Editorial

In the ever-expanding landscape of research, diverse disciplines converge to push the boundaries of knowledge and innovation. Here, we spotlight a selection of recent 6 studies that delve into distinct realms, from public health crises to cutting-edge technological solutions. These investigations not only shed light on pressing issues but also exemplify the breadth and depth of contemporary research endeavours.

The Ebola virus remains a formidable threat to global health security, and understanding its epidemiological dynamics is paramount. In a rigorous analysis utilizing Factorial analysis techniques, researchers scrutinized reported cases and death rates across multiple countries. Their findings, communicated through graphical representations, provide invaluable insights into effective strategies for combating this lethal disease. By employing advanced analytical tools such as R Studio, this study contributes to the arsenal of knowledge aimed at mitigating the impact of Ebola outbreaks worldwide [1].

Amid growing concerns about radiation hazards, a meticulous investigation focused on measuring background radiation levels in radiation facilities. Through systematic monitoring and dose calculations, researchers evaluated the safety of operating console areas for radiation workers. Encouragingly, the study affirms compliance with international safety standards, alleviating undue anxiety surrounding radiation-related risks in cancer hospitals. These findings underscore the importance of stringent regulatory protocols in safeguarding the well-being of healthcare professionals [2].

Decision-making under uncertainty pervades various domains, from artificial intelligence to consumer preferences in the mobile phone market. Leveraging fuzzy matrix theory, researchers devised a methodology for selecting preferred mobile phones with diverse features within budget constraints. This innovative approach not only enhances consumer satisfaction but also exemplifies the applicability of fuzzy logic in addressing real-world decision-making challenges. In an era marked by technological proliferation, such endeavours pave the way for more informed choices and enhanced user experiences [3].

In the realm of quantum computing, error correction poses a formidable challenge. Researchers devised a novel machine learning-based decoding scheme to rectify errors in surface code, achieving significant improvements over conventional techniques. By formulating the decoding problem as a classification task and employing sophisticated ML models, this study demonstrates remarkable enhancements in decoder performance. These findings hold promise for advancing the reliability and scalability of quantum computing systems in the face of complex noise models [4].

Real-time monitoring of respiratory patterns holds immense potential for early detection of pulmonary abnormalities. A pioneering study introduces a novel method utilizing off-the-shelf anti-snoring devices equipped with microphones for data collection. By recording and analysing audio patterns, researchers offer a user-friendly tool for self-diagnosis and data interpretation. This innovative approach facilitates early intervention and complements traditional diagnostic methods, heralding a new era in respiratory healthcare [5].

Amidst the digital transformation of educational institutions, ensuring data security remains paramount. Leveraging blockchain technology, researchers propose a robust framework for securing student records within university systems. By employing SHA-256 encryption and adopting a cascade development process, this study exemplifies the potential of blockchain in enhancing data integrity and privacy. These insights not only bolster institutional resilience but also pave the way for streamlined administrative processes and improved service delivery [6].

In conclusion, these research endeavours exemplify the multifaceted nature of contemporary scholarship, spanning from public health crises to technological innovations and beyond. As researchers continue to push the boundaries of knowledge, their collective efforts pave the way for a brighter, more informed future.

References:

[1]   V. Paritala, H. Thummala, "Factorial Analysis to Categories Spread and Effect of Ebola Virus from Various Countries," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 1–6, 2022, doi:10.55708/js0106001.

[2]   M. Waqar, T.A. Afridi, Q. Soomro, A.S. Abbasi, M. Shahban, "Measurement of Ambient Ionizing Radiation Exposure in Operating Consoles of Radiation Modalities in Cancer Hospital NORIN Nawabshah, Pakistan," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 7–12, 2022, doi:10.55708/js0106002.

[3]   J.A. Shah, "Fuzzy Matrix Theory based Decision Making for Machine Learning," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 13–20, 2022, doi:10.55708/js0106003.

[4]   D. Bhoumik, P. Sen, R. Majumdar, S. Sur-Kolay, L.K. KJ, S.S. Iyengar, "Machine-Learning based Decoding of Surface Code Syndromes in Quantum Error Correction," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 21–35, 2022, doi:10.55708/js0106004.

[5]   A.M. Khatkhate, V. Raut, M. Jadhav, S. Alva, K. Vichare, A. Nadkarni, "Identification of Basic Respiratory Patterns for Disease-related Symptoms Through a Microphone Device," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 36–44, 2022, doi:10.55708/js0106005.

[6]   O. Sarjiyus, I. Isaiah, "Blockchain Based Framework for Securing Students' Records," *Journal of Engineering Research and Sciences*, vol. 1, no. 6, pp. 45–54, 2022, doi:10.55708/js0106006.

Editor-in-chief

Prof. Paul Andrew

JOURNAL OF ENGINEERING RESEARCH AND SCIENCES

CONTENTS

Factorial Analysis to Categories Spread and Effect of Ebola Virus from Various Countries

Venu Paritala*, **Harsha Thummala**

Department of BioTechnology Vignan's Foundation for Science, Technology & Research, Guntur, A.P, India

*Corresponding author: Venu Paritala, Contact No: +91 7799509079, Email: vvenuparitala@gmail.com
ORCID Author 1: 0000-0003-0385-5322, ORCID Author 2: 0000-0002-2924-723X

ABSTRACT: The main objective of this research analyze and correlates the number of cases and death rates reported on the Ebola virus in many countries. Ebola virus is one of the most lethal diseases to infect humans. This approach Proceeding uses the Factorial analysis technique in the Ebola Virus dataset. This method takes the largest common variance from all criteria and combines it into a single score. The analysis is applied to reported and categorize the most effective way countries using the counts of cases and deaths. To this investigate get it, and communicate data in a way of graphic factual properties. The analytic results on the Ebola virus dataset are not accessible anywhere. The Analysis was done by R Studio.

KEYWORDS: Ebola Virus, Factorial analysis, Correlation, Cases

1. Introduction

Ebola virus is also called Ebola hemorrhagic fever. This virus leads to bleeding and calamitous consequence on organ failure which leads to death in several countries. The virus spread from fruit bats accidentally its admittance into the human inhabitants through various contact such as selection of organs and body fluids contaminate animals like monkey, forest antelope porcupine, etc[1]. Ebola spread through human-to-human by-way-of unclean needles touching a contaminated surface and through saliva [2]. The Ebola virus is the impact on the social human development and Lades to the crisis on education, health, economy, and quality of the living system [3]. This crisis eventually carries and differ in several countries. The virus is rapidly increasing in several countries the degree of fear, concern increased inside the people [4]. Mostly the pathogenic doctors of USA they create the awareness within the children, adults threw this virus.

The virus is highly affected in many nations, it is not always the case that the server has the highest death rate [5]. Later global great effect on scientific, economical and sociological. The fundamental aim in maximum of this research is to track and count on how the virus will circulate throughout the country. Researchers have worked difficult to estimate how long the outbreak would remain [6]. As a result, many studies organizations have used numerous modeling techniques for prediction and feature provide you with many intriguing conclusions [7]. The main aim of this research study is to investigate the correlations between the number of Ebola cases and the number of fatalities caused by the disease in several nations that have been badly affected by this pandemic epidemic. To begin, the correlation coefficients are computed to discover the links between these nations. The factor analysis is then used to classify these nations based on the number of cases and fatalities.

2. Materials and Methods

Whatever the study design, the effect of Ebola should first be described (description). For example, Positive and Death cases can be given as the mean value, Coefficient Correlation and standard deviation (for normal distribution), or as the median and range, or in a histogram. Studies on Effect of Ebola Virus in Various Countries also examine influence factors such as Death rate status and annual affect. These variables should be included in the analysis, as they may be risk factors for Ebola Virus and are potential confounders [8]. Risk factors may also be effect modifiers. Effect modification means that the influence of one factor (for example, Ebola Virus) on a Positive cases (for example, death rate) is modified by the presence of another factor (for example, Annual impact). In other words, there is an interaction between the two factors. The effects should be examined in

different subgroups (stratification), each with the same analysis.

Factor analysis is a statistical method applied to the Ebola virus dataset it initial set of input variables that are cases, deaths,and country to have mutual correlations to find a smaller set of factors that describe the underlying interrelationships and mutual variability [9]. On a co-variance data matrix, the function conducts maximum-likelihood factor analysis. The argument factors provide the number of elements to be fitted.

2.1. Datasets

This study, count the number of cases and number of Deaths due to Ebola virus reported in Guinea, Congo, Italy, Spain, UK, USA, Senegal, Mali, Nigeria, Sierra Leona, Liberia, Guinea, Uganda, Sudan, Gabon, South Africa (ex-Gabon) and Cote from the start the year of 1976 to 2021 to be considered. This investigation is to estimate Ebola virus potential dispersion with an accentuation of the number of positive unused occasions, mortality rate and recuperations. Information retrieve from the World Health Organization [10]. This dataset contains information about 41 instances of positive and 41 are negative instances, 6 attributes are comprised of every record, which all are numeric values. The Ebola virus dataset parameter observed in Table 1.

Table 1: Dataset Parameter Description

S. No	Description
1	Year
2	Country
3	Ebola virus species
4	Cases
5	Deaths
6	Case fatality

It can be observed that cases and deaths present minimum in Guinea and maximum present in Congo due to the Ebola virus [11]. The plots for the counts of the cases with the Ebola virus and the deaths due to it are also demonstrated in Figure 1. The relationships between the rates of the positive cases, death, and spread of Ebola among these countries have been studied using coefficient correlation by the factors, it examines associations between variable measurements to understand how the variables themselves group together and are related. It can be seen in Table 2 that the rate of cases, death, and spread of the virus in various countries is very strong [12]. In order of correlation factor observer in the overall analysis of 1 quality and 3 rd quality of arithmetic means and

median of cases and deaths due to Ebola virus [13]. It also observes the minimum and maximum rofe in the effect of the virus. The order of analysis cases reported in countries in cumulative total Minimum reports is 0 and Maximum is 14124 the death.
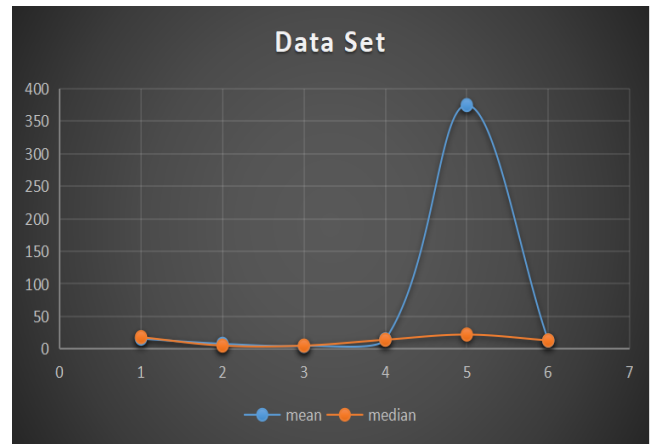
Figure 1: Mean and median of Ebola Virus dataset

Table 2: Mean and Median for the number of cases and deaths due to Ebola Virus

Correlation Factor 1	Cases		Death
Min	0		0
1st Quality	4		1
Median	34		22
Mean	10675		374.5
3rd Quality	857.5		55
Maximum	14124		4809

2.2. Factorial Analysis Model

Factor analysis is a method for condensing a large number of variables into a smaller number of components [14]. This method takes the largest common variance of all variables and combines them into a single score, as shown in Table 3. We may use this score as an index of all variables for future investigation [15]. Exploratory factor analysis (EFA) is a technique for determining a measures factor structure and assessing its internal reliability [16]. When researchers have no theories regarding the nature of the underlying factor structure of their measure, EFA is frequently advised. The goal of this study is to determine the primary factors that affect the Ebola virus data set. An objective strategy incorporating multiple methodologies (statistics, R, GIS, descriptive analysis, and graphical visualization) was used to examine such a complex algorithm [17, 18]. FA is concerned with the correlations of variables so that the variables within a

factor are highly linked with one another, while the variables within distinct factors are substantially uncorrelated [19]. Remember the core premise behind the factor analysis before interpreting the findings. To isolate the variables underlying commonality, factor analysis builds linear combinations of components. Fewer factors represent the majority of the variation in the data set to the degree that the variables have an underlying commonality [19]. This lets us describe an underlying nation in a model by aggregating a large number of observable variables, making the data easier to grasp. Our data variability, XX, is provided by, and its estimate is made up of the variability described by Σ variables explained by a linear combination of factors (commonality) and the variability that cannot be explained by a linear combination of factors (uniqueness).

$$\Sigma = XX^T + \Psi \tag{1}$$

To, calculates an initial estimate of S−Ψ and factors S−Ψ^S−Ψ^, or R−Ψ^R−Ψ^ for the equivalence array. Regrouping the estimated co-variance and equivalence array with the estimated XX array yields:

$$S-\Psi=X$$

$$X=S-\Psi \tag{2}$$

$$R-\Psi=X$$

where S is a pp x 1 vector of observed variable scores, R is a p x q matrix of factor loading's , Ψ is a qq x 1 vector of common factor scores and Σ is a p x 1 vector of unique factor scores.

The S−Ψ^i or R−Ψ^i. Ψ^Ψ^i is a diagonal array is the principal component technique, the h^2ih^i2, is equal to S−Ψ^i for S−Ψ^S−Ψ^ and 1−ψ^i1−ψ^i for R−Ψ^R−Ψ^. The correlation of HH or RR is replaced by their respective commonalities in ψ^iψ^i which gives us the following forms:

$$S-\Psi^i = \begin{bmatrix} h1, i1, s1 \\ s2, h2, i2 \\ i3, s3, hp \end{bmatrix} \tag{3}$$

$$R-\Psi^i = \begin{bmatrix} h1, r1, r12 \\ r2, h2, r21 \\ r3, ..., hp \end{bmatrix} \tag{4}$$

The squared multiple correlations between the observation vector of 1 and the other h variables are used to make an initial estimate of the commonalities. In the instance of r, the squared multiple correlations are comparable to the following:

$$h = 1 - \frac{1}{r} \tag{5}$$

3. Result

This segment reports the consequences of the Factorial analysis way to deal with several cases and deaths along with number of nations dependent on exploration the factors annual as observed in Table 5. It must be mentioned the number of primary factors in FA was computed the number of eigenvectors of the correlation matrix with greater affinity than one. Furthermore, the P values were greater than 0.8 as shown in Table 4, confirming the speed of the FA technique. Bartlett's test using Ebola virus dataset to confirm the value is p-value = 0.05, Then the data set is suitable for normally distributed. The Analysis was done by R Studio.
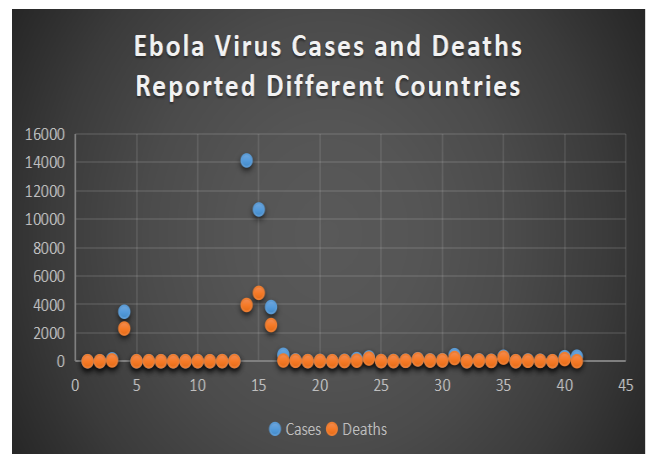
Figure 2: Counts of the Positive cases and Deaths report due to Ebola virus observed in many countries

Table 3: Calculate the Factorial classification of positive cases and deaths due to Ebola in various countries.

Country	Cases	Deaths
Guinea	1.0000	1.00000
Democratic Republic of the Congo	6.47219	2.00000
Democratic Republic of the Congo	2.3171	3.00000
Democratic Republic of the Congo	4.0304	4.00000
Democratic Republic of the Congo	1.0000	5.00000
Italy	1.0000	6.00000

Spain	1.00000	7.00000
UK	2.40001	8.00000
USA	1.00000	9.00000
Senegal	4.03004	1.00001
Mali	2.43018	1.10001
Democratic Republic of the Congo	4.05076	1.20001
Democratic Republic of the Congo	5.04003	1.30001
Uganda	6.20023	1.40001
Uganda	1.00000	1.50001
Uganda	2.63035	1.60001
Democratic Republic of the Congo	3.810260	1.70001
Democratic Republic of the Congo	4.79008	1.80001
Congo	3.56014	1.90001
Sudan	1.03040	2.00001
Congo	3.850247	2.10001
Congo	1.39080	2.20001
Congo	8.25090	2.30001
Uganda	1.00000	2.40001
South Africa (ex-Gabon)	8.32081	2.50001
Gabon	8.22033	2.60001
Democratic Republic of the Congo	1.00000	2.70001
Côte d'Ivoire	8.07067	2.80001
Gabon	2.95038	2.90001
Sudan	1.00000	3.00001

Table 4: Validation of p factor and accuracy of factorial analysis in cases and deaths due to ebola virus

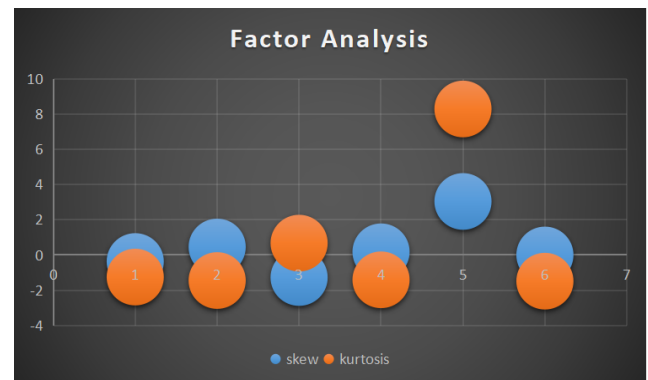| SI.No	Q1	Q2	Q3	Parameters
MSA	0.8	0.8	0.8	Cases
MSA	0.8	0.8	0.8	Deaths
MSA	0.8	0.8	0.8	Case fatality
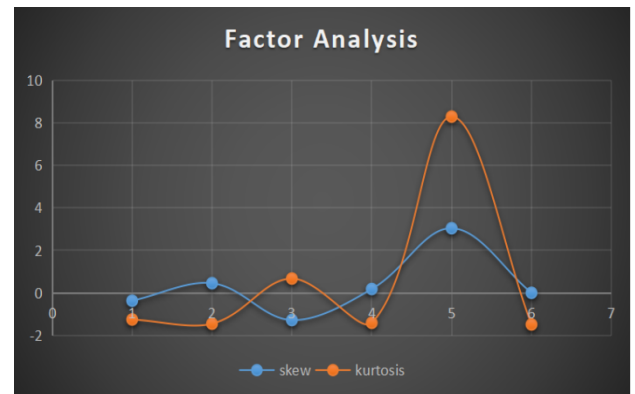
(a)

(b)

Figure 3: Ebola virus cases and Deaths are categorized using the FA method

Table 5: Ebola virus spread of cases and deaths correlated by the annual

Name	vars	n	Mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Year*	1	41	15.7317	7.79430	18	16.06060	7.413	1	27	26	-0.37170	-1.25495	1.217266
Country*	2	41	7.78048	5.71188	5	7.515152	4.4478	1	18	17	0.45374	-1.45162	0.892046
Ebola virus species*	3	41	4.43902	1.14124	5	4.606061	0	1	6	5	-1.28130	0.67151	0.178232
Cases*	4	41	14.6585	10.9718	14	14.18181	14.826	1	33	32	0.18049	-1.41134	1.713507
Deaths	5	41	374.487	1061.42	22	52.93939	32.617	0	4809	4809	3.03906	8.29111	165.7670
Case.fatality.*	6	41	12.7804	8.83604	13	12.63636	11.860	1	27	26	0.00049	-1.48798	1.379957

3.1. *Cumulative number of deaths due to Ebola virus*

Figure 5 shows the results of using the FA approach to categorize the increasing mortality rate in several nations based on the cumulative counts of Ebola virus deaths. The outputs show the statistical differences in the mortality rates associations.
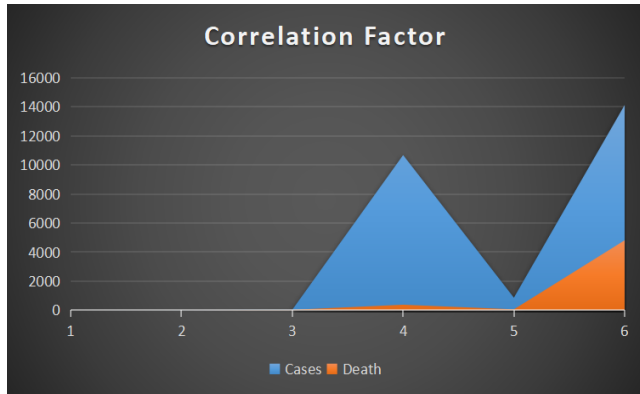
Figure 5: Correlation technique to categorize a cumulative number of of the cases and deaths due to Ebola virus

4. Conclusion

Ebola Virus is still an unclear infectious disease, the outbreak spreads are largely influenced by each country's policy and social responsibility. As data transparency is crucial inside the government, it is also our responsibility not to spread unverified news and to remain calm in this situation. In the fight against the Ebola virus, understanding preventative measures, population mobility, viral propagation. The study has shown the importance of information dissemination that can help to identify the most affected countries and its improving response time and help planning in advance to help reduce risk. Further studies need to be done to help contain the outbreak as soon as possible.

Because the Ebola virus has several effects on the environment, health, society, and economy, it is critical to investigate the spread of disease transmission to compare it across nations. The objectives of this paper were to research Ebola cases and deaths caused by this pandemic disease in different countries that have been significantly affected by it. The count number of cases and mortality due to Ebola virus reported in Guinea, Congo, Italy, Spain, UK, USA, Senegal, Mali, Nigeria, Sierra Leon, Liberia, Guinea, Uganda, Sudan, Gabon, South Africa (ex-Gabon), and Cote from the start the year of 1976 to 2021 to be considered. To begin, the correlation coefficients were calculated to discover the links between these nations. The results revealed that the number of positive cases and death rates in different nations had substantial positive associations.
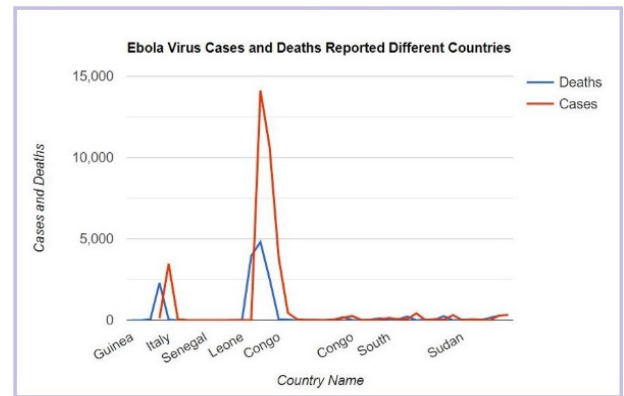
Figure 6: FA technique to categorize the countries on basis of the cumulative counts of the deaths due to the Ebola Virus

The factor analysis was done the count of positive cases and mortality rate. In the perspective of Ebola virus cases, Congo, Liberia, and Italy were distributed similarly to one another but different from other nations as shown in Figure 6. Meanwhile, for deaths, Ongo, Leone, and Italy were distributed similarly to one other but different from other nations. In a future study, the authors propose classifying Ebola virus datasets from more countries using the FA approach or applying this technique to categorize locations for other epidemic or pandemic illnesses.

Conflict of Interest

The authors state that they do not have any conflict of interest.

Authors' contributions

Both authors have an equal contribution.

Acknowledgment

References

[1] K. AP, Appelgren K, Chevalier MS, McElroy A. Ebola Virus Disease: Focus on Children. Pediatr Infect Dis J. 2015 Aug;34(8):893-7. doi: 10.1097/INF.0000000000000707. PMID: 25831417;

[2] R. Burke RM, Midgley CM, Dratch A, et al. Active Monitoring of Persons Exposed to Patients with Confirmed COVID-19 — United States, January–February 2020. MMWR Morb Mortal Wkly Rep 2020;69:245–6. https://doi.org/10.15585/mmwr. mm6909e1.

[3] D. Hunter DJ. Covid-19 and the stiff upper lip—the pandemic response in the united kingdom. N Engl J Med 2020.

[4] Razai MS, Doerholt K, Ladhani S, Oakeshott P. Coronavirus disease 2019 (covid-19): a guide for UK GPs. BMJ. 2020 Mar 5;368:m800. doi: 10.1136/bmj.m800.

[5] P. Lillie PJ, Samson A, Li A, Adams K, Capstick R, Barlow GD, Easom N, Hamilton E, Moss PJ, Evans A, Ivan M, Phe Incident Team, Taha Y, Duncan CJA, Schmid ML, The Airborne Hcid Network. Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. J Infect. 2020 May;80(5):578-606. doi: 10.1016/j.jinf.2020.02.020. Epub 2020 Feb 28. PMID: 32119884; PMCID: PMC7127394.

[6] F. Ahmad, F., Jhajj, A. K., Stewart, D. E., Burghardt, M., & Bierman, A. S, "Single item measures of self-rated mental health: A scoping review". BMC Health Services Research, vol. 14, no. 1, pp. 398, 2014. https://doi.org/10.1186/1472-6963- 14-398.

[7] D. Ahorsu, D. K., Lin, C. Y., Imani, V., Saffari, M., Griffiths, M. D., & Pakpour, A. H, "The fear of COVID-19 Scale: Development and initial validation". International Journal of Mental Health and Addiction, vol. 20, pp. 1537–1545, 2022. https://doi.org/10.1007/s11469-020-00270-8

[8] J. Arbuckle, J. L. IBM SPSS AMOS 23 user's guide. IBM.

[9] I. Arpaci, I., Karatas¸, K., & Baloglu, M., "The development and initial tests for the psychometric properties of the COVID-19 Phobia Scale (C19P-S). Personality and Individual Differences, vol. 164, pp. 110108, 2020. https://doi.org/10. 1016/j.paid.2020110108.

[10] H. Ming H, Abuja N, Kriegman D: Face detection using mixtures of linear subspaces. Proceedings Fourth International Conference on Automatic Face and Gesture Recognition 2000, 4:70-76.

[11] O. Aguilar O, West M: Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economic Statistics 2000, 18:338-357.

[12] M. Reza Mahmoudi, Dumitru Baleanu, Shahab S. Band,Amir Mosavi, Factor analysis approach to classify COVID-19 datasets in several regions.

[13] M. Yıldırım & Abdurrahim Güler, "Factor analysis of the COVID-19 Perceived Risk Scale: A preliminary study", Death Studies, 2020.

[14] R. M, Blettner M, Klug SJ. Data analysis of epidemiological studies: part 11 of a series on evaluation of scientific publications. Dtsch Arztebl Int. Mar; vol. 107, no. 11, pp. 187-92, 2010 doi: 10.3238/arztebl.2010.0187. Epub 2010 Mar 19. PMID: 20386677; PMCID: PMC2853157.

[15] S. Jacob ST, Crozier I, Fischer WA 2nd, et al., "Ebola virus disease" Nat Rev Dis Primers. vol. 6, no. 1, pp. 13, 2020, DOI: 10.1038/s41572-020-0147-3.

[16] Z. Zhang, D. Li, X. Wang, Y. Wang, J. Lin, S. Jiang, Z. Wu, Y. He, X. Gao, Z. Zhu, Y. Xiao, Z. Qu, Y. Li, "Rapid detection of viruses: Based on silver nanoparticles modified with bromine ions and acetonitrile". Chemical Engineering Journal vol. 438, pp. 135589, 2022.

[17] Y. Luo Y, Wu J, Lu J, Xu X, Long W, Yan G, Tang M, Zou L, Xu D, Zhuo P, Si Q, Zheng X. Investigation of COVID-19-related symptoms based on factor analysis. Ann Palliat Med, vol. 9, no. 4, pp. 1851-1858, 2020, doi: 10.21037/apm-20-1113

[18] A. Cori A, Donnelly CA, Dorigatti I, et al. "Key data for outbreak evaluation: building on the Ebola experience", Philos Trans R Soc Lond B Biol Sci. vol. 372, no. 1721, pp. 20160371, 2017, doi:10.1098/rstb.2016.0371.

[19] H. Feldmann, H., Jones, S., Klenk, HD. et al. Ebola virus: from discovery to vaccine. Nat Rev Immunol, vol. 3, pp. 677–685, 2003, https://doi.org/10.1038/nri1154.

Venu Paritala has done her bachelor's degree from Vignan's Foundation for Science, Technology & Research (Deemed to be University) in 2021.

He has research experience in Toxicological studies, Data Analysis, Software Development, and Bioinformatics, He Has total of 4 publications in Springer, Journal of Microbial & Biochemical Technology and journal of Applied Bioinformatics & Computational Biology.

Harsha Thummala has been pursuing a bachelor's degree from Vignan's Foundation for Science, Technology & Research (Deemed to be University) since 2018.

His Research experience in Toxicology and Bioinformatics.

Measurement of Ambient Ionizing Radiation Exposure in Operating Consoles of Radiation Modalities in Cancer Hospital NORIN Nawabshah, Pakistan

Muhammad Waqar [1,*], Touqir Ahmad Afridi [1], Quratulain Soomro [2], Abdul Salam Abbasi [2], Muhammad Shahban [3]

[1] Medical Physics Division, Nuclear Medicine Oncology & Radiotherapy Institute Nawabshah (NORIN), 67450, Pakistan

[2] NM & Allied Division, Nuclear Medicine Oncology & Radiotherapy Institute Nawabshah (NORIN), 67450, Pakistan

[3] Medical Physics Division, Atomic Energy Cancer Hospital Islamabad (NORI), Islamabad, 44800, Pakistan

* Corresponding author: Muhammad Waqar (MS), NORIN Cancer Hospital, Nawabshah, Mob: +92-3355815274, Fax: +92-2449370556, phy.waqar@gmail.com

ABSTRACT: Elevated exposure from background radiations and health hazards for radiation workers have recently grabbed the attention of researchers. This study targeted to measure the background radiation levels in the operating console areas radiation facilities of NORIN, Nawabshah Pakistan. Ten operating consoles of different treatment and diagnostic machines were surveyed using a calibrated RM1001-RD LAMSE radiation monitor for the period of one year periodically and AEDR was calculated using standard formulas. The organ doses were calculated using recommended occupancy and conversion factors. The highest point with increased AEDR was found to be the operating console of cobalt-60 teletherapy machine (0.876 ± 0.03 mSv/yr), while the lowest at the Digital Radiography operating console (0.730 ± 0.03 mSv/yr). The standard error ranged between 0.02-0.03 %. These findings affirm a statistically significant difference in T-test values at a level of significance of 5% (P<0.05). The testes received the maximum dose (0.718 mSv/yr) followed by bone-marrow (0.604 mSv/yr) at Co-60 Teletherapy operating console. **Conclusion:** Based on these results, it was deduced that radiation levels are well within the permissible radiation limit of 1.0 mSv/yr prescribed by the ICRP and hardly about 37% of UNSCEAR limit of 2.4 mSv/yr. Therefore, all radiation workers are radiologically safe in operating console areas because all radiation protection and regulatory protocols are strictly observed in the working environment. This study eliminates the undue anxiety about the hazardous nature of radiation in the radiation workers of cancer hospitals.

