Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 3(3) May 2025
DOI: https://dx.doi.org/10.4314/jobasr.v3i3.15

Deploying Machine Learning Methods for Health Monitoring and Cardiovascular Disease Prediction

Anikwe, C. V.[1]; Nweke, H. F.[1, 2*]; Ikegwu, A. C.[3] and Ndukwe, O. E.
[1]Computer Science Department, David Umahi Federal University of Health Sciences, P.M.B 211, Uburu, Ebonyi State, Nigeria.
[2]Institute for Machine Learning, Robotics and Artificial Intelligence Research, David Umahi Federal University of Health Sciences, P.M.B 211, Uburu, Ebonyi State, Nigeria.
[3]Software Engineering Department, Veritas University, Abuja, Nigeria.
[4]Computer Science Department, University of Port Harcourt, Rivers State, Nigeria.
*Corresponding Author Email: nwekehf@dufuhs.edu.ng
[1]anikwevc@dufuhs.edu.ng, [3]ikegwua@veritas.edu.ng, [4]ogeking2004@gmail.com

ABSTRACT
Smart healthcare has increased to meet the needs of the growing human population and medical expenses. People are all hurrying to catch up with work schedules, academic appointments, and social engagements, especially in this jet age. These often happen at the detriment of our health. Healthcare services, especially those that provide optimal healthcare delivery, face many problems, such as the ineffective provision ofhealth monitoring applications andless emphasis on disease prediction systems.Severalresearchstudies have been carried out in an attempt to proffer solutions to the peculiar problems; however, the problem persists. Therefore, this paper develops a Cardiovascular disease prediction system with specific objectives: implement data analysis for disease prediction using a k-Nearest Neighbors (k-NN)-based machine learning system; evaluate the performance of the developed cardiovascular disease prediction system with existing health monitoring systems. The k-Nearest Neighbors was utilized using a 1025 dataset and 18 attributes collected from the UCI machine learning repository. The results show that k-NN achieved an accuracy of 99.21%. k-Nearest Neighbors algorithm is a non-parametric machine learning that majority voting to classify new case of cardiovascular disease, and non-sensitive to noise and outlier. The proposed model is higher than the existing system, which shows an average accuracy result of 84.63%. The developed machine learning approach will guide healthcare practitioners on the use of machine learning for cardiovascular disease diagnosis and prediction.

Keywords:
Machine Learning,
Mobile Sensors,
Health Monitoring,
Disease Prediction,
Artificial Intelligence.

INTRODUCTION

Mobile sensor-based, especially the wearable components, are a crucial part of the new telemedicine paradigm that is deployed to promote the notion that the use of effective information and digital technologies will lead to an improvement of high-quality healthcare delivery services(Vesnic-Alujevic *et al.*, 2018; Deshpande *et al.*, 2017). Wearable devices are integrated Internet of Things (IoT) hardware elementsembedded in electronic devices such as mobile phones, smartwatches, headphones, etc. that can record several kinds of signals, including activity, physiological, and environmental (De Fazio *et al.*, 2023; El-deep *et al.*, 2025). In addition, mobile-based sensors for health monitoring have been widely used in several areas, such as diabetes, human activity recognition, stroke, fall detection, stress detection, cancer, cardiovascular disease management, and disease diagnosis of well-being (Anikwe *et al.*, 2022; Enshaeifar *et al.*, 2018; Islam *et al.*, 2020; Singh *et al.*, 2021; Thaung *et al.*, 2020). In addition to helping people engage with their bodies and minds, these have the potential to transform medical practice and the roles that patients and healthcare providers play. Furthermore, following the onset of illnesses or other life-threatening and emerging viral infectious diseases such as the coronavirus (COVID-19) (Wu *et al.*, 2020), Zika virus (ZKV) (Ozdener *et al.*, 2020), Ebola virus disease (EBOV), etc., patients can now obtain diagnosis, treatment, and monitoring from home or in other remote areas.

Furthermore, several problems havehampered the successful implementation ofmobile healthcare service delivery and disease prediction systems. These problems include ineffective computing technology-based sensors for data collection, transmission, integration, and system analysis of the societal population in providing accurate information to prevent diseases, ensure the diagnosis of common illnesses, and raise health awareness. Additionally, in both rural and urban areas, there is a shortage of medical workers, which leads to delayed diagnosis and a lack of knowledge about medical health information, particularly when caring for senior patients over the age of 55 (Ayon *et al.*, 2020).

Numerous mobile applications, including fitness and health trackers with calorie counters, Health Mate—a comprehensive health tracking app, an instant heart rate monitor and pulse checker, and cardiograph, which have been created to track various physiological signals and health-related issues, none have been used to cardiovascular disease especially among elderly population (Anikwe *et al.*, 2022; Fisk *et al.*, 2020). For instance, utilizing a study from a pilot telemedicine project in Southern Italy, Marino *et al.* (2020) maintained mobile screening units for the early diagnosis of breast cancer and cardiovascular disease. The medical records from the diagnosed diseases are needed for proper disease prediction. Nadakinamani *et al.* (2022) utilized supervised machine learning methods like Random Forest (RF), Naïve Bayes (NB), and linear regression (LR) to clinically analyze patient data related to cardiovascular illness (e.g., hypertension) to predict the development of HBP. In this sense, machine learning (ML) keeps performing well in various areas of human endeavors (Sada *et al.*, 2024) especially in disease prediction. For instance, Menon *et al.* (2023) used machine learning techniques like advanced spatial vector-based random forest (ASV-RF) in conjunction with an intelligent system to monitor 62 diabetic patients. Islam *et al.* (2023) utilized machine learning algorithms to predict the risk level of cardiovascular (CVD) diseases using the UCI dataset of 920. The study achieved F1 scores of 80.4% and 91% for high and low-risk CVD, respectively. Nonetheless, the proposed system required real-time dataset testing.

Nigeria populations are faced with lots of health challenges such as rheumatism, arthritis, waist pain, cardiovascular diseases, stroke, Alzheimer's disease, Parkinson's disease, accidental falls, M-pox disease, etc. (Olopade *et al.*, 2024), which may hamper their general well-being. Adversely, providing a mobile-basedapplication system and prediction of disease is important to manage and provide necessary health information for them.

