ISSN No:-2456-2165

Machine Learning Survival Analysis Model for Diabetes Mellitus

Maureen I. Akazue[1], Geofrey A. Nwokolo[1], Clement O. Ogeh[1] Emmanuel Ufiofio[1]
[1]Department of Computer Science,
Delta State University, Abraka, Nigeria

Okpako A. Ejaita[2]
[2]Department of Cybersecurity,
University of Delta, Agbor, Delta State, Nigeria

Abstract:- Developing effective survival analysis models would help guide the decision-making in managing major health challenges. Model development can be achieved through various approaches. Diabetes is a health challenge in Nigeria that has attracted the interest of researchers thus much research has been carried out as regards its management necessitating the development of models. This study carried out a machine learning analysis on diabetes data collected from Central Hospital, Warri, Delta State implementing Cox-PH Model due to the role both play in survival analysis. A dataset of 100 diabetic patients' records was collected. The dataset was used for training multiple machine learning algorithms, namely, SupportVector (SVM), K-nearestneighbors (KNN) classifier, etc., and the proposed model (Cox-PH Hybrid or CPH-SML). The performance evaluation of the machine learning algorithms and the proposed model gave accuracy levels as follows: KNN-47%, SVM; 74%, and Cox-PH Hybrid-96%. The concordance index was used to evaluate the proposed model and it had an index of 0.7204, on several covariates such as Age, Gender, Education, Marital Status, history of smoking, SBP, DBP, etc. From this study's analysis of the diabetic data, it was able to conclude that the variables associated with diabetes mortality are; the age of the patient and diabetes types. The patients' hazard ratio reduces when they are young compared to when they are old. The patient's hazard ratio is also dependent on the diabetes type. Thus, early diagnosis and proper health management of diabetics can prolong the age of diabetic patients.

Keywords:- *Survival Model, Machine Learning, Cox Proportional Hazard, Diabetes.*

I. INTRODUCTION

Diabetes patients are increasing at a rapid rate, and it is estimated that more than 90-95 percent of people globally have Type 2 diabetes, which is one of the leading causes of death and contributes to a large number of deaths each year in an unnoticed manner [1]. Diabetes Mellitus (DM) has been defined as a condition that is induced by unregulated diabetes that may lead to multi-organ failure in patients [2]. Diabetes has become one of the biggest health challenges in the world hence the need to control and manage it.

Related works in the healthcare information systems show that the increasing number of healthcare data requires the need for effective means of extracting information to aid the delivery of healthcare services to patients [3,4,5,6& 7]. The development of a prediction model to guide clinical decisions about whether to continue therapy or take an alternate action that affects the life of patients receiving treatment is required[8, 9 & 10].However, to respond to this need, many studies have concentrated on prediction models in traditional techniques, but computer scientists have focused on it using machine learning methodologies, to construct prediction models [11, 12].

Machine learning has played a great role in survival analysis helping out in clinical forecasting. Work done in this area includes "machine learning for survival analysis: a case study on recurrence of prostate cancer" [13], "ComoRbidity: an R package for the systematic analysis of disease comorbidities"[14], "Survival model for Diabetes Mellitus Patients' using support vector machine"[15], etc. This field has drawn much attention in the past and has become a dominant technology in the AI community [16].

Survival Analysis is one of the most popular methods of data mining that deals withthe estimation of the time to an event such as death, childbirth, radioactive decay, etc [3, 17, 18 & 19].

Due to the increasing rate of DM in the world, there arises the needfor more medical attention.One area which has responded to such need is the area of developing survival analysis models by researchers using machine learning algorithms. One algorithm in Machine learning is the Support Vector Machine algorithm used in a recent study for survival analysis on DM patients in Nigeria [20].This algorithm is said to be a fault in the area of handling large datasets for analysis. This, therefore calls for development of an enhanced model for diabetes survival analysis. This study developed a survival analysis model using a machine learning algorithm for diabetes mellitus patients while implementing the Cox Proportional Hazard model. Dataset for diabetes data was collected from Central Hospital in Delta State of Nigeria.