KEYWORDS: Background Radiation, Equivalent Dose, Effective Dose, SSDL, Survey Monitor, Operating console, AEDR

1. Introduction

Radiation has a multifaceted nature; it is beneficial as well as dangerous for living beings. We come across numerous types of radiation having different intensities in daily activities. The common hazards of radiation include cancer, cataract, genetic mutation, degradation of skeletal bones and blood cells. If imparted in enough quantity, radiation dose can cause death of an individual as reported by [1]. As far as the background radiation is concerned, the major contributors to background radiation are the materials used for construction. They also cause the transfer of radionuclides into the environment, enhancing the background radiation levels. Radon gas, formed in earth crust, is the most abundant culprit of natural background radiation in the environment. Uranium-238 yields 222Rn after radioactive decay, which has a half-life of 3.82 days [2]. It's inhalation results in its absorption and penetration into the lung tissues. This absorption entails severe damage to lungs and also a mutation which consequently causes lung cancer [3]. Internationally recommended annual safe exposure limit is set by International Commission on Radiation Protection (ICRP) for Ionizing radiation. Its value by ICRP

is 1mSv/yr for general public as reported by [4]. On the other hand, the United Nation Scientific Committee on the Effects of Atomic Radiations (UNSCEAR) set the effective dose rate limit of 2.4mSv/yr [5]. These safe limits are suggested for indoor facilities including research labs, offices, conference rooms, lecture halls, ete. Many studies have been reported previously which show that areas with increased ambient radiation are found in Kerele, India; Yangjiang, China; and Ramsar, Iran as in [6]; and in Asia as well. Highest levels of outdoor background radiation were reported in Malaysia and the highest indoor levels were reported in Hong Kong and Iran by [7].

Along with natural background radiation, man-made radiation sources are also there. Among these artificial radiation sources, X-Radiation serves as the largest contributor to the world population. The effects of radiation have been divided broadly into two categories: the stochastic effects and non-stochastic or deterministic effects (IAEA 2009) [8]. In a study consisting of 114 documented cases of cancer, it is shown that the technical and medical staff were affected mostly. In another documented case, 359 radiologist expired because of skin and bone cancer induced by radiation [9]. The harmful effects of high level ionizing radiation have been ascertained but there is an ongoing debate about whether its low levels are beneficial or not.

The Nuclear Medicine Oncology and Radiotherapy Institute Nawabshah (NORIN) is a comprehensive healthcare facility for diagnosis, treatment, and research of malignant tumours [10]. It was established with the objective to adopt the latest research methodologies for cancer management. Radiotherapy, Radiology and Nuclear medicine are the main departments of a cancer hospital where all radiation modalities are used for diagnoses and treatment [11]. Special suites are designed for these modalities in which radiation protection is made sure though all aspects. Wall thicknesses, mazes and doors are designed in such a way that there is no leakage radiation outside the suite. Operating consoles, from where the technician operates and controls the function of machine, are the areas which are designed in a special manner for monitoring the machines. The radiation protection protocols are implied in such a way that the operator is safe from the hazards of radiation inside the console during operation of modality [12].

2. Methods and Materials

NORIN cancer hospital Nawabshah is situated in the rural area of Sindh, Pakistan. On average, this centre facilitates about 400 cancer patients per week. The three core departments of this hospital are radiotherapy, nuclear medicine and radiology. Ten operating consoles of radiation modality suites were selected for this study including operating consoles of CT-Scanner, DEXA,

Conventional Simulator, Brachytherapy, Co-60 Teletherapy, Mammography, X-Ray unit, Gamma Camera-I, Gamma Camera-II and Digital Radiography. The data was periodically collected over the span of one year and was analysed for calculation of AEDR.

The study was carried out by using RM1001-RD LAMSE survey meter calibrated from Secondary Standard Dosimetry Laboratory PINSTECH Islamabad as shown in Figure 2 . This model of survey meters is well-suited for the survey of background radiation in hospitals. The readings were taken during working hours with the workers performing their routine task. The equivalent dose readings were recorded in μSv/hr directly from the display screen of the radiation meter. The results were then converted into micro-Sievert per year (μSv/yr) and then finally converted to milli-Sievert per year (mSv/yr). An occupancy factor of 0.8 was used as recommended by the UNSCEAR (2000) [13]. The Annual Effective dose rate (AEDR) was calculated by using the following expressions.

$$\text{Annual Effective Dose Rate} \left(\frac{\mu Sv}{yr}\right) = X \left(\mu \frac{Sv}{hr}\right) \times T \times OF$$

$$\text{AEDR} \left(\frac{\mu Sv}{yr}\right) = X \left(\mu \frac{Sv}{hr}\right) \times 8760 \times 0.8 \times 10^{-03}$$

Where; X=hourly dose rate, T=total number of hours in a year (8760 hrs) and OF=occupancy factor (indoor = 0.8). Based on 24 hours a day and 365 days in a year; the number of hours in a year was 24 x 365 = 8760 hours [14]. AEDR is the total annual effective dose rate (mSv/yr).

Figure 1: RM1001-RD LAMSE survey meter used at NORIN, Nawabshah

Radiation doses to some body organs/tissues such as the lungs, ovaries, bone marrow, testes, kidneys, liver and whole body due to inhalation were computed using equation mentioned below.

$$\text{Organ Dose} \left(\frac{\mu Sv}{yr}\right) = \text{AEDR} \times CF$$

Where CF is the conversion factor of organ doses from air dose. The value of CF is 0.64 for the lungs, 0.69 for bone

marrow, 0.58 for ovaries, 0.62 for kidneys, 0.82 for testes, 0.46 for liver and 0.68 for whole-body [15]. For the statistical analysis of the data, independent T-test on SPSS 17 statistical software (SPSS Inc. USA) [16] was used and values at a level of significance of 5% ($P<0.05$).

Table 1: Annual effective dose rate values (mSv/yr) with standard deviation at different operating console areas at NORIN

Consoles	Mean (µSv/hr)	Mean (mSv/yr)	AEDR (mSv/yr)	P-Value (P < 0.05)
CT-scan	0.112	0.9782	0.785 ± 0.03	0.020
DEXA	0.110	0.913	0.771 ± 0.02	0.0101
Simulator	0.110	0.963	0.772 ± 0.19	0.0019
Brachytherapy	0.104	0.912	0.735 ± 0.03	0.0032
Teletherapy	0.125	1.095	0.876 ± 0.03	0.0120
Mammography	0.110	0.963	0.773 ± 0.03	0.0194
X-ray Unit	0.113	0.985	0.788 ± 0.019	0.0044
Gamma Camera-I	0.114	0.993	0.794 ± 0.024	0.0170
Gamma Camera-II	0.109	0.956	0.765 ± 0.025	0.007
Digital Radiography	0.104	0.922	0.730 ± 0.03	0.0107

3. Results and Discussion

The 10 operating console of radiation modalities are required to assess the radiation risks associated with the scanning examinations and treatment of patients and worker in cancer Hospital Nawabshah, Pakistan. The results are presented in Tables 1 with mean values and AEDR (mSv/yr) and standard error with P values. Results are lower due to the strict following radiation protection protocols and all machines suites & operating Consoles rooms are prepared with PNRA & IAEA guidelines. Moreover, radiological burden of the hospital is within the permissible value of UNSCEAR in most of departments in hospital.

Figure 2 graphically represents the values of AEDR on ten selected operating console radiation suites and their comparison with globally accepted dose limits of ICRP and UNSCEAR. Each bar is associated with a suite and AEDR values are shown respectively. The maximum dose rate was shown at the console of Teletherapy machine (Cobalt-60) with value of 0.876 ± 0.03 mSv/yr. While the minimum does rate was observed in the operating console of the radiography room with AEDR of 0.730 ± 0.02 mSv/yr. It is clear from the graph that even the highest AEDR value (0.876 mSv/yr) is not more than that of annual dose limit of general public. While the annual background dose limit for a radiation worker is much higher with the value of 2.4 mSv/yr as shown the Figure 2 . Also, there is no area in the study where AEDR value goes above 1mSv/yr.
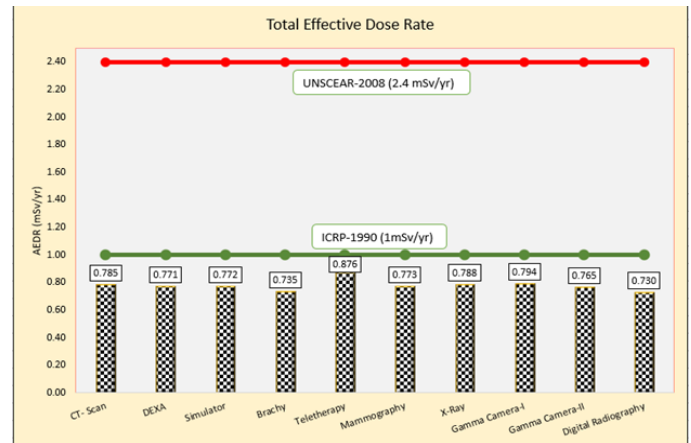

Figure 2: AEDR in different stations with respect to ICRP & UNSCEAR

Organ/Tissue Specific radiation doses (lungs, ovaries, bone marrow, testes, kidneys, liver and whole body) to the radiation workers working in the operating consoles of various radiation modalities due to inhalation were computed and are graphically represented in Figure 3. It was observed that the most vulnerable organ prone to receiving maximum AEDR were testes with the highest mean AEDR of 0.718 mSv/yr at Cobalt-60 teletherapy console and minimum at CT-scan console. The lungs would receive at most (0.561 mSv/yr), whole body (0.596 mSv/yr), Ovaries (0.508 mSv/yr) Bone Marrow (0.604 mSv/yr) and kidneys (0.543 mSv/yr) recorded at Co-60 Teletherapy machine. All AEDRs were found well below the ICRP recommended limit of 01mSv/yr.

The Results of this study are lower compared with previous work. Some of the previous work is graphically summarized in Figure 4. In [17], the author reported 1.22 mSv/yr in a private medical diagnostic centre at Nigeria. A team of researchers found higher values of background ionizing radiation of 2.11 mSv/yr in some hospitals in Jos Plateau state, Nigeria [18]. Another study shows slightly higher values of 1.54 mSv/yr in the experimental labs of Plateau state university, Nigeria [2]. Pharmaceuticals facilities in Nigeria had dose rate of 1.60 mSv/yr [19]. In 2015, authors investigated the indoor background radiation levels at Plateau University Bookos and reported 1.54 mSv/yr [2]. Mean Indoor Annual Effective dose rate in the radiation labortries of Federal University KATSINA state, Nigeria documented 1.41 mSv/yr [9]. while the ICRP (1990) and UNSCEAR (2008) recommended annual dose level are 1.0 mSv/yr and 2.4 mSv/yr respectively [15]. However, a study carried out in Nigeria (2021) reported lower reading 0.80 mSv/yr than the results obtained from current study, while [20] reported the same (0.88 mSv/yr) ambient ionizing radiation level in Keffi General Hospital, Nigeria. This could be due to variations of equipment and office where no radiations are used.
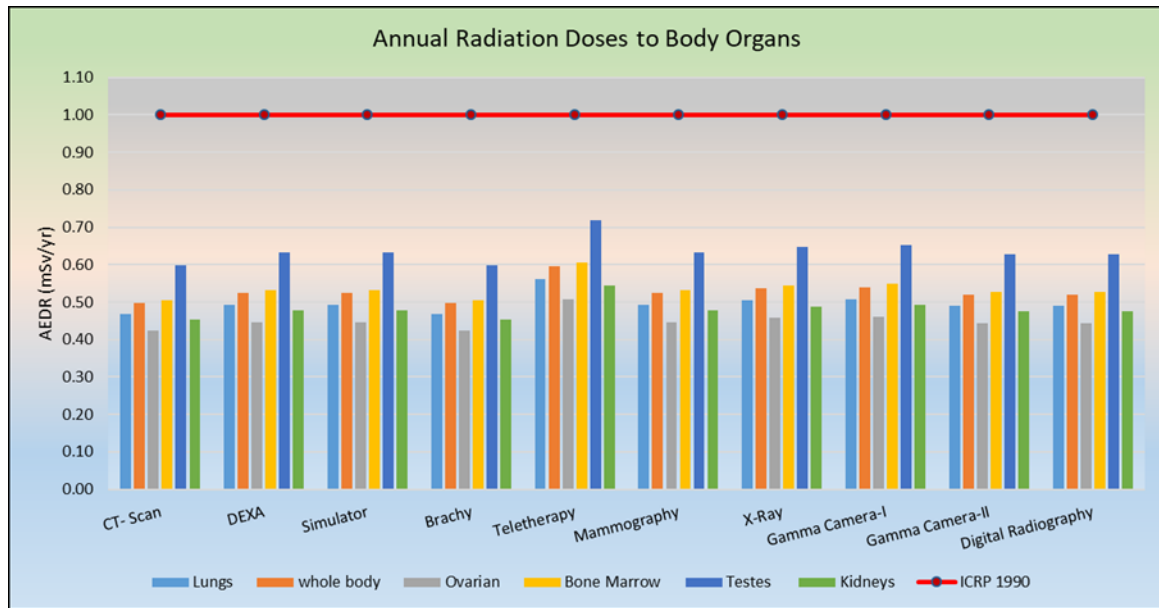
Figure 3: Graphical Representation of Organ-wise Annual effective dose rate at different operating console at NORIN.
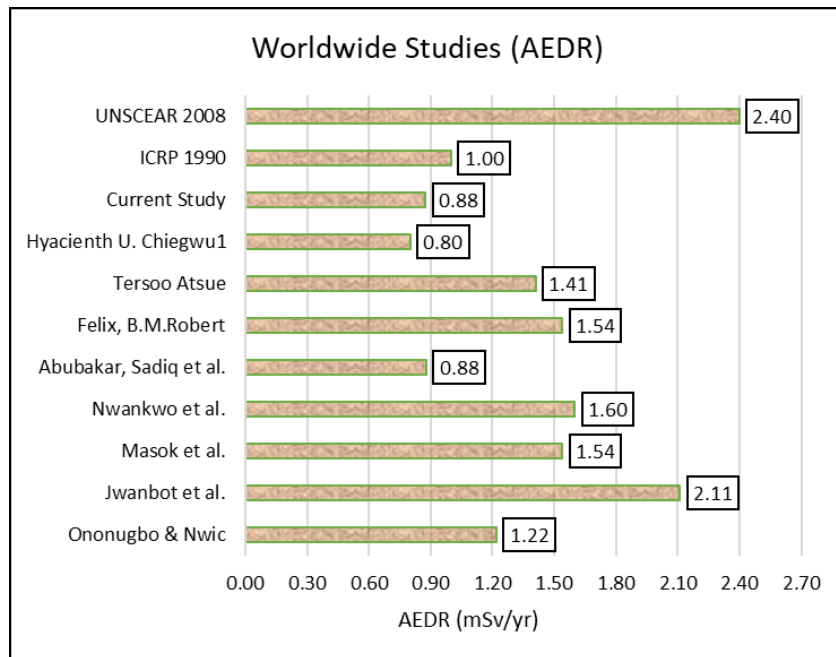

Figure 4: Comparison of current study with previously conducted studies in the world

The mean AEDR for the studied control rooms is comparable with the ICRP recommended annual limit of 1.0 mSv/yr for the general public but well below the UNSCEAR recommended world average value of 2.4 mSv/yr. This implies that the workers and people in those offices are radiologically safe.

Better knowledge about radiation due to exposure from patients is important for deciding on reasonable and appropriate precautions against unnecessary radiation exposure for radiation employees working in operating console and next-of-kin. Our results reaffirm that "undue anxiety among hospital staff with regard to exposure to radioactive patients must be placed in the proper perspective through education and training". However, we observed that significant anxiety was still present.

Perhaps particularly amongst staff not directly involved in hospital, but still in contact with the patients in other departments [21].

4. Conclusion

The findings of this study showed that there were no significant health hazards for radiation workers. The background exposure rate in different operating consoles of radiation modalities in this cancer hospital was well under the internationally recognized limits. Based on the aforementioned findings, it was deduced that radiation levels are within the permissible radiation limit as stipulated by the ICRP is 1.0 mSv/yr and UNSCEAR of 2.4 mSv/yr. The highest AEDR recorded in this study was in the console of Co-60 teletherapy machine which was 0.875

mSv/yr. Usually this area is considered as the area with the maximum ambient radiation levels in cancer hospitals and yet the AEDR calculated here was hardly 37% of the UNSCEAR recommended limit. This indicates that all radiation protection protocols are followed strictly as per regulatory guidelines. This study eliminates the undue anxiety about the hazardous nature of radiation in the radiation workers of cancer hospitals. Hence, all radiation workers in NORIN are radiologically safe in console areas and also, this facility does not elevate the radiation levels in the surrounding. Public health around the center is not on stake and there is no significant impact on the radiation burden of the environment.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] P. E. Ogola, W. M. Arika, D. W. Nyamai, K. O. Osano, and H. O. Rachuonyo, "Determination of background ionizing radiations in selected buildings in Nairobi County, Kenya," *J Nucl. Med. Radiat. Ther*, vol. 7, no. 1, p. 289, 2016.

[2] F. B. Masok, R. R. Dawam, and E. W. Mangset, "Assessment of indoor and outdoor background radiation levels in Plateau State University Bokkos Jos, Nigeria," 2015.

[3] C.-U. E. Yehuwdah, A. Martins, and H. O. Soibi, "Evaluation of indoor background ionizing radiation profile of a physics laboratory," *Facta Univ. Work. Living Environ. Prot.*, vol. 3, no. 1, pp. 1–8, 2006.

[4] J. T. Baraya, M. H. Sani, and M. Alhassan, "Assessment of indoor and outdoor background radiation levels at school of technology, Kano State Polytechnic, Kano State-Nigeria," *J. Appl. Sci. Environ. Manag.*, vol. 23, no. 3, pp. 569–574, 2019.

[5] U. N. S. C. on the E. of A. Radiation, "Sources and effects of ionizing radiation, ANNEX B, Exposures from natural radiation sources," *UNSCEAR 2000 REPORT, New York*, vol. 1, pp. 97–99, 2000.

[6] S. S. Althoyaib and A. El-Taher, "The measurement of radon and radium concentrations in well water from Al-Jawaa, Saudi Arabia," *J Radioanal Nucl Chem*, vol. 304, pp. 547–552, 2015.

[7] F. Alshahri, A. El-Taher, and E. A. Elzain, "Measurement of radon exhalation rate and annual effective dose from marine sediments, ras tanura, Saudi Arabia, using Cr-39 detectors," *Rom. J. Phys.*, vol. 64, p. 811, 2019.

[8] IAEA. and IAEA., *Security of Radioactive Sources*. International Atomic Energy Agency, 2009.

[9] T. Atsue and J. Adegboyega, "Assessment of the Ambient Background Radiation Levels at the Take-Off Campus of Federal University Dutsin-Ma, Katsina State-Nigeria," *FUDMA J. Sci. Maiden Ed.*, vol. 1, pp. 58–68, 2017.

[10] M. Waqar, M. Shahban, Q. Soomro, and M. N. Abro, "Institution-based assessment of cancer patients treated by external beam radiotherapy in the rural area of Sindh, Pakistan: Five years of data analysis," *Middle East J. Cancer*, vol. 9, no. 3, pp. 217–222, 2018.

[11] M. Waqar, M. Shahban, Q. Soomro, T. A. Afridi, and N. Med, "Assessment of residual activities of Technetium-99m radiopharmaceuticals to plastic syringes in nuclear medicine scans : 1-year experience at NORIN," vol. 12, no. 1, pp. 1–6.

[12] A. Razaq, M. Waqar, Q. Soomro, and M. A. Javed, "Evaluation of radiation workers occupational doses for newly established medical center NORIN Nawabshah in Pakistan," *J Med Phys Biophys*, vol. 3, no. 1, pp. 61–65, 2016.

[13] N. Y. United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) NY (United States), *Sources and effects of ionizing radiation UNSCEAR 2000 report to the General Assembly, with scientific annexes Volume I: Sources*. United Nations (UN): UN, 2000.

[14] U. Rilwan, A. U. Maisalatee, and M. Jafar, "Assessment of Indoor and Outdoor Radiation Exposure in Nasarawa General Hospital, Nasarawa State, Nigeria," *J. Radiat. Nucl. Appl. An Int. J.*, vol. 6, no. 3, pp. 245–249, 2021.

[15] H. U. Chiegwu *et al.*, "Assessment of background ionizing radiation exposure levels in industrial buildings in Nnewi, Anambra State, Nigeria," *Int. J. Res. Med. Sci.*, vol. 10, no. 2, p. 305, 2022.

[16] M. Waqar, M. Shahban, Q. Soomro, and M. E. Khan, "Performance of Tc-99m Generators in Terms of Molybdenum Contents Measured in NORIN Nawabshah, Pakistan (5 years' Experience)," *Iran. J. Sci. Technol. Trans. A Sci.*, vol. 43, no. 1, pp. 285–289, 2019.

[17] C. P. Ononugbo and A. A. Nwii, "Indoor and Outdoor Effective doses from Background Ionizing Radiation in Private Medical Diagnostic Centers in Bori, Rivers State," *J. Environ. earth Sci.*, vol. 5, pp. 25–29, 2015.

[18] D. I. Jwanbot, M. M. Izam, G. G. Nyam, and I. S. Agada, "Evaluation of Indoor Background Ionizing Radiation Profile in Some Hospitals in Jos, Plateau State-Nigeria," 2012.

[19] L. I. Nwankwo, D. F. Adeoti, and A. J. Folarin, "Ionizing radiation measurements and assay of corresponding dose rate around bottling and pharmaceutical facilities in Ilorin, Nigeria," *J. Sci. Technol.*, vol. 34, no. 2, pp. 84–92, 2014.

[20] A. Abubakar, A. A. Sadiq, M. G. Musa, J. Hassan, and D. F. Malgwi, "Assessment of indoor ionizing radiation profile in radiology department FMC Asaba Delta State Nigeria," *Internafional Organ. Scientific Res. J. Dent. Med. Sci. 16 98*, vol. 101, 2017.

[21] S. Harb, "Evaluation of radiation doses and radiation risk in Teaching Sohag Hospital, Egypt," *J. Nucl. Part. Phys.*, vol. 6, no. 4, pp. 88–93, 2016.

Muhammad Waqar Qureshi is Head of Medical Physics Division and Radiation Protection Officer at NORIN Cancer hospital, Nawabshah, Pakistan. Hailing from TandoAllahyar, Rural area of Pakistan, the author received bachelor degree in Health Physics from University of Sindh and secured Chancellor Gold Medal & Vice Chancellor Silver Medal. Afterwards, author received Master's Degree in Medical Physics from Pakistan Institute of Engineering & Applied Sciences Islamabad.

He has published 06 Research Articles in different international journals of good repute. His main research areas include Radiotherapy, Dosimetry, Nuclear Medicine & Radiation Protection. He has also attended numerous workshops, Trainings & Seminars at National & International level as participant as well as guest speaker.

Mr. Waqar is Publication Secretary of Pakistan Society of Nuclear Medicine (PSNM) and Life-time member of Radiological Society of Pakistan. Mr. Waqar has membership of many revered societies including Radiation Oncology Society of Pakistan (ROSP), Pakistan Organization of Medical Physics (POMP), Pakistan Nuclear Society (PNS). At domestic level, he is Executive council committee of Cancer Patients Welfare Society Nawabshah.

Touqir Ahmad Afridi is a Medical Physicist at NORIN Nawabshah. He received his MSc Physics degree from University of Lahore and MS (Medical Physics) from PIEAS Islamabad. He has authored 01 research article as first author and co-authored 03 national and international publications.

Dr. Quratulain Soomro is a Medical Oncologist and Director of NORIN cancer hospital Nawabshah. She received her MBBS degree from Chandka Medical College Larkana and DMRT from Sheikh Zaid Hospital Lahore. She is the first author in 02 research articles and co-author in 07 international research publications. She has more than 30 years' experience in the field of Medical Oncology.

Dr. Abdul Salam Abbasi is radiologist of NORIN cancer hospital Nawabshah. He received his MBBS degree from Chandka Medical College Larkana and DMRD from PUHMSW, Nawabshah. He has more than 07 years' experience in the field of Radiology.

Muhammad Shahban is a Medical Physicist at NORI cancer hospital, Islamabad. He completed his MSc (Physics) from Quaid-e-Azam University Islamabad and MS in Medical Physics from PIEAS. He is the first author in 03 research articles and co-author in 07 international research publications. He has more than 07 years' experience in the field of Radiation Dosimetry.

Fuzzy Matrix Theory based Decision Making for Machine Learning

Javaid Ahmad Shah *

Department of Mathematics, Dr. A. P. J. Abdul Kalam University Indore (M.P) India

* Corresponding author: Javaid Ahmad Shah, Email: Jsjavaidjs@gmail.com
ORCID: 0000-0002-6234-3231

ABSTRACT: The Fuzzy set theory has numerous real-life applications in almost every field like artificial intelligence, pattern recognition, medical diagnosis etc. There are so many techniques used for solving decision-making problems given by various researchers from time to time. To be able to make consistent and correct choices is the essence of any decision process pervade with uncertainty. Fuzzy matrix theory plays an important role in scientific development under uncertain conditions. Nowadays there are huge varieties of mobile phones with varying features available in the market. Everyone wants to purchase such a mobile phone which has as many features as possible within it but under his/her budget. This has become an important issue in this modern era where everyone wants to have the most preferred mobile handsets for himself/herself as compared to others. So, in this paper, the Fuzzy matrix approach is used in a decision-making problem where a number of buyers can be able to choose their preferred mobile phones with varying features.

KEYWORDS: Decision Making, Fuzzy Logic Technique, Fuzzy Sets, Max-Min Composition, Fuzzy Matrix Multiplication

1. Introduction

Fuzzy set theory tries to follow more closely the vagueness that is inherent in most natural language and in decision-making processes. In a conventional logic approach, this inherent fuzziness of membership and categorization is not included. Fuzzy logic has found many real-world applications that involve imitating or modeling human behavior for decision making in the real world. The dynamics and complexity of social systems are being explained and modeled through the use of fuzzy set theory. In recent years, computational intelligence has been used to solve many complex problems by developing intelligent systems. Fuzzy logic has proved to be a powerful tool for decision making systems, such as expert systems and pattern classification systems [1, 2]. In [3], the author first introduced the concept of fuzzy matrices, and discussed the convergence of powers of fuzzy matrix. Matrices whose matrix operation defined by fuzzy logical operations and whose entries lies in [0, 1] are called fuzzy matrices. Every fuzzy matrix is a matrix but not conversely. Fuzzy matrices play a fundamental role in fuzzy set theory. They provide us a rich framework within which many problems of practical applications of the theory can be formulated. Fuzzy matrices can be successfully used when fuzzy uncertainty occurs in a problem. These results are extensively used for cluster analysis and classification problems of static patterns under subjective measure of similarity [4]. The theories of fuzzy matrices were developed in [5] as an extension of Boolean matrices. With max-min operation the fuzzy algebra and its matrix theory are considered by many authors [6-11]. In [12] the author studied the canonical form of a transitive fuzzy matrix. In [13], the author presented the canonical form of a strongly transitive matrix. In [14], the author studied the controllable fuzzy matrix. In [15], the author presented some properties of the min-max composition of fuzzy matrices. In [16], the author presented some important results on determinant of square fuzzy matrices. In [17], the author investigated iterates of fuzzy circulant matrices. Determinant theory, powers and nilpotent conditions of matrices over a distributive lattice are considered in [18, 19]. As for as the decision making is concern, it is very important in terms of machine learning or imparting artificial intelligence into machines which work upon the traditional logic theory. It is a process which helps a machine to think like a human being and

for a human being to ease his/her difficult decision making process. In this paper fuzzy matrices and fuzzy logic techniques are utilized in order to ease the process of decision making.

2. Review of literature

Lotfi A. Zadeh in 1965 first introduced the concept fuzzy sets in one of his research papers under the heading fuzzy logic or Fuzzy Sets, where he proposed that some ideas either belongs to or does not belongs to in the area of general consideration which disproves the traditional assumptions [20, 21]. Various researchers have used fuzzy set theory to prove real life decision making problems having uncertain information. Fuzzy matrices introduced by M. G. Thomason have become popular in the last two decades. Fuzzy matrix which is an extension of classical matrix with entries lies between 0 and 1. Numerous scholars have proposed new methods in fuzzy matrix and these methods lead fuzzy matrix a new area of fuzzy mathematics in [22] the authors gave a brief concept of Basic matrix theory and fuzzy matrix theory in their book "Elementary Fuzzy matrix theory and fuzzy models for social scientists". Madhumangal Pal introduced two new binary fuzzy operators, $\oplus$ and $\odot$, for fuzzy matrices, he presented several properties on these new binary operations. He also proposed some results on existing operators along with these new operators. In [23] the author discussed Fuzzy Matrix Theory and applied it for recognizing the Qualities of Effective Teacher. They have used FRM (Fuzzy relational maps) in their research to determine the components of teacher quality and to apply the Teacher Quality Index (TQI) for recognizing the qualities of effective teacher, as well as for the developments of educational institutes. In [24] the author describes Fuzzy Symmetric Solutions of Fuzzy Matrix equations by using system of linear equations according to the Kronecker product of matrices. Fuzzy relations are fuzzy subsets of the Cartesian product of two universes of discourse say X and Y these relations are combined with each other by the operation called Composition operation. Various composition operations (methods) available are max-min operation, max-product operation, Min-max composition operation, max-max composition operation, min-min composition operation, max-average composition operation and sum-product composition operation. But max-min composition method is best known in fuzzy logic applications.

3. Preliminaries

3.1. Fuzzy matrix

A matrix $A = [a_{ij}]_{mxn}$ is said to be a fuzzy matrix of order mn, where $a_{ij} \in [0, 1]$, $1 \le i \le m$ and $1 \le j \le n$. For example

$$A = \begin{pmatrix} 1 & 0.4 & 0.2 & 0.1 & 0.9 \\ 0.7 & 0 & 0.7 & 0.9 & 0 \\ 0.3 & 0.4 & 0.8 & 0.1 & 0.5 \\ 0.5 & 0.3 & 0.2 & 0.3 & 0.4 \end{pmatrix} \text{ is a } 4 \times 5 \text{ fuzzy}$$

matrix.

