International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 2 (2024) 61–74

Paper Type: Original Article

Survey of Heart Disease Prediction Using Various Machine Learning Technique

Samar L. Elhawy [1,*]

[1] Faculty of Computers and Informatics, Zagazig University, Sharqiyah, Zagazig, 44519, Egypt; SLAlhawy@fci.zu.edu.eg.

Received: 18 Dec 2023 **Revised:** 20 Feb 2024 **Accepted:** 18 Mar 2024 **Published:** 20 Mar 2024

Abstract

Many prediction techniques targeted at early identification and intervention have been developed since heart disease continues to be a major cause of death worldwide. The predominant approaches used in cardiac disease prediction are examined in this review of the literature, with an emphasis on statistical models, hybrid approaches, and machine learning (ML) techniques. Because they provide comprehensible findings and are often easy to execute, traditional statistical approaches like logistic regression and decision trees have served as the foundation for predictive modelling. Through the use of sophisticated algorithms like Random Forest, Support Vector Machine (SVM), and Neural Networks (NN), machine learning has, nevertheless, dramatically improved forecast accuracy. Recent research highlights deep learning methods that enhance early diagnosis by recognising complex patterns in large datasets. Furthermore, the integration of clinical data, such as ECG signals, patient history, and genetic markers, with ML algorithms is increasingly recognized for its potential to elevate prediction accuracy. Challenges persist, particularly regarding data quality, model interpretability, and the clinical application of advanced models. The review concludes by highlighting the importance of personalized models and real-time prediction systems in future research, aiming to bridge the gap between algorithmic development and clinical utility.

Keywords: Machine Learning; Support Vector Machine; Neural Networks.

1 | Introduction

In the burgeoning field of healthcare informatics, the extraction of medical association rules stands as a pivotal endeavor, offering insights into complex relationships within medical datasets [1, 2]. Medical association rules elucidate patterns and correlations among various medical variables, aiding in clinical decision-making, disease diagnosis, treatment planning, and healthcare management. Leveraging semantic data mining algorithms in this pursuit not only enhances the accuracy and interpretability of extracted rules but also facilitates the integration of domain knowledge into the mining process [3]. The significance of extracting medical association rules lies in its potential to uncover hidden knowledge from vast repositories of medical data. By identifying meaningful associations between medical variables, healthcare practitioners can gain valuable insights into disease etiology, patient outcomes, and treatment efficacy [4, 5]. This, in turn, can inform evidence-based medical practices, optimize resource allocation, and ultimately improve patient care.

Heart disease continues to be a significant global health challenge, contributing to a substantial portion of morbidity and mortality worldwide. Early detection and precise risk prediction are critical for effective prevention and management of heart conditions. In recent years, the field of medical informatics has

Corresponding Author: SLAlhawy@fci.zu.edu.eg

increasingly focused on leveraging data mining techniques to extract association rules from cardiovascular datasets [6]. This introduction aims to provide an overview of heart disease association rules extraction, highlighting its importance, challenges, and potential applications, supported by relevant references. The significance of heart disease association rules extraction lies in its potential to uncover hidden patterns and relationships within cardiac data, offering insights into disease mechanisms, risk factors, and treatment outcomes [7]. By analyzing associations between clinical variables such as demographic factors, medical history, biomarkers, and lifestyle habits, researchers can identify novel risk predictors and inform personalized interventions [8]. This approach holds promise for improving risk stratification, guiding clinical decision-making, and ultimately enhancing patient outcomes in the management of heart disease [9, 10].

CVDs or heart diseases are one of the most difficult, deadly, and life-threatening illnesses that affect people worldwide. CVDs in [57] are classified into various categories as represented in Figure 1. Also, Figure 1 illustrates the factors causing the disease and its consequences.

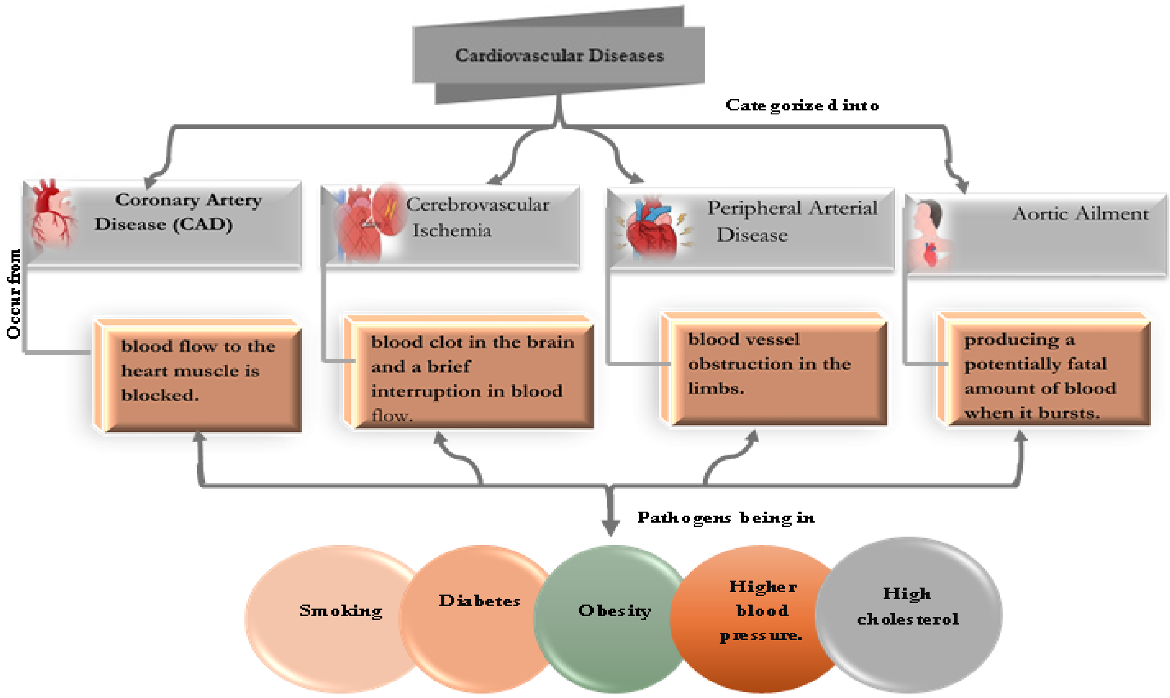

Figure 1. Cardiovascular diseases categories.

