A Pharmacy System Integrated with A Machine Learning Algorithm for Cardiovascular Disease Prediction

Kumar Shwetabh [1*], F Rahman [1], Sushree Sasmita Dash [1]

Abstract

Cardiovascular diseases are widely acknowledged as highly challenging and contribute significantly to global fatalities. The widespread use of medications has strained healthcare systems worldwide. In this review, we describes a method, Machine Learning-based Cardio Vascular Disease Prediction (ML-CVDP), designed for accurately predicting cardiovascular illnesses. The model addresses issues like missing values using the mean substitution approach and tackles data imbalances with the Synthetic Minority Over-sampling Technique (SMOTE). It employs an ensemble of Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN) for categorization. The Feature Significance approach is then applied to select features. Experimental results demonstrate the strategy's superiority over single-modal techniques, achieving an accuracy of 93.2%, precision of 94.1%, specificity of 89.6%, sensitivity of 94.1%, Mean Squared Error (MSE) of 6.3%, and Root Mean Squared Error (RMSE) of 4.8%. The True Positive Rate (TPR) is 91.5%, and the False Positive Rate (FPR) is 92.5%.

Keywords: Cardiovascular Disease, Machine Learning, Pharmacy System, Diagnosis

Significance | The review explores an inclusive ML structure, vital for efficient CVD forecasting. The proposed model showcases superior accuracy, emphasizing the need for comprehensive predictive algorithms in cardiovascular health research.

*Correspondence. Kumar Shwetabh, Faculty of CS & IT, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.kumarshwetabh@kalingauniversity.ac.in

Editor Md Shamsuddin Sultan Khan And accepted by the Editorial Board Jan 26, 2024 (received for review Nov 24, 2023)

1. Introduction

The demanding schedule of contemporary times contributes to an unsafe lifestyle, resulting in heightened levels of anxiety and melancholy. To combat these problems, there is a propensity to engage in excessive tobacco use, alcohol consumption, and substance abuse. These factors are the primary catalysts for several hazardous illnesses, especially cardiovascular disorders and cancer. Cardiovascular diseases (CVDs) have the most significant global mortality rates, as stated by the World Health Organization (WHO) Townsend et al. (2022) Campbell et al. (2021). CVDs account for over 31% of global mortality. Timely detection of such disorders is crucial to enable the implementation of preventive measures before the occurrence of severe consequences Ullah et al. (2023).

CVDs are medical conditions that impact the heart or blood vessels. The four primary categories of CVDs encompass Coronary Heart illness, Stroke/Transient ischemic incident, Peripheral arterial illness, and Aortic illness Hsu et al. (2022). The precise etiology of CVD remains unclear. However, several risk factors contribute to the development of these conditions, including hypertension, tobacco use, diabetes, Body Mass Indexes (BMI), cholesterol levels, age, and family history Bays et al. (2021). These parameters vary among individuals. Age, gender, anxiety,

Author Affiliation.
[1] Faculty of CS & IT, Kalinga University, Naya Raipur, Chhattisgarh, India.

Please cite this article.
Kumar Shwetabh, F Rahman, Sushree Sasmita Dash, (2024). A Pharmacy System Integrated with A Machine Learning Algorithm for Cardiovascular Disease Prediction, Journal of Angiotherapy, 8(1), 1-9, 9490

and a poor lifestyle are significant contributing factors to CVDs. The primary obstacle is accurately forecasting these illnesses in time to decrease the death rate through efficient medicine and other preventative measures Mathur Hsu et al. (2020).

This article introduces an approach called Machine Learning Cardiovascular Disease Prediction (ML-CVDP) to accomplish specific goals. It effectively addresses vital variables contributing to major cardiovascular illnesses, including hypertension, familial predisposition, psychological stress, age, gender, cholesterol levels, BMI, and bad habits. Unlike previous studies primarily concentrating on choosing features and standard categorization approaches, ML-CVDP aims to enhance general precision by addressing deficiencies and unbalanced data. The missing data have been handled by substituting them with the average values for a relevant characteristic Azmi Hsu et al. (2022). ML-CVDP suggests employing Synthetic Minority Over-sampling Techniques (SMOTE) to address data imbalance Sinha et al. (2023). After the data has been weighed, ML-CVDP utilizes the feature significance approach to determine the optimal set of features. A combination of Logistic Regression (LR) Yang et al. (2021), K-Nearest Neighbors (KNN) Wazery et al. (2021), Support Vector Machines (SVM) Sabanci et al. (2022), Decision Trees (DT) (Ghiasi, Zendehboudi, 2021), and Artificial Neural Networks (ANN) Imtiaz et al. (2021) is suggested to enhance the accuracy of predictions.

