https://doi.org/10.48047/AFJBS.6.7.2024.3658-3668

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Review Article Open Access

An Intelligent Cardio Vascular Disease Prediction Using KNN with Data Mining

[1]R. Leela Jyothi, [2]Naga Malleswara rao .P, [3]D.V.N. Bharathi, [4]Kocherla Anusha, [5]K Vijay Kumar

[1]Assistant Professor, Dept. of Computer Science and Engineering, SRKR Engineering College (A), SRKR Marg, china Amiram, Bhimavaram, A.P, India

[2]Assistant professor, Dept. of Computer Science and Engineering, Kallam Haranadh Reddy Institute of Technology, chowdavaram, Guntur, A.P, India

[3]Assistant Professor, Dept. of electronics and communication engineering, SRKR Engineering College (A), SRKR Marg, china Amiram, Bhimavaram, A.P, India

[4]Assistant Professor, Dept. of Computer Science and Engineering, Sri Mittapalli College of Engineering (A), Tummalapalem, Guntur, A.P, India

[5]Associate Professor, Dept. of Computer Science and Engineering, Sri Mittapalli College of Engineering (A), Tummalapalem, Guntur, A.P, India

Volume 6 issue 7 2024

Received:01 June 2024

Accepted:30June2024

doi:10.48047/AFJBS.6.7.2024.3658-3668

ABSTRACT: A disorder known as heart disease is when the heart stops working due to blocks in the blood vessels. Multiple research studies have come to the conclusion that this condition is now the leading cause of death in patients. Early diagnosis is essential for providing the proper treatment and could save the lives of many patients. Based on the symptoms or characteristics of a specific sample or instance, heart disease prediction requires accurately classifying a given sample as either heart disease positive or heart disease negative. As a result, artificial intelligence and machine learning have a significant impact on the healthcare industry, familiarizing people with data processing methods suitable for numerical health data. The process of searching through large amounts of data for important information is known as data mining. In this work, an intelligent cardiovascular disease prediction using KNN with data mining is presented. Datasets collected from several sources, including Kaggle and the machine-learning repository at UCI (University of California, Irvine), are used to analyze this model. The cognitive process of extracting hidden approach patterns from massive data sets is called data mining. Heart disease is predicted using K-nearest Neighbor. The efficacy of prediction models was evaluated using evaluation performance such as F-measure, accuracy, precision, and recall.

KEYWORDS: Heart, Cardio Vascular Disease (CVD), Machine Learning, Data Mining and K-Nearest Neighbor.

1. INTRODUCTION

The term "heart disease" refers to an extensive variety of conditions that affect the heart. Cardiovascular diseases (CVDs) include heart disease, which is the leading cause of death worldwide [1].A major financial burden and a deadly worldwide health issue is cardiovascular disease. The diseases which represent the greatest threat to life and have the highest death rate worldwide are thought to be cardiovascular diseases. In recent times, cardiovascular diseases (CVDs) have been identified as the primary cause of morbidity and mortality worldwide [2]. Every organ in the human body has a specific purpose. One such organ that pumps blood throughout the body is the heart; a malfunction could result in serious health problems. Worldwide, heart disease represents among the leading causes of death [3]. Chest pain, fatigue, high blood pressure, arm pain, and dizziness are among the main signs and symptoms of a heart attack [4]. In basic terms, heart failure is a condition in which the heart is unable to sufficiently pump blood to the body's organs. Typically, conditions like diabetes, hypertension, or other heart conditions like HIV, thyroid issues, alcoholism, or genetic problems. The heart's capacity to pump blood is limited as the heart's muscles deteriorate. It is important to note that heart disease is one of the most prevalent diseases among middle-aged people. The World Health Organization (WHO) has determined that 31% of deaths worldwide annually are caused to cardiovascular disease (CVD), a disorder that affects the heart and blood vessels. Therefore, it is crucial to identify the problem accurately and early in order to be able to provide the correct type of therapy and save the lives of many individuals [5].

They have grown more common over time and are currently overstretching countries healthcare systems. High blood pressure, stress, age, gender, cholesterol, Body Mass Index (BMI), and an unhealthy lifestyle are the main causes of cardiovascular diseases [6]. Acute cardiovascular events, including ST segment elevation myocardial infarction (STEMI) and other cardiovascular emergencies such as non-STEMI (NSTEMI), arrhythmia, hypertensive crisis, acute heart failure (HF), and pulmonary embolism (PE), are the most serious type of cardiovascular disease (CVD) and are life-threatening situations that need to be diagnosed and treated immediately [7].

Reduced risk of cardiovascular deaths can result from controlling cardiovascular risk factors, as recommended by the WHO. Noninvasive ultrasound imaging methods can be used to study atherosclerosis, a common type of CVD. The CVD biomarker intima-media thickness (IMT) has been validated. However, it does not significantly improve the prediction of future cardiovascular events as indicated by conventional risk variables, as a recent meta-analysis has shown [8]. CVDs generate significant lifetime disabilities and a non-negligible economic burden on their patients. But with the right precautions, 90 percent of CVDs are thought to be avoidable. Accordingly, it is important in the medical field to predict when a person would develop CVDs [9].