Even though stressing that heart disease is among the most common diseases in humans, [58] also acknowledged that the diagnosis of heart disease can be complicated and delay the right diagnosis decision due to several factors, including symptoms of the disease and the relationship between the disease's pathological and functional manifestations and human organs other than the heart. In this context [59] indicated that improving clinical outcomes and preventing significant adverse cardiac events is possible with early detection of CVDs, which also enables the escalation of guideline-directed medical treatment. Arguably early identification of cardiovascular disorders is imperative. From the perspective of [60], forecasting with precision is a challenge, particularly in developing and Asian nations where resources, technology, and peripheral devices are few. In the same context [61], was revealed that the capacity for controlling CVD is currently inadequate. Individuals lack knowledge about CVDs and possible detrimental practices that contribute to the disease. Another challenge stated in [62] where a wealth of information on heart disease in the healthcare sector needs to be analyzed to aid with decision-making. The mentioned challenges are catalysts for the notion of diagnosing CVD early to lessen the risk embraced by [63] through using appropriate and precise diagnostic techniques. Moreover, earlier studies tackle the difficulties associated with CVDs and diagnostics by developing a variety of diagnostic methods for the techniques-based prediction of heart disease [64]. For instance [65] where medical practitioners are integrating their experiences as physicians with artificial intelligence (AI) techniques to automate the diagnosing process. Recently, several machine learning

(ML)techniques of AI have been constructed to improve the prediction of cardiovascular diseases. Utilizing Extreme Gradient Boosting (XGBoost) classifier [66] to predicate CVDs. On the other hand [67] adapted the uncertainty theory of the Complex intuitionistic fuzzy set (CIFS) for choosing the best method for diagnosing cardiovascular diseases.

2 | Traditional Approaches to Heart Disease Analysis

Traditional approaches to heart disease analysis have evolved, primarily relying on clinical expertise, epidemiological studies, and statistical methods. Here's an overview of some traditional approaches:

- Clinical risk factors such as age, gender, smoking status, blood pressure, cholesterol levels, and diabetes status have long been recognized as important predictors of heart disease risk [20].
- Longitudinal population-based studies like the Framingham Heart Study have been instrumental in identifying and understanding the epidemiology of heart disease [14].
- Statistical regression techniques, including logistic regression, have been commonly used to model the relationship between various risk factors and the likelihood of developing heart disease [17].
- Clinical guidelines provided by organizations such as the American Heart Association (AHA) and the European Society of Cardiology (ESC) offer evidence-based recommendations for heart disease prevention and management [11, 18].
- Risk scoring systems like the Framingham Risk Score and QRISK algorithm integrate multiple risk factors to provide personalized risk assessments for cardiovascular events [13] [16].
- Diagnostic tests such as electrocardiography (ECG), stress testing, and echocardiography play a crucial role in diagnosing heart disease and assessing its severity [15].
- Evidence-based treatment guidelines recommend lifestyle modifications, pharmacotherapy, and interventional procedures based on the severity of heart disease and individual patient characteristics [19].
- Population Health Interventions: Public health initiatives aimed at promoting cardiovascular health and preventing heart disease, including smoking cessation programs and dietary interventions, have contributed to reducing the burden of heart disease at the population level [12].

  Traditional approaches to heart disease analysis have evolved, and researchers have explored various methods for predicting cardiovascular disease (CVD) risk. Let's delve into some of these approaches:

- Single-Stage Models:

  These models use one data point per patient, typically at baseline. For instance, the widely used Framingham Risk Score predicts the risk of coronary heart disease based on variables like systolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, and smoking status1 [13]. However, single-stage models may not fully utilize available longitudinal data.

- Two-Stage Approaches:

  These models incorporate estimated longitudinal parameters in survival models. They consider multiple time points and account for changes over time [19]. As more data becomes available, two-stage approaches are gaining popularity for CVD risk prediction.

- Joint Models:

  Joint models simultaneously analyze both longitudinal and survival data. They integrate information from repeated measurements and time-to-event outcomes [15]. These models offer a comprehensive view of CVD risk by considering both risk factor trajectories and mortality outcomes.

While single-stage models are still widely used, future studies should explore two-stage and joint approaches to better leverage available longitudinal data for accurate CVD risk prediction1. Machine learning techniques have also been explored in this context, offering promising avenues for further research.

3 | Machine Learning in Heart Disease Analysis

Machine learning (ML) has emerged as a powerful tool in heart disease analysis, offering the potential to improve risk prediction, diagnostic accuracy, and treatment outcomes.

Types of fields that used Machine Learning in Heart Diseases:

- Machine learning algorithms can integrate diverse data sources, including clinical variables, imaging data, and genetic information, to develop more accurate risk prediction models for cardiovascular events [21].
- ML techniques, such as deep learning and ensemble methods, enable the development of diagnostic support systems capable of analyzing medical images (e.g., echocardiograms, angiograms) to detect and classify heart disease with high accuracy [22].
- ML algorithms aid in identifying relevant features and reducing the dimensionality of large datasets, enhancing the interpretability and efficiency of heart disease analysis models [23].
- ML-based approaches facilitate the identification of patient subgroups likely to benefit from specific treatments or interventions, enabling personalized and targeted management of heart disease [24].
- ML algorithms can analyze electronic health records and assist clinicians in making informed decisions regarding patient management, treatment selection, and risk stratification in real-time [25].

