



OPEN ACCESS

EDITED BY

Balu Kamaraj,
Imam Abdulrahman Bin Faisal University,
Saudi Arabia

REVIEWED BY

Udhaya Kumar S.,
Baylor College of Medicine,
United States
Karthick Vasudevan,
Reva University,
India

*CORRESPONDENCE

Rohini Karunakaran
✉ rohini@aimst.edu.my

SPECIALTY SECTION

This article was submitted to
Precision Medicine,
a section of the journal
Frontiers in Medicine

RECEIVED 25 January 2023

ACCEPTED 09 March 2023

PUBLISHED 17 April 2023

CITATION

Subramani S, Varshney N, Anand MV,
Soudagar MM, Al-keridis LA, Upadhyay TK,
Alshammari N, Saeed M, Subramanian K,
Anbarasu K and Rohini K (2023) Cardiovascular
diseases prediction by machine learning
incorporation with deep learning.
Front. Med. 10:1150933.
doi: 10.3389/fmed.2023.1150933

COPYRIGHT

© 2023 Subramani, Varshney, Anand, Soudagar,
Al-keridis, Upadhyay, Alshammari, Saeed,
Subramanian, Anbarasu and Rohini. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Cardiovascular diseases prediction by machine learning incorporation with deep learning

Sivakannan Subramani¹, Neeraj Varshney², M. Vijay Anand³,
Manzoore Elahi M. Soudagar⁴, Lamya Ahmed Al-keridis⁵,
Tarun Kumar Upadhyay⁶, Nawaf Alshammari⁷, Mohd Saeed⁷,
Kumaran Subramanian⁸, Krishnan Anbarasu⁹ and
Karunakaran Rohini^{10,11,12*}

¹Department of Advanced Computing, St. Joseph's University, Bengaluru, Karnataka, India, ²Department of Computer Engineering and Applications, GLA University, Mathura, Uttar Pradesh, India, ³Department of Mechanical Engineering, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India, ⁴Department of VLSI Microelectronics, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India, ⁵Faculty of Science, Princess Norah Bint Abdulrahman University, Riyadh, Saudi Arabia, ⁶Department of Biotechnology, Parul Institute of Applied Sciences and Centre of Research for Development, Parul University, Vadodara, India, ⁷Department of Biology, College of Science, University of Hail, Hail, Saudi Arabia, ⁸Centre for Drug Discovery and Development, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India, ⁹Department of Bioinformatics, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India, ¹⁰Unit of Biochemistry, Centre of Excellence for Biomaterials Engineering, Faculty of Medicine, AIMST University, Semeleing, Bedong, Malaysia, ¹¹Centre for Excellence for Biomaterials Science AIMST University, Semeling, Bedong, Malaysia, ¹²Department of Computational Biology, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India

It is yet unknown what causes cardiovascular disease (CVD), but we do know that it is associated with a high risk of death, as well as severe morbidity and disability. There is an urgent need for AI-based technologies that are able to promptly and reliably predict the future outcomes of individuals who have cardiovascular disease. The Internet of Things (IoT) is serving as a driving force behind the development of CVD prediction. In order to analyse and make predictions based on the data that IoT devices receive, machine learning (ML) is used. Traditional machine learning algorithms are unable to take differences in the data into account and have a low level of accuracy in their model predictions. This research presents a collection of machine learning models that can be used to address this problem. These models take into account the data observation mechanisms and training procedures of a number of different algorithms. In order to verify the efficacy of our strategy, we combined the Heart Dataset with other classification models. The proposed method provides nearly 96 percent of accuracy result than other existing methods and the complete analysis over several metrics has been analysed and provided. Research in the field of deep learning will benefit from additional data from a large number of medical institutions, which may be used for the development of artificial neural network structures.

KEYWORDS

cardiovascular disease, AI-based technologies, internet of things, machine learning, computational method

1. Introduction

Cardiovascular disease (CVD), which is the leading cause of death globally, has become a significant problem in public health all over the world. As a result, patients, their families, and the governments of these countries have incurred substantial socioeconomic expenses. Patients at high risk for CVD can be identified by prediction models that use risk stratification. After that, measures that are tailored to this group, such as dietary changes and the use of statins, can help reduce that risk and contribute to the primary prevention of CVD (1).

Several guidelines for the evaluation and management of CVD have suggested using predictive models as a means of identifying patients at high risk and assisting with clinical decision-making. The Pooled Cohort Equations and the Framingham CV risk equation⁶, for example, have both been subjected to independent evaluations in a variety of populations; however, the findings indicated that both of these equations were only weakly discriminating and had a poor level of calibration (2).

As a direct consequence of this, the predictive power of the majority of the models that are now in use is restricted, and there is room for advancement. For instance, the assumption of linearity is necessary for logistic regression, while the assumption of predictor independence is necessary for the Cox proportional hazard model (3).

In the area of study pertaining to the cardiovascular system, machine learning (ML) algorithms have been demonstrated to be extremely helpful predictors. They are more adept than standard statistical models at capturing the complex interactions and nonlinear linkages that exist between the variables and the results (4). Several different investigations (5–15) came to the conclusion that random forests (RF) and support vector machines (SVM) performed better than traditional models.

Cardiovascular diseases such as coronary artery disease (CAD), atrial fibrillation (AF), and other cardiac or vascular ailments continue to be the leading cause of death in the world (10). As people living standards improve and their stress levels continue to rise, the number of people who suffer from CVD is growing at an alarming rate.

According to the most recent estimations (16, 17), CVD will be responsible for the deaths of about 23 million people by the year 2030. Infarction of the myocardium, atrial fibrillation, and heart failure are all instances of different types of CVD (18, 19). The incidence of cardiovascular disease can be influenced by a number of factors, including racial or ethnic background, age, gender, body mass index (BMI), height, and length of torso, as well as the outcomes of blood tests that evaluate factors such as renal function, liver function, and cholesterol levels (20, 21) which is shown in Figure 1.

