

Heart Disease Predictive Analysis Using Association Rule Mining



Fatima D. Mulla alias Fatima M. Inamdar, NaveenKumar JayaKumar, and Bhushan Bari

Abstract Heart disease prediction is a challenging task that is under research from many decades. There are several factors that cause heart attacks in patients. These factors can be used to analyse and predict if a patient is having a risk of getting heart attack. This paper presents a risk factor analysis of factors that result in heart attack and put forth the association between different factors. The analysis of the association can help doctors personalize the treatment based on the patient condition. The rules of association, namely support, confidence and lift, have been used to find out how different factors, single and combined, can play a role in causing heart attack to the patients. The risk factor parameters under study are: thal, age, exang, restecg, chol, sex, cp, fbs, trestbps, and thalach. Finding the association between parameters can help us analyse what factors, when combined, can have the highest risk in causing heart attack.

Keywords Heart disease · Risk factor · Association · Support · Confidence · Trestbps · Restecg · Cp · Sex · Ca · Exang · Thal · Thalach · And Chol

F. D. M. F. M. Inamdar (✉) · N. JayaKumar · B. Bari
Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be) University
College of Engineering, VIIT, Pune, India
e-mail: fatima.inamdar@viit.ac.in

N. JayaKumar
e-mail: naveenkumar@bvucoep.edu

B. Bari
e-mail: bhushan.a.bari@bvucoep.edu

VIIT, Pune, India

N. JayaKumar
Department of Computer Engineering, Bharati Vidyapeeth (Deemed To Be) University
College of Engineering, Pune, India

B. Bari
Pune, India

1 Introduction

One of the major causes of death in the world is cardiovascular disease [1]. According to WHO, 17.9 million people die in India due to cardiovascular diseases. Eighty-five per cent of these deaths were due to heart stroke. Both Indian men and women have high mortality rate due to heart diseases. Cardiovascular disease (CVD) is the manifestation of atherosclerosis with over 50% diametric narrowing of a major coronary vessel [2]. The infestation pattern can be localized or diffused. Frequently, regardless of the degree of severity, an endothelial dysfunction, even if the findings are localized angiographically often additional plaques are present, is shown on the angiogram as wall irregularities are impressive.

Basically, the diagnostic-methodical procedure is the same as for nondiabetic. The stress ECG, stress echocardiography, and SPECT myocardial scintigraphy are available as noninvasive diagnostic procedures [3]. Electron beam tomography (EBCT) or magnetic resonance imaging (MRI) or multi-line computed tomography (MSCT) are a lot of promising new imaging procedures that are not yet part of the official guideline recommendations for routine diagnostics of the cardiovascular heart disease. EBCT or MSCT can be used to detect coronary calculus, which is a sensitive marker for the early stages of cardiovascular disease. Although the MSCT has a high diagnostic potential for risk assessment [4], the severity of an underlying coronary obstruction has yet to be reliably detected. On the other hand, a negative calcium score includes cardiovascular disease and in particular unstable plaques.

A group of blood vessels and heart disorders [5] is the cardiovascular disease (CVD), including:

- Cerebrovascular diseases: The blood vessels' diseases that supplying the brain
- Coronary heart disease: The blood vessels' disease that supplying the cardiac and the muscle.
- Congenital heart disease: This is the malformation of the heart that is present from birth I.
- Rheumatic heart disease: It is caused due to the bacteria called nothing strep, which causes rheumatic fever and injuries to heart valves and heart muscle.
- Peripheral arteriopathies: The blood vessels' diseases supply to the lower and upper limbs.
- Deep vein thrombosis and pulmonary embolisms: In the legs' veins, the blood is clotted (thrombi), which lodge and dislodge dense (emboli) in the heart and lungs' vessels.
- Angina pectoris: It is a clinical syndrome that causes insufficient supply of blood to the heart muscle. Symptoms are sore throat, pain in back, nausea, chest pain, choking, salivation, sweating, and pressure pain Zion. These symptoms appear quickly and intensify with movement.
- Cardiac infarction: A serious coronary disease is infarction with necrosis of the myocardium. Coronary stenosis develops first; it is suddenly blocked the pericardium, causing severe pain in the heart. The symptoms are tightness in the centre of the chest, a pain that radiates from the pericardium in chest, more and

more frequent chest pains, constant pain in the upper stomach, shortness of breath, sweating, feeling of annihilation, fainting, nausea, vomiting.