A comprehensive machine learning architecture that simultaneously addresses data balance, feature selection, and categorization to enhance and expedite the forecasting of CVD is suggested. Especially missing values and the information balance are controlled by mean and SMOTE correspondingly. The feature significance approach determines the most compelling feature sets. The enhanced forecasting is attained by combining LR, KNN, SVM, DT, and ANN classifiers into an ensemble Miller Hsu et al. (2020).

The following sections are organized in the given manner: Section 2 focuses on the literature review of CVD and the strategies used for its diagnosis and prediction. Section 3 introduces the ML-CVDP approach, which integrates LR, KNN, SVM, DT, and ANN algorithms to classify and predict CVD. Section 4 provides a detailed discussion of the software analysis and its outcomes. The research is concluded in Section 5, which also explores the potential for future development.

2. Literature Survey and Findings

The research provides a sophisticated healthcare system that utilizes a Swarm-Artificial Neural Network (S-ANN) technique to forecast CVD accurately Nandy et al. (2023). The initial step of the S-ANN technique involves randomly generating a certain number of Neural Networks (NNs). These NNs are then used to train and evaluate the structure, taking into account their solution reliability. The neural network communities undergo two phases of weight adjustments, and their mass is changed using a newly developed heuristic formulation. The neurons' mass is adjusted by distributing the globally optimal weight across other neurons, enabling the prediction of CVD precision Krishna Prasad Hsu et al. (2021).

The objective was to create and assess a two-phase Machine Learning (ML) for predicting the simultaneous presence of Diabetic Mellitus (DM) and CVD Abdalrada et al. (2022). During the initial phase, the research employed two ML methods (LR and Evimp function) inside a multivariate adaptable regression splines framework to identify the critical shared risk variables for DM and CVD. The research utilized a correlation network to eliminate unnecessary duplication of information. The study employed a categorization and regression approach for the second phase to construct the framework.

The research introduced an integrated decision support method that can aid in the timely identification of cardiac disease by utilizing the participant's clinical characteristics Rani et al. (2021). An integrated feature selection approach, which combines the Genetic Approach (GA) with recurrent feature removal, has been employed to choose appropriate features from the given dataset. The research used support vector machines to develop the suggested hybrid structure: naive Bayes and Adaboost classifiers Karunamurthy Hsu et al. (2022).

The proposed CardioHelp utilizes a Convolutional Neural Network (CNN) Deep Learning (DL) technique to predict the likelihood of CVD in a patient accurately Mehmood et al. (2021). The suggested technique focuses on modeling historical information by employing CNN to predict Heart Failure (HF) during its first phase.

The research presents the creation and verification of DL algorithms for the computerized assessment of retinal vessel size in retinal photos Cheung et al. (2021). These models were trained and tested using a wide range of multicultural and multicountry data sets comprising over 70k images. The results strongly support the creation of practical and interpretable DL algorithms that can accurately predict CVD by analyzing the characteristics of retinal arteries in retinal pictures.

The study evaluated the categorization and calibration of current CVD risk prediction methods and devised prediction algorithms using ML techniques Cho et al. (2021). This work showcased the efficacy of ML methods in enhancing CVD risk forecasting in statin-naïve young Korean individuals without heart failure, surpassing the performance of current CVD risk models. The paradigm is readily adaptable for risk evaluation and medical decision-making.

This cross-sectional investigation of CVD was undertaken at the Affiliated Hospital of Guangzhou Medical College from the beginning of 2015 to the end of 2018 Jiang et al. (2021). Data on demographics, vital signs, blood sugar levels, and other triage ratings were gathered. According to the feature significance analysis conducted by XGBoost, the factors that had the most impact on decision-making during triage were blood pressure, heart rate, saturation of oxygen, and aging.

This study aims to improve the accuracy of cardiac disease prediction by employing ensemble learning techniques Gao et al. (2021). The extraction approaches, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA), are used to identify crucial characteristics from the dataset. The findings indicate that the bagged group learning approach has demonstrated the highest efficiency when combined with a decision tree.

Various ML and DL methods were created in the provided study setting to forecast CVDs. Scholars mainly concentrate on feature selection approaches and classifying for classification, disregarding the problem of class imbalance. The issue of imbalanced classes significantly impacts the accuracy of the categorization system. Many characteristics are necessary for accurate prediction when the information is imbalanced. This significantly augments the computational intricacy and renders the method unfeasible for real-world settings.

Enhancing the current feature selection algorithms is necessary to decrease the computational burden while maintaining satisfactory accuracy. Improving the precision of existing classifications is essential to attaining dependable outcomes. In summary, there is a significant requirement for a comprehensive algorithmic learning structure for CVD that can effectively address data balance, optimal choice of features, and systematically enhanced categorization. This improves the accuracy of forecasts for heart disease and decreases the computational complexity. Locating a comprehensive strategy for heart disease in current research and commercial efforts takes time and effort.

