Research Paper: A Data Mining Approach for Coronary Artery Disease Prediction in Iran

Azam Dekamin[1*], Ahmad Sheibatolhamdi[2]

1. Department of Information Technology Management, Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
2. Department of Industrial Management, Faculty of Humanities, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran.

Use your device to scan and read the article online

Citation: Dekamin A, Sheibatolhamdi A. A Data Mining Approach for Coronary Artery Disease Prediction in Iran. Journal of Advanced Medical Sciences and Applied Technologies. 2017; 3(1):29-38. https://doi.org/10.18869/nrip.jamsat.3.1.29

doi: https://doi.org/10.18869/nrip.jamsat.3.1.29

Article info:
Received: 13 Aug. 2016
Accepted: 29 Nov. 2016

Keywords:
Coronary artery disease, Angiography, Data mining, Stacking, Random forest

ABSTRACT

Objectives: Coronary artery disease (CAD) is caused by atherosclerosis in coronary arteries and results in cardiac arrest and heart attack. Angiography is one of the most accurate methods to diagnose Heart disease, it incurs high expenses and comes with side effects. Data mining is the extraction of hidden predictive information and unknown data, patterns, relationships and knowledge by exploring the large data sets which are difficult to find and detect with traditional statistical methods. One of the biggest problems that prevent pattern recognition from functioning rapidly and effectively are the noisy and inconsistent data in databases. The present study intends to provide a data preparation method based on clustering algorithms for diagnosis Coronary artery disease with higher efficiency and fewer errors.

Materials: In this study, the data under investigation was collected from a number of 303 persons referring to the heart unit in one of Tehran-based hospitals within the time interval 2011 to 2013. It included 54 features. K-means algorithm is used for clustering based data preprocessing system for elimination of noisy and inconsistent data and Naïve Bayes, K nearest neighbor and Decision tree are used for classification. Another two feature subset selection methods for cleaning data are also used to make a comparison between clustering based method and attribute selection method. Rapid Miner Software was adopted to conduct this study.

Results: Findings of this research indicated that the suggested model will have the highest efficiency, 90.91. According to the results, the proposed method of performance is highly successful compared to other results attained and seems effective for pattern recognition applications.

Conclusion: With these results, the proposed method can be used in the diagnosis of coronary artery disease.

*** Corresponding Author:**
Azam Dekamin, MSc.
Address: *Department of Information Technology Management, Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran.*
Tel: *+98 (912) 3788205*
E-mail: *dekamin@gmail.com*

1. Introduction

Coronary Artery Disease (CAD) is a chief cause of mortality in industrial countries. Every year, 17.3 million people lose their lives due to heart diseases. Many of these deaths occur during the first heart attack. Many patients are unaware of their heart disease until they experience their first heart attack. Our goals include notifying patients in a timely manner by means of diagnostic systems that can readily be used during routine-controls, thereby preventing sudden deaths. While angiography is presently applied to diagnose clogged arteries, it is sought by scholars to be replaced by alternatives due to the complications it poses to the human body and the costs it incurs. Hence, diagnosis of this disease by means of noninvasive methods is both important and helpful [8].

Data mining is the process for finding knowledge out of a huge amount of data saved in databases. Applying some algorithms, data mining intends to discover relationships and patterns among data. Some applications of data mining lie in banking, discovery of cheats and crimes, marketing, and medication that would lead to reduced risk and cost levels [4].

The growth of medical databases is very high. This rapid growth is the main motivation for researchers to mine useful information from these medical databases. Data mining techniques play an important role in finding patterns and extracting knowledge to provide better patient care and effective diagnostic capabilities. Data mining provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with traditional statistical methods. Robert Detrano [16], who has organized a heart disease database, makes use of logistic regression algorithm to predict Ischemic Heart Diseases (IHD). He obtained an accuracy of 77%. Newton Cheung [17] made use of C4.5 algorithm and a Naive Bayes classifier to categorize diseases associated with heart and blood vessels, achieving accuracies 81.11% and 81.48%, respectively.