Many characteristics have been linked to an increased risk of heart disease: I Smoking Use: ii) Higher cholesterol: iii) inappropriate diet habits; iv) A lack of activity; v) Dangerous alcohol consumption; vi) Elevated blood sugar levels; vii) Extreme Overstress; and viii) Blood pressure. Age, gender, and a family history of diabetes can all contribute to blood pressure, a rare condition in which blood powerful enough for the walls and ultimately cause health issues [10].

Heart failure, stroke, coronary artery disease, and hypertension (high blood pressure) are among the most prevalent CVDs. CVDs come in a number of types and can present with an extensive range of symptoms. Chest pain or discomfort, fatigue, irregular heartbeats, and shortness of breath are some of the symptoms of some CVDs, including heart failure and coronary artery disease. A person can suffer from weakness or numbness on one side of their body, difficulty speaking or understanding others, vision loss, and severe headaches from other CVDs such stroke and aneurysms. In some cases, the symptoms of different CVDs may overlap, making it challenging to differentiate between them just by looking at the symptoms. For example, discomfort or discomfort in the chest may indicate heart failure along with coronary artery disease [11].

It is essential for medical experts to predict cardiovascular disease early in order to make informed decisions. Heart failure, heart attacks, strokes, and coronary artery disease can all be avoided with an accurate diagnosis of heart disease [12].The traditional diagnosis of heart disease is based on medical history, reports, and an expert medical examination of the symptoms.

This method takes a long time to complete and isn't always successful. Decision-making tools like machine learning (ML) have recently been created to help practitioners in accurately diagnosing patients [13]. By learning from medical data, making decisions, and predicting the disease, an automatic prediction system can be developed as an analytical tool using machine learning (ML) algorithms to handle complex health problems. Indeed, machine learning methods have recently been used to develop systems for diagnosing cardiac disease [14].
Data mining is becoming more and more common in today's world of medical treatment since it provides a wide range of complex knowledge related to patients, medical devices, medications, healthcare facilities, and disease diagnosis. It is necessary to process and assess such complex data in order to retrieve information that, when made into decisions, is price-effective and beneficial. Based on the symptoms or characteristics of a particular sample or instance, heart disease prediction requires accurately classifying a given sample as either heart disease positive or heart disease negative class [15]. There are models for heart disease prediction in the literature, and numerous researchers have created intelligent models with the goal of enhancing the performance of heart disease prediction models. Still, there is more possibility to improve the current model's efficiency in predicting heart disease [16].

In this work, an intelligent cardiovascular disease prediction using KNN with data mining is presented. The following is the way the work is organized: The literature survey is included in section 2. The section 3 demonstrates an intelligent cardiovascular disease prediction using KNN with data mining. The section 4 evaluates the result analysis. The conclusion is presented in section 5.

2. LITERATURE SURVEY

Y. Fu and J. Guo et. al., [17] explains using a smartphone to monitor blood cholesterol as some electrochemical analyzer to prevent cardiovascular disease. The first medical smartphone to be used as an electrochemical analyzer for blood lipid monitoring is described in this publication. It is possible to monitor the current produced by the enzymatic reaction with the total cholesterol test strip by integrating an electrochemical analyzer into a smartphone. Through an electrochemical process, the biochemical signal is transformed to an electrical signal using a disposable test strip. The clinical biochemical analyzer is not as accurate as the suggested medical smartphone in evaluating a patient's blood lipid level. Because of its portability, reliability, lower cost, convenience, and internet-based medical data interaction, the suggested medical smartphone system is a promising platform as a point-of-care device for blood total cholesterol (TC) monitoring. This platform can be applied for long-term prevention of cardiovascular disease.

S. K. Jain and B. Bhaumik et. al., [18] explains that to use a smartphone to detect cardiovascular diseases with an energy-efficient ECG signal processor. For the purpose of diagnosing cardiovascular disease on smartphones, an innovative forward-search based ECG signal processing algorithm has been integrated into an Application Specific Integrated Circuit (ASIC). Using the Physionet PTB (Pulmonary Tu-Berculosis) diagnostic ECG database, the ASIC and Android application are validated for the identification of bundle branch block, hypertrophy, arrhythmia, and myocardial infarction. The failure rate in detecting bundle branch block, hypertrophy, arrhythmia, and myocardial infarction is 0.69%, 0.69%, 0.34%, and 1.72%.

G. Joo, Y. Song, H. Im and J. Park et. al., [19] explains that machine learning is being used clinically to predict the occurrence of cardiovascular disease using big data from Korea's nationwide cohort. The features of big data and machine learning (ML) for predicting the risk of CVD were examined and examined using data from the Korean National Health Insurance Service-National Health Sample Cohort (KNHSC). Different measures, including receiver operating characteristic curves, precision-recall curves, sensitivity, specificity, and F1 score, were used to verify a number of ML-based prediction models, which included logistic regression, deep neural networks, random forests, and LightGBM. The use of ML approaches was demonstrated by the experimental results, which indicated that all ML models outperformed the baseline approach for estimating the 10-year CVD risk, which was obtained from the ACC/AHA guidelines.