II. RELATED WORKS

Several models have been built to solve different human problems. These problems ranges from customers challenges at airport, performing online transactions, advisory systems, intrusion detection systems, and so onOkpeki et al.[21]; Okofu[22]; Okpeki & Omede[23]; Okofu et al. [24]; Efozia et al. [25]; Akazue[26]; Oijoe [27]; Ojugo&Otakore [28]; Ojugo& Yoro [29].Thus, it is good to provide survival models for health challenges. A survival model to predict the survival of pediatric Sickle Cell Disease (SCD) was developed using clinical variables by Idowu et al., [30]. The predictive model works with fuzzy logic.Three (3) clinical variables were used and the rules for the inference engine were elicited from an expert pediatrician [31]. The study report on the non-validation of the fuzzy logic-based model using live clinical datasets.Furthermore, relevant variables for SCD survival could have been easily identified using feature selection methods from a larger collection of variables monitored for pediatric SCD survival as observed, but the study found that the researchers did simulate nor validate the rule-base of the classification model for SCD survival [32].

Machine learning was applied to the prediction of the survival of pediatricHIV/AIDS patients [33]. The naïve Bayes' classifier was employed. The model was trained, evaluated, and gave a result that showed that the classifier was able to predict the survival of HIV/AIDS patients by the model's accuracy of 68%.

A pilot study on type 2 diabetes patients, showed the use of Deep learning to develop a novel adherence detection. This was based on simulated Continuous Glucose Monitoring (CGM) signals. CGM signals were simulated in a large and diverse amount for T2D patients with the aid of a T2D-adapted version of the Medtronic Virtual Patient (MVP) model for T1D [34]. Due to the signals, different classification algorithms were compared using a comprehensive grid search. The researchers for the study contrasted a standard logistic regression baseline to Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs). It was reported that the best classification performance having an average accuracy of 77:5% came as a result of employing CNN. This study, therefore, confirms the potential of Deep Learning as regards adherence detection systems for Type 2 disease patients [34].

Predicting factors for the survival of breast cancer patients using machine learning techniques was done in 2019. The study developed models for detecting and visualizing significant prognostic indicators of breast cancer survival rate [35]. The datasets were a hospital-based breast cancer dataset from the University of Malaya Medical Centre, Kuala in Malaysia with diagnostic information between 1993 and 2016. To determine the predicting factors, models were developed with a decision tree, random forest, neural networks, extreme boots, logistic regression, and support vector machine. The study affirms that all algorithms produced close outcomes, with the lowest obtained from the decision tree (accuracy = 79.8%) and the highest from the random forest (accuracy = 82.7%).

More work done on breast cancer survival includesKalafi et al., [36] using 4,902 patient records from the University of Malaya Medical Centre Breast Cancer Registry (UMMCBCR).The prediction modelswere designed and implemented by machine learning(SVM, RF, and DT) and deep learning MLP techniques.Findings show that the multilayer perceptron (MLP),random forest (RF) and decision tree (DT) classifierscould predict survivorship, respectively, with 88.2 %, 83.3 %, and 82.5 % accuracy in the tested samples. And Support vector machine (SVM) was recorded lower with 80.5 %.

Tachkov et al.,[37] conducted a study to evaluate the expected life expectancy in patients with diabetes and to compare it to the expected life expectancyof persons living without diabetes in the country.Analysis of confounders wasdone by age, sex, and type of diabetes. The causes of mortality in diabetic patients were analyzed by making use of Kaplan-Meier survival curves for each age cluster. Lon-rank analysis was not also left out. Findings in the study include; male diabetics patients showing slightly longer life expectancy than their counterparts in the non-diabetic population, by a marginal gain of 0.6 years for the entire observed period. Furthermore, life expectancy indiabetic women was said to increaseby 1.3 years, which was not observed in the non-diabetic population. Diabetes was said to occur more in women.

The Support Vector Machine was used by Bamidele et al., [38] to develop a survival model for diabetes mellitus. Identification of variables monitored during the management of diabetes mellitus patients was done.