3.2. Fuzzy composition operation

Suppose $\tilde{R}$ be a fuzzy relation on the Cartesian product space X×Y, $\tilde{S}$ be a fuzzy relation on space Y× Z and $\tilde{T}$ be a fuzzy relation on space X× Z, then fuzzy max-min and max- product is defined in terms of set and membership function notation as [1]:

$$\tilde{T} = \tilde{R} * \tilde{S}$$

$$\mu_{\tilde{T}}(x, z) = \underset{y \in Y}{\max} \{\min (\mu_{\tilde{R}}(x, y), \mu_{\tilde{S}}(y, z))\}$$

$$= \max \{(\mu_{\tilde{R}}(x, y)\}, \text{ if } \mu_{\tilde{R}}(x, y) < \mu_{\tilde{S}}(y, z)$$

$$= \mu_{\tilde{R}}(x, y).$$

This operation is one of the common and best among all the composition operations.

3.3. Multiplication of Fuzzy Matrices s

Let $A = [a_{ij}]_{mxn}$, where $a_{ij} \in [0, 1]$, $1 \le i \le m$ and $1 \le j \le n$, and $B = [b_{jk}]_{nxp}$, where $b_{jk} \in [0, 1]$, $1 \le j \le n$ and $1 \le k \le p$, be two fuzzy matrices then $AB = [C_{ik}]_{mxp}$ is also a fuzzy matrix with respect to the operations defined by max-min operation or min-max operation.

For example, $A = \begin{pmatrix} 0 & 0.9 & 0.7 \\ 0.5 & 0.6 & 0.1 \\ 0.8 & 0.4 & 0.9 \\ 0.3 & 0.7 & 0.2 \\ 0.4 & 1 & 0.3 \end{pmatrix}$ be a 5× 3 fuzzy

matrix.

$B = \begin{pmatrix} 1 & 0.2 & 0.4 & 0.5 \\ 0.4 & 0.3 & 0.8 & 0 \\ 0 & 0.2 & 0.6 & 0.7 \end{pmatrix}$ be a 3× 4 fuzzy matrix.

Then AB is defined using max-min operation as

C = AB = $[C_{ik}]_{5x4}$

where C_{11} = max {min (0, 1), min (0.9, 0.4), min (0.7, 0)} = max {0, 0.4, 0} = 0.9

C_{12} = max {min (0, 0.2), min (0.9, 0.3), min (0.7, 0.2)} = max {0, 0.3, 0.2} = 0.3

C_{13} = max {min (0, 0.4), min (0.9, 0.8), min (0.7, 0.6)} = max {0, 0.8, 0.6)} = 0.8

And so on.

$$\text{Thus } C = \begin{pmatrix} 0.9 & 0.3 & 0.8 & 0.7 \\ 0.5 & 0.1 & 0.6 & 0.5 \\ 0.8 & 0.2 & 0.4 & 0.7 \\ 0.4 & 0.3 & 0.7 & 0.3 \\ 0.4 & 0.3 & 0.8 & 0.4 \end{pmatrix}$$

4. Decision making by using multiplication of fuzzy matrix

Consider a case in which a person needs to decide related to purchasing a mobile phone with specific highlights that he fundamentally needs among a list of numerous such things with discernable yet similarly contending features and henceforth he/she faces a difficulty in the decision-making that which one should he/she really purchase. In such case, a fuzzy matrix multiplication approach based on max-min operation is utilized for such a decision making problem. In spite of a commitment in machine learning, it can be used to rank the multi-dimensional options of an option set. It is frequently used as a part of engineering for settling on design choices, however, can likewise be used to rank speculation alternatives, seller choices, item choices or some other set of multidimensional substances. The proposed decision-making process in view of a fuzzy matrix approach is described as follows.

Let us suppose that B = {b_1, b_2…b_n} be the set of buyers and M = {m_1, m_2…m_k} be the set of alternatives and each alternative is associated with a set of features defined as F = {f_1, f_2…f_p}. The main aim is to provide buyer the best option according to his/her choice. Furthermore the two matrices, BF (buyer features) and FM (mobile features) are constructed. The max-min composition operation is utilized in matrix multiplication method in order to find the BM (buyer mobile) matrix in which the highest value corresponding to the row gives the most appropriate alternative (choice) to the buyer.

4.1. Construction of BF and FM matrices for Decision Making

4.1.1. The BF Matrix:

In order to construct the BF matrix the survey is conducted by asking different questions to the buyers about their choices in terms of features, and converted their choices into membership values, also consider some mobile phones with varying features. Furthermore organize the data collected from the survey into membership valued matrices. Also inquiries have been made to the purchaser about his decision regarding features of the mobiles and converted their choice into membership values using fuzzy logic. Thus, BF matrix is arranged containing buyers along rows and feature choices along columns as shown below.

$$BF = \begin{pmatrix} & f1 & f2 & \dots & fp \\ b1 & . & . & . & . \\ b2 & . & . & . & . \\ . & . & . & . & . \\ bn & . & . & . & . \end{pmatrix}$$

4.1.2. The FM Matrix

Secondly, reviews about the most common reference mobile phones with varying features are collected, such that each reference mobile phone in terms of its features in membership valued sets are ranked, thus arranging the FM matrix in such a way that the reviewed features are taken along rows and reference mobile phones are taken along columns as shown below.

$$FM = \begin{pmatrix} & m1 & m2 & \dots & mk \\ f1 & . & . & . & . \\ f2 & . & . & . & . \\ . & . & . & . & . \\ fp & . & . & . & . \end{pmatrix}$$

4.1.3. The BM Matrix

Finally, matrix multiplication based on max-min composition operation is performed between matrix BF and matrix FM to get the resultant matrix BM as shown below.

$$BM = \begin{pmatrix} & m1 & m2 & \dots & mk \\ b1 & . & . & . & . \\ b2 & . & . & . & . \\ . & . & . & . & . \\ bn & . & . & . & . \end{pmatrix}$$

We will follow the technique that the maximum value along the row corresponding to the buyer in the multiplication matrix gives most appropriate result (mobile phone). Note that the similar questionnaire was accommodated for the overview survey as well as for the purchaser.

Table 1: Mobile Phones

S.No	M	Mobiles Phones
1	m_1	Samsung
2	m_2	Apple
3	m_3	Google Pixel
4	m_4	Xiaomi
5	m_5	Oppo
6	m_6	One Plus
7	m_7	Motorola
8	m_8	Sony
9	m_9	Lenovo
10	m_{10}	Realme

Table 2: Features

S.No	M	Mobiles Phones
1	f_1	Battery Life
2	f_2	Memory (RAM-ROM)
3	f_3	Design and Build quality

4	f_4	Screen
5	f_5	Camera
6	f_6	Processor Power
7	f_7	Price
8	f_8	Operating System

5. Analysis and Results

A The survey is conducted on 8 measure features of mobile phones such as Battery life, Memory (Ram-Rom), Design and builds quality, Screen, Camera, Processor Power, Price and Operating system. The examples of mobile phones considered for this survey are Samsung, Apple, Google, Realme, Oppo, One plus, Motorola, Sony, Lenovo, and Huawei. Depending upon how much preference is to be given to each asset, in a percentage score from 0 to 100(basically in a range). Data will then be modified from a scale of 0-100 to a scale of 0-1, to avoid any inconvenience regarding multiplication of two matrices, and such that the whole operation results into a binary operation. To the test the model we have considered 10 buyers represented as B = {b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}}, 10 mobile phones represented as M = {m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8, m_9, m_{10}} and 8 features of mobile phones represented as F= {f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8}. Table 1 and 2 shows mobile phones and their features considered for the survey respectively.

5.1 Questionnaire used to conduct the survey

A questionnaire is designed which is used to conduct the survey among buyers before decision making and the same is used to conduct a survey among the reviewers which can be shown in table 3.

5.2 Analyzing the survey

In the survey we had asked questions to the buyers about their choices in terms of features, and converted the buyer's choice into values between 0-1, and then we have taken some mobile phones as references, now what we have done is that we organized the data collected through survey into 0-1 valued matrices. First, BF matrix containing buyers in rows and feature choices in columns, further we have collected reviews about the most common reference mobile phones, such that we can rank each reference mobile phone in terms of its features in 0-1 valued sets, thus arranging the 2nd matrix FM as reviewed features in rows and reference mobile phones in columns, and finally, we have performed fuzzy matrix multiplication method to get the matrix BM.

A relation matrix between buyers and features BF is created as:

$$BF = \begin{pmatrix} 0.7 & 0.2 & 0.6 & 0.8 & 0 & 0.1 & 0.4 & 0.5 \\ 0.2 & 0.8 & 0.3 & 0.4 & 0.1 & 0.6 & 0.3 & 0.9 \\ 0.5 & 0.6 & 0.4 & 0.6 & 0.9 & 0.4 & 0.5 & 0.8 \\ 0.6 & 0.9 & 0.9 & 0.5 & 0.3 & 0.2 & 0.4 & 0.6 \\ 0.3 & 0.8 & 0.7 & 0.2 & 0.8 & 0.9 & 0.6 & 0.2 \\ 0.6 & 0.1 & 0.8 & 0.4 & 0.2 & 0.1 & 0.8 & 0.4 \\ 0.1 & 0.2 & 0.6 & 0.3 & 0.5 & 0.7 & 0.9 & 0.9 \\ 0.7 & 0.3 & 0 & 0.5 & 0.6 & 0.9 & 0.4 & 0.2 \\ 0.9 & 0.4 & 0.7 & 0.1 & 0.5 & 0.4 & 0.3 & 1 \\ 0.6 & 0.5 & 0.4 & 0.8 & 0.6 & 0.3 & 0.1 & 0.9 \end{pmatrix}$$

Further a relation matrix between features and mobile phones is created as:

$$FM = \begin{pmatrix} 0.4 & 0 & 0.4 & 0.1 & 0.1 & 0.9 & 1 & 0.9 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.3 & 0.3 & 0.3 & 0.7 & 1 & 0.3 & 0.2 & 0.2 \\ 1 & 0.9 & 0.6 & 0.2 & 0.2 & 0.4 & 0.3 & 0.4 & 0.4 & 0.9 \\ 0.3 & 0.8 & 0.7 & 0.1 & 0.2 & 0.5 & 0.2 & 0.1 & 0.5 & 0.7 \\ 0.9 & 0.3 & 0.5 & 0.9 & 0.4 & 0.7 & 0.7 & 0.2 & 0.9 & 0.5 \\ 0.5 & 0.7 & 0.9 & 0.8 & 0.1 & 0.9 & 0.5 & 0.4 & 0.3 & 0.2 \\ 0.4 & 0.6 & 0.3 & 0.5 & 0.9 & 0.2 & 0.1 & 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.7 & 0.6 & 0.5 & 0.4 & 0.9 & 0 & 1 & 0.4 \end{pmatrix}$$

The resultant multiplication matrix BM is obtained using max-min operation as:

$$BM = \begin{pmatrix} 0.6 & 0.8 & 0.7 & 0.5 & 0.5 & 0.7 & 0.7 & 0.7 & 0.5 & 0.7 \\ 0.5 & 0.6 & 0.7 & 0.6 & 0.5 & 0.7 & 0.9 & 0.4 & 0.9 & 0.4 \\ 0.4 & 0.6 & 0.7 & 0.6 & 0.5 & 0.7 & 0.7 & 0.5 & 0.9 & 0.6 \\ 0.9 & 0.5 & 0.6 & 0.6 & 0.5 & 0.7 & 0.9 & 0.6 & 0.9 & 0.6 \\ 0.8 & 0.7 & 0.9 & 0.8 & 0.6 & 0.9 & 0.8 & 0.5 & 0.8 & 0.7 \\ 0.8 & 0.8 & 0.6 & 0.5 & 0.8 & 0.6 & 0.6 & 0.6 & 0.4 & 0.8 \\ 0.6 & 0.7 & 0.7 & 0.7 & 0.9 & 0.7 & 0.9 & 0.5 & 0.9 & 0.6 \\ 0.6 & 0.7 & 0.9 & 0.8 & 0.4 & 0.9 & 0.7 & 0.7 & 0.6 & 0.5 \\ 0.7 & 0.7 & 0.7 & 0.6 & 0.5 & 0.9 & 0.9 & 0.9 & 1 & 0.7 \\ 0.6 & 0.8 & 0.7 & 0.6 & 0.5 & 0.6 & 0.9 & 0.6 & 0.9 & 0.7 \end{pmatrix}$$

From above relation matrix BM we conclude that the best choice of buyer b_1 is mobile m_2. Also best choice of buyer b_2 is m_7 and m_9 mobile phones. This means that buyer b_1 prefers apple than the other mobile phones and buyer b_2 prefers Motorola and Lenovo mobiles phones.

5.3 Survey

Below mentioned in table 4 is the survey which is obtained by asking questions to different persons in the Questionnaire.

6. Conclusion

In this paper we have used fuzzy matrix multiplication method for the purpose of decision making. We have considered an example of decisions making process to buy a mobile phone among a list of mobile phones and which are evaluated by eight features. Also we have conducted a survey among the buyers and reviewers using a same questionnaire. The application of fuzzy matrix multiplication approach is utilized here to construct a BF and FM matrices which is further utilized to analyze the preference of buyers among alternative mobile phones.

Table 3: Questionnaire

Questionnaire

Name

Address

Tick the mobile phones do you own:

1. Samsung…….   2. Apple…….   3. Google pixel……….   4. Xiaomi……...   5. Oppo……….

6. One plus…….   7. Motorola……   8. Sony………   9. Lenovo………   10. Realme…….   11. Other……...

    1.   Price of mobile phone (thousands):

(a) 1 – 10 ……   (b) 10 – 20 …….   (c) 20 – 30 …..……   (d) 30 – 40 ………   (e) 40 – 50…….   (f) 50 - 60…….   (g) 60 - 70……...

    2.   Battery life (months):

(a) 6 – 12 …..…..   (b) 12 – 18 …..…...   (c) 18 – 24 …..…...   (d) 24 – 30 …..…...   (e) 30- 36 ……..   (f) More …..…...

    3.   Memory:

(a) 0 – 20 ………..   (b) 20 – 40 …..…...   (c) 40 – 60 …..…...   (d) 60 – 80 ……..   (e) 80 - 100 ……….

    4.   Design and build quality:

(a) 0 – 2 …..…..   (b) 2 – 4 …..…...   (c) 4 – 6……..   (d) 6 – 8……...   (e) 8 - 10 …………….

    5.   Screen:

(a) 0 – 20…..   (b) 20 – 40…..   (c) 40 – 60……   (d) 60 – 80……

    6.   Camera:

(a) 0 – 20…..   (b) 20 – 40……   (c) 40 – 60……   (d) 60 – 80……   (e) 80 - 100 ………

    7.   Processor Power:

(a) 0 – 5……   (b) 5 – 10……...   (c) 10 – 15……...   (d) 15 – 20……   (e) 20 – 25……...

    8.   Operating System:

(a) 0 – 20 ……….   (b) 20 – 40 …..……   (c) 40 – 60 ………….   (d) 60 – 80 …..……   (e) 80 – 100 ….....……

    Sig.   Date.

Table 4: Survey

NAME	MOBILE PHONE	MOBILE PHONE FEATURES							
		PRICE	BATTERY LIFE	MEMORY	BUILD AND DESIGN QUALITY	SCREEN	CAMERA	PROCESSER POWER	OPERATING SYSTEM
ZAINAB JAVAID	APPLE	0.8	0.7	0.8	1	0.9	0.8	0.8	0.9
MOHAMMMAD UZAIR	GOOGLE	0.8	0.9	0.8	0.9	1	0.8	0.9	0.8
ZAID BIN NISAR	SAMSUNG	0.9	0.8	1	0.5	0.6	0.7	0.8	0.8
HANNAN SAJAD	OPPO	0.4	0.3	1	0.3	0.9	0.5	0.4	0.3
TARIQ AHMAD	XIAOMI	0	0.2	0.9	0.8	0.3	0.7	0.6	0.4
SAJAD AHMAD	LENOVO	0.4	0.3	0.6	0.7	0.5	0.9	0.3	0.7
MAJID ISHAQ	ONE PLUS	1	1	0.3	0.2	0.7	0.5	0.1	0.9
RAKESH KUMAR	SAMSUNG	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
MONU SINGH	MOTOROA	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
INAM ILLAHI	XIAOMI	0.1	0.3	0.2	0.2	0.4	0.1	0.9	0.5
ZUBAIR AHMAD	GOOGLE	1	0.7	0.5	0.9	0.2	0.1	0	1
NEELA SHUKLA	SAMSUNG	0.4	0.3	0.6	0.7	0.2	0	1	0.4
MUDASIR AHMAD	XIAOMI	0.9	0.7	0.4	0.5	0.7	0.9	0.2	0.4
FAYAZ AHMAD	OPPO	0.4	0.3	1	0.3	0.9	0.5	0.4	0.3
BASIT RIYAZ	SAMSUNG	0	0.2	0.9	0.8	0.3	0.7	0.6	0.4
TANVEER JAN	GOOGLE	0.4	0.3	0.6	0.7	0.5	0.9	0.3	0.7
BHAWNA	APPLE	1	1	0.3	0.2	0.7	0.5	0.1	0.9
HUDAYA	REALME	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
RAVI KUMAR	OTHER	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
NEELIMA	APPLE	0.1	0.3	0.2	0.2	0.4	0.1	0.9	0.5
ADIBA QURESHI	OPPO	1	0.7	0.5	0.9	0.2	0.1	0	1
BASARAT	OTHER	0.4	0.3	0.6	0.7	0.2	0	1	0.4
DANISH	XIAOMI	0.9	0.7	0.4	0.5	0.7	0.9	0.2	0.4
NASREEN NOOR	SONY	0.6	0.1	0.8	0.4	0.2	0.1	0.8	0.4
JYOTSNA SAHU	XIAOMI	0	0.9	0.8	0.6	0.4	0.2	0.8	0.1
MEHRUNISA KHAN	SONY	0.1	0.2	0.3	0.4	0.6	0.9	0.5	0.7
SHAHEEN SHIKH	APPLE	0.4	0.3	0.6	0.7	0.2	0	1	0.4
SARWAT ALAM	SAMSUNG	0.9	0.7	0.4	0.5	0.7	0.9	0.2	0.4
REHMAT ANSARI	OPPO	0.6	0.1	0.8	0.4	0.2	0.1	0.8	0.4
PRIYA DAVE	REALME	0.9	0.8	0.6	0.5	0.8	0.8	0.4	0.1

DEEPAK KUMAR	GOOGLE	0.4	0.5	0.9	0.5	0.6	0.8	0	1
KHALID AHMAD	SAMSUNG	0.6	0.5	0.4	0.8	0.6	0.3	0.1	0.9
BHAWNA KASBE	GOOGLE	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
TABASUM	APPLE	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
SUMIRA	REALME	0.5	0.6	0.4	0.6	0.9	0.4	0.5	0.8
ZEBA PATEL	OTHER	02	0.3	0.8	0.5	0.7	0.61	0	0.9
ASHIQ HUSSAIN	APPLE	0.2	0.8	0.3	0.4	0.1	0.6	0.3	0.9
ARSHID AH.	OPPO	0	0.8	0.7	0.9	1	0.5	0.7	0.5
PRITI SHUKLA	OTHER	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
MUZAFAR AH.	XIAOMI	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
RIKAZ RATHORE	SONY	0.9	0.3	0.4	0.1	0.2	0.4	0.5	0
GITABAI	XIAOMI	1	0.9	0.7	0.5	0.4	0.3	0.2	0.4
GEETA RAO	SONY	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
SARIKA SHARMA	APPLE	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
TARIQ AH.	SAMSUNG	0.1	0.2	0.4	0.5	0.9	0.3	0.2	1
NIHARIKA TIWARI	OPPO	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
SUSHMA SINGH	REALME	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
CHITRA SONI	GOOGLE	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
POOJA TIWARI	XIAOMI	0.1	02.	0.6	0.3	0.5	0.7	0.9	0.9
SUSHILA ATHYA	SONY	0.3	0.8	0.7	0.2	0.8	0.9	0.6	0.2
SEEMA KHAN	OPPO	0.8	0.5	0.3	0.2	0.8	0.8	0.7	0.3
ILYAS AH.	MOTOROLA	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
SHOWKAT	SAMSUNG	0.1	0.2	0.9	0.7	0.5	0.2	0.3	0.4
S. PAWAR	XIAOMI	0.1	0.3	0.2	0.1	0.9	0.8	0.5	0.6
NIDHI KHATRI	OPPO	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
MEHZABI ANSARI	SAMSUNG	0.9	0.4	0.7	0.1	0.5	0.4	0.3	1
MAJID ANSARI	GOOGLE	0.3	0.8	0.7	0.2	0.8	0.9	0.6	0.2
NAMRATA JI	APPLE	0.8	0.5	0.3	0.2	0.8	0.8	0.7	0.3
DOLLY	REALME	0.3	0.8	0.7	0.2	0.8	0.9	0.6	0.2
PRINCE	OTHER	0.8	0.5	0.3	0.2	0.8	0.8	0.7	0.3
RASHMI	APPLE	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
SHWETA JAIN	OPPO	0.6	0.9	0.9	0.5	0.	0.2	0.4	0.6
DIVYA TRIPATHI	OTHER	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
NEEDHI	XIAOMI	0.7	0.3	0	0.5	0.6	0.9	0.4	0.2
SHAGUFTA	SONY	1	0.5	0.6	0.3	0.2	0.7	0.9	0.6
RASHMI	APPLE	0.5	0.6	0.4	0.6	0.9	0.4	0.5	0.8
VEENA JOSHI	SAMSUNG	02	0.3	0.8	0.5	0.7	0.61	0	0.9
ARIHANT JAIN	OPPO	0.2	0.8	0.3	0.4	0.1	0.6	0.3	0.9
NIGHAT	REALME	0	0.8	0.7	0.9	1	0.5	0.7	0.5
IRFAN AH.	GOOGLE	0.7	0.2	0.6	0.8	0	0.1	0.4	0.5

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Limitation of the Proposed Technique

It is necessary for max-min function operation to be defined on two matrices; the number of columns of one matrix should be equal to the number of rows of another matrix. Otherwise max-min operation i.e., max {min (a_{ij}, b_{ij})} will not be defined.

References

[1] T. J. Ross, *Fuzzy Logic with Engineering Applications*, Third Edition © 2010 John Wiley & Sons, Ltd. ISBN: 978-0-470-74376-8.

[2] H. J. Zimmermann, *Fuzzy set theory--and its applieations*-4th ed. Includes bibliographical references and index. ISBN 978-94-010-3870-6.

[3] M. G. Thomason, convergence of powers of a fuzzy matrix, *J, mathAnal.Appl*. Vol. 57, pp. 476-480, 1977.

[4] M. Dhar, "Representation of Fuzzy Matrices Based on Reference Function", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol. 5, No. 2, pp.84-90,.DOI: 10.5815/ijisa.2013.02.10.

[5] K. H. Kim, F. W. Roush, Generalised fuzzy matrices, *fuzzy sets and systems, Vol.* 4, pp. 293-315, 1980.

[6] M. Bhowmik, M. Pal. Generalized intuitionistic fuzzy matrices, *Far-East Journal of Mathematical Sciences*, Vol. 29, No. 3, pp. 533-554, 2008.

[7] M. Gavalec, Computing matrix period in max-min algebra, *Discrete Appl. Math.*, Vol. 75, pp. 63-70, 1997.

[8] S. K. Khan, M. Pal, Some operations on intuitionistic fuzzy matrices, *Acta Ciencia Indica*, Vol. XXXII, No. 2, pp. 515-524, 2006.

[9] M. Pal, "Intuitionistic fuzzy determinant", *V. U. J. Physical Sciences*, Vol. 7, pp. 87-93, 2001.

[10] M. Pal, S. K. Khan, A. K. Shyamal, "Intuitionistic fuzzy matrices", *Notes on Intuitionistic Fuzzy Sets,* Vol. 8, No. 2, pp. 51-62, 2002.

[11] A. K. Shyamal, M.Pal, "Two new operations on fuzzy matrices", *Journal of Applied Mathematics and Computing,* Vol. 15, No. (1-2), pp. 91-107, 2004.

[12] H. Hashimoto, "canonical form of a transitive fuzzy matrix", *fuzzy sets and systems*, Vol. 11, pp. 157-162, 1983.

[13] W. Kolodziejczyk, "convergence of powers of s.transitive fuzzy matrices", *fuzzy sets and systems;* Vol. 26, pp. 127-130, 1988.

[14] L. J. Xin, controllable fuzzy matrices, *fuzzy sets and systems*, Vol. 45. pp. 313-319, 1992

[15] M. Z. Ragab, E.G. Emam, The determinant and adjoint of a square fuzzy matrix, *fuzzy sets and systems,* Vol. 61, pp. 297-307, 1994.

[16] J. B. Kim, Determinant theory for fuzzy and Boolean matrices, *congressusnumerantium*, pp. 273-276, 1988.

[17] R. Hemasinha, N. R. Pal, J.C. Bezdek, "Iterates of fuzzy circulant martices", *Fuzzy Sets and Systems*, Vol. 60, pp. 199-206, 1993.

[18] K. L. Zhang, "Determinant theory for D01-lattice matrices", *Fuzzy Sets and Systems*, Vol. 62, pp. 347-353, 1994.

[19] Y. J. "Tan, Eigenvalues and eigenvectors for matrices over distributive lattice, *Linear Algebra Appl.*, Vol. 283 , pp. 257-272, 1998.

[20] R. K. Tripathi, J. A. shah, "Fuzzy Logic Controller for Accurate Diagnostics in X-Ray Film". In: Mai C.K., Reddy A.B., Raju K.S. (eds) *Machine Learning Technologies and Applications. Algorithms for Intelligent Systems. Springer, Singapore.* https://doi.org/10.1007/978-981-33-4046-6_9. 2021.

[21] K. Valaskova, T. Kliestik, M. Misankova, "The role of fuzzy logic in decision making process". *In 2014 2nd international conference on management innovation and business innovation* .Vol. 44, pp. 143– 148, 2014.

[22] W. B. V. Kandasamy, F. Smarandache, K. Ilanthenral "*Elementary Fuzzy matrix theory and fuzzy models for social scientists*", Los Angeles: Automaton, 2007.

[23] V. V. Raich, A. Gawande, R. K. Tripathi, "Fuzzy Matrix Theory and its Application for Recognizing the Qualities of Effective Teacher", *International Journal of Fuzzy Mathematics and Systems.* Vol. 1, No. 1, pp. 113-122, 2011.

[24] X. Guo, D. Shang, "Fuzzy Symmetric Solutions of Fuzzy Matrix Equations", *Advances in Fuzzy Systems*, Vol. 2012, Article ID 318069, 9 Page, 2012. https://doi.org/10.1155/2012/318069.

The author has done his bachelor's and master's degrees from the University of Kashmir Srinagar. He has completed his M. Phil degree from Vikram University Ujjain and completed his PhD degree from Dr. A. P. J. Abdul Kalam University Indore.

He has ten years of teaching experience and his research interest is in fuzzy mathematics and control. He has published various papers in this field.

Machine-Learning based Decoding of Surface Code Syndromes in Quantum Error Correction

Debasmita Bhoumik[*,1], **Pinaki Sen**[2], **Ritajit Majumdar**[1], **Susmita Sur-Kolay**[1], **Latesh Kumar K J**[3], **Sundaraja Sitharama Iyengar**[3]

[1] Advanced Computing & Microelectronics Unit, Indian Statistical Institute, Kolkata, 700108, India

[2] Department of Electrical Engineering, National Institute of Technology, Agartala, India

[3] KFSCIS, Florida International University, Miami, Florida, USA

[*] Corresponding author: Debasmita Bhoumik, Indian Statistical Institute, Email : debasmita.ria21@gmail.com

ABSTRACT: Errors in surface code have typically been decoded by Minimum Weight Perfect Matching (MWPM) based method. Recently, neural-network-based Machine Learning (ML) techniques have been employed for this purpose, although how an ML decoder will behave in a more realistic asymmetric noise model has not been studied. In this article we (i) establish a methodology to formulate the surface code decoding problem as an ML classification problem, and (ii) propose a two-level (low and high) ML-based decoding scheme, where the first (low) level corrects errors on physical qubits and the second (high) level corrects any existing logical errors, for various noise models. Our results show that our proposed decoding method achieves ∼ 10× and ∼ 2× higher values of pseudo-threshold and threshold respectively, than for those with MWPM. We also empirically establish that usage of more sophisticated ML models with higher training/testing time, do not provide significant improvement in the decoder performance.

KEYWORDS Quantum Error Correction, Surface code, Error decoding, Machine learning decoder

1. Introduction

Quantum computers are expected to provide faster and often more accurate solutions to some of the problems of interest such as factorization [1], database searching [2], Hamiltonian simulation [3], finding the lowest energy configuration of molecular systems [4]. Quantum computers make use of properties such as superposition, entanglement etc., which are not observed in macroscopic world, to achieve the speedup. A general quantum bit, or qubit, is mathematically represented as $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$, $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$ [5].

Quantum states are, however, very prone to errors. Being vectors in Hilbert space, even the slightest unwanted rotation occurring due to interaction with the environment introduces error in the quantum system. It was shown by [6], that any unitary quantum error can be expressed as a linear combination of the Pauli matrices (I, X, Y, Z) [1]. Hence, if a quantum error correcting code (QECC) can correct the Pauli errors, then it can also correct any unitary error. The 9-qubit code [6], 7-qubit code [7] and 5-qubit code [8] are some early QECCs. The 5-qubit code is optimum in the number of qubits .

The circuit realization of the above-mentioned QECCs comprise multiple operations involving qubits which are not adjacent to each other. Operation on two non-adjacent qubits is both slow and error-prone, due to the multiple swap operations required. Surface code was introduced to overcome this drawback, known as the Nearest Neighbour (NN) problem, by placing the qubits in a 2D grid-like structure [9], and the operations for error correction are performed only between adjacent qubits. Protocols are formulated for error recovery, and the efficacy of these protocols were studied in [10] which is reviewed in a simplied manner by [11]. An improved decoding algorithm for the surface code is formulated in [12].

A QECC encodes $n > 1$ physical qubits into $m < n$ logical qubits, where the latter are expected to be more secure under noise. A decoder (which is a classical process), on the other hand, detects the error present in the logical qubit. Decoding is followed by another step where the correction is applied physically and classically (or in some cases noted logically only [13]). A distance d QECC can correct $\lfloor \frac{d}{2} \rfloor$ errors on the physical qubits, keeping the logical qubit error free. However, the logical qubit can become erroneous as well if more errors occur. This is termed as *logical error*. The errors may occur due to interaction with the environment, or faulty decoding. While the first issue may be tackled with a QECC having a larger d, the latter can pose a serious threat towards building error corrected qubits. The performance

---

[1]An operator U is unitary operator if $U^\dagger U = UU^\dagger = I$, where I is the identity element and $U^\dagger$ is the adjoint of U. An operator is hermitian if $U = U^\dagger$. Note that, hermitian operator is a subset of unitary operator. Pauli matrices are hermitian.

of a decoder for a QECC is assessed by two parameters [14], namely: (i) *pseudo-threshold*, which is the probability of physical error below which error-correction leads to a lower logical error probability, and (ii) *threshold*, which is the probability of physical error beyond which increasing d leads to higher logical error probability.