The development of a wide variety of health problems can be influenced by the complex interactions that take place between these factors. Standard statistical approaches are incapable of accounting for all of the intricate causal links that exist between risk factors because there are so many of them (22, 23). In this day and age of big data, the Internet of Things (IoT) has been shown to be of critical importance. It has made it possible for patients to use smart drugs and smart bracelets to monitor and collect accurate data during a pandemic (24).

Researchers are employing artificial intelligence (AI) in an effort to mine new medical information that can be used by clinicians to better understand the symptoms of various diseases and, as a result,

make more informed decisions for patients (25). This comes as the prevalence of data from the internet of things (IoT) grows within healthcare systems. In order to investigate previously unknown risk factors, current efforts to standardise medical data, and efforts to organise national health screening data (26–28), we will first standardise medical data. These risk variables may have a correlation with the occurrence of the disease, which means that they could offer insights into the mechanisms underlying the disease. Furthermore, accurate disease incidence prediction models necessitate the analysis of large amounts of data (29, 30). The use of artificial intelligence (AI) and massive amounts of data in the prediction of CVD models is becoming increasingly common.

The main contribution and novelty of this research is mentioned below:

- To extract a total of 11 distinct characteristics from the dataset.
- After that, we started by normalising the data and then proceeded to divide the Heart dataset into training and testing sets using an 8:2 split.
- Afterwards, the incorporated GBDT is utilised in the SHAP method for the purpose of selecting features.
- It helps to construct a stacking model consisting of a base learner layer in addition to a meta learner layer.
- Finally, we will achieve the results over several performance metrics analyses and method in the stacking model.

2. Background

Weng et al. (31) tested four different models using clinical data from over 300,000 homes in the United Kingdom. According to the findings, NN was the method that produced the most accurate CVD prediction results for the larger amount of data that were analysed.

The three traditional machine learning models that were tested and evaluated by Dimopoulos et al. (32) based on ATTICA data with 2020 samples for the little CVD dataset were K-Nearest Neighbour (KNN), Random Forest (RF), and Decision Tree. When compared, RF was shown to have produced the best results by using the HellenicSCORE tool, which is a calibration of the ESC Score.

In view of the growing popularity of machine learning techniques in IoT applications, Mohan et al. (15) have proposed a hybrid HRFLM strategy as a means of further improving the accuracy of the model predictions in light of the aforementioned popularity of machine learning methods.

An IoT-ML method was investigated by Akash et al. (33) with the goal of predicting the condition of the cardiovascular system in the human body. The algorithm model uses machine learning (ML) techniques to compute and predict the patient cardiovascular health after it has obtained essential data from the human body. This data include the patient heart rate, ECG signal, and cholesterol.

Within the framework of Yang et al. (34) examination of local locations with separate prediction models, LR was utilised to evaluate 30 cardiovascular disease-related characteristics utilising more than 200,000 high-risk participants in eastern China. The results of the experiments led to the development of an RF model that is more suited to eastern China.

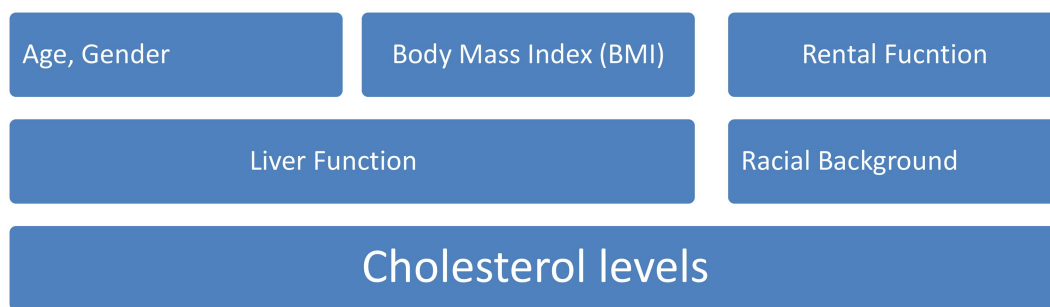


FIGURE 1
Several factor influencing incidence in cardiovascular disease.

For the first time in the study of CVDs, the idea of a stacking model was presented for the very first time by Yang et al. (35). The data on air pollution and weather were considered in order to have a better understanding of how the stacking model influences the daily hospitalisation rate for CVDs. In order to assist in the construction of the stacking model, a grassroots level of five basic learners was first constructed.

During this period, digital, fully automated ecosystems as well as cyber-physical systems are fast growing and finding applications all over the world. The creation of smart healthcare, which offers tools and processes for early diagnosis of life-threatening disorders, is one example of the innovative concepts and technical compositions that are being implemented in nearly every business. As the fourth industrial revolution moves towards a society that is more technologically advanced, there is an urgent requirement for additional research into CVD Zheng et al. (36).

3. Proposed method

The first thing that needs to be done is to combine the data from the Heart Dataset, which already contains information from Cleveland, Hungarian, and Swizerlang, as well as data from Long Beach VA and Stalog (Heart). From the five sources, we extracted a total of 11 distinct characteristics. After that, we started by normalising the data and then proceeded to divide the Heart dataset into training and testing sets using an 8:2 split. Afterwards, the incorporated GBDT is utilised in the SHAP method for the purpose of selecting features.

In the following stage, we will construct a stacking model consisting of a base learner layer in addition to a meta learner layer. The study uses RF, LR, MLP, ET, and CatBoost classifiers to serve as our base learners. LR is utilised in the role of the meta learner. Finally, the suggested stacking model is assessed with regard to its accuracy, precision, recall, F1 score, and area under the curve (AUC). In order to evaluate the model adaptability to new contexts, we made use of a publicly available dataset known as the Heart Attack Dataset.

The Cleveland, Hungarian, Swizerlang, Long Beach VA, and Stalog (Heart) datasets, together with others from the machine learning repository at the University of California, Irvine (UCI), were combined to form the Heart Dataset. We began with a total of 1,190 samples, and after deleting 272 duplicates, we were left with 918 unique sample datasets. We started with 1,190 samples. The whole Heart dataset is displayed in Table 1, and it consists of 11 features that

were taken from five different datasets that contained significant relevant features.

