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

Hamza Aldabbas[1], Zaid Mustafa[2]

[1]Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt- Jordan, email: aldabbas@bau.edu.jo
[2]Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt- Jordan, E-mail: z_lami@bau.edu.jo

Corresponding author: Hamza Aldabbas

Abstract: The Internet of Things (IoT) and Machine Learning (ML) models are emerging technologies that are changing our daily lives. These are also considered as 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 due to several risk factors like an unhealthy diet, sugar, high Blood Pressure (BP) smoking, etc. Preventive treatment and early intervention for those at risk 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 are capable of handling sequential time-series data. In remote places 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 is used that contains data of 70000 patients gathered at the time of 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 utilizing the support system.

Keywords: Internet of things, Machine Learning, heart disease, Decision tree, KNN, Disease detection, Naïve Bayes, Support Vector Machine (SVM).

I. 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 are used to 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 [1, 3, 4]. An estimated 20 million people dies due to cardiovascular diseases in 2020 which is 35% of all global deaths. 85% of cardiovascular deaths happen due to heart attacks and strokes [5]. 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 get blocked because of fatty buildup, which stops blood flow to the heart and brain. Heart attack and stroke are also caused when blood from blood vessels starts bleeding in the brain. Cardiovascular disease is also caused due to many other risk factors like no exercise, tobacco use, oily food items, unhealthy diet, obesity, unsafe use of alcohol, high blood pressure, diabetes, etc. People with these risk factors are more likely to have cardiovascular disease or get cardiovascular disease [7]. To save the lives of humans, we need to address this issue by early detection of cardiovascular disease [8].

Early prediction of cardiovascular disease can save humans' lives and also minimize the cost of treatment people can get better counseling [9]. The ability to predict CVD may significantly affect how it is treated and perhaps save lives [10]. By addressing behavioral 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 [8, 11]. This study offers a method for predicting cardiovascular illness's presence or absence based on machine learning approaches.

To provide universal, high-quality, and reasonably priced medical services, we used an IoT-based platform. A significant quantity of clinical data is produced by our ubiquitous healthcare application. For further processing and analysis, this data has to be maintained appropriately. The integration of 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 a multitude of suggestions gathered from the cardiologists, the system gives the patient the most appropriate advice. We proposed an IoT-based secure system in which different type of human health data is collected and analyzed using popular machine learning algorithms. Our study aims to improve accuracy and system efficiency such that it can forecast the likelihood of a heart attack using standard medical equipment. A CNN model with the best-performing ML algorithms was chosen to meet the aims and acquire an effective high rank. For the real-time data collected, an IoT system was designed to forecast the likelihood of developing CVD. Finally, the effectiveness, efficiency, and satisfaction of the system were reviewed by doctors and patients.

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 to detect cardiovascular attacks by using different machine learning models. Performance evaluation of the proposed approach is mentioned in Section IV and last section conclude the paper.

II. LITERATURE REVIEW

[12] evaluated six machine learning techniques, Support Vector Machine, artificial neural network, Logistic regression, Naıve Bayes (NB), and 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. [13] proposed an approach to predict CVD. CVD causes a high number of causalities anywhere in the world. Data is collected to predict CVD that 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%, 85%, 86.1%, and 75.8%, respectively.

[14] analyze and conclude that CVD is the leading cause of death in the world. 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. The investigation is done on 18 types of research, most from 2018 and 2021. They made an investigation using DT-J48, NB, SVM, ANN, KNN, LR, RF in 11, 10, 9, 8, 3, 3, 1, types of research, respectively. [15] [15] propose a fuzzy logic inference system (HFIS) applied to CVD data that shows effective results in predicting cardiovascular disease to predict the level of CVD. A hierarchical fuzzy inference system (HFIS) is used. CVD represents 31% of deaths around the globe. More people die due to cardiovascular disease than any other cause. [16] 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. Cardiovascular disease is caused due to several risk factors 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. [17] used neural networks to predict human (driver) heart activity to check the possibility of an accident and recommend automatically breaking or stopping the car. Ordinary people’s lives is full of difficulties, especially for drivers that affect their health badly. This study developed an algorithm integrated with the driver support system embedded with the onboard automobile. CNN is used

to train the model on the original data, giving poor accuracy to the two weakest classes, 8% and 19 %. [18] 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 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.