Apart from the accuracy of decoding, the time required is also important. In a fault-tolerant quantum computer, the qubits are encoded only once at the beginning of the computation, whereas they are decoded several times during the computation. Therefore, decoding time is critical [5, 15]. The most popular decoding algorithm for surface codes is Blossom Decoder [16] based on $O(N^4)$ time Minimum Weight Perfect Matching (MWPM) algorithm, N being the number of qubits. Recently machine learning (ML) has been used for decoding in linear time [17]. A baseline decoding algorithms complemented by different kinds of deep neural decoders was introduced by [18] and applied to analyze the common fault-tolerant error correction protocols such as the surface code. The decoding problem is reduced to a classification problem that a feedforward neural network can solve, in [19], for small code distances. Reinforcement Learning based decoders for Fault-Tolerant Quantum Computation were proposed in [20]. It has been observed that the MWPM based decoder performs satisfactorily when the error probability of the system is low, as it always tries to find the minimum number of errors that can generate the observed syndrome (see Section II below). But occurrence of error(s) in the system during decoding is ignored, which the ML based approaches do consider. Therefore, ML based decoders are expected to perform at least as well as MWPM based method and in less time.

ML based decoders can tackle errors incorporated due to faulty decoding upto some extent. This is achieved by introducing two-level decoding, where the low level is a traditional decoder (need not be an ML decoder [19]), and the high level (necessarily ML decoder) predicts any logical errors that may have resulted during decoding. In [17] the authors have used ML for both low and high-level decoders. However, it is unclear whether their noise model considers errors in a single or multiple steps in the error correction cycle of surface code (see Fig: 2). Moreover, the performance of ML decoders for asymmetric noise (which is a more realistic noise model [21]), and whether the usage of more sophisticated ML models can significantly enhance the performance of the decoder, remains largely unanswered.

This article aims to address these unanswered problems by using machine learning based low and high level decoders for both symmetric and asymmetric noise model. We want to emphasize here that the primary focus of this paper is the efficient mapping of the surface code decoding problem to ML classification, showing its decoding performance, and experimentally verifying whether the sophistication of the used ML model has any significant effect on the decoding performance. In this work we are applying machine learning for quantum error correction, hence the machine learning tasks are purely classical, not quantum. In other words, quantum machine learning is not involved in our approach. As a preliminary research, we have assumed that only the data qubits may be noisy, but

both the measurement and the stabilizers are noise-free. In our follow-up work, we will consider noisy data, stabilizer, as well as non-ideal measurement.

The rest of the paper is organized as follows: we have summarized our contributions here in Section 2. In Section 3, we have outlined the stabilizer formulation of the surface code. In Section 4, we have established the formulation of the decoding problem of a surface code as an ML classification problem. In Section 5 we have presented our results with various error models and our concluding remarks appear in Section 6. In the Appendix we have provided the list of the acronyms used in this paper.

2. Summary of the contributions

The main contributions of our work are to:

- design a well-defined step-by-step methodology to formulate the decoding problem in surface code as an ML classification problem;

- study the performance of ML based low and high level decoders, for distance 3, 5 and 7 surface codes, where error can occur in one or more of the eight steps in the surface code QECC cycle with equal probability. We show that our ML based decoder achieves ~ 10× higher pseudo-threshold and ~ 2× higher threshold as compared with MWPM based decoder.

- establish that our ML-based decoder outperforms the one based on MWPM for varying degrees of asymmetry in the noise model as well;

- experimentally show that varying the level of sophistication (i.e., number of layers, nodes in each layer, etc.) for Feed Forward Neural Network (FFNN) and Convolutional Neural Network (CNN) does not provide significant improvement in the performance of the syndrome decoder;

- show that when distance d is small, and hence the total number of distinct errors is also small, the ML based decoder can learn the most probable errors within a small subset of the training data, which is generated uniformly at random. Therefore, a small train-test ratio suffices to obtain a good decoding performance.

- empirically determine the degree of asymmetry of the noise channel below which an ML decoder trained using symmetric noise model retains optimal pseudo threshold. This provides an estimate of re-usability of a pre-trained ML decoder on a variety of noise model.

3. Stabilizer Formulation of Surface Code

In [22], the author proposed the stabilizer formulation for error correction. A set of mutually commuting operators $M_1, \ldots, M_r$, where each $M_i \in \{I, X, Z, Y\}^{\otimes n}$, is said to stabilize an n-qubit quantum state $|\psi\rangle$ if $M_i |\psi\rangle = |\psi\rangle$, $\forall i$ [22]. An error E is said to correctable by a QECC, if there exist stabilizers $M_e \subseteq \{M_1, \ldots, M_r\}$, such that $M_e (E |\psi\rangle) = -E |\psi\rangle$.

A QECC is called degenerate if there exist errors $e_1 \neq e_2$ such that $e_1 |\psi\rangle = e_2 |\psi\rangle$ where $|\psi\rangle$ is the codeword. It is not possible to distinguish between such errors in a degenerate code. Surface code is a degenerate stabilizer code. Surface code is implemented on a two-dimensional array of physical qubits. The data qubits (in which the quantum information is stored) are placed on the vertices, and the faces are the stabilizers (refer Fig. 1). The qubits associated with the stabilizers are also called measure qubits. These are of two types : *Measure-Z* (*M-Z*) and *Measure-X* (*M-X*). Each data qubit interacts with four measure qubits — two *M-Z* and two *M-X*, and each measure qubit, in its turn, interacts with four data qubits (Fig. 1). An *M-Z* (*M-X*) qubit forces its neighboring data qubits a, b, c and d into an eigenstate of the operator product $Z_a Z_b Z_c Z_d$ ($X_a X_b X_c X_d$), where Z_i (X_i) implies Z (X) measurement on qubit i. Pauli-X and Pauli-Z errors are detected by the Z- and X- stabilizers respectively (Fig. 1). An X (Z) logical operator is any continuous string of X (Z) errors that connect the top (left) and bottom (right) boundaries of the 2D array. The number of measure qubits, and hence the number of stabilizers, is one less than the number of data qubits when encoding a single logical qubit of information. An error-correcting code can correct up to t errors if its distance $d \geq 2t+1$. A distance 3 surface code consists of 9 data qubits and 8 measure qubits (Fig. 1). Thus a total of 17 qubits encode a single logical qubit, and hence the distance 3 surface code is also called SC17.

The circuit representations of the decoding corresponding to a single *M-Z* qubit and an *M-X* qubit are shown in Fig. 2. Since the same measure-qubit is shared by multiple data qubits, different errors can lead to the same syndrome in surface code. Hence the mapping from syndrome to error is not one-to-one, as illustrated in Fig. 3. This often leads to poor decoding performance by decoders. In fact, if a decoder misjudges an error e_1 for some other error e_2, it can so happen that $e_1 \oplus e_2$ leads to a logical error. Therefore, not only the presence of physical errors, but also incorrect decoding can lead to uncorrectable logical errors as well. The goal of designing a decoder, thus, is to reduce the probability of logical error for some physical error probability.

As defined in the Introduction section, the performance of a decoder is measured in terms of pseudo-threshold and threshold. With increasing code distance, the pseudo-threshold for a particular decoder also increases, which supports the intuition that using larger distance gives better protection from noise. On the other hand, the threshold does not change with respect to the distance because a decoder for a particular surface code yields a fixed threshold. The higher are the values of these parameters, the better is the performance of the decoder. Of these two parameters, the pseudo threshold is lower than the threshold for a decoder. The reason is that error correction is effective below the pseudo-threshold point, and coding theory asserts [23] that in this region, increasing the distance of the code leads to higher suppression of logical errors. Therefore, if the threshold point is below the pseudo-threshold point it violates coding theory. Hence, it is more important for a decoder to have a higher pseudo-threshold than a higher threshold, since, beyond this error probability, QECC no longer provides any improvement in suppression of errors.

Let there be n physical qubits in a logical qubit (eg. $n = d^2$ for surface code). A logical error can occur only when at least d of the n physical qubits are erroneous. Nevertheless, the presence of d or more physical errors does not necessarily imply the presence of a logical error. If p and p_L are respectively the probability of physical and logical error, then

$$p_L \leq \Sigma_i p^i, \text{ for } d \leq i \leq n$$

Moreover, incorrect decoding itself can lead to logical errors. This can happen when the decoder fails to detect the actual physical errors and thus incorporates more errors during correction. Once again, not every incorrect decoding leads to a logical error. Therefore, if p_d is the probability of failure of the decoder, then

$$p_L \leq \Sigma_i p^i + f(p_d), \text{ for } d \leq i \leq n$$

where $f(p_d)$ is a function of the probability of failure of the decoder. The function $f(p_d)$ may vary with the decoder, hence the logical error probability may differ, resulting in different values of pseudo-threshold and threshold.

4. Machine learning based syndrome decoding for surface code

Machine Learning is a branch of artificial intelligence where a machine learns without being explicitly programmed. Depending on the type of training data (labeled/unlabeled/combined), the ML algorithm can vary (supervised/unsupervised/semi-supervised). It has a plethora of applications domains such as soil properties prediction [24], human pose estimation [25], object recognition [26], video tracking [27], prediction of the efficacy of online sale [28] etc. Here, we employ machine learning to decode error syndrome(s) for quantum error correction.

4.1. Advantages of Machine learning based syndrome decoder

Classical algorithms for decoding, such as Minimum Weight Perfect Matching (MWPM), may perform poorly in certain cases. For example, MWPM tries to find a minimum number of errors that can recreate the error syndrome obtained without considering the probability of error.

If $|\psi\rangle_{1,2,\ldots,n}$ is an n-qubit codeword, and we consider the error generation on this codeword is a stochastic map $S(p_1, p_2, \ldots, p_n)$, where p_i is the probability of error on qubit i (we can further write p_i in terms of the probability of Pauli errors), then the error state $|\psi\rangle_e = S(p_1, p_2, \ldots, p_n)|\psi\rangle_{1,2,\ldots,n}$. Now, for a distance d surface code with t types of errors ($t=4$ for depolarization, 2 for bit/phase flip), there are t^{d^2} possible errors and t^{d^2-1} possible syndromes. Therefore, multiple errors $\mathcal{E} = \{e_1, e_2, \ldots, e_l\}$ lead to the same syndrome, and detecting a syndrome cannot uniquely specify the type of error causing it. Since Pauli errors are hermitian, correction is simply applying the same error once more. If the choice of error is not perfect, then the system ends up with (probably) more error than before after the correction step.

MWPM, being a deterministic algorithm, does not consider the Stochastic map. It assumes that the error probability is low, and always finds the minimum weight error
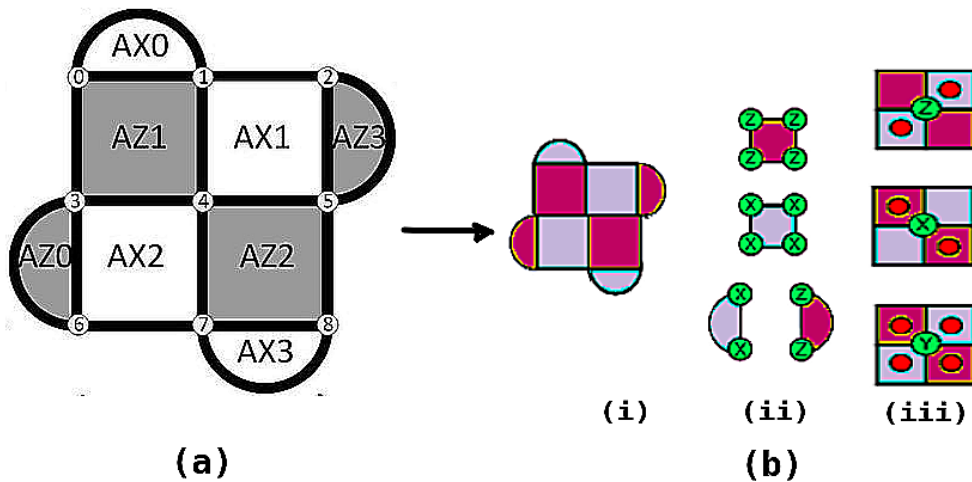
Figure 1: (a) Distance 3 surface code, where the numbered circles (0 - 8) are the physical qubits, white plaquettes are *X* stabilizers (i.e., *M-X* qubits *AX*0, *AX*1, *AX*2, *AX*3), gray plaquettes are *Z* stabilizers (i.e., *M-Z* qubits *AZ*0, *AZ*1, *AZ*2, *AZ*3); (b) the syndromes are defined in a $d \times d$ lattice (d=3), with physical qubits on the vertices and plaquette stabilizers (measure qubits) as faces: (i) pink (purple) plaquettes indicate stabilizers which check the *Z* (*X*) parity of qubits on the vertices of the plaquettes as shown in (ii), (iii) green circles indicate errors and red circles violated stabilizers (i.e., syndromes [19])
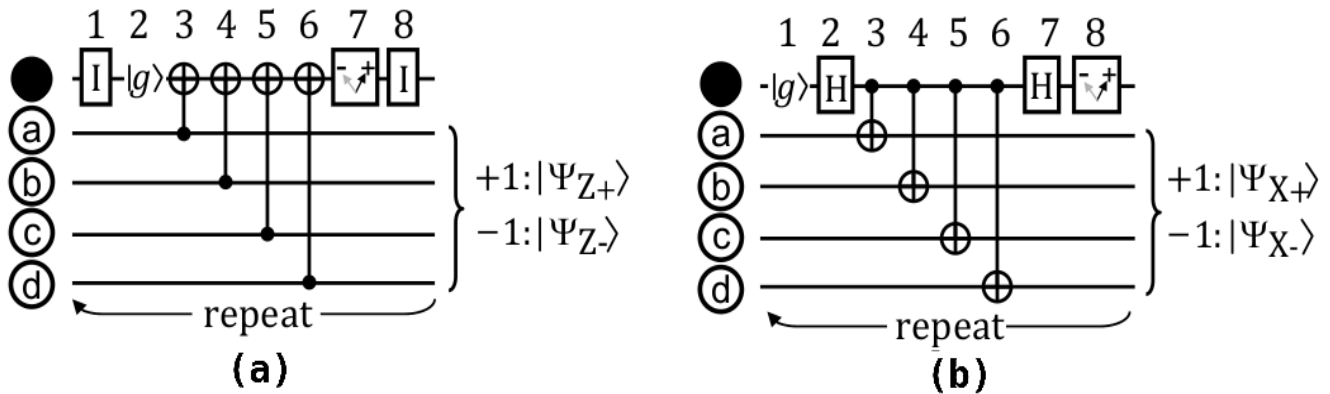
Figure 2: Quantum circuit for a single cycle of surface code. (a) circuit for M-Z qubit, (b) circuit for M-X qubit [14]

$e_{min} \in \mathcal{E}$ that creates the observed syndrome. On the other hand, an ML decoder learns the probabilities $p_1, p_2, \ldots, p_n$ from the training phase. Thus, this decoder finds most likely error $e_{ml} \in \mathcal{E}$ that can cause observed syndrome depending on the Stochastic Map.

Furthermore, the time complexity of MWPM grows as $O(N^4)$ where N is the number of qubits. Lookup Tables have been used for decoding as well [29]. While Lookup Table Decoder is sometimes better than MWPM in performance, its complexity scales as $O(4^N)$ which becomes infeasible even for moderate values of N. To overcome such drawbacks, ML techniques have been applied to learn the probability of error in the system and propose the best possible correction accordingly with comparatively lower time complexity. For example, [19] reduced the decoding problem to a classification problem that a feed-forward neural network can solve, for small code distances. A deep neural network based decoder is proposed by [30] for Stabilizer Codes. Therefore, supervised learning techniques, such as Feed-forward neural network (FFNN), Recurrent Neural Network (RNN) show that these are capable of outperforming the traditional

decoding techniques.

As discussed earlier, surface code is degenerate, i.e., there exist errors $e_1 \neq e_2$ such that $e_1 |\psi\rangle = e_2 |\psi\rangle$, where $|\psi\rangle$ is the codeword. This leads to any decoder failing to distinguish between some errors e_1 and e_2. Nevertheless, that does not always lead to a logical error. For example, bit-flip error in bit 1 and bit 2 are indistinguishable. But error in decoding these two will not lead to a logical error (Refer Fig. 4 (a)). On the other hand, it is possible that $e_1 \oplus e_2$ leads to a logical error, i.e., the decoder may itself incorporate logical errors while correcting physical errors. For example, bit-flip error on qubit 4 is indistinguishable from those on qubits 1 and 7 together. But failure to distinguish between these two bit-flip errors leads to logical error (refer Fig. 4 (b)).

In general, *usually* the decoder incorporates logical errors when it fails to distinguish between $\lfloor (d-1)/2 \rfloor$ and $\lceil (d+1)/2 \rceil$ errors. Broadly speaking, ML can learn the probability of error and predict which of those two are more likely. This makes ML-decoder outperform other traditional decoders.

Since a decoder itself can incorporate logical errors, two stages of decoders, namely low level followed by high level
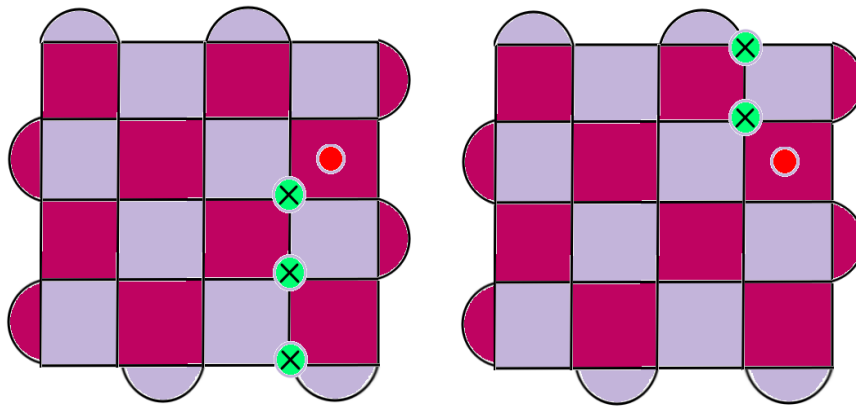
Figure 3: Example of surface code for d=5, where two errors produce the same syndrome [19]. Pink (Purple) plaquettes indicate stabilizers which check the Z (X) parity of qubits on the vertices of the plaquette. Green circles are used to indicate errors and red circles to indicate violated stabilizers.

decoder, have been applied where

- Low level decoders search for exact position of errors at the physical level.

- High level decoders attempt to correct any logical error incorporated by the correction mechanism of low level decoders.

4.1.1. Design methodology of our ML based decoder

Artificial neural networks (ANN) are made to emulate the way human brains learn, and are one of the most widely used tools in ML. Neural networks consist of one input layer, one output layer, and one or more hidden layers consisting of units that transform the input into intermediate values from which the output layer can find patterns that are too complex for a human programmer to teach the machine. The time complexity of training a neural network with N inputs, M outputs and L hidden layers is $O(N \cdot M \cdot L)$. In this paper we are using neural networks as both low-level and high-level decoder for distance 3, 5, and 7 surface code.

In order to apply ML techniques to surface code decoding, we first map the decoding problem to the classification problem as follows. Given a set of data points, a classification algorithm predicts the class label of each data point. These techniques are purely classical. Next, we describe in detail the formulation of a decoder for surface code as a classification problem.

4.1.2. Mapping Surface Code onto a square lattice

For ease of implementation, we have mapped the surface code to a square lattice (refer Fig. 5) in this work. This has been achieved by padding a few dummy nodes (labelled as 0_D in the figure). A distance d surface code is converted into a (d+1) × (d+1) square lattice which has $d^2 − 1$ stabilizers, when encoding a single logical qubit. Therefore, 2(d+1) dummy nodes are required for this square lattice. The dummy nodes are basically don't care nodes, and their value is always 0 irrespective of the error in the surface code. The syndrome changes the values of the stabilizers only.

4.1.3. Error injection and syndrome extraction

Once the distance d surface code is transformed to a (d+1) × (d+1) square lattice, the next step is to extract the syndrome for errors. First, we create a training dataset, where in each data we randomly generate errors on each physical qubit. If p_{phys} is the probability of error on a physical qubit, the total probability of error after the 8 steps of surface code cycle (Fig. 2) is $1 - (1-p_{phys})^8$. We have trained the networks with p_{phys} ranging from 0.0001 to 0.25. One can argue that 0.25 is an unreasonably high error probability. However, we have ranged the error probability that far to show an interesting observation regarding the ML decoder performance (Sec 4).

For generating the training data we have considered bit flip errors, symmetric and asymmetric depolarizing noise models. We have not separately considered phase flip errors since they are similar to bit flips and have a rotational symmetry (i.e., the logical errors of bit flip and phase flip model are equivalent up to a rotation by $\frac{\pi}{2}$).

From the training data (which may or may not contain errors), we generate the syndrome (measured by ancilla qubits of the surface code) (Fig. 5). The syndrome, in our implementation, contains both the ancilla and the dummy nodes. However, the dummy nodes are always 0, whereas the values of the ancilla changes with different errors. Henceforth, in terms of implementation only, syndrome for a distance d surface code will imply (d+1)^2 values including ancilla and dummy nodes. The final training data contains the syndrome, and its corresponding label is the true set of errors that have occurred in the system. Note that this method can lead to multiple labels having the same syndrome. This agrees with the fact that surface code does not have one-to-one mapping from error to syndrome.

Ideally, the dataset to achieve the best decoding performance should include all possible error syndromes. But as the code distance increases, the state space also increases exponentially. Therefore, we can at most include only a small percentage of the entire input dataset. The dataset size that we have used is 100000 from which 70000 is used for training and the rest for testing purpose.
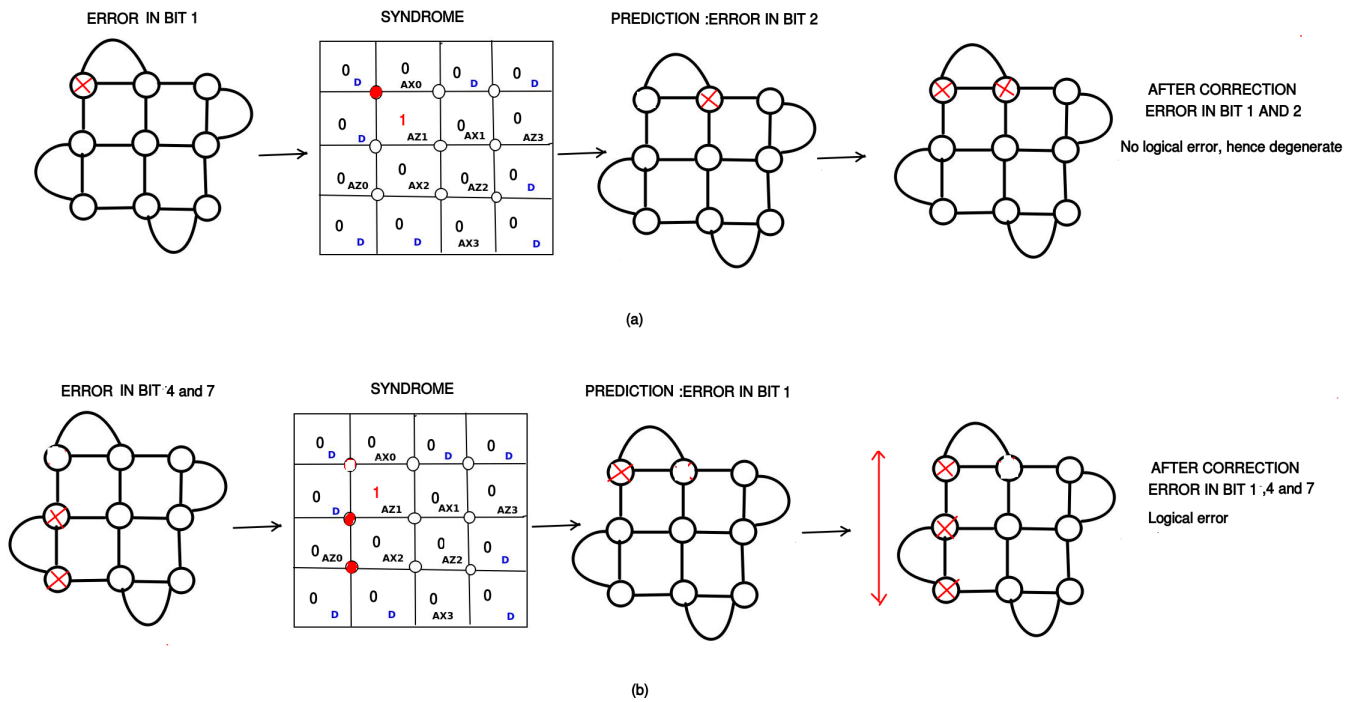
Figure 4: (a) No logical error and (b) Logical error due to mis-classification in low level decoding

4.1.4. Training our ML model

For the low level decoder, we train a neural network where the input layer is the syndrome and the output layer denotes the types of errors along with the physical data qubit where each error has occurred. For a distance d surface code, the number of input nodes are $(d + 1)^2$ containing $d^2 - 1$ measure qubits and $2(d+1)$ dummy nodes. For example, if we consider a distance-3 surface code (SC17), it has 8 ancilla qubits and 8 dummy nodes. Therefore, in the input layer, there are 16 nodes (Fig. 5). In the output layer, there are 2 nodes for each data qubit to differentiate among I, X, Y and Z errors. The size of the hidden layer can be adjusted by trial-and-error.

We have used two types of neural networks, (i) Feed Forward Neural Network (FFNN) and (ii) Convolutional Neural Network (CNN). In our reported results,

(i) FFNN consists of 2 hidden layers having 32 and 16 nodes respectively. For the cost function we have used the mean squared error rate, and as the activation function we have used Rectified Linear Unit (ReLU).

(ii) For CNN, the first layer is a 64 dimension convolution layer where input is a 4×4 matrix and the kernel size is also 4×4. Then we flatten it and add two fully connected layers of dimension 64 and 32. After that we add the fully connected output layer of dimension 9. For the first 3 layers (convolution, dense, dense) we have used ReLu as the activation function and for the output layer we have used sigmoid activation function since it will be a multi-label classification problem.

These values were adjusted after multiple trial-and-errors. We later show in the result section that building a more complex neural network cannot provide any significant improvement in the performance of the decoder, but requires significantly more decoding time. Therefore, we stick to these parameters.

The high-level decoder simply tries to predict any logical error that has been incorporated by the low level decoder. Therefore, its input remains the same as the low-level decoder (i.e., the syndrome) whereas it has 4 nodes in the output, each corresponding to a logical Pauli operator.

First, the network is trained for low-level decoder. After the low-level decoding is done, the predicted corrections are applied, and rechecked by using the high-level decoder whether any logical error has been inserted by the low-level decoder. The entire workflow is given in Fig. 6.

5. Results

First, we focus on the decoding performance of an ML-based low-level and the high-level decoder for surface codes of distances 3, 5, and 7 for both symmetric and asymmetric depolarizing noise models with varying degrees of asymmetry. Our model outperforms the performance of the existing decoders for symmetric noise model. We also show that although the performance of ML is slightly poorer for asymmetric noise models than that for the symmetric one, it still outperforms MWPM. Furthermore, we provide an empirical study to estimate the minimum train-test-ratio needed for optimal accuracy to obtain a better estimate of the minimum number of training data required to obtain the best (or near best) decoding results with ML decoder.

In the following subsections, we first introduce the noise model that we have considered, followed by the parameters

Bit flip in 5th bit marked by X

Dummy nodes marked with D are always 0 (in blue).
Note: A bit flip error is detected by measure Z
qubits. These return "-1" eigenvalue for error and
"+1" for no error, which are coded as "1" and "0"
respectively for implementation purpose.
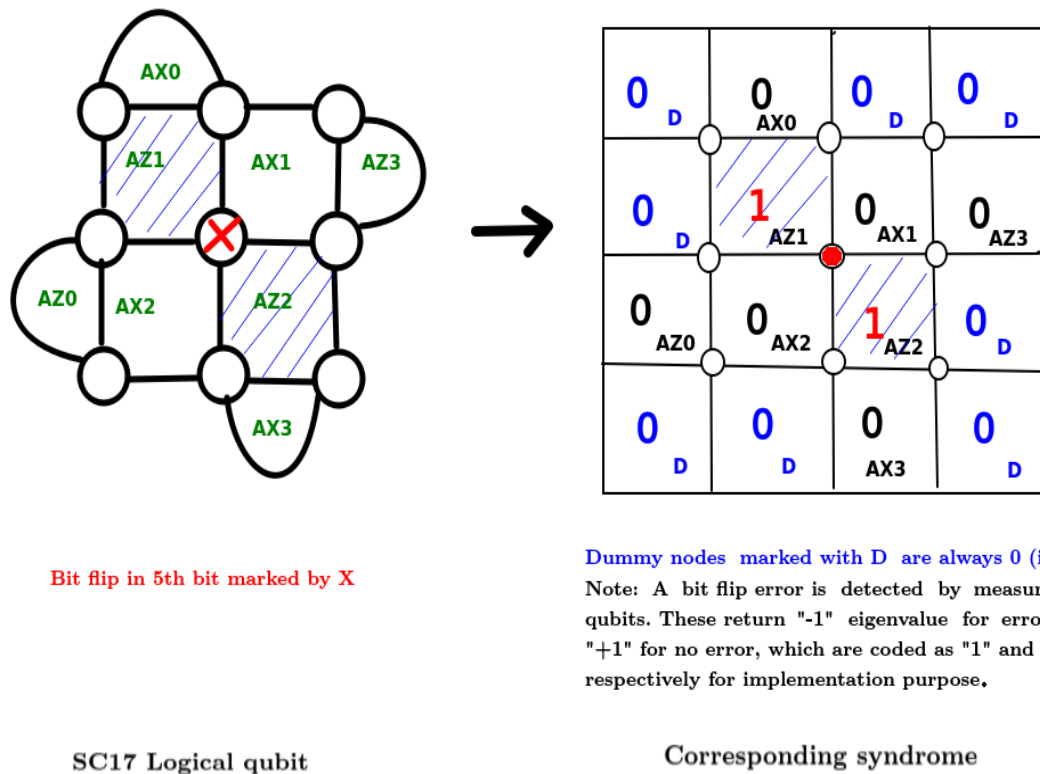
SC17 Logical qubit

Corresponding syndrome

Figure 5: SC17 to syndrome generation

of our ML decoder. Finally, we show the results of our decoder and compare its performance with the traditional MWPM decoder.