Mohd Zubir Suboh, Muhyi Yaakop, Mohd Shaiful Aziz Rashid Ali, Mohd Yusoff Mashor, Abdul Rahman Mohd Saad, Mohd Azlan Abu, Mohd Syazwan Md Yid, Aizat Faiz Ramli et. al.,[20] explains a portable electronic stethoscope-based screening tool for heart valve problems. The heart sound signal is segmented, features are extracted, and the system is classified automatically using PC (Personal Computer) hardware platforms. An electronic stethoscope can be used as the system's input, and a multimedia board provided with a single board computer, audio codec, and graphic LCD (Liquid Crystal Display) is utilized to create a portable heart valve disease screening device. 96.3% specificity was recorded by both systems. Compared to the PC platform, which offers sensitivity and accuracy of more than 90%, the portable device only has 77.78% sensitivity and 87.04% accuracy.

Sashikanta Prusty, Srikanta Patnaik, Sujit Kumar Dash et. l., [21] explains the prediction and comparative analysis of coronary heart disease. A common procedure for the prognosis of cardiovascular disease has been suggested. In consider, we examine nine common classifiers for the comparative study and prediction of coronary heart failure in this paper using both machine learning and deep learning technologies. These models are simple to develop and have low computing costs. Additionally, a confusion matrix in the Jupyter notebook is used to test and evaluate different classifiers, producing classification measures including accuracy, f1-score, recall, and precision. The maximum accuracy, precision, and f1-score that the logistic regression classifier can provide are 90.78%, 90.24%, and 91.35%, respectively.

Adari Ramesh, Ceeke Kalappagowda Subbaraya, Ravi Kumar Guralamata Krishnegowda et. al., [22] explains a system for remote health monitoring that uses an advanced artificial intelligence algorithm to predict diabetes and heart disease. By choosing the best features based on the global fitness function, the Unified Levy Modeled Crow Search Optimization (ULMCSO) algorithm helps the classifier become more accurate and requires less training time. Finally, the classification model based on the probabilistic guided naïve distribution (PGND) is applied to predict the label of whether the patient has been diagnosed with a disease or not. Two different datasets, such as PIMA and Hungarian, are used in an evaluation in order to validate and contrast the results of the suggested model using a number of performance measures.

Kehinde Marvelous Adeniyi, Olasunkanmi James Oladapo, Timothy Oluwaseun Araoye, Taiwo Felix Adebayo, Sochima Vincent Egoigwe, Matthew Chinedu Odo et. al., [23] explains Machine learning classifier-based outcome prediction for heart disease. In order to choose the variables for the logistic regression model and determine its sensitivity, specificity, accuracy, and area under the characteristic curve (AUC), this study used the forward, backward, and enter methods. The risk factors connected to heart disease have an accuracy of 87.9% when the logistic regression model is applied with the enter technique at the 5% level of significance. The model from forward, which has an 88.6% weight, was the favored model for the variable selection approach utilized. The analysis model's sensitivity and specificity were 85.6% and 91.4%, respectively. In addition, the misclassification rate was 11.4%; the positive and negative predicted values were 87% and 90.5%, respectively.

Zarkogianni K, Athanasiou M, Thanopoulou AC, et. al., [24] explains the comparison of machine learning methods for determining the possibility that a long-term diabetes complication will be cardiovascular disease. The study aims to explore the use of advanced machine learning techniques to the creation of personalized models capable of predicting the incidence of fatal or nonfatal cardiovascular disease (CVD) in patients with Type 2 diabetes (T2DM). Considering the outputs of both the HWNN- and SOM-based primary models yields the best discrimination performance (Area Under the Curve (AUC): 71.48%). The suggested approach outperforms the Binomial Linear Regression (BLR) model, supporting the requirement for the use of more advanced methods to provide reliable CVD risk scores.

Nitalaksheswara Rao Kolukula, Prathap Nayudu Pothineni, Venkata Murali Krishna Chinta, Venu Gopal Boppana, Rajendra Prasad Kalapala, Soujanya Duvvi et. al., [25] explains that machine learning algorithms are used to predict the presence of cardiac disease and the significance of its features. For identifying cardiac issues, the random forest (RF) classification algorithm offers a more dependable method. Given that this application has a 95% accuracy rate across training data, data analysis is essential. They've discussed about the experiments and results of the RF classifier approach, which raises the diagnostic accuracy of heart disease research.

N. Sabri et al., [26] presents HeartInspect: The Naïve Bayes algorithm is used to predict an individual's risk of heart disease. To address these issues, a Naive Bayes-based heart disease prediction system is presented. The system uses a dataset that contains data on an individual's age group, physical activity level, general health, height, weight, physical health, difficulty walking, and amount of sleep. The dataset is divided into 80/20 for testing and training. To determine the possibility of heart disease, the dataset is analyzed using the Naive Bayes technique. The heart disease prediction system based on the Naive Bayes algorithm presented in this paper shows promise with 71–73% accurate predictions.