Kemal Polat [5] proposed a method called artificial immune system (AIS), achieving an accuracy of 84.5% in his classifications. Applying a method similar to that adopted by Kemal Polat, SalihGunes achieved an accuracy of 87%. Afterwards, other different results were obtained by Weka and RA software, the highest of their accuracy being 77%. My Chau Tu compared bagging with C4.5 and Naive Bayes to diagnose heart disease, obtaining an accuracy of 81.41% [19]. Rajkumar and Reena compared Naive Bayes and K-nearest neighbor for prediction of heart diseases, achieving an accuracy of 52.33% [6]. Nihat Yilmaz and Onur Inan [8] used a new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases.

Sitair-Tau examined J48 algorithm in decision tree to predict heart disease using Weka software [18]. In 2015, Kemal Akyol and Elif Calik used randomForest classification method and they achieved the ratio of the Correct classification was 97.72% [7]. Parthiban and Subramanian combined neural network data mining algorithm and learning capabilities of fuzzy logic that have a qualitative approach in order to utilize it aimed at diagnosis of heart disease [13]. Srinivas demonstrated that heart disease rates calculated by patients' self-reports are higher in the India-located Andhra Pradesh state where Singareni coal mines exist compared to other states [1]. Their prediction of heart disease was performed by such variables as age, ethnicity, individuals' academic degrees, income levels, Body Mass Index (BMI) values, etc. assisted by clustering data mining technique and data mining algorithms like neural networks, Bayesian networks, decision trees, and Support Vector Machine (SVM). In the meantime, decision trees could present a higher accuracy level compared to other algorithms.

Subalakeshmi applied Bayesian data mining algorithm to classify data [15]. Ordonez et al. made use of C4.5 decision tree, correlation regulations algorithm, and 25 risk factors in order to predict heart disease, concluding that correlation regulations generally generate simpler prediction rules than decision trees. Vanisreeand Singaraju presented a decision support system using a neural network to diagnose congenital heart disease [2]. Nihat Yilmaz and OnurInan proposed a new modified K-means technique for clustering based data preparation method for the elimination of noisy and inconsistent data and Support Vector Machines [8]. Safdari and Ghazi Saeidi undertook to compare performance of decision tree and neural network in prediction of infection by myocardial infarction [11]. Using K-means clustering techniquIn, Parchi Paliwal and Mahesh Malviya showed that the percentage of classify correct records by proposed method is more az compared to the previous methods [10].

In the paper presented by Rajeswari and colleagues, they study the heart disease using Neural Network [8]. They have studied the influence of feature selection for neural network algorithm in identifying patients with Ischemic heart disease. The result of their study shows that when all the features are applied, the precision rate in training mode 89.4% and in test mode is 82.2%. An interesting point in the conclusion is that any reduction in features entry causes the precision decrease in both training and test modes.

As cited here in above, nosocomial big data usually includes useful information on demographic characteristics of patients, on the one hand, the features relevant to the manner by which they are treated, on the other. Studies under investigation indicated that some variables such as EF, Region RWMA, Q Wave, and Twave inversion applied here intended to diagnose CAD have been either left unheeded or less noticed.

2. Materials and Methods

Dataset

In this study, the data under investigation was collected from a number of 303 persons referring to the heart unit in one of Tehran-based hospitals within the time interval 2011 to 2013. It includes 54 features containing 53 features of disease symptoms and 1 attribute of disease diagnosis called class field. The value zero indicates absence of the disease and the value one shows existence of disease. Features of disease symptoms are categorized into four classes: Demographic information of patients, symptoms and impacts investigatable by physicians,electrocardiogram features, and laboratorial post-eco features. Features of this study are rarely adopted by prior research.

Pre-processing and preparation of data

This phase is one of the most important and time-consuming stages of research. Due to various noises, there are data in databases that are inconsistent, duplicate and uncertain in comparison to similar data [22]. Since information achieved from this phase is served as inputs for further phases and. low-quality data would result in low-grade outcomes, pre-processing stage was undertaken in order to guarantee accuracy of data [12].