5.1. Noise models

Given a quantum state ρ in its density matrix formulation [5], the evolution of the state in a depolarization noise model is given as

$$\rho \to (1 - p_x - p_y - p_z)\rho + p_x X \rho X^\dagger + p_y Y \rho Y^\dagger + p_z Z \rho Z^\dagger$$

where p_x, p_y, p_z represent the probability of occurrence of unwanted Pauli X, Y, and Z error. In symmetric depolarization noise model, $p_x = p_y = p_z$. Moreover, quantum channels are often asymmetric or biased, i.e., the probability of occurrence of Z error is much higher than that of X or Y error. Furthermore, each error correction cycle in surface code requires eight steps. We have considered that an error can occur on one or more of the d^2 data qubits in each of the eight steps, where d is the distance of the surface code. Therefore, if $p_x + p_y + p_z = p$, then the overall probability of error for each error correction cycle is $1 - (1 - p)^8$ (Refer Fig. 2). This error model is in accordance with [11]. We assume noise-free measure qubits (which are almost half the total number of qubits) and ideal measurements.

5.2. Machine Learning Parameters

For our study, we have trained the ML model with batches of data, not the entire data set at once. This is often beneficial in terms of training time as well as memory capacity. We have used batch size = 10000, epochs = 1000, learning rate = 0.01

(with Stochastic Gradient Descent), and we have reported the average performance of each batch over 5 instances. This is repeated for each value of the p_{phys} considered here.

5.2.1. Low and high-level decoder

In Fig. 7, we show the increase in the logical error probability with physical error probability p, which is the probability of error per step in the surface code cycle. The results of MWPM and CNN-based low-level decoder for both symmetric and asymmetric noise models are shown. In Tables 1 and 3, we depict the performance of FFNN decoder as well. In Fig. 7, the blue, yellow, green, and red lines respectively are the decoder curves which show the probabilities of logical error for symmetric depolarization, bit flip (X), phase flip (Z), and Y errors. The cyan straight line consists of the points where the probabilities of physical and logical error are equal.

The point where the decoder curves and the straight line intersects, defines the value of pseudo-threshold for the decoder. As expected, the pseudo threshold improves with increasing distance of the surface code. Nevertheless, the threshold value is the probability of physical error beyond which increasing the distance leads to poorer performance. Therefore, threshold is independent of the distance and is a property of the surface code and the noise model only. In Tables 1 and 3, we show the pseudo-threshold and threshold of the low and high-level decoders for distance 3, 5 and 7 surface code in symmetric and asymmetric noise models respectively. Fig. 7 shows the pseudo-thresholds for MWPM and CNN decoder for a distance 3 surface code using low-level decoder (LLD) only. From Table 1 we observe

~ 10× increase in the pseudo threshold for ML-decoders as compared to MWPM.

Figure 6: Outline of the ML based syndrome decoding for surface code

Fig. 8 shows the thresholds and decoder accuracy for MWPM and ML-decoders surface codes of distance 3, 5 and 7. Table 3 depicts the threshold values for MWPM and ML-decoders. From Table 3 we observe ~ 2× increase in the threshold for ML-decoders as compared to MWPM.

As already mentioned earlier, this result assumes error-free stabilizers, and ideal measurements. For a distance d surface code, there are $d^2 - 1$ stabilizers. Therefore, in our setting, nearly half of the total qubits in the surface code structure are considered ideal. We focused more on the mapping of decoding to Machine Learning in this research. A separate study is being carried out on the performance of the ML decoder in the presence of erroneous stabilizers and measure qubits to determine the threshold and pseudo-threshold. Our conjecture is that ML decoders will still outperform MWPM decoder in that scenario, but the increase in performance will be much lower.

In Fig. 7, we observe that at very low error probability the accuracy remains good, then it falls drastically. However, for ML decoders, it again increases beyond a certain physical error probability (0.15). On the other hand, the logical error also decreases in most of the cases for both symmetric and asymmetric ML decoders after more or less that same value of physical error probability. This is due to the bias in the back-end working principle of any machine learning model. When the error probability is low or high, the ML decoder effectively learns the probability and in most of the cases can avoid logical errors. But when the error probability is in the mid range, the ML model gets confused. For example, if in a training set, out of 12 events with same value of the features, 10 events are certainly in class A, and the rest in B, then the ML definitely learns it with high accuracy. Similarly, if those same 10 events are in class B, accuracy will be high. But the ML is confused when 6 of them are in class A and 6 of them in class B. This is an interesting observation in the ML-decoder which is absent in MWPM-decoder.

5.3. More sophisticated ML models

A natural question is whether the use of more sophisticated ML models (e.g, adding more hidden layers, increasing the number of nodes in each layer, etc.) can improve the performance of the decoder. We have addressed this issue as reported below.

In Table 2, Simple FFNN has 1 hidden layer (dense) whereas Complex FFNN has 5 hidden layers (dense) and Simple CNN has 1 convolution (64 dimensions) followed by 2 dense layers of dimension 256 and 64 respectively before the output layer whereas Complex CNN has 3 convolutions (64 dimensions) layers followed by 4 dense layers of dimension 512, 256, 128 and 64 respectively before the output layer. The more sophisticated models naturally require more time for training and prediction. But from Fig. 9, we see that the decoder graphs are more or less overlapping for the simple and complex ML models. Therefore, it can be concluded that using more sophisticated model does not lead to a better performance for the decoder. This can be further verified by the accuracy plots in Fig. 10. Since the more complex models are performing almost at par with the simpler models for $d=3$ and 5 and the complex models are significantly more time-consuming, we performed the experiments for $d = 7$
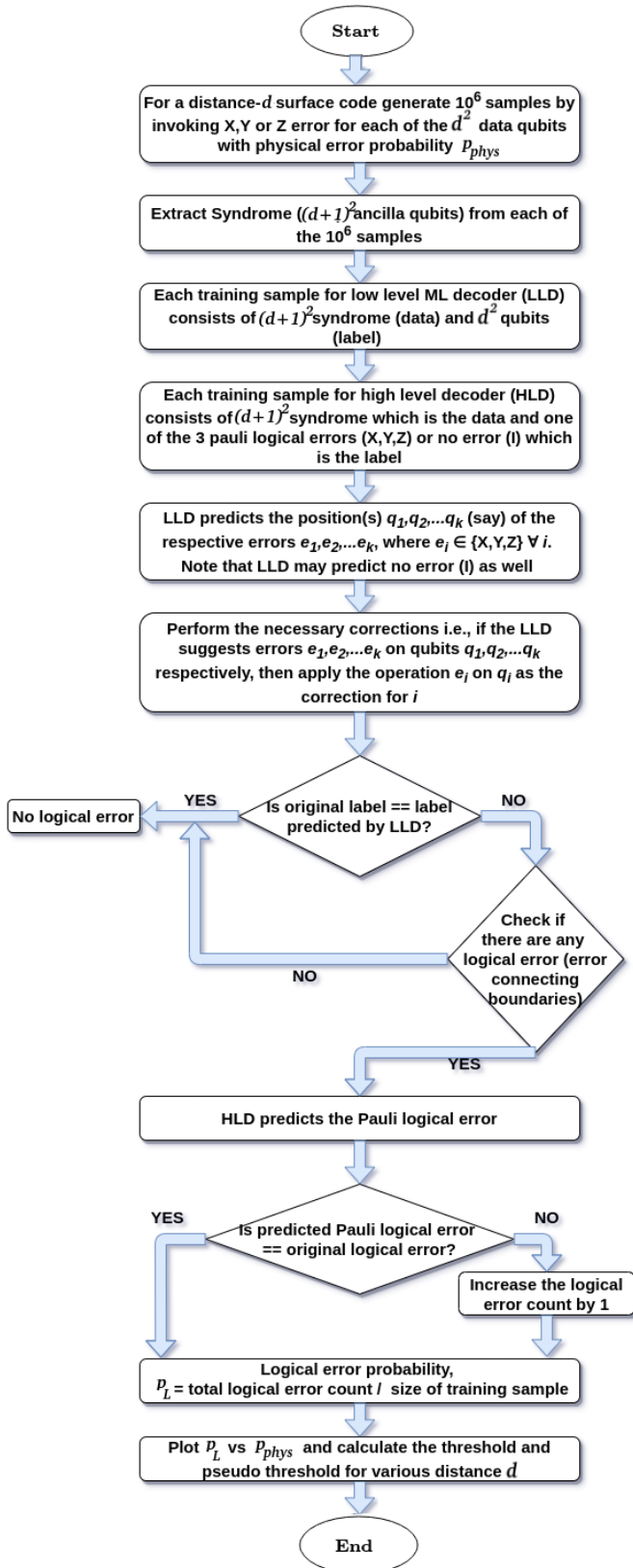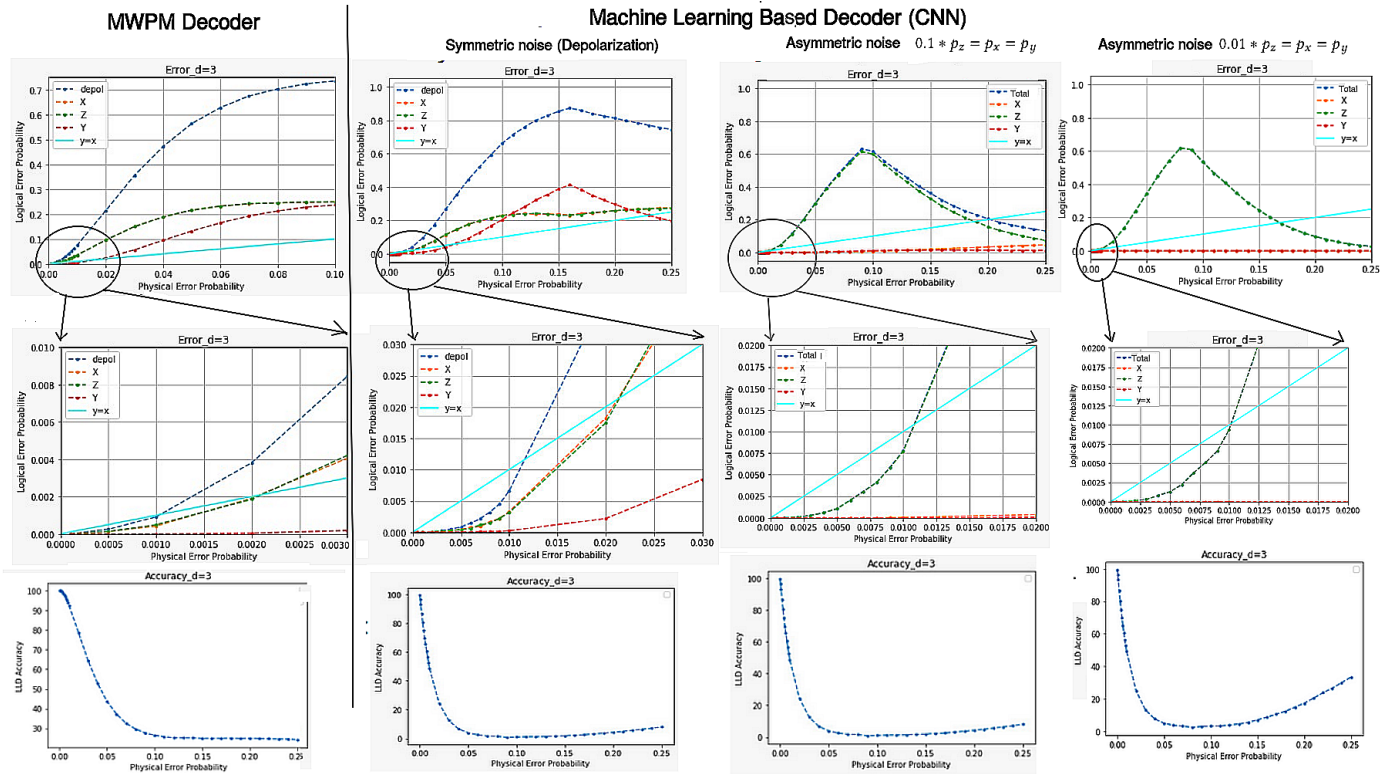
Figure 7: Pseudo-threshold and accuracy — MWPM vs ML-based decoder for distance 3 surface code
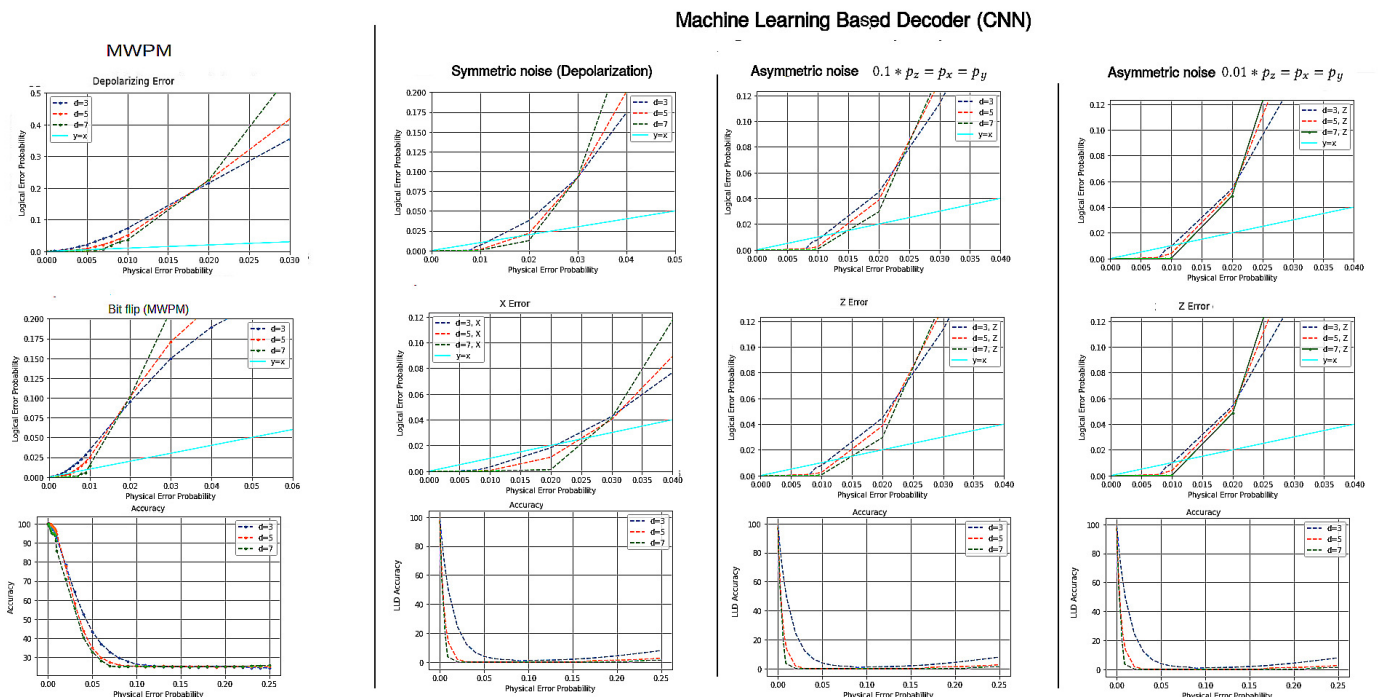

Figure 8: Threshold and accuracy — MWPM and ML-based decoder for surface code with $d = 3, 5, 7$

Table 1: Pseudo-Threshold of the low and high level decoders for distance 3, 5 and 7 surface code

Noise Model	→		Symmetric			Asymmetric					
			p_z p_x p_y			$0.1p_z$ p_x p_y			$0.01p_z$ p_x p_y		
Decoder ↓	$d \to$	3	5	7	3	5	7	3	5	7
MWPM	LLD	0.0011	0.0038	0.0075	0.0012	0.0041	0.0072	0.00098	0.0038	0.0067
	HLD	-	-	-	-	-	-	-	-	-
Our FFNN	LLD	0.012	0.0205	0.0219	0.0109	0.0121	0.0152	0.0120	0.0122	0.0131
	HLD	0.0143	0.0234	0.0241	0.0124	0.0164	0.0189	0.0123	0.0165	0.0189
Our CNN	LLD	0.0121	0.0211	0.0228	0.0112	0.0125	0.0151	0.0111	0.0121	0.0132
	HLD	0.0152	0.0241	0.0247	0.0134	0.0161	0.0192	0.0121	0.0162	0.0195

Table 2: Comparison of training times for different ML models

ML Model		d = 3			d = 5		
		Parameter space	Training time (sec)	Prediction time (sec)	Parameter space	Training time (sec)	Prediction time (sec)
FFNN	Simple	2258	53.12	2.1×10^{-5}	5618	103.18	3.5×10^{-5}
	Complex	84754	324.9	3.55×10^{-5}	88114	394.99	3.72×10^{-5}
CNN	Simple	165650	785.27	5.27×10^{-5}	429874	1852.74	7.4×10^{-5}
	Complex	240246	1485.69	6.02×10^{-5}	504370	4241.58	9.74×10^{-5}

Table 3: Comparison of Threshold of the low and high level decoders

Threshold (LLD)	Threshold (HLD)	Decoder Model	Error model
0.0181	N/A	MWPM	Symmetric
0.0302	0.035	Our FFNN	Symmetric
0.0218	0.025		Asymmetric $0.1 * p_z = p_x = p_y$
0.0221	0.0279		Asymmetric $0.07 * p_z = p_x = p_y$
0.0216	0.0257		Asymmetric $0.04 * p_z = p_x = p_y$
0.0213	0.0251		Asymmetric $0.01 * p_z = p_x = p_y$
0.0311	0.034	Our CNN	Symmetric
0.0225	0.026		Asymmetric $0.1 * p_z = p_x = p_y$
0.0229	0.0281		Asymmetric $0.07 * p_z = p_x = p_y$
0.0223	0.0258		Asymmetric $0.04 * p_z = p_x = p_y$
0.0212	0.0252		Asymmetric $0.01 * p_z = p_x = p_y$

with only Simple CNN and FFNN models.

5.4. Empirical train-test-ratio for optimal accuracy

In general, the higher the number of training samples, better is the accuracy of the ML model up to a certain threshold, beyond which increasing the number of training samples does not improve the performance of the model [31]. However, generation of training data is a humongous task in current quantum devices since it takes up a significant amount of device lifetime. Therefore, lower the size of the training sample required, higher is its usability. But naively reducing the size of the training set may lead to performance degradation. We explore this requirement by studying the minimum train-test-ratio required to obtain the optimal decoder performance.

Given a distance d code, with t types of errors possible on each qubit, the total number of distinct error patterns is t^{d^2}. When the probability of error is very low, most of the test cases will have no error. This observation is reflected in the curve [Fig 11] with $p = 0.001$. As p increases, it is natural that the performance of the decoder will degrade. However, when d is small, the total number of distinct errors is also small. Since the training and testing data is generated uniformly at random, we expect that for a given error probability p, the most likely error patterns are all exposed to the ML decoder for a reasonable size of training set.

To test this hypothesis, in Fig. 11 we have varied the train-test ratio for the simple CNN decoder for a distance 3 surface code. We originally generated 10^5 error data uniformly at random, and varied the train-test ratio starting from 90:10 and moving up to 10:90, lowering the training proportion by 10% in each step. The obtained accuracy for increasing p_{phys} is plotted in Fig. 11. We observe that
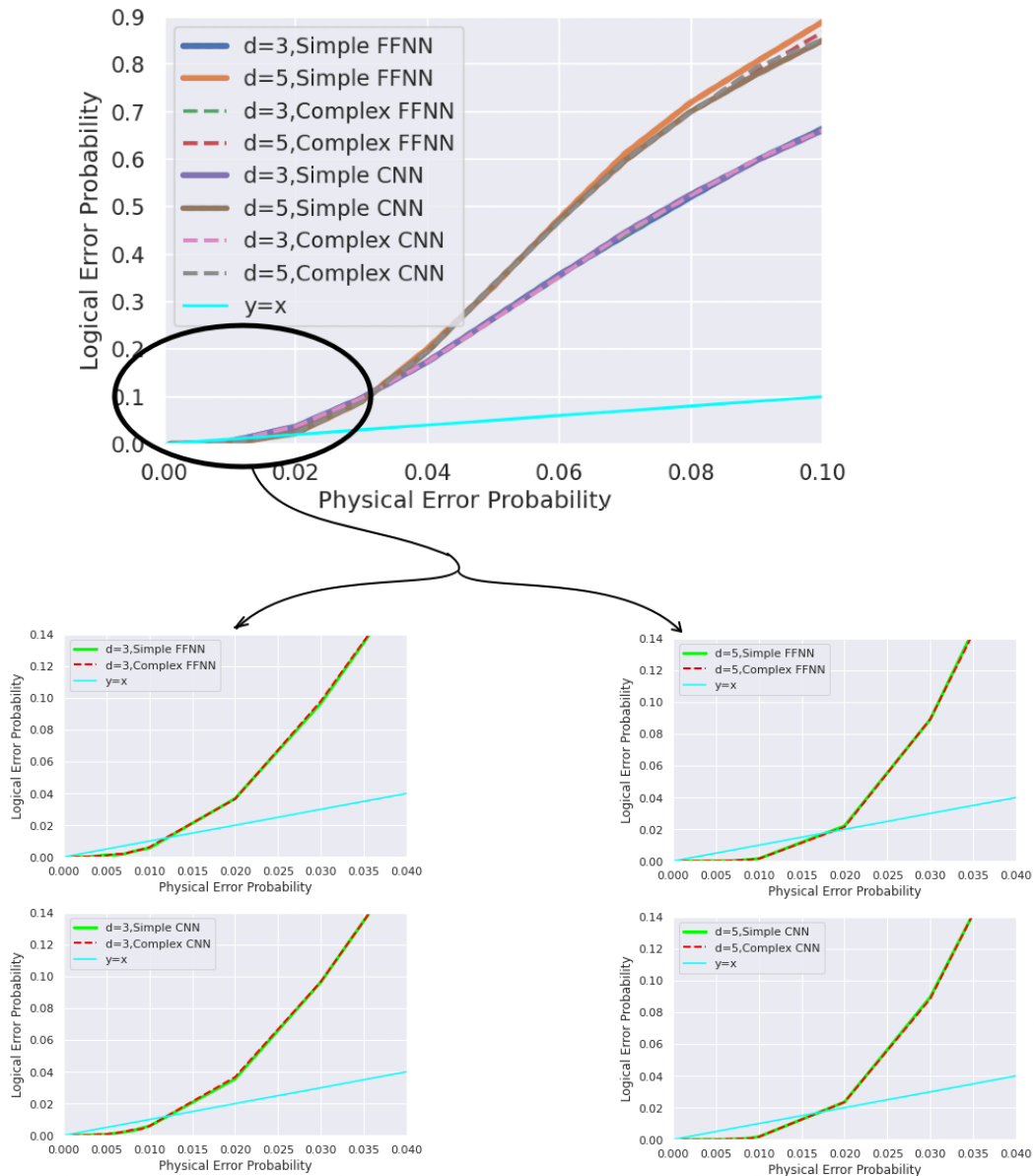
Figure 9: Logical vs physical error probability for various ML models in d= 3 and 5 surface code
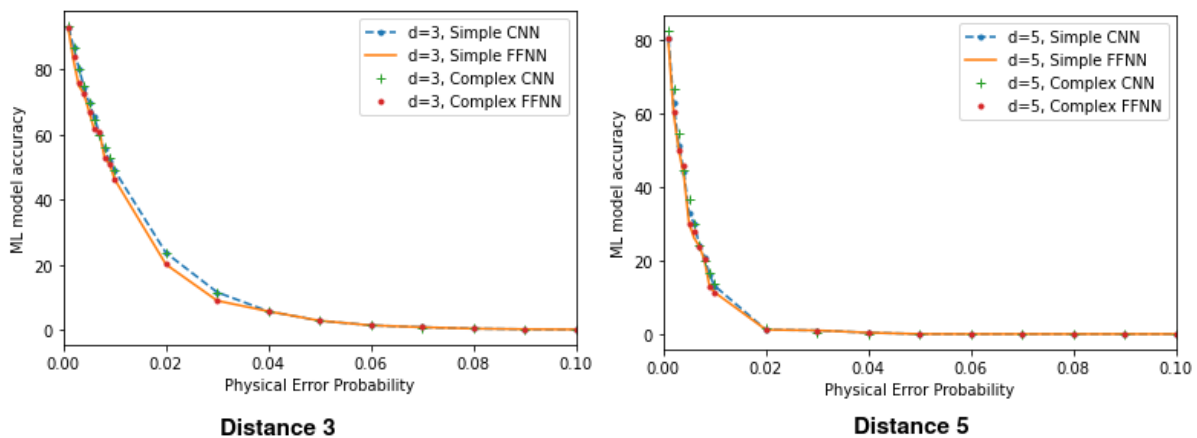

Figure 10: ML model accuracy vs physical error probability for various ML models in d = 3 and 5 surface code

even when we use up $\simeq 50 - 60\%$ of the data as test set, the performance of the decoder remains more or less constant for a given p. The performance takes a dip downwards beyond this value. Therefore, we posit that for $d = 3$ and $t = 4$ (depolarizing noise model), this decoder is exposed to all of the most likely errors within a small fraction of the training set, which is generated uniformly at random.

Since this is a ML based method, and Fig. 11(a) shows the mean value only, in Fig. 11(b) we have also plotted the standard deviation (SD) with a few values of physical error probability for all the test-ratio as an *error bar* plot. We observe that for p=0.001 the accuracy varies between 95.32 to 99.11 (min SD = 0.52, max SD = 1.40). For $p = 0.02$ the accuracy varies between 60.54 to 82.94 (min SD = 1.12, max SD = 2.79) and for $p = 0.08$ the accuracy varies between 32.51 to 51.92 (min SD = 0.28, max SD = 3.86). With increasing p_{phys}, the SD also increases. This supports intuition because as the p_{phys} increases, the decoding performance decreases due to the capacity of the machine learning model to correctly classify the errors. Hence, the performance of ML (which depends on the errors in the dataset), varies more with higher value of physical error probability (p_{phys}).

Moreover, as d increases, for same p, the set of probable errors increases exponentially. Therefore, for the same number of generated error data, we expect that to retain the same accuracy, a much larger portion of the data will need to be devoted for training. In our future research, we shall explore this direction in a more extensive manner and determine the size of the training sample required to retain an accuracy ϵ for a given train-test ratio.

5.5. Performance on Training with Symmetric Noise Models and Testing with Asymmetric Noise Models

As we have discussed, the real life noise models are asymmetric. But this asymmetry can change and we may not know the exact level of asymmetry beforehand always. It would be beneficial if the decoder can be trained once with symmetric noise dataset and tested with different asymmetric noise datasets. Now we analyze how the performance of a decoder trained with symmetric noise model behaves while testing with asymmetric noise model with increase in asymmetry (Δ).

An increase in asymmetry (Δ) in the depolarizing error channel is denoted by $p_x = p/(\Delta+2)$, $p_y = p/(\Delta+2)$, $p_z = p\Delta/(\Delta+2)$, where, p is the value of physical error rate for a given physical error. For example, $\Delta = 10$ denotes $0.1 * p_z = p_x = p_y$. For symmetric noise model, $\Delta = 1$ and for asymmetric noise model, $\Delta > 1$. We define *crossover* point to be the value of Δ beyond which increasing Δ leads to lower pseudo-threshold, when the decoder is trained with symmetric noise. We say that the channel is weakly asymmetric if $\Delta <$ cross-over point, strongly asymmetric otherwise. And we find the crossover point empirically.

As the asymmetry increases in the testing data, stepwise, we see a decrease in the performance of the decoder, as inferred from their decreased pseudo-threshold values. We check for Δ=1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 2, 10.
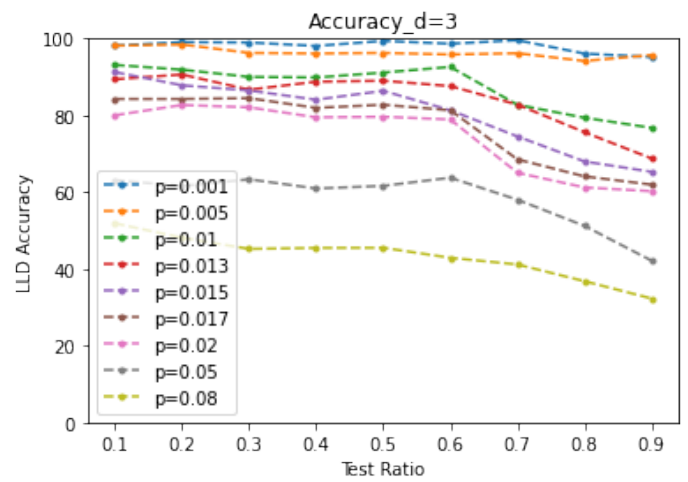
(a) Average accuracy of our low level decoder
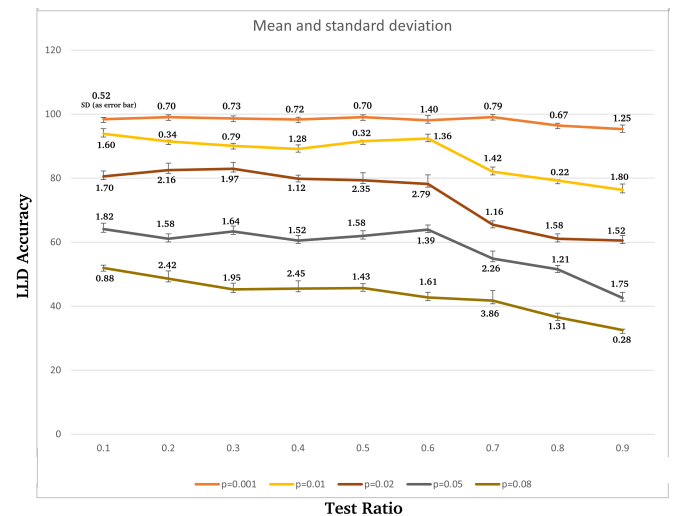
(b) Average Accuracy with its standard deviation as error bar of our low level decoder

Figure 11: Average accuracy (along with its standard deviation) of our low level decoder vs Test Ratio for different values of p_{phys} in distance 3 surface code

From Fig. 12 we can say that $\Delta = 1.3$ is the crossover point as upto this point, the logical error more or less overlaps with the symmetric noise model (i.e. Δ=1). Hence upto $\Delta = 1.3$, the channel is weakly asymmetric and above $\Delta = 1.3$ it is strongly asymmetric. Hence, upto $\Delta = 1.3$ we can train the decoder with symmetric noise model and test with the desired symmetric or asymmetric noise model, as it will not degrade the performance much. However, beyond this degree of asymmetry, the model must be trained with the asymmetric noise model if the optimum pseudo-threshold is to be obtained.