4 | Association Rules

To model and uncover the interdependencies between database entries, association rules are used. Their expressions reflect, respectively, hesitance and skepticism. The importance of an organization is determined by the extent of support and trust its members place in each other. Support, confidence, and lift are criteria to show the importance of associations [68, 69].

The criteria of association Rules:

Support

This metric provides insight into how often a certain collection of products appears in all trades. Let's pretend that Set1 is bread and Set2 is shampoo. There will be a lot more bread purchases than shampoo purchases. You correctly predicted that the support for set1 would be greater than that for set2. Let's say set 1 is "bread and butter" and set 2 is "bread and shampoo." Bread and butter are common cart items, but how often do you see bread and shampoo? Not really. In this situation, set1 is more likely to be preferred than set2 in terms of popularity. In mathematical terms, the amount of backing for an item set is the share of all transactions that include those objects.

$\text{support}\{\{x\}\rightarrow\{y\}\}=(\text{Transactions containing both x and y})/(\text{total number of transactions})$

Using support value, we can determine which rules are worth investigating further. If there are 10,000 transactions, for instance, it may be useful to focus on the subset of item sets that appears at least 50 times or has support = 0.005. Without additional data, we cannot draw any firm conclusions about the nature of the relationships among the items in a very poorly supported item set.

Confidence

This metric describes the probability that the consequent will be present on the cart, assuming that the antecedents are present. That is to say, of all the purchases that included the term "Captain Crunch," how many also included the word "Milk?" It's well known that the "Captain Crunch" vs. "Milk" guideline should be taken very seriously. Confidence, in technical terms, is the chance that the consequence will occur given the antecedent.

Confidence ({x}→{y})=(transactions containing both x and y)/(transaction containing x)

First, let's take a moment to think about a few additional situations. How sure are you that "Butter" and "Bread" are synonymous? To clarify, what percentage of purchases included both butter and bread? Extremely high, or very near to 1? Yeah, you nailed it. What about milk and yogurt? Back on top of the world. Milk for your toothbrush? Still unsure? Since "Milk" is such a common commodity, it is safe to assume that this rule will always hold.

When determining the conditional probability of occurrence of Y given X, Lift accounts for the support (frequency) of consequent. The word "lift" is used to describe this metric rather literally. Imagine this as the *boost* to our self-assurance that comes from having Y in the shopping basket thanks to the presence of X. To restate, lift is the increase in the chance of Y being on the cart due to the knowledge of X's existence relative to the probability of Y being on the cart due to ignorance of X's presence.

Lift ({x}→{y})=((transactions containing both x and y)/(Transaction containing x))/(Fraction of transactions containing y).

5 | Explainable Machine Learning in Healthcare

Explainable machine learning (XAI) techniques are increasingly important in heart disease prediction to enhance the interpretability and transparency of predictive models.

Types of Explainable Machine Learning Models in Heart Disease Prediction:

- Model Transparency: XAI ensures transparency by providing insights into how predictive models arrive at their decisions. This transparency allows clinicians to understand the factors driving each prediction and evaluate the model's reliability for individual patients [28].
- Feature Importance: XAI methods, such as SHAP (Shapley Additive exPlanations) values or LIME (Local Interpretable Model-agnostic Explanations), quantify the importance of features in predicting heart disease risk. These techniques help clinicians identify the most influential risk factors and tailor interventions accordingly [26].
- Interpretable Models: XAI promotes the development of interpretable machine learning models, such as decision trees or rule-based models, that provide clear and understandable rules for predicting heart disease. Clinicians can easily interpret these models and trust their predictions based on clinically relevant criteria [27].
- Clinical Actionability: Explainable predictions empower clinicians to take actionable steps in preventing or managing heart disease. By understanding the rationale behind each prediction, clinicians can tailor interventions to address specific risk factors identified by the model, leading to more effective patient care [25].

6 | Literature Review of Heart Disease Prediction

Recent years have seen significant advancements in the prediction of cardiac disease utilizing various machine-learning techniques.

Table 1. Summary of literature review.