Data preparation based on clustering

After primary examinations, it was revealed that data was unbalanced and a number of 87 records had the objective variable 0 and 216 ones had the objective variable 1. Therefore, sub-sampling method was adopted for balancing data in order to achieve better results. It is the case that some of the problems (noise, duplicate data …) which were mentioned in last paragraph can be solved in the proposed system by using K-means algorithm. In this way, the performance of the classifiers which were tested increased. Data preparation based on clustering as it is seen in Figure 1 is comprised of two phases. In the first phase, the K-means algorithm was applied for the dataset. In the database separated into K sets, eliminating is carried out based on that samples in the same set should also belong to the same class. In the second phase, the data of the more the same class remains within a set. The other data are eliminated as shown in Figure 1.

Testing stage

After data preparation,the accuracy rates of proposed system with the obtained and filtered data are determined by using them in classifiers. In classification phase,in order to prove the reliability of the data cleaning method, the model which is shown in Figure 2 is proposed. In this model after balancing data, two data preprocessing methods were used. One of them was data preparation based on clustering system which was mentioned before, and another one is Attribute selection method which is based on weighting by Gini and information Gain. In the next section, this method will be explained. As a result, two filtered datasets were produced. After this phase, classifiers [Decision Tree (DT), K Nearest Neighbor (KNN) and Naïve Bayes (NB)] were applied to both datasets and results of them were compared with each other. To categorize educational and experimental data, the Hold Out method was applied: 80% of data was used for education stage and the rest 20% for test phase. Culled data was investigates using Rapid Miner Software. In classification methods, one field is held as output field. Here, the variable Cath was regarded as output and 53 other variables as inputs.