Lee et al., [39] did work in Hong Kong with the aim of developing a predictive risk modelfor all-cause mortality in patients with diabetes. The study used a multi parametric approach with data from different domains. The association of risk variables and all-causemortality was assessed using Cox proportionalhazardsmodels. Machine learning approaches were used in the study to improve overall survival prediction and were evaluatedwith a fivefold cross-validation method. Age, male gender, baseline comorbidities, etc were significant predictors of all-cause mortality found through Multivariate Cox regression. The study affirms that a multi parametric model incorporating variables from different domains predicted all-cause mortality accurately in type 2 diabetes mellitus.

Survival analysis has also been done on coronavirus patients with the introduction of two models called Cox_COVID_19 and Deep_ _Cox_COVID_19. These models were developed to help hospitals select patients with better chances of survival and to predict the most important features affecting the rate of survival [39]. One of the survival models; Cox_COVID_19 is based on Cox regression while the second model; Deep_Cox_COVID_19 is a hybrid model, i.e a combination of autoencoder deep neural network and Cox regression. The study affirms that both systems (i.e the Cox_COVID_19 and Deep_

Cox_COVID_19) can predict the survival likelihoodand also present significant symptoms that differentiate severe cases and death cases.

Ojie, et al., [41] applied classification algorithm in their proposed hybrid model of Genetic Algorithm and Data value Metric (DVM) as an information theoretic metric for quantifying the quality and utility for feature selection. They proposed that this can be applied to traditional data.

III. METHODS AND MATERIALS

To carry out survival analysis and develop a predictive model on diabetes mellitus using a machine learning algorithm, the following steps were adopted (i) Data collection, (ii) Data preparation, (iii) Implementing the proposed model, and (iv)Evaluation.

A. Data Collection

After ethical approval, the researcher collected various variables from the records of 100 diabetes patients from the state central Hospital, Warri-Delta State following due procedures for collecting patient data in the health facility. The diabetes data were recorded using a spreadsheet with the assistance of the health workers in the unit. Most of the features relating to diabetes that needed to be collected for the study were outlined by medical personnel that the researcher contacted.

Features relating to the survival or mortality of diabetes mellitus patients were collected from the Hospital. Table 1 below is a description of the variables collected and used for the proposed system.

Table 1: Identified Variables for determining Diabetes Mellitus

S/N	Names of Variables	Labels
1	Gender	Male, Female
2	Present Age (in years)	Numeric
3	Highest Education	Primary, Secondary, Tertiary, Others, Nil
4	Occupation	Business, Civil Servant, Teacher, Electrician, Trader, Carpenter, Farmer, Cleaner, Nil
5	Marital Status	Single, Married, Widow, Widower, Divorced
6	Ethnicity	Urhuoba, Yoruba, Igbo, Hausa, Itsekiri, Ijaw
7	History of Smoking	Yes, No
8	Diabetes Type	Type 1, Type 2, Gestational
9	Age at Diagnosis (in years)	Numeric
10	Systolic Blood Pressure (SPB)	Numeric
11	Diastolic Blood Pressure (DPB)	Numeric
12	BMI Class	Underweight, Normal, Overweight, Obese
13	Haemoglobin A1c	Numeric
14	Treatment Plan	Insulin, others
15	History of Drug Resistance	Yes, No
16	Complication	Neuropathy, Nephropathy, Retinopathy, CVD, Stroke, Peripheral Artery Disease
17	Mortality	Numeric

B. Data Preparation

After the collection of data, the data was analyzed and cleaned. This means that data necessary for use in the various machine-learning algorithms were processed. The process involved the following:

- The ethnicity (Tribe) column was dropped.
- Yes / No type of columns were respectively converted and cleaned.
- Single observation on target variable value was found to be missing and hence dropped.
- Missing observations on a few categorical columns were detected and thus filled with their respective column mode value.
- All the categorical features were one-hot encoded.
- All the values in the dataset were scaled between 0 and 1 as a standardization technique.
- Dataset was split into training and testing sets ina ratio of 80:20.
- An extreme imbalance was detected on the target variablein the ratio of 85:14. It was handled by applyingSynthetic Minority Oversampling Technique (SMOTE).