3. Proposed Machine Learning-based Cardiovascular Disease Prediction

This research presents a new framework called ML-CVDP for predicting cardiovascular disorders. This study aims to introduce a precise machine-learning model that can effectively and dependably detect CVD by analyzing patients' clinical data. The suggested method is founded upon a series of sequential phases. 1) Data preparation includes eliminating outliers, replacing missing values, and handling imbalanced classes. 2) Feature selection employs the concept of feature significance, and 3) Classifier utilizes an Ensemble of LR, KNN, SVM, DT, and ANN. The process of training the algorithm is carried out, and later, the model that has been prepared will be utilized for making forecasts. The primary objective is to get enhanced outcomes while minimizing the amount of characteristics and computational complexity. The ML-CVDP System begins with the initial step of data preparation. During the information preprocessing stage, the primary focus is examining the information for anomalies. An exception is a sample in a database that significantly departs from the typical performance of the database.

3.1 Architecture of the Model

The model's construction is seen in Figure 1. The necessary procedures are delineated as follows:

The initial phase is extracting coronary artery disease prediction data from the UCI Archive (https://www.nhlbi.nih.gov). The UCI archive is a renowned website or archive for collections of this nature.

- The research utilized two distinct databases sourced from the UCI library.
- Following the information extraction, the subsequent phase in the study involves data pre-processing.

Pre-processing involves normalizing data, which requires selecting a narrow range of values.

- During the pre-processing stage, a selection is made from 75 characteristics (characteristics), with only 15 features being picked. These 15 characteristics are discussed in the preparation of the experiment sub-section.
- Imputing missing values using the NaN value in a Python module.
- Once the user has chosen which rows they want from the database, any contents that are not numeric are substituted with NAN.
- The converted information is now prepared for the training and evaluation of different algorithms or classifications.

Various ML methods are initially used on the information independently and combined for evaluation. Later, the outcome is examined, and the most proficient method is identified.

3.2 Data Pre-Processing

The data on coronary artery disease is subjected to pre-processing once it has been collected from various databases. The database comprises 300 records of patients, of which six entries have missing information. The database has been purged of 5 entries, leaving 290 medical records for pre-processing. The collection has characteristics that are classified into many classes and binary categories. The multi-class factor assesses the presence or lack of heart disease. If an individual has heart illness, the value is assigned as 1; otherwise, it is given as 0, showing no evidence of coronary artery disease in the individual. The information is pre-processed by transforming healthcare records into diagnostic values. The information pre-processing findings for 290 patient files reveal that 130 recordings exhibit a value of 1, showing the existence of coronary heart disease. In contrast, the remaining 150 records display a value of 0, signifying no evidence of heart illness.

3.3 Feature Selection And Reduction

Out of the 15 characteristics in the database, the participant's private data is determined using two variables related to age and sex. The following 12 qualities are deemed significant due to their inclusion of crucial clinical data. Accurate medical records are essential for diagnosing and assessing the extent of cardiac conditions.

3.4 Classification Modelling

Dataset grouping is performed based on the parameters and constraints of DT elements. The classifications are used on each grouped dataset to gauge their efficacy. Those findings determine The most effective algorithms based on their minimal error rate. The accuracy is enhanced by selecting the DT group with a higher error rate and extracting its matching classifier characteristics. The classifier's efficiency is assessed to optimize loss on this information set.

3.4.1 Logistic Regression

LR is a guided learning algorithm applied to regression and categorization issues. LR utilizes likelihood to forecast the categorization of qualitative information. The inputs can be blended linearly using a sigmoid or logistic equation and parameter values to predict the output. The sigmoid shape is employed in maximal likelihood estimating to determine the most plausible information. The resulting probability is bounded by a value ranging from 0 to 1, indicating the possibility of an event occurring. Utilizing the choice criterion transforms the problem into a categorization task. The numerous forms of categories include Binary (with two possible values: 0 or 1), Multinomial (with three or more categories and no specific order), and Ordinal (with three or more categories and a particular order). This approach is easy to carry out and can offer accurate predictions, as shown in Equation (1).

$$p = \frac{1}{1+\exp(-(k_1+k_2X))} \quad (1)$$

The variable p represents the likelihood, k_1 and k_2 are the model parameters, and X is a variable.

3.4.2 K-Nearest Neighbour

KNN is a supervised machine training technique that solves categorization and regression issues. Within the KNN algorithm, when the target value of an information point is absent, the approach involves identifying the k nearest data points in the learning set and assigning it the mean value of those discovered data points. In categorization, the mode of k labels is given or returned, while in regression, the mean of k labels is given back. The method is a fundamental method employed for categorization when prior knowledge of the data is absent. Distance measurements such as Euclidean or Manhattan length can be used to calculate distances and determine the closest data points. It can yield superior outcomes and forecasts, even in the presence of noisy and extensive information.

