Middle-East Journal of Scientific Research 24 (S1): 374-381, 2016
ISSN 1990-9233
© IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.S1.72

Effective Diagnosis of Heart Disease Using Inter Quartile Range Filter and Decision Tree Classifier

[1]R. Delshi Howsalya Devi and [2]M. Indra Devi

[1]K.L.N College of Engineering, Madurai, India
[2]Kama raj College of Engineering, Madurai, India

Abstract: Heart failure is the leading causes of death in worldwide. In recent years, data mining as one of the most widely used methods in health care. Diagnosing of the heart disease is one of the important issue and many researchers investigated to expand the smart medical decision support systems for improve the ability of the physicians. This paper presents a new technique for diagnosis of heart disease. The proposed model is based on a hybrid method that uses IQR filter for pre-processing the original data set by removing outliers and j48 decision tree classifier is used for diagnose the heart data into healthy and a patient who is focus to possible heart disease. The robustness of the proposed method is examined using classification accuracy and confusion matrix. The experimental results demonstrate that the obtained classification accuracy 99.67% is very promising compared to the previously reported classification techniques.

Key words: Outlier Detection · Decision tree · IQR · Data Mining · Heart Disease

INTRODUCTION

Heart disease has significantly improved over the last decade and has become the primary cause of death for people in the majority countries around the world. There are many features of heart disease disturbing the structure or function of the heart [1]. These might be hard for doctors to diagnose fast and accurately. Therefore, it is necessary to employ computerized tools [2] in heart disease diagnosis to assist doctors to diagnose earlier with higher accurately. Due to many features of heart datasets, which contain related as well as unrelated and redundant features? [3] Irrelevant features do not control description of the target class. Redundant features do not give to anything but they make outliers towards description of target class Those features not only involve the results of classification but also make the system job slowly [4]. Therefore, removing those features before applying classifier method is necessary [5]. For this purpose, outlier detection and removal is required in the heart disease analysis system. This reduces the risk of over fitting, improves simplification ability of the model, provides better preventability and requires less computation causing smaller features. Having so many factors to evaluate the diagnose the heart disease of a patient makes the physician's job complicated [6]. A physician typically makes decisions by estimating the current test results of a patient and by referring to the earlier decisions she prepared on other patients with the same condition. The former method depends strongly on the physician's knowledge. On the other hand, the later depends on the physician's awareness to compare her patient with her earlier patients [7]. This work is not easy considering the number of features she has to estimate. In this crucial step, she may need an accurate tool that lists her earlier decisions on the patient having same (or close to same) factors. Data mining have played an important role in heart disease research. To find the unknown medical information from the different expression [8] between the healthy and the heart failure persons in the existed clinical data is a noticeable and powerful approach in the study of heart disease classification. In this paper, we propose two steps to predict the heart disease status for presenting a more efficient and exact heart disease prediction system. We first apply inter quartile filter is a pre-processing step to the heart disease dataset for outlier detection and removal. After pre-processing, we run the main decision tree j48 classifier algorithm on the preprocessed dataset.

Corresponding Author: R. Delshi Howsalya Devi, K.L.N College of Engineering, Madurai, India.

We obtained 99.67% classification accuracy from the experiments made on the data Cleveland heart disease taken from UCI. We also obtained 99.3% and 99.7% precision and false positive values in heart disease diagnosis.

The remaining of the paper is prepared as follows. We present the related work in the next section. In Section III, we give a brief background inter quartile range filtering and J48 decision tree classifier [9]. We present the proposed method in Section IV. In Section V, we give the experimental data to show the effectiveness of proposed method. Finally, we conclude this paper in Section VI.