The stepwise variable feature selection with iteration between the ‘forward and backward’ steps to obtain the best patient final Cox proportional hazard model was then implemented, tested, and evaluated for optimum performance.

IV. IMPLEMENTING THE PROPOSED MODEL

A. Data Training and Learning

The datasets are prepared for training and learning the features of the datasets using the various machine learning methods, including Random Forest, Gradient Boosting, Decision Tree, Support Vector, Multi-Layer Perceptron, and K-nearest Neighbors' classifier. During the training process, the model makes an effort to comprehend the properties and instance representation of a certain dataset that is used as input. The text input is converted by the feature extractor into a feature vector that categorizes its polarity. The programmer chose to divide the datasets into two halves for this investigation, with the ratio being 80% to 20% for training and learning, respectively. The training data is represented using the attributes after the instance and attributes have been chosen.

B. Data Testing and Classification

The output of the training datasets was compared with that of the testing datasets in order to check all conceivable combinations and evaluate how effectively a model will predict the intended or expected results. Were the expected result far different from the output result, the input was adjusted and the model was fine-tuned based on the results of the test data set. This was accomplished by comparing the attributes of the training and testing datasets, computing the probability for each hypothesis based on the attributes, and categorizing the attributes that were most similar to the outcome. Then Real-world datasets were then fed into the classifier for the categorization of tweets and spam emails while taking into account all of the classifying methods either as a single selection or bulk selection. First, the classifier is fed with testing datasets to check the correctness of the algorithm.

C. Building the Model

Multiple models were trained, namely, Random Forest, Gradient Boosting, DecisionTree, Support Vector, Multi-Layer Perception, and K-nearest neighbors' classifier. The Cox proportionality Hazards model was also implemented and evaluated on the dataset. This model was introduced by Cox and takes into account the effect of several variables at a time and the relationship of their survival distribution to these variables. The Random Forest algorithm yielded the best result with an accuracy of 82% against other algorithms. Then, a model with Random Forest Classifier and Cox Proportional Hazard algorithm was then developed with the title CPH-MLS Prediction model and deployed to the web app with a model performance accuracy of 96%.

From the fitted COX PH Model the variables that are associated with the mortality are the age of the patient, diabetes, and education level. The patients' hazard ratio increases by 1.95 when they have secondary education compared to when they have primary education. The patients' hazard ratio reduces by 0.96 when they are young compared to when they are old. The patients' hazard ratio reduces by 0.53 when they have diabetes compared to when they have no diabetes.

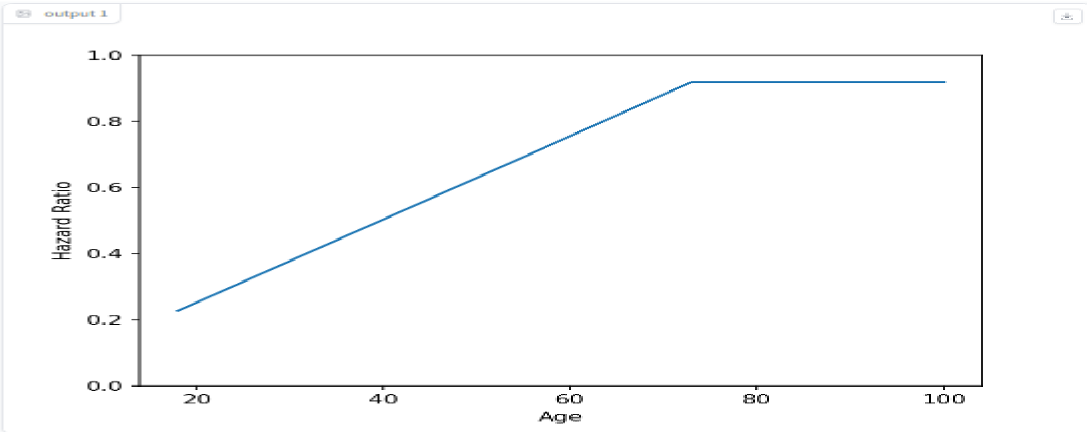

Fig. 1: Graph of Model prediction

The above chart shows the usability of the model in predicting the probability of having diabetes or not and the survival time of new patients. Thus the chart gives the survival time and probability. On the left is the prediction, 0, 1. When the chart reads 1, it means that it is predicting possible cases of diabetes mortality for the patient record analyzed and when it reads 0, it indicates no diabetes mortality prediction.