3.4.3 Support Vector Machine

SVM is a supervised ML technology that uses labeled data to make predictions. In SVM, a hyperplane is constructed to maximize the margin between distinct categories of information or separate comparable details of one type from similar information of another. SVM is utilized for forecasting and categorizing tasks. In forecasting heart attack, various categories can be distinguished by a hyperplane, with one side representing the likelihood of experiencing cardiac illness and the other representing the absence of such possibility.

SVM is categorized into two main types: linear SVM and non-linear SVM. Linear SVM can separate data using a straight line, while non-linear SVM is employed when data cannot be separated using a line. In cases where the data is intricate and linear SVM is insufficient for data separation, the kernel function is employed for non-linear SVM. The kernel's function transforms data from various categories that cannot be segregated linearly into higher degrees. This transformation allows for linear separation of the information in the additional dimension. The equation for the kernel performance is expressed in Equation (2).

$$K(P_x, P_y) = \varphi(P_x)\varphi(P_y) \quad (2)$$

Kernel operations can vary in kinds, including polynomial, straightforward, Radial Basic Function (RBF), and logistic. Kernel operations accept input and transform them into a desired format. Kernel operations can transform a lower-dimensional input field into a higher-dimensional one.

3.4.4 Decision Trees

A decision tree is a guided learning algorithm employed to classify and predict data using Figure 2. The research has demonstrated the fundamental framework of the decision tree. A decision tree is a storage of data that hierarchically represents facts. The process involves calculating the degree of entropy of every characteristic and then splitting the root into sub-trees or branching depending on the characteristic that provides the most significant information gain, following certain constraints. The procedure is executed repeatedly, considering the characteristics, and the leaf component offers the ultimate result, as shown in Equation (3).

$$E(x) = -\sum_{p=0}^{N-1} pr(x_p)\log\left(pr(x_p)\right) \quad (3)$$

Equation (3) calculates the entropy of a category, where n represents the set of categories in X. X refers to the present-day data set being determined, and $pr(x_p)$ represents the proportion of the number of components in category n to the total number of elements in set X. The information gain is shown in Equation (4).