3.1. Feature select and analysis

It is feasible to increase model performance and save a considerable amount of runtime by selecting the ideal subset of features that have a significant impact on the prediction outcomes. This process is referred to as feature selection, and it is possible to accomplish both of these goals.

The three most common methods for picking characteristics are called filters, wrappers, and embedding. The research we conducted utilised the embedded approach known as GBDT as a means of selecting feature variables. This was due to the fact that embedded techniques offer superior prediction performance compared to filter methods and are noticeably quicker than wrapper methods.

GBDT makes use of an additive model and a forward stepwise algorithm in order to achieve learning. These two components work together to accomplish this. For non-leaf nodes, the significance of the features increases proportionately with the magnitude of the reduction in weighted impurity that occurs during splitting.

Because of this, it is not possible to provide a detailed explanation of the role that each attribute plays in determining the overall accuracy of the predictions made by the integrated GBDT. In order to find a solution to this issue, we make use of a technique known as feature imputation, in which the explanatory model is a linear function of the values produced by feature imputation.

$$l(z' = \emptyset_0 + \sum N_i = 1 \emptyset_i Z'_i) \quad (1)$$

where N —features; \emptyset_i —feature attribute value, and Z'_i —feature is valid or not.

The Φ_i value of Equation (1) can be determined by employing a tree-valued estimate methodology (also known as the SHAP method), which is founded on the concepts of game theory and used as the feature attribute values. Below is a formulation for a model f and a set S of non-zero Z' indices, as well as the conventional spherically valued attribute \emptyset_i for each feature.

$$\emptyset_i = \sum S \in M \{i\} | S | (N - |S| - 1)! N! [f(S \cup \{i\}) - f(S)] \quad (2)$$

TABLE 1 Heart dataset features.

Feature	Detailed Information
Age	Age of the patient
Sex	Sex of the patient (Male: 0 or female: 1)
Chest pain type	Four chest pain types
	• ATA: atypical angina
	• TA: typical angina
	• ASY: asymptomatic
	• NAP: non-angina
Resting BP	Value of blood pressure during fasting (Unit mm hg)
Cholesterol	Concentration of serum cholesterol (Unit mm/dL)
Fasting BS	Value of blood glucose during fasting (1: blood glucose >120 mg/dL, 0: other)
Resting ECG	Resting electrocardiogram
Max HR	maximum heart rate
Exercise angina	Presence of exercise angina
Old peak	ST value decision
ST_Slope	Slope of ST section at the movement peak (up, flat, and down)

where M —input feature set.

It is essential to keep in mind that the SHAP strategy is adapted to the specific context and tailored to individual needs. In contrast to the tree model gain, this method can produce consistent results for global feature attributes. This is an advantage over the tree model gain. In the course of our study, we make use of the SHAP methodology in order to isolate and assess several individual characteristics.

In addition to this, we investigate the ways in which various characteristics interact with one another in order to improve our ability to predict outcomes. We differentiate between the contributions of individual features and the contributions of feature interactions by referring to the former as individual feature contribution and the latter as joint feature contribution $\Phi_{i,j}$. In the same way as the value, the Shapley interaction index is calculated using a formula, and the joint feature contribution i and j can be found by doing the calculation as follows.

$$\Phi_{i,j} = \sum_{S \in M \setminus \{i,j\}} |S|! (N - |S| - Z)! Z (N - 1)! \nabla_{i,j}(S) \quad (3)$$

When $i \neq j$:

$$\nabla_{i,j}(S) = f(S \cup \{i,j\}) - f(S \cup \{i\}) - f(S \cup \{j\}) + f(S) \quad (4)$$

where Z represents the indices. i, j represent the feature contributions. S represents the Shapley interaction Index.

Equations (3) and (4) in the GBDT model quantify the twinning relationships between joint features. So, when judging the model, you can get a good idea of how the different factors that interact with each other contribute together.

3.2. Model building

To the extent that the model predictions are accurate, each individual in the base population has a stronger capacity for learning, and the degree of correlation between them decreases. When the individual learners are already more accurate, a fusion of models will be more successful if the individual learners come from a diverse range of backgrounds. This is the foundation upon which the concept of error-ambiguity decomposition is built.

This suggests that when picking the foundation learners for our organisation, we should take into account the performance of individual learners while also taking into account the distinctiveness of each individual learner. Theoretically, it is conceivable to build layers of the stacking model indefinitely as long as their fundamental classifier is operational. This, of course, results in an increase in the level of complexity represented by the model.

To ensure accuracy while reducing the level of complexity exhibited by the model, we solely employ the stacking model, which is comprised of a two-tiered structure consisting of base learners and meta-learners. As a direct consequence of this, SVM, KNN, LR, and ET were decided upon as the possible models for base learners to utilise in the prediction of CVDs. XGBoost, LightGBM, CatBoost, and MLP were some of the other options that were thought about. Following the selection of the most reliable models as the foundation for our education, we restricted the pool of potential candidates to five people who exemplified a comprehensive representation of the population as a whole. The optuna framework was used in order to get the optimal values for the model parameters.

After running a 5-fold CV, this model may generate a large number of features. 5-fold CV is a technique that is frequently utilised in the first layer of a stacking framework to collect input features for the second layer. This paper employs linear regression (LR) as the

classifier for the fusion model predictions since generalised linear regression, also known as GLR, has historically been employed in the second layer of the stacking architecture. Because adjusting the complexity of the output layer of a neural network does not require the employment of more complex functions, this example makes use of functions that are simpler in nature.

The primary learners are the LR, RF, DT, MLP, and CatBoost protocols. At the beginning, we give the training sets eight times as many points as the testing sets. Within the training package that we provide for each of the five foundational learners, we utilise a 5-fold CV. A single base learner is capable of producing five predictions, and each of these five predictions is arranged in a vertical column within a one-dimensional matrix. It is possible that the second stage of training will be based on a five-dimensional matrix that has been developed using the data of five different learners as its foundation.