- **Arrhythmia:** The heart rhythm disorder when the coordination of the electrical activity of the heartbeat is not working properly, for what our heart begins to beat too fast, too slow, or irregularly. Arrhythmia may develop as a consequence of scarring. The heart muscle after a heart attack, diseases valves or coronary arteries of the heart, and abnormal thickness of the walls ventricular.

1.1 Risk Factors Causing Heart Disease

The most widely known cardiovascular risk factors are: tobacco, blood cholesterol, diabetes, high blood pressure, obesity, lack regular physical exercise (sedentary lifestyle), family history of cardiovascular disease and stress. In women, there are specific factors such as ova-polycystic rivers, oral contraceptives, and self-oestrogens. How much higher the level of each risk factor, the greater the risk of having a disease cardiovascular [6, 7].

Tobacco: Smoking increases the rate of heart, causing alterations in the heartbeat rhythm and constricts the main arteries. It makes the heart much harder. The blood pressure is also increased by smoking that increases the strokes' risk. This indicates that a smoker has twice the risk of myocardial infarction than a nonsmoker.

Cholesterol in the blood: When cholesterol levels rise, usually due to an inadequate diet when high in saturated fat, too much cholesterol is transported to the tissues, form that these cholesterol molecules can be deposited in the arteries, on all those that are around the heart and brain and get to plug them thus deriving serious heart problems.

Diabetes: Whether there is insufficient insulin production or resistance to the action, and accumulation of glucose in the blood, accelerating the arteriosclerosis process, acute myocardial infarction, increasing the cardiovascular disease risk: angina, and sudden death due to cardiac arrest.

Blood pressure: It is the pressure with which the blood circulates through the blood vessels when it leaves the heart (systolic blood pressure known as high pressure) or when the heart fills up with blood returning to the heart (diastolic blood pressure: commonly known as low pressure). The appearance of hypertension usually indicates that there is a cardiovascular risk.

Obesity: Being overweight or obese exponentially increases the risk of suffering a cardiovascular disease. Currently, overweight and obesity are considered as important as other classic risk factors related to the disease coronary. Adipose tissue not only acts as a store for fat molecules, but also synthesizes and releases numerous hormones related to metabolism into the blood of immediate principles and the regulation of intake.

Stress: Mental stress induces endothelial dysfunction, increased blood viscosity, stimulated aggregation of platelets, and promoted arrhythmogenesis by haemoconcentration and stimulates factors involved in inflammation. Cardiovascular hyper-responsiveness from blood pressure to mental stress tests, performed in the laboratory, has been related to the risk of improving high blood pressure.

Age: With age, the activity of the heart results is to be deteriorated. Thus, the heart walls' thickness can be increased, and the flexibility of arteries may lose. Before the body muscles, the heart can't pump the blood efficiently if this happens. The risk of suffering from heart disease would be higher for older people. Due to heart disease, about 4 out of 5 deaths can be occurred in the people whose age is above 65 years.

Sex: Men are generally at higher risk of having a seizure than women to the heart, but this difference is small if women started the phase of menopause. One of the feminine hormones, oestrogen, helps to protect the women from heart disease based on the researchers. The same cardiovascular risk is there for both women and men after 65 years of age.

This paper presents a risk factor analysis of heart disease. The goal is to analyse the causes of the heart attacks to individual patients. This would help the doctors treat individual patients with targeted approaches. This increases the chances of their survival.