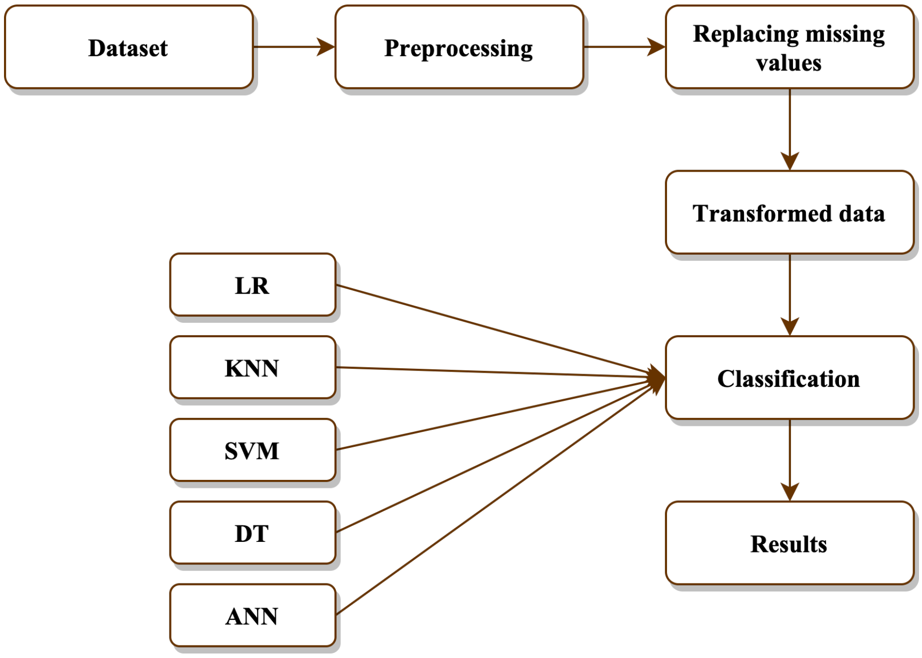

Figure 1. Workflow of the proposed ML-CVDP system

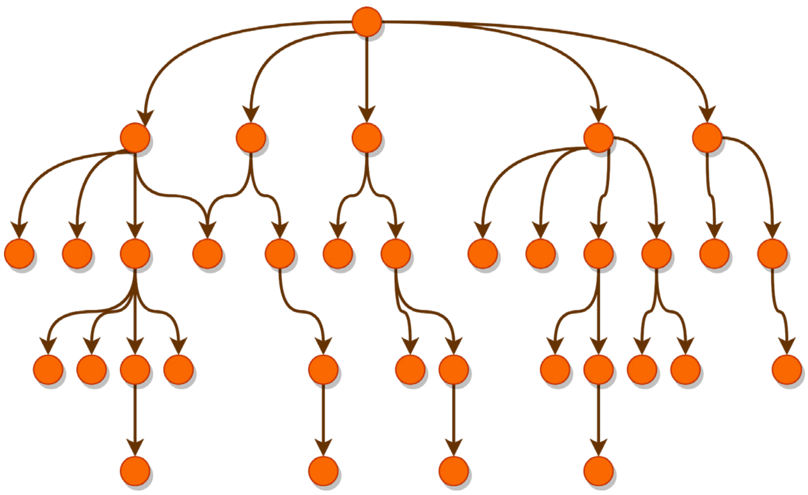

Figure 2. Architecture of the proposed IIP-BCDC

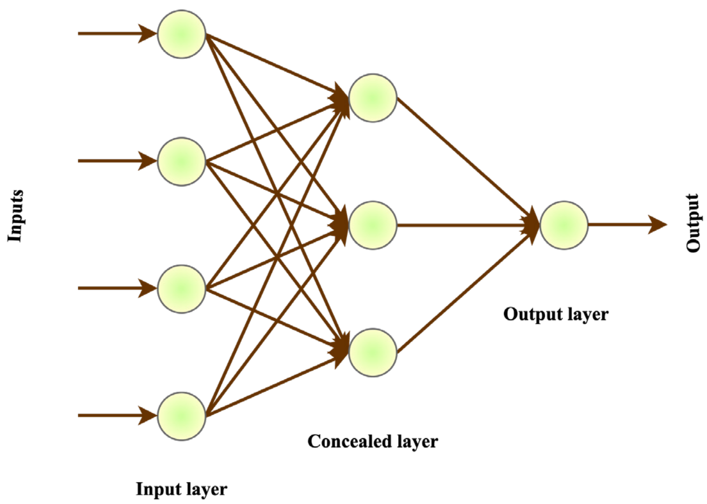

Figure 3. Architectural diagram of ANN

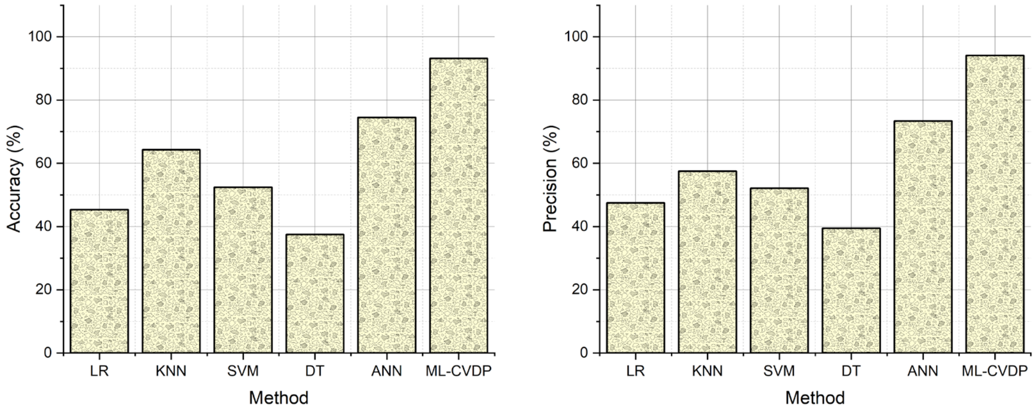

Figure 4. (a) Accuracy and (b) Precision analysis of CVD prediction

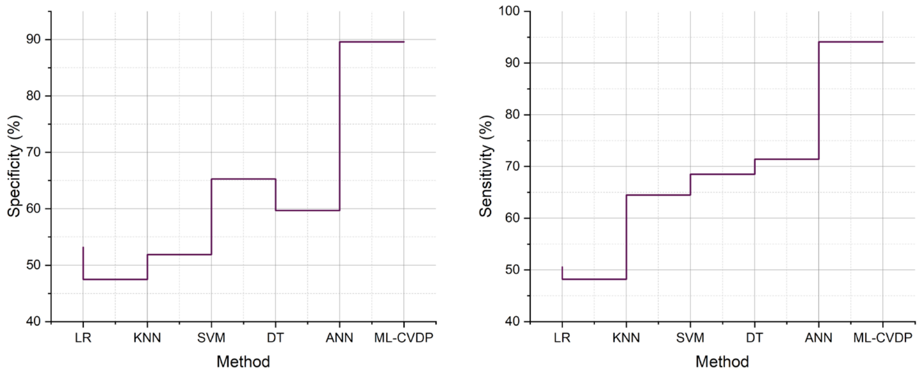

Figure 5. (a) Specificity and (b) Sensitivity analysis of CVD prediction

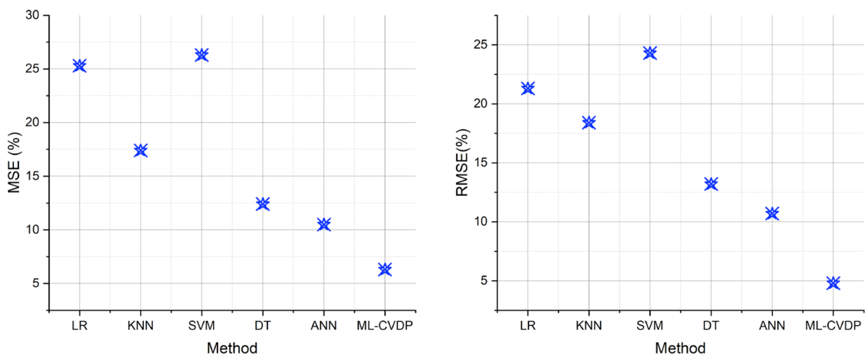

Figure 6. (a) MSE and (b) RMSE analysis of CVD prediction

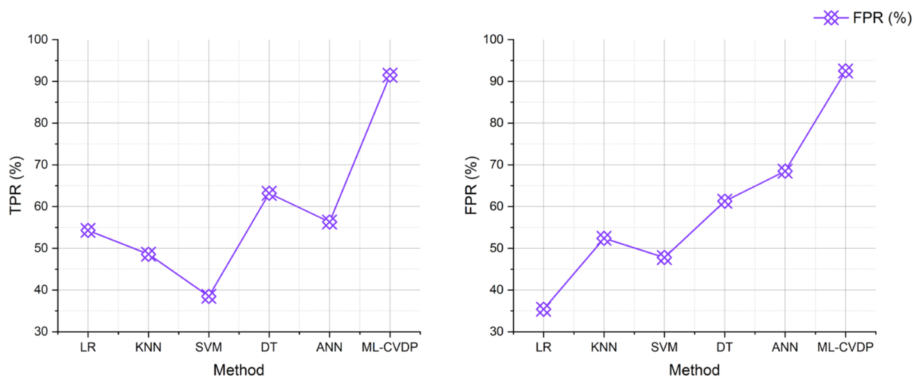

Figure 7. (a) TPR and (b) FPR analysis of CVD prediction

$$IG = C_E - A_E \quad (4)$$

Information gain is denoted as IG, category entropy is represented as C_E, and entropy attributes is abbreviated as A_E. The entropy of every branch is deducted from the group's entropy to calculate the data gain. DT applies to category and continuous information, representing definite information as binary values (0 or 1) or yes/no. It offers exceptional precision by considering and analyzing every feature using a tree-like structure. The comprehension of models and the generation of rules is achieved effortlessly. This approach experiences overfitting, making it challenging to calculate and comprehend if a tree contains several branches.

3.4.5 Artificial Neural Network

ANNs are a component of neural networks within ML. ANNs operate in a manner analogous to the functioning of the human brain. The ANN is modeled after a human neuron cell, which receives data and generates a response. The ANN learns from input and uses this knowledge to classify and foresee outputs. The non-linear statistical design is designed to uncover solutions for difficult problem-solving. An ANN construction comprises three main components: the data input level, one or more hidden levels, and the output level. These layers consist of numerous nodes, likened to neurons in the human mind, as seen in Figure 3.