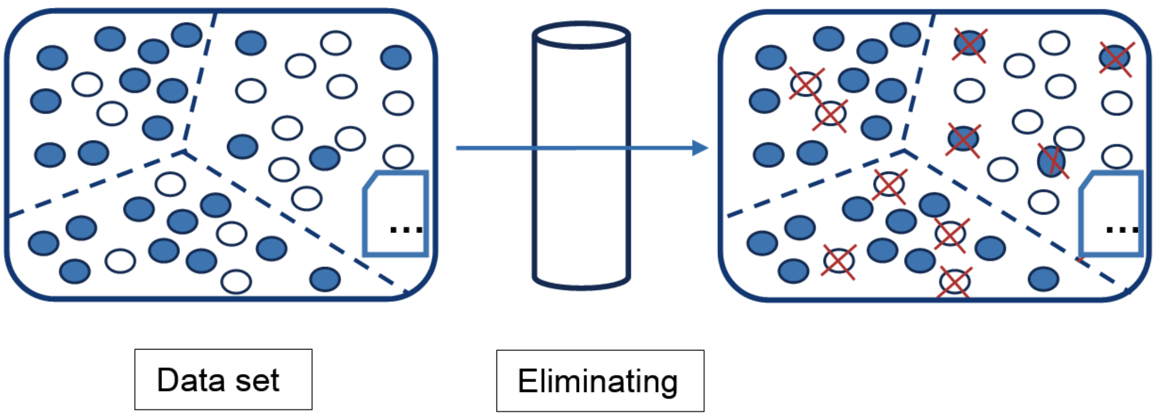

Figure 1. Proposed data preprocessing system

JAMSAT
Journal of Advanced Medical Sciences and Applied Technologies

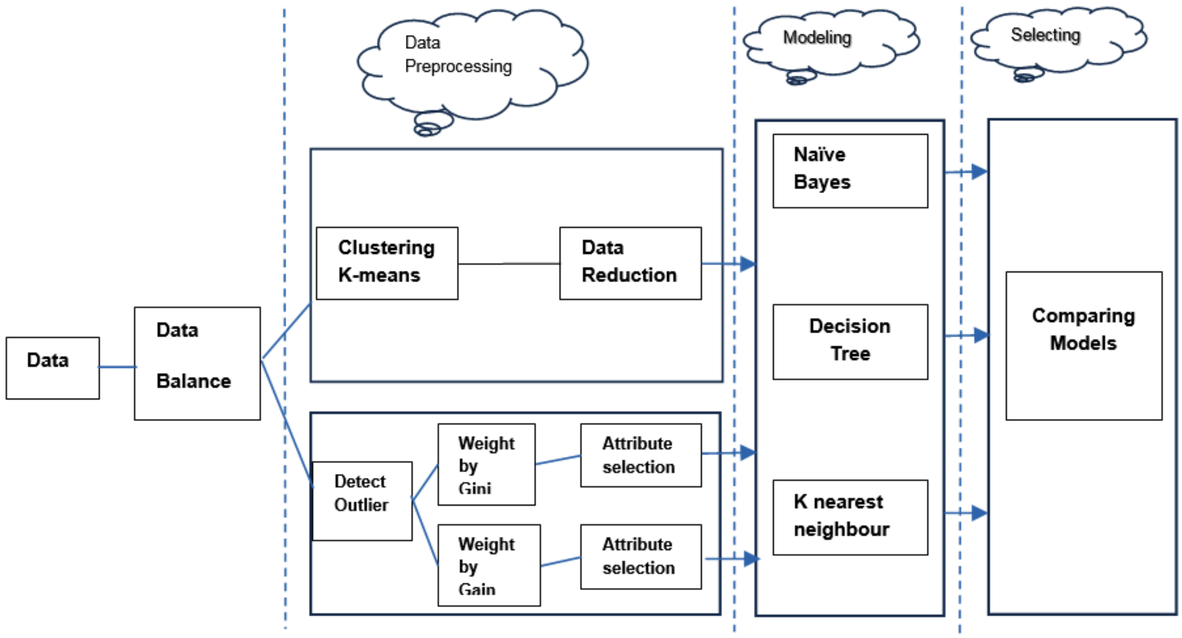

Figure 2. Proposed model of this study

JAMSAT

Feature subset selection

In order to preprocess dataset, outliers were detected and filtered at first. One of the important factors in development of a high-accuracy model is proper selection of features. To reach at this purpose, it is mandatory to select a subset of desirable features and finding suitable characteristics. In this study, filtering methods such as weighting by Information Gain and Gini, were applied, influential features were selected, and inserted into the model.

Weighting based on Gini Index

A subset of features with highest influence was selected to participate in modeling. After parameter adjustment, 10 features were selected.

$$GINI(t)=1-\sum_{i}[p(j\mid t)]^2 \quad GINI_{split}=\sum_{i=1}^{k}\frac{n_i}{n}GINI(i)$$

Weighting based on Gain Index

This criterion is one of the most well-known criteria that is applied for weighting, which makes use of another criterion called entropy. After parameter adjustment, a number of 10 features were selected.

$$Entropy(t)=-\sum_{j}p(j\mid t)\log p(j\mid t) \quad GAIN_{split}=Entropy(p)-\left(\sum_{i=1}^{k}\frac{n_i}{n}Entropy(i)\right)$$

Evaluation

Evaluation criteria F-measure, sensitivity, specificity, recall, precision, and accuracy relevant to the research model were measured and cited in Table 3 and 4.

Confusion Matrix

The Confusion Matrix (Table 1) illustrates the manner the classification technique performs with respect to input dataset as separated by different types of classifi-

Table 1. Confusion matrix

		Predicted Class	
		Class = Yes	Class = No
ACTUAL CLASS	Class = Yes	a (TP)	b (FN)
	Class = No	c (FP)	d (TN)

JAMSAT

cation issue. The concepts TN, FP, FN, and TP in this matrix are defined as follows:

TN: This value indicates the number of records whose real class was negative, and classification algorithm could accurately diagnose their class to be negative.

FP: This value indicates the number of records whose real class was negative, and classification algorithm wrongly diagnosed their class as positive.

FN: This value indicates the number of records whose real class was positive, and classification algorithm wrongly diagnosed their class as negative.

TP: This value indicates the number of records whose real class was positive, and classification algorithm could accurately diagnose their class to be positive.

Now, different types of important evaluation criteria pertinent to classification technique with regard to confusion matrix are expounded. Accuracy of the model shows that all items that have to be observed by a classification in order to possess a suitable performance are considered in the criterion Accuracy. Clearly, the values TP and TN are the most important values that have to be maximized in a problem. Since both TP and TN are located in nominator, it is safe to assert that all classes existing in classification problem are considered in above relation. Thus, the criterion Accuracy of Classification is the commonest and most renowned criterion for calculation of efficiency of classification algorithms. The criterion Recall (x) expresses accuracy of classification of class x with respect to all records with the label x.