2 Literature

Several researchers have attempted to solve the problem of heart disease classification using machine learning algorithms. Linear regression was used to analyse the heart diseases in [8, 9]. K-means clustering has been used in [10]. Authors used K-nearest neighbour for heart disease prediction in [11, 12]. Feature extraction-based techniques have been presented in [13–15]. Support vector machine (SVM) has been used in [11, 16]. Decision tree-based heart disease classification is used in [8, 11, 17, 18]. Deep learning techniques have been used in [18–22].

MontherTarawneh et al. [23] have proposed a new heart disease prediction system with the integration of all techniques into the single algorithm, known as hybridization. The accurate diagnosis results are achieved based on the combined model from all methods.

IlliasTougui et al. [24] have studied the comparison of six data mining tools based on ML techniques to predict the heart disease. The artificial neural network was proved as the best method with the use of MATLAB simulation tool.

Ching-seh Wu et al. [25] have studied different classifiers and effect of data processing methods based on the experimental analysis. The higher accuracy results were achieved with the Naïve Bayes and logistic regression based on random forest and decision tree algorithms.

Mohammad Shafenoor Amin et al. [26] have made researches on detection of significant features and data mining methods for improving the prediction accuracy of

cardiovascular disease. The accuracy of 87.4% in heart disease prediction is achieved using the best data mining technique, known as vote.

Nowbar et al. [27] have discussed the leading cause of death that is remained as IHD in all income groups of countries. The higher risk factors of cardiovascular issues have been contributed by the globalization in developing countries.

Bechthold, Angela, et al. [28] have showed the important lower risk of CHD, HF, and stroke with the optimal intake of vegetables, fruits, and nuts. The results proved that the key component of deriving the dietary food-based guidelines is used to prevent the CVD.

Kai Jin et al. [29] have described that the interventions of telehealth have the potential of improving the cardiovascular risk factors. The current evidence practice gap could be narrowed down to attend the centre-based cardiac rehabilitation.

Khandaker et al. [30] have been evident that the comorbidity between CHD and depression raised from the shared environmental factors. For CHD, the risk factors are CRP, IL-6, and triglycerides linked with depression causally.

Pencina et al. [31] have demonstrated the cardiovascular risk models and PAFs decrease with age. The potential and absolute risk reductions increase with age with respect to the treatment.

Latha and Jeeva [32] have investigated the ensemble classification method to enhance the prediction accuracy for heart disease. The detection of heart disease risk factors exhibits the satisfactory results based on the ensemble techniques like boosting and bagging.

Haoxiang et al. [33] presented a solution for data security optimal geometric transformation. Chen, et al., (2021) [34] presented a method for early detection of CAD for a higher accurate value. Table 1 shows a few literature analysis.

3 Proposed Method

The notion of data mining varies according to the literature. The definition of this concept can range from the extraction of patterns to the overall process of extraction of knowledge from data. Data mining is the process of extracting the knowledge from data. Its goal is to find patterns linking the data together, knowing that they must be interesting and for that new useful and nontrivial. The image regularly used is that of the mountain in which nuggets are buried. Data mining therefore consists in extracting them despite the vastness of the mountain.

The algorithmic approach aims to extract association rules. From a transaction basis, it first looks for frequent reasons that satisfy threshold conditions. Depending on the amount of data and the desired goals, many algorithms are studied, some of which are presented in this document. From these patterns, association rules are extracted, allowing links to be made enter the items that compose them. They are characterized by measures of interest, putting highlights certain aspects.