Neurons interact with each other using nodes in an ANN. These nodes serve as the input for the input layer, which receives information from the external environment and transfers it to the concealed layer. The concealed level discerns the trend by executing computational operations on the information. A neural network can have either a single or several concealed layers. When a neural network has several hidden layers and uses backpropagation, it is called a Multi-Layer Perceptron (MLP). Upon completing all manufacturing, the categorized data is transmitted to the result layer. Activation is employed to transform an input function into a result. It can take on several forms, including linear, sigmoid shape tanh, logistic, and others. ANNs have had a surge in popularity in recent times. They are being applied in many domains like medicine, image identification, recognition of voice, and recognition of faces. However, achieving superior forecasting results with ANNs necessitates the selection of appropriate variables and activation features.

4. Simulation Analysis and Outcomes

This section is subdivided into two segments. The first portion offers an overview of the dataset utilized, while the second details the tools and strategies employed in this study. The research used the Framingham database as the central database to showcase the effectiveness of the architecture (https://www.nhlbi.nih.gov). The dataset was gathered throughout three distinct iterations. The initial round occurred in 1950, during which data was collected from 5200 male and female individuals aged 31 to 64. In 1975, the second stage took place, with the participation of 5125 individuals required to undergo the identical testing procedure.

Figures 4(a) and 4(b) display the accuracy and precision analysis of CVD prediction results obtained from various approaches, including LR, KNN, SVM, DT, ANN, and the suggested ML-CVDP. The proposed method surpassed all previous methods by employing many techniques to anticipate outcomes, achieving an accuracy of 93.2% and a precision of 94.1%. The results are validated using the provided dataset with various samples.

