Ayatollahi et al. BMC Public Health (2019) 19:448
https://doi.org/10.1186/s12889-019-6721-5

BMC Public Health

RESEARCH ARTICLE Open Access

Predicting coronary artery disease: a comparison between two data mining algorithms

Haleh Ayatollahi[1], Leila Gholamhosseini[2,3*] and Masoud Salehi[4]

Abstract

Background: Cardiovascular diseases (CADs) are the first leading cause of death across the world. World Health Organization has estimated that morality rate caused by heart diseases will mount to 23 million cases by 2030. Hence, the use of data mining algorithms could be useful in predicting coronary artery diseases. Therefore, the present study aimed to compare the positive predictive value (PPV) of CAD using artificial neural network (ANN) and SVM algorithms and their distinction in terms of predicting CAD in the selected hospitals.

Methods: The present study was conducted by using data mining techniques. The research sample was the medical records of the patients with coronary artery disease who were hospitalized in three hospitals affiliated to AJA University of Medical Sciences between March 2016 and March 2017 ($n = 1324$). The dataset and the predicting variables used in this study was the same for both data mining techniques. Totally, 25 variables affecting CAD were selected and related data were extracted. After normalizing and cleaning the data, they were entered into SPSS (V23.0) and Excel 2013. Then, R 3.3.2 was used for statistical computing.

Results: The SVM model had lower MAPE (112.03), higher Hosmer-Lemeshow test's result (16.71), and higher sensitivity (92.23). Moreover, variables affecting CAD (74.42) yielded better goodness of fit in SVM model and provided more accurate result than the ANN model. On the other hand, since the area under the receiver operating characteristic (ROC) curve in the SVM algorithm was more than this area in ANN model, it could be concluded that SVM model had higher accuracy than the ANN model.

Conclusion: According to the results, the SVM algorithm presented higher accuracy and better performance than the ANN model and was characterized with higher power and sensitivity. Overall, it provided a better classification for the prediction of CAD. The use of other data mining algorithms are suggested to improve the positive predictive value of the disease prediction.

Keywords: Coronary artery disease (CAD), Data mining algorithms, Artificial neural network (ANN), Support vector machine (SVM)

Background

Cardiovascular diseases are among the common diseases in both developed and developing countries and regarded as the main cause of death throughout the world [1]. In fact, any condition or disease that affects the heart, its vessels [2], and the blood circulatory system can be related to coronary vascular diseases (CVDs) [3]. In general, the clinical spectrum of CVDs ranges from asymptomatic ischemia to chronic stable angina pectoris, unstable angina (UA), acute myocardial infarction (AMI), ischemic cardiomyopathy and sudden death [4]. They are sometimes associated with conditions such as hypertension, stroke, coronary artery diseases, chronic heart failure, congenital heart disease, rhythm disorders, subclinical atherosclerosis, valvular disease, and peripheral arterial disease [5]. In recent years, in addition to

* Correspondence: gholamhosseini.l@tak.iums.ac.ir
[2]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran
[3]School of Paramedical Sciences, AJA University of Medical Sciences, Tehran, Iran
Full list of author information is available at the end of the article

the main risk factors, other factors such as infection, inflammatory and chronic diseases have been discussed as other risk factors of cardiovascular diseases [6].

At the beginning of the twentieth century, 10% of all the deaths were attributed to cardiovascular diseases. At the end of this century, the mortality caused by CVDs increased to 25%. It is estimated that, considering the present increasing trend, over 35–60% of deaths worldwide would be due to cardiovascular diseases by 2025 [7]. Based on the report by WHO, in 2017, more than half (54%) of the deaths around the world were caused by 10 leading causes, and cardiovascular diseases which led to 15 million deaths in 2015 constituted the largest group of fatal diseases [8]. Cardiovascular diseases kill millions of people annually and this value may be increased up to 24.8 million by 2020 if preventive measures are not taken [9].

In Iran, the Ministry of Health reported that 39.9% of the mortality rate in the country is due to cardiovascular diseases and their risk factors, among them CAD is the most prevalent type and is greatly increasing [10]. CAD is a multi-causal disease, in which a series of risk factors, e.g. increased cholesterol, hypertension, diabetes and smoking should be taken into account [11]. According to the results of an epidemiological study with the aim of examining coronary artery disease mortality rate, 63 out of 6537 death cases were due to CAD in 2015 [12]. CAD is more prevalent among men than women, and the symptoms of the disease may appear in women 10 years later than in men [13]. Therefore, considering the great increase in cardiovascular diseases which imposes a heavy financial burden on the society, medical communities attempt to find a way for the accurate and timely prediction of CAD by using new statistical techniques, such as data mining [14]. It is noteworthy that the healthcare domain is filled with data. However, the data required for effective decision-making and discovery of hidden patterns are not extracted. By extracting useful data and discovering knowledge from the large volume of medical data, the causes of incidence, growth or the spread of diseases can be identified and physicians can be equipped with valuable information for better decision making. Therefore, many healthcare centers are seeking practical solutions for knowledge discovery by means of data mining techniques [15]. These techniques can help to recognize the patterns and factors influencing diseases [16].

