IOSR Journal of Mobile Computing & Application (IOSR-JMCA)
e- ISSN: 2394-0050, P-ISSN: 2394-0042.Volume 11, Issue 4 (Jul. - Aug. 2024), PP 01-09
www.iosrjournals.org

Machine Learning-Based Cardiovascular Disease Prediction System to reduce the incidence of occurrence

Sylvester Agbo Igwe[1] and Chukwu Jeremiah[2]
[1]*Computer Science Department, Ebonyi State University, P.M.B 053, Abakaliki, Ebonyi State*
[2]*Computer Science Department, Ebonyi State University, P.M.B 053, Abakaliki, Ebonyi State.*
Email: igwe.sylvester@ebsu.edu.ng (S. A. Igwe), chukwu.jeremiah@ebsu.edu.ng (J. Chukwu)
Correspondence: igwesylvesteragbo@gmail.com

ABSTRACT
Health is a paramount factor when considering any nation's sustainable growth and development. The concentration of the larger population of Nigerians is in the rural areas where public health infrastructures are either inadequate or not available. This has affected the larger labor force, leaving society vulnerable to high health. In recent times, one of the major health challenges facing rural and urban dwellers is heart disease occasioned by the non-existence of proper medical diagnosis. This study develops a machine learning model to predict the onset of heart disease to minimize onset cases of occurrence. The health prediction system is integrated with a prediction module to help in providing public healthcare services to rural dwellers. The developed system was comparatively achieved using hybrid machine learning techniques, which include a support vector machine and random forest in the design of the predictive module. The system integrates rural dwellers' health records with the global health database to form part of the Big Data for global access and considerations. The performance evaluation was achieved with the help of Random Forest. The system developed shows an average accuracy of 98% using machine learning algorithms. It also provides a convenient means for rural dwellers to access public health and enables them to receive health predictions to guide against fatal heart disease.
KEYWORDS*: Machine-Learning, Cardiovascular-Disease, occurrence, Prediction.*

Date of Submission: 03-07-2024 Date of Acceptance: 15-07-2024

I. INTRODUCTION

Disease prediction is a way to recognize patient health using appropriate techniques such as patient treatment history, deep learning, machine learning, etc. In recent times, humans have faced various life-threatening diseases such as malaria, cholera, COVID-19, dementia, cardiovascular (e.g. heart, hypertension, stroke, etc.), and diabetes, etc. due to the current environmental condition and their living habits (Alanazi, 2022; Alo *et al.*, 2022; Anikwe *et al*., 2022). Most rural dwellers normally have limited access to communication networks (e.g. internet, signal, etc.), poor living conditions, and inadequate healthcare facilities and services. In most cases, medical workers or specialists are not available always. Hence, the quality of healthcare services in that area is extremely low. Nonetheless, the earlier identification and prediction of chronic diseases are highly sorted to prevent life loss. The error due to humans and the inefficiency of some technology, it difficult for health workers or doctors to manually or technologically identify the diseases accurately in most cases. Therefore, appropriate techniques are needed to assist health workers medical personnel, and individuals to automatically detect and predict the future occurrence of diseases in urban and rural dwellers. Some of the major health challenges faced by rural and urban dwellers in recent times are hypertension, diabetes, and strokes.

The occurrence of cardiovascular diseases, hypertension, and diabetes in stroke patients was evaluated in a retrospective epidemiological study in the Ebonyi State. During the years 2019 through 2022, 5522 new stroke cases (3010 males and 2022 females) were diagnosed and included in the study. Cerebral ischemia was diagnosed in 509 patients (33%), 181 patients (12%) had an intracerebral or subarachnoid hemorrhage and 832 patients (55%) had a stroke of undetermined type. For the total stroke series, 42% had hypertension. Almost the same percentage was found for males (41%) and females (43%). There was almost no sex predominance in the hypertensive stroke cases in the different age groups and for the various types of stroke. The frequency of hypertension among the stroke cases was low in the 40 to 49 age group, higher in the 50 to 59 age group, maximal in the 60 to 69 age group, and declining in the above 70 age group. The percentage of hypertensive was about the same for the ischemic and the undetermined types of stroke and for the total stroke series in the different age groups. It was found to be slightly higher in the hemorrhage type.

Heart disease (Singh *et al*., 2021) refers to several types of heart conditions. The most common type of heart disease is coronary artery disease (CAD) affects the blood flow to the heart. Decreased blood flow leads to

a heart attack. The symptoms associated with heart diseases include heart attack (e.g. chest pain, upper back/neck pain, indigestion, heartburn, etc.), heart failure (e.g. shortness of breath, fatigue, neck veins, etc.), and or/ an arrhythmia (e.g. palpitations). The risk factors include high blood pressure, high blood cholesterol, and smoking. Heart disease has become a silent killer disease that affects a lot of people living in rural and urban areas. However, severally medical attention and brazen research efforts have been made or predicted using various techniques such as machine learning, deep learning, etc. to address this instant killer disease. For instance, Singh et al., (2021) predicted heart disease using machine learning. However, more predictive algorithms or techniques are needed. Heart disease remains a silent killer disease that requires more research efforts to address it. Machine learning algorithms are also utilized for the detection or prediction of heart-related diseases in (Wehner *et al*., 2017; Yadav and Jadhav, 2019; Katarya and Meena,2021).