Related Work: Aha & Kibler (1988) have proposed an instance-based algorithms and achieved 77% and 74.8% accuracy For Antigrowth and C4.5 techniques. Detrain *et al.* (1989) have investigated a probabilistic algorithm to diagnose the risk of coronary artery disease and finished that patients experiencing chest ache and transitional disease occurrences are the higher risk subjects. Genera, Langley, & Fisher (1989) have explored a conceptual clustering system and gained an acceptable accuracy (78.9%).. Nishara *et al.* (2014) have proposed a C4.5 algorithm, MAFIA and K means clustering and achieved 89% classification accuracy. To, *et al*.,(2009) have proposed a decision tree j48 algorithm and obtained 78.9% accuracy. Also they has proposed a bagging algorithm and achieved classification accuracy 81.41%. Plat *et al*(2007) has developed a Fuzzy-AIRS–nearest Neighbor and has achieved a classification accuracy 87%. Das, *et al*., (2009) have obtained classification accuracy 89.01% with neural network ensembles. Rajkumar, A.et (2010) has proposed a decision list algorithm and has obtained classification accuracy 52%. Srinivas (2010) has achieved a classification accuracy 80.46% with one dependency augmented naïve bays classifier M. Anabasis (2010) have developed a genetic with classification via clustering and have obtained 88.3% classification accuracy. Robert Detrain (2008), who assembled the Cleveland heart disease database, used logistic regression algorithm and obtained 77.0% classification accuracy Newton Cheung used Naïve Bays, BNNF, BNND, C4.5 and BNNF algorithms and reached the classification accuracies of 80.96%, 81.11% of 81.11% and 81.48% respectively (Cheung, 2001) Plat *et al*.(2005) proposed a method (AIS) Artificial Immune System and achieved 84.5% classification accuracy (Plat *et al*., 2005). Then, a similar model was used by Olsen and Guns and obtained 87.0% classification accuracy. Pander *et al* (2013) have proposed a Decision Tree with Reduced Error Pruning Method and achieved 75.73% classification accuracy. Basher *et al* (2014) achieved classification accuracy 81.82% with combination of Naïve Bays, Decision Tree and Support Vector Machine. Caucasia *et al* (2013) have used the commonest types of decision tree algorithms for the prediction of heart diseases. CARD, ID3 and DT decision trees were applied with the same dataset available at [10] and evaluated using 10-fold cross validation method. CARD decision tree has presented the highest classification accuracy with 83.49%, followed by DT with 82.50% and finally 72.93% for ID3.

Upping *et al*. (2014) have proposed C4.5 decision tree classifier for predicting heart disease. The data set in [10] were also used within this experiment. Their strategy aimed to reduce the number of parameters within the data set in order to avoid the redundant features that are not important in the classification. Therefore, 7 of 13 parameters have only been used and the C4.5 classifier showed 85.96% of accuracy. Mahmud and Kuppa (2010) has proposed a new pruning method with the aim of improving classification accuracy of heart diseases and reducing the tree size. A combination of pre-pruning and post-pruning was used for pruning C4.5 decision tree classifier. The new decision tree has been compared with the benchmark algorithms using dataset available online at [11]. The results showed that the new method significantly reduced the tree size and achieved 76.51% of accuracy. Showman *et al*(2012) have focused on the improvement of decision tree accuracy for diagnosis of heart disease. K-means clustering was integrated into the decision tree in order to enhance the diagnosis of heart disease. The dataset mentioned in [12] has been utilized. The highest accuracy obtained was 83.9% by applying the inliers method with two clusters. Melilla *et al* [13] developed a model for risk assessment in patients suffering from congestive heart failure. ECG recording for long-term heart rate variability has been used as a dataset, which is derived from two different Congestive Heart Failure databases. A CART decision tree algorithm is used with the aim of classifying patients into two groups based on the risk factor and achieved 85.4% of accuracy. Bohacik (2013) applied an alternating decision tree for the prediction of heart failure and obtained 77.65% classification accuracy.