3. AN INTELLIGENT CARDIOVASCULAR DISEASE PREDICTION USING KNN

In this section, an intelligent cardiovascular disease prediction using KNN with data mining is presented. The Figure 1 shows the block diagram of presented approach. The datasets related to cardiovascular disease were gathered from the UCI machine learning repository and Kaggle. Thirteen features are covered by the seventy thousand entries in the first dataset. The "cardio" variables in the dataset are used to categorize patients with heart disease on a scale from zero to one, where zero denotes the absence of cardiovascular illness and one indicates those who have the condition. The Cleveland dataset has 76 features, whereas the second dataset was gathered from the Cleveland Clinic Foundation in Switzerland and Long Beach. Only 14 features and 303 records make up the dataset available in the UCI repository. One feature, NUM, is used as the predicted attribute; one indicates that a patient has heart disease, and zero indicates.

Preparing the data is thought to be the most crucial step in using data mining techniques to improve accuracy of predictions. Through the use of cleaned data and the removal of features from the dataset that are either unnecessary or only partially relevant, data preparation is essential in improving prediction performance. This stage involves removing from the dataset any unnecessary, pointless, repetitive, spelled incorrectly, and wildly inaccurate data that neither increases nor decreases the accuracy of the prediction. The databases on cardiovascular disease did not have any missing data records throughout this investigation. The duplicate records in the first dataset were removed even though they had the same information values. Furthermore, features with varying scales across datasets have been rescaled to match. The rescaling procedure used the range 0 to 1. Preprocessing data offers a number of advantages: Removing redundant data reduces the chance of making conclusions based on noise, which in turn reduces overfitting. ii) It reduces the training period because algorithms learn more quickly with less data. All clinical features obtained from various medical examinations made up the residual characteristics.

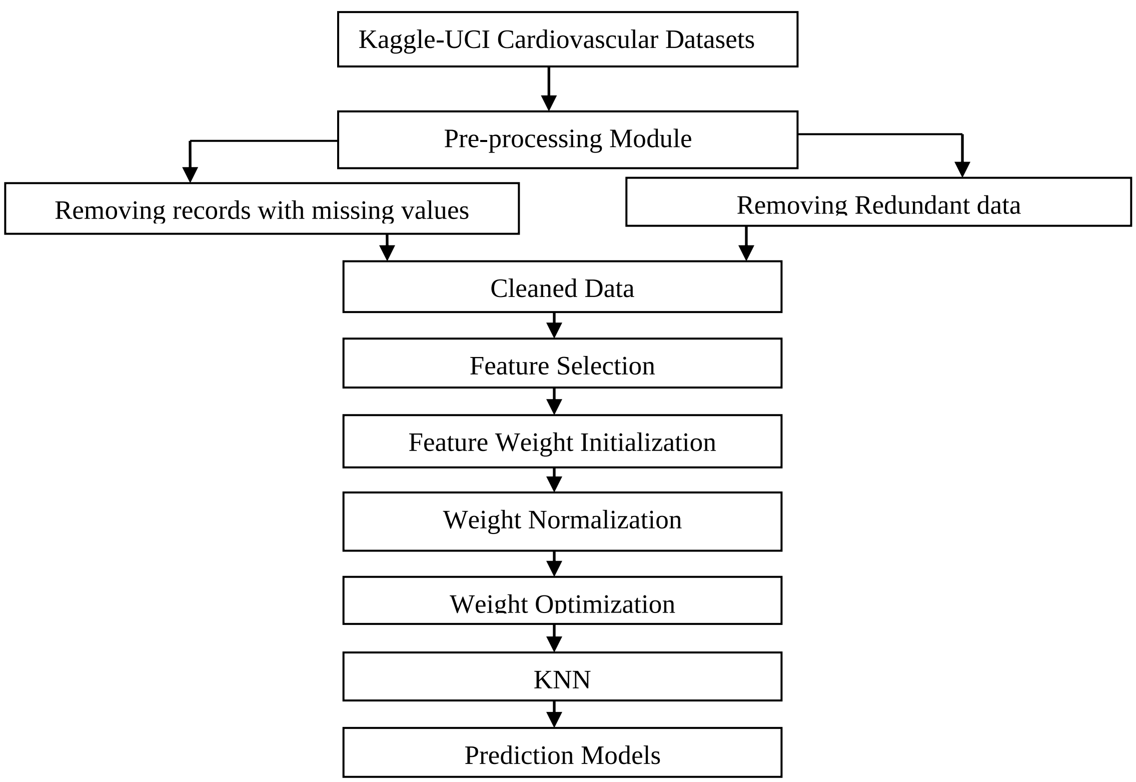


Figure 1. Block diagram of an intelligent cardiovascular disease prediction using KNN with data mining