The criterion Precision (x) articulates accuracy of classification x with regard to the whole items for which the label x has been proposed by classifier for examined record. Remember that the criterion Recall (x) expresses efficiency of classifier with regard to the number of events of class x. While, the criterion Precision (x) is basically grounded upon prediction accuracy of classification, demonstrating the degree by which outputs of classification are reliable. Another important point is that denominator of Recall (x) relations should be held as equal to the total number of records with label x in classifications where label of some records is determined to be unknown. The criterion F-Measure (x) exemplifies a combination of the criteria Precision (x) and Recall (x) that is utilized when special importance could be attached neither to Precision (x) nor Recall (x).

$$\text{Accuracy}=\frac{a+d}{a+b+c+d}=\frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision (p)}=\frac{a}{a+c}$$

$$\text{Recall (r)}=\frac{a}{a+c} \quad \text{(F-measure)}=\frac{2rp}{r+p}=\frac{2a}{2a+b+c}$$

$$\text{Sensitivity}=\frac{d}{a+d}=\frac{TP}{TP+FN} \quad \text{Specificity}=\frac{d}{b+c}=\frac{TN}{TN+FP}$$

3. Results

This section presents experimental results using our model on the dataset which was randomly selected from among medical cases of 303 patients referring to the heart unit in one of Tehran-based hospitals, from which 216 subjects were infected by CAD and the rest were not. In this research, primary studies showed that a number of 87 records had the objective variable 0 and 216 ones had the objective variable 1. It is obvious from these observations that data was unbalanced, and improved results could be achieved by balancing methods. Hence, data was balanced using a sub-sampling method. The information about the dataset is shown in Table 2.

As it can be seen in Figure 2,after sub-sampling, datawas being balanced. As mentioned in last section, in data preprocessing phase, two separate methods were used.

Method 1

K-means clustering based method was utilized as an instrument for finding the relevant or irrelevant datum inside themselves. Data reduction was done in order to eliminate disruptive data from highly constant sets. The number of clusters in this method was adjusted as a result of comparing Davies-Bouldin index. K=6 was selected in k-means algorithm. (Table 2).Afterwards, data was prepared for introduction into modeling section where training and testing data was divided with a ratio of 80 to 20. In this phase, Naive Bayes algorithm, decision tree, and K-nearest neighbor were applied for modeling. Relevant results are cited in Table 3. From these results, it is evident that the accuracy of KNN was 90.91 which was the most accurate algorithm between Naïve Bayes, decision tree and K-nearest neighbor. The recall index was 93.33 which means that we can predict 93.33% of patients with coronary artery disease in that hospital.

Method 2

After sub-sampling, outlier detection by distance was carried out,in order to reduce the number of features and the computational time and also the cost of the experi-

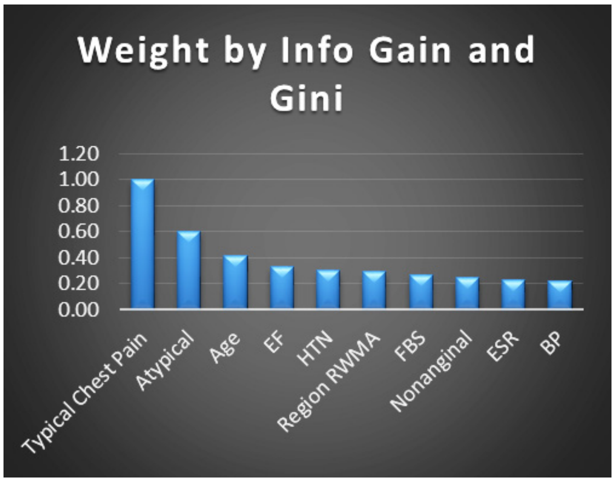

Figure 3. Comparison of variables influential in model using weighteby Gain and Gini index.

JAMSAT

TG=Triglyceride
DM=Diabetes
HTN=Hyper Tension
FBS=fasting blood sugar
ESR=erythrocyte sedimentation rate
FH=family history
Region RWMA=regional wall motion abnormality
BP=blood pressure
EF =ejection fraction

ments, the most effective features which increase the risk of heart disease, were extracted and selected using Feature subset selection, as cited in Table 2. The two methods weight by Information Gain and weight by Gini were applied to select features influential in the model. In each method, a number of 10 features with higher weights were selected for modeling process after parameters were regulated. In Figure 1 , ten features which are more influential is sorted.

