International Journal of Computing and Digital Systems
2025, VOL. 17, NO. 1, 1-12

http://dx.doi.org/10.12785/ijcds/1571033487

An Efficient IoT-based Prediction and Diagnosis of Cardiovascular Diseases for Healthcare Using Machine Learning Models

Hamza Aldabbas[1] and Zaid Mustafa[1]

[1]*Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt- Jordan.*

Received 27 May 2024, Revised 22 August 2024, Accepted 4 August 2024

Abstract: The Internet of Things (IoT) and Machine Learning (ML) models are emerging technologies that are changing our daily lives. These are also considered game-changing technologies in recent years, catalyzing a paradigm change in traditional healthcare practices. Cardiovascular disease (CVD) is considered a major reason for the high death rate around the world. Cardiovascular disease is caused by several risk factors like an unhealthy diet, sugar, high Blood Pressure (BP), smoking, etc. Preventive treatment and early intervention for at-risk people depend heavily on the prompt and accurate prediction of illnesses. Developing prediction models with improved accuracy is essential, given the increasing use of electronic health records. Recurrent neural network variations of deep learning can handle sequential time-series data. In remote places, they often lack access to a skilled cardiologist. Our proposal aims to develop an efficient community-based recommender system using IoT technology to detect and classify heart diseases. To address this issue, machine learning techniques are applied to a dataset to predict patients with cardiovascular disease because it's difficult for the medical team to identify CVD effectively. A public dataset contains data of 70000 patients gathered during medical examination, and each row has 13 attributes. The risk groups were determined by their likelihood of developing cardiovascular disease as it works successfully in forecasting diseases by utilizing the support system.

Keywords: Internet of things; Machine Learning; Heart disease; Decision tree; Disease detection; Naïve Bayes; Support Vector Machine (SVM);

1. Introduction

Online apps built on the cloud and the Internet of Things are more effective than standard cloud-based apps. It may be used in new fields, including account management, the military, and pharmaceuticals [1]. In particular, the cloud-based IoT system would be useful for supplying pharmaceutical applications with efficient administrations for observing and receiving information from several distant sites. IoT-enabled healthcare applications gather important data, update the severity of therapeutic parameters over a defined period, and track sufficient changes in health limitations, for example, [2]. Many people are dying due to cardiovascular disease (CVD) around the world annually. Cardiovascular diseases cause many deaths annually [3]. An estimated 20 million people died from cardiovascular disease in 2020, which is 35% of all global deaths. 85% of cardiovascular deaths happen due to heart attacks and strokes [4].
In low- and middle-income nations, cardiovascular disease (CVD) accounts for more than 75 percent of fatalities [6]. Heart attacks, strokes, and cardiovascular disease (CVD) are caused by obstructions that stop the heart and brain from receiving the needed blood. Blood arteries are blocked due to fatty build-up, stopping blood flow to the heart and brain. Heart attack and stroke are also caused when blood from blood vessels starts bleeding in the brain. Many other risk factors like no exercise, tobacco use, oily food items, unhealthy diet, obesity, unsafe use of alcohol, high blood pressure, diabetes, etc, also cause cardiovascular disease. People with these risk factors are more likely to have cardiovascular disease or get cardiovascular disease [5], [6]. To save lives, we must address this problem by early detection of cardiovascular disease [7]. With an estimated 17.9 million deaths annually, cardiovascular diseases (CVDs) are the world's leading cause of mortality. A collection of diseases known as cardiovascular diseases (CVDs) includes rheumatic heart disease, coronary heart disease, and cerebrovascular disease. Heart attacks and strokes are the primary cause of over four out of every five CVD fatalities, with one-third of these deaths occurring prematurely in

E-mail address: aldabbas@bau.edu.jo, z_lami@bau.edu. jo

those under the age of 70.
With the backing of several academic disciplines, the Internet of Things (IoT) is a revolutionary platform that improves user experience by generating premium connections across numerous interdependent sensors to detect the CVD [8], [9]. The healthcare system may be greatly enhanced by combining cutting-edge technology like AI, IoT, and cloud computing. Because it lessens stress for patients and physicians, the Internet of Things is essential to the medical field. The system helps patients and doctors monitor, measure, and record important medical data by integrating devices, applications, and networked systems. Smartphone sensors connected to CVD and Wi-Fi dongles are two examples of these gadgets. Apps for smartphones aid with medical record-keeping, emergency support, and timely alerts [10]. Early prediction of cardiovascular disease can save human lives and minimize the treatment cost so people can get better counselling. The ability to predict CVD may significantly affect how it is treated and perhaps save lives [8]. By addressing behavioural risk factors such as cigarette use, poor eating and obesity, inactivity and hazardous alcohol use, hypertension, diabetes and hyperlipidemia, blood pressure, sugar, and other conditions, the majority of cardiovascular illnesses may be detected or avoided [11]. This study offers a method for predicting the presence or absence of cardiovascular illness based on machine learning approaches.
Obtaining data to forecast diseases is still a difficult challenge. Accurate data collection is essential for decision-making, particularly when diagnosing CVD. The e-health system aims to identify CVD early on to lower the risk of illness and death from it [12]. It also seeks to correctly identify illness and offer suitable advice for enhancing the patient's health. It is necessary to provide patients with CVD with personalized and optimal advice to enhance their health in isolated locations, particularly when a cardiologist is unavailable. Currently, CVD detection systems depend on traditional devices to detect the attacks and other symptoms. Using community-based recommender systems to provide relevant and personalized suggestions based on demographic data is still a key area of study [13].
The study's goal is to ascertain how to identify CVD in humans using contemporary instruments and technologies. Every conventional CVD risk-assessment model assumes that every risk factor has a linear relationship to the outcome of CVD. These models frequently oversimplify intricate relationships, such as those involving several risk variables and non-linear interactions. More subtle correlations between the risk factors and outcomes should be found, and more risk factors should be appropriately included. This study aims to identify which ML algorithm output had relatively good brevity and investigate whether ML may improve cardiovascular risk prediction accuracy in general population primary care. Several machine learning (ML) models for CVD detection have been put forth in recent years. We used an IoT-based platform to provide universal, high-quality, and reasonably priced medical services. Our ubiquitous healthcare application produces a significant quantity of clinical data. For further processing and analysis, these data must be maintained appropriately. Integrating IoT with cloud computing offers a viable approach to effectively handle sensor data in healthcare, removing the requirement for technical infrastructure knowledge by abstracting technical aspects. Additionally, it makes it simple and inexpensive to automate the process of gathering and transmitting data. Patients with CVD may get lifestyle, nutritional, and exercise suggestions from the suggested fog-based recommender system. The system gathers many characteristics from the patients and determines the illness. Then, based on many cardiologist suggestions, the system gives the patient the most appropriate advice. We proposed an IoT-based secure system in which different types of human health data are collected and analyzed using popular machine learning algorithms. Our study aims to improve accuracy and system efficiency by using standard medical equipment to forecast the likelihood of a heart attack. A CNN model with the best-performing ML algorithms is chosen to achieve the objectives and acquire an effective high ranking. For this purpose, real-time data is collected, and an IoT system is designed to forecast the likelihood of developing CVD. Finally, doctors and patients reviewed the system's effectiveness, efficiency and satisfaction. The main contributions are as follows:

- The authors describe an innovative Internet of Things-based smart heart disease prediction system in which different types of sensors are used to get patient data.
- Acquiring relevant data from commonplace devices to process data by learning specifics about medical history and other ailments
- To propose an autonomous heart disease diagnostic system utilizing several machine learning models, such as CNN, NB, Random Forest, Bayes net, Decision tree, and NB, with parameter optimization.
- Verifying the effectiveness of the cleverly constructed smart heart disease prediction model. Seventy thousand patient records collected during medical examinations are available as a public dataset with thirteen attributes per row. Based on their propensity to develop cardiovascular disease, the risk groups were established. Because it effectively uses the support system to forecast diseases.