The selection of features reduces the quantity of related parameters in the dataset that is utilized to evaluate the model. It is ideal to limit the total of features in order to lower the computational modeling expenses; in some cases, this also improves the model's performance. When the in-class labeling is possible and independent factors have different effects on the class label, balancing the features to improve performance is recommended. The feature weight between each feature and class label is determined by the Normalized Mutual Information (NMI). Regarding the class label c, this dataset has two parameters, a and b, for the samples i, j. In this case, the normalized mutual information can only be determined by knowing the class label of the instances. To scale the results between 0, which denotes no mutual information, and 1, which denotes perfect correlation, the mutual information score is computed. The starting weights for the features are estimated by equations (1) and (2).

$$i_w(\propto) = \frac{mi(\alpha,c)}{avg\ (e\ (\beta),e(c))} \quad (1)$$

$$i_w(\beta) = \frac{mi(\beta,c)}{avg(e(\beta),e(c))} \quad (2)$$

The function $i_w()$ is used to evaluate the feature's initial weights based on equations (2) and (3), and the mutual information is denoted by the variable mi. The entropy related to the feature is represented by the variable e. The mean of the numbers is determined using the denominator's avg() method. For a more consistent training dataset distribution, a feature normalization technique is used. Variable values during training comprise the feature set connected with the algorithm, which reduces the loss function. With each iteration, the algorithm grows and shrinks until it reaches the optimal value, either locally or worldwide. In the given approach, the feature values are scaled from 0 to 1 using the min-max normalization.

If the weights of the minority samples were increased, they might become more significant and harder to disregard. In general, it could be challenging to decide effectively without knowing the weight attached to each of these classes. The most common method for figuring out the weight distribution is to look at how frequently the classes appear in the training data. Creating a weight distribution that will enable us to get good performance in the training and testing sets is the aim. The k samples that are categorized as the closest hits and misses would be iterated over a number of rounds by the explainable weight optimization model. Equation (3) provides the best explanation for the weight optimization when utilizing the gradient descent technique with a constant learning rate of $\frac{1}{k\times s}$.

$$O_w = \sum_{i=0}^{k}\sum_{j=0}^{q} w_j\ \rho^r\left(a_j^s, a_{j_{hit}}^{s_i}\right) - \sum_{i=0}^{k}\sum_{j=0}^{q} w_j\ \rho^r(a_j^s, a_{j_{miss}}^{s_i}) \quad (3)$$

The function ρ^r calculates the similarity between the features based on the equation above, and wj indicates the weight that is currently applied to feature j. The value of the j^{th} feature associated with i^{th}, the nearest hit over the instance m, is represented by the variable $a_{j_{hit}}^{s_i}$. In the same manner, the jth feature value linked to the ith nearest miss is represented by the variable $a_{j_{miss}}^{s_i}$. The number of predictors is selected with the target variable's importance in mind using the feature selection wrapper technique. In order to maximize the performance of the model, the predictors are added and deleted using this procedure. The features' Shapley values in the summary plot. In order to get the result, the predictors "age" and "sex" are selected, and the "cholesterol" feature is connected to "age" and "sex."

The process of examining a large batch of unprocessed data to find patterns and extract useful information is known as data mining. In this work, class prediction models are created based on specified features through the use of data mining. The majority of a space's nearest neighbors determines the classification in a KNN. In this instance, the data class is assigned according to which closest points have the greatest number of representatives. Using uniform weights, the KNN generates its value from the closest neighbors. There are class labels connected to every vector in the training dataset. Positive and negative classes occur. The local neighbors are used by K-nearest neighbors to make predictions, and the distance functions are utilized to compare the similarity examples. Equations (4) and (5) explain that the distance formula is generalized in higher dimensions. In one dimension, it is the rounded magnitude of the numerical variance of their coordinates.

$$d(\alpha,\beta) = |\alpha - \beta| \quad (4)$$

$$d(\alpha,\beta) = \sqrt{(\alpha - \beta)^2} \quad (5)$$

The Euclidean distance is calculated in the KNN, and Equation (6) displays the distance between p and q with Cartesian coordinates (m1, n1) and (m2, n2).

$$d(\alpha,\beta) = \sqrt{(\alpha_1 - \beta_1)^2 + (\alpha_2 - \beta_2)^2} \quad (6)$$

The KNN algorithm predicts the CVD presence and absence very effectively and accurately. Equations (7), (8), (9) & (10) express the accuracy, recall, precision, and F1-score, which are used to evaluate the performance of the presented model.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} X100 \quad (7)$$

$$Recall = \frac{TN}{(TN+FN)} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

where False Negative (FN), True Negative (TN), True Positive (TP), and False Positive (FP) are utilized.