Nguyen, Abbes, Douglas and Saied (2015a) presented a medical diagnosis system which was mutual genetic fuzzy logic system with wavelet. The wavelet transformation was engaged to extract discriminative patterns for high-dimensional datasets from UCI. Then fuzzy standard additive trained by genetic algorithm

(GSAM) was used to classify medical dataset. This proposed method was evaluated using Cleveland heart disease datasets from UCI. The experimental results proved that GSAM became highly capable when deployed with small number of wavelet patterns as its computational load was reduced. However, this proposed approach had a shortcoming regarding selection of the best number of wavelet features and the accuracy of this proposed model was 78.78% for Cleveland heart disease datasets. Nguyen, Abbes *et al*. (2015b) have proposed an automated medical data classification Fuzzy c-mean clustering algorithm was proposed to create fuzzy rule based of the fuzzy system and genetic algorithm was used to tune parameter of the fuzzy system. The WT was used to locate a reduction of features therefore that reduces computational burden and enhances performance of the proposed methods. It was measured using Cleveland heart disease datasets from UCI. Experiments Results proved that a significant dominance of the wavelet–IT2FLS approach compared to other machine learning techniques including probabilistic neural network, fuzzy ARTMAP, support vector machine and adaptive neuron-fuzzy inference system However, this proposed method did not select optimal number of features and the accuracy of this proposed method was 81.01% for Cleveland heart disease datasets.

METHODS AND MATERIALS

In this section, The Inter Quartile Range(IQR) is introduced briefly for outlier detection and J48 classification algorithm is introduced for classifying the Cleveland heart disease data set in to two types of risk level ie. Low risk (or) High risk.

Inter Quartile Range [IQR]: An outlier filtering approach commonly uses a distance measures to detect outlier instances that are at a significant distance from the others. It is a challenging task to eliminate outliers in order to improve the performance of classifiers In this paper, Inter- Quartile Range to detect Outliers and Extreme Values in a Cleveland heart disease data set. The steps for detecting the outliers in the data using IQR are outlined in Table 1.

Consider the following data points 32,26,27,11.6,28.5,33.2,18.9,48,41.2,25,36.1,24.6.In step 1, order the sample data points in ascending order. The order of the data points follows the sequence. 11.6,18.9,24.6,25,26,27,28.5,32,33.2,36.1,41.2,48. In step 2 and 3, The first and third quartile value are calculated. The value of IQR is computed by following

Q1 = 24.6
Q2 = 27
Q3 = 33.2
Q4 = 48

In step 4 The inter quartile range(IQR) is calculated by using Eq(1)

$$\begin{aligned} IQR &= Q3 - Q1 \quad (1) \\ &= 33.2 - 24.6 = 8.6. \end{aligned}$$

where $Q3 = 33.2$
$Q1 = 24.6$

The lower boundary value is calculated using Eq(2).

$$\begin{aligned} \text{Lower boundary} &= Q1- (1.5 * IQR) \quad (2) \\ &= 24.6 - (1.5 * 8.6) \\ &= 24.6 - 12.9 = 11.7 \end{aligned}$$

So the value of lower boundary is 11.7. In step 4, The value of upper boundary is calculated by using Eq(3).

$$\begin{aligned} \text{Upper boundary} &= Q3 + (1.5 * IQR) \quad (3) \\ &= 33.2 + (1.5 * 8.6) \\ &= 33.2 + 12.9 \\ &= 46.1 \end{aligned}$$

Table 1: Steps for Outlier detection using IQR.

Step 1: Arrange data points in ascending order
Step 2: Calculate the first quartile value (Q1)
Step 3: Calculate the third quartile value (Q3)
Step 4: Calculate inter quartile range (IQR) =Q3-Q1
Step 5: Calculate lower boundary value based on the following formula. Lower boundary = Q1-(1.5*IQR)
Step 6: Calculate upper boundary value based on the following formula Upper boundary = Q3+ (1.5*IQR)
Step 7: Data points anything outside the lower and upper boundary value is an outlier

Table 2: Steps for extreme value detection using IQR

Step 1: Arrange data points in ascending order
Step 2: Calculate the first quartile value (Q1)
Step 3: Calculate the third quartile value (Q3)
Step 4: Calculate inter quartile range (IQR) =Q3-Q1
Step 5: Find the extreme value factor (EVF) from the list of data points
Step 6: Calculate lower boundary value based on the following formula.
Lower boundary = Q1-(EVF*IQR)
Step 7: Calculate upper boundary value based on the following formula
Upper boundary = Q3+ (EVF*IQR)
Step 8: Any data points outside the lower and upper boundary value are extreme values.

