Advances in Computational Design, *Vol. 10, No. 4 (2025) 405-414*
https://doi.org/10.12989/acd.2025.10.4.405

An innovative approach that uses machine learning algorithms to detect heart problems

V. Kamakshi and S. Prasanna*

Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India

(Received July 28, 2025, Revised September 30, 2025, Accepted October 19, 2025)

Abstract. Early diagnosis and treatment are essential to minimizing the effects of heart based disease, which are the main causes of number death rates in and around the world. The nature of work focuses on using machine learning methods to predict cardiac problems early stage. Machine learning algorithms are implemented to analyze for large number of datasets and finding the complex patterns related to heart disease risk. Generally diseases are also many types of heart - based diseases. By integrating various data sources such as demographic information, medical history, genetic data, and lifestyle factors, machine learning models can effectively predict the likelihood of developing heart diseases. Papers represent an overview of the current research and its highlight the role of major machine learning in enhancing for early prediction capabilities. The results imply that machine learning algorithms may enhance risk classification and enable individualized therapies for the prevention and treatment of cardiac disorders. Future research and major studies are required to start the reliability and medical utility of these models in real-world healthcare settings. The attributes will be support for the finding the early stages of various heart diseases.

Keywords: data mining; echocardiogram; genetic data; machine learning; prediction; validation

1. Introduction

Heart disease continues to be a serious global health issue, greatly increasing the rates of sickness and death. In the medical or research field, machine learning and deep learning techniques are utilized to determine whether or not having diseases or not with the help of medical records, such as patient photo files or CSV files from the electronic health records, indicate that coronary heart disease exists or not various algorithms are also applied (Organisation 2017). Early prediction of heart diseases plays a critical role in effective prevention, timely interventions, and improved patient outcomes. Traditional risk assessment models have limitations in accurately identifying each individual at high risk for developing heart diseases, as they often deal with a limited set of risk factors and may not capture the complex interplay of various contributing factors. To overcome these challenges, researchers and clinicians are turning to machine learning techniques to develop more accurate and personalized prediction models.

Machine learning algorithms have clearly demonstrated about the major capabilities in handling large number of dataset and diverse datasets, identifying intricate patterns, and

*Corresponding author, Ph.D., E-mail: prasanna.scs@velsuniv.ac.in

http://www.techno-press.org/?journal=acd&subpage=7 ISSN: 2383-8477 (Print), 2466-0523 (Online)

calculating predictive models. By leveraging advanced computational algorithms, these techniques have the potential to incorporate a wide range of data sources including clinical data, genetic information, lifestyle factors, and imaging data. This integration allows for a more comprehensive and nuanced assessment of an own risk profile.

The use of machine learning in predicting heart diseases offers several advantages. Firstly, it enables the exploration of complex interactions between multiple risk factors, leading to more accurate risk stratification. Secondly, machine learning algorithms can easily adapt and learn from new type of data that allow for continuous model to train development and refinement. This dynamic nature enhances the predictive power of the models and enables them to adapt to emerging risk factors.

To provide an overview of the current many research on early stage prediction of heart diseases applying many machine learning algorithms. It explores the various types of data sources utilized in these models and the different machine learning algorithms employed. The potential outcomes and challenges link with implementing machine learning in clinical practice are also discussed. Moreover, the paper highlights the need for validation studies and real-world applications to assess the performance and clinical utility of these models.

Machine learning approaches have the potential to alter the practice of cardiovascular care by enhancing early prediction capabilities. Accurate detection of those with a high risk of developing heart disease can lead to focused therapies, dietary changes, and individualized treatment programs (Amin *et al*. 2013). Ultimately, the goal of machine learning in predicting heart diseases can contribute to reducing the burden of heart diseases and improving patient outcomes. The primary objective of the paper is to predict of heart based disease with possible number of attributes by utilizing several algorithms, including SVM, LR, NB, K-NN, and RF. But in the Random forest technique provide the best accuracy result with less number of attributes used.

2. Background:

Heart based disease will be the main major causes of death rate in worldwide, affecting millions of people. An accurate medical diagnosis is skilled, dependable, and supported by computer systems to low the actual price of diagnostic examinations. Data mining also enables system to build and categorize with different quantities and qualities. This study analysis various categorization methods to find cardiac disease (Xu *et al*. 2017). This section contains a brief introduction to related topics like machine learning and its methodology, along with details on data cleaning, assessment measures, and some clinical dataset applied in this study.