D. Evaluation

The performance metrics were used to evaluate the performance of the various algorithms in this study. This gives the overall performance summary of the learning algorithms on the datasets and weighted (average) performance rate which is shown below in table 2, fig 2, and fig 3.

Table 2: Summary of Model Performance

Model	Precision	Recall	F1-Score	Accuracy
KNN	0.45	0.56	0.5	47%
	0.5	0.4	0.44	
Logistic	0.62	0.56	0.59	63%
	0.64	0.7	0.67	
SVM	1	0.44	0.62	74%
	0.67	1	0.8	
Cox-PH Hybrid	0.93	1	0.96	96%
	0.96	0.96	0.96	

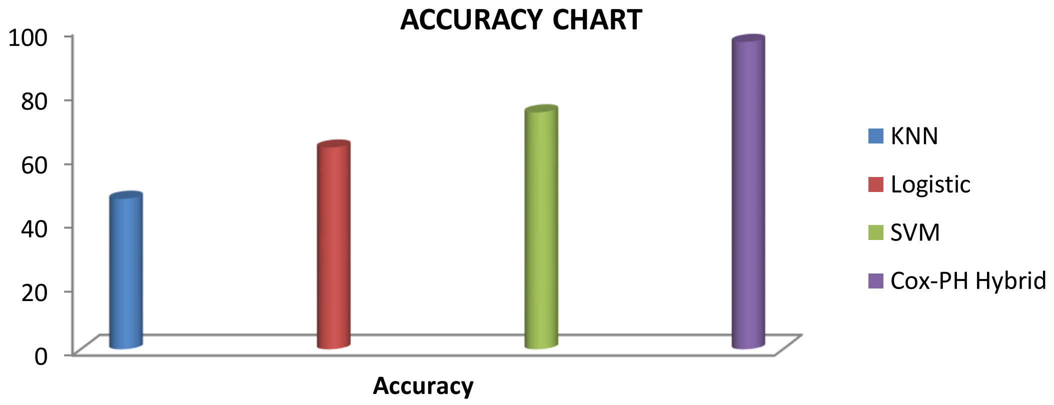

Fig. 2: Accuracy Performance Analysis on Diabetes Datasets

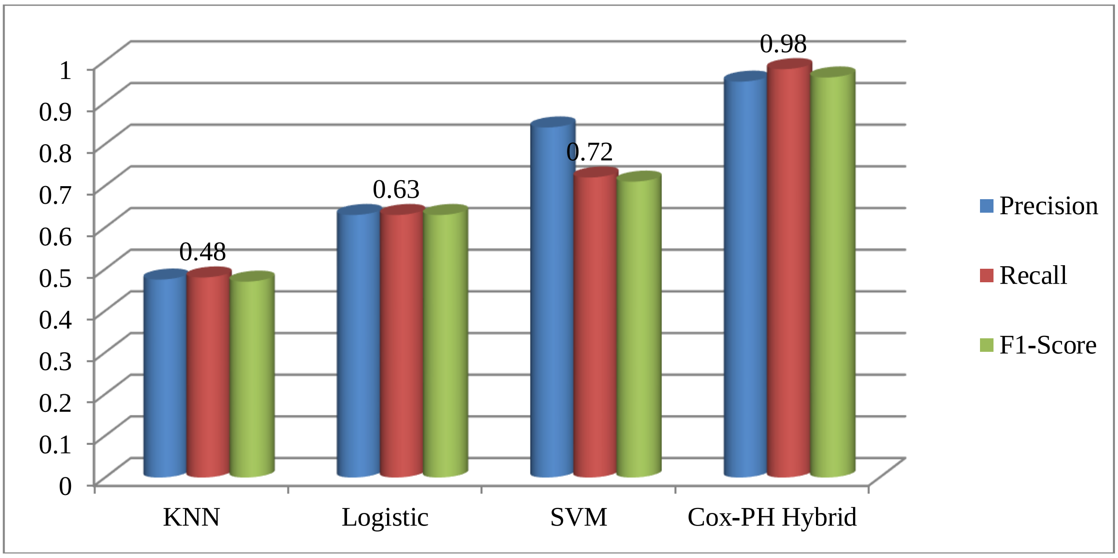

Fig. 3: Precision, Recall Value, and F-Measure Analysis on Diabetes Datasets

V. RESULT AND DISCUSSION

To perform the survival analysis of diabetes mellitus, this study has developed a hybrid model of implementing the Random Forest algorithm Cox-PH. From the performance and evaluation of the designed model, this study has shown that the integration of Machine Learning and the Cox-proportional hazard model in survival analysis is achievable. The result of the model on the dataset showed that the variables that are associated with diabetes mortality are; the age of the patient and diabetes types. The study shows that patients' hazard ratio reduces by 0.96 when they are young compared to when they are old. The patient's hazard ratio reduces by 0.53 when they have diabetes compared to when they have no diabetes.

Evaluating the algorithm's predictions using a statistic tool known as the concordance index or c-index on the Cox-PH concordance Index for the models with several covariates the best model with increasing index was used and deployed. The result from the fitting of the CoxPH model on the dataset shows that the variables that are associated with diabetes mortality are; the age of the patient and diabetes types. The patients' hazard ratio reduces by 0.96 when they are young compared to when they are old. The patient's hazard ratio reduces by 0.53 when they have diabetes compared to when they have no diabetes.