6. Conclusion

In this paper, we have proposed an ML-decoder to correct both symmetric and asymmetric depolarizing noise on surface codes. Our decoder has two levels — in the low-level it tries to accurately predict the error on the qubits, followed by the high level that tries to detect any logical error that
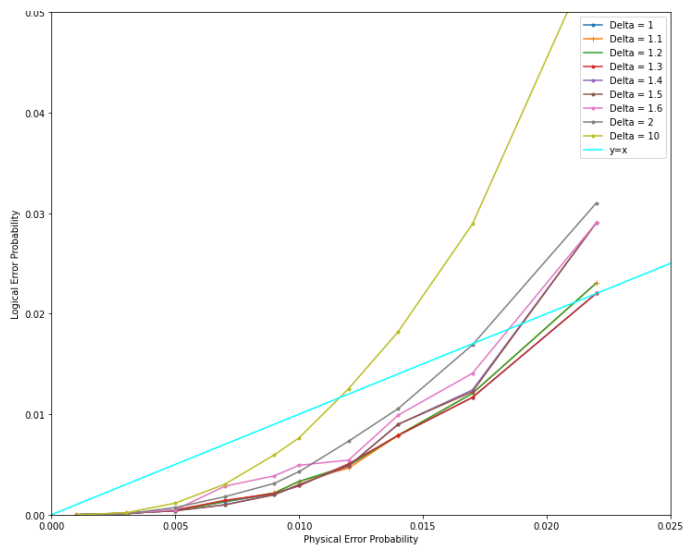
Figure 12: Training with Symmetric Noise Models and Testing with Asymmetric Noise Models

may have been introduced by the low-level decoder. Both these decoders have been implemented using neural network (FFNN and CNN) for surface code of distances 3, 5 and 7. Our proposed ML-decoder outperforms MWPM, and we observe ~ 2× increase in threshold and ~ 10× increase in pseudo threshold. We further show that the decoder performance is equally good for asymmetric errors as well, which is more realistic in quantum devices.

We have used ML models with different levels of sophistication, (i.e. varying number of hidden layers and node-density of each layer). Our results show that the mere increase of complexity in ML model requires an increased amount of time for decoding but hardly yields any better performance.

In this work, we have assumed, noise-free measure qubits and ideal measurements. A future prospect of this research can be to consider noisy measure qubits and imperfect measurements.

Data and Code availability

Data from the numerical simulations and the code can be made available upon reasonable request.

Conflict of Interest   The authors declare no conflict of interest.

References

[1] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", *SIAM J. Comput.*, vol. 26, no. 5, pp. 1484–1509, 1997, doi:10.1137/S0097539795293172.

[2] L. K. Grover, "A fast quantum mechanical algorithm for database search", "Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing", STOC '96, pp. 212–219, ACM, New York, NY, USA, 1996, doi:10.1145/237814.237866.

[3] A. M. Childs, N. Wiebe, "Hamiltonian simulation using linear combinations of unitary operations", *Quantum Info. Comput.*, vol. 12, no. 11–12, p. 901–924, 2012.

[4] J. R. McClean, J. Romero, R. Babbush, A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms", *New Journal of Physics*, vol. 18, no. 2, p. 023023, 2016.

[5] M. A. Nielsen, I. Chuang, *Quantum computation and quantum information*, AAPT, 2002.

[6] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", *Phys. Rev. A*, vol. 52, pp. R2493–R2496, 1995, doi: 10.1103/PhysRevA.52.R2493.

[7] A. M. Steane, "Error correcting codes in quantum theory", *Phys. Rev. Lett.*, vol. 77, pp. 793–797, 1996, doi:10.1103/PhysRevLett.77.793.

[8] R. Laflamme, C. Miquel, J. P. Paz, W. H. Zurek, "Perfect quantum error correcting code", *Phys. Rev. Lett.*, vol. 77, pp. 198–201, 1996, doi:10.1103/PhysRevLett.77.198.

[9] S. B. Bravyi, A. Y. Kitaev, "Quantum codes on a lattice with boundary", *arXiv preprint quant-ph/9811052*, 1998.

[10] E. Dennis, A. Kitaev, A. Landahl, J. Preskill, "Topological quantum memory", *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4452–4505, 2002.

[11] A. G. Fowler, A. M. Stephens, P. Groszkowski, "High-threshold universal quantum computation on the surface code", *Physical Review A*, vol. 80, no. 5, p. 052312, 2009.

[12] D. S. Wang, A. G. Fowler, L. C. Hollenberg, "Surface code quantum computing with error rates over 1%", *Physical Review A*, vol. 83, no. 2, p. 020302, 2011.

[13] L. Riesebos, X. Fu, S. Varsamopoulos, C. G. Almudever, K. Bertels, "Pauli frames for quantum computer architectures", "Proceedings of the 54th Annual Design Automation Conference 2017", pp. 1–6, 2017.

[14] A. G. Fowler, A. C. Whiteside, L. C. Hollenberg, "Towards practical classical processing for the surface code", *Physical review letters*, vol. 108, no. 18, p. 180501, 2012.

[15] R. Majumdar, S. Basu, P. Mukhopadhyay, S. Sur-Kolay, "Error tracing in linear and concatenated quantum circuits", *arXiv preprint arXiv:1612.08044*, 2016.

[16] J. Edmonds, "Paths, trees, and flowers", *Canadian Journal of Mathematics*, vol. 17, p. 449–467, 1965, doi:10.4153/CJM-1965-045-4.

[17] S. Varsamopoulos, K. Bertels, C. G. Almudever, "Comparing neural network based decoders for the surface code", *IEEE Transactions on Computers*, vol. 69, no. 2, pp. 300–311, 2019.

[18] C. Chamberland, P. Ronagh, "Deep neural decoders for near term fault-tolerant experiments", *Quantum Science and Technology*, vol. 3, no. 4, p. 044002, 2018.

[19] S. Varsamopoulos, B. Criger, K. Bertels, "Decoding small surface codes with feedforward neural networks", *Quantum Science and Technology*, vol. 3, no. 1, p. 015004, 2017.

[20] R. Sweke, M. S. Kesselring, E. P. van Nieuwenburg, J. Eisert, "Reinforcement learning decoders for fault-tolerant quantum computation", *Machine Learning: Science and Technology*, vol. 2, no. 2, p. 025005, 2020.

[21] L. Ioffe, M. Mézard, "Asymmetric quantum error-correcting codes", *Phys. Rev. A*, vol. 75, p. 032345, 2007, doi:10.1103/PhysRevA.75.032345.

[22] D. Gottesman, "Stabilizer codes and quantum error correction", *arXiv preprint quant-ph/9705052*, 1997.

[23] R. Hill, *A first course in coding theory*, Oxford University Press, 1986.

[24] 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.

[25] H.-Y. Tran, T.-M. Bui, T.-L. Pham, V.-H. Le, "An Evaluation of 2D Human Pose Estimation based on ResNet Back-", *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 59–67, 2022.

[26] M. A. Danlami, A. O. Oluwaseun, M. S. Adam, S. M. Abubakar, E. Ayobami, "Cascaded Keypoint Detection and Description for Object Recognition", *Journal of Engineering Research and Sciences*, vol. 1, no. 3, pp. 164–169, 2022.

[27] S. K. Pal, D. Bhoumik, D. Bhunia Chakraborty, "Granulated deep learning and z-numbers in motion detection and object recognition", *Neural Computing and Applications*, vol. 32, no. 21, pp. 16533–16548, 2020.

[28] A. V. Singhania, S. L. Mukherjee, R. Majumdar, A. Mehta, P. Banerjee, D. Bhoumik, "A machine learning based heuristic to predict the efficacy of online sale", "Emerging Technologies in Data Mining and Information Security", pp. 439–447, Springer, 2021.

[29] S. Varsamopoulos, K. Bertels, C. G. Almudever, "Designing neural network based decoders for surface codes", *arXiv preprint arXiv:1811.12456*, 2018.

[30] S. Krastanov, L. Jiang, "Deep neural network probabilistic decoder for stabilizer codes", *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.

[31] S. Shalev-Shwartz, S. Ben-David, *Understanding machine learning: From theory to algorithms*, Cambridge university press, 2014.

Debasmita Bhoumik has received the B.Sc degree in Computer Science from Bethune College in 2013. She has received the B.Tech degree from the University of Calcutta in 2016 and M.Tech degree from the University of Calcutta in 2018 in Computer Science and Engineering. She is currently pursuing her PhD degree in the "Application of machine learning techniques in quantum computing" from the Indian Statistical Institute.

Her research interests are in quantum computing, quantum error correction, quantum circuit placement, machine learning techniques. She is a Gold medalist in her B.Sc from Bethune College (Calcutta University).
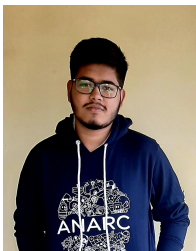
Pinaki Sen is a final year electrical engineering undergraduate at National Institute of Technology Agartala.

He is interested in Machine Learning and Quantum Computing. He has experience of working in Quantum-dot Cellular Automata as well. He has been volunteering as Technical lead of Girlscript, Agartala and previously volunteered as Vice President of Robotics Club, NIT Agartala.

Ritajit Majumdar is a PhD scholar at the Indian Statistical Institute. His research interests are in Noisy Quantum systems, and Near term Quantum Algorithm.

He was a Fulbright Nehru Doctoral Research Fellow at IBM Quantum, Thomas J. Watson Research Lab. He was a former Assistant Professor in the Department of Computer Science and Engineering, B. P. Poddar Institute of Management and Technology. He is a Gold medalist in M.Tech. from Calcutta University and a recipient of the DST Inspire Fellowship.

Susmita Sur-Kolay is currently a Professor at Indian Statistical Institute. She received the B.Tech. degree in Electronics and Electrical Communications Engineering from IIT Kharagpur, and the Ph.D. degree in Computer Science and Engineering from Jadavpur University. From 1980 to 1985, she was a Graduate Research Assistant in the Laboratory for Computer Science at Massachusetts Institute of Technology. In 1992, she was a Postdoctoral Fellow at the University of Nebraska–Lincoln. From 1993 to 1999, she was a Reader at Jadavpur University.

In 2002, she was a Visiting Faculty Member at Intel Corporation, USA and in 2012 as Visiting Researcher at Princeton University. Her current research interests include EDA for physical design and testing, hardware security, in-memory computation and synthesis of quantum computers. She was a Distinguished Visitor (India) of the IEEE Computer Society, an Associate Editor of the IEEE Transactions on Very Large Scale Integration (VLSI) Systems and ACM Transactions on Embedded Computing Systems. She is a recipient of the President of India Gold Medal, Distinguished Alumnus Award of IIT Kharagpur and Women in Technology Leadership Award of the VLSI Society of India.

Latesh Kumar K J is presently a Visiting Professor at the Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, USA. He obtained his Bachelors and Masters degrees in Computer Science and Ph.D in Computer Science and Engineering from AeU Malaysia.

He has published numerous papers in International Journals, Conference and Technical articles published by Springer, IEEE, ACM, Elsevier and IT next, NetApp Technical Library worldwide. He is a Subject Matter Expert in Network Security of IT companies. He has received several awards including distinguished "Customer Excellency for Technical Consultancy and Subject Matter Expert" at Hewlett Packard, and Best International Journal award at PSRC, Indonesia. He is currently researching and working in the Discovery Lab, FIU, USA on Security systems and Predominant work Quantum Key Distribution Techniques.

Sundaraja Sitharama Iyengar is currently the Distinguished University Professor, Ryder Professor of Computer Science and Director of the School of Computing and Information Sciences at Florida International University, Miami. He has been involved with research and education in high-performance intelligent systems, Data Science and Machine Learning Algorithms, Sensor Fusion, Data Mining, and Intelligent Systems. He has received his Ph.D. degree in 1974 from MSU, USA.

He has published more than 600 research papers, has authored/co-authored and edited 26 books. His h-index is 61 with over 18000 citations. His books are published by MIT Press, John Wiley and Sons, CRC Press, Prentice Hall,

Springer Verlag, IEEE Computer Society Press, etc. Dr. Iyengar has been a Visiting Professor or Scientist at Oak Ridge National Laboratory, Jet Propulsion Laboratory, Naval Research Laboratory, and has been awarded the Satish Dhawan Visiting Chaired Professorship at the Indian Institute of Science, the Homi Bhaba Visiting Chaired Professor (IGCAR), and a professorship at the University of Paris-Sorbonne.

Identification of Basic Respiratory Patterns for Disease-related Symptoms Through a Microphone Device

Amol M Khatkhate [*,1], **Varad Raut** [1], **Madhura Jadhav** [2], **Shreya Alva** [2], **Kalpesh Vichare** [1], **Ameya Nadkarni** [1]

[1] Department of Mechanical Engineering, Rizvi College of Engineering, Mumbai, 400050, India
[2] Department of Biotechnology Engineering, Rizvi College of Engineering, Mumbai, 400050, India
*Corresponding author: Amol M Khatkhate, +91-8007977166 & amolmk@eng.rizvi.edu.in (ORCID ID: 0000-0001-7025-1372)

ABSTRACT: The research in this paper focuses on the capture of respiratory patterns in real-time through a minor modification of an anti-snoring device that is readily available off the shelf or can be bought online and attaching a microphone to it. The audio pattern is recorded for the various lung conditions at different stages. The data recorded is generated in graphical form to understand the pattern. The setup for recording and generating data patterns is explained in the paper. The procedure for using the device and gathering data is mentioned in detail in this paper and sample raw data is presented for researchers. The purpose of this device is easy data collection and interpretation for self-diagnosis of the anomalies in the breathing pattern through audio data and display of the patterns in smartphone applications that can further be investigated by a physician.

KEYWORDS: Respiratory patterns, Anti-snoring device, Breathing pattern, Sensor lab, Audio Pattern

1. Introduction

Breathing is the most important respiratory activity which can change according to human health conditions. It is the amount of air that enters and exits with each respiratory cycle. A healthy human being experiences a stress-free Respiratory cycle. Inhaled oxygen is responsible for various bodily functions such as digestion of food, peaceful sleep, mental stability, improved body immune system, reduced stress levels, etc.

According to the health conditions the respiratory pattern changes along with the change in oxygen inhaled and the amount of carbon dioxide thrown out of the body.

During cough and cold, the Inhaled oxygen amount and its pattern change along with the change in voice disrupting the homeostasis. Usually cold, cough and throat conditions are symptoms of various infections, pulmonary problems, sleep apnea, and other minor health conditions that cause ineffective breathing, some major circumstances could be Heart failure, lymphomas, Paralysis, liver, and kidney problems, etc. Abnormal respiration can be in the form of hyperventilating due to stress, anxiety, a panic attack or due to heart or lung disease, or hypoventilation due to lung conditions like bronchitis or cystic fibrosis, or Obstructive sleep apnea [1]. This disruption in normal breathing patterns is disturbing and stressful hence, the Clearance of airways in the respiratory system is a priority, as it affects the oxygen supply to tissues.

Sometimes depending upon the breathing pattern, the audio pattern of voice during speaking, and coughing, problems can be detected. Generated patterns can be used to detect the affected region of the respiratory system and help in early diagnosis and treatment. These patterns are generated in the graphical form of a sinusoidal wave, where Rate, depth, timing, and consistency of breaths are all important to maintaining a delicate balance of respiration and metabolism. In case of illness, these patterns change typically to increased, decreased respiratory rate or clusters of rapid respiration.

In normal breathing conditions at rest, there is inhalation followed by exhalation and a pause of 1-2 seconds. While in case of abnormal breathing different patterns are formed according to respiratory problems.

The given Figure 1 and [2] show the pattern formed during various conditions and a description of what exactly can be causing it. As you can see the normal breathing pattern is a sinusoidal wave.

In this paper along with breathing patterns, we have also gathered patterns for a sore throat condition, cough, and chewing. This audio pattern can be used to examine underlying health conditions with an easy setup. We are trying to highlight the cheapest method that can be

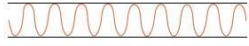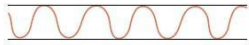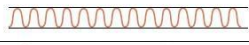further developed from an audio pattern to a graphical format to observe and understand for early diagnosis.
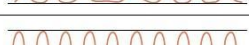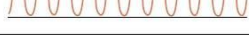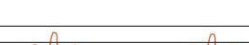
	Definition	Graphic Representation
Eupnea	Normal, breathing at 12-18 breaths/minute	
Bradypnea	Slower than normal rate (< 10 breaths/minute), with normal depth and regular rhythm	
Tachypnea	Rapid, shallow breathing > 24 breaths/minute	
Hypoventilation	Shallow, irregular breathing	
Hyperventilation	Increased rate and depth of breathing (called Kussmaul's respiration if caused by diabetic ketoacidosis)	
Apnea	Period of cessation of breathing. Time duration varies; apnea may occur briefly during other breathing disorders, such as with sleep apnea. Life threatening if sustained.	
Cheyne-Stokes	Regular cycle where the rate and depth of breathing increase, then decrease until apnea (usually about 20 seconds) occurs.	
Biot's respiration	Periods of normal breathing (3-4 breaths) followed by a varying period of apnea (usually 10 seconds to 1 minute).	

Figure 1: Graphical representation of breathing patterns for various health conditions. (Brokenshire.weebly.com)

We recorded the breathing pattern using a device with an easy setup from scratch. We have used readily available materials such as anti-snore devices and microphones. Audio files are created as raw data to record audio patterns for speaking, coughing, chewing, and breathing from a healthy person and with a sore throat.

A mobile application namely Sensor Lab is used which can generate an audio file that can be transformed into graphical patterns for better examination. This application can be downloaded easily on android as well as iOS systems.

Figure 2: Anti Snoring device

Anti Snoring device [3] Figure 2 fits properly in the nose with no gaps for the leak. It essentially is used for people with sleep apnea in which case it prevents deep sleep due to snoring. To maintain continuous positive air pressure in the respiration this device is used to provide enough extra air pressure to prevent soft tissue from partially obstructing airflow.

Through this paper, we are trying to promote the idea of regularly observing the breathing pattern by recording it into audio form and generating and storing data for comparison.

A habit of recording data over a specific time interval can result in early detection of any problem and maintaining good health conditions. Nowadays people have a variety of health conditions and data collected using wearables during sports mode has become regular. Similarly, the breathing pattern and audio pattern for cough, throat, and chewing can be noticed periodically or regularly to maintain a healthy lifestyle.

The data can be saved successfully in the form of an audio file or graphical pattern for future references. Through this, the detection of abnormal breathing patterns, cough sounds, and throat conditions can be detected for early diagnosis and treatment easily anytime anywhere. The setup of detail vice enables the user to record data in a fraction of a minute.

Users can use this setup to record breathing patterns while performing sports, at rest, during sleep, etc. Depending upon the graph he/she can seek further doctor consultation over it.

This could be the cheapest and easiest way to understand and generate data on breathing patterns from audio files.

2. Breathing pattern sensors

Breathing patterns can be monitored by pressure sensors that can be attached as a belt to the ribcage, humidity sensor that is attached to a mask, oximetry sensor to measure oxygen saturation and hence, detect respiratory anomalies due to decreased oxygen intake, acceleration sensors that monitor muscular contraction during breathing and are attached to the chest wall, resistive sensors, the thoracic or abdominal expansion, and relaxation manipulate the resistance of the stretch sensor [4].

Lung sounds can be monitored by acoustic sensors as well [4], and this paper intends to explore this sensing technique due to its convenience and availability.

3. Design of Anti-snoring air-purifying Device

The proposed device for sampling input is a simple modification to the commercially available anti-snoring device on the market. The detailed diagram of the anti-snoring device is shown in Figure 2.

3.1. Description of the original device

The anti-snoring device as seen in Figure 2, consists of a medical-grade silicone and plastic insertion module that snugly fits in the nose of the person and an activated

carbon filter to purify the air while breathing to remove contaminants. The manufacturer claims that the device also does PM2.5 filtering [1].

3.2. Working on the original device

Some types of snoring are caused due to obstruction of the nasal passage. The anti-snore device as shown in Figure 2 claims to open this restriction by clearing the air and purifying it at the same time thereby maximizing the airflow and eliminating the need to gasp for air. However, there is insufficient data to reinforce its efficiency in eliminating snoring completely, as there are several causes including nasal congestion, obesity, sleep apnea, anatomy, sleeping position, and consumption of relaxants such as alcohol, etc [2] and this device fails to address other causes of snoring. But this device can certainly be modified for another purpose.

3.3. Cleaning mechanism of the modified device

The modified device can be opened using a minus-type screwdriver. The device needs to be cleaned and sanitized after usage to avoid infection. To clean the device, prepare a solution of soap. Cleaning can be carried out by immersing a cotton swab and wiping or dipping the silicon-plastic device in this cleaning solution to ensure no allergen or pathogen triggers an immune response. This will also ensure the drastic reduction of any kind of viral load, including SARS-COV2 on the inner surface of the device. Then it needs to be rinsed with water. The cleansing solution can be stored in a spray bottle and can be used intermittently. Make sure to remove the activated charcoal filter before cleaning can be attached and assembled once the device is cleaned and dry [5].

3.4. Design and working of the modified device for breathing input

A] Device setup

The device as seen in Figure 3, has two frontal openings that are to be inserted into the nostrils and one opening at the top for airway passage. The steps for assembly are as follows:

- The device can be opened from the inside by removal of two + screws using a screwdriver. This will separate the silicone module from the plastic casing.
- The original device is dismantled, and an activated carbon filter is added, due to its effectiveness as a barrier or an entrapment against pathogens as small as a virus with sizes in nanometers [5].

- The device is reassembled and is fitted into a cloth mask or surgical mask, that is easily available and can be discarded after use [6].
- A small microphone is attached to the mask to capture audio signals. This microphone can be attached to any smart device like a smartphone, computer, or laptop via a wired/ wireless network such as Bluetooth.
- Applications like sensorlab need to be downloaded to visualize the interpreted data.

Figure 3 – Assembly of the mask with an Anti-snoring device

B] Data interpretation

- Connect your microphone to your smart device and open the sensor lab application to sample your audio data while making a sound of coughing or sneezing. You can also sample your data during the sore throat and normal conditions and compare the graph in the form of waves.
- Ensure that the microphone is at least 1-2 cm away from your modified anti-snoring device to get a clear sound. Record the audio for 1 minute.
- The data is generated in the form of audio signals and the application saves this data anywhere on your device in a .zip file extension.
- To locate the file, rename the file and upload it to your drive, cloud, or in an accessible location on your device.
- Now the recorded sample needs to be interpreted in the form of a wave graph. For this, we need to open the file in a CSV file extension. This would allow data to be saved in a tabular format in Microsoft excel.
- This tabular data can be plotted as a wave graph using a chart tool in excel to display the final pattern with the crests and troughs for a given sample.
- The method that we adopted was manual but there is a scope for automation to directly take the recorded sample as an input and generate the patterns and scalar value of the anomaly as an output. This can be done by using various software such as MATLAB.

- Machine learning algorithms can be used in the software that would collect the samples for each respiratory condition in individuals and predict the outcome. The algorithms that perform as per the calculations can be incorporated into the sensor lab application as well.

3.5. Audio data recording for Breathing pattern

The anti-snoring device combined with a microphone attached to a smartphone for real-time data collection and post-processing to give levels of anomalies in respiratory patterns of individuals like sore throat, cough, and cold which can act as precursors to the disease. Also, in addition to this, accelerometer data can be gathered using a mobile application namely Sensor Lab, and correlated with the respiration patterns to indicate other symptoms like shortness of breath, etc.

A sensor lab is used to generate the audio data patterns. The recorded audio files in CSV format can be used with the sensor lab. The sensor lab has an accelerometer that is used to measure the vibration of motion, it has a transducer that converts mechanical force caused by vibration or change in motion into electric current using the piezoelectric effect.
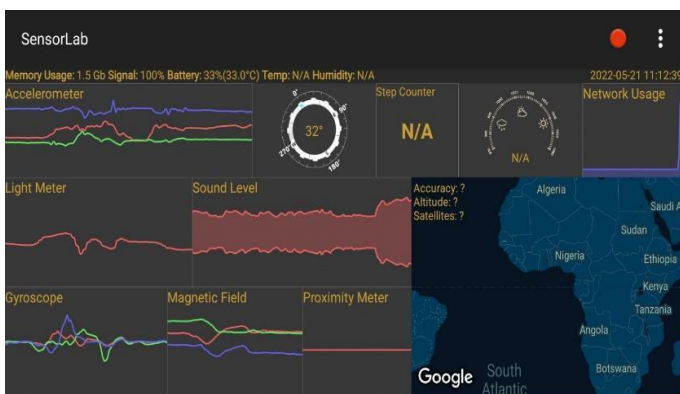
Figure 4: a screenshot of the Sensor lab mobile application. This dashboard appears as soon as the user opens the application. For audio recording simply click on sound level and the sound meter will open up.
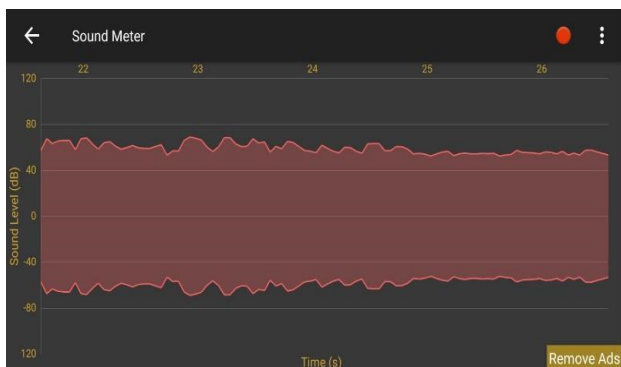
Figure 5: This window opens up once a user clicks on the Sound lever tab. Audio recording starts by clicking the red circle on the top right corner. The red circle turns into a white square box and on the left-hand top corner timer starts along with it. To stop the recording click on the white square box.

Installation of sensor lab on android:

1. Open google play and search for sensorlab.
2. Look for the 1.3 sensor lab version offered by LP Ellis
3. Click on the install button to initiate the download
4. Open the application and start using it.

The Sensor lab app is initiated after installation in the smartphone and must record the acceleration data along with the sound data. This selection must be made by going into settings and checking the box next to the sound, accelerometer. Other data may also be recorded and stored in CSV format for offline analysis.

4. Sample Data Recorded in Audio pattern and charts

Figure 2 shows blank data generated with no sound. Usually, this blank is recorded as a reference or control before any further data collection.
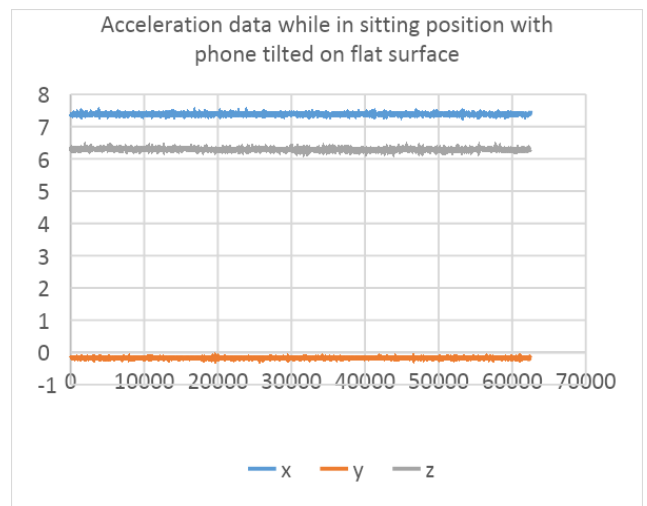

Figure 6: Data recorded by the accelerometer in the Sensor lab in a sitting position.
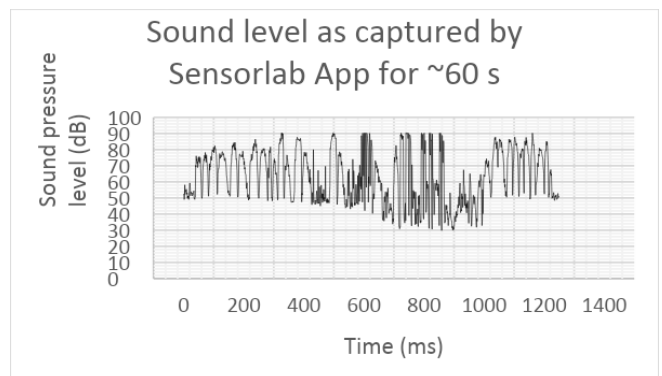

Figure 7: Breathing sample captured from the subject using the Sensor lab App

The given pattern is breathing data recorded by the user in a sitting position at rest. The user simply held the device close to the nose while breathing for 1 min. The graph shows sinusoidal waves which are very much like each other. There is no disturbance in the wave pattern recorded for a 1 min time duration. From this Figure 7, we

can easily interpret the user being a normal healthy human being at rest whose respiratory pattern is undisturbed.
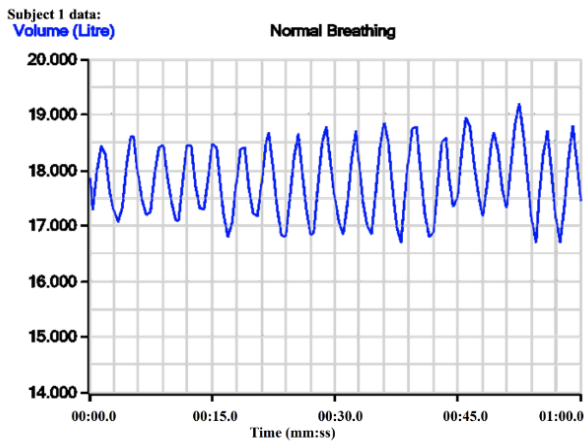

Figure 8: Normal breathing pattern (sample)

Users can record breathing patterns periodically to compare the pattern generated. Depending upon the stored data user can understand if any anomalies are detected in the waveform. Figure 8 (taken from Chegg.com as a reference or for pictorial representation) clearly explains the breathing pattern. Normal breathing is a sinusoidal wave with a crest and trough representing exhalation and inhalation respectively.
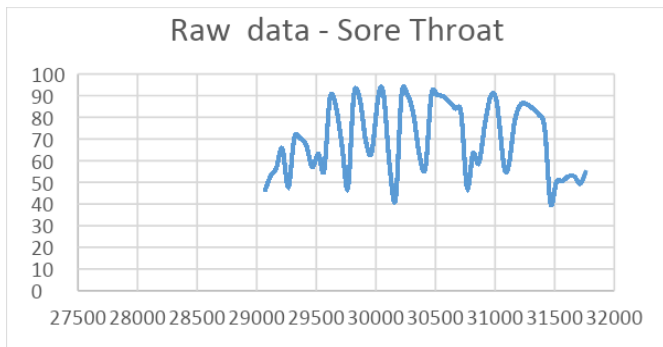
Figure 9: Sample data for a user speaking in sore throat condition

During a sore throat, the quality of the voice becomes gravelly-sounding and sometimes too quiet or soft to hear. The vocal cords swell up which affects the way they vibrate and cause hoarseness. The pattern generated for sore throat must change pitch.
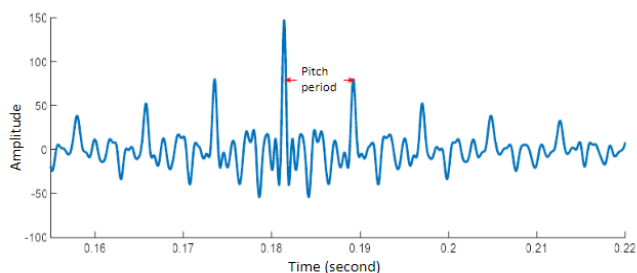
Figure 10: Reference image for Sore throat from Research gate by author Hendra Setiawan, Sutono Sutono.