3. Machine learning:

A major area of artificial intelligence called “machine learning” focuses on developing systems, letting them learn, and then using what they’ve learned to make predictions. Built using machine learning algorithms, the model uses new type of attributes to predict heart disease and also for different types of patterns in the input data set to create that easily predict new accuracy data.

4. Overview of different techniques

To develop a computational analogy of heart muscles, the cardiologist might have employed

Table 1 Comparative literature review

Author 1: (Otoom *et al*. 2015) used techniques such as Support Vector Machine (SVM) and Bayes Net. The main purpose is focusing on monitoring and effective diagnosis of heart disease. These techniques output the various accuracy levels, such as SVM (Support Vector Machine), accuracy is 88.3%, and Bayes Net, which accuracy is 84%.
Author 2: (Ayoub Khan 2020) used techniques such as artificial neural networks (ANN), support vector machines (SVM), and the RIPPER classifier. The main purpose is focusing on data mining classification techniques in cardiovascular disease. In this method, SVM (Support Vector Machine) is the only one to predict the heart disease accurately.
Author 3: (El Hamdaoui 2020) used techniques such as the K-NN algorithm, support vector machine (SVM), and neural network. The main purpose is focusing on data mining techniques to predict heart disease. They considered that the SVM (Support Vector Machine) is the most efficient method to find heart based disease.
Author 4: (Raju *et al*. 2018) used techniques such as Naive Bayes and support vector machine. The main purpose is focusing on methods in diagnosing heart disease for diabetic patients. The method Naïve Bayes gives 74%, and SVM gives 94.60%. So, SVM is the most accurace compared to the Naïve Bayes technique.
Author 5: (Kumar and Rani 2020) Thomas used various techniques, such as four algorithms: KNN, SVM, modified KNN, and decision tree. The main purpose is focusing on prediction using ensemble learning algorithms and methods. These methods give the accuracy for techniques such as KNN 91.21%, SVM 92.31%, modified KNN 92.31%, and decision tree 87.91%.

various techniques from the field of computational modeling and simulation, such as mathematical models or computational simulations based on physiological principles (Ayon *et al*. 2022). These approaches aim to mimic the behavior of the heart and analyze the impression of different variables the development and progression of heart based diseases.

In the context of machine learning and data analysis, there are several algorithms commonly used for heart disease prediction, including but not limited to:

1. Logistic Regression:

Based on the input type of variables, is known as logistic regression, which is usually used to solve binary classification problems, may be able to forecast whether or not a person has heart disease.

2. SVM:

Support vector machines are a potent approach for multiple-class as well as binary data classification. It locates the best hyper plane to divide various classes in a high dimensional space.

3. Random Forest:

It is an ensemble type of method that merges multiple decision trees to produce number of predictions. It is generally applied for both classification and regression problems because of its robustness and capacity to manage data.

4. Naïve Bayes:

This approach is a statistical technique grounded in Bayes' theorem. It works well for classification problems, such as predicting heart disease, and operates under the assumption that the features are independent of one another given the class label.

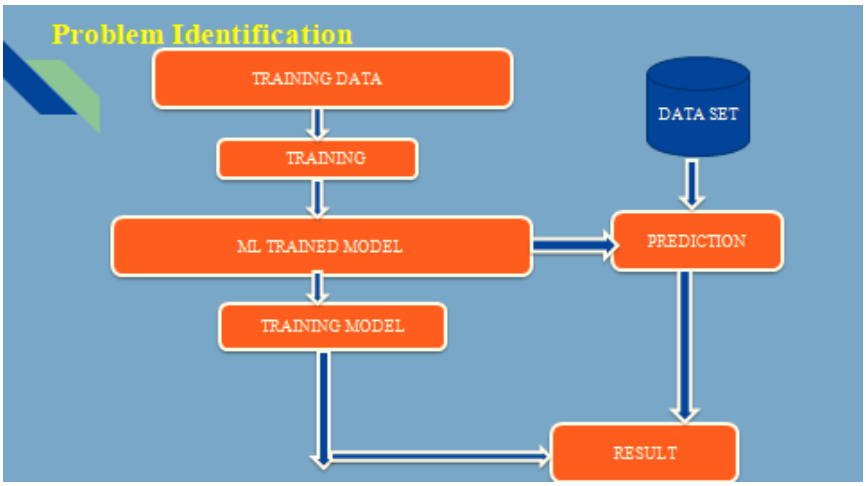

Fig. 1 Proposed model

5. K-Nearest Neighbors (KNN):

It is a non-parametric technique that categorizes occurrences based on their properties and methods in the feature space. When underlying data distribution is not clearly defined, it can be effective for classification tasks.