When applied to the testing set, the 5-fold CV model is utilised once more to make predictions about our initial testing set, which results in the production of five predictions once more. The base learners can be concatenated into a matrix for the second iteration. We use LR on the meta-learner so that it does not become too good at its job. In the second step of the process, we use these predictions to build the final results.

4. Results and discussion

The outcomes of the experiments will be discussed here in order to illustrate the benefits of the stacking paradigm that was recommended by us. Python version 3.9.7 was used throughout each and every test. In this investigation, the sklearn 1.0.2 toolbox is used for model prediction. The SHAP 40.0 toolbox is used for feature selection, and the Optuna 2.10.0 framework is used to determine the optimum values for the model parameters which is shown in Table 2. We executed 10 splits of the data set using various random seeds in order to account for the small sample size of this study and the aforementioned randomisation. After doing so, we averaged the results of all 10 experiments.

Before we started the feature selection process, our dataset contained a total of 11 features. Using the Tree SHAP approach, you are able to determine the contribution value that corresponds to each feature that is contained inside the sample dataset. The ranking of the feature contributions is determined by using the average SHAP value for all of the samples. In accordance with the GBDT model, the contributions of the global features are depicted. The ST Slope and Chest Pain Type have a significant influence on the patient condition (CVD) in patients with cardiovascular disease. In order to cut the model operating time even more, some features that aren't necessary will have to be eliminated. We chose to adopt a cutoff of 0.02, which led to the elimination of the Resting ECG characteristic

while permitting the retention of the other 10 features. We used the AUC to evaluate the performance both before and after the feature selection process. Even though the AUC values of GBDT went down, the drop was not substantial at all, and there was not any difference that could be considered statistically significant by performing metrics such as AUC, Threshold, Sensitivity, Specificity which is shown in Figures 2–5.

The incidence of CVD was quite low in this experiment, resulting in poor PPV and NPV values for each of the seven different ML models. Because of this, their therapeutic value may suffer as a result of an increase in the number of false-positive results. The probabilities that were predicted by each machine learning model were unique, and the risk distribution for LR was comparable to that of SVM. Patients who had a CVD episode had estimated risks that were greater, across all ML models, than those patients who had not had a CVD episode. The plots also demonstrated that all ML models overestimated the risks of those individuals who had not experienced any CVD events. This finding suggests that this variable may also affect how well a model predicts what will happen.

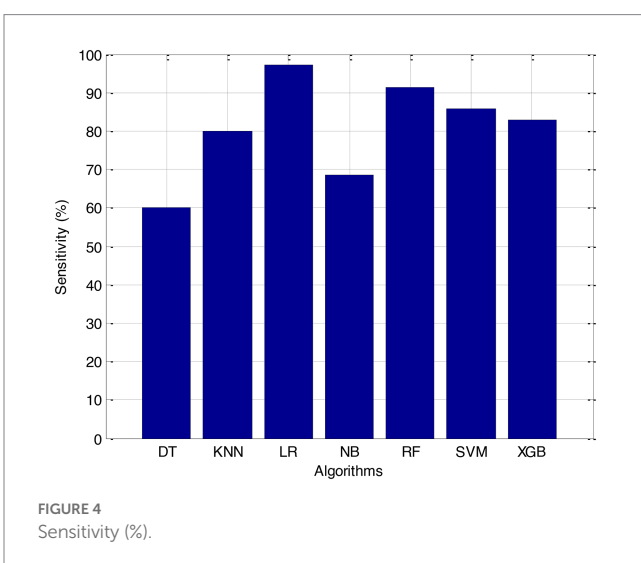
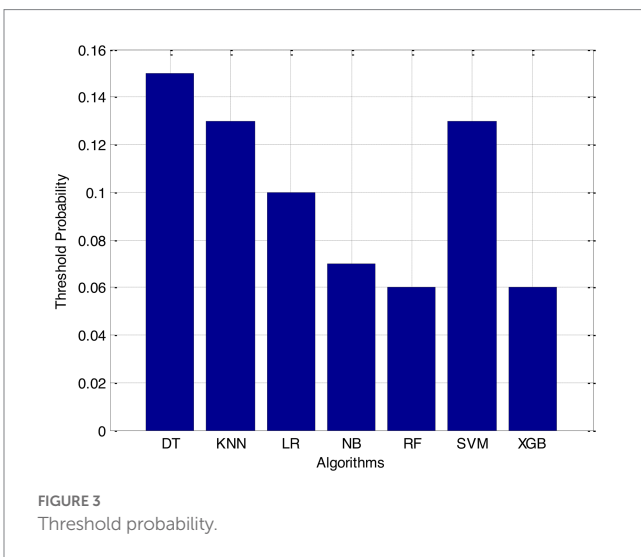
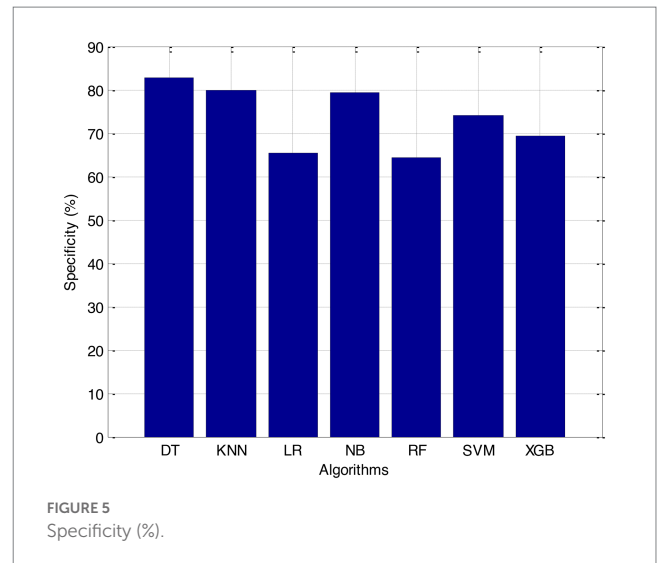
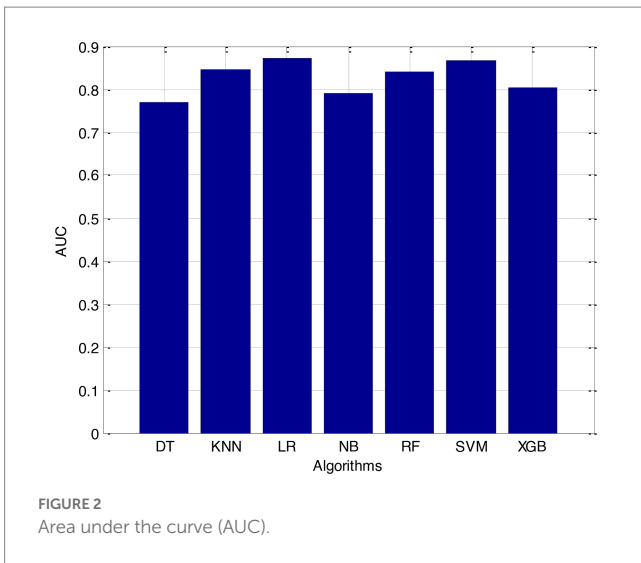
It is necessary to have a risk model in order to determine whether individuals have a high probability of developing CVD. We intended to test seven machine learning (ML)-based models to evaluate how correctly they could predict the risk of CVD. The findings demonstrated that each one of them had good to excellent discrimination and that they were all accurately calibrated. When it came to forecasting the risk of CVD, penalised LR performed better than other machine learning models, just like SVM did. The specificity of the SVM was higher than that of the LR, while the LR had a lower level of sensitivity. It is possible that a higher level of specialisation was favoured in this Kazakh Chinese group because the majority of the population was nomadic and there were few medical services available. In addition to this, when taking calibration and DCA into consideration, SVM fared marginally better than LR. Because of this, SVM and LR can be used to find people in this group who are at a higher risk of CVD and to find out if putting risk-mitigation interventions in place for these people will improve their CVD outcomes during the clinical decision-making process.