The contribution of this work is bestowed in the following:

- implement a mobile-based health monitoring and disease prediction system;
- perform data analysis for disease prediction using machine learning algorithms such as K-Nearest Neighbors from the data obtained from the mobile-based sensor;
- evaluate the performance of the proposed system with baseline studies.

The other sections of the paper include Section 2– the methodology; Section 3 – results and discussions; and Section 4 – conclusion.

MATERIALS AND METHODS

Different methods utilised to implement mobile sensor-based health monitoring and disease prediction are discussed in a subsection. The process deployed in the implementation of the envisaged system is depicted in Figure 1.

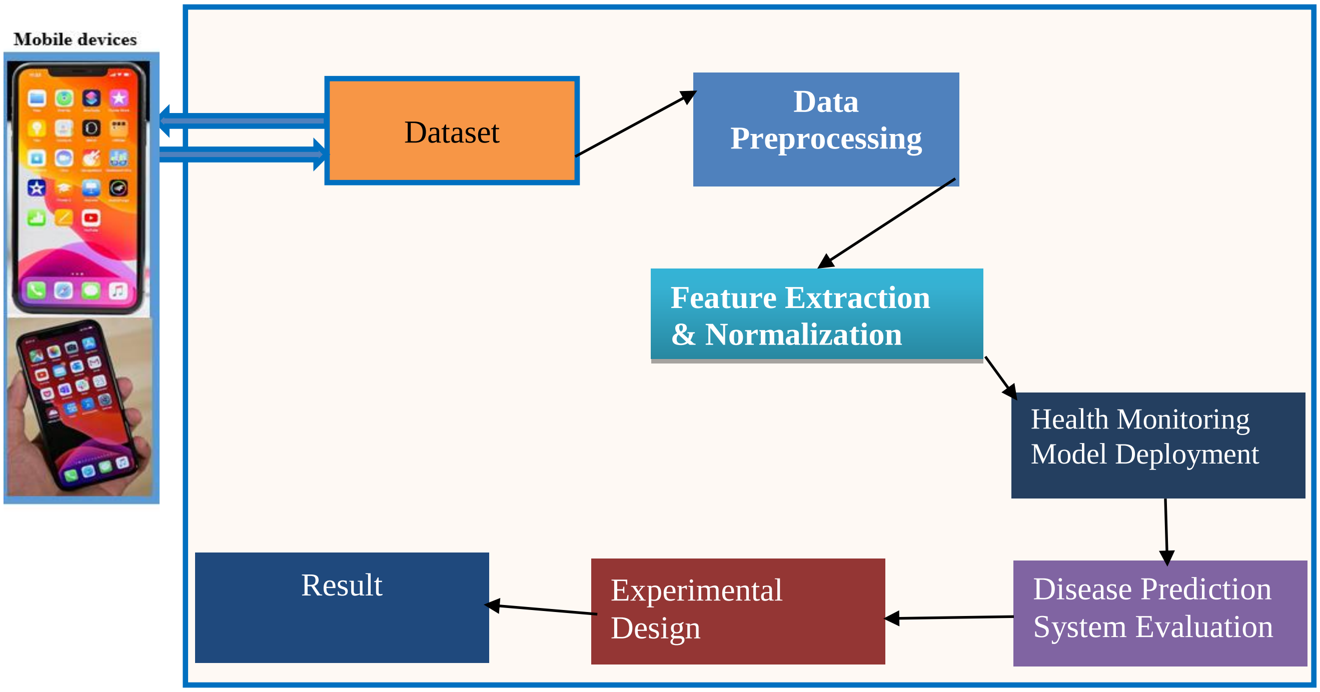

Figure 1. Deployment Method

Implementation approach

The mobile-based system for disease prediction and health monitoring. The implementation of a mobile-based health monitoring system is specifically used for the obtain vital information from patients. This system focused primarily on gathering the patient's pertinent data, vital signs, medical historical records, demographic data, etc. As a result, it mostly addresses illnesses that impact the general pollution and older adult. Heart disease, or cardiac disease, is one of the most common conditions that is challenging to identify because of its links to other fatal conditions like high blood pressure, diabetes, and stroke. Heart illness affects the heart and can be either asymptomatic (having no symptoms) or symptomatic (having some symptoms). It causes significant unexplained mortality, particularly in older men and women, and it may have been irreversible before medical professionals discovered it. As a result, cardiac illness can go undiagnosed and untreated, raising the fatality rate. This paper deployed a machine learning approach to predict the early occurrence of heart disease among the population using data obtained from mobile apps and patient information.

In this section, we focused on the implementation of a disease prediction system using data collected with the developed mobile-based application. The section is divided into various subsections. These include data description, preprocessing, feature analysis and normalization, prediction model development, evaluation of the developed model using commonly used performance metrics, comparison with existing methods, experimental setting and presentation of the results of the system.

Dataset Description

The dataset obtained from the UCI (University of California, Irvine) machine learning repository and the activity data obtained from employing sensors implanted in mobile phones were combined for the suggested disease prediction system using a machine learning model. The magnetometer, gyroscope, and accelerometer are some of the sensors. These sensors collect data that represent the movement and orientation of users and ensure step counts to maintain a healthy heart and reduce cardiovascular disease. The data were collected from both male and female patients between the ages of 40 and 75 years. These are the age ranges with the highest number of cardiovascular disease patients. The data contains 1025 instances with 18 attributes/features and two target classes. The target classes show the presence or absence of heart disease in the patients. A sample of the data is shown in Figure 2.