Specificity and sensitivity analyses of the results of CVD prediction derived from various methods are displayed in Figures 5(a) and 5(b), respectively. These methods include LR, KNN, SVM, DT, ANN, and the recommended ML-CVDP technique. With a specificity of 89.6% and a sensitivity of 94.1%, the recommended strategy outperformed all of the ways that had been used in the past. This was accomplished by utilizing various strategies to predict the results in ML-CVDP.

In Figures 6(a) and 6(b), respectively, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) analyses of the results of CVD prediction produced from various approaches (LR, KNN, SVM, DT, ANN, and the ML-CVDP) are presented. These analyses were performed on the findings of the CVD prediction. The proposed technique surpassed the previously utilized methods, with an MSE of 6.3% and an RMSE of 4.8%. They used various approaches to forecast the outcomes, allowing for the completion of this ML-CVDP task.

Figures 7(a) and 7(b) illustrate, respectively, True Positive Rate (TPR) and False Positive Rate (FPR) analyses of the results of CVD prediction generated from various approaches. These analyses are displayed for LR, KNN, SVM, DT, ANN, and the ML-CVDP techniques. This analysis was performed on the findings of CVD prediction. With a TPR of 91.5% and an FPR of 92.5%, the ML-CVDP approach that was developed exceeded all of the methods that had been utilized in the past with its superior performance. When it came to predicting the outcomes of the ML-CVDP, this was achieved by employing a variety of different methodologies.

5. Conclusion and Future Studies

This article introduces the ML-CVDP structure, which aims to predict and diagnose CVD at an early stage. The design consists of four main phases, with the first stage focusing on managing value deficiencies through a mean replacement approach. The information imbalance problem is addressed in the second phase using the Synthetic Minority Over-sampling Technique (SMOTE). During the third step, choosing features is carried out utilizing the approach of feature significance. A combination of Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN) is suggested to enhance the accuracy of the prediction. The ML-CVDP is implemented using the Python programming language and is publicly accessible on the GitHub source. The comparison analysis demonstrates that ML-CVDP surpasses the current leading research by obtaining enhanced accuracy with a smaller collection of characteristics. ML-CVDP exhibits high dependability and is effectively implemented in real-world settings to detect CVD ailments. The proposed ML-CVDP shows the accuracy of 93.2%, the precision of 94.1%, specificity of 89.6%, the sensitivity of 94.1%, MSE of 6.3%, and RMSE of 4.8%. TPR of 91.5% and FPR of 92.5%. The ML-CVDP Platform attains superior precision and is consistently utilized on various databases to forecast CVD. The assessment of ML-CVDP is carried out using cutting-edge databases. Researchers are now collaborating with many hospitals to evaluate the feasibility of using ML for cardiovascular disease prediction in a real-world setting—the research plan to provide the real-time assessment outcomes of ML-CVDP on multiple participants in the future.

Author contribution

K.S., F.R., S.S.D. wrote, reviewed and edited the article. All authors read and approved for publication.

Acknowledgment

The authors are grateful to the Kalinga University to support their study.

Competing financial interests

The authors have no conflict of interest.

References

Abdalrada, A.S., Abawajy, J., Al-Quraishi, T., Islam, S.M.S. (2022). Machine learning models for predicting co-occurrence of diabetes and cardiovascular diseases: a retrospective cohort study. J. Diabetes Metab. Disord. 21(1), 251-261.
https://doi.org/10.1007/s40200-021-00968-z

Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., & Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Medical Engineering & Physics, 105, 103825.
https://doi.org/10.1016/j.medengphy.2022.103825

Bays, H.E., Taub, P.R., Epstein, E., Michos, E.D., Ferraro, R.A., Bailey, A.L., Toth, P.P. (2021). Ten things to know about ten cardiovascular disease risk factors. Am J Prev Cardiol. 5, 100149.
https://doi.org/10.1016/j.ajpc.2021.100149

Campbell, N.R., Ordunez, P., Giraldo, G., Morales, Y.A.R., Lombardi, C., Khan, T., Varghese, C. (2021). WHO HEARTS: a global program to reduce cardiovascular disease burden: experience implementing in the Americas and opportunities in Canada. Can J Cardiol. 37(5), 744-755.
https://doi.org/10.1016/j.cjca.2020.12.004

Cheung, C.Y., Xu, D., Cheng, C.Y., Sabanayagam, C., Tham, Y.C., Yu, M., Wong, T.Y. (2021). A deep-learning system for assessing cardiovascular disease risk via the measurement of retinal-vessel caliber. Nat. Biomed. Eng. 5(6), 498-508.
https://doi.org/10.1038/s41551-020-00626-4

Cho, S.Y., Kim, S.H., Kang, S.H., Lee, K.J., Choi, D., Kang, S., Chae, I.H. (2021). Pre-existing and machine learning-based models for cardiovascular risk prediction. Sci. Rep. 11(1), 8886.
https://doi.org/10.1038/s41598-021-88257-w