Without further information about the specific algorithms used in the heart based disease some prediction methods mentioned in the paper, it is difficult to provide a precise answer (Ansari *et al*. 2020). It's possible that the cardiologist utilized one or more of these algorithms or employed a custom computational model specifically designed for the task. There are different techniques and accuracy levels applied to find the comparative study discussed below.

5. Proposed model

In the proposed model we can classify easily using the training data the machine learning models and various trained model from them the huge number of dataset the prediction can be identified considered as a result. The models are divided into different stages shown in Fig. 1

1. Data Collection:

At this stage, we can collect the attributes from the data sources which are available in the formatted as a CSV file by implementing the required codes and processes as inputs for the recommended model.

2. Data Pre-processing:

During this phase, we can manage on handling missing values, minimizing the noise factors and the selecting features for the particular training purpose.

3. Pattern Recognition:

During this stage, we identify patterns and methods that embody knowledge based on specific metrics. Performance evaluation is also conducted to achieve optimal results.

4. Outputs/Results:

At this point, we are able to determine the results we wanted in the analysis step.

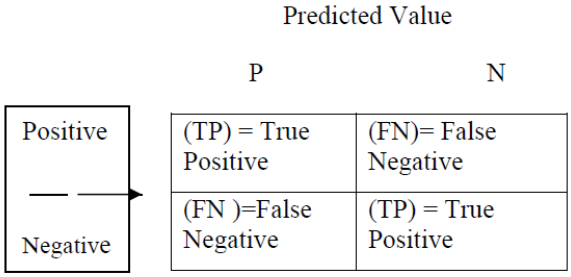

Fig. 2 Confusion matrix

6. Using random forest classifer

Using supervised classification, the random forest technique can be used. Several trees make up a forest in this algorithm. In a random forest, every tree predicts class likelihood, and the final prognosis is determined by the class that obtains the huge number of votes. More trees added to the random forest will increase its accuracy. It can manage missing variables and is applicable for both classification and regression tasks, though it performs particularly well in classification applications (Otoom *et al*. 2015).

In the below data visualization the five attributes to find the solution with the help of graph used to plot easily understand the data in the percentage format. There are some attributes identified with values if the blood pressure can be calculated by the both woman and Men. The normal blood pressures for the age group are from the age of 8 – 39years for women normal pressure 110/68 mm hg for the men 119/70 mm hg. If the age group of above 40-59years the women have 122/74 mm Hg for men 124/77 mm Hg. If age 60+years for women 139/68 mm Hg and Men 133/66 mm Hg.

Among other vital functions, cholesterol is a unique lipid (fat) that supports your body. High blood cholesterol levels can be hazardous. Atherosclerotic plaque, or hardened deposits, can develop as a result of it entering your arterial wall, compromising the wall's integrity (Raju *et al*. 2018). Atherosclerosis is the name for this process of plaque accumulation. It may result in significant issues like: Coronary artery disease: A blockage of the blood supply to the heart. Peripheral artery disease: Your arms and legs' blood flow is restricted. Carotid artery disease leads to an obstruction in the blood flow to the brain.

7. Evaluation process

The evaluation processes apply accuracy, precision, recall, sensitivity, and F1 score to evaluate interpretation can be obtained. A confusion matrix is a tabular representation that displays true and predicted values, referred to as true positives and true negatives. There are four components to its definition. The first is True Positive (T.P), which identifies values as true. The second type is known as False Positive (F.P), values are false but are recognized are found. The third type of False Negative (F.N) occurs when a valid value is reported as negative. The fourth one, known as True Negative (T.N), had a negative value that was accurately classified as negative (Kumar and Rani 2020). The confusion matrix table is shown in Fig. 2.

An accuracy score is used to assess a prototype performance level. It also used to calculate by

