and chicken manure under anaerobic wet, hemi-solid, and solid state conditions," *Bioresource Technology*, vol. 149, pp. 406–412, 2013, doi:http://dx.doi.org/10.1016/j.biortech.2013.09.091.
- [89] M. Hassan et al., "Methane enhancement through co-digestion of chicken manure and oxidative cleaved wheat straw: Stability performance and kinetic modeling perspectives," *Energy*, vol. 141, pp. 2314–2320, 2017, doi:https://doi.org/10.1016/j.energy.2017.11.110.
- [90] Y. K. Li, X. M. Hu, L. Feng, "Characteristics of biogas production via high-temperature dry fermentation of chicken manure," *Applied Ecology and Environmental Research*, vol. 18, no. 4, pp. 4883–4895, 2020, doi:http://dx.doi.org/10.15666/aer/1804_48834895.
- [91] D. Cheong, J. Kim, C. Lee, "Improving biomethanation of chicken manure by co-digestion with ethanol plant effluent," *International Journal of Environmental Research and Public Health*, vol. 16, no. 5023, pp. 1–10, 2019, doi:10.3390/ijerph16245023.
- [92] L. Zhang, K. Loh, J. Zhang, "Enhanced biogas production from anaerobic digestion of solid organic wastes: Current status and prospects," *Bioresource Technology Reports*, vol. 5, pp. 280–296, 2019, doi:10.1016/j.biteb.2018.07.005.
- [93] B. Shamurad et al., "Stable biogas production from single-stage anaerobic digestion of food waste," *Applied Energy*, vol. 263, no. 114609, pp. 1–37, 2019, doi:http://dx.doi.org/10.17632/6wnfvkz6gb.1.
- [94] M. Tabatabaei et al., "A comprehensive review on recent biological innovations to improve biogas production, Part 2: Mainstream and downstream strategies," *Renewable Energy*, vol. 146, pp. 1392–1407, 2019, doi:https://doi.org/10.1016/j.renene.2019.07.047.
- [95] A. Abdelhay, L. Al-Hasanat, A. Albsoul, "Anaerobic co-digestion of cattle manure and raw algae: Kinetic study and optimization of methane potential by RSM," *Pol. Journal of Environmental Studies*, vol. 30, no. 2, pp. 1029–1037, 2021, doi:10.15244/pjoes/125523.
- [96] S. Zhao et al., "Anaerobic co-digestion of chicken manure and cardboard waste: Focusing on methane production, microbial community analysis and energy evaluation," *Bioresource Technology*, vol. 321, no. 12449, pp. 1–11, 2021, doi:https://doi.org/10.1016/j.biortech.2020.124429.
- [97] N. A. D. Cahyono et al., "Anaerobic co-digestion of chicken manure with energy crop residues for biogas production," *IOP Conference Series: Earth and Environmental Science*, pp. 765, 2021, doi:10.1088/1755-1315/765/1/012044.
- [98] E. Mehryar et al., "Anaerobic co-digestion of oil refinery wastewater and chicken manure to produce biogas, and kinetic parameters determination in batch reactors," *Agronomy Research*, vol. 15, no. 5, pp. 1983–1996, 2017, doi:https://doi.org/10.15159/AR.17.072.
- [99] C. Li et al., "Assessment of regional biomass as co-substrate in the anaerobic digestion of chicken manure: Impact of co-digestion with chicken processing waste, seagrass and *Miscanthus*," *Biochemical Engineering Journal*, pp. 1–38, 2016, doi:http://dx.doi.org/doi:10.1016/j.bej.2016.11.008.
- [100] J. Shena et al., "Biogas production from anaerobic co-digestion of durian shell with chicken, dairy, and pig manures," *Energy Conversion and Management*, pp. 1–10, 2018, doi:https://doi.org/10.1016/j.enconman.2018.06.099.
- [101] A. Bayrakdar et al., "Biogas production from chicken manure: Co-digestion with spent poppy straw," *International Biodeterioration & Biodegradation*, pp. 1–6, 2016, doi:http://dx.doi.org/10.1016/j.ibiod.2016.10.058.