VI. CONCLUSION

The study has been able to establish that a machine learning model for survival analysis and prediction can be implemented along parameter or non-parameter tools for modeling time-to-event data. This is because machine learning tools and algorithms and efficient in building prediction and analysis models.

REFERENCES

[1.] S. I. Ayom and I. Milon, "Diabetes Prediction: A Deep Learning Approach. *I.J. Information Engineering and Electronic Business*, Vol2, pp.21–27, 2019.

[2.] J. Chaki, S.T. Ganesh, S. K. Cidham and S. A. Theertan, "Machine learning and artificial intelligence-based Diabetes Mellitus detection and self-management: A systematic review, *Journal of King Saud University – Computer and Information Sciences*, volxxx, no. xxxx, pp. 1–22, 2020.

[3.] E. U. Omede, “Optimization of determinant diagnostic symptoms for Febrile Diseases using Genetic Algorithm”, *Communication in Physical Sciences*, vol. 8 no. 4, pp. 556 – 572, 2022.

[4.] I. Ajenaghughrure, P. Sujatha, and A. Maureen, “Fuzzy based multi-fever symptom classifier diagnosis model”, *International Journal of Information Technology and Computer Science*, vol 9, no. 10, pp 13 – 28, DOI: 10.5815/ijitcs.2017.

[5.] B. O. Ojeme and M. Akazue, “Human immunodeficiency virus (HIV) diagnosis using Neuro-Fuzzy Expert System”, *Orient. J. Comp. Sci. and Technol.*, vol. 7, no. 2, pp 207 – 218, 2014.

[6.] E. Domingos, B. Ojeme and O. Daramola, “Experimental analysis of Hyper-parameters for Deep Learning-based Churn prediction in the banking sector”, *Computation*, vol. 9, no. 3, 2021.

[7.] B. Ojeme and A. Mbogho, “Selecting learning algorithms for simultaneous identification of depression and comorbid disorders, *Procedia Computer Science*, vol. 96, pp 1294 – 1303, 2016.

[8.] A. O. Amoo, T. O. Oyegoke, J. A. Balogun, S. A. Bamidele and P. A. Idowu, “Survival Model for Diabetes Mellitus Patient Receiving Treatment”,*International Journal of Computers*, Vol5,pp 1–13, 2020.