[19] apply ML techniques to predict CVD that is caused due to 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. [20] analyze the performance of the logistic regression algorithm used by previous researchers between 2000 to A multivariate technique called logistics regression is used to determine the correlations between independent and dependent variables. thirty-seven research publications (published between 2000 and 2018) that included logistic regression as a primary model in their study, as well as six book reviews. This research identifies flaws in the use and reporting of LR. Goodness-of-fit metrics, regression diagnostics, and validation analysis were not reported in several research. 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.

[21] used a machine-learning algorithm to predict cardiovascular diseases and did a comparison of two data mining algorithms to find the best algorithm. The dataset contains medical data of hospitalized patients with cardiovascular diseases between March 2016 and March 2017. After data were normalized and cleaned in SPSS (V23.0) and Excel 2013, 25 characteristics that impact CVD were added to the dataset, and R 3.3.2 was used to do statistical calculations. 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. [22] 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.

III. PROPOSED METHODOLOGY

To address the problem of heart disease, we proposed an IoT-based system in which several components work together to get the data from different devices and perform analysis to indicate the 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 and forecast a patient's risk of coronary heart disease. The input data set is the one related to heart disease. Substituting non-available values with column means is the next preprocessing step. Figure 2 illustrates the various tactics that were used. The machine learning models' output represents their accuracy metric. Predictions using the model are then possible. The medical specialists assisted in gathering the dataset from a reputable hospital. The following features are included in the dataset: blood pressure, heart rate, age, gender, history of diabetes, family history of coronary artery disease, heartbeat rhythm, ECG readings, chest discomfort, smoker-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.

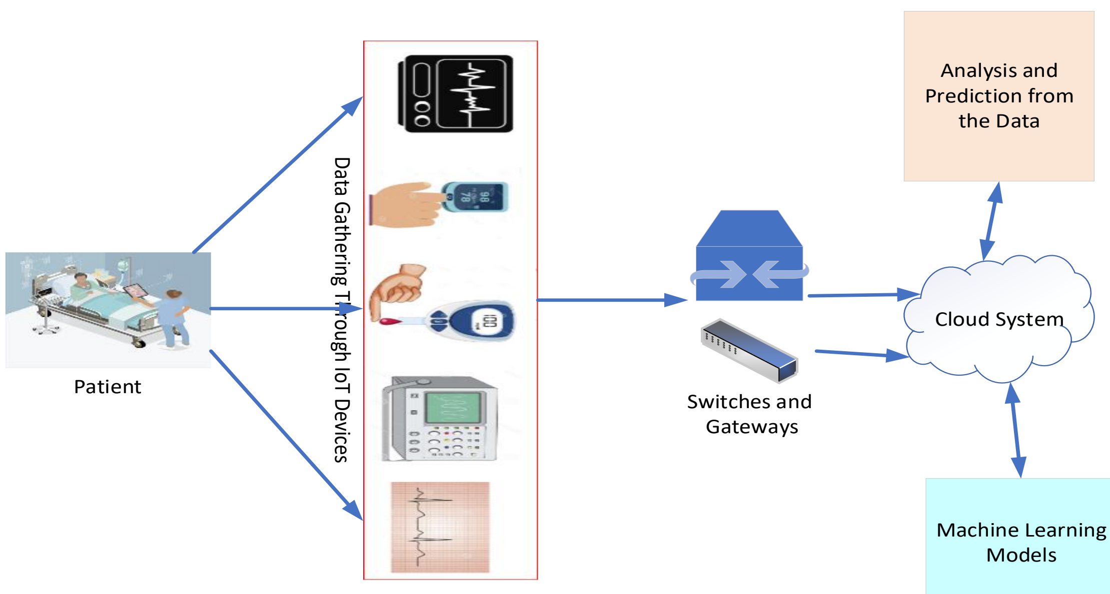


Figure 1: IoT-based Proposed system model for patient diagnosis.

3.1 Data Acquisition (Data Set)

In any machine-learning-based problem, the first step is getting the dataset [23] processed in the next stages. In this research, we selected the cardiovascular disease (CVD) dataset because it’s new in the market, so not so much work was 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 publicly available on kaggle.com for research purposes [24]. The Cardiovascular diseases dataset was released on kaggle.com in 2018. 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 findings' accuracy 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.