The novel science of data mining is among the 10 developing sciences which have made the next decade face enormous technological evolutions. Using specialized knowledge, it will have extensive applications in the domain of medicine. [17, 18]

The literature review showed that different algorithms such as clustering, classifications, regression and association rules, decision trees, Bayesian network, neural network, multi-layer perceptron with error back propagation algorithm, scaled conjugate gradient (SCG) and support vector machine (SVM) have been used for predicting CAD [19–31]. However, the comparison between the algorithms has not received adequate attention. Among these algorithms, artificial neural network has some advantages, such as high speed, simplicity and capability of solving complex relationships between variables and extracting the non-linear relationships by means of training data. Another algorithm is support vector machine which is the most common and effective machine learning algorithm. SVMs have a powerful theoretical background that used in different activities, such as classification, recognition and prediction in supervised learning [32, 33]. Therefore, the present study aimed to compare the PPV of CAD using artificial neural network (ANN) and SVM algorithms and their distinction in terms of predicting CAD in the selected hospitals.

Methods

Study design and setting

The present research was conducted using data mining techniques. The research setting was three selected hospitals affiliated to AJA University of Medical Sciences.

Participants and sampling

In this study, only medical records of patients with coronary artery disease who were hospitalized in three teaching hospitals between March 2016 and March 2017 were used ($n = 1324$). Other diseases, such as arrhythmia, angina pectoris, acute myocardial infarction, chronic rheumatic heart diseases, congenital heart disease, dilated cardiomyopathy, heart failure, hypertrophic cardiomyopathy, hypertensive heart diseases, ischemic heart diseases, myocardial infarction, mitral regurgitation, mitral valve prolapse, pulmonary stenosis, and pulmonary heart disease were excluded. A unique dataset including the same CAD predicting variables was used for both SVM and ANN techniques.

Instruments

The data collection instrument was a checklist designed based on the variables used in the guideline of the Cleveland heart disease dataset policy in UCI (University of California) repository. [34] The checklist included 25 variables for predicting CAD. These variables were gender, age, weight, marital status, occupation, address, family history, smoking, comorbidity, diabetes, pulse rate, T.S.T waves, high blood pressure (HBP), cholesterol, triglyceride (TG), hemoglobin (Hgb), blood glucose level, creatinine, systolic blood pressure, diastolic blood pressure, chest pain, low density lipoprotein (LDL), high

density lipoprotein (HDL), CAD diagnosis, and the length of hospitalization.

The collected data were controlled by using different methods, such as data preparation, integration, cleaning, normalization and reduction.

Statistical analysis

After normalization, processing and cleansing, data were entered into SPSS (V23.0) and Microsoft Excel 2013. Moreover, R 3.3.2 was used for statistical computing. The dataset was divided into training and testing sets and to do so, the standard randomized allocation method was used. Consequently, 70% of the records was used for training and 30% was used for testing the models.

Ethical consideration

The study protocol was approved by the Ethical Clearance Committee of AJA University of Medical Sciences. The data were used anonymously and were kept confidential.

Results

Initially, the research variables were analyzed in each hospital separately. The results showed that the majority of the patients were men ($n = 829$, 62.7%) with the mean age of 54–62 years old. The rest of the patients were women ($n = 494$, 37.3%) with the mean age of 61–64 years old. The weight comparison between the patients showed that there was a significant difference between CAD and mean weight in Hospitals A and B.

Socio-demographic predictors of CAD

The frequency distribution of CAD and the type of occupation showed that there was a significant difference between having and not having the occupation. In addition, the frequency distribution of CAD and the place of residence suggested that the majority of the patients ($n = 1082$, 98%) resided in cities. Similarly, the frequency distribution of CAD and a family history indicated that there was a significant difference between having and not having a family history of CAD ($n = 1049$, 79.3%) ($p < 0.001$). Moreover, the results showed that 77.3% ($n = 1024$) of patients were non-smokers and there was a significant difference ($p < 0.001$) among the hospitals in terms of smoking and CAD.

The predicting variables

The main objective of this study was to determine the PPV of CAD using ANN algorithm and compare the results with the results of the SVM model. Therefore, 25 predicting variables were extracted from the database of cardiovascular patients in the selected hospitals and were used as the input variables and the weight of each was calculated by running algorithms in order to fit the multi-layer ANN model (Fig.1). Based on the calculated weights, the following variables were selected as CAD predicting variables: gender, occupation, place of residence, family history, smoking status, comorbidity, mean value of pulse rate, TST waves status, hypertension history, chest pain, cholesterol, triglyceride, blood glucose level and creatinine level.

In the present study, 70% of the data was used for training and 30% was used for testing the ANN model. The results revealed that the goodness of fit was appropriate in ANN model with the PPV (Available from: https://en.wikipedia.org/wiki/Positive_and_negative_predictive_values) of 0.798, the smaller mean squared error (MSE) and relative error in the test dataset (Table 1).

$$\mathrm{PPV} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false positives}} = \frac{\text{number of true positives}}{\text{number of positive calls}}$$

Figure 2 illustrates the receiver operating characteristic (ROC) curve for CAD patients. The PPV of the model depends on the extent, to which the test has correctly distinguished CAD patients (sensitivity). This PPV is calculated by computing the area under the ROC curve. The closer this value is to 1, the higher the PPV of the model. Moreover, the closer the value of this ratio is to the left corner, the larger the area under the curve would be. The results showed that the ANN model had high PPV when predicting CAD.