From the above lower and upper boundary values, the data points below and above the lower and upper boundary values should be considered as outliers. Here data point 11.6 is below the lower boundary values so it should be a outlier point and the data point 48 is above the upper boundary value so it is also considered as a outlier point. The steps for detecting extreme values in the data using IQR are outlined in Table 2.

From the above steps lower and upper boundary values can be calculated by using Eq(4) and Eq(5)

$$\text{Upper Boundary} = Q3 + (EVF*IQR) \qquad (4)$$

$$\text{Lower boundary} = Q1 – (EVF*IQR) \qquad (5)$$

where, EVF – Extreme Value Factor.

From the above upper and lower boundary calculations, the data points above and below are considered as extreme values. These extreme values are also considered as outliers

Decision Tree: Larger programs are usually split into more than one class. A decision tree is a predictive machine-learning method that decides the target value (dependent variable) of a new sample based on various attribute values of the available data.

The interior nodes of a decision tree denote the different attributes, the branches between the nodes notify that the possible values that these attributes can have in the observed samples, while the terminal nodes notify that the final value (classification) of the dependent variable. The attribute that is to be calculated is known as the dependent variable, since its value depends upon, or is determined by, the values of all the other attributes. The other attributes, which help in calculating the value of dependent variable, are known as the independent variables in the dataset.

Decision Tree : J48 Algorithm: Decision tree J48 is the implementation of algorithm ID3 (Iterative Dichotomies 3) developed by the WEKA project team. It is an open source Java implementation of the C4.5 algorithm in the Weak data mining tool. The Decision tree J48 classifier provides the following simple steps.. In order to classify a new item, it first needs to create a decision tree based on the characteristic values of the available training data. So, whenever it encounters a set of items it classifies the attribute that differentiates the various instances most clearly. This feature that is clever to tell us most about the data request so that we can classify them the best is said to contain the highest information gain. Now, among the achievable values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its group have the same value for the target variable, then we stop that branch and assign to it the target value that we have attained. For the other cases, we then look for another attribute that gives us the highest information gain. Hence we carry on in this way until we either get a clear decision of what arrangement of attributes gives us a particular target value, or we run out of attributes. In the event that we run out of attributes, or if we cannot get an unambiguous answer from the available information, we give this branch a target value that the majority of the substances under this branch possess. By checking all the respective attributes and their values with those seen in the decision tree representation, we can assign or predict the target value of this new instance. The following Figure 1 shows a example of decision tree for predicting whether the heart disease is low risk or high risk.

Proposed System: In this paper, a hybrid technique is used to diagnose the heart disease. First, inter quartile filtering approach is implemented. The outliers in a heart disease instances are detected by using IQR. After pre-process the data, the decision tree J48 algorithm is

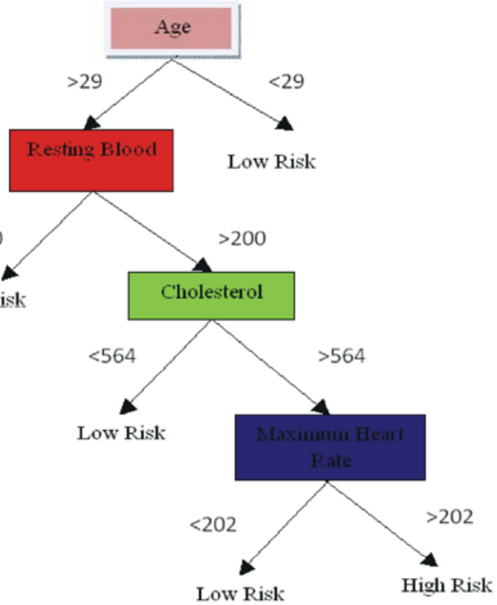

Fig. 1: Decision Tree classifier

executed on pre-processed data instances. The proposed algorithm for diagnosis of heart disease is outlined in following steps.