The rest of the paper is organized as follows: Section II represents the related work on this topic and the motivation for the proposed approach. Next, section III elaborates on the proposed model for detecting cardiovascular attacks using different machine learning models. Performance evaluation of the proposed approach is mentioned in Section IV, and the last section concludes the paper.

2. Literature Review

In this section, we discussed the different studies on heart disease detection using various techniques. Hussain et al. [14] evaluated six machine learning techniques: Support Vector Machine, artificial neural network, Logistic regression, Naıve Bayes (NB), k-nearest neighbor, and Classification trees for heart disease prediction. Because the number of heart patients is increasing rapidly, early identification and early treatment can help save many heart patients' lives. Islam et al. [15] proposed an approach to predict CVD. CVD causes a high number of causalities anywhere in the world. Data is collected to predict CVD, which contains 301 samples. Each sample contains 12 attributes. Different machine learning algorithms, i.e., Naive Bayes, Logistic regression, SVM, and Decision tree classification algorithms, have been applied to predict heart disease. The results were also compared with the UCI Heart Disease dataset. Logistic regression, Decision tree, SVM, and Naive Bayes have been applied to the UCI Heart Disease dataset, giving 82.9%, 86.1%, and 75.8%, respectively.

Ahmed et al. [16] presents a hybrid approach that uses classification as a means of prediction analysis. A hybrid approach for predicting heart disease was developed using actual patient data. In this study, KNN and SVM classification algorithms were used. Regarding heart disease prediction analysis, a hybrid method performs better than other algorithms. Many researchers have published and proposed approaches to predict CVD to save the lives of human beings. Most researchers use the UCI Heart Disease dataset, and few use the CVD dataset to predict CVD using different ML algorithms. [17] presents a fuzzy logic inference system that is applied to CVD data and demonstrates effective outcomes in predicting cardiovascular disease, specifically in predicting the severity of CVD. A hierarchical fuzzy inference system (HFIS) is employed. Cardiovascular disease (CVD) accounts for 31% of global mortality. Cardiovascular disease is responsible for a greater number of deaths compared to any other cause. [18] applied the machine learning technique logistic regression on the Heart Disease UCI dataset to predict cardiovascular disease. The Heart Disease UCI dataset consists of fourteen variables: age, cholesterol level, and unhealthy lifestyle. Several risk factors cause cardiovascular disease, and ML techniques are used to predict CVD. Logistic regression shows some effective results. The accuracy reaches 85% with an error rate of 0.1406565.

Guo et al. [19] presented an approach that combines CNN with interpretable machine-learning algorithms. Magnetic resonance imaging (MRI) helps to detect cardiovascular disease. CNN has success in image segmentation and gives effective results, but it requires large datasets and provides suboptimal results that require further processing. They developed a continuous cut segmentation algorithm by combing normalized cuts and continued regularization in a united framework. The result shows that the new approach improved CNN segmentation and reduced the variability of CNN segmentation. Maiga et al. [20] apply ML techniques to predict CVD that is caused by several risk factors like cholesterol level, glucose level, and blood pressure. The dataset contains 70000 patients' records or rows and each row has 13 attributes or risk factors. Different algorithms achieved good results and high accuracy 73%, 70%, 72 %, 71%, respectively. [21] analyze the performance of the logistic regression algorithm used by previous researchers between 2000 to A multivariate logistics regression technique is used to determine the correlations between independent and dependent variables. thirty-seven research publications included logistic regression as a primary model in their study and six book reviews. This research identifies flaws in the use and reporting of LR. Several research did not report goodness-of-fit metrics, regression diagnostics, and validation analysis. This research provides a correct application of logistic regression as well as an illustration of how modeling approaches should be used in order to compute and evaluate the model's coefficients.

[22] used a machine-learning algorithm to predict cardiovascular diseases and did a comparison of two data mining algorithms to find the best algorithm. Cardiovascular disorders were predicted using two data mining techniques, SVM and ANN, and SVM performs better than ANN. SVM is more accurate than ANN because the ROC curve area in SVM is greater. Gupta et al. [23] developed 12 layers CNN model and applies a pixel-wise, patch-based procedure to predict Breast arterial calcifications (BACs) in mammograms. Subsequently, the efficacy is assessed using calcium mass quantification analysis and free-response receiver operating characteristic (FROC) analysis on 840 full-field digital mammograms from 210 people. Calcium mass quantification analysis yields more realistic findings than free-response receiver operating characteristic (FROC) analysis, although both methods provide comparable results. This finding indicates that deep learning is effective in identifying breast cancer patients and BACs in mammograms. In [24], the authors use machine learning approaches to focus on early identification of cardiac illnesses, including myocardial infarction. It addresses the problem of unbalanced datasets by carrying out an exhaustive literature analysis to pinpoint practical solutions. Various machine learning and deep learning classifiers were utilized to improve the accuracy of heart disease predictions. These included K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Convolutional Neural Network, Gradient Boost, XGBoost, and Random Forest. The results of the study highlight how well an XGBoost model can be precisely adjusted to treat cardiovascular disorders. Regulating frameworks, technological difficulties, and concerns about data privacy are only a few of the drawbacks and difficulties that accompany them [25]. According to this assessment, despite these difficulties, there is hope for the future of smart technologies in the management of CVD, as advances in fabrication methods, telemedicine platforms, and AI algorithms create new avenues for effective and personalized therapy. This article delves into the role of smart technologies in CVD management, including their benefits and drawbacks, limitations, present uses, and smart future.

Based on tiny datasets from testing, classifiers and popular machine learning encoding techniques have shown a wide range of unexpected results when applied to diagnose heart illness [26]. The early study employed convolutional

neural networks (CNNs), a kind of deep learning model, to extract features without establishing an understanding of sequence information. In this paper, an efficient deep learning-based system is suggested to predict cardiovascular disease from patient data, namely a CNN with bidirectional long- / short-term memory. Using feature selection, which involved ranking and choosing characteristics that received high ratings in the provided illness dataset, only the most pertinent features were chosen. Subsequently, the CNN + BiLSTM hybrid deep learning approach was utilized to forecast cardiovascular disease.
Previous studies have identified several persistent issues. These include the inability to achieve cost-optimal execution with different models and QoS characteristics, the lack of improvement in performance despite various optimization and feature selection techniques, and the failure to enhance precision by incorporating advanced artificial intelligence. The suggested approach is significantly enhanced by integrating many advanced technologies, including artificial intelligence (AI), Internet of Things (IoT), and cloud computing. The Internet of Things is crucial in the medical profession due to its ability to reduce stress for both patients and clinicians. The system facilitates the monitoring, measurement, and recording of crucial medical data for patients and clinicians by combining various devices, apps, and networked systems. Smartphone sensors, such as those attached to CVD and Wi-Fi dongles, exemplify these devices. Smartphone applications facilitate the management of medical records, provide emergency assistance, and provide timely notifications.