The PPV was measured by using the SVM algorithm (Fig. 3).

In this phase, 70% of the data were considered as training data and the remaining 30% was used as test data to run the SVM algorithm. Then, PPV of the model is presented in Table 2.

$$\text{Cohen}'\text{s kappa coefficient} = (\text{Accuracy-expAccuracy})/(\text{1-expAccuracy})$$

$$\text{F-measure} = 2 * ((\text{PPV} * \text{Recall})/(\text{PPV} + \text{Recall}))$$

As Table 2 shows, F-measure and Cohen's Kappa coefficients were used to determine the PPV of the SVM model. The result showed that the SVM model had a moderate to high power and sensitivity for predicting CAD patients. Moreover, the SVM model had higher PPV in classifying and predicting CAD. Furthermore, comparison between the accuracy indices showed that, the SVM model had higher accuracy compared to the ANN model and presented better classification (Table 2).

That the findings also showed that the area under the ROC curve was larger in the SVM model than in the

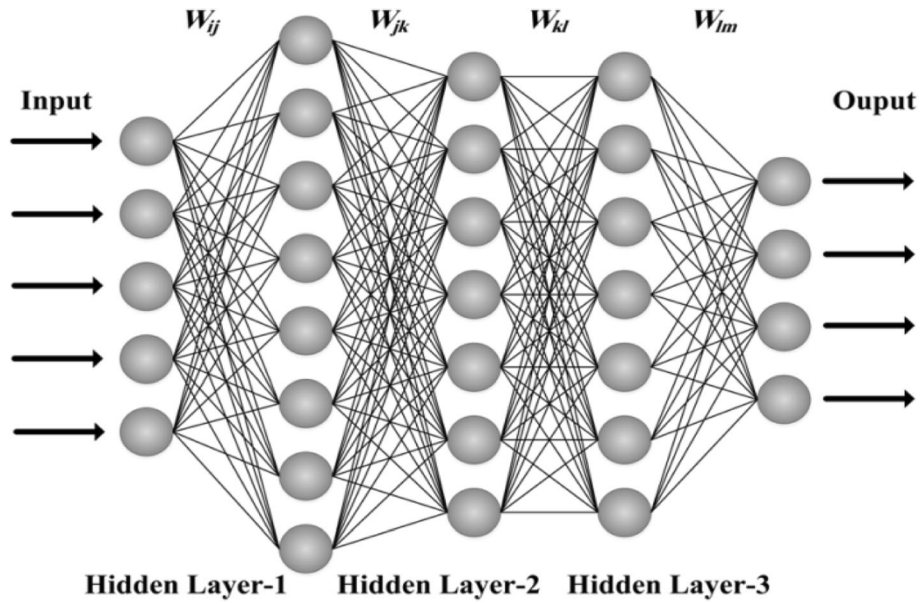

Fig. 1 Multi-Layer Perceptron in ANN

ANN model (Fig. 4). As a result, SVM had better performance in predicting patients with CAD.

Other statistical tests which were used to compare the performance of the ANN and SVM algorithms were Hosmer-Lemeshow goodness-of-fit test, MAPE (mean absolute percentage error), sensitivity and specificity indices. The tests' results are presented in Table 3.

According to the results, the smaller value of MAPE in the SVM model indicated better fitness of data with less error. Moreover, the larger value of Hosmer-Lemeshow goodness-of-fit test showed the superiority of the SVM model. Furthermore, the SVM model was superior to the ANN model in terms of sensitivity and specificity.

Discussion

Based on the results, the most important factors affecting the incidence of CAD were gender, occupation, family history, smoking, co-morbidity, mean value of heart rate, TST wave status, hypertension, chest pain, cholesterol, triglyceride, blood glucose level and creatinine. Similarly, previous studies introduced numerous factors affecting the disease and the progress of cardiovascular diseases. These factors were divided into six general groups: environmental factors, daily habits, risk factors, underlying diseases, mental-personality factors and social factors [35].

Table 1 PPV indices in ANN algorithm

	Sample	MSE	Relative error	Positive Predictive Value
Training	70%	5.39	0.002	0.798
Testing	30%	3.84	0.002	

Other common risk factors associated with CAD include hypertension, lifestyle [36], high level of cholesterol [37], diabetes [38], obesity [39] and smoking [40].

The results of the present study showed that the incidence of the disease was higher in men than women, and the risk of CAD could increase by an increase in age and weight. Similarly, according to another study, age, gender (male) and smoking had significant correlations with CAD [41]. In the study conducted by Masethe and Masethe, a system was proposed for predicting heart attack and included the variables of gender, age, type of chest pain, heart rate, cholesterol, smoking, blood glucose level, blood pressure, diet and alcohol consumption [42].