Authors	Title	year	Tools and methods	Achievements
Palaniappan and Awang [29]	Intelligent heart disease prediction system using data mining techniques	2008	Naive Bayes Decision Tree, Genetic Algorithm	proposed Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, in which they used genetic algorithms to identify the characteristics that are more helpful in the diagnosis of heart conditions, thereby reducing the number of tests that a patient must undergo. By employing genetic search, 13 traits are whittled down to 6 attributes. With the same accuracy as before the reduction in the number of characteristics, three classifiers—Naive Bayes, Classification by Clustering, and Decision Tree—are then employed to predict the diagnosis of patients.
Guan et al. [30]	In mixed-integer support-vector machine	2013	SVM	proposed a system that could accurately forecast cardiac illness and was entirely based on support vector machines. The proposed system's accuracy rate is 76.5%. The system compared various models, including approximated L0-norm SVM methods, Recursive Characteristic Elimination, and Standard SVM.
Shilaskar and Ghatol [31]	Feature selection for medical diagnosis: evaluation for cardiovascular diseases,	2013	Feature selection algorithms and classification methodologies and SVM	employed multiple feature selection algorithms and classification methodologies to forecast coronary heart disease. Both the forward characteristic optimisation and selection as well as the back-elimination characteristic selection were done using an SVM classifier. Their research demonstrates that it decreased the number of input variables and increased accuracy. The algorithm discovered an accuracy rate of close to 85%.
Pawlovsky [32]	"Deep ensemble detection of congestive heart failure using short-term rr intervals	2019	The k closest neighbor (KNN) approach	developed an ensemble model employing distance and the k nearest neighbor (KNN) method for the diagnosis of heart disease. A configuration of three and five distances was used to implement the suggested model. Additionally, a weight is put at the base of the average accuracy that was calculated using KNN. Using the proposed method, an average accuracy of 85% was reported using the Cleveland, UCI dataset used in this experiment.
Anooj [33]	Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules	2011	Fuzzy rule and a neural network-based system	utilized a weighted rule to create a risk model, and a neural network-based system then assessed the model's performance using the UCI heart disease dataset.
Shao et al. [34]	Hybrid intelligent modeling schemes for heart disease classification	2014	Logistic regression, rough set techniques, and adaptive regression splines	proposed a system to determine the precision of predicting coronary heart disease using machine learning techniques like logistic regression and to choose the most effective and significant features. To reduce the number of explanatory characteristics needed to diagnose heart disease, the system used rough set techniques and multivariate adaptive regression splines. The accuracy of the suggested system was 82.14%.
Methaila et al. [35]	Assessing cardiovascular risks from a mid-thigh ct image: a tree-based machine learning approach using radio densitometric distributions	2020	ML techniques such as decision trees, NB, and NN. Apriori algorithm and frequent pattern mining with MAFIA	established a technique for forecasting heart disease using data mining technologies. The proposed method used machine learning (ML) techniques such as decision trees, NB, and NN to predict cardiac disease. An online dataset from the Cleveland Heart Disease database was used in the study. To reduce the feature dimension, the Apriori method and frequent pattern mining using MAFIA were applied. The significance weight computation of the features was evaluated for improved feature selection. The results of the proposed research showed that the decision tree performed better than the other ML techniques, using 15 features and an accuracy of 99.62%.

Archana Singh et al. [36]	Heart disease prediction using machine learning algorithms	2020	Linear regression, decision trees, support vector machines, and K-NN	created a model for predicting cardiac disease using machine learning classifiers. The dataset has 14 features that are used in training and testing models to get the highest level of accuracy. The classifiers produced the following results: 78% for linear regression, 79% for decision trees, 83% for support vector machines, and 87% for K-NN. Using 15 characteristics, K-NN demonstrated a maximum accuracy of 99.62%, according to the data.
Purushottam et al. [37]	Efficient heart disease prediction system	2016	Support vector machines, C4.5, neural networks, PART, multilayer perceptrons, and radial basis functions	To determine the connection between specific patients and the underlying cause of cardiac illness, researchers used radial basis functions, C4.5, neural networks, support vector machines, PART, and multilayer perceptrons.
Kim and Kang [38]	Neural network-based coronary heart disease risk prediction using feature correlation analysis	2017	A neural network using feature correlation analysis.	The 4146 entries in a Korean heart disease dataset were retrieved to find any connections between feature relations and important risk variables. Next, a neural network was trained on this dataset using feature correlation analysis. The recommended model performed better than the Framingham risk score.
Amin et al. [39]	Genetic neural network-based data mining in prediction of heart disease using risk factors	2013	Neural Networks System and the Genetic Algorithm	proposed a hybrid model that uses the most crucial risk factors to categorize heart disease. They used the Neural Networks system and the Genetic Algorithm, two well-known techniques, for their system. They initialized the weight of each neuron on the neural networks using a genetic algorithm and a global optimization technique. Their research showed that their model is rapid in comparison to other models and that it has an 89% accuracy.
Khatibi and Montazer [40]	A fuzzy-evidential hybridinference engine for coronary heart disease risk assessment	2010		Proposed a fuzzy-based system employing the Dempster-Shafer theory and fuzzy set ideas. The suggested approach consists of two steps. Fuzzy units are used to first characterise the inputs, and then the fuzzy inference system is used to carry out the fuzzy sets. Second, it created the interval of beliefs for the hybrid inference engine and used combination rules to mix the various pieces of knowledge. It was carried out with a 91.58% accuracy.
Barik et al. [41]	Heart disease prediction using machine learning techniques	2020	Decision trees, random forests, and other algorithms and WEKA tools	To forecast cardiac disease in its early stages, decision trees, optimized decision trees, random forests, and other algorithms were used. These risk models were created with the help of the RapidMiner and WEKA tools, and their accuracy, precision, sensitivity, and specificity were examined.
Temurtas and Tanrikulu [42]	An approach to probabilistic neural network for diagnosis of mesothelioma's disease	2012	Neural network approaches.	Suggested a model to categorize a dataset of cardiac disease using neural network approaches. The dataset was divided into three separate cross-validations, and the outcome was then determined using neural network techniques. In the experiment, their model's accuracy was 96.30%.
Yumusak and Temurtas [43]	Chest disease diagnosis using artificial neural networks	2010	Neural networks	employed multilayer neural networks to forecast the development of coronary heart disease. They observed that their model had an average accuracy of 91.60% using two hidden layers between the input and output layers.
Dulhare et al. [44]	Prediction system for heart disease using ¨ıve Bayes and particle swarm optimization	2018	Particle swarm optimization (PSO) and	combined the particle swarm optimization (PSO) and naive Bayesian common feature selection methods to produce a reliable cardiac disease prediction. The UCI repository of the VA Long Beach machine learning dataset, which