3.2 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 apply to the CVD dataset.

3.3 Operation of Naive Bayes algorithm (NB)

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 remove 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. Consideration is given to linear time rather than the expensive iterative 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.

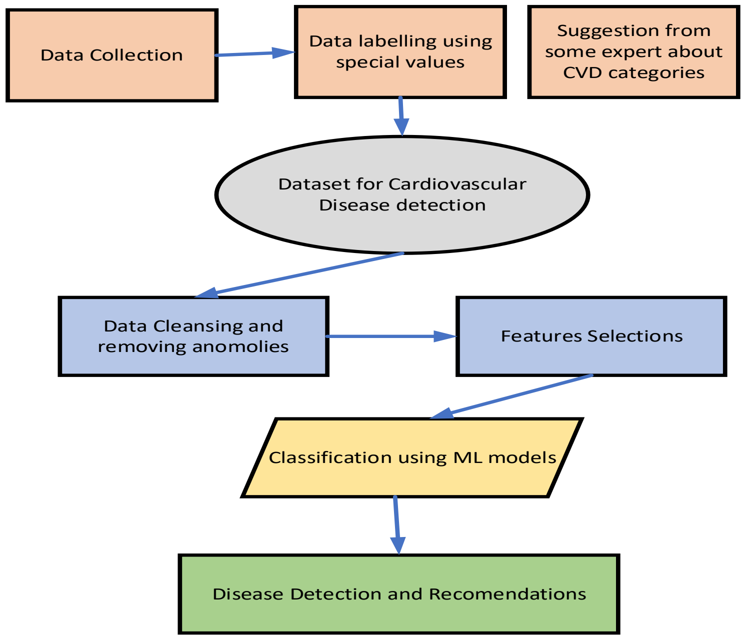


Figure 2: Entire system model to detect CVD

3.4 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 other categories. 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. The following Equations 1 and 2 are used to show the decision stamps operations.

$$H = \{h_{a,c} : a \in R \; c \in +1, -1\} \tag{1}$$

$$h_{(a,c)}(x)\left\{ c = -1 \; if \; x \leq a, c = 1 \; if \; x > a \right. \tag{2}$$

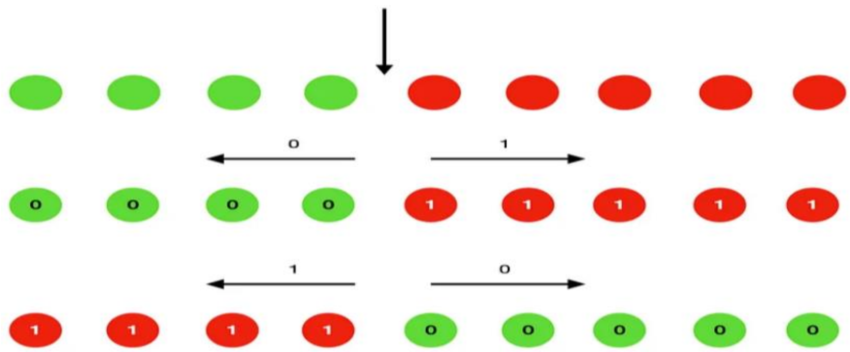


Figure 3: Operation of decision stamp

3.5 Bayes Net

Bayesian networks, sometimes referred to as Bayes networks, belief networks, or decision networks, are possible graphic representations of group 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 given the chance for any number of different elements to have 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 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. The working of 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.

Figure 4 represents the working of Bayes algorithm for predicting heart disease.