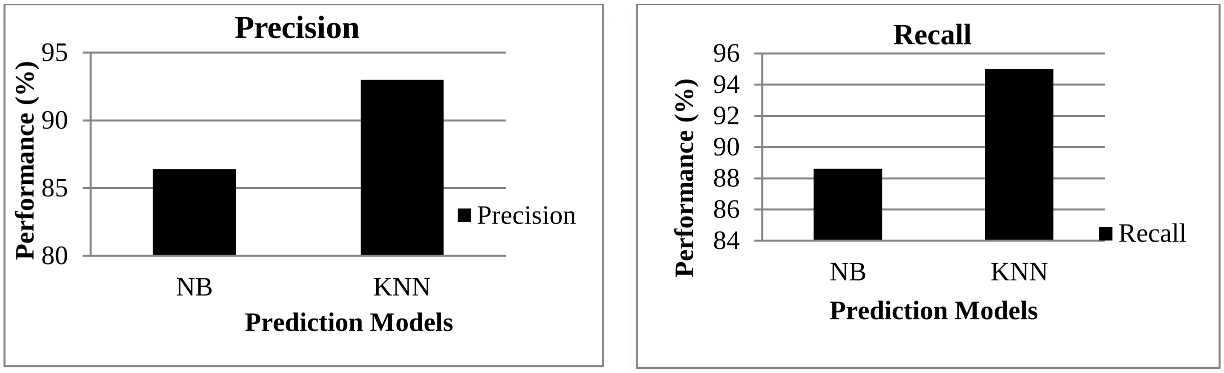
4. RESULTS AND DISCUSSION

In this section, an intelligent cardiovascular disease prediction using KNN with data mining is implemented. The results and discussion of presented model is evaluated here. The accuracy, precision, sensitivity, and F1-score of the proposed method are validated. The Table 1 shows the performance comparison.

Table 1. Performance Comparison

Metrics/Algorithms	Naïve Bayes (NB)	KNN with Data Mining
Precision (%)	86.40	93
Recall (%)	88.6	95
Accuracy (%)	89.2	96.5
F1-score (%)	88	96

Presented KNN model with Data Mining has obtained better performance for CVD prediction than NB algorithm. Figures 2(a) and 2(b) respectively demonstrate the precision and recall comparisons.



(a) (b)

Figure 2. Performance Comparison in terms of (a) Precision and (b) Recall

The performance numbers are displayed on the y-axis in figure 2, while the x-axis displays prediction models. The KNN algorithm with data mining has obtained better precision than NB algorithm. Compared to NB classifier, the KNN has obtained better sensitivity for CVD prediction. The Figure 3 shows Accuracy Performance comparison.

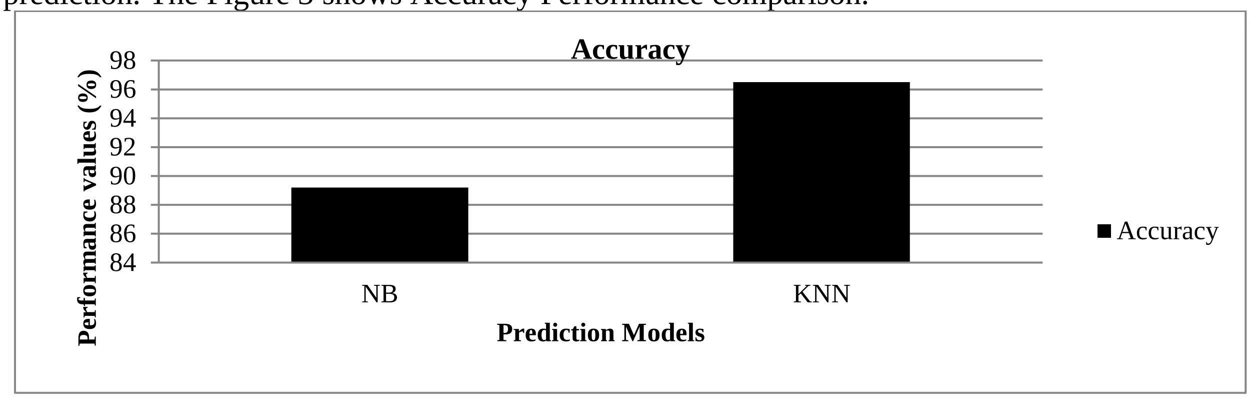


Figure 3. Accuracy Comparison

The KNN with DM has achieved high accuracy than NB for CVD prediction. The Figure 4 shows F1-score comparison.