Afterwards, data was prepared for introduction into modeling section where training and testing data were divided with a ratio of 80 to 20. In this phase, first Naive Bayes algorithm, decision tree, and K-nearest neighbor were applied for modeling per Figure 2. Relevant results are cited in Table 4. Then these results were compared with proposed method's result (Figure 3).

From these results, it is obvious that Naïve Bayes accuracy and recall rate were 83.33% which was higher than other algorithms in Table 4. It means that K-means clustering based method performance is highly successful compared to other results attained, and seems very promising for pattern recognition applications. Figure 4 represents accuracies of classifiers in this research. It is obvious that k1 NB is the most accurate classifiers in this research.

Figure 5 compares the criteria Recall and F-Measure for evaluation of algorithms. As it can be seen, both recall and f-measure increased dramatically by k-means method. For all classifiers,accuracy, precision, recall, f-measure, sensitivity and specificity increased by using clustering based method in preparation data phase.

Table 2. Information about dataset and features in this study

An Iranian hospital dataset samples	303
Number of classes	2
Number of features	54
Number of clusters in K-means algorithm	6
Balanced data samples	173
Cleared data K-means samples	100
Number of feature selected by Gini	10
Number of feature selected by Information Gain	10

JAMSAT

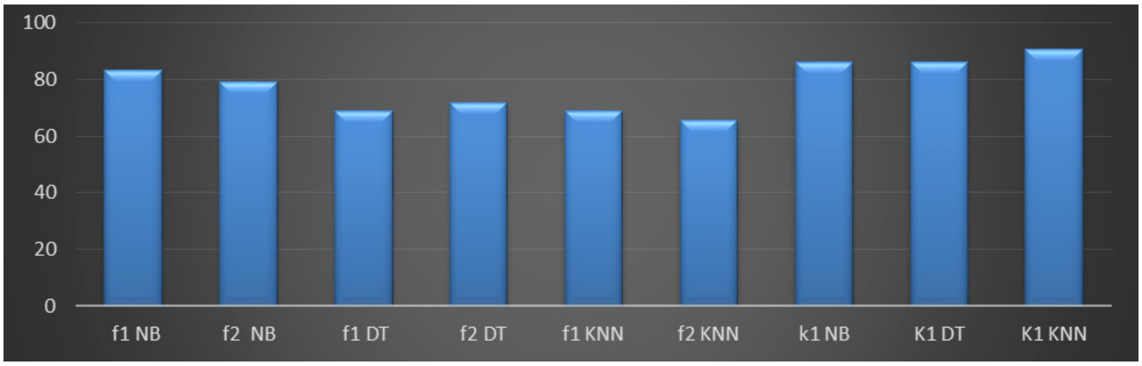

Figure 4. Comparison of accuracies of classifiers

JAMSAT

NB=Naïve Bayes
DT=Decision tree
KNN=K nearest neighbor

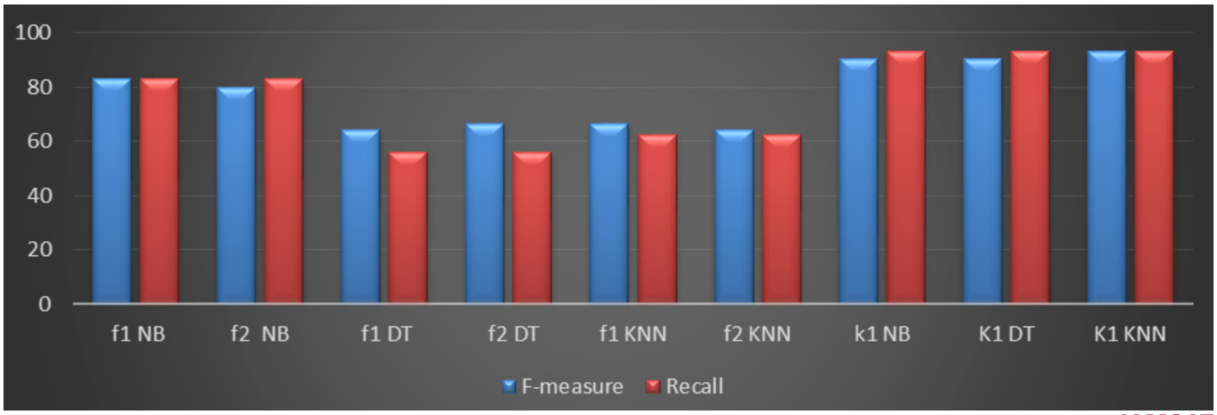

Figure 5. Comparison of the criteria recall, and F-measure for evaluation of algorithms