The proposed association rule mining approach is implemented on the dataset parameters to analyse the risk each patient has. The risk factor analysis is useful to

Table 1 Literature analysis

Ref. No.	Method	Description
Tarawneh and Embarak [23]	Hybridization	Heart disease prediction system with the integration of all techniques into the single algorithm, known as hybridization. The accurate diagnosis results are achieved based on the combined model from all methods
Ilias et al. [24]	Six machine learning techniques	The comparison of six data mining tools based on ML techniques to predict the heart disease. The artificial neural network was proved as the best method with the use of MATLAB simulation tool
Wu et al. [25]	The Naïve Bayes and logistic regression based on Random Forest and decision tree algorithms	Different classifiers and effect of data processing methods based on the experimental analysis. The higher accuracy results were achieved with the different machine learning techniques
Amin et al. [26]	Seven classification techniques	The detection of significant features and data mining methods for improving the prediction accuracy of cardiovascular disease. The accuracy of 87.4% in heart disease prediction is achieved using the best data mining technique, known as vote
Jin et al. [29]	Determine whether contemporary telehealth interventions	The current evidence practice gap could be narrowed down to attend the centre-based cardiac rehabilitation

identify the cause of heart attack to the patient. This approach can later be used to predict the factor that can cause heart attacks in individual patients. The proposed framework is implemented based on the rules of support confidence and lift.

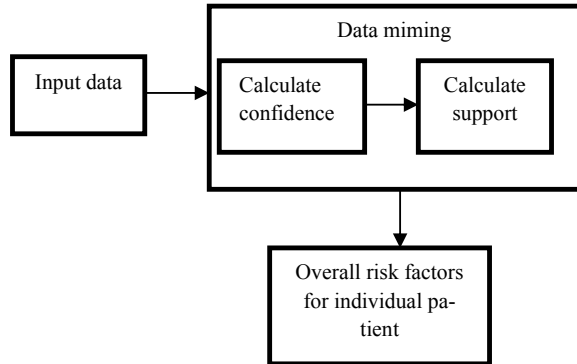
Fig. 1 Proposed system

Figure 1 shows the proposed system architecture. The input parameters are read from the database. Each individual parameter is taken for consideration, and the abnormal values in the list are selected. The confidence and support values for the remaining abnormal parameters are calculated, and risk assessment is performed.

Algorithm

Step 1: Read the input data.

Step 2: Calculate the confidence value with each individual parameter present in the dataset which have caused heart attack.

Step 3: Calculate the support value with each individual parameter present in the dataset which have caused heart attack.

Step 4: Calculate the overall risk factor by combining the confidence and support metric for individual patients.

Initially, two communities approached data mining differently. On the one hand were the supporters of information visualization, the aim of which was to give the user a general view of the data, while allowing a detailed view. On the other hand, the defenders of the algorithmic approach argued from the sufficiency of statistical and learning methods to find interesting patterns. Today, even if these two philosophies still exist, a third approach has emerged by the search for patterns by combining the visual approach and the algorithmic approach. There are four recommendations for the development of future knowledge research systems:

A mining is carried out automatically, manually, or by a combination of these two approaches.

3.1 Association Rule

A rule of association is defined, from an item set I , by the relationship in Eq. (1):

$$X \Rightarrow Y \quad (1)$$

where $X \cup Y = I$ and $X \cap Y = \emptyset$.

This can be translated as: “If X is present in the transaction, then Y is also”. Note that X and Y can be composed of several attributes, but an attribute cannot appear simultaneously in both parts of the rule. The left part of the ruler is called the *premise* or *antecedent* or *body*. The right part is called the *conclusion* or the *consequent* or the *head*. Support describes the probability that both the body and the head of the rule are in of a transaction as measured by all transactions. Support is a measure that describes the frequency with which a rule is in the database occurs. Both terms, support and coverage, describe exactly the same thing. But also want to provide a generally applicable definition of support in Eq. (2).

$$\text{sup}(A \Rightarrow B) = \frac{|\{t \in D | (A \cup B) \subseteq t\}|}{|D|} \quad (2)$$

where A is the body of the rule, B is the rule header, t is a transaction, and D is the transaction database. Since the support is a probability, it can have values between 0 and 1 take.