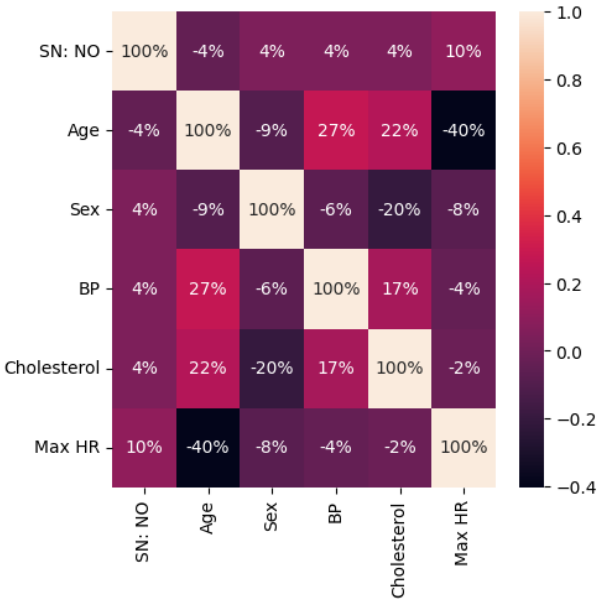

Fig. 3 Correlations heatmap

dividing the sum of the True Positive and Negative values divided by the sum of the true positive and negative values as well as false positive and negative. The Eq. is (1).

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

A classifier's ability to detect negative cases is measured by its specificity, which comes after accuracy and the percentage of true negative that were labeled as negative. It is also known as the negative rate. The formula is TN stands for True Negative and FP for False Positive.

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

The percentage rate of real positive events that were expected to be positive (or true positive) is therefore referred to as sensitivity. Another name for sensitivity is recall. Stated differently, an unhealthy individual was projected to be unhealthy. The equation is

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

8. Data visualization

The correlations coefficients can be viewed based on several attributes in a dataset are shown pictorially structure in a correlation heatmap in Fig 3. This kind of heatmap shows how strongly and in which direction two variables are correlated, as indicated by the color's intensity (Hasan and Bao 2020).

Correlation quantity values can be range from -1 to 1.

There are some values such as:

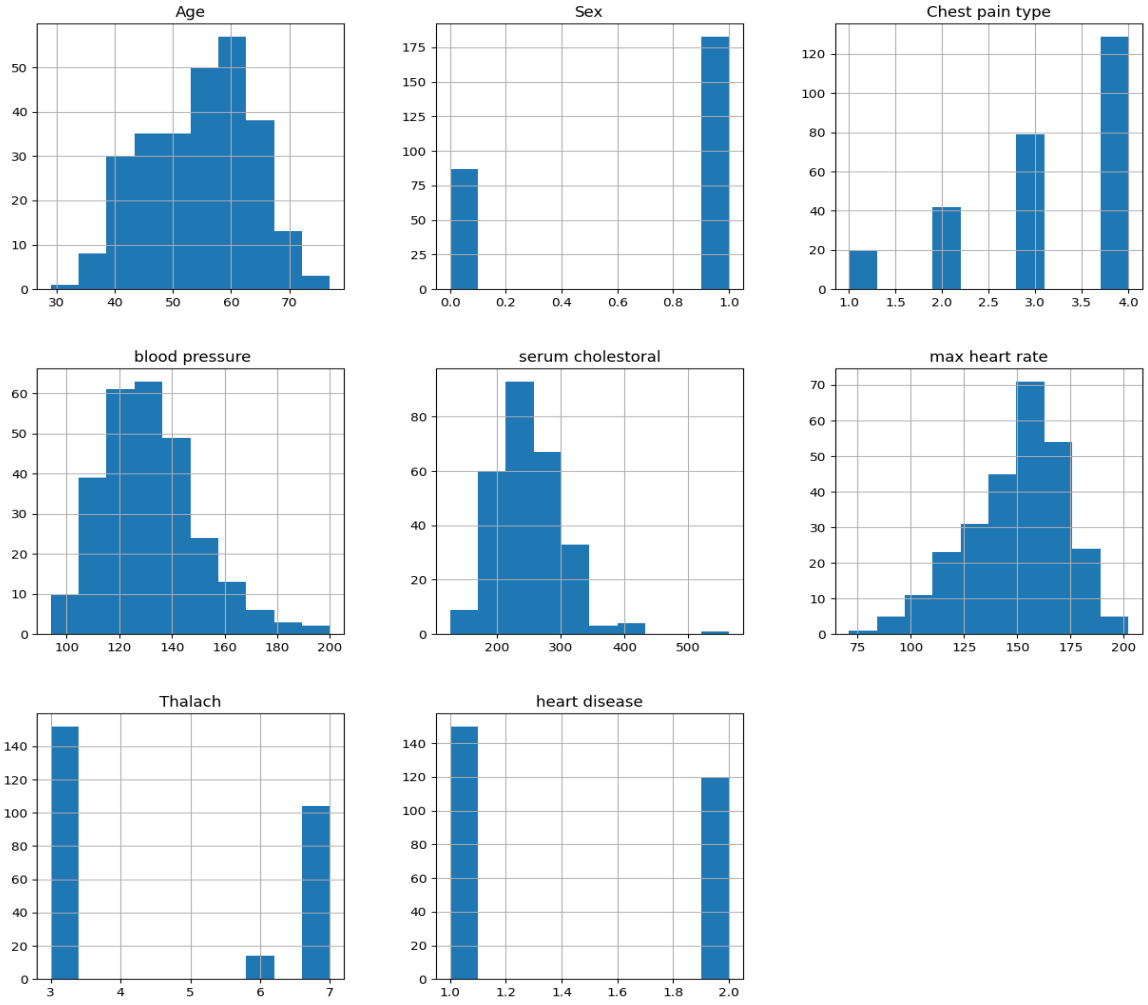

Fig. 4 Bar charts

+1: This range denotes Perfect positive correlation, meaning that when one variable rises, so does the other.

0: The variables do not relate to one another or have a linear relationship.

-1: This range denotes the perfect negative correlation, of one variable increases, the other variable decreases (Ouf and ElSeddawy 2021).

9. Color coding:

Correlation's durability and direction are embodied by colors. Generally speaking, positive correlations are symbolized by shades of red or orange, whereas negative correlations are symbolized by shades of blue. The power of the hue demonstrates the level of relationship. Below we can see the six different solutions with the various values. Easily identified the each parameters easily, displays values.

Matplotlib is a well-liked Python data visualization toolkit for making animated, interactive, and static visualization (Khan and Mondal 2020). Plotting functions are extensive, and it is particularly well-suited for producing charts and graphs of publishing quality. We can easily plot and identified different levels of various dataset.

Visualizations assist in identifying trends, patterns, and abnormalities in heart health datasets.

There are different types of tools are available such as

-> Examine the numbers of patients with and without cardiac disease to compare the prevalence of various illnesses using bar charts.

->Examine the distribution of variables such as chest pain type levels using histograms plots.

->Examine correlations between variables, such as age and cholesterol levels, using scatter plots (Palaniappan and Awang 2008).

10. Visualizing risk factors

Comparing the distributions of heart based disease risk factors, such as pressure or cholesterol, between groups (such as those with and without heart disease) is possible with box plots in Fig. 5.

Emphasize the geographic distribution of risk factors for diseases (Dangare and Apte 2012) There is possible to anticipation the heart disease count along with the number of patients in easy way. These methods are used by healthcare providers to identify and count cases of heart disease in individuals:

11. Physical examination and medical history:

• Gather data regarding risk factors (such as smoking, family history, and diabetes) and symptoms (such as shortness of breath, chest pain, etc.) (Zriqat *et al*. 2017)

• Physical examinations to look for symptoms such as elevated blood pressure, edema in the limbs, or irregular heartbeats.

12. Assessment tests:

• Electrocardiogram (ECG/EKG): ECG signal captures the activity of the heart to identify anomalies such as irregular heartbeats or heart attacks.

• Echocardiogram: Produces images of the heart using ultrasonography to evaluate its structure and function (Liu *et al*. 2017).

• Stress Test: Tracks cardiac activity while exercising or taking medicine. Blood tests measure cardiac biomarkers such as troponin, which indicate heart damage, as well as triglyceride and cholesterol levels (Aggrawal and Pal 2021).

• Coronary Angiography: This technique visualizes coronary artery blockages using X-rays (Alvanou *et al*. 2022).

13. Conclusions

Estimating a patient risk affected heart oriented disease with limited number of attributes can

easily predict the diseases risk is the aim of this study (Rahim *et al*. 2021). This work applied supervised machine learning classification techniques with the help of UCI repository, many techniques and algorithms are applied such as Naive Bayes, Decision Trees, Random Forests, and K-Nearest neighbour. The data was categorized and divided into a test set and a training set (Spencer *et al*. 2021). The data is pre-processed using supervised classification methods like Nave Bayes, SVM and decision tree also find the accuracy values.

References

Aggrawal, R. and Pal, S. (2021), "Prediction of heart disease with different attributes combination by data mining algorithms", *Proccedings of the Computational Vision and Bio-Inspired Computing: ICCVBIC 2020*, 469-482, Singapore, Springer Singapore. https://doi.org/10.1007/978-981-33-6862-0_38.

Alvanou, A.G., Stylidou, A. and Exarchos, T.P. (2022), "Web-based decision support system for coronary heart disease diagnosis", *Proceedings of the GeNeDis 2020: Computational Biology and Bioinformatics*, 31-38, Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-78775-2_5.