- [102] A. M. Abubakar, M. U. Yunus, "Reporting biogas data from various feedstock," *International Journal of Formal Sciences: Current and Future Research Trends (IJFSCFRT)*, vol. 11, no. 1, pp. 23–36, 2021, doi:10.5281/zenodo.6366775.
- [103] C. C. Ngumah et al., "Biogas potential of organic waste in Nigeria," *Journal of Urban and Environmental Engineering (JUEE)*, vol. 7, no. 1, pp. 110–116, 2013, doi:10.4090/juee.2013.v7n1.110116.
- [104] S. Theuerl, J. Klang, A. Prochnow, "Process disturbances in agricultural biogas production – Causes, mechanisms and effects on the biogas microbiome: A review," *Energies*, vol. 12, no. 3, pp. 1–20, 2019, doi:10.3390/en12030365.
- [105] W. Fuchs et al., "Tackling ammonia inhibition for efficient biogas production from chicken manure: Status and technical trends in Europe and China," *Renewable and Sustainable Energy Reviews*, vol. 97, pp. 186–199, 2018, doi:10.1016/j.rser.2018.08.038.
- [106] P. Baltrėnas, V. Kolodinskij, D. Urbanas, "Biogas production from chicken manure at different organic loadings using a special zeolite additive (ZeoVit sorbent)," *Journal of Renewable Sustainable Energy*, vol. 11, no. 063101, pp. 1–11, 2019, doi:10.1063/1.5119840.
- [107] B. Stürmer et al., "Agricultural biogas production: A regional comparison of technical parameters," *Renewable Energy*, vol. 164, pp. 171–182, 2021, doi:10.1016/j.renene.2020.09.074.
- [108] J. Maroušek et al., "Advances in the agrochemical utilization of fermentation residues reduce the cost of purpose-grown phytomass for biogas production," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–11, 2020, doi:10.1080/15567036.2020.1738597.
- [109] B. K. McCabe, T. Schmidt, *Integrated Biogas Systems: Local Applications of Anaerobic Digestion Towards Integrated Sustainable Solutions* (Queensland, Australia: IEA Bioenergy, 2018).
- [110] HomeBioGas, "What is Biogas? A Beginner's Guide." www.homebiogas.com/what-is-biogas-a-beginners-guide- (accessed: 13-Aug-2021).
- [111] G. Cayci, C. Temiz, S. S. Ok, "The effects of fresh and composted chicken manures on some soil characteristics," *Communications in Soil Science and Plant Analysis*, vol. 00, no. 00, pp. 1–11, 2017, doi:10.1080/00103624.2017.1373794.
- [112] K. A. Korys et al., "The review of biomass potential for agricultural biogas production in Poland," *Sustainability*, vol. 11, no. 6515, pp. 1–13, 2019, doi:10.3390/su11226515.
- [113] I. F. S. dos Santos et al., "Assessment of potential biogas production from multiple organic wastes in Brazil: Impact on energy generation, use, and emissions abatement," *Resources, Conservation & Recycling*, vol. 131, pp. 54–63, 2018, doi:10.1016/j.resconrec.2017.12.012.
- [114] Y. Y. Choong, K. W. Chou, I. Norli, "Strategies for improving biogas production of palm oil mill effluent (POME) anaerobic digestion: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2993–3006, 2018, doi:10.1016/j.rser.2017.10.036.
- [115] N. Izzah et al., "A review on life cycle assessment of biogas production: Challenges and future perspectives in Malaysia," *Biomass and Bioenergy*, vol. 122, pp. 361–374, 2019, doi:10.1016/j.biombioe.2019.01.047.
- [116] N. I. H. A. Aziz, M. M. Hanafiah, "Life cycle analysis of biogas production from anaerobic digestion of palm oil mill effluent," *Renewable Energy*, vol. 145, pp. 847–857, 2020, doi:10.1016/j.renene.2019.06.084.
- [117] J. Piekutin et al., "The efficiency of the biogas plant operation depending on the substrate used," *Energies*, vol. 14, no. 3157, pp. 1–12, 2021, doi:https://doi.org/10.3390/en14113157.
- [118] I. Yerima et al., "The influence of proximate composition of cow dung on the rate and volume of biogas generation in Maiduguri, North Eastern Nigeria," *International Journal of Environment, Agriculture and Biotechnology (IJEAB)*, vol. 4, no. 1, pp. 146–153, 2019, doi:http://dx.doi.org/10.22161/ijeab/4.1.24.
- [119] A. Bulnes, "Anaerobic digestion and biogas plants in Africa: Integrated organic matter management for a sustainable agroindustrial sector." 2017.
- [120] U. Brémond et al., "Biological pretreatments of biomass for improving biogas production: An overview from lab scale to full-scale," *Renewable and Sustainable Energy Reviews*, vol. 90, pp. 583–604, 2018, doi:10.1016/j.rser.2018.03.103.
- [121] A. M. Wanjohi, E. K. Irungu, "An overview of consumption of biogas in the context of Somalia," *International Journal of Environmental and Health Sciences (JEHS)*, vol. 2, pp. 1–4, 2021.
- [122] J. O. Egbere et al., "Generation of biogas from segregates of municipal solid wastes in Jos, Nigeria," *Global Journal of Pure and Applied Sciences*, vol. 17, no. 1, pp. 41–45, 2011.
- [123] K. A. Adeniran et al., "Relative effectiveness of biogas production using poultry wastes and cow dung," *Agricultural Engineering International: CIGR*, vol. 16, no. 1, pp. 126–132, 2014.
- [124] A. J. Eferi, A. P. Aderemi, "Potential, barriers and prospects of biogas production in North-Central Nigeria," *3rd International Engineering Conference (IEC 2019)*, pp. 1–6, 2019.
- [125] I. A. Rufai, "A review of the evolution and development of anaerobic digestion technology," *Journal of Engineering and Technology (JET)*, vol. 5, no. 1, pp. 100–111, 2010.
- [126] A. Z. Abdul, A. M. Abubakar, "Potential swing to natural gas-powered electricity generation," *International Journal of Natural Sciences: Current and Future Research Trends (IJNSCFRT)*, vol. 10, no. 1, pp. 27–36, 2021.
- [127] L. Ioannou-tfofa et al., "Life cycle assessment of household biogas production in Egypt: Influence of digester volume, biogas leakages, and digestate valorization as biofertilizer," *Journal of Cleaner Production*, vol. 286, no. 125468, pp. 1–14, 2021, doi:10.1016/j.jclepro.2020.125468.
- [128] M. Tanczuk et al., "Assessment of the energy potential of chicken manure in Poland," *Energies*, vol. 12, no. 1244, pp. 1–18, 2019, doi:10.3390/en12071244.
- [129] A. M. Wanjohi, E. K. Irungu, H. C. Gicheru, "Biogas program in Kenya: History, Challenges and Milestones," *International Journal of Environmental and Health Sciences (JEHS)*, vol. 2, pp. 1–3, 2022.
- [130] P. K. Nimame, I. P. Nimame, R. A. Ekemube, "Importance of biogas utilization as alternative source in an energy deficit economy," *International Journal of Academic Information Systems Research (IJAIRS)*, vol. 4, no. 12, pp. 17–21, 2020.
- [131] R. Venkateshkumar, S. Shanmugam, A. R. Veerappan, "Anaerobic co-digestion of cow dung and cotton seed hull with different blend ratio: Experimental and kinetic study," *Biomass Conversion and Biorefinery*, pp. 1–111, 2020, doi:https://doi.org/10.1007/s13399-020-01006-3.
- [132] A. M. Abubakar, B. Iliyasu, Z. M. Sarkinbaka, "Detailed overview on POLYMATH software for chemical engineering analysis," *Journal of Engineering Research and Sciences (JENRS)*, vol. 1, no. 3, pp. 133–147, 2022, doi:https://doi.org/TBA.
- [133] T. R. T. Yusof et al., "Evaluation of hydrogen and methane production from co-digestion of chicken manure and food waste," *Polish Journal of Environmental Studies*, vol. 28, no. 4, pp. 1–11, 2019, doi:10.15244/pjoes/86222.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



ABDULHALIM MUSA ABUBAKAR has a degree in Chemical Engineering from University of Maiduguri-Nigeria (2018). He is a Masters student of the same institution (enrolled in November, 2019).