Linear regression has been widely used in the clinic to construct predictive models due to the ease with which it may be interpreted and its general straightforwardness. In a study aimed at predicting myocardial ischemia, both LR and SVM were shown to have the same level of predictive ability, which was consistent with our findings. A recent and exhaustive study concluded that there is no performance benefit to be gained from using ML in clinical prediction models over using LR. It was determined that when machine learning algorithms were applied to small datasets with a limited number of predictors, LR models might perform better than ML models in terms of overall performance. It is possible that the small sample size and the L1 penalised technique used in this work are to blame for the superior performance of LR in comparison to other machine learning models, with the exception of SVM.

The Support Vector Machine (SVM) is a well-known supervised machine learning approach that has found applications in a wide variety of business sectors. The fundamental idea behind support vector machines (SVM) is to locate the hyperplane that has the

TABLE 2 Software specifications.

Language	Python Version 3.9.7
Operating system	Windows 11
Tool box for model prediction	Sklearn 1.0.2
Feature selection	SHAP 40.0
Optimum values	Optuna 2.10.0 framework

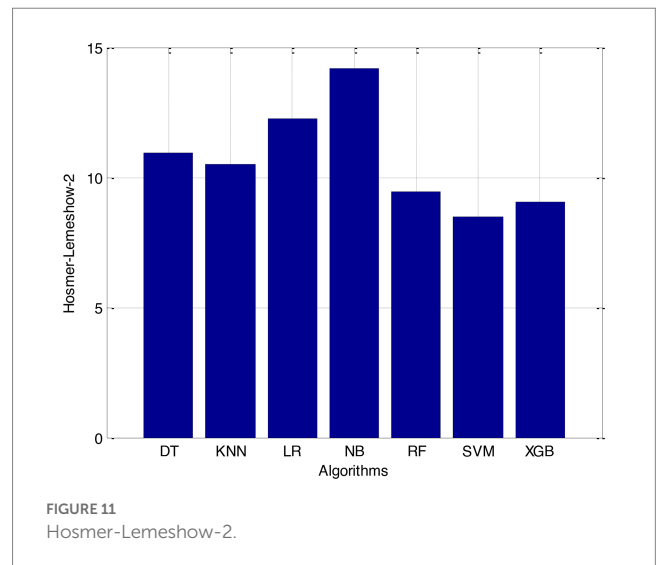
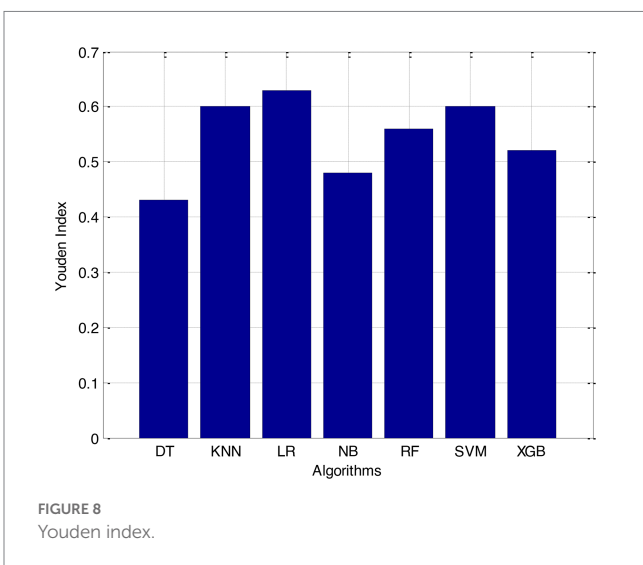
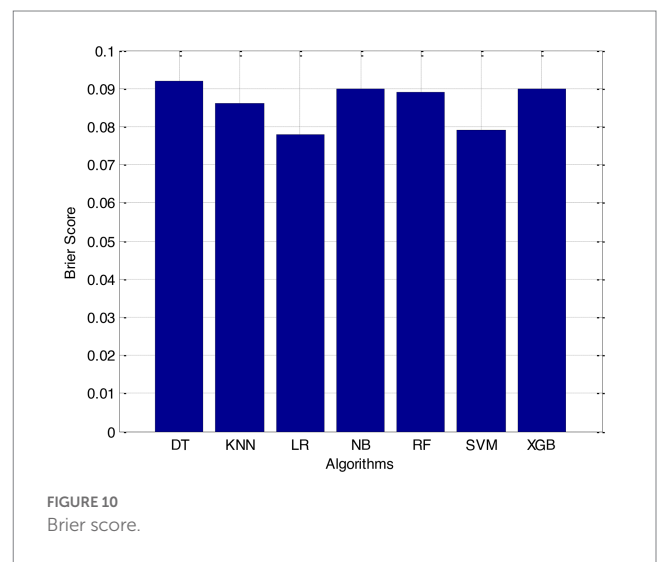
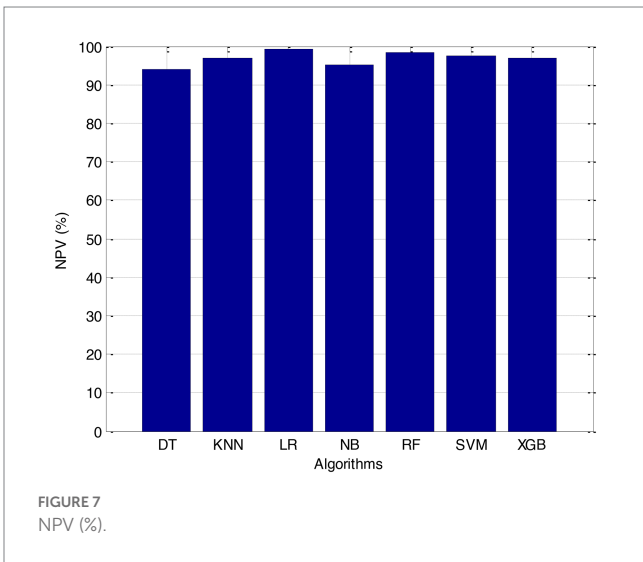
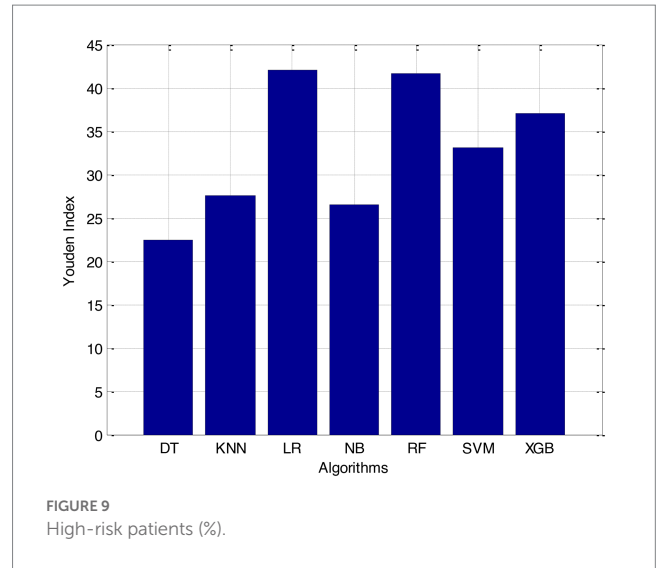
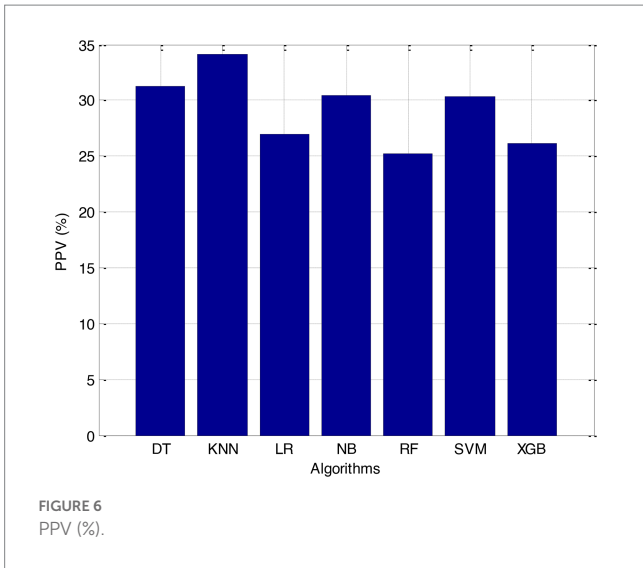


capacity to effectively classify the data while also providing the biggest geometric margin. In addition to this, it possesses significant kernel capabilities, which simplify the process of dealing with nonlinear classification issues. The outstanding performance of SVM demonstrates that it is a great tool for tackling classification challenges on small, non-linear, and high-dimensional datasets. This demonstrates that SVM is an excellent tool. In our experiment, we observed that the SVM performed significantly better than other machine learning models.