Amin, S.U., Agarwal, K. and Beg, R. (2013), "Genetic neural network based data mining in prediction of heart disease using risk factors", *Proceedings of the 2013 IEEE Conference on Information & Communication Technologies*, 1227-1231, IEEE, April.

Ansari, M.F., Alankar, B. and Kaur, H. (2020), "A prediction of heart disease using machine learning algorithms", *Proceedings of the International Conference on Image Processing and Capsule Networks*, 497-504, Cham: Springer International Publishing, May. https://doi.org/10.1007/978-3-030-51859-2_45.

Ayon, S.I., Islam, M.M. and Hossain, M.R. (2022), "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques", *IETE J. Res*., **68**(4), 2488-2507. https://doi.org/10.1080/03772063.2020.1713916

Ayoub Khan, M. (2020), "An IoT Framework for Heart Disease Prediction based on MDCNN Classifier", *arXiv e-prints*, arXiv-2012. https://doi.org/10.48550/arXiv.2012.05999.

Dangare, C.S. and Apte, S.S. (2012), "Improved study of heart disease prediction system using data mining classification techniques", *Int. J. Comput. Appl*., **47**(10), 44-48.

El Hamdaoui, H., Boujraf, S., Chaoui, N.E.H. and Maaroufi, M. (2020), "A clinical support system for prediction of heart disease using machine learning techniques", *Proceedings of the 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 1-5, IEEE, September. https://doi.org/10.1109/ATSIP49331.2020.9231760.

Hasan, N. and Bao, Y. (2020), "Comparing different feature selection algorithms for cardiovascular disease prediction", *Health Technol*., **11**, 49-62. https://doi.org/10.1007/s12553-020-00499-2

Khan, I.H. and Mondal, M.R.H. (2020), "Data-driven diagnosis of heart disease", *Int. J. Comput. Appl*., **176**, 46-54.

Kumar, R. and Rani, P. (2020), "Comparative analysis of decision support system for heart disease", *Adv. Math. Sci. J*., **9**(6), 3349-3356. https://doi.org/10.37418/amsj.9.6.15.

Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q. and Wang, Q. (2017), "A hybrid classification system for heart disease diagnosis based on the RFRS method", *Comput. Math. Methods Med*., **2017**(1), 8272091. https://doi.org/10.1155/2017/8272091.

Otoom, A.F., Abdallah, E.E., Kilani, Y., Kefaye, A. and Ashour, M. (2015), "Effective diagnosis and monitoring of heart disease", *Int. J. Softw. Eng. Appl*., **9**(1), 143-156. https://doi.org/10.14257/ijseia.2015.9.1.12.

Ouf, S. and ElSeddawy, A.I.B. (2021), "A proposed paradigm for intelligent heart disease prediction system using data mining techniques", *J. Southwest Jiaotong Univ*., **56**, 220-240.

Palaniappan, S. and Awang, R. (2008), "Intelligent heart disease prediction system using data mining techniques", *Proceedings of the 2008 IEEE/ACS International Conference on Computer Systems and Applications*, 108-115, IEEE, March.

Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M.A. and Muzaffar, A.W. (2021), "An integrated machine learning framework for effective prediction of cardiovascular diseases", *IEEE Access*, **9**, 106575-106588. https://doi.org/10.1109/ACCESS.2021.3098688

Raju, C., Philipsy, E., Chacko, S., Suresh, L.P. and Rajan, S.D. (2018), "A survey on predicting heart disease using data mining techniques", *Proceedings of the 2018 Conference on Emerging Devices and Smart Systems (ICEDSS)*, 253-255, IEEE, March. https://doi.org/10.1109/ICEDSS.2018.8544333.

Spencer, R., Thabtah, F., Abdelhamid, N. and Thompson, M. (2020), "Exploring feature selection and classification methods for predicting heart disease", *Digital Health*., **6**, 2055207620914777. https://doi.org/10.1177/2055207620914777.

World Health Organization. (2017), "New initiative launched to tackle cardiovascular disease, the world number one killer", Intra-Health International. http://www. who. int/cardiovascular_diseases/en/.

Xu, S., Zhang, Z., Wang, D., Hu, J., Duan, X. and Zhu, T. (2017), "Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework", *Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 228-232, IEEE, March.

Zriqat, I.A., Altamimi, A.M. and Azzeh, M. (2017), "A comparative study for predicting heart diseases using data mining classification methods", *Int*. *J. Comput*. *Sci*. *Inform*. *Security (IJCSIS)*, **14**(12), 868-879.

PR