age	sex	cp	trestbps	chol	BP	fbs	restecg	avacc	avmag	avgyro	thalach	exang	avg_glucor	oldpeak	slope	ca	thal	target
52	1	0	125	212	130	0	1	2.7213	0.75605	1.0873	168	0	228.69	1	2	2	3	Abscence
53	1	0	140	203	115	1	0	2.6965	1.3724	1.3459	155	1	202.21	3.1	0	0	3	Abscence
70	1	0	145	174	124	0	1	2.9957	0.72863	1.1618	125	1	105.92	2.6	0	0	3	Abscence
61	1	0	148	203	128	0	1	2.4695	0.99001	0.97457	161	0	171.23	0	2	1	3	Abscence
62	0	0	138	294	120	1	1	3.0585	0.76068	1.172	106	0	174.12	1.9	1	3	2	Abscence
58	0	0	100	248	120	0	0	2.7853	0.82254	1.0066	122	0	186.21	1	1	0	2	Presence
58	1	0	114	318	130	0	2	2.7904	0.88707	1.0031	140	0	70.09	4.4	0	3	1	Abscence
55	1	0	160	289	110	0	0	2.6861	0.7177	1.0701	145	1	94.39	0.8	1	1	3	Abscence
46	1	0	120	249	140	0	0	2.5967	0.609	1.06	144	0	76.15	0.8	2	0	3	Abscence
54	1	0	122	286	150	0	0	2.7889	0.76963	1.0406	116	1	58.57	3.2	1	2	2	Abscence
71	0	0	112	149	135	0	1	2.3248	0.73213	1.078	125	0	80.43	1.6	1	0	2	Presence
43	0	0	132	341	142	1	0	3.2827	0.6958	1.2871	136	1	120.46	3	1	0	3	Abscence
34	0	1	118	210	140	0	1	1.9152	1.0065	0.93132	192	0	104.51	0.7	2	0	2	Presence
51	1	0	140	298	134	0	1	1.0686	1.0122	0.54971	122	1	219.84	4.2	1	3	3	Abscence
52	1	0	128	204	128	1	1	2.3793	0.69669	0.91167	156	1	214.09	1	1	0	0	Abscence
34	0	1	118	210	112	0	1	2.6892	0.79238	1.0371	192	0	167.41	0.7	2	0	2	Presence
51	0	2	140	308	140	0	0	2.6684	1.0939	1.0908	142	0	191.61	1.5	2	1	2	Presence
54	1	0	124	266	140	0	0	1.6301	0.71356	0.90188	109	1	221.29	2.2	1	1	3	Abscence
50	0	1	120	244	110	0	1	2.5581	0.65082	1.1745	162	0	89.22	1.1	2	0	2	Presence
58	1	2	140	211	140	1	0	2.8326	0.93125	1.229	165	0	217.08	0	2	0	2	Presence
60	1	2	140	185	120	0	0	2.5671	0.74668	1.0584	155	0	193.94	3	1	0	2	Abscence
67	0	0	106	223	130	0	1	2.7183	1.0637	1.1871	142	0	233.29	0.3	2	2	2	Presence
45	1	0	104	208	115	0	0	2.4246	1.4741	1.0385	148	1	228.7	3	1	0	2	Presence
63	0	2	135	252	112	0	0	2.59	0.71539	1.2438	172	0	208.3	0	2	0	2	Presence
42	0	2	120	209	132	0	1	2.4032	0.64662	1.1091	173	0	102.87	0	1	0	2	Presence
61	0	0	145	307	130	0	0	2.8565	0.98698	1.0476	146	1	104.12	1	1	0	3	Abscence
44	1	2	130	233	138	0	1	2.4639	0.93156	0.98836	179	1	100.98	0.4	2	0	2	Presence
58	0	1	136	319	120	1	0	2.5918	0.74332	1.1299	152	0	189.84	0	2	2	2	Abscence
56	1	2	130	256	112	1	0	3.068	0.722	1.1512	142	1	195.23	0.6	1	1	1	Abscence

Figure 2. Sample Dataset with Features Used for the Heart Disease Prediction

While the absent sample is regarded as a negative class, the sample that is present is seen as a positive class. As seen in Figure 3, the dataset contains 526 examples of positive classes and 499 samples of negative classes. In this instance, 48.68% of the patients do not have heart-related disease, but 51.32% of the patients do.

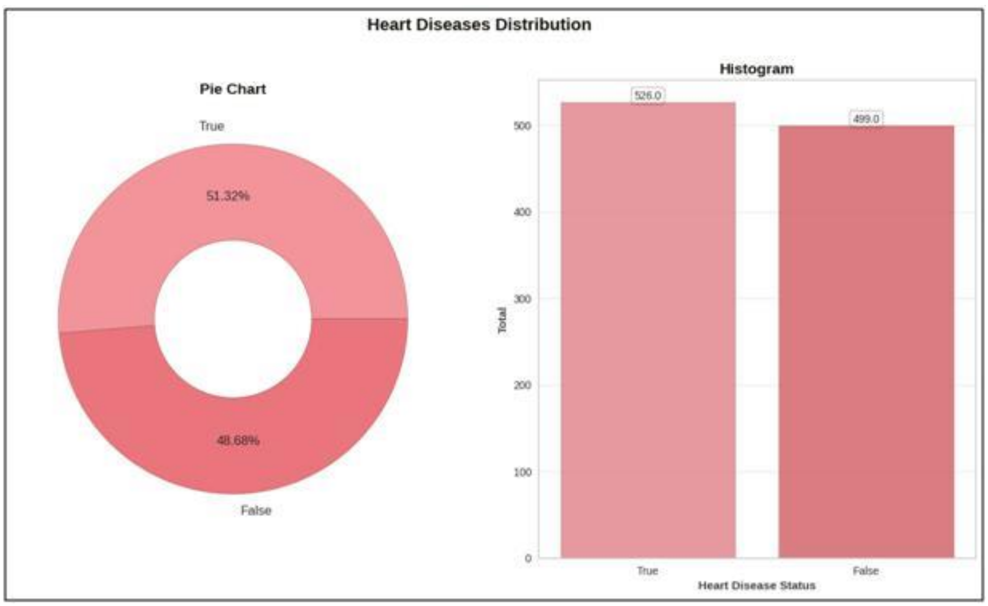

Figure 3. Distribution of Patients with Heart Disease and those without heart disease

These include 713 females where 300 patients have heart disease and 312 males with 226 patients having heart disease as shown in Figure 4.