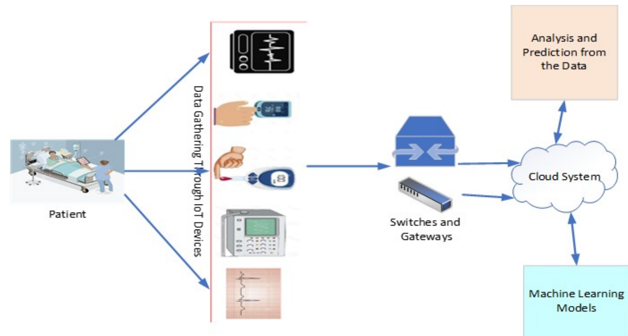

Figure 1. IoT-based Proposed system model for patient diagnosis.

3. Proposed Solution

To address the problem of heart disease, we proposed an IoT-based system in which several components work together to obtain data from different devices and perform analysis to indicate normal and abnormal conditions. The suggested system is shown in Figure 1, where the patient's body is equipped with hardware such as activity, medical, and ambient sensors. Blood pressure, heart rate, EEG, blood oxygen level, EMG, respiration rate, ECG, and other measures are used to gather information about the body. Following data processing by gateway devices, the worker or broker nodes in the cloud system get the processed data for the purpose of predicting cardiac disease. The suggested model uses patient medical data to assess different machine-learning techniques. The healthcare sector quickly embraced the IoTnput data set, which is related to heart disease. Substituting non-available values with column means is the next preprocessing step. Figure 1 illustrates the various tactics that were used. The machine learning models' output represents their accuracy metric. The predictions using the model are then possible. The medical specialists assisted in gathering the dataset from a reputable hospital. The following characteristics are included in the data set: blood pressure, heart rate, age, gender, history of diabetes, family history of coronary artery disease, heartbeat rhythm, ECG readings, chest discomfort, and smoking status. To deal with missing values, duplicate records, and noise, preprocessing techniques are used to the training dataset. The data is then subjected to feature selection and preprocessing using the suggested system.

The IoT was quickly embraced by the healthcare sector as it improves service quality and efficiency by incorporating IoT features into medical devices. For the elderly, those with long-term illnesses, and those in need of steady care, this offers amazing benefits. IoT devices continuously create enormous volumes of health data because IoT-based healthcare applications are used to acquire vital data, such as real-time changes in health parameters and updates of medical parameters' severity within a specified period. Industry interest in the Internet of Things (IoT) is growing, with many considering it to be the most important technology of the future. The patient has a great deal of sensors attached to them. These sensors gather various data, such as blood sugar level, heart rate, ECG, and pulse rate. The patient's connected sensor gadget transmits sensor information continually. These are categorized according to the training results, which implies that the IoT sensor readings are compared to the training phase values. After comparing the values, the system produces categorized results. The Entire system model is reflected in Figure 2

IoT devices continuously create enormous volumes of health data because IoT-based healthcare applications are used to acquire vital data, such as real-time changes in health parameters and updates of medical parameters' severity within a specified period. Industry interest in the Internet of Things (IoT) is growing, with many considering it to be the most important technology of the future. The patient has a great deal of sensors attached to them. These sensors gather various data, such as blood sugar level, heart rate, ECG, and pulse rate. The patient's connected sensor gadget transmits sensor information continually. These are categorized according to the training results, which implies that the IoT sensor readings are compared to the training phase values. After comparing the values, the system produces categorized results. The system employs Machine Learning methodologies to achieve the goals of categorization and anomaly detection. The analytics component of the system

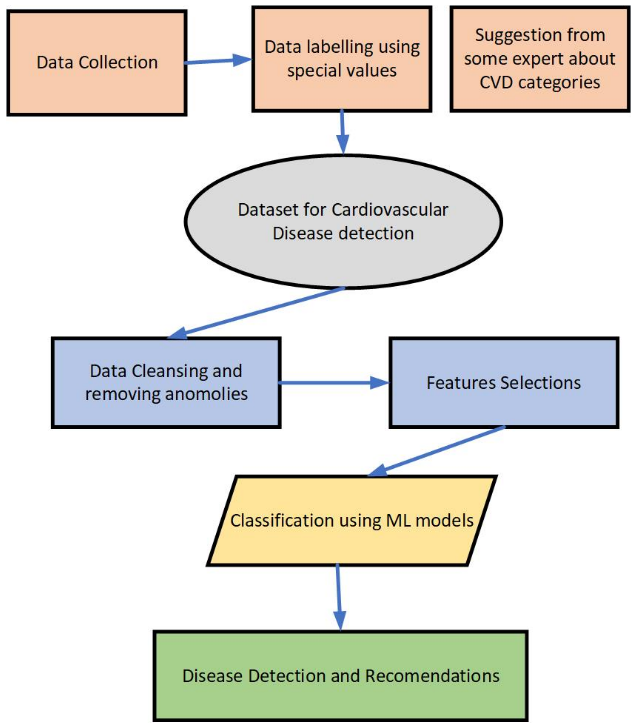

Figure 2. Entire system model to detect CVD.

enables the supply of actionable insights, diagnoses, and the development of disease treatments. The technology also enables the deployment of robust authentication and data security protocols to ensure the secure flow of data across many platforms. Security protocols like TLS and SSL are used to encrypt the data during transmission while the data is safely kept in databases. The security standards used in these IoT devices guarantee the effectiveness of the whole system.

A. Data Acquisition (Data Set)

In any machine-learning-based problem, the first step is getting the dataset [27] processed in the next stages. In this research, we selected the cardiovascular disease (CVD) dataset because it's new in the market, so not much work has been done on it, and the dataset is in good condition. This dataset contains 70000 CVD patients' data, and every data flow in the dataset consists of 11 flow features. All dataset values were collected at the moment of medical examination and are publicly available on the following link for research purposes. https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data The main benefit of this dataset is that it offers a foundation for more reliable findings and is widely accepted for early detection of cardiovascular illnesses, which may save lives and lower treatment expenses. The majority of researchers use this dataset to study CVD. We begin using this strategy with the one characteristic that performs the best. The next best-performing feature is then added, and the correctness is verified. Up until the accuracy of the findings starts to rise, the method keeps adding features. When the rate of recognition begins to decline, it will end. The dataset's twelve selected variables are age, height, weight, gender, systolic and diastolic blood pressure, glucose, cholesterol, alcohol use, physical activity, and the existence or absence of cardiovascular diseases.

B. Data Preprocessing

After acquiring the dataset, preprocessing is a major step in supervised machine learning. The CVD dataset contains Numeric values. We use the Weka tool to convert Numeric values into Nominal Values using the Numeric Nominal converter. After converting the attribute from Numeric to nominal, several ML techniques were applied to the CVD dataset. Aiming to address issues and problems with raw data, including unbalanced, high dimensionality, missing, noisy, and inconsistent data, is data preparation. Data cleaning, data reduction, data transformation, and data integration are the four categories into which the preprocessing activities may be divided. Data balancing has also been thought of as an extra preprocessing step. The issue of unequal class distribution is its primary focus. Dealing with missing data, finding and removing outliers, identifying and removing noise, and fixing inconsistencies are all considered aspects of data cleaning. The term "data reduction" describes procedures intended to minimize the quantity of data inside a dataset. Data reduction can be accomplished by converting continuous characteristics into a small collection of nominal values, selecting features, or extracting features.