He works as teaching/research assistant at Modibbo Adama University, Yola-Nigeria. His publications are titled: "Neural Network Based Performance Evaluation of a Waterflooded Oil Reservoir", "Newton's Method Cubic Equation of State C++ Source Code for Iterative Volume Computation", "Simplex C++ Syntax for Solving Chemical Engineering Cost Optimization Problems", "Potential Swing to Natural Gas-Powered Electricity Generation", "History, Adverse Effect and Clean Up Strategies of Oil Spillage", "Reporting Biogas Data from Various Feedstock", and "Detailed Overview on POLYMATH Software for Chemical Engineering Analysis". His Masters Thesis dwells on poultry waste management through biogas generation and kinetic study.

Received: 13 January 2022, Revised: 09 March 2022, Accepted: 17 March 2022, Online: 28 March 2022

DOI: <https://dx.doi.org/10.55708/js0103020>

Length-Weight Relationships (LWRs) and Condition Factor of Seven Fish Species in River Nyangweta Tributary, Kenya

Fredrick Mang'era Ondemo ^{*1,2}, Albert Getabu ¹, Zipporah Gichana ¹, Job Ombiro Omweno ¹

¹ Department of Environment, Natural resources and Aquatic Sciences, Kisii University, P.O Box 408-40200, Kisii, Kenya.

² Department of Agriculture, Fisheries, Livestock & Cooperatives Development, Kisii County Government, P.O Box 700 – 40200 Kisii, Kenya

*Corresponding Author: Fredrick Mang'era Ondemo, Telephone: +254 723 836 762/+254 733 528 144, amangesh69@yahoo.com

Corresponding Author ORCID: 0000-0001-8581-510X

ABSTRACT: Length-weight relationships (LWRs) and condition factors are important for effective management of riverine fisheries. This study investigated the LWRs and condition factors of the dominant fish species from Nyangweta tributary of River Kuja, in Lake Victoria Catchment, Kenya. A total of 615 fish of seven dominant species were analyzed: *Enteromius altianalis*, *Enteromius neumeyeri*, *Clarias theodora*, *Labeo victorianus*, *Labeobarbus altianialis*, *Chiloglanis* species and *Amphilius jacksonii*. The fish were sampled from five locations from October 2020 to March 2021. The length and weight measurements were taken using a measuring board and an electronic balance respectively and used to determine Fulton's (KF) and allometric (Ka) condition factors. The largest species was *E. altianalis* ($70.74 \pm 8.72\text{g}$), followed by *L. altianialis* and *L. victorianus* with mean body weights of $62.14 \pm 3.48\text{g}$ and $56.43 \pm 11.26\text{g}$ respectively. With exception of *E. altianalis* and *C. theodora*, the Fulton's condition factors (KF) of all the species were greater than (1) and not significantly different ($p = 0.43$) among the species. Majority (four) of the species exhibited positive allometry ($Ka > 3$), while only one species exhibited isometric ($Ka = 3$) growth, due to different body profiles which may not allow uniform growth of all individuals of the same species, a large percentage of females in adult fish populations and high nutrient influxes. The LWRs were significant and showed strong linear relationships between total and standard lengths and body weight ($p < 0.05$, $R^2 > 0.5$). These findings suggest that conservation measures need to be taken to improve the fish health and physiological condition in upstream areas of the Nyangweta tributary.

KEYWORDS: Riverine fish, Length-weight relationship, Fulton's condition factor, Allometric condition.

1. Introduction

Fishery researchers use Length-Weight relationships to determine the physiological condition and growth patterns of fish in rivers and other aquatic ecosystems [1]. [2], report that LWR can be used in fisheries biology to estimate body weight from a given length class of fish. This relationship also provides an estimate of fishery potential and variability of individual fish populations by indicating the fishery exploitation rates determined by fishing effort [3], [4]. It is assumed that heavier fish relative to their body length reflect a healthier physiological state

than lighter ones because of high energy investments for growth and reproduction which increases the fishery potential to withstand high fishing pressure [5]. However, only few organisms have been reported to grow isometrically because the aquatic ecosystem is subject to a large number of changing environmental and physiological factors [6]. Rivers make crucial contributions towards lacustrine fish populations. Nevertheless, the success of the fish populations depends on various factors such as habitat quality, heterogeneity and water quality that affect the well-being of riverine fishes [7].

Despite a high global interest in riverine fisheries, in Kenya they are largely regarded as commercially inexpressive. However, some potamodromous species such as *Labeo victorianus*, which form the mainstay of Lake Victoria's river-mouth fisheries [8]. Besides, rivers provide refugia for survival and breeding of both migratory and endemic populations [9]. It is therefore necessary to evaluate the length-weight relationships (LWRs) and Condition factors of riverine species to provide baseline information for the management of riverine fisheries in rivers found in Lake Victoria Basin.[10–12]. Hence, the main aim of the present study was to investigate the LWRs and condition factors of fish species in the Nyangweta tributary of River Kuja located in the Kenyan sector of Lake Victoria catchment.