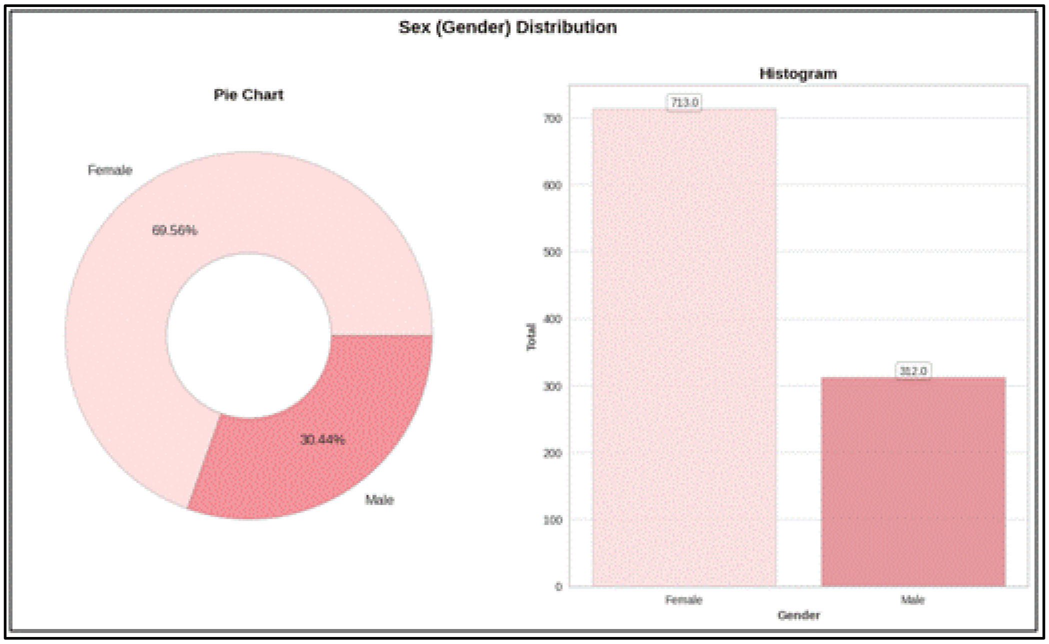

Figure 4. Gender Distribution among Patients with Heart Disease

Compared to their male counterparts, female patients are more likely to acquire cardiac disease (Figure 4). All of the gathered data was stored for analysis and preparation.

Data Preprocessing

Most of the time, noise, abnormalities, missing values, and data duplication influence the patient information that is gathered. As a result, data preparation is a crucial technique for eliminating anomalies and duplicate data from patient data. Additionally, one crucial technique to guarantee better illness prediction system performance is to enter missing values using the average of each column. The data's mean was used to fill in the gaps left by the missing values. Duplicate values in the data were eliminated, and noise and errors were located and eliminated. For feature analysis, the pre-processed data was saved in a comma-separated values (CSV) file format.

Feature Analysis and Normalization

The method of finding significant and distinctive characteristics that might forecast the course of heart disease in senior hospital patients is known as feature analysis. Various patient data were gathered and examined for discriminative characteristics to guarantee a thorough examination of the elements that lead to heart disease. Sex, age, chest pain, cholesterol, resting blood pressure, fasting blood sugar, resting electrocardiography, and physical activity levels as determined by accelerometer, gyroscope, and magnetometer measurements in tri-axial data are some of the traits that are crucial for identifying patients with heart disease. Other factors include blood artery blockage, the maximal heart rate, exercise-induced angina, exercise-induced depression, the slope of the peak by exercise segment, and the diagnostic angiographic disease status of heart disease. The dataset for predicting heart disease was created by combining these characteristics with the demographic data of the patients. Table 1 highlights the specifics of a few of the features.

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

Attributes	Description	Types	Values
Age	Age of the patient	integer	[40-77]
sex	The 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]

Avacc	Average accelerometer value	Float	[0.0-3.0]
Avgyro	Average gyroscope value	Float	[0.0-2.0]
Avmag	Average magnetometer	Float	[0.0-250.0]
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]
Slope	The 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	Coronary heart 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 heart disease prediction system. Here, the heatmap as shown in Figure 5 was used to visualize the relationship between the data values in the dataset. In the diagram, there is an indication that heart disease symptoms such as high blood pressure, cholesterol and depression have a moderate relationship with the age of patients. Other important factors that might be indicative of heart disease in older adults include fasting blood sugar level, maximum heart rate, chest pain, and blood vessel status.