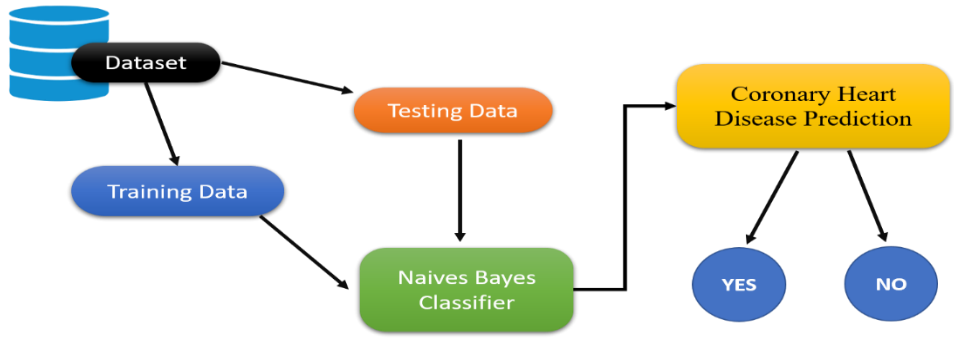


Figure 4: Naïve Bayes Classifier for Heart Disease Prediction

3.6 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. A probability distribution across the potential classes is generated by CNN's output layer, and the class with the greatest probability is chosen as the final prediction.

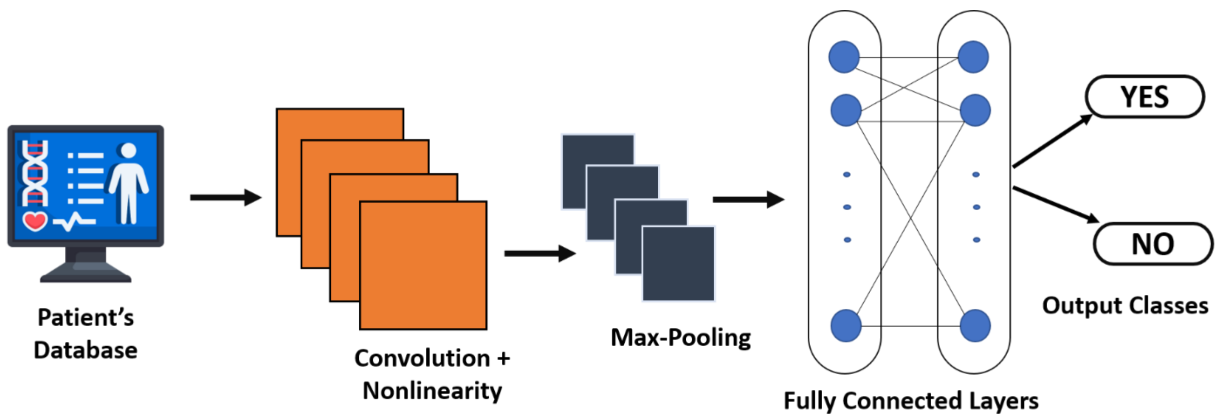


Figure 5: CNN Classifier for Heart Disease Prediction

3.7 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. 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.

3.8 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 the Figure 6.

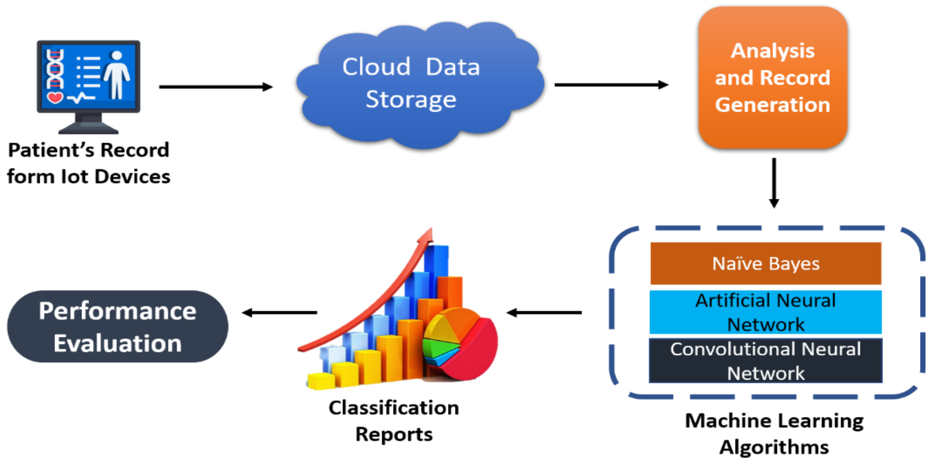


Figure 6: Overall operations of the proposed model

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 carried out 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:

4.1 Model Evaluation

4.1.1 Accuracy computation

Over the whole number of packets, it is the total of the true positives (TP) and true negatives (TN). As stated and counted by the total number of packets, accuracy is the number of properly categorized packets and the number of incorrectly classified packets. It is used to determine the overall dataset's correctness.