The findings revealed that the risk of CAD was higher among the employed patients compared to the unemployed and the retired ones. Similarly, the results of a cohort study represented that the risk of cardiovascular diseases was about 40%, because of job strain, and an increase in work load doubled the risk of these diseases. Therefore, the type of job can be a risk factor for cardiovascular diseases [43]. Moreover, the incidence of the disease was higher among those who were living in the urban than the rural areas. In another study, kermani et al. examined the relationship between the mortality rate caused by cardiovascular and chronic obstructive pulmonary diseases (COPD) due to nitrogen dioxide air pollutants in Tehran and reported a significant relationship between these risk factors [44]. Another study investigated the relationship between spatial dispersion of particulate matter and mortality among patients with cardiovascular diseases in Beijing and reported that an

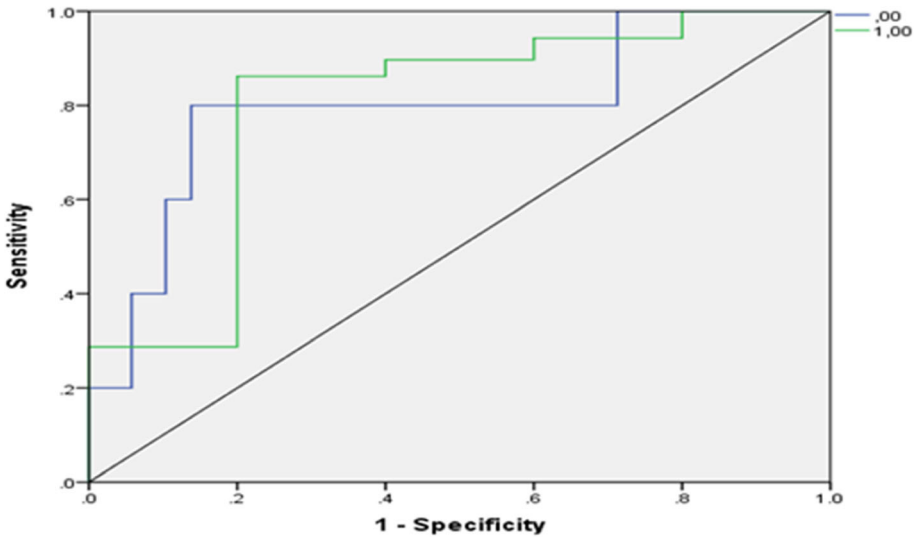

Fig. 2 The ROC curve in ANN

increase in particulate matter increased the rate of death among those residing in cities [45]. In another project, researchers evaluated the risk of death by air pollution in 10 cities in Canada and found that there were significant relationships between mortality among patients with cardiovascular respiratory diseases, urban residence and urban air pollutants [46]. However, the results of another study showed that cardiovascular programs have not been implemented in the rural areas; therefore, the mortality rate caused by cardiovascular diseases were increased in the rural areas compared to the big cities [47].

The results also revealed that there was a significant difference between family history and CAD. 193(20.2%) and 215(22.5%) patients had paternal and maternal positive family history (father, mother and siblings) of CAD; there was a possibility to be diagnosed with the disease before 55 and 65 years old in men and women, respectively [48]. As mentioned before, family history of the disease and other risk factors such as blood glucose level, HDL, LDL, cholesterol, systolic and diastolic blood pressure as well as age and gender have been highlighted in the literature [49].

According to the results of the present study, only 58 of the participants were smokers and 142 were non-smokers. The results of a Chi-squared (X^2) test showed that mean plasma levels of NO was significantly lower in smoker patients ($P = 0.004$). According to the literature, smoking has an increasing trend in Asian countries compared to the rest of the world [50].

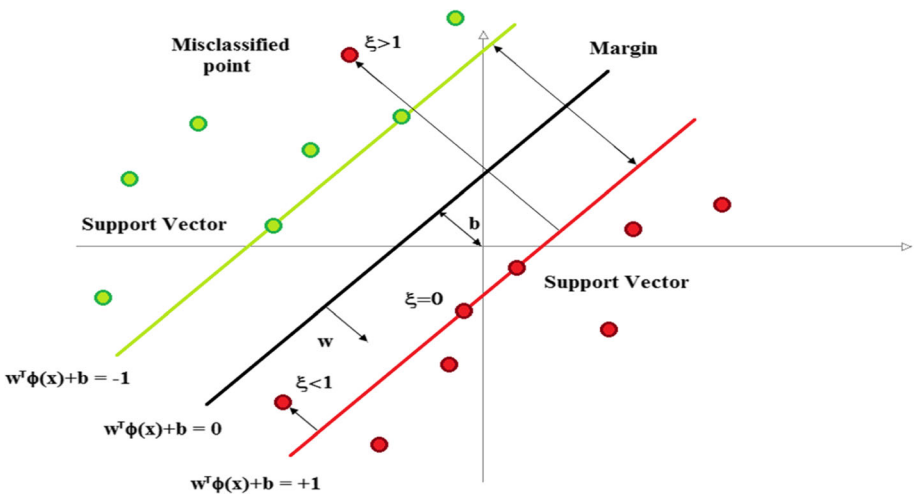

Fig. 3 An overview of the SVM model

Table 2 PPV indices in the SVM algorithm

	Sample	F-measure	Kappa coefficient	Positive Predictive Value
Training	70%	0.761	0.706	0.871
Testing	30%	0.696	0.636	