Step 1: Pre-process data for applying Inter Quartile Range filter

a) Calculate the first and third quartile value(Q1,Q2)
b) Calculate Lower boundary = Q1-(1.5*IQR)
c) Calculate Upper boundary = Q3+ (1.5*IQR)
d) Data points anything outside the lower and upper boundary value is an outlier.

Step 2: Calculate extreme values by following steps

a) Find the extreme value factor (EVF)
b) Calculate Lower boundary = Q1-(EVF*IQR)
c) Calculate Upper boundary = Q3+ (EVF*IQR)
d) Any data points outside the lower and upper boundary value are extreme values

Step 3: Decision tree J48 classification algorithm is applied in pre-processed data using WEKA

Step 4: Diagnosis of heart disease patient either high risk or low risk with better accuracy using J48.

Experimental Study and Results

Data SetL The Cleveland Clinic Foundation heart disease dataset has been used in this study, which is available online at [10]. It consists of 303 constant instances and without missing values. Each instance contains 14 attributes in addition to the output class, 54.5% of the instances are for patients with low risk of developing a heart failure, while the remaining 45.5% are for patients with different risk levels. Details of the heart disease dataset are presented in Table 3.

Validation Method: In order to examine the overall performance of the decision tree classifier, the 10-fold cross-validation method is used to measure the classifiers' performance. This method is usually utilized to maximize the use of the data set. The data set is arbitrarily partitioned into 10 equal subsets. Each one of them contains approximately the same proportion of different class labels. Of the 10 subsets, a single subset is retained as a testing data and the remaining 9 subsets as the training data. The cross-validation method is then repeated 10 times, until each one of the 10 subsets was used accurately once as a testing set. The results can then be averaged to estimate the classifier’s performance. The advantage of this model is that all subsets are used for both training and testing and each subset is used for testing exactly once [13, 14].

Table 3: Heart Disease Attributes Description

S.No	Attributes	Types	Descriptions
1	Age	Continuous	Age in years
2	Sex	Discrete	Gender (1: male and 0: female)
3	CP	Discrete	Chest pain1: typical angina 2: atypical angina 3: non-anginal pain 4: asymptomatic
4	Trestbps	Continuous	Resting blood pressure in mm Hg
5	Chol	Continuous	Serum cholesterol in mg/dl
6	Fbs	Discrete	Fasting blood sugar > 120 mg/dl 1:true 0:false
7	Restecg	Discrete	Resting electrocardiography results 0: normal 1: having ST-T wave abnormality 2: left ventricular hypertrophy
8	Thalach	Continuous	Maximum heart rate
9	Exang	Discrete	Exercise induced angina (1: yes and 0: no)
10	Oldpeak	Continuous	Depression induced by exercise relative to rest
11	Slope	Discrete	The slope of the peak exercise ST segment 1: up sloping 2: flat 3: down sloping
12	Ca	Discrete	Number of major vessels coloured by fluoroscopy (from 0 to 3)
13	Thal	Discrete	3: normal 6: fixed defect 7: reversible defect
14	Class	Discrete	The predicted attributes 1: No risk 2: High risk

Table 4: The Confusion Matrix Of The Decision Tree Model After Removing Outlier

		Predicted Class	
		a	b
Actual Class	a	132	2
	B	37	107

Classifier Assessment: The overall performance of the predictive model for heart failure risk assessment can be calculated using a range of statistical methods including sensitivity, specificity and accuracy [14]. These calculations can be made based on the numbers of a correctly and incorrectly predicted risk levels, which are presented in the confusion matrix as integer values [15]. The confusion matrix plots the true class of instances (i.e. gold standard) in a classification problem against the predicted class, which generated by the predictive model. These will be represented as true positive (TP), false positive (FP), true negative (TN) and false negative (FN) [14, 15].Let's consider that (a, b) are matched (low, High) risk levels respectively. Then the following Table 4 presents the confusion matrix of the heart failure predictive model.