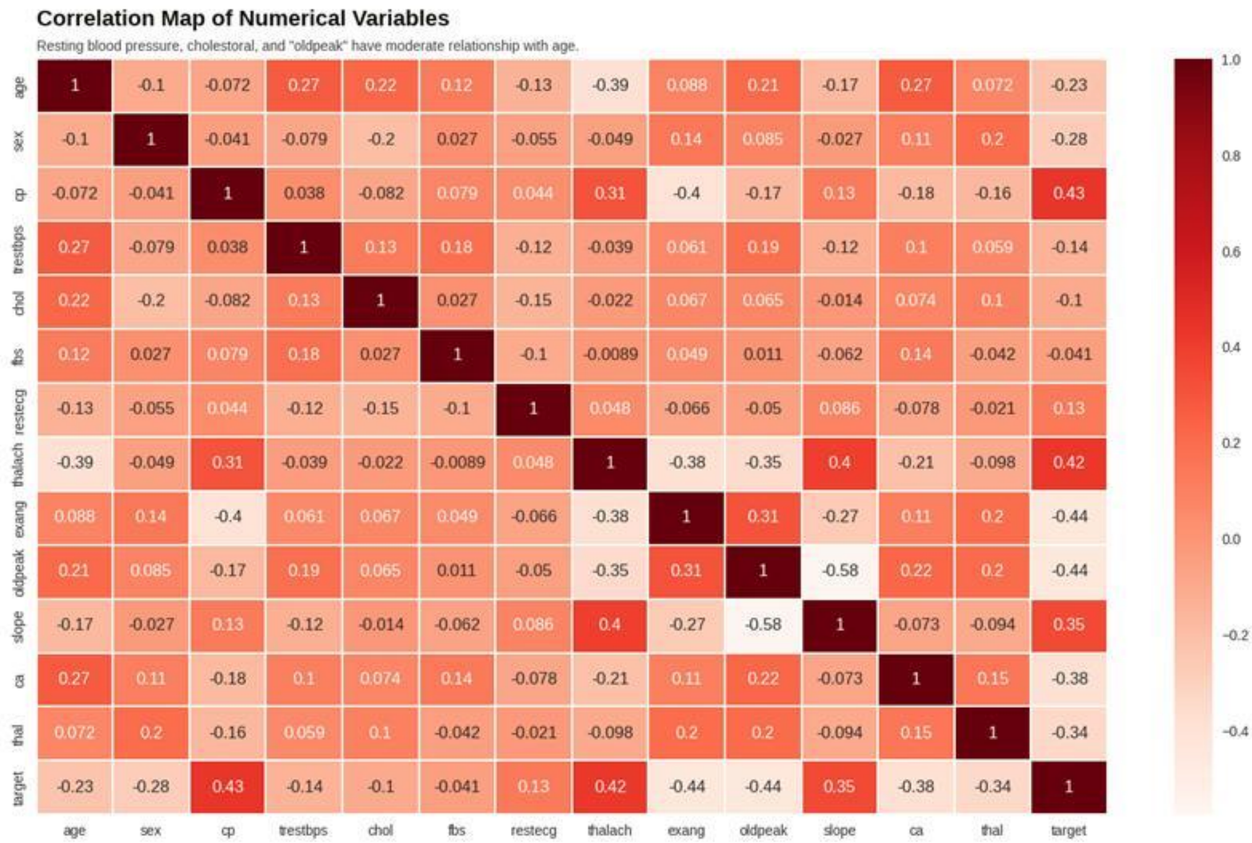

Figure 5: Relationship between each Feature Values and Target Classes

After being identified, the features were reduced to a certain range by standardizing them to have a zero mean and unit variance. Feature normalization is essential to machine learning since it improves the prediction system's efficacy. The Z-score normalization was calculated by subtracting the mean values of the feature vectors from the individual feature value points and dividing the result by the standard deviation. Equation 1 below shows the formula for computing Z-score normalization where $\acute{x}$ represent the compute-score, $\bar{x}$ the mean value, x is the individual features and α is the standard deviation, $\acute{x} = \frac{\bar{x}-x}{\alpha}$ (1)

The normalized features were combined into a master feature vector and saved as .csv for heart disease prediction system implementation.

Disease Prediction Model Deployment

To develop the proposed disease prediction system, k-Nearest Neighbours machine learning algorithms were utilized. A non-parametric and lazy learning algorithm called K-Nearest Neighbors (KNN) uses instance learning techniques to store instances and classify incoming training data using similarity index measures like Euclidean distances. It is an efficient machine learning algorithm that is effective for disease prediction due to its competitive performance and simplicity. k-NN has been widely utilized in pattern recognition and image classification (Ali *et al.*, 2021). The algorithm is effective for handling large datasets and provides good classification in disease prediction as shown in recent studies (Singh *et al.*, 2021; Ramalingam *et al.*, 2018). This paper implements a k-NN-based machine-learning algorithm to detect cardiovascular heart disease.

Disease Prediction System Evaluation

Different performance criteria were used to assess the effectiveness of the suggested prediction system in identifying heart disease using the gathered data. Accuracy, Area Under the Curve, Training Curve, Precision, and Confusion Matrix are some of these performance measurements. The percentage of heart disease that was accurately predicted from the overall number of both the presence and absence of heart disease in the patient group is known as accuracy. The performance of the disease prediction system is measured by the area under the curve (AUC), which shows how well the system can predict cases of heart disease compared to cases without heart 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. The confusion matrix then combines a collection of test data with known true values with one or more assessment or classification models. Furthermore, these performance metrics are mathematically represented in equation (2) to (6)

$$\text{Accuracy (ACC)} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$\text{Sensitivity (SEN)} = \frac{TP}{(TP+FN)} \quad (3)$$

$$\text{Specificity (SPC)} = \frac{TN}{(TN+FP)} \quad (4)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (5)$$

$$\text{KAPPA (KAP)} = \frac{ACC}{(1-ACC_r)} \quad (6)$$

where ACC is the expected accuracy.

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

Figure 7 representing the AUC-ROC plots the true positive rate (TPR) against the false positive rate (FPR) achieved by the machine learning models. The true positive rate measures the actual correctly patients who have heart disease within the data sample, while the false positive rate computes the number of patients with heart disease incorrectly predicted. The performance values of the AUC-ROC curve are in the range of 0 and 1. As shown in Figure 6, k-Nearest Neighbors-based heart disease prediction outperformed other machine learning models in the implementation. The result obtained is promising for the prediction of heart disease in older adults using the collected sample data. The k-Nearest Neighbors (k-NN) based heart disease prediction achieved an AUC-ROC of 0.99, and this shows the applicability of the implemented model for heart disease prediction.