For example, the support usually shoes \rightarrow socks 2%. It can be seen that both shoes and socks come together in one transaction 20,000 times before. Measured against all transactions, here 1,000,000, that’s exactly 2%. Mathematically, it looks like this for the example described above:

$$\text{sup}(\text{shoes} \rightarrow \text{socks}) = p(\text{shoes} \cup \text{socks})$$

In the dataset, the support value identifies the combination of parameters that can cause heart attack. For each individual parameter, the remaining parameters which are abnormal are identified, and the support value is calculated.

The confidence of a rule is defined as the proportion of transactions that Rule body and rule header contain the amount of transactions that meet the rulebook.

The confidence c of the association rule $A \Rightarrow B$ is defined by Eq. (3):

$$c(A \Rightarrow B) = \frac{s(A \cup B)}{s(A)} \quad (3)$$

For our example, the confidence is 10%. That’s the 20,000 transactions in which both shoes and socks occur divided by the 200,000 transactions in which only appear in shoes. It is the conditional probability of A given B rule $A \rightarrow B$. The formal definition looks like in this Eq. (4):

$$\text{conf}(A \Rightarrow B) = \frac{|\{t \in D | (A \cup B) \subseteq t\}|}{|\{t \in D | A \subseteq t\}|} = \frac{\text{sup}(A \Rightarrow B)}{\text{sup}(A)}$$

Or

$$\text{conf}(A \Rightarrow B) = p(B|A) \tag{4}$$

As regards the basic measure characterizing an association rule, it can be assimilated the conditional probability $P(B|A)$, that is to say the probability of Y knowing X . The trust can therefore be written as follows Eq. (5):

$$c(A \Rightarrow B) = \frac{P(AB)}{P(A)} \tag{5}$$

Confidence is intended to measure the validity of an association rule. The confidence value presents a ratio of the intersection of individual parameters abnormal and the total number of abnormal cases.

4 Experimental Results

For experimental analysis, UCI heart repository has been used (<https://archive.ics.uci.edu/ml/datasets/heart+disease>). The repository has 13 parameters, namely Age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, and thal. The characteristics to demographic data, including behaviour, medical history, clinical examinations and are described as follows:

- Age in years
- Cp—type of chest pain
- Ca—number of major vessels (0–3) coloured based on fluoroscopy
- Exang—angina induced the exercise (1 = yes; 0 = no)
- Thalach—achieved maximum heart rate, > 100 abnormal
- Fbs—fasting blood sugar (> 120 mg/dl) (1 = true; 0 = false)
- Chol—cholesterol, > 200 abnormal
- Trestbps—resting blood pressure (in mm Hg) > 120 abnormal
- Sex—1 that indicates male and 0 indicates the female
- Thal—type of defect

Each attribute can fall under a specific category. The characteristics of which their values can take any value within a range called continuous. Table 2 shows the abnormal association value between cp and exang, thalach, restecg, fbs, chol, trestbps, thal. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

From Table 2, the association between cp, chol, and thalach has very high association. Thal also has good association with cp.

Table 3 shows the abnormal association value between trestbps and cp, ca, restecg, chol, thal, exang, restecg, thalach, and fbs. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 2 Abnormal association for cp

	Support (%)	Confidence (%)	Lift (%)
Trestbps	56.46	62.53	0.19
Chol	73.95	81.89	0.19
Fbs	11.22	12.42	0.17
Restecg	56.27	62.32	0.20
Thalach	89.73	99.37	0.19
Exang	11.79	13.05	0.18
Ca	18.82	20.84	0.19
Thal	76.24	84.42	0.19

Table 3 Abnormal association for trestbps

	Support (%)	Confidence (%)	Lift (%)
Cp	56.46	88.66	0.19
Chol	52.85	82.99	0.19
Fbs	10.46	16.42	0.23
Restecg	34.41	54.03	0.17
Thalach	63.69	100.00	0.19
Exang	7.98	12.54	0.18
Ca	15.59	24.48	0.22
Thal	54.56	85.67	0.20