When it comes to classification, RF is among the most successful ensemble learning strategies that may be used. Its predictions were not nearly as accurate as those generated by the LR and SVM algorithms, which were the other two options. It is likely that the limited number of participants in this study will prevent RF from achieving its full potential as a prediction tool. The concept of variable importance was utilised in order to locate potential indicators of CVD. Some studies suggest that RF may be capable of revealing crucial but undisclosed predictions.

According to the results of feature selection that was based on RF, the age of the patient was the most significant predictor in the classification of CVD. In this study, it was discovered that certain risk factors, such as smoking and alcohol intake, were not as predictive as previously believed. Previous studies have shown that the synthetic indices BAI and LHR are both very good indicators of cardiovascular disease. Inflammation plays a significant part in the formation of atherosclerotic plaques as well as the progression of cardiovascular disease is shown in [Figures 6–11](#).

There is evidence that inflammatory cytokines, such as high-sensitivity CRP and interleukin-6, are associated with an elevated risk of cardiovascular disease. The Hs-CRP inflammatory marker was included in the Reynolds Risk Score in order to account for its role as a potential contributor to cardiovascular disease. hs-CRP has been shown in a number of other epidemiological studies to be an important predictor of CVD. These studies have also shown that hs-CRP acts as a mediator in the pathogenesis of vascular disease and is a marker of endothelial dysfunction. These findings are consistent with the findings of the aforementioned studies. It was discovered that Hs-CRP increases the risk of atherosclerotic plaque rupture in



addition to destabilising atherosclerotic plaques *via* NO, IL-6, and prostacyclin.