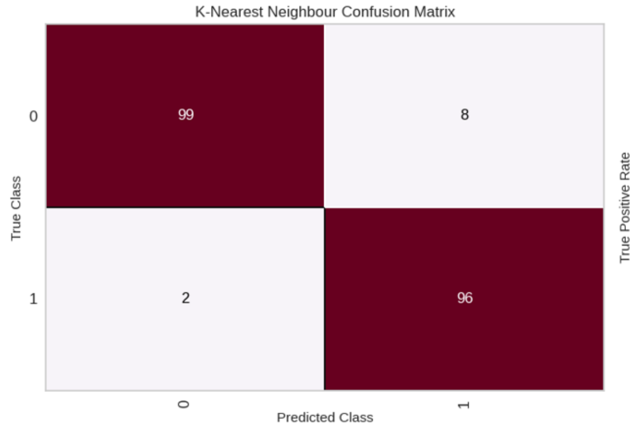

Figure 6(a): Confusion matrix of k-NN

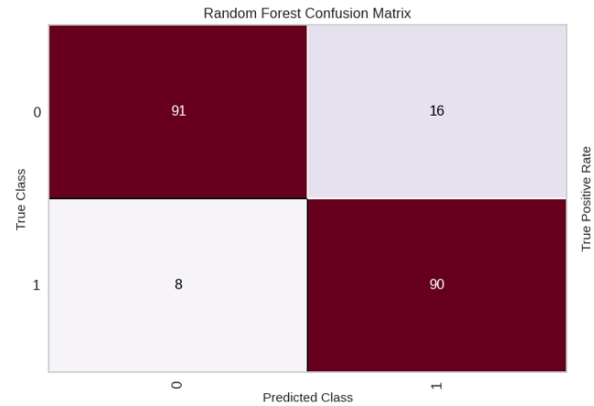

Figure 6(a): Confusion matrix of RF

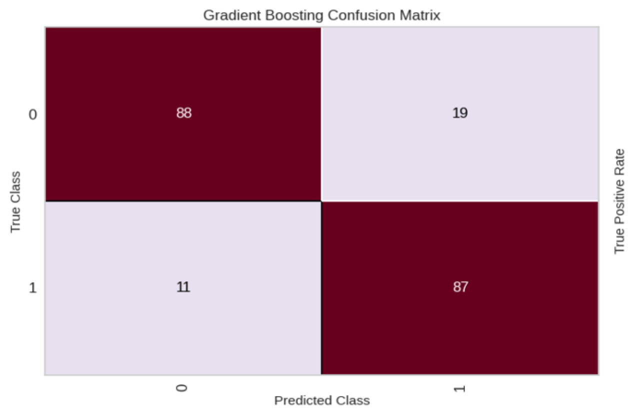

Figure 6(c): Confusion matrix of GR

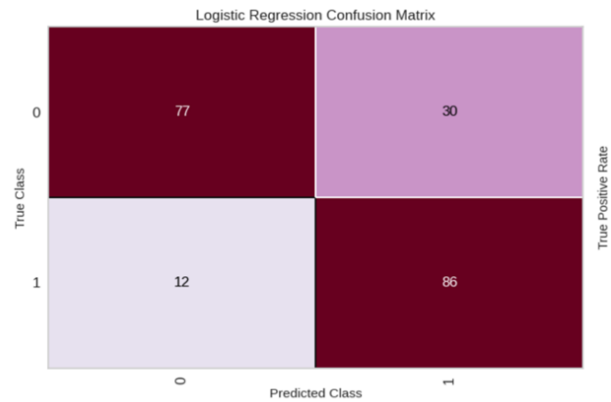

Figure 6(d): Confusion matrix of LR

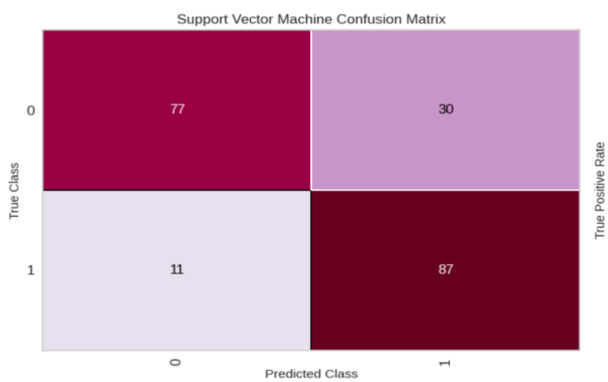

Figure 6(e): Confusion matrix of SVM

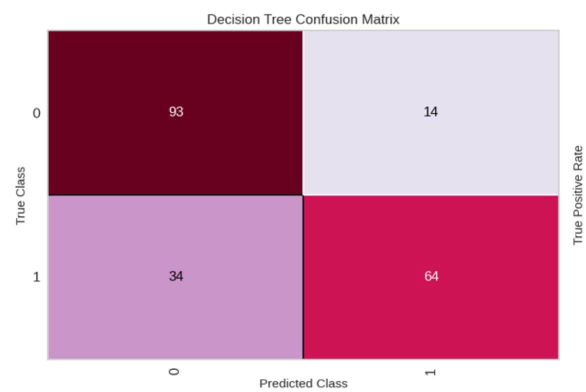

Figure 6(f): Confusion matrix of DT

Figure 6: Confusion matrix of the machine learning models

The confusion matrices of the proposed cardiovascular heart disease prediction system and the compared machine learning models are shown in Figures 6(a) to 6(d). The confusion matrix helps to visualize and represent the prediction outcome of the proposed system. It represents the true positive (actual number of heart disease cases) and false positive (actual number of heart disease cases incorrectly predicted) using the implemented machine learning. In Figure 36, "*0*" represents the absence of heart disease, while "*1*" represents the presence of heart disease. From the figure, it can be seen that k-NN correctly identified ninety-six (99) cases of patients with heart disease and ninety-nine (99) cases of patients without heart disease in the utilized data sample. Similarly, the implemented model only misdiagnosed two (2) and eight (8) cases, respectively. The results obtained by the proposed heart disease prediction system are of great improvement when compared with other machine learning models such as support vector machine, logistic regression, and decision tree. In general, the results obtained with the implementation of the proposed heart disease prediction system suggest that the use of a heart disease prediction system is effective for predicting heart disease using health and demographic information, and activity details obtained using mobile phones.

Experimental Setting

Each phase, including preprocessing, feature analysis, and algorithm creation, was carried out on a machine running the Windows 10 operating system. The disease prediction system was implemented in Python. With a Random-Access Memory (RAM) installed capacity of 6GB, the machine uses an Intel Core 2 Duo processor running at 3.400 GHz. Parts for training and parts for testing were taken from the entire pre-processed data set. Here, 70% of the data was used for training the disease prediction system, while 30% was utilized to test the developed system. Using the train-test data partitioning approach ensures uniformity and reduces complexity. Important hyperparameter settings for both k-NN and other machine learning algorithms are outlined in Table 2. These parameter settings are the defaults and were selected following the results of an empirical evaluation of machine-learning algorithms for the prediction of cardiac disease (Ayon *et al.*, 2020). Additionally, applying some of the machine learning models' default settings would guarantee the algorithms' reproducibility in similar situations.

Table 2: Parameter Values for each Machine Learning Model

Machine learning algorithms	Parameter Tuning
k-Nearest Neighbours (K-NN)	k=5;weight=uniform;algorithm=kd_tree;metric=minkowski;leaf_size=30;
Support vector machine (SVM)	Kernel=rbf;C=1;gamma=scale;degree=3;max_iter=-1;random_state=None;verbose=false
Random forest (RF)	n_estimator=100;criterion=gini;max_depth=None;min_sample_split=2;random_state=None;verbose=0
Naïve Bayes (NB)	priors=None;var_smoothing=1e-9
Decision Tree (DT)	max_depth=None;min_sample_split=2;max_feature=None;random_state=None
Logistic Regression (LR)	Penalty=L1;dual=True;fit_intercept=True;random_state=None;max_iter=100;verbose=0
Gradient Boosting (GB)	max_depth=8;sub_sample=0.8;mas_feature=sqrt;learning_rate=0.05;random_state=None;verbose=0

RESULTS AND DISCUSSION

The outcomes of the suggested machine learning model-based heart disease prediction system are shown in this portion of the paper. Here, data from the UCI machine learning repository and data from the built mobile-based health monitoring and disease prediction system were used to build k-Nearest Neighbors (k-NN). In addition, the mobile-based health monitoring system collected the demographic and health status of people. This information was integrated with the UCI machine learning dataset to implement the proposed heart disease prediction system. The performance results obtained using the proposed k-NN algorithms and the compared machine learning model are shown in Table 3. The table presents the accuracy obtained using the proposed k-NN-based heart disease prediction system and compares it with other machine learning prediction models. From the table, the proposed k-NN-based heart disease prediction system outperformed other machine learning models and achieved an accuracy of 99.21%. This was slightly followed by random forest and gradient boosting classifiers that obtained an accuracy of 94.05% and 93.44%, respectively. There is a high improvement in the use of a k-NN-based heart disease prediction system by 5% to 15% for the heart disease prediction system, as shown in Table 3.