Table 3 Comparison between the ANN and SVM algorithms

Test	ANN	SVM
MAPE	125.17	112.03
Hosmer-Lemeshow	12.4	16.71
Sensitivity	88.01	92.32
Specificity	73.64	74.42

Similarly, another study showed that obesity, family history, co-morbidities and smoking can increase the risk of CAD [51]. As smoking is a strong and independent risk factor for cardiovascular diseases, all patients with these diseases must stop smoking [52]. Doctors emphasize that the risk of CAD can be considerably reduced in future by limiting smoking. Therefore, the status of smoking must be systematically evaluated in patients with cardiovascular diseases [39]. According to the results of a hospital-based observational study, there is a direct association between the smoking status and CAD among the young adults. In general, the incidence of CAD had a higher mean value among smokers and the age of patients was lower than or equal to 35 years old [53].

According to the results, 28.6% of the patients had one or multiple co-morbidities. In another study, the results showed that patients with ischemic heart disease (IHD) and chronic obstructive pulmonary disease had the most severe complications compared to those with only one of the noted diseases [54]. Furthermore, according to the results of other studies; obesity, hypertension, diabetes mellitus, metabolic syndrome, high levels of LDL, low level of HDL, high fat diet, lack of regular exercise and dyslipidemia are the risk factors for the mentioned diseases [55, 56].

In terms of the relationship between the mean value of heart rate and the incidence of CAD, a significant relationship was seen which showed, the risk of CAD increases by increasing the mean value of heart rate. The findings of the present study indicated that only 45% of patients had abnormal TST waves and there was no relationship between TST waves' status and CAD. In a research on the diagnosis of ventricular cardiomyopathy using ANN algorithms, the results showed that a reduction in the dimensions of cardiac signals had a positive effect on the cardiac sound classification [57].

Another finding of the current study was related to the relationship between the incidence of CAD and level of triglyceride and creatinine. In fact, the risk of CAD could increase due to an increase in these variables. Moreover, a significant relationship was seen between the chest pain and CAD [39]. However, the results of the study conducted by wertli et al. showed that there was no significant relationship between these variables. The chest pain has a subjective nature which cannot be used for predicting CAD and panic disorders should be considered in recognizing types of chest pain. [58].

According to the literature review, numerous studies have been conducted to predict CAD by using data mining algorithms. For instance, Kurt et al. used logistic regression, decision trees, classification and neural networks and finally, the multi-layer perceptron ANN with the PPV of 78.8% was introduced as the best model [59]. In another study, Sajja employed a simple Bayesian

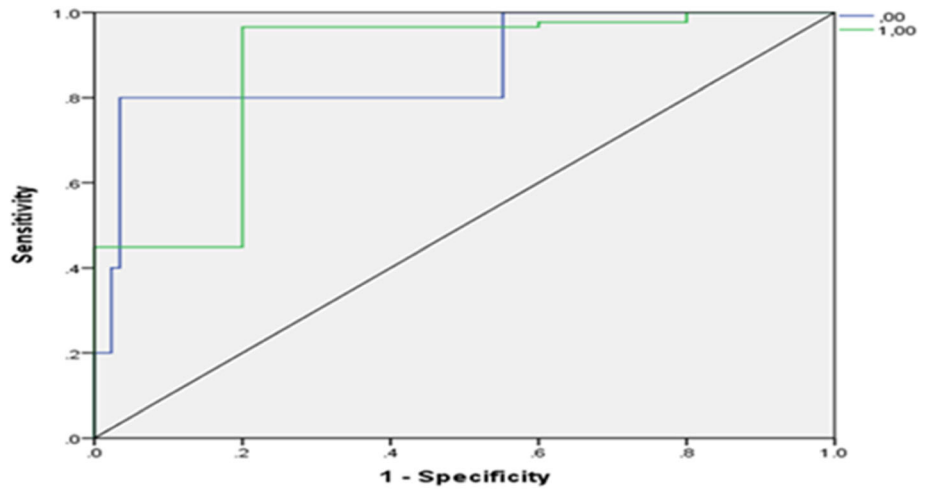

Fig. 4 The ROC curve in SVM

algorithm, decision tree and multi-layer perceptron ANN on a dataset. The results showed that the precision of the multi-layer perceptron ANN algorithm was 91.75, indicating the best performance [60]. In the present study, CAD was selected as the output variable and 25 variables were used as input variables. The results showed that the ANN model could be appropriate for fitting these data with the total PPV of 0.798. On the other hand, the SVM algorithm fitted the data with smaller MAPE and error. The larger value of Hosmer-Lemeshow goodness-of-fit test also showed the superior performance of the SVM model on the data and provided better prediction for CAD diagnosis. Furthermore, the SVM algorithm predicted CAD patients with higher PPV and sensitivity than the ANN model.