Table 3 shows that trestbps has high confidence with thalach. The remaining parameters have not contributed to the occurrence of the heart attack in combination with this parameter,

Table 4 shows the abnormal association value between chol and cp, ca, fbs, thal, exang, restecg, trestbps, thalach, thal. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 4 Abnormal association for chol

	Support (%)	Confidence (%)	Lift (%)
Cp	73.95	91.10	0.19
Trestbps	52.85	65.11	0.19
Fbs	10.65	13.11	0.18
Restecg	44.11	54.33	0.17
Thalach	81.18	100.00	0.19
Exang	12.36	15.22	0.21
Ca	17.87	22.01	0.20
Thal	67.49	83.14	0.19

Table 5 shows the abnormal association value between fbs and cp, thal, ca, exang, restecg, thalach, chol, and trestbps. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 6 shows the abnormal association value between restecg and cp, thal, exang, thalach, chol, trestbps, chol, ca, and thal. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 7 shows the abnormal association value between thalach and cp, thal, chol, exang, trestbps, and ca. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 5 Abnormal association for fbs

	Support (%)	Confidence (%)	Lift (%)
Cp	11.22	83.10	0.17
Trestbps	10.46	77.46	0.23
Chol	10.65	78.87	0.18
Restecg	6.84	50.70	0.16
Thalach	12.93	95.77	0.18
Exang	1.14	8.45	0.12
Ca	4.75	35.21	0.32
Thal	10.08	74.65	0.17

Table 6 Abnormal association for restecg

	Support (%)	Confidence (%)	Lift (%)
Cp	56.27	94.87	0.20
Trestbps	34.41	58.01	0.17
Chol	44.11	74.36	0.17
Fbs	6.84	11.54	0.16
Thalach	58.75	99.04	0.19
Exang	6.65	11.22	0.16
Ca	13.88	23.40	0.21
Thal	47.91	80.77	0.19

Table 7 Abnormal association for thalach

	Support (%)	Confidence (%)	Lift (%)
Cp	89.73	90.25	0.19
Trestbps	63.69	64.05	0.19
Chol	81.18	81.64	0.19
Fbs	12.93	13.00	0.18
Restecg	58.75	59.08	0.19
Exang	13.50	13.58	0.19
Ca	21.10	21.22	0.19
Thal	82.32	82.79	0.19

Table 8 shows the abnormal association value between exang and cp, fbs, thalach, trestbps, thal, chol, and restecg. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 9 shows the abnormal association value between Ca and cp, thal, thalach, exang, restecg, chol, fbs, and trestbps. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Table 10 shows the abnormal association value between thal and cp, restecg, fbs, ca, exang, thalach, and chol. In the table, column 1 shows Support, column 2 shows Confidence, and column 3 shows the Lift.

Based on the Tables 2, 3, 4, 5, 6, 7, 8, 9, and 10, the relationship with the risk of getting heart attack due to individual parameter and the combination of parameters can be analysed. Thalach is the parameter with highest association will other parameters in the abnormal range. The second and third highest parameters are found out as CP, Chol, respectively.

Table 8 Abnormal association for exang

	Support (%)	Confidence (%)	Lift (%)
Cp	11.79	87.32	0.18
Trestbps	7.98	59.15	0.18
Chol	12.36	91.55	0.21
Fbs	1.14	8.45	0.12
Restecg	6.65	49.30	0.16
Thalach	13.50	100.00	0.19
Ca	3.04	22.54	0.20
Thal	9.89	73.24	0.17

Table 9 Abnormal association for ca

	Support (%)	Confidence (%)	Lift (%)
Cp	18.82	89.19	0.19
Trestbps	15.59	73.87	0.22
Chol	17.87	84.68	0.20
Fbs	4.75	22.52	0.32
Restecg	13.88	65.77	0.21
Thalach	21.10	100.00	0.19
Exang	3.04	14.41	0.20
Thal	16.73	79.28	0.18