The contributions of this study to the body of knowledge are:

- Undertake a comprehensive approach to a machine learning-based heart disease prediction system;
- Outline important features for health disease prediction;
- Implemented multiple classifier systems for heart disease prediction using a random forest algorithm;
- Evaluate the implemented heart disease prediction system using various machine learning evaluation approaches.

II. METHODOLOGY

This section discusses the implementation of the proposed cardiovascular disease prediction using machine learning algorithms. Here, we discuss the various stages implemented for the system. These stages include data collection, preprocessing, feature extraction, building the prediction model, and evaluation of the prediction model. In addition, the prediction model was implemented using a dataset collected from patients with cardiovascular diseases and patients without cardiovascular diseases. The dataset has different characteristics to enable the developed system to accurately detect the diseases.

2.1 DATASET DESCRIPTION

There are various diseases faced by rural dwellers in Ebonyi state and beyond. These diseases are outlined in chapter two of this thesis and they include malaria, high blood pressure, stroke, cardiovascular disease, Lassa fever, tuberculosis, cancer, etc. Among these, cardiovascular disease (CVD) is difficult to detect due to its interconnection with other killer diseases such as high blood pressure, diabetes, and stroke. Cardiovascular disease affects the heart and blood vessels and may be symptomatic (showing some signs in the patients) or asymptomatic (having no signs of the disease). It accounts for major unexplained deaths in Ebonyi state as before detection by doctors, it might have become irreversible. A recent study by (Eke *et al*, 2020) shows the high prevalence of cardiovascular among rural dwellers which are farmers, petty traders, and laborers. The study noted that risk factors such as hypertension, diabetes, alcohol abuse, family history of stroke, and previous stroke that have become predominantly among rural dwellers have contributed to the high incidence of disease. Although the popular belief is that cardiovascular diseases are common in urban areas, recent studies on the accessibility of medications for treating hypertension (Acton *et al*, 2018) show an increase in the number of people with cardiovascular disease among rural dwellers in Ebonyi state. Consequently, cardiovascular disease may go undetected and untreated thereby resulting to an increased death rate. This thesis developed a machine learning system to detect the early occurrence of cardiovascular disease among rural dwellers using data collected from mobile apps and health centers.

To access the performance of the proposed cardiovascular disease prediction system, data from the developed mobile-based public health information system, and patient health history were integrated with other data crawled from patients' information. Here, the implementation specifically focused on cardiovascular disease prediction as it has become one of the diseases ravaging the rural areas, especially in Ebonyi state. One major challenge of detecting cardiovascular disease manually by doctors is the lack of visible symptoms.

The data for developing the model consisted of seventy-six (76) features, but some of these features were not useful for cardiovascular disease prediction. Consequently, the implementation utilized fourteen (14) features/attributes that are mainly linked with the onset of cardiovascular disease as shown in Figure 1.

Age	Sex	Chest pain	BP	Cholesterc	FBS over 1	EKG result	Max HR	Exercise a	ST depress	Slope of S	Number o	Thallium	Heart Disease
70	1	4	130	322	0	2	109	0	2.4	2	3	3	Presence
67	0	3	115	564	0	2	160	0	1.6	2	0	7	Absence
57	1	2	124	261	0	0	141	0	0.3	1	0	7	Presence
64	1	4	128	263	0	0	105	1	0.2	2	1	7	Absence
74	0	2	120	269	0	2	121	1	0.2	1	1	3	Absence
65	1	4	120	177	0	0	140	0	0.4	1	0	7	Absence
56	1	3	130	256	1	2	142	1	0.6	2	1	6	Presence
59	1	4	110	239	0	2	142	1	1.2	2	1	7	Presence
60	1	4	140	293	0	2	170	0	1.2	2	2	7	Presence
63	0	4	150	407	0	2	154	0	4	2	3	7	Presence
59	1	4	135	234	0	0	161	0	0.5	2	0	7	Absence
53	1	4	142	226	0	2	111	1	0	1	0	7	Absence
44	1	3	140	235	0	2	180	0	0	1	0	3	Absence
61	1	1	134	234	0	0	145	0	2.6	2	2	3	Presence
57	0	4	128	303	0	2	159	0	0	1	1	3	Absence
71	0	4	112	149	0	0	125	0	1.6	2	0	3	Absence
46	1	4	140	311	0	0	120	1	1.8	2	2	7	Presence
53	1	4	140	203	1	2	155	1	3.1	3	0	7	Presence
64	1	1	110	211	0	2	144	1	1.8	2	0	3	Absence
40	1	1	140	199	0	0	178	1	1.4	1	0	7	Absence
67	1	4	120	229	0	2	129	1	2.6	2	2	7	Presence
48	1	2	130	245	0	2	180	0	0.2	2	0	3	Absence
43	1	4	115	303	0	0	181	0	1.2	2	0	3	Absence
47	1	4	112	204	0	0	143	0	0.1	1	0	3	Absence
54	0	2	132	288	1	2	159	1	0	1	1	3	Absence
48	0	3	130	275	0	0	139	0	0.2	1	0	3	Absence
46	0	4	138	243	0	2	152	1	0	2	0	3	Absence
51	0	3	120	295	0	2	157	0	0.6	1	0	3	Absence
58	1	3	112	230	0	2	165	0	2.5	2	1	7	Presence