Similarly, the results of the previous studies showed that the use of the SVM algorithm predicts the disease and distinguishes patients from non-patients with higher accuracy [61–63]. Other studies have also confirmed the superior performance and precision of SVM. Nevertheless, there are few studies which do not confirm the efficiency of this algorithm and suggest other methods, such as binary particle swarm optimization (BPSO) and genetic algorithm as the best model of choice for CAD determination [64]. Although input variables were selected based on the literature review and related guidelines, there might be other risk factors which can be studied in the future to gain a bigger picture of the disease risk factors. Moreover, in this study, the results of two algorithms were compared. The data can be used to test other algorithms, such as genetic algorithm to recognize the best performance model.

Conclusion

The process of disease prediction in medical sciences is as an important process for decision-making and physicians need to know the risk factors for different diseases. This process can be facilitated by using logical and purposeful methods, such as machine learning methods and data mining algorithms. Currently, due to the considerable increase in cardiovascular diseases and the heavy financial burden imposed by them on the society, healthcare communities are seeking ways to predict, diagnose, and treat these diseases effectively. The results of the current study showed that the use of data mining algorithms, such as the SVM model can be useful in predicting CAD. However, more research is needed to compare the performance of different algorithms and to find the best performance model.

Abbreviations

AMI: Acute Myocardial Infarction; ANN: Artificial Neural Network; BPSO: Binary Particle Swarm Optimization; CAD: Coronary Artery Disease; COPD: Chronic Obstructive Pulmonary Disease; IHD: Ischemic Heart Disease; MAPE: Mean Absolute Percentage Error; MSE: Mean Squared Error; SCG: Scaled Conjugate Gradient; SVM: Support Vector Machine; UA: Unstable Angina

Acknowledgements

The authors gratefully acknowledge AJA University of Medical Sciences for funding this study.

Funding

This study was funded and supported by AJA University of Medical Sciences Grant (AJAUMS-594273/2016). The process of study design and data collection, analysis, and interpretation were completed by the researchers.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly accessible due to the data confidentiality principals related to the patients admitted to the military hospitals, but are available from the corresponding author on a reasonable request.

Authors' contributions

LG conceived and designed the study. LG and HA drafted the manuscript. HA and MS participated in the critical review of the manuscript. The authors declared their final approval of the version of the manuscript submitted for publication.

Ethics approval and consent to participate

This study was granted ethics approval by the Ethics Committee of AJA University of Medical Sciences. Only medical records' numbers were used and all other identifiable information were removed before data storage and reporting.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran. [2]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran. [3]School of Paramedical Sciences, AJA University of Medical Sciences, Tehran, Iran. [4]Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran.

Received: 29 September 2018 Accepted: 28 March 2019
Published online: 29 April 2019

References

1. Healthy environment, healthy heart (Internet). Jakarta: Ministry of Health, Republic of Indonesia; 2014. Available at: http://www.depkes.go.id/article/view/201410080002/lingkungan-sehat-jantung-sehat.html.
2. Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, Budaj A, Bugiardini R, Crea F, Cuisset T, Di Mario C. 2013 ESC guidelines on the management of stable coronary artery disease: the task force on the management of stable coronary artery disease of the European society of cardiology. Euro Heart J. 2013;34(38):2949–3003.
3. Kelly BB, Fuster V, editors. Promoting cardiovascular health in the developing world: a critical challenge to achieve global health. National Academies Press; 2010.
4. Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. Ann Transl Med. 2016;4(13):256.
5. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, De Ferranti S, Després JP, Fullerton HJ, Howard VJ. Executive summary: heart disease and stroke statistics-2016 update: a report from the American Heart Association. Circulation. 2016;133(4):447–54.