Table 3: Performance accuracy of the heart disease prediction system

Methods	Accuracy
Random Forest (RF)	94.05%
Decision Tree (DT)	84.63%
Support vector machine (SVM)	87.32%
Naïve Bayes	86.40%
Logistic Regression (LR)	87.42%

Gradient Boosting Classifier (GBC)	93.44%
Proposed method(k-NN)	**99.21%**

Figure 7 compares the heart disease prediction system's area under the curve (AUC) and receiver operating characteristics (ROC) with other machine learning models. The AUC and ROC curve are performance indicator that shows the overall effectiveness of the predicted system using k-Nearest Neighbors and other machine learning methods.

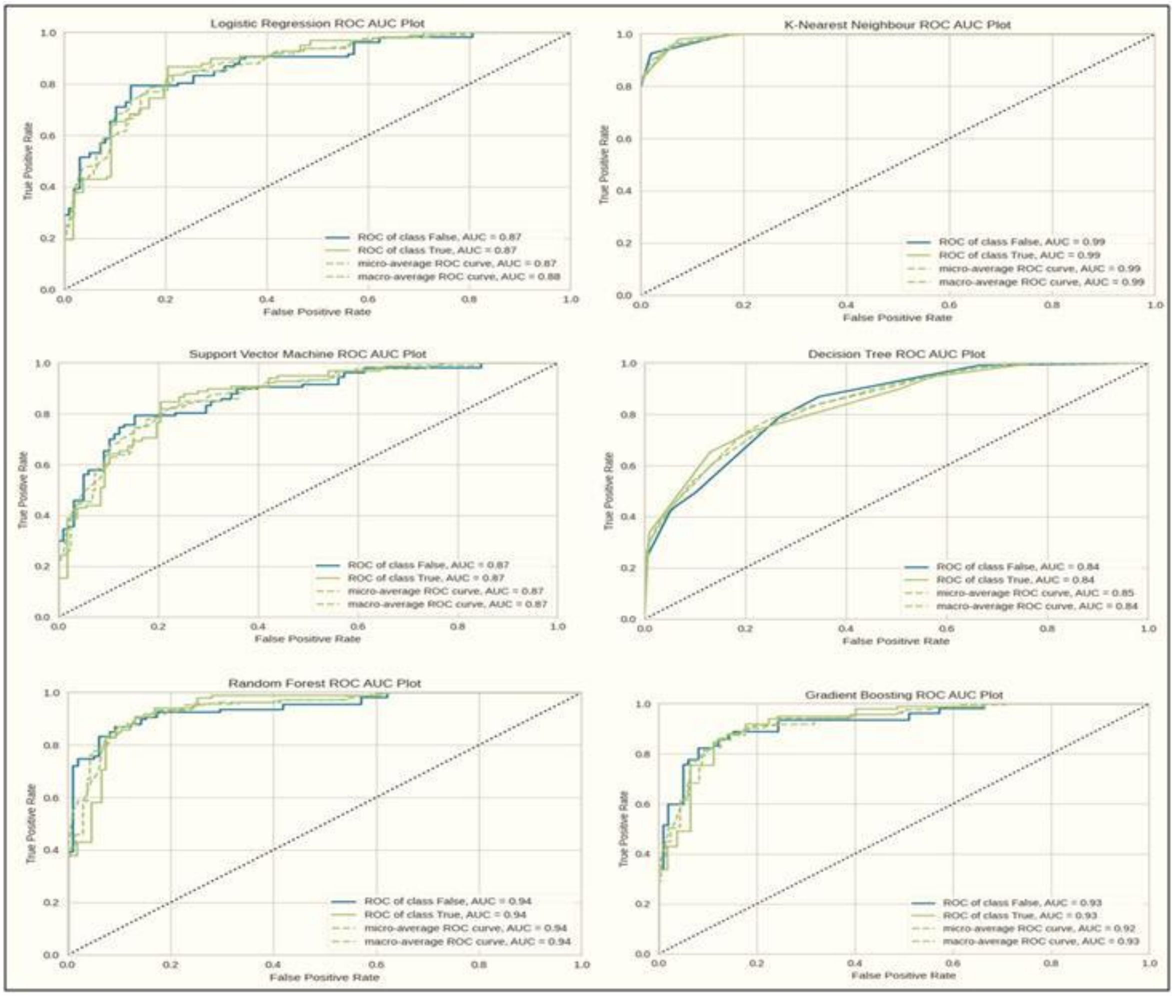

Figure 7: AUC-ROC Performance of the Machine Learning Models

Comparison with other methods

To assess the significance of the proposed heart disease prediction system using k-Nearest Neighbors, the proposed system was compared with five (5) machine learning algorithms that have played vital roles in heart disease prediction (Ayon *et al.*, 2020; Moshawrab *et al.*, 2022). The machine learning algorithms include logistic regression, Naïve Bayes, support vector machine, Random Forest, and gradient boosting. Using the same features used to implement the proposed k-NN-based disease prediction, we ran various experiments and obtained results for comparison. The results are presented in the analysis of the results in section 3, Table 3 of the paper.