2. Materials and Methods

The study was carried out in River Nyangweta a tributary of River Kuja which is located in the eastern part of Lake Victoria Basin and lies between latitude 08°24'N and longitudes 39°01'E (Figure 1).

2.1. The fish condition factor

The monthly length and weight data of all the species in the catch data was used to estimate length- weight relationship of the dominant riverine species. Length-weight relationship was conducted on seven dominant species out of all 13 species sampled with total length and body weight range shown in Table 1. computed using the least squares regression analysis using the Le

Cren (1951) equation, $W = a \cdot L^b$, where W = body weight of fish, L represents total length of fish and 'a' and 'b' are the exponents. To solve for exponents, the Le Cren equation was log transformed using natural logarithms in $\log W = \log a + b \log L$, as recommended by [13]. In this linearized relationship, the value 'log a' is the regression intercept which estimates the initial growth and indicates the unit body weight for which the length is zero, while the slope of the regression 'b' describes the rate of change in body weight at unit length. The degree of association between the body weight and total length variables was determined by computing the correlation co-efficient (r) and the adjusted coefficient of determination (R^2), which exhibit the relationship: $r = \sqrt{R^2}$ [14]. To arrive at these coefficients, all sampled datasets were subjected to ordinary least squares statistical procedures of fitting the length-weight relationship using the [13] and [15]. The length-weight relationships were separately computed and the linear regression lines were fitted into the same plot using R software graphical techniques [16], [17]. The Fulton's condition factor (K) was determined as the ratio of body weight to the cube of total length according to [18] equation: Where L represents the total fish length

measured in centimeters and W , represents the body weight measured in grams. Bailey's t - test = $(b-3) / S_b$, where b = regression coefficient and S_b = Standard error of 'b' [19] was performed to determine whether the slope coefficient 'b' value significantly deviated from the expected cube value of 3. Statistically significant of the b value from 3 implies either a positive or negative allometric growth ($P < 0.05$). An isometric growth is when b is not significantly different from 3 ($P > 0.05$) [20]. The coefficient of determination, r^2 value, was used to determine whether significant correlation exists between length and weight measurements. All data was analyzed using R software [16] and the significance of observed differences judged at the level of $p < 0.05$. Shapiro – Wilk normality test was used to determine the normality of b distribution.

3. Results

3.1. Growth parameters

A total of 696 specimens representing 13 species were collected from Nyangweta tributary during the study period. A total of 615 individuals of seven dominant fish species belonging to 4 families (*Amphiliidae*, *Machokidae*, *Clariidae* and *Cyprinidae*) were used for Length-Weight analyses. The number of fish sampled ranged from 4 specimens for *A. Jacksonii* to 254 specimens for *L. altianalis* as shown in Table 1

The lowest standard length of fish species was 3.3 cm in *Chiloglanis sp* while the largest was 26.3 cm in *L. altianalis*. The smallest specimen had a body weight of 0.8 g while the largest specimen had 371.3 g. The largest species was *E. altianalis* (70.74 ± 8.72 g), followed by *L. altianalis* and *L. victorianus* with mean body weights of 62.14 ± 3.48 g and 56.43 ± 11.26 g respectively. The maximum and minimum total lengths recorded for *E. altianalis* species in the Nyangweta tributary was 6.0 cm and 26.0 cm respectively, which were comparatively lower than the recorded maximum (32.3 cm) and minimum (6.5 cm) length for *L. altianalis* species which was second most important in terms of mean total length. *Labeo victorianus* showed the largest variation in both body weight and total length measurements in all stations during the sampling period ($P < 0.05$). The total length recorded for the *L. victorianus* species ranged from 10.1 - 25.3 cm, while body weight ranged from 10.29 g to 172.77 g.

3.2 The Fulton's condition factor

The values of Fulton's condition factors (the KF values) showed that all the species were in good condition during the sampling period and comparison among the fish species condition factor is shown in Table 2.