In addition, hs-CRP has been demonstrated to enhance thrombosis and cardiomyocyte apoptosis in response to hypoxia, which provides more support for its position as a risk factor for cardiovascular disease. It has been demonstrated that IL-6 is a factor in the course of atherosclerosis and that it promotes the creation of atherosclerotic plaques. This factor may have contributed to the increase in the number of cases of CVD. Taking statins, which can reduce a person's chance of acquiring CVD, is beneficial for a great number of people and can help them avoid developing the condition. In clinical practice, Hs-CRP and IL-6 can be used as biomarkers for the early diagnosis of patients who have an increased likelihood of developing cardiovascular disease.

According to the findings of our study, a decrease in ADP was associated with an increased risk of developing cardiovascular disease. The adipose hormone ADP, which is secreted by adipocytes, has anti-inflammatory effects. These effects manifest themselves as a reduction in the levels of CRP and lymphocyte recruitment in atherosclerotic lesions, a reduction in the expression of TNF-, and an increase in the production of cytokines that are protective against inflammation. On the other hand, there is evidence from a small number of studies that suggests an increase in ADP may assist in avoiding an ischemic stroke. Increased NEFA concentrations have been associated with an increased risk of cardiovascular disease in previous research, and our study came to the same conclusion. The possible effects of NEFA on cardiovascular disease include the potential to exacerbate or worsen a number of illnesses, including type 2 diabetes, hypertension, the metabolic syndrome, and endothelial deterioration, to name a few. Patients can have a lower chance of developing cardiovascular disease if they are treated to have a lower ADP (CVD).

The risk prediction models that are currently being used in CVD domains were built using traditional statistical methodologies, as many studies have revealed. Nevertheless, these models have been proven to be erroneous in external populations. In the field of cardiology, machine learning algorithms have proven to be superior methods for deriving predictions from big datasets that are notoriously difficult to understand. No prior assumptions are made by machine learning algorithms, which means that any data can be used to develop accurate and resilient models. Because of this, ML is able to model more complex relationships between outcomes and predictors, which are typically more challenging to express using standard statistical methods. Discovering interactions between marginal predictors can help improve risk-management strategies when using ML.

Machine learning has the potential to identify new genotypes and phenotypes for a variety of CVDs, as well as novel risk factors for CVDs. All aspects of medical picture recognition, diagnosis, outcome prediction, and prognosis evaluation can be improved with the application of more sophisticated machine learning techniques such as deep learning and artificial neural networks (ANN). It is possible that in the future, cardiologists will make better clinical decisions if they use machine learning models rather than the CVD risk stratifications that are currently used. On the other hand, most ML models may be hard for medical professionals to understand and use, which may limit how often they can be used in clinical settings.