CONCLUSION

Conclusively, Deploying Machine Learning for Mobile Sensor-based Health Monitoring and Disease Prediction System checksisessential in human health, living, and well-being. The use of smartphone health monitoring apps enabled with sensors that are driven by machine learning algorithms has made it possible for effective disease diagnosis, treatment, and monitoring. As our world rapidly changes, we cannot afford to leave the health of our loved ones, especially our family members and elderly, behind. This paper presenteda mobile-based health monitoring applicationto check the health status of people living in urban and rural areas. The primary reason is to seamlessly collect information about the health of the patients for analysis, counselling, recommendation, and quick decision-making by healthcare professionals who may be near or far away. The mobile-based application has the capability and components to collect data, transmit vital signs, integrate the data, pre-process, extract the feature, and ensure the data is analysed effectively. The smartphone-enabled sensors that were utilized for health monitoring include electrocardiography (ECG), pulse oximeter, gyroscope, temperature, GPS, accelerometer sensor, etc. Some underlying disease or illness was selected, such as diabetes, high blood pressure, etc., to check the patient for such cardiovascular diseases. Nevertheless, the problem of the high cost of medical support, inefficient tools, lack of health awareness, and the burden of providing optimum healthcare to elderly people will be solved. Future studies will focus on the implementation of mobile applications for disease diagnosis and predictions, and providing security for data generated using mobile health applications.

REFERENCE

Ali, M. H., Khan, D. M., Jamal, K., Ahmad, Z., Manzoor, S., & Khan, Z. (2021). Prediction ofMultidrug-Resistant Tuberculosis Using Machine Learning Algorithms in SWAT, Pakistan. *Journal of healthcare engineering*, *2021*(1), 2567080.https://doi.org/10.1155/2021/2567080

Anikwe, C. V., Nweke, H. F., Ikegwu, A. C., Egwuonwu, C. A., Onu, F. U., Alo, U. R., & Teh, Y. W. (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

Ayón, C., Ramos Santiago, J., & López Torres, A. S. (2020). Latinx undocumented older adults, health needs and access to healthcare. *Journal of Immigrant and Minority Health*, *22*, 996-1009. https://doi.org/10.1007/s10903-019-00966-7

De Fazio, R., Mastronardi, V. M., De Vittorio, M., & Visconti, P. (2023). Wearable sensors and smart devices to monitor rehabilitation parameters and sports performance: an overview. *Sensors*, *23*(4), 1856.https://doi.org/10.3390/s23041856.

Deshpande, U. U., & Kulkarni, M. A. (2017). IoT-based real-time ECG monitoring system using Cypress WICED. *International Journal of advanced research in electrical, electronics and instrumentation engineering*, *6*(2).https://doi.org/10.15662/IJAREEIE.2017.0602035.

El-deep, S. E., Abohany, A. A., Sallam, K. M., & El-Mageed, A. A. A. (2025). A comprehensive survey on impact of applying various technologies on the Internet of Medical Things. *Artificial Intelligence Review*, *58*(3), 86.https://doi.org/10.1007/s10462-024-11063-z.

Enshaeifar, S., Zoha, A., Markides, A., Skillman, S., Acton, S. T., Elsaleh, T., ... & Barnaghi, P. (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), e0195605.https://doi.org/10.1371/journal.pone.0195605

Fisk, A. D., Czaja, S. J., Rogers, W. A., Charness, N., &Sharit, J. (2020). *Designing for older adults: Principles and creative human factors approaches*. CRC press. https://doi.org/10.1201/b22189.

Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science*, *1*(5), 290.https://doi.org/10.1007/s42979-020-00305-w.

Islam, M. N., Raiyan, K. R., Mitra, S., Mannan, M. R., Tasnim, T., Putul, A. O., & Mandol, A. B. (2023). Predictis: an IoT and machine learning-based system to predict risk level of cardio-vascular diseases. *BMC Health Services Research*, *23*(1), 171.https://doi.org/10.1186/s12913-023-09104-4.

Marino, M. M., Rienzo, M., Serra, N., Marino, N., Ricciotti, R., Mazzariello, L., ... & Caracciolo, A. L. (2020). Mobile screening units for the early detection of breast cancer and cardiovascular disease: a pilot telemedicine study in Southern Italy. *Telemedicine and e-Health*, *26*(3), 286-293.https://doi.org/10.1089/tmj.2018.0328.

Menon, S. P., Shukla, P. K., Sethi, P., Alasiry, A., Marzougui, M., Alouane, M. T. H., & Khan, A. A. (2023). An intelligent diabetic patient tracking system based on machine learning for E-health applications. *Sensors*, *23*(6), 3004.https://doi.org/10.3390/s23063004.

Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., & Raad, A. (2022). Cardiovascular events prediction using artificial intelligence models and heart rate variability. *Procedia Computer Science*, *203*, 231-238.https://doi.org/10.1016/j.procs.2022.07.030.

Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A.

W. (2022). [Retracted] Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Computational intelligence and neuroscience*, *2022*(1), 2973324.https://doi.org/10.1155/2022/2973324.

Olopade, I. A., Akinwumi, T. O., Philemon, M. E., Mohammed, I. T., Sangoniyi, S. O., Adeniran, G. A., ... & Adesanya, A. O. (2024). Analyzing Global Stability of M-Pox Disease Dynamics: Mathematical Insights into Detection and Treatment Strategies. *Journal of Basics and Applied Sciences Research*, *2*(1), 61-76. https://doi.org/10.33003/jobasr-2024-v2i1-25

Ozdener, M. H., Donadoni, M., Cicalese, S., Spielman, A. I., Garcia-Blanco, A., Gordon, J., &Sariyer, I. K. (2020). Zika virus infection in chemosensory cells. *Journal of neurovirology*, *26*, 371-381. https://doi.org/10.1007/s13365-020-00835-2

Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology*, *7*(2.8), 684-687.https://doi.org/10.14419/ijet.v7i2.8.10557.

Sada, I., Obunadike, G. N., & Abubakar, M. (2025). Machine learning-based framework for predicting user satisfaction in e-learning systems. *Journal of Basics and Applied Sciences Research*, *3*(2), 78-85. https://doi.org/10.33003/jobasr

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosis of disease using machine learning. In *Machine learning and the internet of medical things in healthcare* (pp. 89-111). Academic Press.https://doi.org/10.1016/B978-0-12-821229-5.00003-3.

Thaung, S. M., Tun, H. M., Win, K. K. K., Than, M. M., & Phyo, A. S. S. (2020). Exploratory data analysis based on remote health care monitoring system by using IoT. *Communications*, *8*(1), 1-8.https://doi.org/10.11648/j.com.20200801.11.

Vesnic-Alujevic, L., Breitegger, M., & Guimarães Pereira, Â. (2018). 'Do-it-yourself'healthcare? Quality of health and healthcare through wearable sensors. *Science and engineering ethics*, *24*, 887-904.https://doi.org/10.1007/s11948-016-9771-4.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265-269.