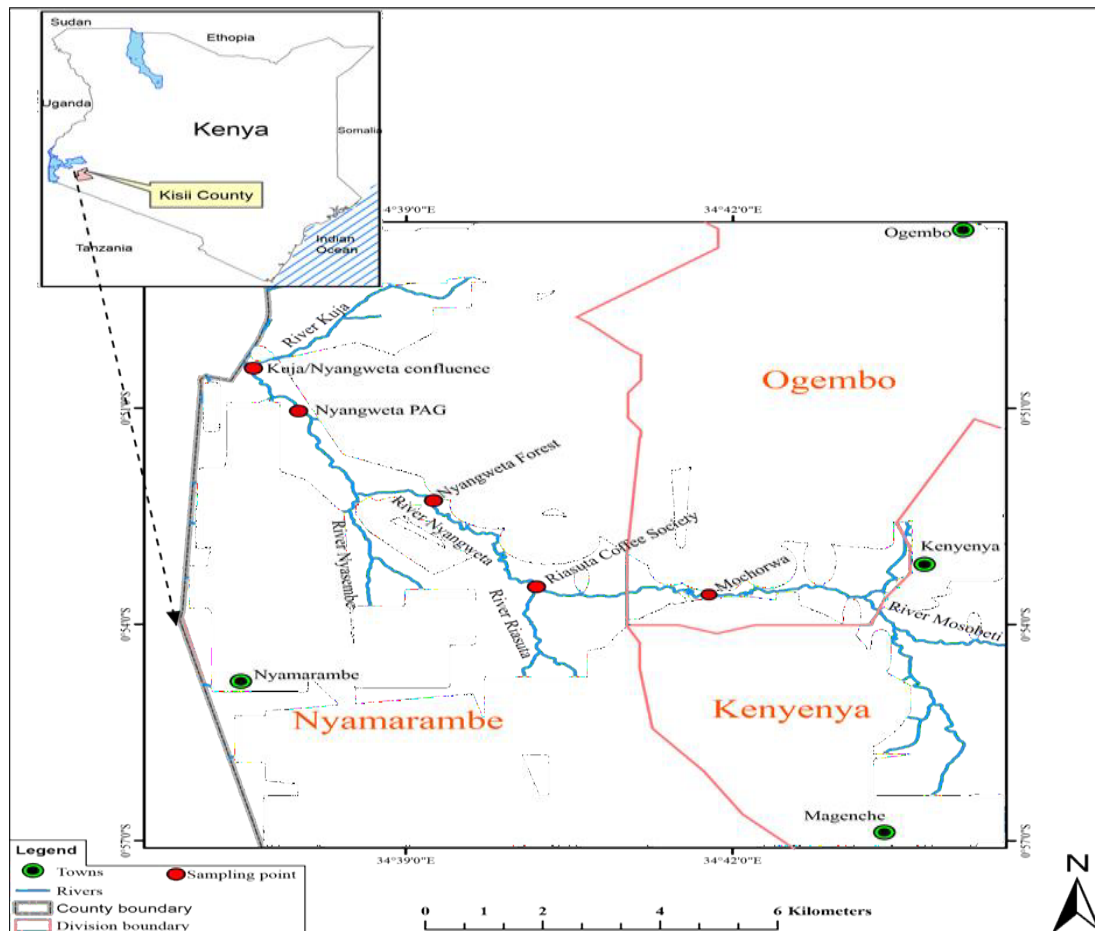


Figure 1: Map showing the sampling stations along the River Nyangweta tributary.

Table 1: Species sampled at each sampling station during the sampling period.

4. Species	Occurrence	Abundance	NKC	PAG	NFO	RCF	MOC
<i>Amphilius jacksonii</i>	1	4	4	0	0	0	0
<i>Chilogranis sp</i>	3	12	8	0	0	3	1
<i>Clarias theodorae</i>	6	210	31	29	70	56	24
<i>Clarius gariepinus</i>	1	1	0	1	0	0	0
<i>Enteromias cercops</i>	1	1	0	0	1	0	0
<i>Enteromius neumayeri</i>	5	142	69	15	15	36	7
<i>Enteromius nyanzae</i>	1	1	1	0	0	0	0
<i>Entoromius altianalis</i>	5	41	13	17	1	4	6
<i>Entoromius apleogramma</i>	1	2	0	0	0	2	0
<i>Labeobarbus altianalis</i>	5	254	53	63	7	73	58
<i>Labeo victorianus</i>	2	25	11	14	0	0	0
<i>Mastacembelus frenatus</i>	1	2	2	0	0	0	0
<i>P.multicolor</i>	1	1	0	0	0	0	1
Total	= 5 times						
Mean ± SE							
No of species	N = 13 species		9	6	5	6	6

Note: NKC is Nyangweta Kuja Confluence, PAG – Nyangweta PAG, NFO -Nyangweta forest, RCF – Riasusa Coffee Society, MOC – Mochorwa

Table 2: Mean (\pm SE) of Body weight, condition factor, total and standard lengths of seven Fish species along Nyangweta Tributary

Species	n	Body weight (g)		Total length (cm)				Standard length (cm)
		Mean \pm SE	Range	Mean \pm SE	Range	KF Mean	KF Range	Mean \pm SE
<i>A. jacksonii</i>	4	6.0 \pm 1.12	3.16 -8.50	8.1 \pm 0.72	6.4 - 9.9	1.11 \pm 0.10	0.88 -1.33	6.38 \pm 0.47
<i>Chiloglanis sp</i>	12	1.43 \pm 0.13	0.8 - 2.3	4.73 \pm 0.10	4.2 - 5.4	1.32 \pm 0.07	0.94 -1.63	4.01 \pm 0.12
<i>C. theodora</i>	211	26.73 \pm 1.65	2.1 - 163.12	13.6 \pm 0.3	6.2 -28.0	0.97 \pm 0.06	0.18 - 11.41	12.25 \pm 0.28
<i>E. neumayeri</i>	142	9.13 \pm 1.87	2.0 - 265.8	8.58 \pm 0.18	5.5 -26.0	1.06 \pm 0.02	0.58 - 1.51	6.80 \pm 0.16
<i>E. altianalis</i>	41	70.74 \pm 8.72	2.7 - 180.80	17.62 \pm 0.84	6.0 -26.0	0.98 \pm 0.02	0.55 -1.25	13.68 \pm 0.71
<i>L. altianalis</i>	254	62.14 \pm 3.48	2.97 - 371.3	17.07 \pm 0.3	6.5 -32.3	1.02 \pm 0.02	0.29 -5.30	13.45 \pm 0.24
<i>L. victorianus</i>	24	56.43 \pm 11.26	10.29 - 172.77	15.77 \pm 0.98	10.1-25.3	1.09 \pm 0.02	0.77 - 1.29	12.75 \pm 0.76

Table 3: Length-Weight relationships for seven dominant Fish Species along the Nyangweta tributary of Kuja River.

Species	<i>E. altianalis</i>	<i>E. neumayeri</i>	<i>C. theodora</i>	<i>L. victorianus</i>	<i>L. altianalis</i>	<i>Chiloglanis Sp</i>	<i>A. jacksonii</i>
Ka	0.008	0.005	0.013	0.006	0.008	0.114	0.050
B	3.17	3.40	2.55	3.14	2.97	3.19	2.27
Sb	0.07	0.07	0.08	0.08	0.04	0.24	0.50
$\frac{Sb}{b-3}$	0.41	0.18	-0.18	0.57	-1.33	1.26	-0.68
r ²	0.984	0.947	0.820	0.986	0.948	0.702	0.911
Adj. R ²	0.983	0.947	0.819	0.985	0.947	0.672	0.866
P	0.000	0.000	0.000	0.000	0.00	0.001	0.046
t-test sig.	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Growth behaviour	Positive allometry	Positive allometry	Negative allometry	Positive allometry	Isometry	Positive Allometry	Negative allometry
IUCN status	Least Concern	Least Concern	Least Concern	Critically Endangered	Least Concern	Least Concern	Least Concern

Note: a and b are length - weight relationship parameters; negative allometry, $Ka < 3$, Positive allometry $Ka > 3$, isometry, $Ka = 3$. Sb is the standard error of b; P is the p value; r² is the coefficient of determination.