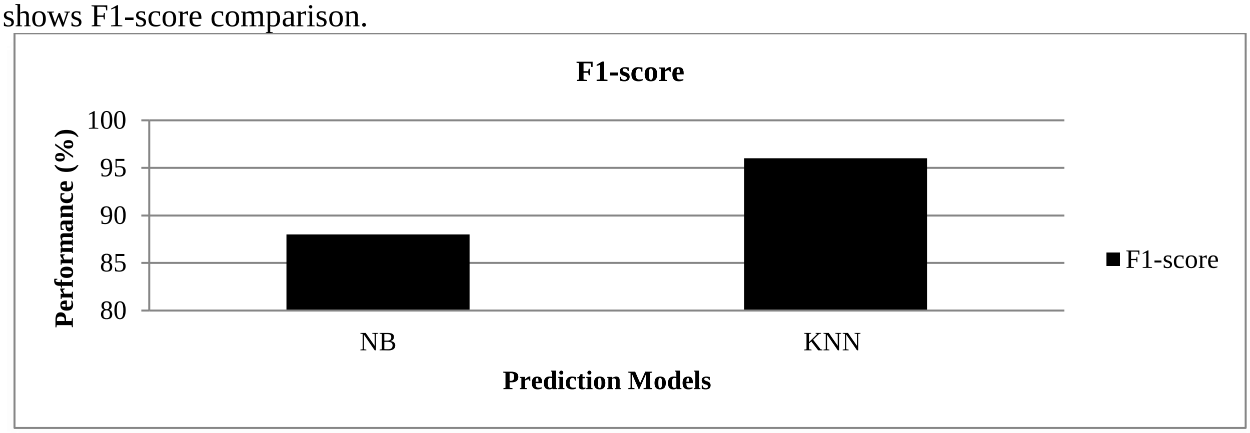


Figure 4. F1-score Comparison

Compared to NB classifier, the KNN has obtained better F1-score. Hence, presented KNN model has shown significant performance for CVD prediction. If CVD is predicted accurately, then effective treatment will be provided. As a result, patient lives will be saved.

5. CONCLUSION

Worldwide, heart diseases are the leading cause of death. Reducing mortality can be achieved through continuous clinical supervision and early detection of cardiac diseases. In this work, an intelligent cardiovascular disease prediction using KNN with data mining is presented. This model is trained and tested on two datasets Kaggle and UCI. The data is pre-processed, null values and missing values are removed to improve the model performance. The described approach uses the min-max normalization to scale the feature values. Data mining combined with KNN is utilized to predict CVD. In order to reduce the fatality rate from cardiovascular disorders, our research creates a model that accurately predicts these conditions. Accuracy, Precision, Recall, and F1-score are used to evaluate the performance of the presented approach. Compared to other algorithms, presented algorithm has shown better performance for the prediction of CVD.

REFERENCES

[1] Manjunathan Alagarsamy, Jemin Vijayaselvan Mariyarose, Nithya Devi Shanmugam, Joseph Michael Jerard Vedam, Mary Dallfin Bruxella Joseph, Kannadhasan Suriyan, “Development of electrocardiogram intelligent and wearable monitoring system-assisting in care,” International Journal of Reconfigurable and Embedded Systems (IJRES), vol. 12, no. 1, pp. 51-59, March 2023, doi: 10.11591/ijres.v12.i1.pp51-59

[2] Abhishek, H. V. Bhagat and M. Singh, "A Machine Learning Model for the Early Prediction of Cardiovascular Disease in Patients," *2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC)*, Puducherry, India, 2023, pp. 1-5, doi: 10.1109/ICACIC59454.2023.10435210.

[3] M. A. Simonyan *et al*., "Spectral Analysis of Photoplethysmography Signal in Patients with Cardiovascular Diseases and Healthy Subjects," *2021 5th Scientific School Dynamics of Complex Networks and their Applications (DCNA)*, Kaliningrad, Russian Federation, 2021, pp. 180-182, doi: 10.1109/DCNA53427.2021.9587125.

[4] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan and S. H. Bouk, "Heart Disease Prediction Using Stacking Model With Balancing Techniques and Dimensionality Reduction," in *IEEE Access*, vol. 11, pp. 116026-116045, 2023, doi: 10.1109/ACCESS.2023.3325681.

[5] Yudi Ramdhani, Cakra Mahendra Putra, Doni Purnama Alamsyah, “Heart failure prediction based on random forest algorithm using genetic algorithm for feature selection,” International Journal of Reconfigurable and Embedded Systems (IJRES), vol. 12, no. 2, pp. 205-214, July 2023, doi: 10.11591/ijres.v12.i2.pp205-214

[6] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim and A. W. Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," in IEEE Access, vol. 9, pp. 106575-106588, 2021, doi: 10.1109/ACCESS.2021.3098688.

[7] N. Ji *et al.*, "Recommendation to Use Wearable-Based mHealth in Closed-Loop Management of Acute Cardiovascular Disease Patients During the COVID-19 Pandemic," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 4, pp. 903-908, April 2021, doi: 10.1109/JBHI.2021.3059883.

[8] C. P. Loizou, E. Kyriacou, M. B. Griffin, A. N. Nicolaides and C. S. Pattichis, "Association of Intima-Media Texture With Prevalence of Clinical Cardiovascular Disease," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 68, no. 9, pp. 3017-3026, Sept. 2021, doi: 10.1109/TUFFC.2021.3081137.

[9] Y. An, N. Huang, X. Chen, F. Wu and J. Wang, "High-Risk Prediction of Cardiovascular Diseases via Attention-Based Deep Neural Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 3, pp. 1093-1105, 1 May-June 2021, doi: 10.1109/TCBB.2019.2935059.