6. Bergh C, Fall K, Udumyan R, Sjöqvist H, Fröbert O, Montgomery S. Severe infections and subsequent delayed cardiovascular disease. Eur J Prev Cardiol. 2017;24(18):1958–66.
7. Longo D, Fauci A, Kasper D, Hauser S. Harrison's principles of internal medicine. 18th ed. New York: McGraw-Hill Professionals; 2011.
8. WHO, No.310 Fs. The top ten causes of death. Geneva: World Health Organization; 2017. (cited 2018 Feb 16). Available at: http://www.who.int/mediacentre/factsheets/fs310/en/
9. World Health Organization. Physical activity and older adults. 2018 (cited 2018 Mar 4) Available at: https://www.who.int/dietphysicalactivity/factsheet_olderadults/en/.
10. Mann DL, Zipes DP, Libby P, Bonow RO. Braunwald's heart disease e-book: a textbook of cardiovascular medicine. Elsevier health sciences. 2014;2136:861–71.
11. U.S. National Library of medicine (NLM). Coronary artery disease (CAD) medlineplus (trusted health information for you); 2017.
12. Kamiya K, Masuda T, Tanaka S, Hamazaki N, Matsue Y, Mezzani A, et al. Quadriceps strength as a predictor of mortality in coronary artery disease. Am J Med. 2015;128(11):1212–9.
13. Charchar FJ, Bloomer LD, Barnes TA, Cowley MJ, Nelson CP, Wang Y, Denniff M, Debiec R, Christofidou P, Nankervis S, Dominiczak AF. Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. Lancet. 2012;379(9819):915–22.
14. Rezaei-hachesu P, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthc Inform Res. 2013;19(2):121–9.
15. Sudhakar K, Manimekalai DM. Study of heart disease prediction using data mining. Int Adv Res Comput Sci Soft Eng. 2014;4(1):1157-1160.
16. Yeh DY, Cheng CH, Chen YW. A predictive model for cerebrovascular disease using data mining. Expert Syst Appl. 2011;38(7):8970–7.
17. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. Wiley interdisciplinary reviews: WIREs data mining and knowledge discovery. 2011;1(5):416–30.
18. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. J Appl Biomed. 2013;11:47–58.
19. Sivagowry S, Durairaj M, Persia A. An empirical study on applying data mining techniques for the analysis and prediction of heart disease. Information communication and embedded systems (ICICES), international conference on 2013 Feb 21;265–270.
20. Sufi F, Khalil I. Diagnosis of cardiovascular abnormalities from compressed ECG: a data mining-based approach. IEEE Trans Inf Technol Biomed. 2011; 15(1):33–9.
21. Amin SU, Agarwal K, Beg R. Genetic neural network based data mining in prediction of heart disease using risk factors. In: Information & Communication Technologies (ICT), IEEE conference on 2013. https://doi.org/10.1109/CICT.2013.6558288.
22. Desai SD, Giraddi S, Narayankar P, Pudakalakatti NR, Sulegaon S. Back-propagation neural network versus logistic regression in heart disease classification. J adv comput commun technol. 2019. https://doi.org/10.1007/978-981-13-0680-8_13.
23. Kausar N, Abdullah A, Samir BB, Palaniappan S, BS AG, Dey N. Ensemble clustering algorithm with supervised classification of clinical data for early diagnosis of coronary artery disease. J Med Imaging Health Inform. 2016; 6(1):78–87.
24. Abawajy JH, Kelarev AV, Chowdhury M. Multistage approach for clustering and classification of ECG data. Comput Methods Prog Biomed. 2013;112(3):720–30.
25. Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Guo Y, Zhang H, Gao Z, Yan X. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artifi Intell Med. 2010; 48(2–3):139–52.
26. Guner LA, Karabacak NI, Akdemir OU, Karagoz PS, Kocaman SA, Cengel A, Unlu M. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion SPECT. J Nucl Cardiol. 2010;17(3):405–13.
27. Orphanou K, Stassopoulou A, Keravnou E. DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis. IEEE J Biomed Health Inform. 2016;20(3):944–52.
28. Kim J, Lee J, Lee Y. Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree. Healthc Inform Res. 2015;21(3):167–74.
29. Karaolis MA, Moutiris JA, Hadjipanayi D, Pattichis CS. Assessment of the risk factors of coronary heart events based on data mining with decision trees. IEEE Trans Inf Technol Biomed. 2010;14(3):559–66.
30. Verma L, Srivastava S, Negi PC. A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. J Med Syst. 2016;40(7):178.
31. Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert Syst Appl. 2009;36(4):7675–80.
32. Acharya UR, Fujita H, Lih OS, Adam M, Tan JH, Chua CK. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. Knowl Based Syst. 2017;132:62–71.
33. Dolatabadi AD, Khadem SE, Asl BM. Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Comput Methods Prog Biomed. 2017;138:117–26.
34. Janosi A, Steinbrunn W, Pfisterer M, Detrano R, Aha WD. UCI machine learning repository. Heart disease data set, 1988. (Cited 2018 Jun 15) Available at: http://archive.ics.uci.edu/ml/datasets/heart+disease.
35. Steenman M, Lande G. Cardiac aging and heart disease in humans. Biophys Rev. 2017;9(2):131–7.
36. Bayturan O, Kapadia S, Nicholls SJ, Tuzcu EM, Shao M, Uno K, Shreevatsa A, Lavoie AJ, Wolski K, Schoenhagen P. Clinical predictors of plaque progression despite very low levels of low-density lipoprotein cholesterol. J of the American Col of Cardio. 2010;55(24):2736–42.
37. Nicholls SJ, Hsu A, Wolski K, Hu B, Bayturan O, Lavoie A, et al. Intravascular ultrasound-derived measures of coronary atherosclerotic plaque burden and clinical outcome. J Am Coll Cardiol. 2010;55(21):2399–407.
38. Murthy VL, Naya M, Foster CR, Gaber M, Hainer J, Klein J, et al. Association between coronary vascular dysfunction and cardiac mortality in patients with and without diabetes mellitus. Circulation. 2012;126(15):1858–68 CIRCULATIONAHA. 112.
39. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, et al. European Association for Cardiovascular Prevention & rehabilitation (EACPR); ESC Committee for practice guidelines (CPG). European guidelines on cardiovascular disease prevention in clinical practice (ver 2012). The fifth joint task force of the European Society of Cardiology and Other Societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Euro Heart J. 2012; 33(13):1635–701.
40. Frey P, Waters DD, DeMicco DA, Breazna A, Samuels L, Pipe A, et al. Impact of smoking on cardiovascular events in patients with coronary disease receiving contemporary medical therapy (from the treating to new targets (TNT) and the incremental decrease in end points through aggressive lipid lowering (IDEAL) trials). Am J Cardiol. 2011;107(2):145–50.
41. Yiu KH, de Graaf FR, Schuijf JD, van Werkhoven JM, Marsan NA, Veltman CE, de Roos A, Pazhenkottil A, Kroft LJ, Boersma E, Herzog B. Age-and gender-specific differences in the prognostic value of CT coronary angiography. Heart. 2012;98(3):232–7.
42. Masethe HD, Masethe MA. Prediction of heart disease using classification algorithms. In: Proceedings of the world congress on engineering and computer science. San Fransico: WCECS; 2014. p. 22–4.
43. Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. Lancet. 2012;380(9852):1491–7.
44. Kermani M, Jonidi Jaffar A, Dowlati M, Rezaei Kalantari R. Number of total mortality, cardiovascular mortality and chronic obstructive pulmonary disease due to exposure with nitrogen dioxide in Tehran during 2005-2014. Urmia Med J. 2017;28(4):22.
45. Zhang Q, Cheng L, Boutaba R. Cloud computing: state-of-the-art and research challenges. J Internet Ser Appl. 2010;1(1):7–18.
46. Vanos JK, Hebbern C, Cakmak S. Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. Environ Pollut. 2014;185:322–32.
47. Garcia MC, Faul M, Massetti G, Thomas CC, Hong Y, Bauer UE, Iademarco MF. Reducing potentially excess deaths from the five leading causes of death in the rural United States. MMWR Surveill Summ. 2017;66(2):1.
48. Hoseini K, Sadeghian S, Mahmoudian M, Hamidian R, Abbasi A. Family history of cardiovascular disease as a risk factor for coronary artery disease in adult offspring. Monaldi Arch Chest Dis. 2016;70(2):84-87.
49. Siervo M, Lara J, Chowdhury S, Ashor A, Oggioni C, Mathers JC. Effects of the dietary approach to stop hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. Br J Nutr. 2015;113(1):1–5.
50. Mahmoodi K, Nasehi L, Karami E, Soltanpour MS. Association of nitric oxide levels and endothelial nitric oxide synthase G894T polymorphism with coronary artery disease in the Iranian population. Vasc Specialist Int. 2016; 32(3):105.