With exception of *E. altianalis* and *C. theodora*, the Fulton's condition factors (KF) of all the species were greater than (1) and not significantly different ($p = 0.43$)

among the species. The calculated mean condition factor (KF) was the highest in *Chiloglanis sp* (1.32 \pm 0.07) and lowest in *Clarias theodora* (0.97 \pm 0.06) and *E. altianalis*

(0.98 ± 0.02), which had the condition factors of less than (1).

4.1. Length-Weight relationships of seven fish species along Nyangweta tributary

The length-weight relationship parameters and information on the growth behavior (isometric or allometric) for the seven fish species ($n = 615$) sampled from Nyangweta tributary are shown in Table 3.

E. altianalis, *E. neumeyeri*, *L. victorianus* and *Chilogramis sp* showed a positive allometric growth pattern during this study indicated by the b values ($b = 3.17, 3.40, 3.14$ and 3.19). The computed 'b' values were lowest and less than 3 for *C. theodora* and *A. jacksonii* species which had b values of 2.55 and 2.27 respectively, while *L. altianalis* ($b = 2.97$) exhibited isometric growth. The fish total length increased correspondingly with body weight resulting to high correlation coefficients due to a large percentage of body weight explained by total length in the log-transformed regression models (Figure 2).

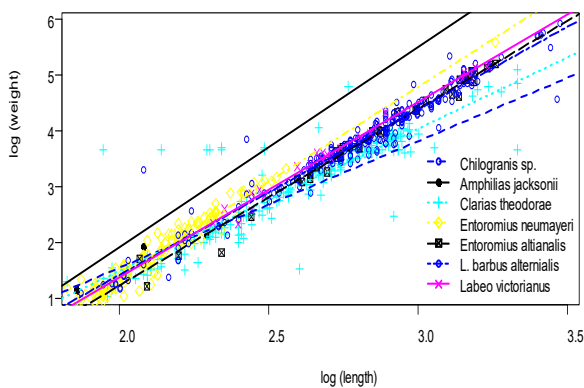


Figure 2: Variations Length-Weight relationships among the seven dominant fish species in River Nyangweta tributary

There was a significant positive relationship between body weight and length which was shown by correlation coefficients ($r^2 > 0.5$, $P < 0.05$). The correlation coefficient was however not found significant in *A. jacksonii* ($R^2 = 0.9107$, $p = 0.046$), showing that individual fish in this species did not grow stouter with increase of total length. Further, the adjusted R^2 values of *E. altianalis*, *E. neumeyeri* and *A. jacksonii* sampled showed a high correlation between the total length and bodyweight with the R^2 values of 0.9835, 0.9474 and 0.9107 respectively. However, *Chilogramis sp* species had the lowest coefficient of differentiation ($R^2 = 0.6721$). With exception of *Chilogramis sp* and *A. jacksonii*, the standard error (Sb) of each length-weight coefficient 'b' was less than 10 times smaller than the corresponding parameter estimate coefficient.

5. Discussion

The location of the stream in a high agriculturally potential area characterized by regular and frequent flooding might have played a pivotal role in determining the biological condition of the Nyangweta tributary, besides replenishing the available food with overland influxes that provide a cue for spawning and create a food base for fish and other aquatic organisms. With exception of *C. theodora*, *E. neumayeri*, *E. altianalis*, *L. altianalis*, other species showed preferred dominating certain areas of the tributary and were completely absent in other sampling stations. The size distribution along the stream gradient was influenced by the variation in river bio-physico-chemical parameters such as nutrient levels, the habitat quality and heterogeneity, stream width, water depth and temperature as well as energy dynamics which contributed to species succession.

There were heterogeneous fish habitats in the riverbanks, substrate and slow-moving pools of water among which were reflected in the diverse composition and structure of the biological community. The variations in fish allometric and Fulton's condition values depended on many factors such as the number of specimens examined, health condition of the habitat, even the season during which sampling was conducted [21]. However, the change in b values depends primarily on the individual body profiles of the species, physico-chemical parameters, food availability, sex of the individual as well as maturity stage of the individuals [13], [22], [23]. Some of these factors did not account for the variations in the fish condition factors observed in the present study. Most K_a values were greater than or less than 3 and highly significant showing that the increase in body weight of fish corresponds with an increase in fish total lengths. Consequently, most b -coefficients deviated from isometry ($b = 3$) observed for most species and ranged between 2.27 and 3.19. This may be because these fishes exhibited changes in shape and density during the sampling period which might have caused the slope coefficients of the weight - length regression to deviate from 3. In addition, the electrofisher sampling equipment had low efficiency with several fish escaping along the sampling transect. Nevertheless, these values lie within the range of 2.5 and 3.5 which has been reported for most species [6], [24]. The low 'Sb' to 'b-3' ratio in this study indicates low variability of the slope estimate coefficients from isometry, which further implies that the observed data yielded best-fitting regression models which could have a high precision in making predictions outside the observed datasets [17].

Moreover, environmental changes resulting from increased anthropogenic activities, such as land use changes in the riparian zone may have been caused

deterioration in water quality, consequently reducing habitats and food availability for the species. This is evidenced by increased riparian activity in which land was cultivated close to the river bank. For instance, *L. victorianus* and *M. frenatus* had low abundance in those areas with high riparian activity. *C. theodora* showed negative allometric growth because majority of the individuals sampled from areas with high anthropogenic activity were small sized. Fish that occupy stress free aquatic environments are likely to have grown stouter and heavier resulting to 'b' values of 3 and greater than 3 [25]. According to [26], isometric growth occurs when the volume of a fish is constant or taken as a surrogate of a linear measure of the size being modeled. A highly significant linear relationship between total and standard lengths and body weight indicates that body weight measurements in these riverine species could be reliably predicted from total length measurements after linearizing the curvilinear to obtain highly significant intercept and slope coefficients of the regression.