Figure 1: Sample Cardiovascular disease dataset

In the figure, samples with presence are treated as a positive class while the sample with absence is treated as a negative class. The dataset contains 499 samples of negative classes and 526 samples of positive classes as shown in Figure 1. In this case, 51.32% of the patients have the presence of cardiovascular diseases while 48.68% of the patients do not have cardiovascular-related diseases.

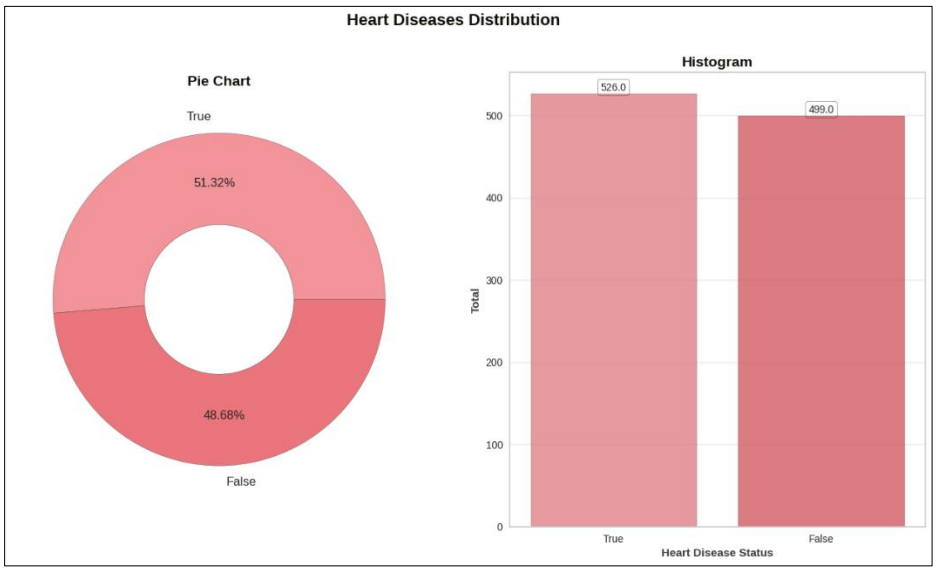

Figure 2: Distribution of patients with heart disease and those without heart disease

These include 713 females where 300 patients having heart disease and 312 males with 226 patients having heart disease as shown in Figure 2. In addition, the diagram shows that the mean age of rural dwellers with cardiovascular disease is 50 years and above.