JAMSAT

NB=Naïve Bayes
DT=Decision tree
KNN=K nearest neighbor

4. Discussion

In this study, the likelihood of CAD affliction was attempted to be predicted using data mining algorithms. In this study, attempts were made to take advantage of K-means algorithm in preprocessing phase that were less applied beforehand. Using this newly method was arranged aimed at reaching at more efficient outcomes and fewer errors. In pre-processing phase, it was made clear that data was unbalanced. To acquire better results in pre-processing phase, therefore, data balancing was undertaken—the point which was mostly left unnoticed in prior studies. The proposed method's result was compared with the derived result of classifiers with feature subset selection methods in data cleaning phase showed that omitting noise and outliers may have deep effect on the result.

Existence of suitable data and proper pre-processing as well as application of ensemble data mining algorithms would provide good outcomes on medical data.

Table 3. Performance of classifiers with clustering based preprocessing

Numbers	Modeling Algorithms	Accuracy	Precision	Recall	F-Measure	Sensitivity	Specificity
1	Naïve Bayes (k1 NB)	86.36	87.50	93.33	90.32	93.33	71.43
2	Decision tree (K1 DT)	86.36	87.50	93.33	90.32	93.33	71.43
3	KNN (K1 KNN)	90.91	93.33	93.33	93.33	93.33	85.71

k1: When K-means is used in preprocessing data for data reduction.

JAMSAT

Table 4. Performance of classifiers with feature selection preprocessing

Numbers	Modeling Algorithms	Feature Subset Selection	Accuracy	Precision	Recall	F-Measure	Sensitivity	Speci-ficity
1	Naïve Bayes (f1 NB)	Gini	83.33	83.33	83.33	83.33	83.33	83.33
2	Naïve Bayes (f2 NB)	Info gain	79.17	76.92	83.33	80.00	83.33	75.00
3	Decision tree (f1 DT)	Gini	68.75	75.00	56.25	64.29	56.25	81.25
4	Decision tree (f2 DT)	Info gain	71.88	81.82	56.25	66.67	56.25	87.50
5	KNN (f1 KNN)	Gini	68.75	71.43	62.50	66.67	62.50	75.00
6	KNN (f2 KNN)	Info gain	65.62	66.67	62.50	64.52	62.50	68.75

f1: When weight by Gini is used in preprocessing data for feature selection.

JAMSAT

f2: When weight by Information Gain is used in preprocessing data for feature selection.

In pre-processing phase, it was made clear that data was unbalanced. To acquire better results in pre-processing phase, therefore, data balancing was undertaken-the point which was mostly left unnoticed in prior studies.

The proposed method's result was compared with the derived result of classifiers with feature subset selection methods in data cleaning phase showed that omitting noise and outliers may have deep effect on the result.

5. Conclusion

This study was conducted in order to diagnose and predict cardiovascular disease. The mentioned dataset has been gathered from one of Tehran's hospitals with 303 records and 54 features that were recorded from patients seeking medical attention. As it was mentioned before, Noise and outliers in datasets cause the classification algorithms to be less efficient; therefore, by omitting defective data and outliers we tried to increase the performance of the algorithms.

The suggested method in this study, is to apply K-means clustering in the preprocessing phase that has significantly increased the performance of the classification algorithms. This method successfully predicted Coronary artery disease with a rate of %90.91 using the KNN algorithm. It could be concluded that the above mentioned method superseded other compared methods in prediction of Cardiovascular disease in this study and could be applied in the medical diagnosis. In this study, this goal is achieved. Finally, larger datasets, more features and also broader data mining approaches, could be used to achieve better and more interesting results.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declared no conflict of interests.