The allometric condition factor of fish is also governed by the feeding behaviour of fish. As the fish changes shape or density as it grows, it can become "plumper" with further increases in length assuming allometric growth [18]. Human activities also affect the river condition and influence the biological condition within the fishery ecosystem by altering water quality, flow regime, physical habitat structure and interactions among species [27]. For instance, the replacement of the natural vegetation cover with exotic *Eucalyptus* spp., may result to siltation, which has been worsened by the planting of annual crops which require frequent cultivation. As a result, large and more stable substrate, such as stones that offer surfaces for the attachment of macrophytes and algae which provide a food source for insects and fish is removed, resulting to loss of fish habitats and food reserves [28].

6. Conclusion and Recommendations

Most species in River Nyangweta Tributary exhibited allometric growth condition (b is greater than or less than 3) and only *L. alternialis* exhibited isometric growth among the seven dominant species. There were no references on LWRs and condition for most of the species sampled in the present study.

Fish sampled from and those with high abundance in upstream areas (MOC, RCS and NFO), which had high riparian activity exhibited negative allometry ($b < 3$), which indicated poor physiological condition. Anthropogenic activity, together with other factors such as inefficient sampling gear contributed to small sample size of fish used in calculating the LWRs and condition factors, the information obtained on native fish species in the Nyangweta tributary of River Kuja, which could limit

generalizations on the fishery potential, recruitment and the physiological condition of the riverine fish populations with respect to environmental changes and human population growth. These findings suggest that conservation measures need to be taken to improve the fish health and physiological condition in upstream areas of the Nyangweta tributary.

7. Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Special thanks to Kenya Marine and Fisheries Research Institute (KMFRI), Kisumu, Kenya for the technical support during collection and measurements of the fish samples. This study was funded by the Germany Exchange Program (DAAD) Masters Scholarship (Personal ID: 91637034) awarded to Mr. Fredrick Ondemo. The authors would also like to thank Kisii University for the institutional support received to obtain the required research permits.

References

- [1] J. E. Beyer, "On length-weight relationship. Computing the mean weight of the fish of a given length class.", *Fishbyte*, vol. 5, pp. 11–13, 1987.
- [2] I. A. Abdel-Raheem, H. A. E., Abeer-Mohamed, H. M. K., Izeldeen, I. A., Mohammed, A. E. & d Mohammed, "Length-Weight Relationship and Condition Factor of Three Commercial Fish Species of River Nile, Sudan", *EC Oceanography*, vol. 1, no. 1, pp. 01–07, 2017, doi:10.13140/RG.2.2.12579.37921.
- [3] P. Marx, K. K., Vaitheeswaran, T., Chidambaram, P., Sankarram, S. and Karthiga, "Length - Weight Relationship of Nile Tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Family: Cichlidae)", *Ind. J. Vet & Anim. Sci. Res.*, vol. 43, no. 1, pp. 33–39, 2014.
- [4] P. Brosset, P. Fromentin, J., Van-Beveren, E., Lloret, J. & Marques, V., "Spatio-temporal patterns and environmental controls of small pelagic fish body condition from contrasted Mediterranean areas.", *Progress in Oceanography, Elsevier*, vol. 151, pp. 149–162, 2017, doi:10.1016/j.pocean.2016.12.002.
- [5] P. L. C. T. Bolger, "The selection of suitable indices for the measurement and analysis of fish condition", *Journal of Fish Biology*, vol. 34, no. 2, pp. 171–182, 1989, doi:10.1111/j.1095-8649.1989.tb03300.x.
- [6] R. Froese, "Cube law, condition factor and weight-length relationships: history, meta- analysis and recommendations.", *Journal of Applied Ichthyology*, vol. 22, no. 4, pp. 241–253, 2006, doi:10.1111/j.1439-0426.2006.00805.x.
- [7] Z. Gichana et al., "Effects of human activities on benthic macroinvertebrate community composition and water quality in the upper catchment of the Mara River Basin, Kenya", *Lakes & Reservoirs: Research & Management*, vol. 20, no. 2, pp. 128–137, 2015, doi:10.1111/lre.12094.
- [8] S. E. Orina N. E, Getabu A, Omondi R, "Ecosystem integrity of River Kuja based on habitat and fish biotic indices.", *International Journal of Tropical Hydrobiology & Fisheries*, vol. 1, no. 1, pp. 7–13, 2018.
- [9] T. Pj, A., Lk, H., Jr, P., Lj, J., Rhodes, T., & Ms, "Spatial patterns in fish biomass and relative trophic level abundance in a waste water enriched river", *Ecology of FRESHWATER FISH*, vol. 16, no. 3, pp.