It can be seen in Figures 9 & 10 [7] pattern in the graphical format generated for sore throat fluctuates considerably. The pitch varies repetitively from high sharp to low pitch. A broken or cracking sound pattern sometimes flat pitch can also be observed. This pattern generated can be used to detect the problem involving the throat.

Figure 11: Audio Pattern recording for sore throat cough.

Figure 11 (taken from the website Pond5) is an audio pattern for human male sore throat coughing that differs from Figure 13 representing normal coughing. The difference can be seen at the end of coughing with a sound, and a sharp fluctuation is seen at the end of the voiced region. Also the difference between the explosion region, intermediate region, and voiced region is unclear in Figure 11. A sore throat can be caused due to smoking, acid reflux, allergies, infections, etc. could be the reason that can be used for prediction.

Early detection of sore throat is possible through this cheap and readily available device set up by observing the pattern generated.
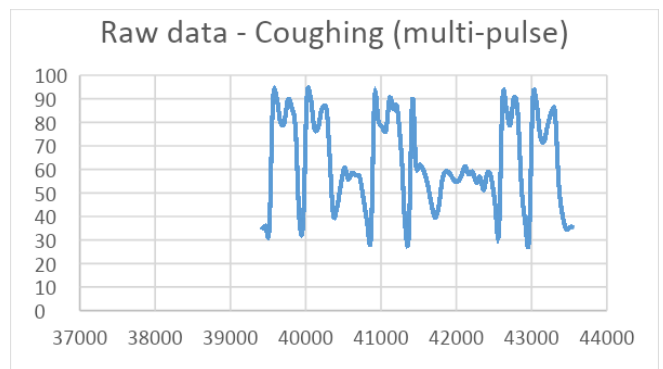

Figure 12: Sample data for normal coughing.

Figure 12 shows the interpreted audio pattern in the graphical one and Figure 13 shows the audio pattern recorded by ERS Publication [8]. In Figure 12 one can notice easily that a square wave is formed. The crest and trough region together forms a square pattern as it is a thick exhalation activity. Coughing is a louder sound with very low inhalation during the period. A small trough formed between crests represents a small inhalation time needed to complete the coughing.
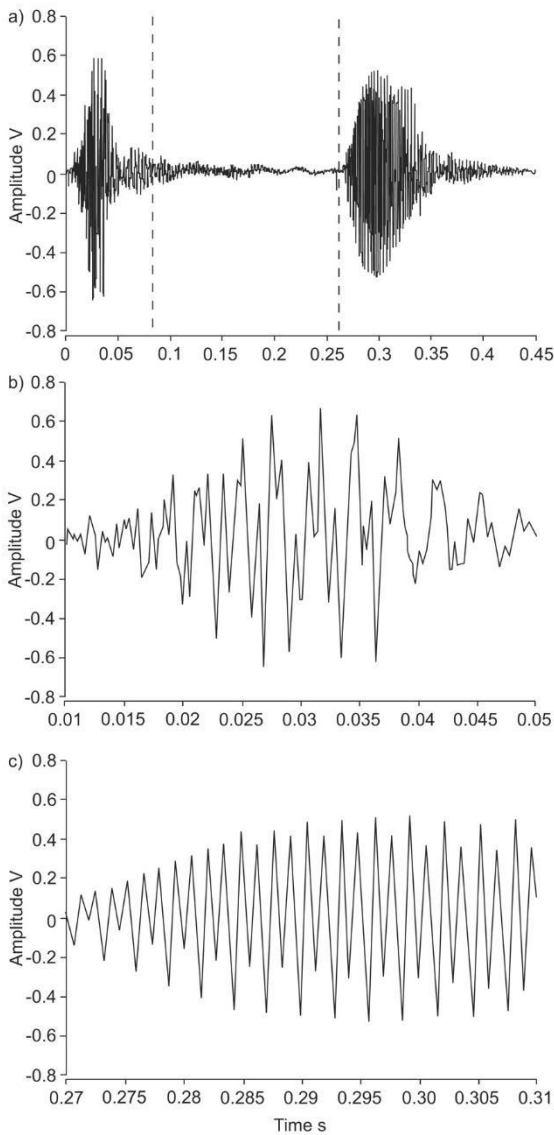
Figure 13: Audio pattern for coughing taken from ERS publication as reference for better understanding

The audio waveform in Figure 13 can be divided into namely explosive (left) which is an enlarged or expanded wave pattern that indicates noise, intermediate (center indicating gate break in the first cough exhale, and inhaling time before the next cough, voiced (right) indicate periodic appearance.
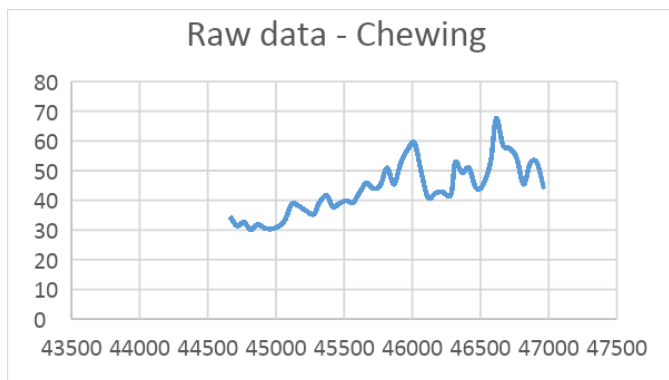

Figure 14: Sample data for chewing

Such patterns can indicate the extent to which cough is present in the throat or lungs. Cough in the lungs called

bronchitis sounds whistling with the noise. So, depending upon sound and audio patterns we can detect the cough condition. We can compare the two images Figure 11 and Figure 13 to spot and study the difference between coughing and sore throat conditions and coughing in normal conditions.
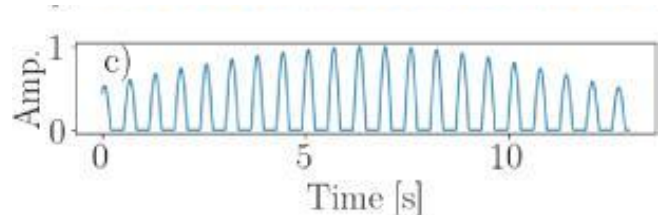

Figure 15: chewing pattern taken from ACM DL

We have also recorded data patterns for Chewing sounds Figure 14. This graph appears to be different from other sound patterns. Chewing in the initial period is slow because of big food bites and becomes steady after the breakdown of big bites into smaller chunks so fluctuation can be seen at the center in Figure 14 and the pattern then starts to move downward representing the gulping of food after a considerable breakdown. In figure 15 [9] it can be observed the sound pattern increases in the initial period, is stable at the center, and starts decreasing towards the end of chewing.

The reason to record chewing patterns is to observe healthy dietary means. The digestion process starts right from your mouth when a person chews. Saliva moistens food and its enzymes break down starches. It is important to chew food properly as it is a healthy activity and helps in better digestion and absorption.

Some people swallow food quickly with no significant chewing. Such activity is not healthy and should be given attention. Recording such activity can increase awareness about the chewing period for better digestion and absorption.

5. Calculation of intensity of anomalies in respiratory patterns using symbol sequences

- The anomalies in respiratory patterns like cough, sore throat, etc, and other patterns like chewing are identified using Sensorlab App data in Dbl.
- The data in DBL is ranging from 0 – 128 and hence, can be represented using 7 bits. Use 1 bit for sign, 7 bit for the pre-decimal part, 1 bit for decimal, and 7 bit for 2 digits after decimal which gives us a string of 16 bits.
- For each value of DBL, calculate the symbol sequence and concatenate each sequence with the previous time version of the DBL sequence. This will then form a string of symbol sequences.
- Convert the string of 0 and 1 into a finite state machine of alphabet size A = 2 and depth D, which is chosen to decide the complexity of the machine.

- Find the probabilities of visits of all the states in the finite state automata and compute the probability vector.
- Use a suitable norm to identify the intensity of the anomalies in the patterns like the intensity of cough, cold, and other plausible symptoms.

Sample calculation

A single value of DBL is parsed into a symbol sequence as shown below and a sample calculation of probability for a 2-state machine (0 and 1) is carried out for the understanding of the readers. The analysis is carried out for a singleton data point, however, in reality, a window of suitable size must be chosen (for eg; 100 data points, in this case, is a likely window) and then the above procedure must be followed to quantify the anomalies. As seen in Table 1, the value of the pressure level P1 is converted to binary after rounding up to 2 digits The probability of occurrence of 0 = 11/16 = 0.6875 and the probability of occurrence of 1 is 5/16 = 0.3125

Table 1 : Conversion of dB to binary string for computation

k	time	dB (SPL)	P1(μpa)	Round up to 2 digits	Binary(P1) after rounding up to 2 digits	0 in 16 bit	1 in 16 bit
0	15	51.90992	787.9995975	787.99	**0000001100010011**	11	5

A similar analysis has to be carried out after each time interval of 100 sample points in column 1 of the data stored via the Sensorlab app.

The results shown via the above analysis indicate that the potential anomalies can be classified and the intensity of the same can be quantified using the above procedure.

The studies conducted by the authors also imply that gaps (as caused by an improper fit of the mask) can result in over a 60% decrease in the filtration efficiency, implying the need for future cloth mask design studies to take into account issues of "fit" and "leakage", while allowing the exhaled air to vent efficiently.

To overcome this drawback, the proposed design in this paper of the device allows the nose to be directly exposed to the environment and breathing to take place through the device itself rather than the mask. The inhalation can occur free from the frontal end and exhalation can take place from the top end. As the device snugly fits into the nose, it will eliminate the problem of gaps to a large extent. This will lead to almost zero leakage and hence, filtration efficiencies will increase dramatically as well as pressure drop due to leakages will be minimal.

6. Results

It is clear from the given information that Audio patterns and graphical patterns can be recorded easily by setting up the device from the readily available materials. The setup hardly takes 1 min and recording the audio files in CSV extension can be done using the Mobile application Sensor lab.

The breathing pattern for the normal condition shown in Figures 7 & 8 clearly shows sinusoidal waves with crest and trough. This data pattern was extracted directly from the Sensor lab in the audio format followed by transformation into Graphical format through Programming (automation). Figures 9 & 10 Show sore throat, and coughing patterns. One can observe the change in the waveform in case of sore throat and coughing. Figure 14 Chewing pattern which can also be observed and examined through the line chart. Similarly, sleep respiratory rate can also be examined by timing the modified device to record data after certain time intervals.

All of the sample recordings clearly show the differences in a pattern according to sound captured in a specific time interval.

The acoustic type sensing of the breathing patterns can be manipulated or impeded by external perturbations and care must be taken while sampling the data. One needs to ensure that they sample their data in a place where they are immune to any kind of auditory disturbances.

While sampling, the microphone also has to be placed at least 1-2 cm away from the modified device to avoid crackling noises which can result in a disturbing pattern.

More data is required from individuals that can be grouped into one category of symptoms related to lung conditions and run the algorithm to generate the probabilities. This data can be used by pulmonologists and cardiologists based on which they can suggest further tests or diagnose a particular anomaly, briskly.

7. Conclusions, Salient features, and Future work

The paper proposes a respiratory data collection procedure and sample data from a modification to the existing anti-snoring device available in the Indian market at the link [4].

The salient features of the work presented in this paper are given below

- The device snugly fits the nose and hence will not cause any disruption to normal breathing.
- The device very well fits the microphone and can be used along with the Sensorlab App available on the play store.

- The device also overcomes the drawbacks of masks like gaps and leaks and eliminates recirculation of CO_2.

- The respiratory patterns are gathered along with accelerometer data which is necessary to know the current body movements of the individual and to know if the person is walking, running, jogging, or doing heavy exercise.

- The sample data is recorded without any voice or any speech during the trial. The patterns are distinguishable and can be classified into various categories of sore throat, cough, and chewing.

- The sample data shown in the Figure above indicates that the respiratory patterns can very well be identified and distinguished using the microphone data. The patterns can be classified into various categories of sore throat, cough, and chewing with their intensities.

More such applications can be created with a highly efficient design of the device that can cut down the disturbance while recording the audio data, interpreting the data, and directly one day diagnose the condition or assist the physicians with treatment. Additions to the features of this system may enable a person to one day diagnose cardiovascular diseases or even concussions. There is tremendous potential for this technology in the future, as scientists are already making innovations and are looking for inspiration in the healthcare sector. Machine learning algorithms have proven to be reliable for more accurate predictions and this system can be implemented universally without any inconvenience.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] D. Starovoytova, "Snoring and its- management (Part 1/2): A review", Innovative Systems Design and Engineering, ISSN 2222-1727 (paper) ISSN 2222-2871 (online), vol. 9, no.12, 2018.

[2] I. Wheatley," Respiratory rate 4: breathing rhythm and chest movement" *Nursing Times*; 114: 9, 49-50,2018.

[3] A. Khatkhate, K. Shirsat, N. Barai, N. Sawant, N. M., M. P. Singh Raj, S.Kumari, " Study of effective respiration through an anti-snoring device modified to restrict the entry of the novel Coronavirus (SARS-Cov2)", 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), 2020, DOI: 10.1109/B-HTC50970.2020.9297985

[4] T. D. D. Costa, M. De F. F. Vara, C. S. Cristino, T. Z. Zanelle, G. N. N. Neto, P. Nohama, "Breathing monitoring and pattern recognition with wearable sensors", Wearable Devices - the Big Wave of Innovation, 2019, DOI: 10.5772/intechopen.85460.

[5] M. S. Reza, ABM K. Hasan, S. Afroze, M. S. A. Bakar, J. Taweekun, A. K. Azad, "Analysis on preparation, application, and recycling of activated carbon to aid in COVID-19 protection", International Journal of Integrated Engineering, vol. 12, no. 5, pp. 233-244, 2020 DOI: 10.30880/ijje.2020.12.05.029.

[6] A. Konda, A. Prakash, G. A. Moss, M. Schmoldt, G. D. Grant, S. Guha, "Aerosol filtration efficiency of common fabrics used in Respiratory cloth masks", ACS Nano, vol. 14, no. 5, pp. 6339-6347, 2020, DOI: 10.1021/acsnano.0c0325

[7] H. Setiawan, S. Sutono, "The Effect of Sore Throat on Changes of Vowel Sounds," ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, vol. 9, no. 1, p. 87, Jan. 2021, doi: 10.26760/elkomika.v9i1.87.

[8] A. Kelsall et al., "How to quantify coughing: correlations with quality of life in chronic cough," European Respiratory Journal, vol. 32, no. 1, pp. 175–179, Mar. 2008, doi: 10.1183/09031936.00101307.

[9] D. Kopyto, R. Zhang, and O. Amft, "Audio-Based Onset Detection applied to Chewing Cycle Segmentation," 2021 International Symposium on Wearable Computers, Sep. 2021, doi: 10.1145/3460421.3478819.

Amol Khatkhate completed his bachelor's degree from VJTI, Mumbai in 2001. He completed his doctoral degree from Pennsylvania State University in 2006. He has been working as a faculty with Rizvi College of Engineering since 2015.

His research interests are in mechatronics, robotics, and control systems. His publications can be accessed at Google scholar and have 150+ citations to his name.

Varad Raut is a graduate student from Rizvi College of Engineering completing his bachelor's degree in Mechanical Engineering. He is currently in his third year. He has won an award in a national-level competition. His third-year project on Innovative masks is currently in progress and expected to be showcased at exhibitions and workshops.

His research interests are in mechatronics and he has excellent creative skills as a mechanical engineer.

Madhura Jadhav completed her Bachelor's degree in Biotechnology Engineering from Rizvi College of Engineering in 2021. She worked in bioplastics as her final year project. She continued this project further and received incubation for her work at the Rizvi COE R&D Incubation center.

She aims to work in the future on the environmental pollution problem

Shreya Alva has done her bachelor's of engineering from Rizvi College of Engineering in 2021. Her research and training experience includes Food technology (Health-Eatos, Mumbai) R&D intern, Chemtech Drug discovery (RASA Life Sciences, Pune)

She has completed her final year project on analyzing components for preparing bioplastics with spent coffee grounds.

Kalpesh Vichare is a graduate student from Rizvi College of Engineering completing his bachelor's degree in Mechanical Engineering. He is currently in his third year. He has participated in various technical competitions and has won awards in national-level competitions. His third-year project on Innovative masks is currently in progress and expected to be showcased at exhibitions and workshops.

His research interests are in mechatronics, finite element analysis, and control systems.

Ameya Nadkarni completed his bachelor's degree from Rizvi College of Engineering, the University of Mumbai in 2009. He completed his Master's degree from VJTI, Mumbai in 2015. His research interests are engineering mechanics, machine design, CAD, and Additive manufacturing.

He is serving as Head of the Mechanical Department at Rizvi COE and is a vibrant faculty guiding students in areas of design.

JENRS

Blockchain Based Framework for Securing Students' Records

Omega Sarjiyus* **, Israel Isaiah**

Department of Computer Science, Adamawa State University, Mubi Nigeria
* Corresponding author: Omega Sarjiyus, Adamawa state university, mubi, sarjiyus@gmail.com

ABSTRACT: Right now, colleges, as focuses of exploration and development, coordinate in their cycles different advances that permit further developing administrations and cycles for their individuals. Among the inventive innovations is the Web of Things that permits getting information from the climate and individuals through various gadgets. Security of information on college grounds is expected to shield basic information and data from unapproved parties. One way of ensuring information is to apply the study of blockchain innovation to perform information encryption. There is a wide assortment of calculations utilized to encrypt information; however, this exploration centers around the SHA-256 encryption. Subsequently, this examination is pointed toward fostering a blockchain-based structure framework for getting understudies' records. These outcomes lead to further developing cycles and settling on better choices that further develop the administrations accessible at the college. Blockchain innovation is otherwise called appropriated record innovation, and It permits members to get the settlement of exchanges, accomplish the exchange, and move resources for a minimal price. In this manner, an examination will be done on the blockchain empowered college framework for getting understudies' records. The framework advancement philosophy taken on to improve this framework is the cascade procedure and Laravel for programming. Blockchain innovation can assist clients with putting away understudy records securely.

KEYWORDS: Blockchain, Encryption, Framework, Records, Security

1. Introduction

As of now, the utilization of arising advancements has dramatically filled in the public arena, and this has permitted working on numerous parts of personal satisfaction [1]. The improvement that these advancements have carried is predominantly because of the entrance of the Web, which has made enormous enterprises support the plan and making of gadgets that cooperate with individuals or with one another, using the Web [2]. These gadgets have better execution, and their dynamic correspondence with individuals considers their association in most everyday exercises [3]. Indeed, a few of these gadgets are not centered around close-to-home use; however, all things considered, they are centered around taking care of issues in conditions where hundreds or thousands of individuals reside. The speed at which society works brings about innovations that perform exercises that individuals would regularly prefer not to do themselves [4]. Starting here of view, the obligation of controlling the exhibition of an action shared with the gadgets should be possible distantly. These qualities of the gadgets make them part of the Web of Things (IoT) [5]. The IoT permits the association of

numerous gadgets to the organization through sensors and actuators that play out an assortment of errands [6]. These assignments can be intended for homegrown use or in exceptionally enormous conditions, such as grounds or urban communities, where their fundamental task is gathering data from the climate and playing out an activity [7]. A proof of this is home mechanization, which has customarily required private innovation to work appropriately. With the utilization of IoT, this methodology has changed. Presently, such stages are accessible, and advancement is centered around the improvement of uses that utilize the administrations of different stages to perform more intricate errands [8]. The IoT has situated itself as one of the mainstays of the advanced climate. This way, conventional conditions use data innovation (I.T.), remembered for designs that work close by, arising advancements to give the climate some knowledge [9]. A progressively apparent model is keen urban communities. These large-scale conditions have an endless number of administrations that need IoT to work on their utilization, particularly in regions like versatility and security [6].

What's more, ideas, for example, human digital actual frameworks to screen, control, and oversee conditions proficiently, are coordinated determined to guarantee sufficient conditions at a lower cost [10,11]. These frameworks are worked from the straightforward coordination of physical and computational parts. These frameworks beat the present straightforward coordinated frameworks as far as limiting flexibility, versatility, strength, security, and ease of use [12].

IoT deals with the administrations of a climate through its design, which is underlying layers and given the assortment of data. The gadgets consistently sense the medium, producing a huge volume of information that goes about as a natural substance for distributed computing [13]. The activity of IoT, even though it presents special benefits to develop client experience further, likewise presents a few issues that have happened definitively in the information procurement layer and how they are shipped [14]. This puts client and association information in danger because, as per reports and exploration on IoT security, there are points of reference in which security levels have been compromised and data has been influenced [15]. These insufficiencies need to do straightforwardly with the incredible second that passes the gadgets that fuse this innovation. This has driven makers to zero in on the offer of an item and not on the nature of the gadget that should fulfill guidelines that permit it to give significant degrees of wellbeing. Large numbers of the security imperfections found in the IoT incorporate makers that leave indirect accesses open to get data about the utilization of their items and, in this manner, work on their business [16]. Data the board is presently an issue that ought to be viewed as a need in sending any engineering that incorporates the utilization of arising advancements. Some arrangements and guidelines are being chipped away at to save information security, and yet, search for models that guarantee the appropriate utilization of data. The blend of innovations has frequently permitted us to work on a framework, design, or show and fortify its utilization in the public eye [17]. This reference is material in the utilization of arising advances, and there are explicit cases that indicate the strength of this mix [2]. This is the situation of college grounds that take on the Web of Things, Enormous Information, and distributed computing to become brilliant grounds [18]. The need is the personalization of the administrations committed to all individuals and producing economic environments [19]. Inside this equivalent climate, the reconciliation of an innovation that, like IoT, has outstanding development, the blockchain, has been thought of. The blockchain is a novel library concurred and appropriated in a few hubs of an organization. This innovation permits the exchange of computerized information with extremely modern coding and protected manner [20]. Each square has a particular and steadfast spot in the chain since each square contains data from the past block's hash. The whole chain is put away in every hub of the organization that makes up the blockchain, so a precise chain is put away for all organization members [21].

Blockchain innovation turns into the ideal accomplice for IoT because it ensures that data treatment and instability in the client vanishes by not uncovering data from the source and the beneficiary. This is ideal for IoT sending, tackling the issues it presents until now [22]. In any case, how a blockchain engineering can be incorporated into the IoT design without influencing the accessible administrations of a climate? The appropriate response is to make test seats, where these advances are executed in a controlled climate, and the information isn't compromised. To do this, the work proposes to make a design that incorporates both the IoT and the blockchain inside college grounds [23]. College fields can be pretty much as extensive as little urban areas and handle comparable administration models. The benefit of making this engineering on college grounds is to give the ideal controlled climate and characterize every one of the genuine factors of a scaled climate.

2. Review of Related works

In [24], the authors stated that to further develop learning, there is a tool stash that can become rehearsed for understanding the hypothesis of IoT and blockchain innovation. The instruments comprise of three sections, the "mind," "muscle," and "cloud." Raspberry Pi is utilized as a "cerebrum" of activity that speaks with the Kaiser Cloud stage. The outcomes demonstrate that the learning instruments effectively associate with the Kaiser Cloud stage and can be applied as a training gadget for learning destinations [24]. This exploration expects to present a solid framework dependent on blockchain to determine the trouble of point synchronization in IoT. This review has some advantage of level inactivity through utilizing an assent gadget dependent on evidence of dependability. Squeezing assessments and reproductions demonstrate that this plan vessel effectively opposes not well-arranged uses and diminishes inactivity from the agreement strategy by coordinating with it with old strategies [25]. A concentrate by [26] on the different strides of key casing extraction from a video upheld by secure and progressed transportation to clients offered a technique that permits clients to audit a video dependent on people and items as boundaries. Cryptographic hashes are utilized, including blockchain, and hashes are made from diminished video hinders, communicated, and moved through blockchain by [26]. Additionally, [27] presented a Blockchain-based calculated construction that gives an explanation for primary squares of the actual Web concerning the

substitution of significant worth and genuine resources in strategic frameworks and decentralized control structures [27].

3. Methodology

3.1 Analysis of the existing system

From the information gathered plainly, most Nigerian colleges experience the ill effects of digital assaults, for example, unapproved admittance to understudies or staff records, control of results without regard for anyone else, and in some cases for monetary profits and other noxious inclinations by the programmer. Subsequently, there is a requirement for pressing upgrades which can be accomplished through the mix of a portion of the components of blockchain Innovation like decentralization, SHA-256 encryption, approval, and straightforwardness to work on the security of the framework.

The current framework utilizes a customer server design. A customer/server engineering is a unified design where every one of the records of information dwells on a solitary P.C., and it is powerless against assaults, and in case of an assault, everything information may be lost in light of the fact that there is no reinforcement server.

- Explicit issues related to the current framework include:
- Overloaded servers when there are incessant concurrent customer demands, server seriously get over-burden, framing gridlock.
- vulnerability to assaults because of complete centralization of framework.
- lack of approval of records.
- All clients are absolutely subject to the maker's information base because of centralization.
- Impact of a brought together organization: since it is concentrated, if a basic server fizzles, customer demands are not refined. Accordingly, customer server comes up short on the vigor of a decent organization.
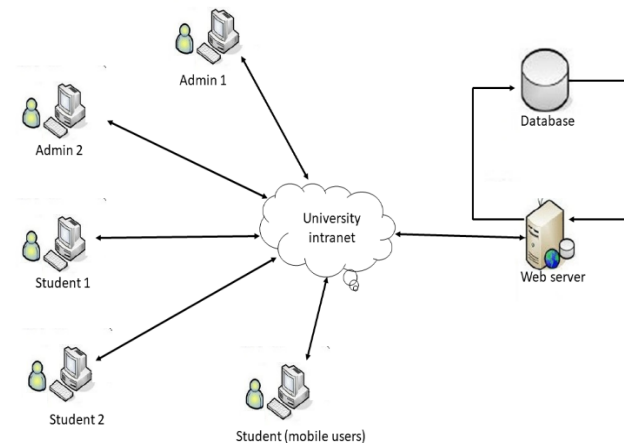

Figure 1: Existing architecture of a university portal

3.2. Proposed system

For the framework to fulfill the necessary guideline of in-weakness to assaults, there is a need to further develop the framework security, effectively open and be blockchain-like framework; it should be a blended organization design. Blockchain is totally decentralized with a "distributed" network design (that is, there is no focal control, and the hubs have equivalent control on the organization) dissimilar to the current college framework which is generally a concentrated framework running on a customer server network engineering.

Thusly, to accomplish the coordination of the blockchain into the college framework, there should be a think twice about both sides. The proposed framework in this exploration is a blended engineering, the mix of shared design and customer server engineering.

Un-like a blockchain framework there is a type of focal control by the chairman, the director can transfer records and make changes in the framework, the head can see the records and subtleties of the understudies and can likewise alter the understudies profile, yet a non-manager can't adjust, change or controls any record or data on the framework. The understudies can only login to the framework to see their outcome and alter their profile which is approved by the blockchain. Complete components of blockchain innovation can't be incorporated into the college framework, just about scarcely any significant provisions like SHA-256 encryption, record approval, straightforwardness and decentralization.

Because of Blockchain permanence, decentralization, circulation and agreement, blockchain gives upgraded security [28]. They see that chronicled information change will be unrealistic, and constant new information will be difficult to control as it is divided between all blockchain hubs with modification effortlessly recognized, followed and observed, forestalling misrepresentation and abuse. Blockchain can give both security and protection. However, with every one of the current dangers, blockchain is safer than the current concentrated framework

The availability of various elements in the proposed design in figure 3.1 depends on blockchain which gives efficient records stockpiling, sharing, and certification, which is made out of training foundations, consortium blockchain, capacity server, and the system administration.

3.2.1. Education Institutions

As the elements compared to the blockchain part hubs, instructive organizations are the fundamental assortment of information stockpiling and sharing on the blockchain.
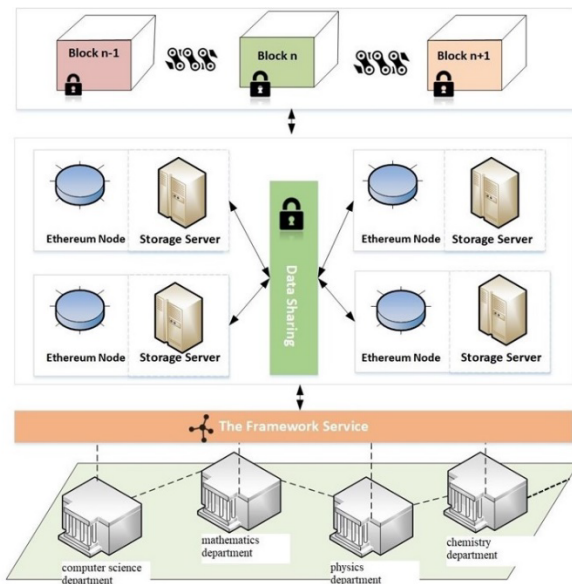
Figure 2: Proposed system architecture

3.3.2. Consortium Blockchain

The pre-owned consortium blockchain is liable for putting away the synopsis data of scrambled records. Once more, Ethereum is taken on to carry out the consortium blockchain. The savvy contracts are sent to control the capacity and recovery of information.

3.2.3. Storage Server

The first records and files are encoded and put away in the capacity server. Xampp information base is picked as the capacity server to efficiently store and recover information and backing scrambled capacity of files.

3.2.4. The Framework Service

Contains a bunch of accessible utilitarian modules, doesn't store any recorded data and offers types of assistance to hubs or clients as Relaxing APIs.

The proposed plan can be profoundly coordinated into the capacity, sharing and the board of instructive records.

3.3. System design

The framework configuration characterizes the plan, parts modules, interface and information for a framework that fulfill explicit necessities. The framework configuration permits the client gain an itemized comprehension of how the framework capacities.

It fosters the structural detail needed to assemble a framework or item.

The between network of various modules in the proposed framework engineering present in figure 3.1 gives a complete comprehension of the proposed framework design.

The Executives in various divisions transfer understudy records and information into the structure administration (the college framework) as data sources,

this present data's transition to the capacity servers where they are being put away. The associations between the capacity server and blockchain is a vehicle for sending data to the blockchain for approval and SHA-256 encryption. After the data is being gotten to by the blockchain, it goes through approval and SHA-256 encryption, and afterward, the information is being send back to the server or information sharing Center module. The information-sharing Center module then, at that point, gives the college structure administration admittance to this data however in SHA-256 scrambled structure. The scrambled data must be seen by various division yet can't be changed on the grounds that the SHA-256 calculation produce irreversible and one-of-a-kind hashes. The bigger the quantity of potential hashes, the more modest the possibility that two qualities will make a similar hash.