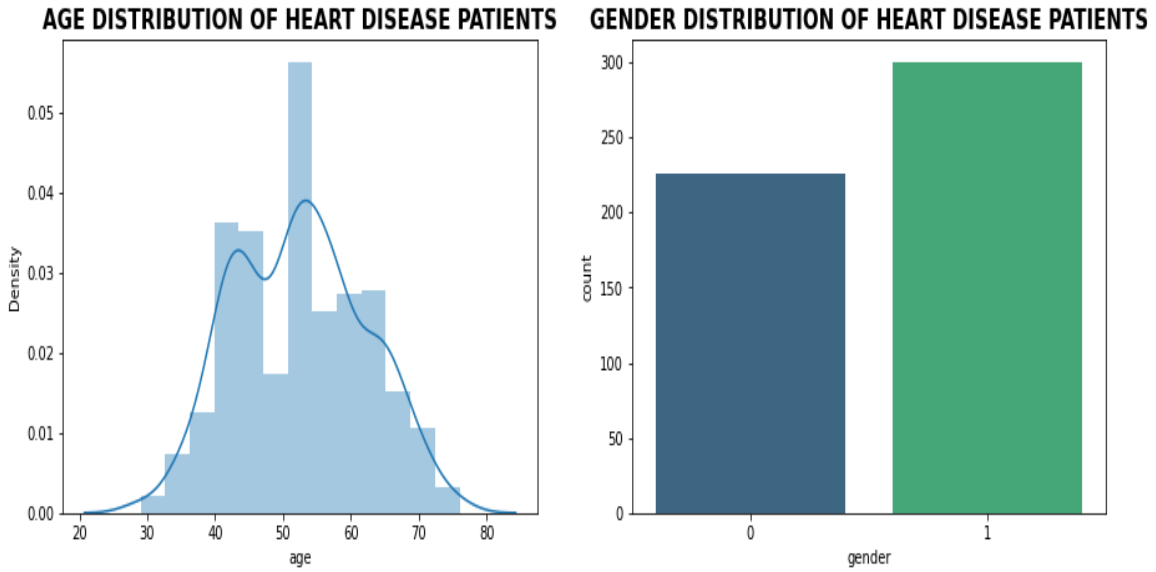

Figure 3: Gender distribution among patient with heart disease

The figure 3 shows that female patients have more tendencies to develop heart disease than their male counterparts. The entire data collected were saved for pre-processing and further analysis.

2.2 Pre-processing

In most cases, the information collected from patients are affected by noise, duplication anomalies, and some missing information. Accordingly, data preprocessing methods are used to remove anomalies and data duplication in patient data. In addition, inputting missing values using an average of each column is also an important method to ensure improved performance of the disease prediction system. Here, the missing values were replaced with the mean of each data. Data with duplicate values were removed, and noise and errors were identified and removed. The pre-processed data were saved as comma-separated values (CSV) file format for feature analysis.

2.3 Feature analysis and normalization

Here, the implementation identified valuable and relevant attributes that would predict the outcome of cardiovascular disease among rural dwellers who visit the hospital. To ensure a comprehensive analysis of factors that contribute to cardiovascular disease, various information was collected and analyzed for their discriminative factors. Some of the attributes that were identified for the detection of cardiovascular disease among rural dwellers include age, sex, chest pain, high cholesterol level, elevated resting blood pressure, high fasting blood sugar level, resting electrocardiography, and lack of physical movement, especially among elderly populations. Others include maximum heart rate, exercise-induced angina, depression by exercise, slope of the peak by exercise segment, blood vessel blockage, and diagnosis of heart disease angiographic disease status. These features were integrated with patients' demographic information, and form the dataset for predicting cardiovascular disease among rural dwellers in Ebonyi state. The details of some of the features are highlighted in Table 1.

Table 1: Attributes/features of the dataset used for disease prediction

Attributes	Description	Types	Values
Age	Age of the patient	integer	[40-77]
Sex	Gender of the patient	Integer	Male =1; female = 0
Cp	Chest pain type	Integer	Angina =1;abnanr=2; notang=3, asympt=4
Trestbps	Resting Blood pressure value	integer	[94-200]
Chol	Cholesterol	Integer	[126-564]
Fbs	Fasting blood sugar	Integer	True=1;false=0
Restecg	Resting electrocardiographic results	Integer	[0-2]
Thalach	Maximum heart rate	Integer	[71-202]
Exang	Angina induced exercise	Integer	[1-4]=yes; 0=no
Avg_glucose	Average glucose level	Float	[50-250]
Oldpeak	Depression induced by exercise	Float	[0-4

Attributes	Description	Types	Values
Slope	Slope of the peak exercise	Integer	Upsloping=1; flat=2; downsloping=3
Ca	Number of major blood vessel	Integer	[0-3]
Thal	Blood vessel status	Integer	Normal=3; fixed defect=7; reversible defect=7
Target class	Cardiovascular disease diagnosed	Integer	Present=1; absent=0

To ensure that the features of the data collected correlated with each other, feature correlation was performed on the data. In this case, data visualization was used to ascertain the discriminative strength of each attribute included in the dataset for implementing the cardiovascular disease prediction system. Here, a heatmap as shown in Figure 4 was used to visualize the relationship between the data values in the dataset. In the diagram, there are indications that cardiovascular disease symptoms such as elevated blood pressure, high cholesterol, and consistent depression have a moderate relationship with the age of patients. Other important factors that might be indicative of cardiovascular disease among rural dwellers in Ebonyi state are fasting blood sugar level, maximum heart rate, chest pain, and blood vessel status.