where, sensitivity and also called the true positive rate (TPR), refers to the classifier's capacity to identify a risk level correctly, while the specificity refers to the classifier's ability to exclude the other risk levels correctly (identifies the negative cases). The classification accuracy is the overall precision of the model, it can be calculated as the sum of true results divided by the total number of the examined test set [16, 17]. The sensitivity, specificity, precision and accuracy of a multi output classification problem can be expressed mathematically as follows

$$Sensitivity = \frac{\sum_{i=1}^{1} TP_1}{\sum_{i=1}^{1}(TP_1 + FN_{\downarrow})}$$

$$Specificity = \frac{\sum_{i=1}^{2} TN_1}{\sum_{i=1}^{2}(TN_1 + FP_1)}$$

$$Precision = \frac{\sum_{i=1}^{1} TP_i}{\sum_{i=1}^{l}(TN_1 + FP_1)}$$

$$Accuracy = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_1 + FN_1 + FP_i + TN_i}}{1}$$

Table 5: Detailed Performance Of The Proposed Model

Output classes	Sensitivity	Specificity	Precision	F-Measure	ROC Area
a	0.805	0.268	0.781	0.793	0.767
b	0.732	0.195	0.759	0.745	0.767
Average	0.772	0.235	0.771	0.771	0.767

Table 6: Comparison of Different Models

S.No	Year	Author	Techniques	Accuracy
1.	1989	Gennari *et al*	Clustering	78.9%
2.	1998	Aha & Kibler	NTgrowth	77%
1.			C4.5	74.8%
3.	2005	Polat *et al.*	Artificial Immune System	84.5%
4.	2007	Polat *et al.*	Fuzzy-AIRS–knearest Neighbor	87%
5.	2008	Robert Detrano	Logistic Regression	77.0%
6.	2009	Tu *et al.*	J48	78.9%
1.			Bagging	81.41%
7.	2009	Das *et al.*	NN	89.01%
8.	2010	Rajkumar *et al.*	Decision List	52%
9.	2010	Srinivas	NB	80.46%
10.	2010	M.Anbarasi	Classification	88.3
11.	2010	Mahmood *et al.*	C4.5	76.51%
12.	2012	Shouman	K means with decision tree	83.9%
13.	2013	Pandey *et al.*	Decision tree with error pruning method	75.73%
14.	2013	Chaurasia *et al*	CARD	83.49%
1.			DT	82.50%
1.			ID3	72.93%
15.	2013	Bohacik	Decision tree	77.65%
16.	2014	M.A.Nishara *et al*	C4.5 algorithm, MAFIA and K means	89%
17.	2014	Bashir *et al.*	NB and SVM	81.82%
18.	2014	Uppin *et al.*	C4.5	85.96%
19.	2015	Nguyen *et al.*	GA	78.78%
20.	2015	Nguyen *et al.*	Wavelet Tranformation	81.01%
21.	2016	Proposed Model	IQR and J48	99.76%

RESULTS AND DISCUSSION

Roughly three hundred and three instances are used in this experiment, which derived from the Cleveland Clinic Foundation heart disease dataset [18]. Among these instances, 53.87% were for healthy individuals with no likelihood of developing heart failure, while the remaining 40.07% was for patients with different risk levels. The major contribution of this study is to adopt a proposed model that is able to early predict the likelihood of developing heart failure using J48 decision tree classifier. In contrast to previous studies, two different risk levels of heart failure can be predicted in this model. The following Table 5 illustrates the detailed performance of each class using 10-fold cross validation method.

As can be seen from the above table, class ‘a’ that refers to healthy individuals with low risk levels. While the lowest sensitivity was in detecting the high risk. This is due to the fact that this stage describes the shift from low risk ratio to a higher level and starting a serious threat that could possibly lead to death. In contrast to the previous models, the proposed model shows impressive results with 99.67% of accuracy in predicting two different risk levels of heart failure using J48 decision tree classifier. The following Table 6 illustrates our proposed model against existing heart disease diagnosis models.