5. Conclusion

According to the findings of this research, a stacking fusion model-based classifier performs better than individual models on all assessment criteria. This finding suggests that stacking models can combine the benefits of a variety of model types to achieve superior prediction performance. The recommended stacking approach offers improved prediction performance, increased resilience, and increased utility for individuals who are at high risk of developing cardiovascular disease. Hospitals can utilize this information to identify patients who are at a high risk of developing cardiovascular disease and provide them with early clinical intervention in order to reduce that risk. Research in the field of deep learning will benefit from additional data from a large number of medical institutions, which may be used for the development of artificial neural network structures or for the usage of deep learning frameworks in the future. In future work, the other deep learning techniques algorithms can be incorporated into Internet of Things (IoT) environments which helps to achieve more accuracy in terms of result and it can be useful to the hospitals and saving several human life.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MA and KS: conceptualization. SS, NV, and MA: methodology and investigation. TU and LA-k: software. MSo, TU, and NA: validation. KA and RK: formal analysis. KA and MA: data curation. SS and NV: writing—original draft preparation. MA, MSo, LA-k, and RK: writing—review and editing. LA-k, NA, and RK: supervision. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors are thankful to the princess Nourah Bint Abdulrahman University researcher, supporting program number (PNURSP2023R82) Princess Nourah bint Abdulrahman University, Riyadh Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Al Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J*. (2019) 40:1975–86. doi: 10.1093/eurheartj/ehy404
- Ghosh P, Azam S, Jonkman M, Karim A, Shamrat FJM, Ignatious E, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*. (2021) 9:19304–26. doi: 10.1109/ACCESS.2021.3053759
- Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. *Sci Rep*. (2020) 10:1–11. doi: 10.1038/s41598-020-72685-1
- Alaa AM, Bolton T, Di Angelantonio E, Rudd JH, Van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK biobank participants. *PLoS One*. (2019) 14:e0213653. doi: 10.1371/journal.pone.0213653
- Mezzatesta S, Torino C, De Meo P, Fiumara G, Vilasi A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Comput Methods Prog Biomed*. (2019) 177:9–15. doi: 10.1016/j.cmpb.2019.05.005
- Brites I. S., Silva L. M., Barbosa J. L., Rigo S. J., Correia S. D., Leithardt V. R. (2022) “Machine learning and iot applied to cardiovascular diseases identification through heart sounds: a literature review.” in *International Conference on Information Technology & Systems*. Springer, Cham, 356–388.
- Dinesh K. G., Arumugaraj K., Santhosh K. D., Mareeswari V. (2018) “Prediction of cardiovascular disease using machine learning algorithms.” in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE, 1–7.
- Li Y, Sperrin M, Ashcroft DM, Van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. *BMJ*. (2020) 371:1–8. doi: 10.1136/bmj.m3919
- Padmanabhan M, Yuan P, Chada G, Nguyen HV. Physician-friendly machine learning: a case study with cardiovascular disease risk prediction. *J Clin Med*. (2019) 8:1050. doi: 10.3390/jcm8071050
- Aryal S, Alimadadi A, Manandhar I, Joe B, Cheng X. Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension*. (2020) 76:1555–62. doi: 10.1161/HYPERTENSIONAHA.120.15885
- Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak*. (2019) 19:1–15. doi: 10.1186/s12911-019-0918-5
- Ponnusamy M, Bedi P, Suresh T, Alagarsamy A, Manikandan R, Yuvaraj N. Design and analysis of text document clustering using salp swarm algorithm. *J Supercomput*. (2022) 78:16197–213. doi: 10.1007/s11227-022-04525-0
- Allan S, Olaiya R, Burhan R. Reviewing the use and quality of machine learning in developing clinical prediction models for cardiovascular disease. *Postgrad Med J*. (2022) 98:551–8. doi: 10.1136/postgradmedj-2020-139352
- Yadav A, Singh A, Dutta MK, Travieso CM. Machine learning-based classification of cardiac diseases from PCG recorded heart sounds. *Neural Comput Applic*. (2020) 32:17843–56. doi: 10.1007/s00521-019-04547-5
- Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*. (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707
- Juhola M, Joutsijoki H, Penttinen K, Aalto-Setälä K. Detection of genetic cardiac diseases by Ca²⁺ transient profiles using machine learning methods. *Sci Rep*. (2018) 8:1–10. doi: 10.1038/s41598-018-27695-5
- Maheshwari V, Mahmood MR, Sravanthi S, Arivazhagan N, ParimalaGandhi A, Srihari K, et al. Nanotechnology-based sensitive biosensors for COVID-19 prediction using fuzzy logic control. *J Nanomater*. (2021) 2021:1–8. doi: 10.1155/2021/3383146
- Maini E., Venkateswarlu B., Gupta A. (2018). “Applying machine learning algorithms to develop a universal cardiovascular disease prediction system” in *International Conference on Intelligent Data Communication Technologies and Internet of Things*. Springer, Cham, 627–632.
- Li Q, Campan A, Ren A, Eid WE. Automating and improving cardiovascular disease prediction using machine learning and EMR data features from a regional healthcare system. *Int J Med Inform*. (2022) 163:104786. doi: 10.1016/j.ijmedinf.2022.104786
- Maiga J, Hungilo G. G. (2019). “Comparison of machine learning models in prediction of cardiovascular disease using health record data.” in *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* IEEE, 45–48.
- Sivasankari S. S., Surendiran J., Yuvaraj N., Ramkumar M., Ravi C. N., Vidhya R. G. (2022). “Classification of diabetes using multilayer perceptron.” in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*. IEEE, 1–5.
- Rahim A, Rasheed Y, Azam F, Anwar MW, Rahim MA, Muzaffar AW. An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access*. (2021) 9:106575–88. doi: 10.1109/ACCESS.2021.3098688
- Maurovich-Horvat P. Current trends in the use of machine learning for diagnostics and/or risk stratification in cardiovascular disease. *Cardiovasc Res*. (2021) 117:e67–9. doi: 10.1093/cvr/cvab059
- Ahn I, Gwon H, Kang H, Kim Y, Seo H, Choi H, et al. Machine learning-based hospital discharge prediction for patients with cardiovascular diseases: development and usability study. *JMIR Med Inform*. (2021) 9:e32662. doi: 10.2196/32662
- Arunachalam SK, Rekha R. A novel approach for cardiovascular disease prediction using machine learning algorithms. *Concurr Comput Pract Exp*. (2022) 34:e7027. doi: 10.1002/cpe.7027
- Kannan S, Yuvaraj N, Idrees BA, Arulprakash P, Ranganathan V, Udayakumar E, et al. Analysis of convolutional recurrent neural network classifier for COVID-19 symptoms over computerised tomography images. *Int J Comput Appl Technol*. (2021) 66:427–32. doi: 10.1504/IJCAT.2021.120453
- Smita, Kumar E. Probabilistic decision support system using machine learning techniques: a case study of cardiovascular diseases. *J Discret Math Sci Cryptogr*. (2021) 24:1487–96. doi: 10.1080/09720529.2021.1947452
- Di Castelnuovo A, Bonaccio M, Costanzo S, Gialluisi A, Antinori A, Berselli N, et al. Common cardiovascular risk factors and in-hospital mortality in 3,894 patients with COVID-19: survival analysis and machine learning-based findings from the multicentre Italian CORIST study. *Nutr Metab Cardiovasc Dis*. (2020) 30:1899–913. doi: 10.1016/j.numecd.2020.07.031
- Shu S, Ren J, Song J. Clinical application of machine learning-based artificial intelligence in the diagnosis, prediction, and classification of cardiovascular diseases. *Circ J*. (2021) 85:1416–25. doi: 10.1253/circj.CJ-20-1121
- Nashif S, Raihan MR, Islam MR, Imam MH. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World J Eng Technol*. (2018) 06:854–73. doi: 10.4236/wjet.2018.64057
- Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. (2017) 12:e0174944. doi: 10.1371/journal.pone.0174944
- Dimopoulos AC, Nikolaidou M, Caballero FF, Engchuan W, Sanchez-Niubo A, Arndt H, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Med Res Methodol*. (2018) 18:1–11. doi: 10.1186/s12874-018-0644-1
- Zaman M. I. U., Tabassum S., Ullah M. S., Rahaman A., Nahar S., Islam A. M. (2019). “Towards IoT and ML driven cardiac status prediction system.” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 1–6. IEEE
- Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep*. (2020) 10:1–8. doi: 10.1038/s41598-020-62133-5
- Hu Z, Qiu H, Su Z, Shen M, Chen Z. A stacking ensemble model to predict daily number of hospital admissions for cardiovascular diseases. *IEEE Access*. (2020) 8:138719–29. doi: 10.1109/ACCESS.2020.3012143
- Zheng H, Sherazi SWA, Lee JY. A stacking ensemble prediction model for the occurrences of major adverse cardiovascular events in patients with acute coronary syndrome on imbalanced data. *IEEE Access*. (2021) 9:113692–704. doi: 10.1109/ACCESS.2021.3099795