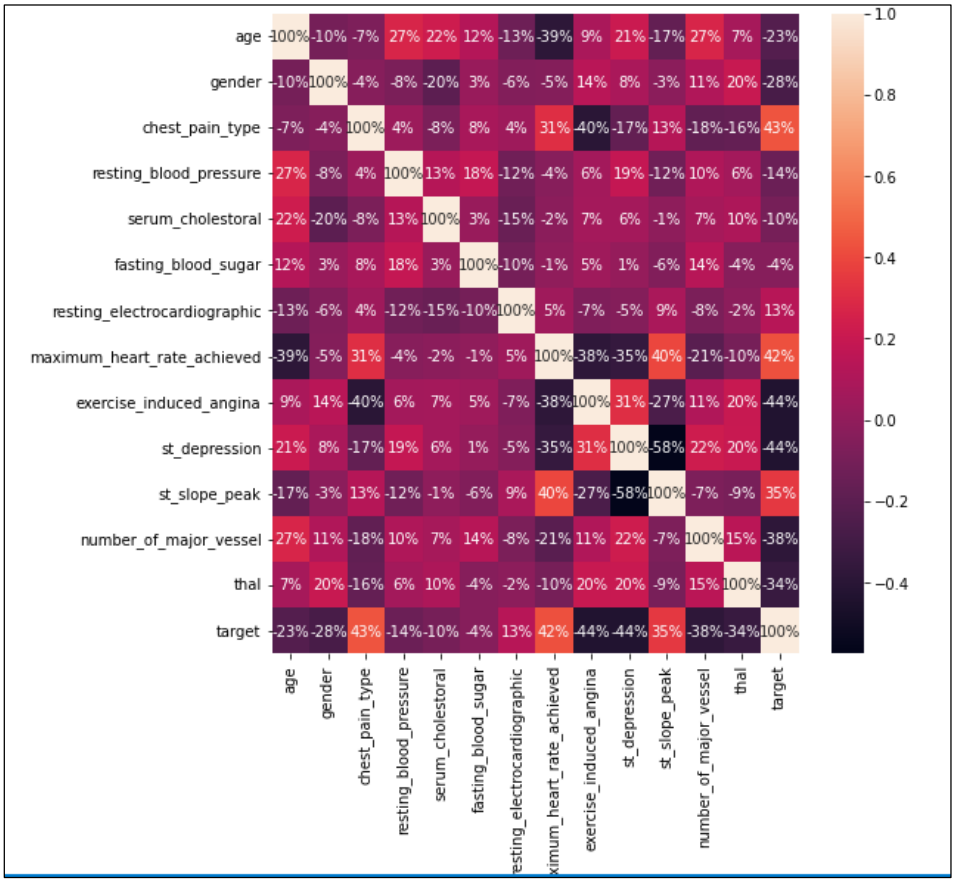

Figure 4: Relationship between each feature values and target classes

After the identification of features, then the features were normalized to zero mean and unit variance to reduce the features to a certain range. Feature normalization is important in prediction system implementation as it helps to improve its performance. The Z-score normalization was used where the mean value of the feature vectors was subtracted from the individual feature value point and divided by the standard deviation. Equation 1 below shows the formula for computing Z-score normalization where $\acute{x}$ represent the computed z-score, $\bar{x}$ is the mean value, x is the individual features and α is the standard deviation.

$$\acute{x} = \frac{\bar{x} - x}{\alpha} \qquad (1)$$

The normalized features were combined into a master feature vector and saved as .csv for heart disease prediction system implementation. In addition, some of the feature cluster distributions such as blood pressure, cholesterol, and age are shown in Figure 5.

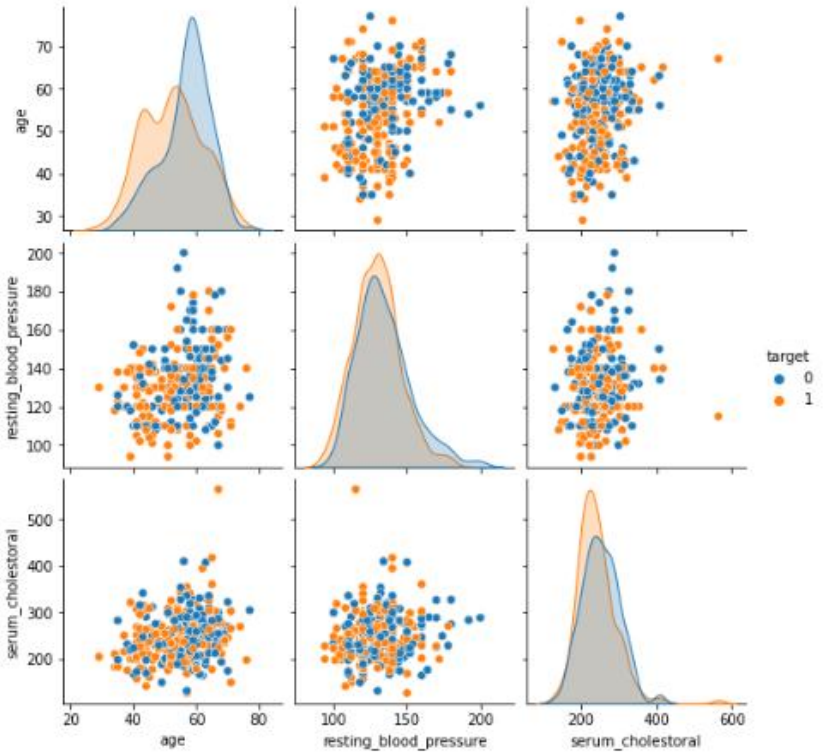

Figure 5: important feature distributions

2.4 Cardiovascular Disease Prediction Model

Supervised machine learning is one of the most widely utilized methods for disease prediction. Therefore, this thesis implemented a supervised machine learning model to predict the onset of cardiovascular disease using the dataset described earlier. Supervised machine learning is the type of machine learning where a labeled dataset is used to train the prediction models. Here, we proposed a random forest-based disease prediction system. Random forest is an ensemble machine learning algorithms that combine multiple decisions and was first developed by (Breiman, 2001). The algorithm is made up of different decision tree structures, where the independent and identically distributed random vectors of each tree cast a unit vote for the most popular class label during training. Random forest can be built in five steps as outlined below:

i. Create the training data;
ii. Choose the random decision tree;
iii. Compute the decision split of each node;
iv. Aggregate the trees to form the random forest algorithms.

Random forest algorithm can improve the prediction of cardiovascular disease using patients' demographic information, medical history, and doctor reports. In this thesis, the random forest-based algorithm was implemented to predict the onset of cardiovascular disease among rural dwellers in Ebonyi state.

2.5 Evaluation of the Cardiovascular Disease Prediction System

To evaluate the performances of the proposed prediction system to detect cardiovascular disease using the collected data, different performance metrics were utilized. These performance metrics include accuracy, recall, precision, f-measures, Area under the curve (AUC), and confusion matrix. Accuracy measures the ratio of correctly predicted cardiovascular disease to the total number of both the presence and absence of cardiovascular-related disease among rural dwellers in Ebonyi state. Area under the curve (AUC) measures the performance of the disease prediction system in which the system can predict cases with cardiovascular disease more than those without cardiovascular disease. The AUC shows the performance of the classification model using two parameters (True Positive Rate and False Positive Rate) at various threshold values. AUC ranges in value from 0 to 1. Then, the confusion matrix combines two or more evaluation or classification models for which set of test data whose true values are known. Furthermore, these performance metrics are mathematically represented as:

$$\text{Accuracy (ACC)} = \frac{(TP+TN)}{(TP+TN+FP+FN)};$$

$$\text{Recall} = \frac{TP}{(TP+FN)};$$

$$\text{Precision} = \frac{TP}{(TP+FP)};$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive and false negative respectively.

2.6 Comparison with other cardiovascular disease prediction Systems

To measure the significance of the Random forest-based cardiovascular disease prediction system, the system was compared with two machine learning-based cardiovascular disease prediction systems that has played vital in recent research (Jiang *et al*, 2021). The machine learning algorithms include Support vector machines and k-nearest Neighbors. Using the same features deployed to implement the proposed Random forest-based cardiovascular disease prediction system, various experiments were run, and results were obtained for the comparison. This was done to ensure consistency in comparison.

III. Experimental setting

The cardiovascular disease prediction system was implemented in Python, and each stage such as pre-processing, feature analysis and algorithm development were implemented in computer on running Window 10 operating system. The system is using Intel Core 2 duo processor @ 3.400 with install Random Access Memory (RAM) capacity of 2GB. The whole pre-processed data were divided into training and testing parts. Here, 70% of the data were used for training the cardiovascular disease prediction system while 30% was utilized to test the developed system. Using the train-test data partitioning approach ensure uniformity and reduce complexity. Significant hyper-parameter settings for Random Forest, support vector machine and k-Nearest Neighbors are as outlined in Table 2. These parameter values are default and were chosen based on empirical evaluation of machine learning algorithms for heart disease prediction (Nahar *et al*, 2013), and reported improved performance results. Furthermore, using some of the default values of the machine learning models would ensure reproducibility of the algorithms in similar scenarios.

Table 2: parameter values for each machine learning model

Machine learning algorithms	Parameter Tuning
Random forest (RF)	n_estimator=100;criterion=gini;max_depth=None;min_sample_split=2;random_state=None;verbose=0
Support vector machine (SVM)	Kernel=rbf;C=1;gamma=scale;degree=3;max_iter=-1;random_state=None;verbose=false
k-Nearest Neighbors (K-NN)	k=5;weight=uniform;algorithm=kd_tree;metric=minkowski;leaf_size=30;

IV. RESULTS AND DISCUSSION

In this section, the results obtained with a random forest-based cardiovascular disease prediction system are presented. The health information collected during patient registration, visitation, and similar data crawled online were saved in the database and utilized to develop the random forest-based cardiovascular disease prediction, and its comparison with support vector machine and k-Nearest Neighbors algorithms. First, the data were saved in comma-separated values (CSV) alongside the target labels, and sent to the listed machine learning algorithms for cardiovascular disease prediction. The results obtained are presented in table 3. From the table, the implemented random forest algorithm achieved prediction accuracy of 98.92% and outperformed other machine learning algorithms. Here, k-Nearest Neighbors achieved accuracy of 76.01% followed by support vector machine which achieved prediction accuracy of 71.28%. When compared with the developed random forest-based disease prediction system, there are high improvement in the system by 23% for the random forest-based cardiovascular disease prediction system as shown in Table 3