It is possible for there to be imbalances in the classes during classification, which would cause the model to favour the majority class. We employed methods like applying algorithms made for unbalanced data oversampling the minority class, or undersampling the majority class. When missing values exist in the dataset, it can negatively impact how well our machine-learning models perform. We used sophisticated imputation techniques, removed rows or columns containing missing values, or used imputation (filling missing values using statistical approaches) to solve this difficulty. to handle outliers that have a large effect on the mean and standard deviation and might cause skewed results and incorrect forecasts. Using statistical techniques (such as the Z-score or IQR) to identify these outliers, it is decided whether to change or eliminate them.

C. Operation of Naive Bayes (NB) Algorithm

The Naive Bayes approach of probabilistic machine learning, used for a variety of classification issues, is built on the Bayes Theorem. In this article, we discuss the Naive Bayes approach in detail to eliminate any possibility of misunderstanding. Numerous linear parameters are required because of the multiscale structure of the Nave Bayes classifier and the substantial amount of variables (characteristics/predictors) in the learning issue. It is feasible to train most probabilistically by analyzing closed phrases. Linear time is considered rather than the expensive iterative

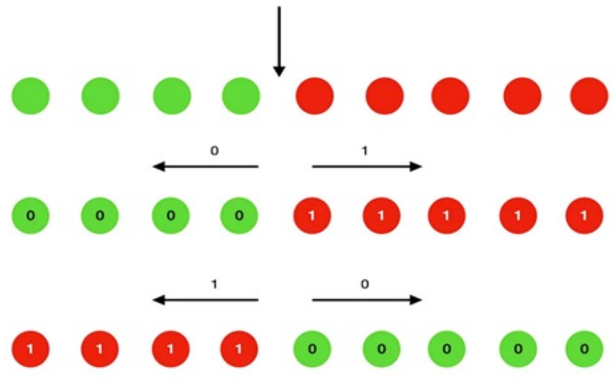

Figure 3. Operation of decision stamp.

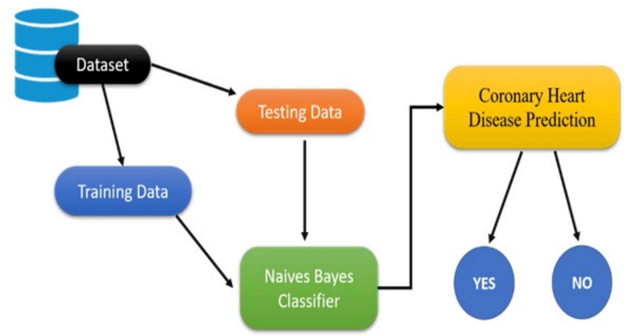

Figure 4. Naïve Bayes Classifier for Heart Disease Prediction.

method used in many other classifiers. It is simple to build classes using naive Bayes. Here is an example of a class structure for problem states that are stated as values. From the terminal, the class name is extracted. Although there isn't a single technique for training these classifiers, there are families of algorithms built on fundamental ideas. Every member of the NB team thinks that the value of one thing influences the value of another.

D. Decision Stump

A one-step decision tree is considered a decision stump model. A decision tree in which the inner(root) node is instantly linked to the terminal node. The decision stump produces predictions depending on the value of one input characteristic, sometimes called a 1- rule. For nominal features, you can create a stump with leaves or a stump with two leaves for each possible sign value. One corresponds to the selected category, and the other to all others. For binary properties, the two schemes are the same. Missing values can be considered as another category. The decision stump's core idea is straightforward. Find a place to divide data effectively and focus on one aspect at a time. First, each time, we just take into account one variable. Second, we need to think about both ways. The left side is not always 0. It may be either the right or the side seen in Figure 3. It is necessary to find one number, (a). Input values that exceed an are classified as 1s. Anything that is less than or equal to an is given the designation "-1." We most efficiently classify our training data by determining the ideal number. Figure 3 shows the decision-stamp operations.

E. Bayes Net

Bayesian networks, sometimes referred to as Bayes networks, belief networks, or decision networks, are possible graphic representations of groups of variables and their interdependencies as determined by regularly administered graphs (DAGs). The event is expected to be covered by the Bayesian Network on the right, and any number of different elements should be given a chance to play a role in it. For instance, the Bayesian network can reflect a potential connection between symptoms and an illness. Using the grid, one may determine the likelihood of certain illnesses based on the symptoms. Figure 4 represents the operation of the Bayes Net algorithm. 3.5.1 Working with Bayes Net The probabilistic machine learning technique, the Bayes net algorithm, is used to solve classification issues. Its foundation is the Bayes theorem. Figure 4 represents the working of the Bayes algorithm for predicting heart disease. The working of the Bayes net classifier is explained in the steps below:

- Step 1: Collection and preprocessing of data.
- Step 2: Dataset splitting into training and testing data.
- Step 3: Calculation of prior probabilities.
- Step 4: Calculation of conditional probabilities.
- Step 5: Calculating the likelihoods.
- Step 6: Calculating posterior probabilities by using Bayes Theorem.
- Step 7: Performance Evaluation of Bayes net classification.

F. CNN Operation

CNN's primary characteristic is its capacity to carry out convolutional operations, which take an input picture and extract its characteristics. The CNN application used to forecast cardiac problems is shown in Figure 5. The convolutional layers of a CNN operate on the picture to extract features like edges, corners, and forms once the input layer gets the image. Following that, these characteristics are sent to the pooling layer, which helps to find the most significant features and lowers the dimensionality of the data. Following the layer for pooling, the fully linked layers handle the features and generate the result. CNN's output layer generates a probability distribution across the potential classes, and the class with the greatest probability is chosen as the final prediction.

G. Decision Table

A decision table is a straightforward visual depiction of specifying the action to be taken in certain circumstances. These algorithms provide a string of definable actions.

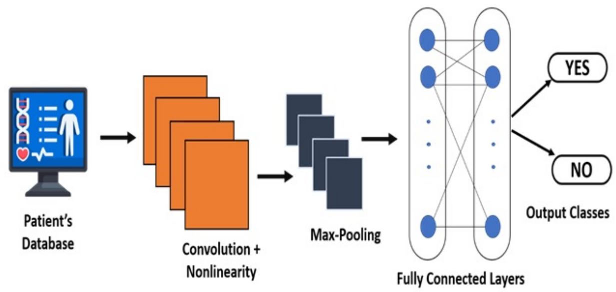

Figure 5. CNN Classifier for Heart Disease Prediction.

Random decision trees or if-then-else and switch-case statements from computer languages may be used to show the information on the decision board. In the case of continuous features, the boundary function is often specified for a range of values, and the bar has two sheets for values above and below the lower limit. Because you may choose numerous frames, the stems sometimes contain three or more leaves.

H. Overview of Proposed Methodology

To tackle the issue of heart disease, we have put forward a system based on the Internet of Things (IoT). This system comprises several components that collaborate to collect data from various devices and analyze both normal and abnormal conditions. A visual representation of the proposed system is provided in Figure 6.