- 343–353, 2007, doi:10.1111/j.1600-0633.2007.00221.x.
- [10] M. Silva, L. M. A., Oliveira, M. S. B., Florentino, A. C., & Tavares-Dias, "Length-weight relationship of 11 fish species from a tributary of the Amazon River system in northern Brazil.", *Journal of Applied Ichthyology*, vol. 1, no. 4, pp. 816–817, 2015, doi:10.1111/jai.12791.
- [11] M. Oliveira, M. S. B., Silva, L. M. A., Prestes, L., & Tavares-Dias, "Length-weight relationship and condition factor of 11 fish species of the Igarapé Fortaleza basin, a tributary from the Amazon River system in eastern Amazon (Brazil)", *Journal of Applied Ichthyology*, vol. 34, no. 4, pp. 1038–1041, 2018, doi:10.1111/jai.13679.
- [12] F. C. Prestes, L., Oliveira, M. S. B., Tavares-Dias, M., Soares, M. G. M., & Cunha, "Length-weight relationship and condition factor of eight fish species from the upper Araguari River, State of Amapá, Brazil.", *Acta Scientiarum Biological Sciences*, vol. 41, no. 1, pp. e46666, 2019, doi:10.4025/actasciobiolsci.v41i1.46666.
- [13] S. C. Sparre, P. & Venema, *Introduction to tropical fish stock assessment. Part 1. Manual* Rome, Italy, , FAO (Food and Agricultural organisation), 1998.
- [14] D. W. Taylor, P., Blackwell, B. G., Brown, M. L., Willis, "Relative Weight (Wr): Status and current use in fisheries assessment and management", *Reviews in Fisheries Science*, vol. 8, no. 1, pp. 37–41, 2010, doi:10.1080/10641260091129161.
- [15] Le Cren E.D., "Length-weight relationship and seasonal cycle in gonad weight and condition in perch (*Perca fluviatilis*).", *Journal of Animal Ecology*, vol. 20, no. 2, pp. 201–219, 1951, doi:10.2307/1540.
- [16] R Core Team, "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna." Vienna, 2020.
- [17] P. M. Omweno J. O., Orina P. S., Getabu, A. and Ondieki, "Comparative growth of Jipe tilapia, *Oreochromis jipe* and Nile tilapia, *Oreochromis niloticus* using regression modeling", *Int. J. Adv. Res.*, vol. 8, no. 09, pp. 984–992, 2020.
- [18] D. W. Blackwell, B. G., Brown, M. L. & Willis, "Relative Weight (Wr): Status and current use in fisheries assessment and management.", *Reviews in Fisheries Science*, vol. 8, pp. 1–44., 2000, doi:10.1080/10641260091129161.
- [19] W. G. Snedecor, G.W. and Cochran, *Statistical methods*. Ames, Iowa, the Iowa state University., 6th Editio. Ames, Iowa, , the Iowa state University, 1967.
- [20] N. Yilmaz, S., Yazmaz, S., O., Erbas_aran, M., Esen, S., Zengin, M., Polat, "Length weight relationship and relative condition factor of white bream, *Blicca bjoerkna* (L., 1758), from Lake Ladik, Turkey", *J. Black Sea/Medit. Environ*, vol. 18, no. 3, pp. 380–387, 2012.
- [21] S. Karna, S. K., Sahoo, D. & Panda, "Length Weight Relationship (LWR), Growth estimation and Length at maturity of *Etioplos suratensis* in Chilika Lagoon, Orissa, India.", *International Journal of Environmental Sciences*, vol. 2, no. 3, pp. 1257–1267, 2012, doi:0.6088/ijes.00202030012.
- [22] O. Cherif, M., Zarrad, R., Gharbi, H., Missaoui, H. & Jarboui, "Length-weight relationships for 11 fish species from the Gulf of Tunis (SW Mediterranean Sea, Tunisia)", *Pan- American Journal of Aquatic Sciences*, vol. 3, no. 1, pp. 1–5, 2008.
- [23] J.O. Omweno, "Aquaculture Potential of Jipe tilapia (*Oreochromis jipe* Lowe, 1955).", Kisii University, 2021.
- [24] D. M. Britton, J. R. & Harper, "Juvenile growth of two tilapia species in Lakes Naivasha and Baringo, Kenya", *Ecology of FRESHWATER FISH*, vol. 17, no. 3, pp. 481–488, 2008, doi:10.1111/j.1600-0633.2008.00298.x.
- [25] G. Prasad and P.H. Anvar Ali, "Length-weight relationship of a cyprinid fish *puntius filamentosus* from Chalakudy River, Kerala.", *Zoos' Print Journal*, vol. 22, no. 3, pp. 2637–2638., 2007.
- [26] D. Ogle, "fishR Vignette - Length-Weight Relationships." Winsconsin, 2013.
- [27] USEPA, "A practitioner's guide to the biological condition gradient:

A framework to describe incremental change in aquatic ecosystem." Washington, DC, 2016.

- [28] J. D. Wersal, R. M., & Madsen, *International Plant Protection Convention- Aquatic Plants and Their Uses and Risk:A review of the global status of aquatic plants*Rome, Italy, , International Plant Protection Convention (IPPC), 2014.

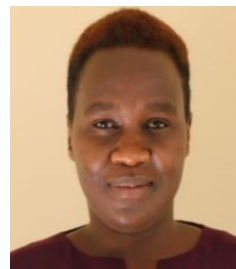
Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).



Fredrick Ondemo received a BSc. degree from Kisii University, Kenya. He is an MSc. student in the same University, Department of Environment, Natural Resources & Aquatic Sciences. He works as a Fisheries Officer in the Kisii County Government. His research interests include Fisheries and Aquatic Ecology.



Albert Getabu is currently an associate Professor at Kisii University, Department of Environment, Natural Resources & Aquatic Sciences. He has a PhD in Fisheries from University of Hull, an MSc and BSc from University of Nairobi. He has a wide experience in fresh water research with bias on fisheries and limnology.



Zipporah Gichana received her PhD in Natural Resources & Life Sciences from BOKU University, Vienna, Austria, MSc. degree in Fisheries & Aquatic Sciences from the University of Eldoret, Kenya, BSc. from Moi University, Kenya. Post graduate training from IHE Delft Institute for Water Education, Netherlands & BOKU University, Vienna, Austria. Her research interests include Fisheries, wastewater management,

sustainable aquaculture production systems and limnology.



Job Omweno received his MSc in Fisheries from Kisii University, Kenya. He has a BSc degree in Applied Limnology from Egerton University, Kenya. His research interests include: Fisheries, aquaculture and aquatic Sciences.