			Na¨ıve Bayesian algorithms	comprises 270 records and 14 attributes, was used for the model training and testing activities. Only seven of the 14 characteristics of heart disease have been demonstrated to be predictive of the illness. PSO and NB together increase NB's performance accuracy to 87.91%. Studies show that accuracy rises by 8.79 percent when compared to net-benefit accuracy.
Kim et al. [45]	Neural network-based coronary heart disease risk prediction using feature correlation analysis	2017	Machine learning algorithms and DT algorithm	making predictions about heart disease using machine learning techniques. The datasets were taken from the repository of machine learning at the University of California, Irvine (UCI), which consists of 303 records and employs 14 attributes. For training and testing, the 10-fold cross-validation approach is applied. The DT algorithm predicts cardiac disease with a higher accuracy of 93.19%.
CarlosOrdonez et al. [46]	Evaluating association rules and decision trees to predict multiple target attributes	2011	Association rules and decision trees	suggested A detailed comparison of limited association rules and decision trees was presented when assessing their ability to predict several target features. There are notable differences between the two tactics for this goal. A large amount of experimental evaluation was conducted on an actual medical data set to determine criteria for disease forecasting on different heart arteries. Association rules include two parts: a consequent that specifies the severity of the disease on one or more arteries, and an antecedent that comprises patient risk factors and medical measures. Limited association rule mining yields more rules that are more numerous and dependable than the prediction rules generated by decision trees.
Anupriya et al. [47]	Enhanced Prediction of Heart Disease with Feature Subset Selection Using Genetic Algorithm	2010	Genetic algorithms and Naive Bayes, Classification by Clustering, and Decision Tree	By utilizing genetic algorithms to identify the qualities that are more helpful in the diagnosis of heart problems, the researchers proposed "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm," which would reduce the number of tests a patient would need to do. Thirteen criteria are reduced to six by the genetic search technique. The patient's diagnosis is then predicted using three classifiers (Naive Bayes, Classification by Clustering, and Decision Tree) with the same level of accuracy as before the feature reduction.
A. Ahmed et al. [49]	Heart Disease Prediction under Machine Learning and Association Rules under Neutrosophic Environment	2023	neutrosophic AHP machine learning random forest (RF) and decision tree (DT) k-nearest neighbors (KNN), and gradient boosting AdaBoosting logistic regression and Naïve Bayes SVM	employed three methods to predict cardiac disease: machine learning models, association rules, and the neutrosophic analytical hierarchy process (AHP) for feature selection. The process of determining feature weights and choosing the best features involves the usage of the neutrosophic AHP approach. All datasets' values are governed by rules that are provided by the association rules. Next, we selected the optimal feature to use using the neutrosophic AHP feature selection technique. Input for models of machine learning. To predict cardiac disease, we employed nine machine-learning models. We found that the decision tree (DT) and random forest (RF) had the highest accuracy at 100%, followed by gradient boosting, k-nearest neighbors (KNN), bagging, and AdaBoosting at 99%, 98%, and 97%, respectively. Logistic regression and Naïve Bayes came in last with 89% accuracy.
Devansh et al. [50]	Heart Disease Prediction Using Machine Learning Techniques	2020	Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithm	Used the existing dataset from the Cleveland database of UCI repository of heart disease patients. The dataset comprises 303 instances and 76 attributes. Of these 76 attributes, only 14 attributes are considered for testing, important to substantiate the performance of different algorithms.

				Results portray that the highest accuracy score is achieved with the K-nearest neighbor. Accuracy of KNN is 90.789%
Umarani et al. [51]	Machine Learning Technology-Based Heart Disease Detection Models	2022	Na¨ıve Bayes, support vector machine (SVM) and XGBoost	uses XGBoost to test alternative decision tree classification algorithms in the hopes of improving the accuracy of heart disease diagnosis. In terms of precision, accuracy, f1-measure, and recall as performance parameters above mentioned, four types of machine learning (ML) models are compared. The accuracy of Na¨ıve Bayes's weighted approach is 86.00%. The accuracy of SVM's and XGBoost is 94.03%. Accuracy of XGBoost is 95.9%. So, the best accuracy when using XGBoost
Safial et al. [52]	Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques	2020	Logistic Regression (LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbour (K-NN)	Used Statlog and Cleveland heart disease dataset which are retrieved from the UCI machine learning repository database with several evaluation techniques. From the study, it can be carried out that the highest accuracy of 98.15% was obtained by a deep neural network with sensitivity and precision of 98.67% and 98.01% respectively.
Shokoofa et al. [53]	Data Mining and Diagnosis of Heart Diseases: A Hybrid Approach to the B-Mine Algorithm and Association Rules	2022	Support vector machine classifications, k nearest neighbor, decision tree, and simple Bayesian	Used real and standard datasets of cardiac patients show that the average accuracy of the proposed method is approximately 98%, which has been tested on the Cleveland database that includes 76 features in the case of the heart disease dataset, 14 features of which are related to heart disease. The accuracy of predicting the results of the support vector machine classifications, k nearest neighbor, decision tree and simple Bayesian is 81.11%, 66.67%, 59.72%, and 19.85%, respectively, which are relatively satisfactory results.
Yahia et al. [54]	Effectiveness of Artificial Intelligence Models for Cardiovascular Disease Prediction: Network Meta-Analysis	2022	Artificial neural network (ANN), Support vector machine (SVM), Random Forest	Used machine learning models, association rules, and the neutrosophic analytical hierarchy process (AHP) for feature selection as three strategies to forecast heart illness. Utilizing the neutrosophic AHP technique is necessary to determine feature weights and select the optimal features. The values of all datasets are determined by rules given by the association rules. The best feature was then chosen utilizing the neutrosophic AHP feature selection method. Were looked at. In all, 17 studies with 285,213 CVD patients were included in the network meta-analysis. With an AUC of 0.843, the DL algorithms did well in the prediction of heart failure, whereas the gradient boosting machine (GBM) in the ML algorithm achieved an average accuracy of 0.843, according to statistical data.
Syed et al. [55]	Significance of Visible Non-Invasive Risk Attributes for the Initial Prediction of Heart Disease Using Different Machine Learning Techniques	2022	Random forest, Na¨ıve Bayes, decision tree, support vector machine, and K nearest neighbor	The dataset was gathered using quantitative data collection techniques from several heterogeneous data sources in Kashmir, India. Findings indicate that the random forest model performs better than other risk assessment models, with an ideal 85% accuracy, 83% specificity, 85% sensitivity, 85% precision, 85% AUROC score, and only 13% misclassification rate. The random forest's prediction accuracy for heart disease is the highest, surpassing that of earlier research.