51. Andria N, Nassar A, Kusniec F, Ghanim D, Qarawani D, Kachel E, et al. Ethnicity of symptomatic coronary artery disease referred for coronary angiography in the galilee: prevalence, risk factors, and a case for screening and modification. Isr Med Assoc J. 2018;20(3):182–5.
52. Meyers DG, Neuberger JS, He J. Cardiovascular effect of bans on smoking in public places: a systematic review and meta-analysis. J Am Coll Cardiol. 2009;54(14):1249–55.
53. Lv S, Liu W, Zhou Y, Liu Y, Shi D, Zhao Y, Liu X. Hyperuricemia and smoking in young adults suspected of coronary artery disease [less than or equal to] 35 years of age: a hospital-based observational study. BMC Cardiovasc Disord. 2018;18(1):178.
54. Campo G, Pavasini R, Malagù M, Mascetti S, Biscaglia S, Ceconi C, et al. Chronic obstructive pulmonary disease and ischemic heart disease comorbidity: overview of mechanisms and clinical management. Cardiovasc Drugs Ther. 2015; 29(2):147–57.
55. Jahangir E, De Schutter A, Lavie CJ. The relationship between obesity and coronary artery disease. Transl Res. 2014;164(4):336–44.
56. Rairikar A, Kulkarni V, Sabale V, Kale H, Lamgunde A. Heart disease prediction using data mining techniques. In Intelligent computing and control (I2C2), IEEE international conference on 2017 Jun: 1–8.
57. Uğuz H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. J Med Syst. 2012; 36(1):61–72.
58. Wertli MM, Ruchti KB, Steurer J, Held U. Diagnostic indicators of non-cardiovascular chest pain: a systematic review and meta-analysis. BMC Med. 2013;11(1):239.
59. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl. 2008;34(1):366–74.
60. Sajja S. Data mining of medical datasets with missing attributes from different sources (PhD thesis), Youngstown State University; 2010. Available at: https://etd.ohiolink.edu/pg_6?::NO::: Cited 2018 Mar 21.
61. Maglogiannis I, Loukis E, Zafiropoulos E, Stasis A. Support vectors machine-based identification of heart valve diseases using heart sounds. Comput Methods Prog Biomed. 2009;95(1):47–61.
62. Ghumbre S, Patil C, Ghatol A. Heart disease diagnosis using support vector machine. International conference on computer science and information technology (ICCSIT'). Pattaya; 2011.
63. Hanbay D. An expert system based on least square support vector machines for diagnosis of the valvular heart disease. Expert Syst Appl. 2009;36(3):4232–8.
64. Babaoglu I, Findik O, Ülker E. A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. Expert Syst Appl. 2010;37(4):3177–83.