From the college structure administration (framework), various divisions in the instructive foundation then, at that point, gains admittance to see this data.

The managers in various divisions are the just once fit for survey and making changes to this current data's. The understudies can just view their outcomes and deal with their profile.

3.4. Data Security Framework using modified RSA Encryption Technique

Given that 'n' = modulus, $\Phi(n)$ = totient function, 'e' = Public Key functionality, 'f' = transform the value of Public Key functionality, 'd' = Private Key which is also the function the modulus, 'C' = encrypted using the transformed 'f' value, N = message in plain form, D = decrypted message.

A.   Key Generation
1.   Select large prime numbers p and q.
2.   Obtain product n =p*q.
3.   Obtain product n = (p-1) *(q-1)
4.   Get e based on the following conditions {p> e >$\Phi(n)$, coprime $\Phi(n)$ and n} and gcd [e, $\Phi(n)$] = 1
5.   Make a random selection 'e' from the list
6.   Compute f= ((e*2) +1).
7.   Make a selection of d from the relation {de mod$\Phi(n)$ = 1}
8.   Send the Public Key (f, n)
      Where 'f' is serving as the new Public Key which is expected to hide the original 'e' value.
9.   Send Private key (d, n)
B.   For data encryption
      $C = M^{((e*2)+1)} \bmod(n)$
C.   For decryption at the receiving end

$$D = C^d \bmod(n)$$

Explanation: Observed that in the modified RSA Algorithm above, the public key functionality 'e' has been tempered with by transforming 'e' to 'f'= (e*2)+1 in bid to obtained an improved security for the student data that is less vulnerable to factorization attack for malicious purpose (due to the additional second security layer obtained by transforming 'e' to 'f'). In this case, students data on transit in a network or at rest in the server is protected to a reasonable extent. Essentially, the existing RSA encryption that depends on public key 'e' has been strengthened in the modified RSA by transforming the 'e' to an additional security layer 'f' to effectively protect student data in terms of confidentiality, integrity and authenticity.
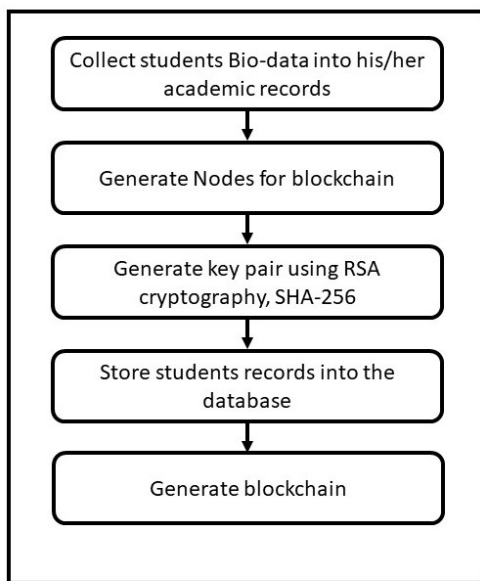
Figure 3: Process model for the system.

3.5. System Design Specification

Framework particular can be formal or casual. The proper framework determination is utilized in this exploration. It is a sort of framework particular that depicts different part of the framework verbally. Casual framework particular can likewise utilize graph to complete framework plan detail. In any case, it isn't obligatory to utilize graphs given the portrayal of the framework is all around given.

In the mean time, the major useful necessity detail of the framework is that it should confirm exchange and encode records utilizing SHA-256 encryption highlights. The framework configuration comprises of interface determination, program detail and information base particular.

3.5.1. Interface specification

The interface is where data can be supplied to and out the system. It has an interface for both the admin and customer login, the interface to create and an account for the customer and for the customer to transfer funds to the other customer, interface for viewing profiles and transactions.

3.5.2. Database specifications

The database consists of admin/bankers, customers and transactions(passbook). The field of the database should be according to the information needed in the fields. The names and other personal details should have variable characters.

3.5.3. Input Interface Design

The input interfaces of the application consist of some input functionality which can be accessible by the administrator and user or customer of the system.

The input design of this software includes
a. **Administrator login:** is the first page the admin sees (index page) when the web system loads on the browser, admin have to input his or her username and password and when he clicks the login button, the input is authenticated before the admin can access the admin home page, where the admin can manage students record. The admin manages the students record by performing actions such as: creating, updating, and editing students record.
b. **Customer login:** is the first page the student sees (index page) when the web system loads on the browser, student have to input his or her username and password. The input is authenticated before the student can access the customer home page, where the student can edit his or her profile.
c. **Create students (this functionality is only available to the administrator)**: in this input page, the admin creates a profile for the students, the genesis block is also created when the student profile is created. The admin inputs the student details (such as name, age, date of birth and other relevant information) in this page and submit in other to create the student profile.

3.6. Database Design

Files held in this project are made up of different data types. These types are integer, Character, Double, Date, etc. Some of the files used are designed and linked with database. Also, in the project design, MySQL database was used. Below is the database specification for the files used.

Table 1: Students; The design layout of database tables for students

File Name	Data Type	Size	Relation
Student_id	Int	11	**Primary key**
First_name	Varchar	30	Not null

Last_name	Varchar	30	Not null
Gender	Varchar	10	Not null
Faculty	Int	20	Not null
Department	Int	20	Not null
DOB	Varchar	30	Not null
Nationality	Varchar	20	Not null
State	Varchar	30	Not null
LGA	Varchar	30	Not null
Marital status	Int	15	Not null
Email id	Int	20	Not null
Phone no.	Int	11	Not null
Uname	Varchar	30	Foreign key
Pwd	Varchar	30	Foreign key

Table 2: ADMINISTRATOR: the design layout of database tables for admin

Filename	Data Type	Size	Relation
Id	Int	11	**Primary key**
Username	Char	25	Foreign key
Pwd	Char	25	Foreign key

Table 3: RECORD/RESULTS: the design layout of database tables for students'records

File name	Data type	Size	Relation
Record_id	Int	11	**Primary key**
Course code	Int	10	Not null
Course title	Int	50	Not null
Credit unit	Int	11	Not null
GP	Int	11	Not null
CGPA	Int	11	Not null
Remark	Int	11	Not null
PRecordhash	Varchar	64	Not null
Recordhash	Varchar	64	Not null

3.7. System Development

Flow chart shows how the data flow in the system. Flow chart depicting the operation of this system and show the operational flows.
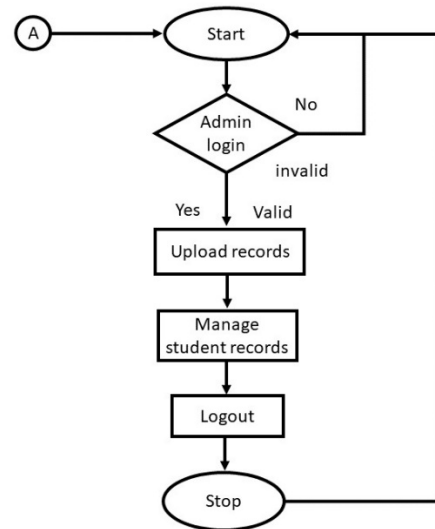
Figure 4: Flowchart for admin officer gaining access to the system
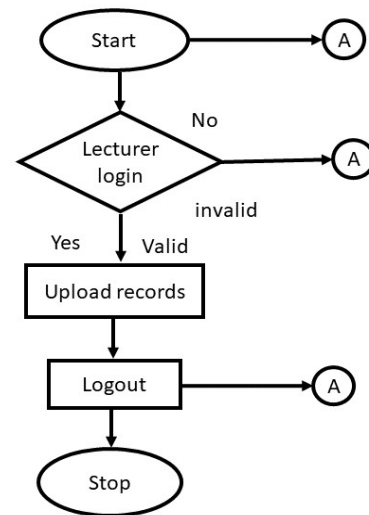
Figure 5: Flowchart for lecturer gaining access to the system
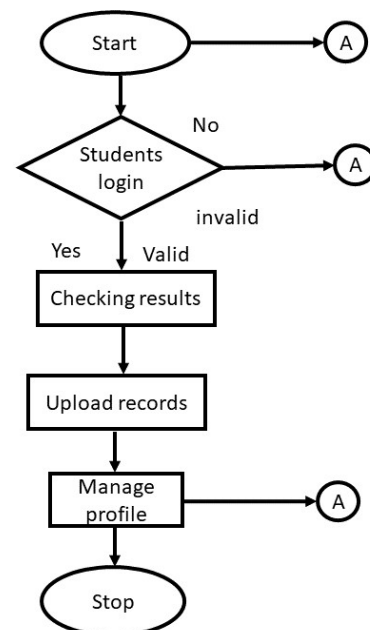
Figure 6: Flowchart for student gaining access to the system

3.8. System Modeling Using Unified Modeling Language (UML).

Various UML tools were used to model the basic functionalities required for a working system. More specifically for this research, UML use cases, class diagram, sequence diagram and entity-relation (E-R) diagram were used to model the system.

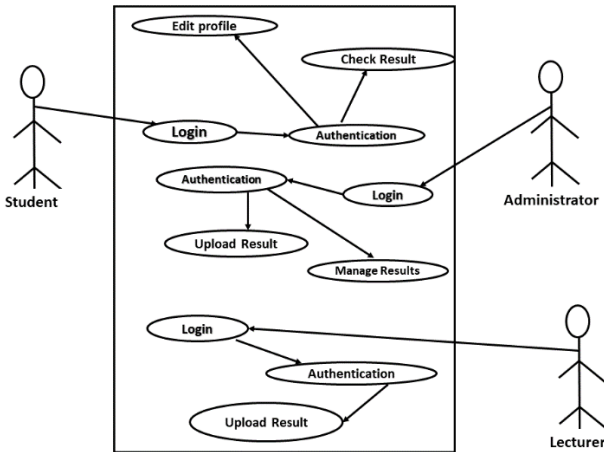The UML use case for the system is as shown in Figure 7 below:

Figure 7: Use case for the system
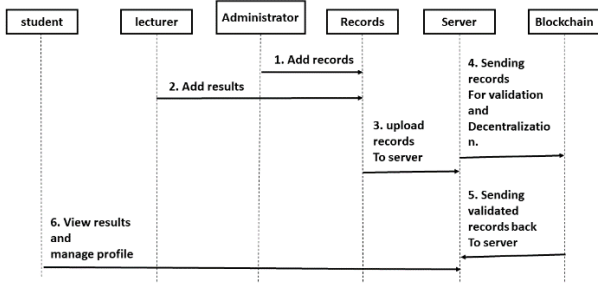
The sequence diagram used to model basic artifacts of the system is given in Figure 8 below:

Figure 8: Sequence diagram for the system

The class diagram that was used in modelling different object classes for the system is as displayed in Figure 9 below:
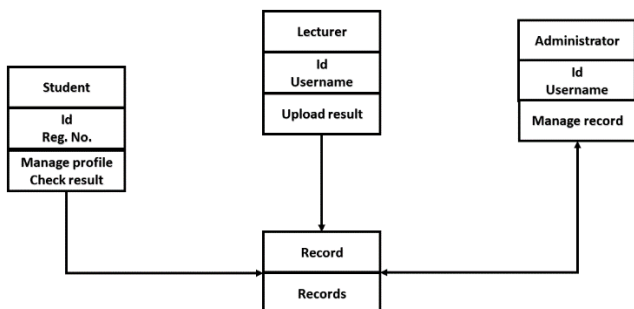
Figure 9: Class diagram of the system

The entity relationship model for the proposed system is described in Figure 10 below:
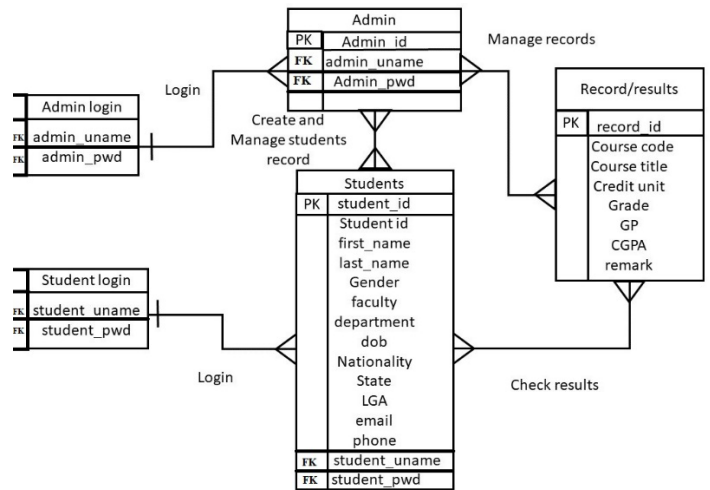
Figure 10: Entity Relation (E-R) Diagram for the system

3.9. Technologies for Development

The developmental tools and technologies used include PHP/MySQL for the database because PHP/MySQL is more secure and robust than another database, it is easy to implement, and I have more experience with PHP/MySQL.

The system was developed using vs code. HTML, JAVASCRIPT and CSS are used to develop the User Interface. PHP (Hypertext Preprocessor) was the programming language used. PHP is a scripting language designed specifically for use on the Web.

4. Implementation

4.1. System Implementation

In this section the modelled system was implemented where details were presented for a smooth execution.

The integration of blockchain-enabled framework for securing students record was a success. After several tests run on the developed system. The administrator can create the genesis block by adding students' function, which is creating students' profile.

The institution framework is secure and reliable since students' records are encrypted, and the system is decentralized. Therefore, the aim of the study is achieved.
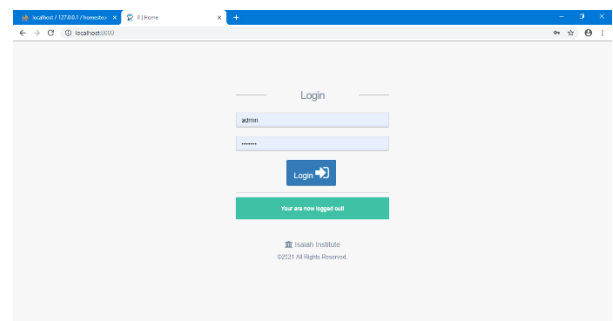

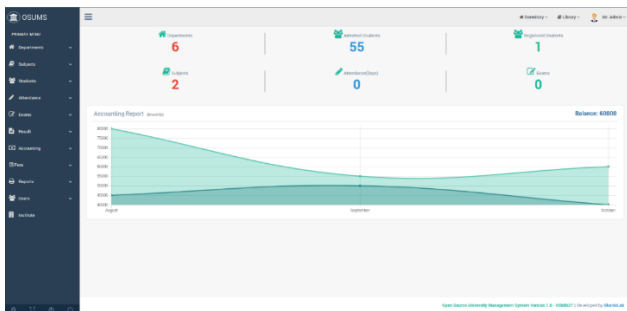Figure 11: login page for administrator where the administrator gets access into the system

Figure 12: The interface above is the Administrators dashboard that is displayed as a result of the Administrator successfully gaining access or logging into the system through the administrator's login page.
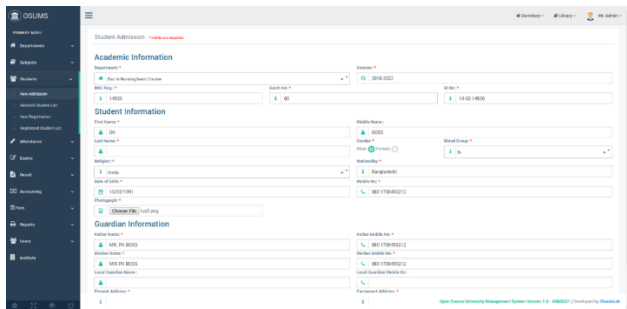

Figure 13: The create students profile interface is used by the administrator to input students Information/Bio-data into the system
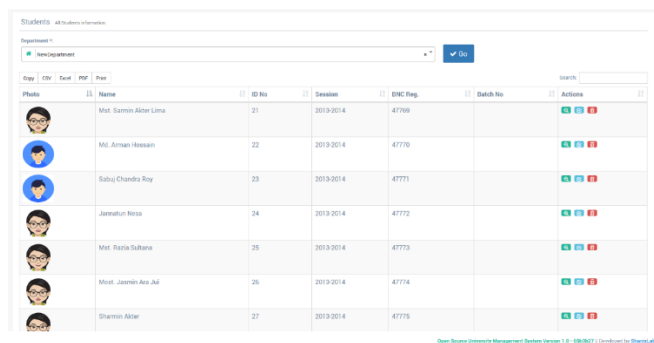

Figure 14: The different department interface is an interface that shows students of different department which can be viewed by administrators of different departments
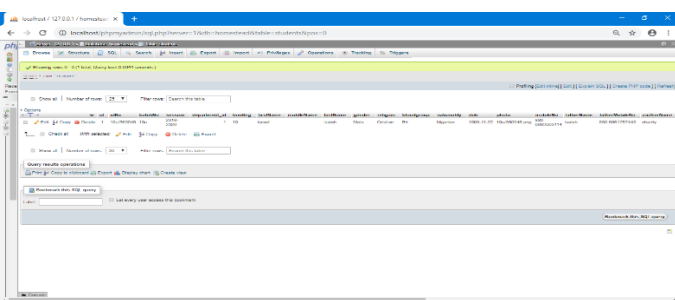

Figure 15: Student profile File System Implementation: this is the student table as it appears in the database
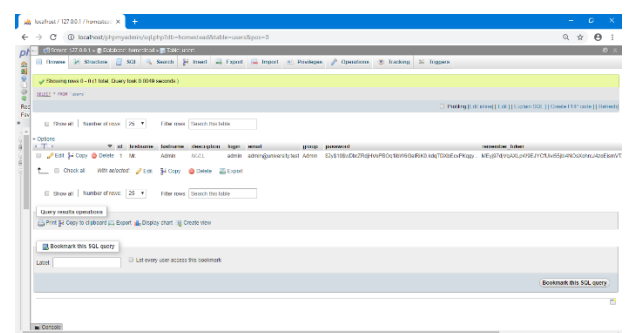

Figure 16: Administrator File System Implementation: this is the administrators table as it appears in the database, the administrators table is not in any way linked to the students table.
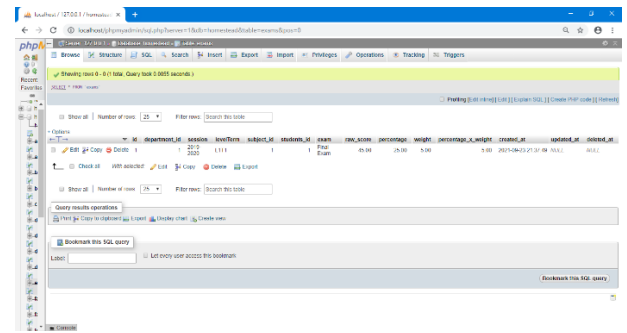

Figure 17: Records (Exams) File System Implementation: this is the student records as it appears in the database.

4.2. Discussion

Blockchain has been accepted in many areas such as banking system, health care, supply chain production and academic institutions.

The results of this research is the blockchain-enabled system for data storage, security and sharing of students record in educational institution. The results obtained shows that the use of blockchain technology will be a great improvement in universities.

From the literature review and related works of this research, the implementation of blockchain in educational institution to improve learning always end up with issues like difficulty of point of synchronization in IoT, inability to tamper with any record after being stored in the system. But mostly, blockchain has not been widely used in academic institutions or universities especially in most African countries like Nigeria.

This research focuses on the practical aspect of blockchain implementation, from the results of the implementation this research eliminates the need of any difficulty of point of synchronization and tamper of proof. The implemented system provides a sharing medium whereby an administrator from the computer science department is capable of viewing students record from mathematics or physics but cannot tamper or change those records.

The system provides an efficient and maximum data storage, sharing of students records, and security using RSA and SHA-256 encryption to record and encrypt students records.

Figure 11 shows vividly the administrators login page, where the administrators can get access into the university management system to store and manage students records and perform other activities. This is the administrators gateway into the system.

Figure 12 the home page or dashboard of the system. After the administrator has get accessed to the system, this landing page is the administrator's dashboard. Here the administrator can see the number of departments, number of students, number of subjects, number of

registered courses and other related activities that are contained or stored on the system.

Figure 13 is an interface which shows a clear view of the create student profile for new students by the administrator. This is where the administrator creates profiles for new students. The administrator is the only person capable of creating a students' profile in the university system. The fields consist of different data to be obtained from the newly admitted student.

Figure 14 shows different students of different department, this interface allows administrators of different department to view the records and details of student from different departments. Here, an administrator from the computer science department can view the results of a student from mathematics and physics or any other department as administrators from other departments too can view the records of students from another department.

Figure 15 shows the student profile File System Implementation; the interface is the student table as it appears in the database. Figure 16 shows the Administrator File System Implementation; the interface is the administrators table as it appears in the database. Figure 17 shows the Records (Exams) File System Implementation the interface is the student records as it appears in the database where the students' data is being stored.

5. Conclusion

The point of this review is to survey IoT and perceive how it tends to be Coordinated with Square chain innovation in college gateways in a bid to further develop security of information store, plan, study and carry out a framework that will upgrade blockchain-empowered. A tremendous degree of potentials for the institutions to take advantage of the implementation of a blockchain-empowered framework. This innovation can specify college development and advancement. The degree of innovation in terms of progress which is taking in division in present day have been both strengthen and overwhelming, in terms of examining how many surges of mechanical developments with immerse specialization and organized information world.

The review explored the utilization of blockchain-empowered instructive organization and the investigation discovers that clients of the proposed gateway system expect; admittance to grades, profile and different provisions on the entry; clients need more admittance to the entryway. In this way, it is reasoned that reviews on client necessities of blockchain-empowered training foundation ought to be consistent and at stretches to get criticism from clients by administrators of the gateway with the end goal of

meeting client prerequisites for better usability. It may be insufficient to start the framework and similar services without taking into consideration, client discernments, prerequisites, needs and difficulties, all of which have the tendency to add to the overall expectations of the organization being upgraded in terms of learning and to simultaneously understudy suitably of students' records. It is much more efficient to lead comparable analysis in this 21st century electronic driven climate.

Blockchain has shown its capacity for modifying conventional organization with its key attributes: decentralization, persistency, secrecy and auditability. In this research, a far-reaching outline on blockchain is presented. An outline of blockchain advances including blockchain engineering and key qualities of blockchain is presented. Also, an examination of the regular agreement algorithms is utilized in blockchain.

References

[1] W. Villegas-Ch; M. Roman-Carnizares; X. Palacious-Pacheco. Improvement of an online educational model with the integration of machine learning and data analysis in LMS. *Appl. Sci. 2020*, vol. 10, pp. 5371, doi: 10.3390/app10155371.

[2] P. Butz; B. Tauscher. Emerging technologies: *Chemical aspects. Food Res. Int. 2002*, vol. 35, pp. 279–284, doi: 10.1016/S0963-9969(01)00197-1.

[3] F. Montori; L. Bedogni; L. Bononi. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. *IEEE Internet Thing J. 2018*, vol. 5, pp. 592–605, doi: 10.1109/JIOT-IEEE.2017.2720855.

[4] M. Alvarez-Campana; G. López; E. Vázquez; V. A. Villagrá; J. Berrocal; Smart CEI moncloa: An iot-based platform for people flow and environmental monitoring on a Smart University Campus. *Sensors 2017*, 17, 2856, http://doi.org/10.3390/j.s2017.171222856.

[5] L. Catarinucci; D. De Donno; L. Mainetti; L. Palano; L. Patrono; M. L. Stefanizzi; L. Tarricone. An IoT-Aware Architecture for Smart Healthcare Systems. *IEEE Internet Things J. 2015*, 2, 515–526, doi: 10.1109/JIOT.2015.2417684.

[6] H. Y. Shwe T. K. Jet; P. H. J. Chong. An IoT-oriented data storage framework in smart city applications. *In Proceedings of the International Conference on Information and Communication Technology Convergence, ICTC 2016*, Jeju, Korea, 19–21 October 2016; pp. 106–108, doi: 10.1109/ICTC.2016.7763446.

[7] S. Madakam; R. Ramaswamy; S. Tripathi. Internet of Things (IoT): *A Literature Review. J.Comput. Commun. 2015*, vol. 3, pp. 164–173, doi: 10.4236/jcc.2015.35021.

[8] P. Bellagente; P. Ferrari; A. Flammini; S. Rinaldi. Adopting IoT framework for Energy Management of Smart Building: A real test-case. *In Proceedings of the IEEE 1st International Forum on Research and Technologies for Society and Industry, Turin, Italy,* 16–18 September 2015; pp. 138–143, dol: 10.1109/RTSI.IEEE.2015.7325084. .

[9] M. Batty; K. W. Axhausen; F. Giannotti; A. Pozdnoukhov; A. Bazzani; N. Wachowicz; G. Ouzounis; Y. Portugali. Smart cities of the future. *Eur. Phys. J.Spec.* Top.2012,214 pp. ,481–518, doi: 10.1140/epjst/e2012-01703-3.

[10] J. Zhou; Y. Zhou; B. Wang; J. Zang. Human–Cyber-Physical Systems (HCPSs) in the Context of New-Generation Intelligent Manufacturing. *Engineering 2019*, vol. 5, pp. 624–636, doi: 10.1016/j.eng.2019.07.015.

[11] J. Zhou; Y. Zhou; B. Wang; J. Zang; L. Meng. Toward New-Generation Intelligent Manufacturing. *Engineering 2018*, 4, 11–20, doi: 10.1016/j.eng.2018.01.002.

[12] W. Baicun; Z. Jiyuan; Q. Xianming; D. Jingchen; Z. Yanhong. Research on New-Generation Intelligent Manufacturing based on Human-Cyber-Physical Systems. *Chin. J. Eng. Sci. 2018*, 20, 29, 10.15302/J-SSCAE-2018.04.006.

[13] M. Aazam; I. Khan; A. A. Alsaffar; E. N. Huh. Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved. *In Proceedings of the 11th International Bhurban Conference on Applied Sciences and Technology, IBCAST, Islamabad, Pakistan*, 14–18 January 2014; pp. 414–419, doi: 10.1109/IBCAST.2014.6778179.

[14] Y. Mahmoodi; S. Reiter; A. Viehl; O. Bringmann; W. Rosenstiel. Attack surface modeling and assessment for penetration testing of IoT system designs. *In Proceedings of the 21st Euromicro Conference on Digital System Design (DSD), Prague, Czech Republic,* 29–31 August 2018; pp. 177–181, doi: 10.1109/DSD.2018.00043.

[15] H. Suo; J. Wan; C. Zou; J. Liu. Security in the internet of things: A review. *In Proceedings of the International Conference on Computer Science and Electronics Engineering, ICCSEE, Hangzhou, China*, 23–25 March 2012; Volume 3, pp. 648–651, http://doi.org/10.1109.ICCSEE.2012.373.

[16] P. P. Ray. A survey of IoT cloud platforms. *Future Comput. Inf. J. 2016*, vol. 1, pp. 35–46, doi: 10.1013/j.fcij.2017.02.001.

[17] R. Khan; S. U. Khan; R. Zaheer; S. Khan. Future internet: The internet of things architecture, possible applications and key challenges. *In Proceedings of the 10th International Conference on Frontiers of Information Technology, FIT 2012, Islamabad, India*, 17–19 December 2012; pp. 257–260, doi**:** 10.1109/FIT.2012.53.

[18] M. Liu; L. Li. The construction of smart campus in universities and the practical innovation of student work. *In Proceedings of the International Conference on Information Management & Management Science, Chengdu, China*, 24–26 August 2018; pp. 154–157, doi: 10.1145/IMMS.3277139.3278307.

[19] W. Villegas-Ch; X. Palacios-Pacheco; S. Luján-Mora. Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: *A Sustainable Smart Campus. Sustainability 2019*, vol. 11, pp. 2857, doi: 10.3390/su11102857.

[20] H. Wang; Z. Zheng; S. Xie; H. N. Dai; X. Chen. Blockchain challenges and opportunities: *A survey. Int. J. Web Grid Serv. 2018*, 14, 352, doi: 10.1504/ijwgs.2018.095647.

[21] D. Minoli; B. Occhiogrosso. Blockchain mechanisms for IoT security. *Internet Things 2018*, 1–2, pp. 1–13, doi: 10.1016/j.iot.2018.05.002.

[22] A. D. Dwivedi; G. Srivastava; S. Dhar; R. Singh. A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors 2019*, 19, 326, doi: 10.3390/s19020326.

[23] V. L. Uskov; J. P. Bakken; A. Pandey. Smart University Taxonomy: Features, Components, Systems. In Smart Education and e-Learning; Springer: *Cham, Switzerland, 2016*; Volume 59, pp. 3–14, doi: 10.1007/978-3-319-39690-3_1.

[24] N. Kamal, M.H.M. Saad, C.S. Kok, and A. Hussain, "Towards Revolutionizing STEM Education via IoT and Blockchain Technology," *International Journal of Engineering and Technology (UAE),* Vol. 7, No. 4, 2018, pp. 189-192, doi: 10.14419/ijet.v7i4.11.20800.

[25] K. Fan, S. Sun, Z. Yan, Q. Pan, H. Li, and Y. Yang, "A Blockchain-Based Clock Synchronization Scheme in IoT," *Future Generation Computer Systems*, Vol. 101, December 2019, pp. 524-533, doi: 10.1016/j.future.2019.06.007.

[26] G. Khan, S. Jabeen, M.Z. Khan, M.U.G. Khan, and R. Iqbal, "Blockchain-Enabled Deep Semantic Video-to-Video Summarization for IoT Devices," *Computers & Electrical Engineering*, Vol. 81, 2020, art. 106524, doi: 10.1016/j.compeleceng.2019.106524.

[27] T. Meyer, M. Kuhn, and E. Hartmann, "Blockchain Technology Enabling the Physical Internet: *A Synergetic Application Framework,"* Computers & Industrial Engineering, Vol. 136, October 2019, pp. 517, doi: 10.106/j.cie.2019.07.006

[28] H. Hassani, and E. S. Silva (2018). Forecasting U.K. Consumer Price Ination Using Ination Forecasts. *Records in Economics, in Press.* doi: 10.1016/j.rie.2018.07.001.

SARJIYUS OMEGA, the corresponding author obtained his bachelor's degree in computer science from Adamawa state university Mubi, Nigeria in 2009. He also obtained his master's degree from the same institution in 2014 and presently on the advanced state of his PhD in Modibbo Adama University of Technology, MAUTECH Yola, Nigeria. He has a flair for research areas such as Cyber security, Information security, Artificial Intelligence with many publications in those fields.
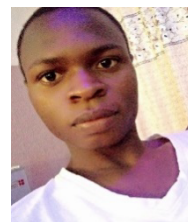
ISRAEL ISAIAH has done his bachelor's degree in computer science from Adamawa state university Mubi, Nigeria in 2021. He is passionate and enthusiastic about research and he has a flair for research areas such as Cyber security, Information security, Artificial Intelligence with many publications in those fields.