Gao, X.Y., Amin Ali, A., Shaban Hassan, H., Anwar, E.M. (2021). Improving the accuracy for analyzing heart disease prediction based on the ensemble method. Complex. 2021, 1-10.
https://doi.org/10.1155/2021/6663455

Ghiasi, M.M., Zendehboudi, S. (2021). Application of decision tree-based ensemble learning in the classification of breast cancer. Comput. Biol. Med. 128, 104089.
https://doi.org/10.1016/j.compbiomed.2020.104089

Hsu, W., Warren, J. R., & Riddle, P. J. (2022). Medication adherence prediction through temporal modelling in cardiovascular disease management. BMC Medical Informatics and Decision Making, 22(1), 1-21.
https://doi.org/10.1186/s12911-022-02052-9

https://www.nhlbi.nih.gov/science/framingham-heart-study-fhs

Imtiaz, S.I., ur Rehman, S., Javed, A.R., Jalil, Z., Liu, X., Alnumay, W.S. (2021). DeepAMD: Detection and identification of Android malware using high-efficient Deep Artificial Neural Network. Future Gener Comput Syst. 115, 844-856.
https://doi.org/10.1016/j.future.2020.10.008

Jiang, H., Mao, H., Lu, H., Lin, P., Garry, W., Lu, H., Chen, X. (2021). Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. Int. J. Med. Inform. 145, 104326.
https://doi.org/10.1016/j.ijmedinf.2020.104326

Karunamurthy, A., Kulunthan, K., Dhivya, P., & Vickson, A. V. S. (2022). A Knowledge Discovery Based System Predicting Modelling for Heart Disease with Machine Learning. Quing: International Journal of Innovative Research in Science and Engineering, 1(1), 14-22.
https://doi.org/10.54368/qijirse.1.1.0005

Krishna Prasad, K., Aithal, P. S., Bappalige, N. N., & Soumya, S. (2021). An Integration of Cardiovascular Event Data and Machine Learning Models for Cardiac Arrest Predictions. International Journal of Health Sciences and Pharmacy (IJHSP), 5(1), 55-54.
https://doi.org/10.47992/IJHSP.2581.6411.0061

Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. Clinical Medicine Insights: Cardiology, 14, 1179546820927404.
https://doi.org/10.1177/1179546820927404

Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. Arab J Sci Eng. 46(4), 3409-3422.
https://doi.org/10.1007/s13369-020-05105-1

Miller, D. D. (2020). Machine intelligence in cardiovascular medicine. Cardiology in Review, 28(2), 53-64.
https://doi.org/10.1097/CRD.0000000000000294

Nandy, S., Adhikari, M., Balasubramanian, V., Menon, V.G., Li, X., Zakarya, M. (2023). An intelligent heart disease prediction system based on a swarm-artificial neural network. Neural. Comput. Appl. 35(20), 14723-14737.
https://doi.org/10.1007/s00521-021-06124-1

Rani, P., Kumar, R., Ahmed, N.M.S., Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. J. Reliab. Intell. Environ. 7(3), 263-275.
https://doi.org/10.1007/s40860-021-00133-6

Sabanci, K., Aslan, M.F., Ropelewska, E., Unlersen, M.F. (2022). A convolutional neural network-based comparative study for pepper seed classification: Analysis of selected deep features with support vector machine. J. Food Process Eng. 45(6), e13955.
https://doi.org/10.1111/jfpe.13955

Sinha, N., Kumar, M.G., Joshi, A.M., Cenkeramaddi, L.R. (2023). DASMcC: Data Augmented SMOTE Multi-class Classifier for prediction of Cardiovascular Diseases using time series features. IEEE Access.
https://doi.org/10.1109/ACCESS.2023.3325705

Townsend, N., Kazakiewicz, D., Lucy Wright, F., Timmis, A., Huculeci, R., Torbica, A., Vardas, P. (2022). Epidemiology of cardiovascular disease in Europe. Nat. Rev. Cardiol. 19(2), 133-143.
https://doi.org/10.1038/s41569-021-00607-3

Ullah, M., Hamayun, S., Wahab, A., Khan, S.U., Rehman, M.U., Haq, Z.U., Naeem, M. (2023). Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective. Curr Probl Cardiol. 48(11), 101922.
https://doi.org/10.1016/j.cpcardiol.2023.101922

Wazery, Y.M., Saber, E., Houssein, E.H., Ali, A.A., Amer, E. (2021). An efficient slime mould algorithm combined with k-nearest neighbor for medical classification tasks. IEEE Access. 9, 113666-113682.
https://doi.org/10.1109/ACCESS.2021.3105485

Yang, Z., Wang, W., Huang, P., Gao, G., Wu, X., Zhang, F. (2021). Local median based linear regression classification for biometric recognition. Comput. Electr. Eng. 96, 107509.
https://doi.org/10.1016/j.compeleceng.2021.107509