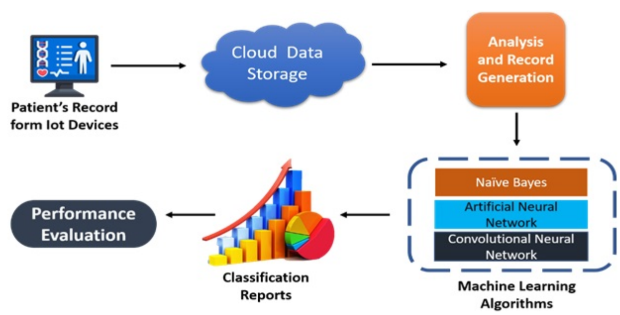

Figure 6. Overall operations of the proposed model.

With the rise of Internet services in recent years, cloud computing and the Internet of Things are important providers of services for many applications. Because latency-sensitive frameworks like surveillance systems and health monitoring require the use of massive amounts of data, centralized IoT-based computing platforms are necessary to address cloud challenges like the inability to meet requirements and limited scalability. A new area is also needed for these kinds of frameworks. The medical sector uses novel edge and fog computing frameworks to provide users with resources that are low-latency and energy-efficient, owing to the volume of data it handles. Fog computing, however, has drawbacks, such as slower reaction times and less accurate findings. These cutting-edge technologies, which include fog, IoT, and cloud computing with edge devices, provide improved solutions for low latency, security, privacy, and mobility, as well as improved communication, computation, and storage for networks. Decision Trees are used to create and assess a framework that identifies the most significant aspects of patient data to accurately forecast cardiac disease. This project seeks to construct a decision-tree model to identify and diagnose heart disease using a dataset provided by Public Health. This methodology encompasses many stages, one of which is data normalization. Unlike other machine learning models available for classification problems, decision trees are chosen for their interpretability, ability to handle both category and numerical data, and efficiency in training. Decision trees (DTs) are suitable for practical applications in healthcare settings due to their ability to give transparency, their low processing requirements, and their capability to handle both categorical and numerical data.

. The system employs machine learning techniques to achieve the objectives of categorization and anomaly detection. The analytics component of the system facilitates the provision of practical insights, diagnostics, and the creation of illness therapies. The technology also facilitates the implementation of effective authentication and data security procedures to guarantee the safe transfer of data across different platforms. Security methods such as TLS and SSL are used to cypher the data during transmission, and the data is securely stored in databases. These IoT devices are protected by security standards that ensure the efficacy of the overall system.

4. The proposed system performance evaluation

We utilized an HP Envoy Notebook with specifications that included an AMD RyzenTM 5 7530U processor (up to 4.5 GHz, 16 MB L3 cache), 8 GB of onboard graphics memory, 512 GB of NVMeTM M.2 SSD, and Windows 11 Home operating system to test the performance of the suggested solution. Experiments are conducted to assess the efficiency of the recommended method for CVD. Numerous machine learning methods, such as ANN and CNN, Decision Table, Zero R, One R, Decision stump, REP Tree, IBK, SGD Text, Bayes net, Naïve Bayes, and Naïve Bayes Multinomial Text, are applied to the CVD dataset with the use of the Weka tool. The precision, recall, F-measure, ROC area, MCC, TP rate, FP rate, and PRC area are used to assess the machine learning models. Cloud servers receive the Internet of Things (IoT) data collected via wireless body sensor networks (WBSNs) and use them for pre-processing and categorization operations. The TensorFlow machine learning package and the Cloud platform were used in the experiment. Apache Spark and Cassandra were used for server and storage infrastructure, respectively. Understanding the model's performance may be done using this. The evaluation metrics of the algorithms are as follows:

A. Model Evaluation

1) Accuracy computation

It is the total of the true positives (TP) and true negatives (TN) over the whole number of packets. As stated and counted by the total number of packets, accuracy is the number of properly categorized packets and metric that considers both precision and recall and is mostly used in information retrieval and classification

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2) Precision computation

True positives (TP) divided by the sum of false positives (FP) and true positives (TP) is how it is represented. According to the recommended classification technique, accuracy is the percentage of instances correctly assigned to a certain target application class.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3) Recall computation

It is the proportion of true positives (TP) to the total of false negatives (FN) and true positives (TP). It is the proportion of successfully categorized instances in an application class according to the suggested classification technique.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4) F1-score computation

It can be defined as the harmonic mean between precision and recall. It is a metric that considers both precision and recall and is mostly used in information retrieval and classification.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

5) Fitness Function Evaluation

Accuracy is explained as a measure of the total number of correctly classified packets from the actual dataset. In Eq (1), TP is true positive, which means how many times the number of packets are classified as yes when it was actually yes. TN is the true negative means: how many times the number of packets classified as no when it was no. FP is false positive, which means how many times the classifier had classified as yes when it was no. FN is false negative, which means how many times the classifier had classified as no when it was actually yes.

A decision table is a simple visual representation of defining the action to be performed under certain conditions. These are algorithms that lead to a series of actions. The information on the decision board can be displayed randomly as a decision tree or in a programming language such as if-then-else and switch-case statements. Evaluation metrics Precision, FP Rate, Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, ROC Area obtained by applying the decision table is graphically shown in Figure 7.

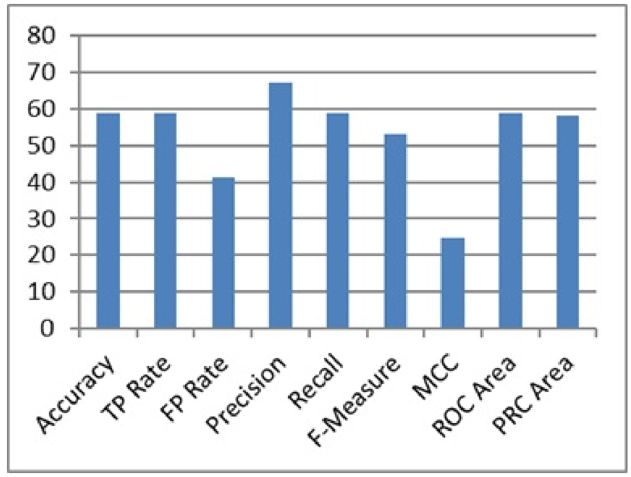

Figure 7. EVALUATION METRIC GRAPH FOR DECISION TABLE.

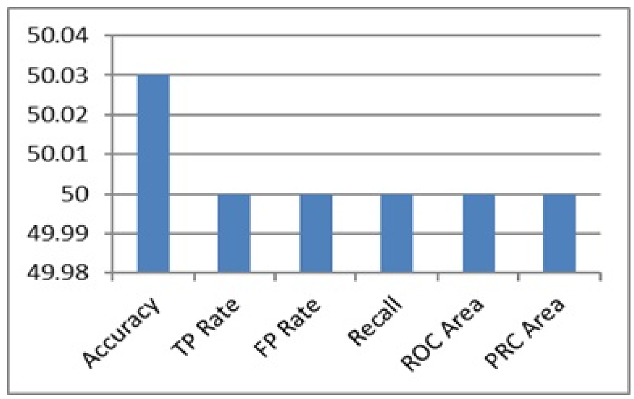

Figure 8. EVALUATION METRIC GRAPH FOR ZERO(R).