References

[1] Srinivas K, Rao GR, Govardhan A. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. Paper presented at: The 5th International Conference on Computer Science & Education. 2010 August 24-27; Hefei, China. doi: 10.1109/iccse.2010.5593711

[2] Vanisree K, Singaraju J. Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks. International Journal of Computer Applications. Foundation of Computer Science; 2011; 19(6):6–12. doi: 10.5120/2368-3115

[3] Elsayad A, Fakr M. Diagnosis of cardiovascular diseases with Bayesian classifiers. Journal of Computer Science. 2015; 11(2):274-82.

[4] Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. A data mining approach for diagnosis of coronary artery disease. Computer Methods and Programs in Biomedicine. 2013; 111(1):52–61. doi: 10.1016/j.cmpb.2013.03.004

[5] Polat K, Güneş S. A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS. Computer Methods and Programs in Biomedicine. 2007; 88(2):164–74. doi: 10.1016/j.cmpb.2007.07.013

[6] Rajkumar M, Reena GS. Diagonsis of heaer disease using datamining algorithm. Global journal of Computer Science and Technology. 2010; 10(10):38-43.

[7] Akyol K, Çalik E, Bayir Ş, Şen B, Çavuşoğlu A. Analysis of demographic characteristics creating coronary artery disease susceptibility using random forests classifier. Procedia Computer Science. 2015; 62:39–46. doi: 10.1016/j.procs.2015.08.407

[8] Yilmaz N, Inan O, Uzer MS. A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases. Journal of Medical Systems. 2014; 38(5):48. doi: 10.1007/s10916-014-0048-7

[9] Nahar J, Imam T, Tickle KS, Chen YPP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications. 2013; 40(4):1086–93. doi: 10.1016/j.eswa.2012.08.028

[10] Paliwal P, Malviya M. An efficient method for predicting heart disease problem using fitness value. International Journal of Computer Science and Information Technologies. 2015; 6(2):1290-3.

[11] Safdari R, Ghazi Saeedi M, Gharooni M, Nasiri M, Arji G. [Comparing performance of decision tree and neural network in predicting myocardial infarction (Persian)]. Journal of Paramedical Science and Rehabilitation. 2014; 3(2):26-37.

[12] Rao VSH, Kumar MN. A new intelligence-based approach for computer-aided diagnosis of dengue fever. IEEE Transactions on Information Technology in Biomedicine. 2012; 16(1):112–8. doi: 10.1109/titb.2011.2171978

[13] Parthiban L, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering; 2007; 1(5):278-81.

[14] Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Philadelphia: Elsevier; 2011.

[15] Subbalakshmi G, Ramesh K, Rao MC. Decision support in heart disease prediction system using naive bayes. Indian Journal of Computer Science and Engineering. 2011; 2(2):170-6.

[16] Detrano RC, Wong ND, Doherty TM, Shavelle RM, Tang W, Ginzton LE, et al. Coronary calcium does not accurately predict near-term future coronary events in high-risk adults. Circulation. 1999; 99(20):2633–8. doi: 10.1161/01.cir.99.20.2633

[17] Cheung N. Machine learning techniques for medical analysis. School of Information Technology and Electrical Engineering [PhD dissertation]. Queenland: University of Queenland.

[18] Sitar-Taut VA, Zdrenghea D, Pop D, Sitar-Taut DA. Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics. 2009; 5(3):29-32.

[19] Tu MC, Shin D, Shin D. Effective diagnosis of heart disease through bagging approach. Paper presented at: The 2nd International Conference on Biomedical Engineering and Informatics. 2015 March 30-31; Penang, Malaysia.

[20] Chaurasia V, Pal S. Early prediction of heart diseases using data mining techniques. Caribbean Journal of Science and Technology. 2013; 1:208-17.

[21] Mahdi N, Behrouz M, Amir K. Dynamic recommendation: Disease prediction and prevention using recommender system. International Journal of Basic Science in Medicine. 2016; 1(1):13-7.

[22] Rajeswari K, Vaithiyanathan V, Neelakantan, TR. Feature selection in ischemic heart disease identification using feed forward neural networks. International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012). Procedia Engineering. 2012; 41:1818-23.