Abdul et al. [48]	AMethod for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms	2022	AB, LR, ET, MNB, CART, SVM, LDA, RF, and XGB.	After comparing the performance of several classifiers, we find that the XGB and ET classifiers exhibit generally better accuracy. SVM, on the other hand, achieved an accuracy of 96.72% and demonstrated the best performance while modifying the hyperparameters.
K. Karthick et al. [56]	Implementation of a Heart Disease Risk Prediction Model Using Machine Learning	2022	Support vector machine (SVM), Gaussian Naive Bayes, logistic regression, LightGBM, XGBoost, and random forest algorithm	Used the Cleveland HD dataset from the UCI ML repository. The purpose of the data visualization is to show how the features relate to one another. The experiments show that the random forest algorithm validates 303 data instances with 13 selected attributes of the Cleveland HD dataset with an accuracy of 88.5%. Heart disease risk prediction models have been developed using support vector machine (SVM), logistic regression, Gaussian Naive Bayes, XGBoost, LightGBM, and random forest method; the resulting models have yielded accuracy values of 80.32%, 78.68%, 80.32%, 77.04%, 73.77%, and 88.5%, respectively.

7 | Limitation of Previous Literature Review

Used other data mining techniques such as time series, clustering and association rules, support vector machine, and genetic algorithm. There is a need to implement a more complex combination of models to get higher accuracy for the early prediction of heart disease. Adding more attributes to the heart disease dataset and making it more interactive for the users. It can also be carried out as a mobile application with reduced computing time and complexity. The same methods can be used to predict different diseases and more intelligence methods will be used to predict coronary artery heart disease. Increased performance in knowledge extraction and a thorough comprehension of the challenges associated with measuring and gathering data will result from the diversity of resources. Designed to offer a technique for making predictions. In addition to extracting association rules, we provided a model for future association rule forecasting, which means that the next set of data and associated rules are predicted as accurately as possible based on the extracted rules and the recorded data. Similar studies are urged to focus more on DL models for the same or different outcomes related to cardiovascular diseases, in addition to performing a Bayesian A prior, likelihood, and posterior distribution that are suitable for network meta-analysis. To evaluate the risk of heart disease, investigate deep learning and the effects of additional controllable variables on different age and gender groups. Compare the outcomes to determine if a higher accuracy can be achieved when utilizing XGBoost to predict cardiac disease in youngsters. In terms of classifying and predicting cardiac sickness, effective feature management will yield notable outcomes.

8 | Conclusions

This literature review paper highlights heart disease association rules extraction, emphasizing its significance, challenges, and applications in clinical decision-making and treatment planning. Various frameworks for extracting heart disease association rules using explainable machine learning are discussed, along with the associated challenges, opportunities, and case studies. Traditional approaches to heart disease analysis, including clinical risk factors, population studies, and diagnostic tests, are also examined to provide context. Additionally, the role of machine learning in heart disease analysis, personalized treatment strategies, and clinical decision support systems is highlighted. The importance of explainable machine learning in heart disease prediction is emphasized, focusing on enhancing model transparency, feature importance interpretation, and ensuring clinical actionability to improve patient care and outcomes in the management of heart disease.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22(2), 207–216. https://doi.org/10.1145/170036.170072

[2] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 77(2), 81–97. https://doi.org/10.1016/j.ijmedinf.2006.11.006

[3] Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the F-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association, 12(3), 296–298. https://doi.org/10.1197/jamia.M1733

[4] Wang, J., & Han, J. (2009). BIDE: Efficient mining of frequent closed sequences. Data Mining and Knowledge Discovery, 18(2), 292–320. https://doi.org/10.1007/s10618-008-0114-3

[5] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

[6] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668–2679. https://doi.org/10.1016/j.jacc.2018.03.521

[7] Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: Promise and challenges. Nature Reviews Cardiology, 13(6), 350–359. https://doi.org/10.1038/nrcardio.2016.42

[8] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

[9] Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. P., & Marwick, T. H. (2016). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. Journal of the American College of Cardiology, 73(11), 1317–1335. https://doi.org/10.1016/j.jacc.2019.12.032

[10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765–4774).