Table 10 Abnormal association for thal

	Support (%)	Confidence (%)	Lift (%)
Cp	76.24	76.24	0.19
Trestbps	54.56	54.56	0.20
Chol	67.49	67.49	0.19
Fbs	10.08	10.08	0.17
Restecg	47.91	47.91	0.19
Thalach	82.32	82.32	0.19
Exang	9.89	9.89	0.17
Ca	16.73	16.73	0.18

5 Conclusion

Association rules can be applied to extract the relationship in the data attributes. Here the parameters responsible for heart attack are studied in detail. The risk factor analysis on the attributed can be performed with support and confidence values. Support describes the probability that both the body and the head of the rule are in of a transaction as measured by all transactions. The confidence of a rule is defined as the proportion of transactions that rule body and rule header contain the number of transactions that meet the rulebook. The experimental results show that the thalach is the parameters with highest association will other parameters in the abnormal range. The second and third highest parameters are found out as CP, Chol, respectively. The future scope would include targeted treatments for patients based on their individual risk factor analysis report. This would make the treatment more effective and increase their survival chances.

References

1. Cohen S, Liu A, Wang F, Guo L, Brophy JM, Abrahamowicz M, Therrien J et al (2021) Risk prediction models for heart failure admissions in adults with congenital heart disease. *Int J Cardiol* 322:149–157
2. Indrakumari R, Poongodi T, Jena SR (2020) Heart disease prediction using exploratory data analysis. *Procedia Comput Sci* 173:130–139
3. Nees SN, Chung WK (2020) The genetics of isolated congenital heart disease. *Am J Med Genet Part C Semin Med Genet* 184(1):97–106 (John, Hoboken)
4. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S (2019) A review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol* 234(10):16812–16823
5. Yang J, Tian S, Zhao J, Zhang W (2020) Exploring the mechanism of TCM formulae in the treatment of different types of coronary heart disease by network pharmacology and machining learning. *Pharmacol Res* 159:105034
6. Al-Omary MS, Sugito S, Boyle AJ, Sverdlov AL, Collins NJ (2020) Pulmonary hypertension due to left heart disease: diagnosis, pathophysiology, and therapy. *Hypertension* 75(6):1397–1408