$$Acc = \frac{TP + TN}{TP + FN + FP + FN} \quad (1)$$

4.1.2 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. $Percision = \frac{TP}{TP + FP}$ (2)

4.1.3 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. $\mathrm{Re}call = \frac{TP}{TP + FN}$ (3)

4.1.4 F1-score computation

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

$$F1 - score = 2 * \frac{Percision * \mathrm{Re}call}{Percision + \mathrm{Re}call} \quad (4)$$

4.2 Fitness Function Evaluation

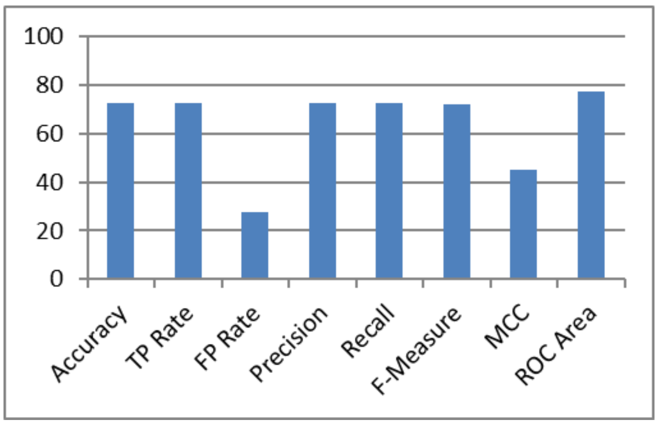
Accuracy is explained as a measure of the total number of correctly classified packets from the actual dataset. In mathematical terms, accuracy is written as:

$$Accuracy = \frac{TP + TN}{Total} \quad (1)$$

$$ErrorRate = \frac{FP + FN}{Total} \; ErrorRate = 1 - Accuracy \quad (5)$$

In Eq (1) and Eq (5) , TP is true positive means: how many times the number of packets that 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 means how many times the classifier has classified as yes when it was no, and FN is false negative means how many times the classifier has 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 Persicion, 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.



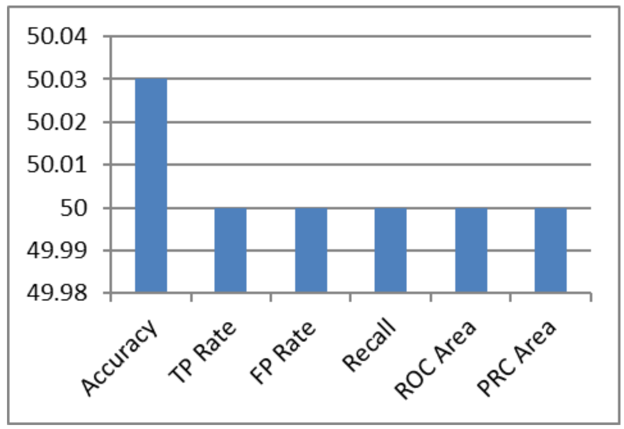
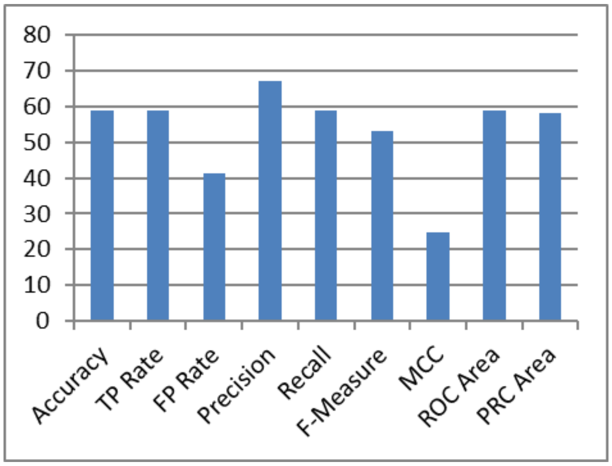


FIGURE 7: EVALUATION METRIC GRAPH FOR DECISION TABLE FIGURE 8: EVALUATION METRIC GRAPH FOR ZERO(R)

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, ROC Area obtained (Figure 8). Machine learning algorithm Decision Stump is applied on CVD dataset using 10-crossfold validation. Evaluation metrics Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, ROC Area obtained by applying decision stump is graphically shown in Figure 9.