[11] Arnett, D. K., Blumenthal, R. S., Albert, M. A., Buroker, A. B., Goldberger, Z. D., Hahn, E. J., ... & Smith Jr, S. C. (2019). 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: Executive summary. Journal of the American College of Cardiology, 74(10), 1376–1414. https://doi.org/10.1016/j.jacc.2019.03.009

[12] Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., ... & Muntner, P. (2019). Heart disease and stroke statistics—2018 update: A report from the American Heart Association. Circulation, 139(10), e56–e528. https://doi.org/10.1161/CIR.0000000000000659

[13] D'Agostino Sr, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. Circulation, 117(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579

[14] Dawber, T. R., Meadors, G. F., & Moore Jr, F. E. (1951). Epidemiological approaches to heart disease: The Framingham Study. American Journal of Public Health and the Nations Health, 41(3), 279–281. https://doi.org/10.2105/AJPH.41.3.279

[15] Fihn, S. D., Gardin, J. M., Abrams, J., Berra, K., Blankenship, J. C., Dallas, A. P., ... & Gibbons, R. J. (2014). 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease. Journal of the American College of Cardiology, 60(24), e44–e164. https://doi.org/10.1016/j.jacc.2012.07.013

[16] Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. BMJ, 335(7611), 136. https://doi.org/10.1136/bmj.39261.471806.55

[17] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons. https://doi.org/10.1002/9781118548387

[18] Piepoli, M. F., Hoes, A. W., Agewall, S., Albus, C., Brotons, C., Catapano, A. L., ... & Wood, D. A. (2016). 2016 European guidelines on cardiovascular disease prevention in clinical practice. European Heart Journal, 37(29), 2315–2381. https://doi.org/10.1093/eurheartj/ehw106

[19] Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... & Muntner, P. (2020). Heart disease and stroke statistics—2020 update: A report from the American Heart Association. Circulation, 141(9), e139–e596. https://doi.org/10.1161/CIR.0000000000000757

[20] Wilson, P. W., D'Agostino Sr, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. Circulation, 97(18), 1837–1847. https://doi.org/10.1161/01.CIR.97.18.1837

[21] Gao, J., Wang, F., Wang, S., & Zhao, P. (2017). Atrial fibrillation detection based on deep learning network with RR interval signals. IEEE Access, 5, 11815–11821. https://doi.org/10.1109/ACCESS.2017.2711060

[22] Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C. P., ... & Ng, A. Y. (2020). Combining universal adversarial training with a conditional variational autoencoder for medical image segmentation. arXiv preprint arXiv:2007.13867.

[23] Bogunovic, H., De Fauw, J., & Aung, T. (2019). Learning feature engineering for classification. arXiv preprint arXiv:1901.04971.

[24] Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: Are we there yet? Heart, 104(14), 1156–1164. https://doi.org/10.1136/heartjnl-2017-311198

[25] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2016). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657–2664. https://doi.org/10.1016/j.jacc.2017.03.571

[26] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765–4774).

[27] Craven, M., & Shavlik, J. W. (1996). Extracting tree-structured representations of trained networks. In Advances in Neural Information Processing Systems, 18(3), 24–30

[28] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

[29] Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. In IEEE/ACS International Conference on Computer Systems and Applications (Vol. 8, No. 8, pp. 108–115). https://doi.org/10.1109/AICCSA.2008.4493524

[30] Guan, W., Gray, A., & Leyffer, S. (2013). In mixed-integer support vector machine. In Mini Symposia & Workshops NIPS (pp. 1–6).

[31] Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. Expert Systems with Applications, 40, 4146–4153. https://doi.org/10.1016/j.eswa.2013.01.010

[32] Wang, L., Zhou, W., Chang, Q., Chen, J., & Zhou, X. (2019). Deep ensemble detection of congestive heart failure using short-term RR intervals. IEEE Access, 7, 69559–69574. https://doi.org/10.1109/ACCESS.2019.2917537

[33] Anooj, P. K. (2011). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. King Saud University - Computer and Information Sciences, 24(1), 27–40. https://doi.org/10.1016/j.jksuci.2011.09.002

[34] Shao, Y. E., Hou, C. D., & Chiu, C. C. (2014). Hybrid intelligent modeling schemes for heart disease classification. Applied Soft Computing, 14, 47–52. https://doi.org/10.1016/j.asoc.2013.09.021

[35] Ricciardi, C., Edmunds, K. J., Recenti, M., & others. (2020). Assessing cardiovascular risks from a mid-thigh CT image: A tree-based machine learning approach using radiodensitometric distributions. Scientific Reports, 10(1), 2863. https://doi.org/10.1038/s41598-020-59669-3

[36] Singh, A., & Kumar, R. (2020). Heart disease prediction using machine learning algorithms. In International Conference on Electrical & Electronics Engineering (pp. 452–457). https://doi.org/10.1109/ICEE.2020.9355879

[37] Purushottam, K. S., Saxena, K., & Sharma, R. (2016). Efficient heart disease prediction system. Procedia Computer Science, 85, 962–969. https://doi.org/10.1016/j.procs.2016.05.251

[38] Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. Journal of Healthcare Engineering, Article ID 2780501. https://doi.org/10.1155/2017/2780501

[39] Amin, S. U., Agarwal, K., & Beg, R. (2013). Genetic neural network-based data mining in prediction of heart disease using risk factors. In Information and Communication Technologies (ICT) (pp. 1227–1231). IEEE. https://doi.org/10.1109/CICT.2013.6558239

[40] Khatibi, V., & Montazer, G. A. (2010). A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. Expert Systems with Applications, 37, 8536–8542. https://doi.org/10.1016/j.eswa.2010.05.029

[41] Barik, S., Mohanty, S., Rout, D., Mohanty, S., Patra, A. K., & Mishra, A. K. (2020). Heart disease prediction using machine learning techniques. In Advances in Electrical Control and Signal Systems (Vol. 665, pp. 879–888). https://doi.org/10.1007/978-3-030-54882-6_79

[42] Temurtas, F., & Tanrikulu, A. C. (2012). An approach on probabilistic neural network for diagnosis of mesothelioma's disease. Computers & Electrical Engineering, 38, 75–81. https://doi.org/10.1016/j.compeleceng.2012.01.001

[43] Yumusak, N., & Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. Expert Systems with Applications, 37, 7648–7655. https://doi.org/10.1016/j.eswa.2010.04.069

[44] Dulhare, U. N. (2018). Prediction system for heart disease using Naïve Bayes and particle swarm optimization. Biomedical Research, 29(12), 2646–2649. https://doi.org/10.4066/biomedicalresearch.29-18-318

[45] Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. Journal of Healthcare Engineering, 2017, Article ID 2780501. https://doi.org/10.1155/2017/2780501

[46] Ordonez, C., & Zhao, K. (2011). Evaluating association rules and decision trees to predict multiple target attributes. Intelligent Data Analysis, 15, 173–192. https://doi.org/10.3233/IDA-2010-0470

[47] Anupriya, E., et al. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology, 2(10), 5370–5376.