[9.] S.C. Chiemeke and E.U. Omede, “Mal-Typho Diagnosis intelligent system (MATDIS): The Auto-diagnostic rule generation algorithm”, vol. 5, no. 4, 2014.

[10.] T. Agbele, B. Ojeme and R. Jiang, “Application of local binary patterns and cascade AdaBoost classifier for mice behavioural patterns detection and analysis”, *Procedia Computer Science*, vol. 159, pp 1375 – 1386, 2019.

[11.] A. Alessandro and S. Stefano, “Information dropout: Learning optimal representations through noisy computation”,*IEEE Transactions on Pattern Analy-Sis and Machine Intelligence*, vol*40, no.* 12, 2897–2905, 2018.

[12.] M. Akazue, O. Ovoh, A. E. Edje, C. O. Ogeh and A. C. Hampo-JohnPaul, “An Enhanced Model for the Prediction of Cataract Using Bagging Techniques”, *International Journal of Innovative Science and Research Technology*,vol 8, no. 2, pp 220-227, 2023.

[13.] B. Zupan, J. Demsar, M.W. Kattan, J.R. Beck and I. Bratko, “Machine learning for survival analysis: a case study on recurrence of prostate cancer”, *Artificial Intelligence in Medicine*, Vol 20, pp. 59-75, 2000.

[14.] A. Gutierrez-Sacristan, A. Bravo, A. Giannoula, M.A. Mayer, F. Sanz and L.I. Furlong, “comorbidity: an R package for the systematic analysis of disease comorbidities, *Bionformatics*, vol 34, no. 18, pp. 3228 – 3230.

[15.] S. A. Bamidele, A. Asinobi, N. C. Egejuru and P. A. Idowu, “Survival Model for Diabetes Mellitus Patients’ Using Support Vector Machine”,*Computational Biology and Bioinformatics*, vol*8, no.* 2, pp.52–61. https://doi.org/10.11648/j.cbb.20200802.14, 2020.

[16.] J. Xie, Y, Liu, X Zeng, W. Zhang and Z. Mei, “A Bayesian network model for predicting type 2 diabetes risk based on electronic health records, *Modern PhysLett B*, vol 31, pp.31-21, 2017.

[17.] M. Akazue and N.F. Efozia, “A Review of Biometric Technique for Securing Corporate Stored Data,” Software Engineering and Intelligent Systems, vol. 1, pp. 329- 342, 2010.

[18.] B. Ojeme, M. Akazue and E. Nwalih, “Automatic Diagnosis of Depressive Disorders using Ensemble Techniques”, *African Journal of Computing & ICT*, 2016.

[19.] E. E. Abel, Ubiquitous computing: An assistive surveillance on in and out patients with mental illness via RFID”, *British Journal of Mathematics & Computer Science*, vol. 17, no. 3, pp 1-15, 2016.

[20.] P. A. Idowu, O. Agbelusi and T. A. Aladekomo, “The Prediction of Pediatric HIV/AIDS Patients’ Survival: A Data Mining Approach”,*Asian Journal of Computer and Information Systems*, vol*4, no* 3, pp. 287–294, 2016.

[21.] U. K.Okpeki, A. S.Adegoke and E. U.Omede, “Application of Artificial Intelligence for Facial Accreditation of Officials and Students for Examinations”,*FUPRE Journal of Scientific and Industrial Research*, vol. 6, no 3, pp 01 – 11, 2022.

[22.] S. N. Okofu, “Users service quality trust perception of online hotel room reservation”, *SAU Journal of Management and Social Sciences*, vol 3, no. 1&2, pp. 1-14.

[23.] U.K. Okpeki and E. U. Omede, “ Design and implementation of auto tech resource sharing system for secondary schools in Delta State”, *Journal of the Nigerian Association of Mathematical Physics*, vol. 51, pp 325 – 332.

[24.] N. F. Efozia, S. O. Anigbogu and K.S. Anigbogu, “Development of a hybrid model for enhancing data integration process of business intelligence”, *Journal of Basic Physical Research*, vol. 9, no. 2, pp 1-16, 2019.