Machine learning algorithm Zero R is applied to the CVD dataset using 10-crossfold validation. The most straightforward classification technique, ZeroR, ignores all predictors and just depends on the target. Predicting the majority category (class) is the ZeroR classifier. ZeroR has no prediction power, but it might help set a performance baseline to compare with other classification techniques. The evaluation metrics generated by using ZeroR visually include Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-measure, and ROC Area obtained Figure 8. Machine learning algorithm Decision Stump is applied to the CVD dataset using 10-crossfold validation. Evaluation metrics Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, and ROC Area obtained by applying decision stump is graphically shown in Figure 9. Machine learning algorithm IBK is applied to the CVD dataset using 10-cross-fold validation. Just-in-time prediction generation for a test instance is what the IBK method performs, as opposed to model building. Each test instance in the training data is found using a distance measure by the IBK method, which then selects k ”close” examples from the training data and utilizes those instances to forecast. Figure 10 presents a visual representation of the evaluation metrics Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, ROC Area that were acquired by using IBK. A Bayesian network, for instance, might be used to show the probability associations between symptoms and illnesses. The network may be used to calculate the probability of the existence of different illnesses given symptoms. assessment metrics Figure 11 presents the results of implementing the Bayesian network regarding Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-measure, and ROC Area.

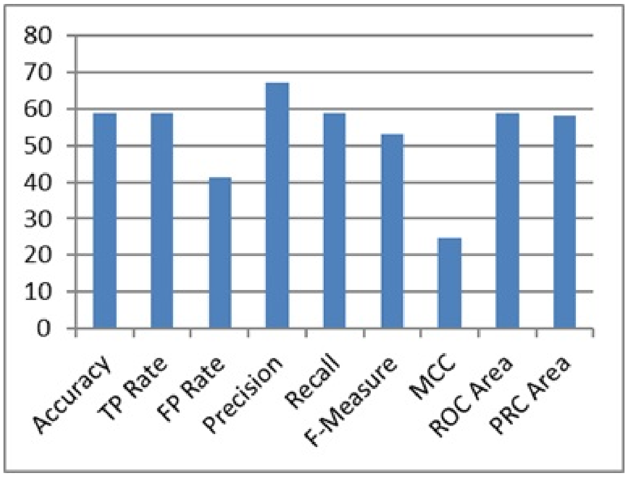

Figure 9. EVALUATION METRIC GRAPH FOR DECISION STUMP.

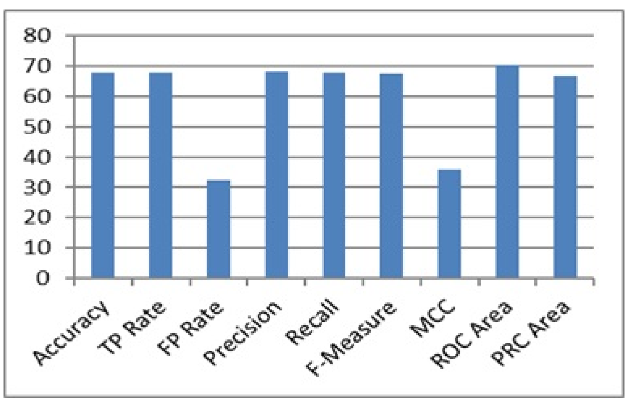

Figure 10. EVALUATION METRIC GRAPH FOR DECISION IBK.

Machine learning algorithm Naive Bayes is applied to the CVD dataset using 10-crossfold validation. Less training data is needed for NB. It can work with discrete and continuous data. When it comes to the quantity of predictors and data points, NB is very scalable. It's quick and useful for making predictions in real time. Assessment metrics. Figure 12 illustrates the results of using the decision table regarding Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, and ROC Area.

Convolutional neural network (CNN) is a deep learning model. CNN divides the data into subgroups using layers. We have a binary classification problem, so we convert our data into binary using OneHotEncoder(). CNN is applied to the dataset using pandas in Python. $train_{t}est_{s}plit$ function from sci-kit-learn and use 67% of the data for training and 33% for testing. A set of patients' cardiac measurements from a dataset are considered when they exhibit signs of CVD. It is thought that three months is enough time to monitor the signs and symptoms of CVD. Patients were receiving care from the departments that handled them. On a standard scale, the disease's symptoms are rated as follows: severely impacted, somewhat affected, not affected, partially improved, substantially improved, and entirely improved. About 150 individuals are randomized to an experimental therapy for observation in this trial. The figure illustrates how the outcome results are distributed based on the statistically significant test. As seen in this figure, more patients have significantly improved while fewer people have severe problems. The patients who are fully recovered and those who are just partially recovered differ little, and these results are shown in Figure 13.

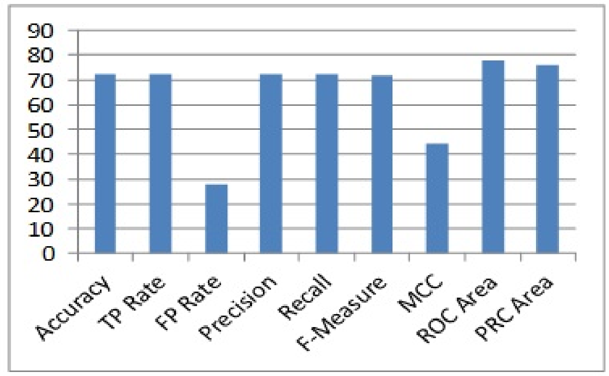

Figure 11. EVALUATION METRIC GRAPH FOR NAÏVE BAYES.

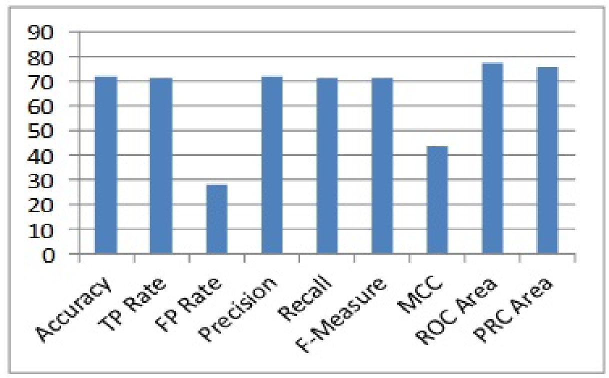

Figure 12. EVALUATION METRIC GRAPH FOR BAYES NET.

This artificial neural network is feed-forward and deep. Because data passes through the models directly, they are referred to as "feed-forward" models. The model does not have any feedback links via which its outputs might be fed back into it. The accuracy obtained by applying ML algorithms Bayes Net, ANN, NB, decision stump, decision

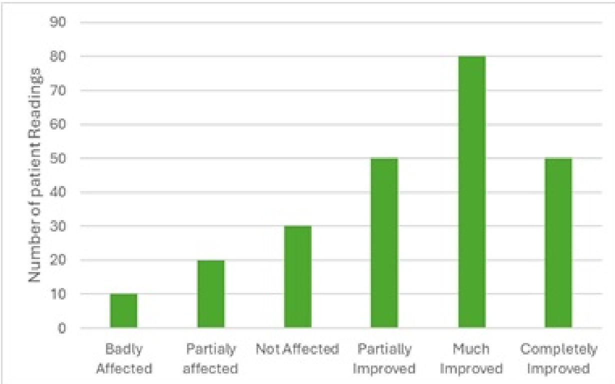

Figure 13. Comparison of CVD affected people.