[10] Rana Riad K. AL-Taie, Basma Jumaa Saleh, Ahmed Yousif Falih Saedi, Lamees Abdalhasan Salman, “Analysis of WEKA data mining algorithms Bayes net, random forest, MLP and SMO for heart disease prediction system: A case study in Iraq,” International Journal of Electrical and Computer

Engineering (IJECE), vol. 11, no. 6, pp. 5229-5239 December 2021, doi: 10.11591/ijece.v11i6.pp5229-5239
[11] S. Ghorashi *et al*., "Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases," in *IEEE Access*, vol. 11, pp. 60254-60266, 2023, doi: 10.1109/ACCESS.2023.3286311.
[12] A. Jafar and M. Lee, "HypGB: High Accuracy GB Classifier for Predicting Heart Disease With HyperOpt HPO Framework and LASSO FS Method," in IEEE Access, vol. 11, pp. 138201-138214, 2023, doi: 10.1109/ACCESS.2023.3339225.
[13] Ban Salman Shukur, Maad M. Mijwil, "Involving machine learning techniques in heart disease diagnosis: a performance analysis," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 2, pp. 2177-2185, April 2023, doi: 10.11591/ijece.v13i2.pp2177-2185
[14] R. Chen, F. Miao, J. Zheng, Y. Wu and Y. Li, "Effects of Consecutive Moderately Cold Days on Cardiovascular Disease Mortality in Shenzhen, China: A Preliminary Study," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA, 2018, pp. 1148-1151, doi: 10.1109/EMBC.2018.8513273.
[15] Raid Luaibi Lafta, Mohanad S. AL-Musaylh, Qahtan Makki Shallal, "Clustering similar time series data for the prediction the patients with heart disease," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 2, pp. 947-954, May 2022, doi: 10.11591/ijeecs.v26.i2.pp947-954
[16] Tamilarasi Suresh, Tsehay Admassu Assegie, Subhashni Rajkumar, Napa Komal Kumar, "A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model," International Journal of Electrical and Computer Engineering (IJECE) Vol. 12, No. 2, April 2022, pp. 1831~1838 ISSN: 2088-8708, DOI: 10.11591/ijece.v12i2.pp1831-1838
[17] Y. Fu and J. Guo, "Blood Cholesterol Monitoring With Smartphone as Miniaturized Electrochemical Analyzer for Cardiovascular Disease Prevention," in IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 4, pp. 784-790, Aug. 2018, doi: 10.1109/TBCAS.2018.2845856.
[18] S. K. Jain and B. Bhaumik, "An Energy Efficient ECG Signal Processor Detecting Cardiovascular Diseases on Smartphone," in IEEE Transactions on Biomedical Circuits and Systems, vol. 11, no. 2, pp. 314-323, April 2017, doi: 10.1109/TBCAS.2016.2592382.
[19] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in *IEEE Access*, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.
[20] Mohd Zubir Suboh, Muhyi Yaakop, Mohd Shaiful Aziz Rashid Ali, Mohd Yusoff Mashor, Abdul Rahman Mohd Saad, Mohd Azlan Abu, Mohd Syazwan Md Yid, Aizat Faiz ,"Ramli, "Portable heart valve disease screening device using electronic stethoscope," Indonesian Journal of Electrical Engineering and Computer Science, vol. 15, no. 1, , pp. 122-132, July 2019, doi: 10.11591/ijeecs.v15.i1.pp122-132
[21] Sashikanta Prusty, Srikanta Patnaik, Sujit Kumar Dash, "Comparative analysis and prediction of coronary heart disease," Indonesian Journal of Electrical Engineering and Computer Science, vol. 27, no. 2, pp. 944-953, August 2022, doi: 10.11591/ijeecs.v27.i2.pp944-953.
[22] Adari Ramesh, Ceeke Kalappagowda Subbaraya, Ravi Kumar Guralamata Krishnegowda, "A remote health monitoring framework for heart disease and diabetes prediction using advanced artificial intelligence model," Indonesian Journal of Electrical Engineering and Computer Science, vol. 30, no. 2, pp. 846-859, May 2023, doi: 10.11591/ijeecs.v30.i2.pp846-859
[23] Kehinde Marvelous Adeniyi, Olasunkanmi James Oladapo, Timothy Oluwaseun Araoye, Taiwo Felix Adebayo, Sochima Vincent Egoigwe, Matthew Chinedu Odo, "Prediction of heart disease outcomes using machine learning classifier," Indonesian Journal of Electrical Engineering and Computer Science, vol. 30, no. 2, pp. 917-926, May 2023, doi: 10.11591/ijeecs.v30.i2.pp917-926
[24] Zarkogianni K, Athanasiou M, Thanopoulou AC. Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication," IEEE J Biomed Health Inform, vol. 22, no. 5, pp. 1637-1647, September 2018, doi: 10.1109/JBHI.2017.2765639.

[25] Nitalaksheswara Rao Kolukula, Prathap Nayudu Pothineni, Venkata Murali Krishna Chinta, Venu Gopal Boppana, Rajendra Prasad Kalapala, Soujanya Duvvi, “Predictive analytics of heart disease presence with feature importance based on machine learning algorithms,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 32, no. 2, pp. 1070-1077, November 2023,doi: 10.11591/ijeecs.v32.i2.pp1070-1077
[26] N. Sabri *et al*., "HeartInspect: Heart Disease Prediction of an Individual Using Naïve Bayes Algorithm," 2023 IEEE 11th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2023, pp. 350-354, doi: 10.1109/ICSPC59664.2023.10420149.