Table 3: Performance results of the cardiovascular disease prediction system

Methods	Accuracy	Precision	Recall	F1-Score
Proposed Method (Random Forest)	**98.92%**	**1.00**	**1.00**	**1.00**
k-Nearest Neighbors (k-NN)	76.01%	0.70	0.72	0.712
Support vector machine(SVM)	71.28%	0.69	0.71	0.73

In addition, table 3 presents the proportions of cardiovascular disease that occurred among rural dweller actually and correctly identified by the proposed random forest-based machine learning. The results obtained showed the ability of the designed prediction system to correctly identify cardiovascular disease among rural dwellers. However, other machine learning models such as k-Nearest Neighbors and Support vector machines achieved lower ability to correctly identify patients with cardiovascular disease as presented in Table 3.

The confusion matrices of the proposed cardiovascular disease prediction system using a random forest algorithm is shown in Figure 6. In a disease prediction system, a confusion matrix helps to visualize and represent the prediction outcome of the designed system. It represents the true positive (actual number of cardiovascular disease cases) and false positive (actual number of cardiovascular disease cases incorrectly predicted) using the implemented random forest algorithms. Here, "*0*" in the figure represents the absence of cardiovascular disease while "*1*" represent the presence of cardiovascular disease.

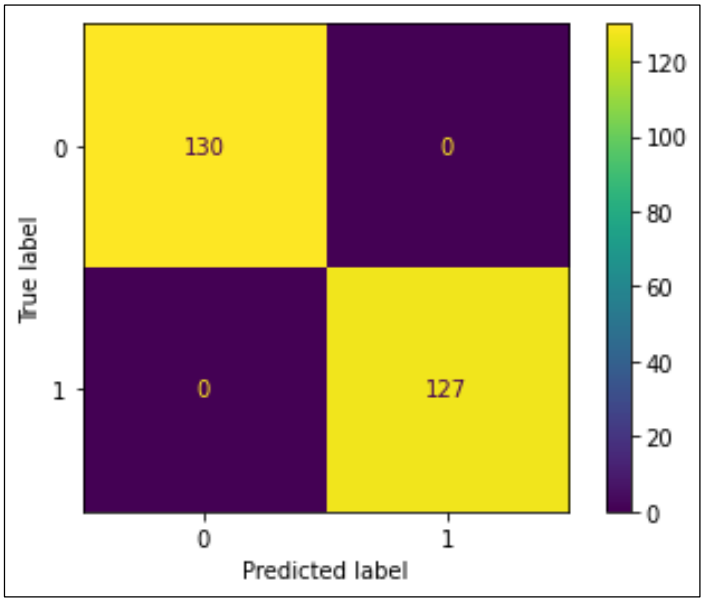

Figure 6: Confusion matrix of the implemented Random Forest based disease prediction

From Figure 6, random forest predicted all patients with cardiovascular disease, 127 in this case and those without a case of cardiovascular disease which included 130 patients. The confusion matrix obtained with the support vector machine is shown in Figure 7.

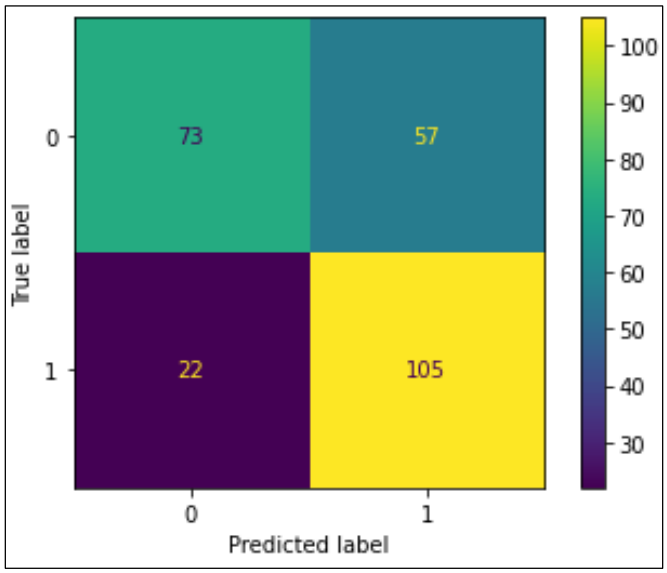

Figure 7: Confusion matrix for Support vector machine

Figure 7 represents the confusion matrix of the support vector machine. Here, the support vector machine learning model performed poorly at predicting cases of cardiovascular disease.
Largely, the implemented random forest-based cardiovascular disease prediction system achieved optimal results for predicting cardiovascular disease among rural dweller using their health and demographic information.