TABLE I. Comparison of Different Algorithms for CVD

ML Models	Accuracy
Bayes Net	71.71
CNN	86
NB	72
Decision Stump	58.83
Decision Table	72.43
One R	50.03
NB Multinomial Text	50.03
NB Updateable	72
REP Tree	50.03

table, IBK, One R, Zero R, NB multinomial text, NB updateable, SGD text, and REP tree by using weka is shown in Table 1.

The accuracy of the CNN algorithm is good when compared to other algorithms. Several ML techniques are used on the CVD dataset, but as mentioned in Table 2, the CVD dataset achieves a higher accuracy of 76% with CNN.

TABLE II. Distribution of Datasets

Technique	Dataset	Accuracy
CNN	CVD	76 %
Naïve Bayes	UCI	84.24 %
CNN	Cleveland	83.2%

TABLE III. Comparison of F1-Score and Recall with other algorithms

Models	F1-Score	Recall
KNN	82.02%	81.80%
RF	81.80%	85.81%
NB	79.01%	84.52%
BN	78.28%	82.10%
DT	76.20%	79.37%
CNN	86.80%	87.50%

Table III shows the comparison of different models for CVD disease detection. We can notice that the CNN model outperforms all other models.

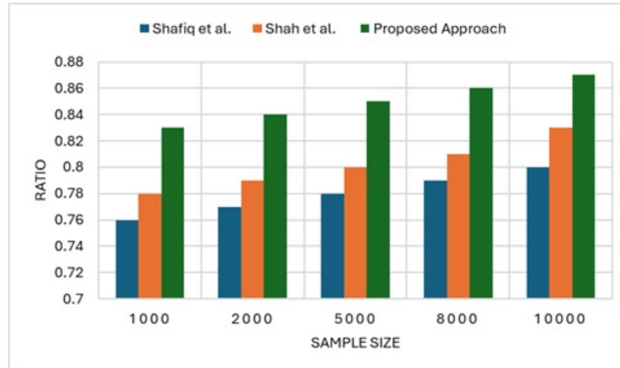

Figure 14. Performance comparison with other approaches.

In Figure 14, we compare the performance of the proposed work with existing approaches. Different samples are used to compare the performance.
When putting forth a CVD framework, we are unable to foresee how the evidence will be interpreted and applied by decision-makers; however, this could have an impact on policy outcomes if the suggested solution involves the policy community as a stakeholder or through other stakeholders. The work is more difficult: there are several pertinent approaches, demanding and often impatient customers, queries that need constant negotiation and rely as much on context as substance, literature with hazy borders, and few universally accepted criteria for excellence. However, some people are stepping up to the plate and working to create techniques for interpretative synthesis that will help decision-makers. Population-wide interventions have been implemented recently to address the risk factors of physical inactivity and poor nutrition, based on the theory that strong contextual contexts make it challenging for a person to maintain a behavior change. Food sources that are mostly composed of processed, packaged, high-calorie meals with no nutritional value and work conditions that promote sedentary behavior are two examples. a few issues with data quality, including unbalanced, excessive dimensionality, missing, noisy, and inconsistent data. The absence of interpretation in these models poses several dangers and limits to their application in real-world diagnostic decision support systems.
Resource-scarce circumstances may arise in any location, even in highly advanced and affluent nations. These inadequacies may be temporary and emerge after broad natural or man-made catastrophes, resulting from local shortages or disruptions in the availability of equipment or persons that are typically accessible. In addition, rural hospitals, clinics, and public health organizations often have regular shortages, while sparsely populated regions typically lack adequate healthcare services. Our Internet of Things (IoT) based cardiovascular disease (CVD) solution may assist in providing life-saving support to patients during crucial sit-

uations. The doctor may engage in communication with the patient and provide medication depending on the patient's condition after receiving data from the IoT-based system. In inaccessible regions, it is not feasible to establish well-equipped medical facilities for patients.
Almost all research conducted to detect indications of heart disease using artificial intelligence used rather tiny datasets. The characteristics were not designed for use in machine learning since a number of them lacked relevance and needed to be improved. Therefore, it was crucial to provide a dataset that contains several pertinent characteristics and a much larger number of samples. Resource-scarce conditions may occur in any place, even technologically sophisticated and wealthy countries. These deficiencies may be transient and arise after extensive natural or man-made disasters caused by local scarcities or interruptions in the accessibility of equipment or personnel that are usually available. Furthermore, rural hospitals, clinics, and public health organizations often have consistent deficiencies, while thinly populated places generally suffer from insufficient healthcare services. Our Internet of Things (IoT) cardiovascular disease (CVD) solution has the potential to provide vital assistance to patients in critical circumstances, potentially saving lives. Upon receiving data from the IoT-based system, the doctor may communicate with the patient and provide medicine based on the patient's condition. Establishing fully equipped medical facilities for patients is not practical in remote areas.

B. Limitation and Future Research Direction

Various forms of pictures are used for the detection of cardiovascular disease (CVD) in humans. These images have the potential to impact the effectiveness of the suggested remedy. Angiography is a diagnostic medical procedure that uses X-rays and a contrast substance to see the blood arteries throughout the body. Both approaches have inherent limitations, including invasiveness (which carries the danger of bleeding or infection), restricted access, high cost, and limited visualization. These restrictions impede the ability to acquire a comprehensive understanding of a certain timeframe. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are often used for the identification and diagnosis of cardiovascular conditions. Both techniques have inherent limitations, such as the need for the patient to remain immobile in a certain posture, exposure to ionizing radiation, costly treatments, and the inability to identify moderate or early-stage cardiac disease. Furthermore, several types of malfunctions in IoT devices might impact the outcomes. To enhance the effectiveness of these predictive classifiers in detecting heart disease, we will do further experiments in the future, using other feature selection approaches and optimization tactics. In addition, we will use generative artificial intelligence and robust machine learning techniques, including XEI. In order to train the classifiers, it is advisable to make use of additional real-time datasets. We will conduct further tests in the future to improve the efficacy of these predictive classifiers for the detection of heart disease by utilizing other feature selection methods and optimization strategies. We will also use generative AI and powerful machine learning methods, such as XEI. More real-time datasets may be used to train the classifiers. We will conduct more tests to enhance the efficacy of these prediction classifiers for diagnosing heart disease by using other feature selection methods and optimization strategies. Additionally, the proposed research would use completely wearable technologies that are now commercially accessible for training and testing purposes.

5. Conclusion

Diabetes, heart disease (HD), cancer, and chronic respiratory diseases are among the world's top causes of mortality. When predicting cardiovascular risk variables, machine learning approaches are quite helpful. It is difficult to determine who will survive cardiac disease. Certain methods for diagnosing cardiac disease were employed by earlier heart attack prediction systems. In this article, an alternative ML classifier-based IoT-based heart disease prediction system was developed. KNN, NB, CNN, decision trees, and random forests are machine-learning models that are used to predict cardiac disease. These methods function in three stages: pre-processing, feature selection from the dataset, and feature-based patient categorization. BMI, hyperglycemia, cholesterol, and diastolic and diastolic blood pressure, among other risk factors, are used to pick the characteristics. A sizable data set was used to assess the suggested system's performance. Additionally, compared to the current methods, the suggested methodology delivers a better degree of accuracy. After achieving a high accuracy of 86%, the study's best model is CNN. This research recommends CNN as a method for CVD prediction because of its greater accuracy achievements.