Machine learning algorithm IBK is applied to 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 in terms of Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, ROC Area.



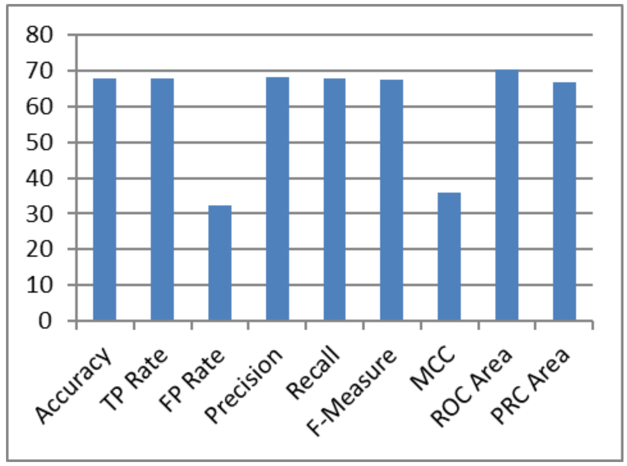
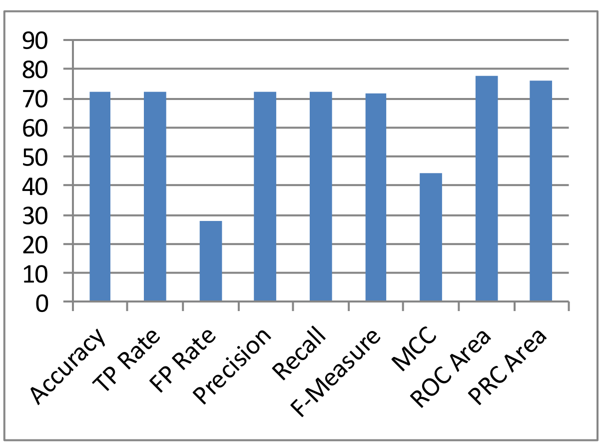


FIGURE 9: EVALUATION METRIC GRAPH FOR DECISION STUMP

FIGURE 10: EVALUATION METRIC GRAPH FOR IBK

Machine learning algorithm Naive Bayes is applied to 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 in terms of Accuracy, Recall, ROC, TP Rate, MCC, Recall, F-Measure, ROC Area.



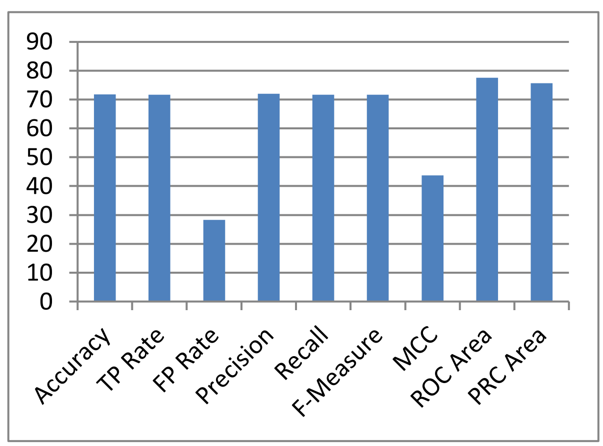


FIGURE 12: EVALUATION METRIC GRAPH FOR NAÏVE BAYES

FIGURE 11: EVALUATION METRIC GRAPH FOR BAYES NET

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_test_split() function from scikit-learn and use 67% of the data for training and 33% for testing. 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 table, IBK, One R, Zero R, NB multinomial text, NB updateable, SGD text, and REP tree by using weka is shown in Table 1.