V. CONCLUSION

Despite numerous machine-learning algorithms, determining the best suitable algorithm that is feasible to suit datasets for reducing cardiovascular disease (CVD) occurrence in rural areas remains a challenge. The CVD prediction model using ML algorithms exhibited superior validation performances compared to those of previously proposed prediction studies. It was also possible to predict CVD occurrence in rural areas of Ebonyi state, readily using existing health screening data and ML algorithms. Therefore, our study verifies that a CVD prediction model using hybrid ML algorithm techniques: Support vector machine and Random forest techniques can predict CVD effectively. The paper also investigates the feasibility and utility of various machine-learning algorithms. Among the factors considered in this study, the preexisting history of CVD was the most important contributing factor to the prediction model performance employing machine learning techniques. The performance of the developed prediction system was evaluated using various metrics to identify the most suitable machine learning model. When it came to predicting cardiovascular disease patients, the Random Forest model performed exceptionally well with the highest accuracy of 98% and the quickest prediction time of 0.01(secs). The new system projects the rural dwellers in the global health image to equally access standard health resources globally. It also provides convenient means for rural dwellers to access public health and enables them to receive health tips and awareness as well as health prediction to guide against fatal health breakdown of the people.

REFERENCES

[1]. Acton, T., Elsaleh, T., Hassanpour, M. and Ahrabian, A. (2018). Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. PloS One, 13(5), 1–20.

[2]. Alanazi, R. (2022). Identification and Prediction of Chronic Diseases Using Machine Learning Approach. Journal of Healthcare Engineering, 2022. https://doi.org/10.1155/2022/2826127

[3]. Alo, U. R., Nkwo, F. O., Nweke, H. F., Achi, I. I. and Okemiri, H. A. (2022). Non-Pharmaceutical Interventions against COVID-19 Pandemic : Review of Contact Tracing and Social Distancing. Sensor Review, 22(1), 280.

[4]. Anikwe, C. V., Nweke, H. F., Ikegwu, A. C., Egwuonwu, A. C., Onu, F. U., Alo, U. R. and Wah, T. Y. (2022). Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and future prospect. Expert Systems with Applications, 202, 117362. https://doi.org/10.1016/j.eswa.2022.117362

[5]. Berman, A. C. and Chutka, D. S. (2016). Assessing effective physician-patient communication skills: "Are you listening to me, doc?" Korean Journal of Medical Education, 28(2), 243–249. https://doi.org/10.3946/kjme.2016.21

[6]. Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R. and Okonkwo, O. R. (2022). Big Data Analytics for Data-driven Industry: A Review of Data Sources, Tools, Challenges, Solutions and Research Directions. Cluster Computing.

[7]. Jiang, H., Huang, Y. and You, Z. (2020). OPEN SAEROF : an ensemble approach for large-scale drug-disease association prediction by incorporating rotation forest and sparse autoencoder deep neural network. Scientific Reports, 10(4972), 1–11. https://doi.org/10.1038/s41598-020-61616-9

[8]. Katarya, R. and Meena, S. K. (2021). Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. Health and Technology, 11(1), 87–97. https://doi.org/10.1007/s12553-020-00505-7

[9]. Nada, W., Panpiemras, J. and Manachotphong, W. (2020). The Impact of a Billing System on Healthcare Utilization : Evidence from the Thai Civil Servant Medical Benefit Scheme. Oxford Bulletin of Economics and Statistics, I12(D12), 0305–9049. https://doi.org/10.1111/obes.12376

[10]. Nweke, H. F., Teh, Y. W., Mujtaba, G. and Al-Garadi, M. A. (2019). Data Fusion and Multiple Classifier Systems for Human Activity Detection and Health Monitoring: Review and Open Research Directions. Information Fusion, 46, 147–170. https://doi.org/10.1016/j.inffus.2018.06.002

[11]. Singh, H., Gupta, T. and Sidhu, J. (2021). Prediction of Heart Disease using Machine Learning Techniques. Proceedings of the IEEE International Conference Image Information Processing, Iceca, 164–169. https://doi.org/10.1109/ICIIP53038.2021.9702625

[12]. Singh, P., Singh, N., Singh, K. K., and Singh, A. (2021). Diagnosing of disease using machine learning. Machine Learning and the Internet of Medical Things in Healthcare, 89–111. https://doi.org/10.1016/B978-0-12-821229-5.00003-3

[13]. Wehner, M. R., Levandoski, K. A., Kulldorff, M. and Asgari, M. M. (2017). Research Techniques Made Simple: An Introduction to Use and Analysis of Big Data in Dermatology. Journal of Investigative Dermatology, 137(8), e153–e158. https://doi.org/10.1016/j.jid.2017.04.019

[14]. Yadav, S. S. and Jadhav, S. M. (2019). Machine learning algorithms for disease prediction using Iot environment. International Journal of Engineering and Advanced Technology, 8(6), 4303–4307. https://doi.org/10.35940/ijeat.F8914.088619