Funding Statement: The research reported in this publication was supported by the Deanship of Scientific Research and Innovation at Al-Balqa Applied University in Jordan. Grant Number: DSR-2022#480.

References

[1] S. Balasubramaniam, C. V. Joe, C. Manthiramoorthy, and K. S. Kumar, "Relieff based feature selection and gradient squirrel search algorithm enabled deep maxout network for detection of heart disease," *Biomedical Signal Processing and Control*, vol. 87, p. 105446, 2024.

[2] M. Aftab, R. Amin, D. Koundal, H. Aldabbas, B. Alouffi, and Z. Iqbal, "Classification of covid-19 and influenza patients using deep learning," *Contrast Media & Molecular Imaging*, vol. 2022, no. 1, p. 8549707, 2022.

[3] H. Aldabbas, D. Albashish, K. Khatatneh, and R. Amin, "An architecture of iot-aware healthcare smart system by leveraging machine learning." *Int. Arab J. Inf. Technol.*, vol. 19, no. 2, pp. 160–172, 2022.

[4] R. Amin, M. A. Al Ghamdi, S. H. Almotiri, M. Alruily *et al.*, "Healthcare techniques through deep learning: issues, challenges and opportunities," *IEEE Access*, vol. 9, pp. 98 523–98 541, 2021.

[5] J. Gong and N. J. Navimipour, "An in-depth and systematic literature review on the blockchain-based approaches for cloud computing," *Cluster Computing*, pp. 1–18, 2021.

[6] H. Kibriya, R. Amin, J. Kim, M. Nawaz, and R. Gantassi, “A novel approach for brain tumor classification using an ensemble of deep and hand-crafted features,” *Sensors*, vol. 23, no. 10, p. 4693, 2023.

[7] C. V. B. Murthy, M. L. Shri, S. Kadry, and S. Lim, “Blockchain based cloud computing: Architecture and research challenges,” *IEEE Access*, vol. 8, pp. 205 190–205 205, 2020.

[8] S. Ramzan, A. Aqdus, V. Ravi, D. Koundal, R. Amin, and M. A. Al Ghamdi, “Healthcare applications using blockchain technology: Motivations and challenges,” *IEEE Transactions on Engineering Management*, vol. 70, no. 8, pp. 2874–2890, 2022.

[9] H. Kibriya and R. Amin, “A residual network-based framework for covid-19 detection from cxr images,” *Neural Computing and Applications*, vol. 35, no. 11, pp. 8505–8516, 2023.

[10] K. Vembandasamy, R. Sasipriya, and E. Deepa, “Heart diseases detection using naive bayes algorithm,” *International Journal of Innovative Science, Engineering & Technology*, vol. 2, no. 9, pp. 441–444, 2015.

[11] A. Garavand, A. Behmanesh, N. Aslani, H. Sadeghsalehi, and M. Ghaderzadeh, “Towards diagnostic aided systems in coronary artery disease detection: a comprehensive multiview survey of the state of the art,” *International Journal of Intelligent Systems*, vol. 2023, no. 1, p. 6442756, 2023.

[12] A. Srivastava, S. Samanta, S. Mishra, A. Alkhayyat, D. Gupta, and V. Sharma, “Medi-assist: A decision tree based chronic diseases detection model,” in *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE, 2023, pp. 1–7.

[13] J. Rwebembera, J. Marangou, J. C. Mwita, A. O. Mocumbi, C. Mota, E. Okello, B. Nascimento, L. Thorup, A. Beaton, J. Kado *et al.*, “2023 world heart federation guidelines for the echocardiographic diagnosis of rheumatic heart disease,” *Nature Reviews Cardiology*, vol. 21, no. 4, pp. 250–263, 2024.

[14] M. I. Hossain, M. H. Maruf, M. A. R. Khan, F. S. Prity, S. Fatema, M. S. Ejaz, and M. A. S. Khan, “Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison,” *Iran Journal of Computer Science*, vol. 6, no. 4, pp. 397–417, 2023.

[15] S. Islam, N. Jahan, and M. E. Khatun, “Cardiovascular disease forecast using machine learning paradigms,” in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2020, pp. 487–490.

[16] R. Ahmed, M. Bibi, and S. Syed, “Improving heart disease prediction accuracy using a hybrid machine learning approach: a comparative study of svm and knn algorithms,” *International Journal of Computations, Information and Manufacturing (IJCIM)*, vol. 3, no. 1, pp. 49–54, 2023.

[17] R. Khan and G. L. Masala, “Detecting breast arterial calcifications in mammograms with transfer learning,” *Electronics*, vol. 12, no. 1, p. 231, 2023.

[18] M. Cui, X. Gang, F. Gao, G. Wang, X. Xiao, Z. Li, X. Li, G. Ning, and G. Wang, “Risk assessment of sarcopenia in patients with type 2 diabetes mellitus using data mining methods,” *Frontiers in endocrinology*, vol. 11, p. 123, 2020.

[19] F. Guo, M. Ng, M. Goubran, S. E. Petersen, S. K. Piechnik, S. Neubauer, and G. Wright, “Improving cardiac mri convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach,” *Medical image analysis*, vol. 61, p. 101636, 2020.

[20] J. Maiga, G. G. Hungilo *et al.*, “Comparison of machine learning models in prediction of cardiovascular disease using health record data,” in *2019 international conference on informatics, multimedia, cyber and information system (ICIMCIS)*. IEEE, 2019, pp. 45–48.

[21] E. Y. Boateng, D. A. Abaye *et al.*, “A review of the logistic regression model with emphasis on medical research,” *Journal of data analysis and information processing*, vol. 7, no. 04, p. 190, 2019.

[22] A. Menshawi, M. M. Hassan, N. Allheeib, and G. Fortino, “A hybrid generic framework for heart problem diagnosis based on a machine learning paradigm,” *Sensors*, vol. 23, no. 3, p. 1392, 2023.

[23] K. Gupta, P. Kumar, S. Upadhyaya, M. Poriye, and S. Aggarwal, “Fuzzy logic and machine learning integration: Enhancing healthcare decision-making,” *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 16, no. 3, pp. 20–20, 2024.

[24] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, “Machine learning-based predictive models for detection of cardiovascular diseases,” *Diagnostics*, vol. 14, no. 2, p. 144, 2024.

[25] M. Ullah, S. Hamayun, A. Wahab, S. U. Khan, M. U. Rehman, Z. U. Haq, K. U. Rehman, A. Ullah, A. Mehreen, U. A. Awan *et al.*, “Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective,” *Current Problems in Cardiology*, vol. 48, no. 11, p. 101922, 2023.

[26] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and Y. D. Alotaibi, “Diagnosis of cardiovascular disease using deep learning technique,” *Soft Computing*, vol. 27, no. 13, pp. 8971–8990, 2023.

[27] K. Kanagarathinam, D. Sankaran, and R. Manikandan, “Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset,” *Data & Knowledge Engineering*, vol. 140, p. 102042, 2022.