[48] Abdul, S., et al. (2022). A method for improving prediction of human heart disease using machine learning algorithms. Mobile Information Systems, 2022, Article ID 1410169. https://doi.org/10.1155/2022/1410169

[49] Ahmed, A., et al. (2023). Heart disease prediction under machine learning and association rules under neutrosophic environment. Neutrosophic Systems with Applications, 10, 2023. https://doi.org/10.1007/s12530-022-01005-4

[50] Devansh, et al. (2020). Heart disease prediction using machine learning techniques. In Springer Nature Singapore Pte Ltd (pp. 665–675). https://doi.org/10.1007/978-981-15-9010-9_64

[51] Umarani, et al. (2022). Machine learning technology-based heart disease detection models. Journal of Healthcare Engineering, 2022, Article ID 7351061, 9 pages. https://doi.org/10.1155/2022/7351061

[52] Safial, et al. (2020). Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. IETE Journal of Research. https://doi.org/10.1080/03772063.2020.1713916

[53] Shokoofa, et al. (2022). Data mining and diagnosis of heart diseases: A hybrid approach to the B-Mine algorithm and association rules. International Journal of Research in Industrial Engineering, 11(1), 77–91. https://doi.org/10.22105/riej.2022.349117.1346

[54] Yahia, et al. (2022). Effectiveness of artificial intelligence models for cardiovascular disease prediction: Network meta-analysis. Computational Intelligence and Neuroscience, 2022, Article ID 5849995, 12 pages. https://doi.org/10.1155/2022/5849995

[55] Syed, et al. (2022). Significance of visible non-invasive risk attributes for the initial prediction of heart disease using different machine learning techniques. Computational Intelligence and Neuroscience, 2022, Article ID 9580896, 12 pages. https://doi.org/10.1155/2022/9580896

[56] Karthick, K., et al. (2022). Implementation of a heart disease risk prediction model using machine learning. Computational and Mathematical Methods in Medicine, 2022, Article ID 6517716, 14 pages. https://doi.org/10.1155/2022/6517716

[57] Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of cardiovascular diseases (CVD) using machine learning and deep learning techniques. ICT Express, 8(1), 109–116. https://doi.org/10.1016/j.icte.2021.04.005

[58] Hassan, D., Hussein, H. I., & Hassan, M. M. (2023). Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. Biomedical Signal Processing and Control, 79, 104019. https://doi.org/10.1016/j.bspc.2022.104019

[59] Khozeimeh, F., Sharifrazi, D., Izadi, N. H., Joloudari, J. H., Shoeibi, A., Alizadehsani, R., & Islam, S. M. S. (2022). RF-CNN-F: Random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. Scientific Reports, 12(1), 1–12. https://doi.org/10.1038/s41598-022-15374-5

[60] Rustam, F., Ishaq, A., Munir, K., Almutairi, M., Aslam, N., & Ashraf, I. (2022). Incorporating CNN features for optimizing performance of ensemble classifier for cardiovascular disease prediction. Diagnostics, 12(6), 1474. https://doi.org/10.3390/diagnostics12061474

[61] Wei, X., Rao, C., Xiao, X., Chen, L., & Goh, M. (2023). Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. Expert Systems with Applications, 219, 119648. https://doi.org/10.1016/j.eswa.2023.119648

[62] 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

[63] Ashraf, M., Salal, Y. K., Abdullaev, S. M., Zaman, M., & Bhut, M. A. (2022). Introduction of feature selection and leading-edge technologies viz. TensorFlow, PyTorch, and Keras: An empirical study to improve prediction accuracy of cardiovascular disease. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021 (Vol. 3, pp. 19–31). Springer. https://doi.org/10.1007/978-3-030-87939-8_3

[64] Ali, L., Niamat, A., Khan, J. A., Golilarz, N. A., Xingzhong, X., Noor, A., & Bukhari, S. A. C. (2019). An optimized stacked support vector machines-based expert system for the effective prediction of heart failure. IEEE Access, 7, 54007–54014. https://doi.org/10.1109/ACCESS.2019.2910802

[65] Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using machine learning algorithms. Materials Today: Proceedings, 80, 3682–3685. https://doi.org/10.1016/j.matpr.2022.09.470

[66] Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost-based diagnostic system for effective prediction of heart disease. Journal of King Saud University-Computer and Information Sciences, 34(7), 4514–4523. https://doi.org/10.1016/j.jksuci.2021.06.013

[67] Alghazzawi, D., Liaqat, M., Razaq, A., Alolaiyan, H., Shuaib, U., & Liu, J. B. (2023). Selection of optimal approach for cardiovascular disease diagnosis under complex intuitionistic fuzzy dynamic environment. Mathematics, 11(22), 4616. https://doi.org/10.3390/math11224616

[68] Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. GESTS International Transactions on Computer Science and Engineering, 32(1), 71–82.

[69] Wen, F., Zhang, G., Sun, L., Wang, X., & Xu, X. (2019). A hybrid temporal association rules mining method for traffic congestion prediction. Computers and Industrial Engineering, 130, 779–787. https://doi.org/10.1016/j.cie.2018.12.028