Table 1: Comparison of Different Algorithms for CVD

ML Technique	**Accuracy**
Bayes Net	71.71
CNN	76
NB	72
Decision Stump	58.83
Decision Table	72.43
IBK	67.73
One R	50.03
Zero R	50.03
NB Multinomial Text	50.03
NB Updateable	72
SGD Text	49.99
REP Tree	50.03

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

Table 2: Distribution of Datasets

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

V. 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. People's lives are impacted by CVD; this makes accurate CVD prediction desirable. An updated approach to early cardiovascular disease prediction is put forward. The data of 70000 CVD patients obtained at the time of a medical examination is included in the CVD dataset in this suggested manner. The purpose of this research was to assess machine learning models for the prediction of cardiovascular disease using several risk variables, including BMI, hyperglycemia, cholesterol, diastolic and diastolic blood pressure, and others. ML algorithms such as KNN, NB, and random forests. After achieving a high accuracy of 76%, 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. Sujith, A., et al., *Systematic review of smart health monitoring using deep learning and Artificial intelligence.* Neuroscience Informatics, 2022. **2**(3): p. 100028.
2. Ramzan, S., et al., *Healthcare applications using blockchain technology: Motivations and challenges.* IEEE Transactions on Engineering Management, 2022.
3. Aldabbas, H., et al., *An architecture of IoT-aware healthcare smart system by leveraging machine learning.* Int. Arab J. Inf. Technol., 2022. **19**(2): p. 160-172.
4. Pramod, A., H.S. Naicker, and A.K. Tyagi, *Machine learning and deep learning: Open issues and future research directions for the next 10 years.* Computational analysis and deep learning for medical care: Principles, methods, and applications, 2021: p. 463-490.
5. Injadat, M., et al., *Machine learning towards intelligent systems: applications, challenges, and opportunities.* Artificial Intelligence Review, 2021. **54**(5): p. 3299-3348.
6. Nandy, S., et al., *An intelligent heart disease prediction system based on swarm-artificial neural network.* Neural Computing and Applications, 2023. **35**(20): p. 14723-14737.
7. Tiwari, A., A. Chugh, and A. Sharma, *Ensemble framework for cardiovascular disease prediction.* Computers in Biology and Medicine, 2022. **146**: p. 105624.
8. Li, J.P., et al., *Heart disease identification method using machine learning classification in e-healthcare.* IEEE access, 2020. **8**: p. 107562-107582.
9. Ali, F., et al., *A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion.* Information Fusion, 2020. **63**: p. 208-222.
10. Ozcan, M. and S. Peker, *A classification and regression tree algorithm for heart disease modeling and prediction.* Healthcare Analytics, 2023. **3**: p. 100130.
11. Shah, D., S. Patel, and S.K. Bharti, *Heart disease prediction using machine learning techniques.* SN Computer Science, 2020. **1**(6): p. 345.
12. Dwivedi, A.K., *Performance evaluation of different machine learning techniques for prediction of heart disease.* Neural Computing and Applications, 2018. **29**: p. 685-693.
13. Islam, S., N. Jahan, and M.E. Khatun. *Cardiovascular disease forecast using machine learning paradigms*. in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. 2020. IEEE.
14. Shafiq, M., et al., *A machine learning approach for feature selection traffic classification using security analysis.* The Journal of Supercomputing, 2018. **74**: p. 4867-4892.
15. Zhen, L. and L. Qiong, *A new feature selection method for internet traffic classification using ml.* Physics Procedia, 2012. **33**: p. 1338-1345.
16. Cui, M., et al., *Risk assessment of sarcopenia in patients with type 2 diabetes mellitus using data mining methods.* Frontiers in endocrinology, 2020. **11**: p. 123.
17. Saeed, A., et al., *Deep Physiological Arousal Detection in a Driving Simulator using Wearable Devices.*
18. Guo, F., et al., *Improving cardiac MRI convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach.* Medical image analysis, 2020. **61**: p. 101636.
19. Maiga, J. and G.G. Hungilo. *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)*. 2019. IEEE.

20. Boateng, E.Y. and D.A. Abaye, *A review of the logistic regression model with emphasis on medical research.* Journal of data analysis and information processing, 2019. **7**(04): p. 190.
21. Menshawi, A., et al., *A Hybrid Generic Framework for Heart Problem diagnosis based on a machine learning paradigm.* Sensors, 2023. **23**(3): p. 1392.
22. Wang, J., et al., *Detecting cardiovascular disease from mammograms with deep learning.* IEEE transactions on medical imaging, 2017. **36**(5): p. 1172-1181.
23. Kanagarathinam, K., D. Sankaran, and R. Manikandan, *Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset.* Data & Knowledge Engineering, 2022. **140**: p. 102042.
24. ULIANOVA, S. *Cardiovascular Disease dataset*. 2019 [cited 2023; Available from: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data.