7. Lutz M, Fuentes E, Ávila F, Alarcón M, Palomo I (2019) Roles of phenolic compounds in the reduction of risk factors of cardiovascular diseases. *Molecules* 24(2):366
8. Singh A, Kumar R (2020) Heart disease prediction using machine learning algorithms. In: 2020 international conference on electrical and electronics engineering (ICEE3). IEEE, pp 452–457
9. Nalluri S, Vijaya Saraswathi R, Ramasubbareddy S, Govinda K, Swetha E (2020) Chronic heart disease prediction using data mining techniques. In: *Data engineering and communication technology*. Springer, Singapore, pp. 903–912
10. Ripan RC, Sarker IH, Minhaz Hossain SM, Anwar MM, Nowrozy R, Hoque MM, Furhad MH (2021) A data-driven heart disease prediction model through K-means clustering-based anomaly detection. *SN Comput Sci* 2(2):1–12
11. Al-Yarimi FAM, Munassar NMA, Bamashmos MHM, Ali MYS (2021) Feature optimization by discrete weights for heart disease prediction using supervised learning. *Soft Comput* 25(3):1821–1831
12. Shah D, Patel S, Bharti SK (2020) Heart disease prediction using machine learning techniques. *SN Comput Sci* 1(6):1–6
13. Gárate-Escamila AK, El Hassani AH, Andrès E (2020) Classification models for heart disease prediction using feature selection and PCA. *Inf Med Unlocked* 19:100330
14. Sarkar BK (2020) Hybrid model for prediction of heart disease. *Soft Comput* 24(3):1903–1925
15. Kavitha M, Gnaneswar G, Dinesh R, Rohith Sai Y, Sai Suraj R (2021) Heart disease prediction using hybrid machine learning model. In: 2021 6th international conference on inventive computation technologies (ICICT). IEEE, pp 1329–1333
16. Diwakar M, Tripathi A, Joshi K, Memoria M, Singh P (2021) Latest trends on heart disease prediction using machine learning and image fusion. *Mater Today Proc* 37:3213–3218
17. Rani P, Kumar R, Sid Ahmed NMO, Jain A (2021) A decision support system for heart disease prediction based upon machine learning. *J Reliable Intell Environ* 1–13
18. Tasnim F, Habiba SU (2021) A comparative study on heart disease prediction using data mining techniques and feature selection. In: 2021 2nd international conference on robotics, electrical and signal processing techniques (ICREST). IEEE, pp 338–341
19. Ali F, El-Sappagh S, Islam SMR, Kwak D, Ali A, Imran M, Kwak K-S (2020) A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf Fusion* 63:208–222
20. Dutta A, Batabyal T, Basu M, Acton ST (2020) An efficient convolutional neural network for coronary heart disease prediction. *Expert Syst Appl* 159:113408
21. Mienye ID, Sun Y, Wang Z (2020) Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Inf Med Unlocked* 18:100307
22. Kirubakaran SS, Santhosh Kumar B, Rajeswari R, Daniya T (2021) Heart disease diagnosis systematic research using data mining and soft computing techniques. *Mater. Today: Proceedings*
23. Tarawneh M, Embarak O (2019) Hybrid approach for heart disease prediction using data mining techniques. In: *International conference on emerging internetworking, data & web technologies*. Springer, Cham, pp. 447–454
24. Ilias T, Jilbab A, El Mhamdi J (2020) Heart disease classification using data mining tools and machine learning techniques. *Heal Technol* 10:1137–1144
25. Wu C-s, Badshah MM, Bhagwat V (2019) Heart disease prediction using data mining techniques. In: *Proceedings of the 2019 2nd international conference on data science and information technology*, pp 7–11
26. Amin MS, Chiam YK, Varathan KD (2019) Identification of significant features and data mining techniques in predicting heart disease. *Telematics Inf* 36:82–93
27. Nowbar AN, Gitto M, Howard JP, Francis DP, Al-Lamee R (2019) Mortality from ischemic heart disease: analysis of data from the World Health Organization and coronary artery disease risk factors from NCD risk factor collaboration. *Circ Cardiovasc Qual Outcomes* 12(6):e005375
28. Angela B, Boeing H, Schwedhelm C, Hoffmann G, Knüppel S, Iqbal K, De Henauw S et al (2019) Food groups and risk of coronary heart disease, stroke and heart failure: a systematic review and dose-response meta-analysis of prospective studies. *Crit Rev Food Sci Nutr* 59(7):1071–1090

29. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, Briffa T, Bauman A, Redfern J, Neubeck L (2019) Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. *Eur J Cardiovasc Nurs* 18(4):260–271
30. Khandaker GM, Zuber V, Rees JMB, Carvalho L, Mason AM, Foley CN, Gkatzionis A, Jones PB, Burgess S (2020) Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort. *Mol Psychiatry* 25(7):1477–1486
31. Pencina MJ, Navar AM, Wojdyla D, Sanchez RJ, Khan I, Ellassal J, D'Agostino RB Sr, Peterson ED, Sniderman AD (2019) Quantifying importance of major risk factors for coronary heart disease. *Circulation* 139(13):1603–1611
32. Latha CBC, Jeeva SC (2019) Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inf Med Unlocked* 16:100203
33. Haoxiang W, Smys S (2021) Big data analysis and perturbation using data mining algorithm. *J Soft Comput Paradigm (JSCP)* 3(1): 19–28
34. Chen JIZ, Hengjinda P (2021) Early prediction of coronary artery disease (CAD) by machine learning method-a comparative study. *J Artif Intell* 3(1): 17–33