[25.] M. I. Akazue and I. B. Ajenaghughrure, "Virtual Examination Supervision System for Nigerian Universities", *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 8, no. 9, pp. 43-50, 2016.

[26.] M. Akazue, “Enhanced hotel management information system for multiple reservation booking”, *Int. Manag. Rev.*, vol. 12, no. 1, pp 52, 2016.

[27.] U. Oijoe, “Authenticated and dynamic website: A sure control against website spoofing attacks”, *Journal of Software Engineering and Simulation*, vol. 9, no. 3, pp 57-60, 2023.

[28.] A.A. Ojugo and D.O. Otakore, “Intelligent cluster connectionist recommender system using implicit graph friendship algorithm for social networks”, *Int. Journal of Artificial Intelligence*,vol. 9, no.

3: pp497~506, doi: 10.11591/ijai.v9.i3.pp497~506, 2020.

[29.] A.A. Ojugo. And R. Yoro.,"Forging deep learning neural network intrusion detection framework to curb the distributed denial of service attack", *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp 1498-1509, 2021.

[30.] P. A. Idowu, T. A. Aladekomo, K.O. Williams and J. A. Balogun, "Predictive Model for Likelihood of Sickle Cell Anemia (SCA) among Pediatric Patients using Fuzzy Logic",*Transactions in Networks and Communications*, vol*31*, *no* 1, pp. 31–44, 2015.

[31.] A. O. Amoo, T. O. Oyegoke, J. A. Balogun, S. A. Bamidele and P. A. Idowu, "Survival Model for Diabetes Mellitus Patient Receiving Treatment",*International Journal of Computers*, Vol*5*,pp 1–13, 2020.

[32.] P. A. Idowu, O. Agbelusi and T. A. Aladekomo, "The Prediction of Pediatric HIV/AIDS Patients' Survival: A Data Mining Approach",*Asian Journal of Computer and Information Systems*, vol*4*, *no* 3, pp. 287–294, 2016.

[33.] A. Mohebbi, T. B. Aradóttir, A. R. Johansen, H. Fraccaro and M. Mørup, "A deep learning approach to adherence detection for type 2 diabetics"*39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)IEEE Engineering in Medicine and Biology Society (EMBC)*,Jeju, Korea (South). https://doi.org/10.1109/EMBC.2017.8037462, 2017.

[34.] M. D. Ganggayah, N. A. Taib, Y. C. Har,P.Lio, and S. K.Dhillon, "Predicting factors for survival of breast cancer patients using machine learning techniques", *BMC Medical Informatics and Decision Making,* vol 19, no. 48, pp. 1-17, 2019.

[35.] E. Y. Kalafi,N. A. M. Nor,N. A.Taib,M. D. Ganggayah,C.Town and S. K.Dhillon, "Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data", *Folia Biologica (Praha),*vol 65, pp. 212-220, 2019.

[36.] K. Tachkov, K. Mitov, Y. Koleva, Z. Mitkova, M. Kamusheva, M. Dimitrova, V. Petkova, A. Savova, M. Doneva et al., "Life expectancy and survival analysis of patients with diabetes compared to the non diabetic population in Bulgaria",*PLoS ONE*, vol 15, no. 5, pp. 1-16, 2020.

[37.] S. Lee, J. Zhou, K. S. K. Leung,W. K. K. Wu,W. T. Wong, T. Liu, I. C. K. Wong, K. Jeevaratnam, Q. Zhang and G.Tse, "Development of a predictive risk model for all-cause mortality in patients with diabetes in Hong Kong", *BMJ Open Diabetes Research & Care,* vol 9, pp. 1-12, 2021.

[38.] M.Atlam,H.Torkey,N.El-Fishawy and H. Salem, "Coronavirus disease 2019 (COVID- 19): Survival analysis using deep learning and Cox regression model", *Pattern Analysis and Applications,* vol 24, 2021.

[39.] D. V. Ojie, M. I. Akazue and A.Imianvan, "A Framework for Feature Selection using Data Value Metric and Genetic Algorithm", *International Journal of Computer Applications,* vol 184, no. 43, pp. 097-8887, 2023.