CONCLUSION

Early detection of heart disease is important to save life. There are number of studies have been reported focusing on heart disease diagnosis. These studies applied different methods to the given problem and achieved high classification accuracies using the Cleveland heart disease dataset taken from UCI machine learning repository. Data mining techniques play an important role in finding samples and extracting knowledge from large volume of data. Understanding the usefulness of data mining for supporting in the diagnosis of heart disease is so important. In this paper presents a risk prediction model of developing heart failure using inter Quartile range filter and j48 decision tree classifier. 10-fold cross validation technique has been applied to a performance evaluation. Statistical measurements (i.e. accuracy, specificity, sensitivity) were used to evaluate the proposed model. The experimental results achieved 99.76% classification accuracy for heart disease diagnosis. The results strongly suggest that proposed IQR and J48 classifier based on a medical decision making method can assist in the diagnosis of heart disease. Also It is very helpful to provide better patient care and effective diagnostic capabilities.

REFERENCES

1. Nguyen, T., K. Abbas, C. Douglas and N. Saeid, 2015a. Classification of healthcare data using genetic fuzzy logic system and wavelets. Expert Systems with Applications, 42(4): 2184-2197.
2. Nguyen, 15. bT., K. Abbas, C. Douglas and N. Saeid, 2015b. Medical data classification using interval type-2 fuzzy logic system and wavelets. Applied Soft Computing, 30: 812-822.
3. Aha, D. and D. Kibler, 1988. Instance-based prediction of heart-disease presence with the Cleveland database. Technical Report, University of California, Irvine, Department of Information and Computer Science, Number ICS-TR-88-07.
4. Detrano, R., A. Janosi, W. Steinbrunn, M. Pfisterer, J. Schmid, S. Sandhu, *et al.*, 1989. International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64: 304-310.
5. Gennari, J.H., P. Langley and D. Fisher, 1989. Models of incremental concept formation. Artificial Intelligence, 40: 11-61.
6. Nishara Banu, M.A. and B. Gomathy, 2014. Disease Forecasting System Using Data Mining Methods.
7. Rajkumar, A. and G.S. Reena 2010. Diagnosis of Heart Disease Using Data mining Algorithm, Global Journal of Computer Science and Technology, 10 (Issue 10).
8. Srinivas, K., 2010. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques, IEEE Transaction on Computer Science and Education (ICCSE), pp: 1344-1349.
9. Anbarasi, M., 2010. Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology, 2(10) 5370-5376.
10. Cleveland database: http:// archive.ics.uci.edu/ ml/ datasets/ Heart+Disease
11. Pandey, A.K., P. Pandey, K.L. Jaiswal and A.K. Sen, 2013. A heart disease prediction model using decision tree, IOSR Journal of Computer Engineering (IOSR-JCE), 12(6): 83-86.

12. Bashir, S., U. Qamar and M.Y. Javed, 2014. An ensemble based decision support framework for intelligent heart disease diagnosis, International Conference on Information Society (i-Society 2014), London, IEEE, 2014.
13. Silva, F.R., V.G. Vidotti, F. Cremasco, M. Dias, E.S. Gomi and V.P. Costa, 2013. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry, Arq. Bras. Oftalmol, 76(3).
14. Gupta, N., A. Rawal, V.L. Narasimhan and S. Shiwani, 2013. Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data, IOSR Journal of Computer Engineering (IOSR-JCE), 11(5).
15. Elkan, “Evaluating Classifiers”, Available at: elkan/ 250Bwinter2012/ classifiereval.pdf, Accessed in: 10 March 2015
16. Chaurasia, V. and S. Pal, 2013. Early prediction of heart diseases using data mining techniques, Caribbean Journal of Science and Technology, 1: 208-217.
17. Uppin, S.K. and M.A. Anusuya, 2014. Expert system design to predict heart and diabetes diseases, International Journal of Scientific Engineering and Technology, 3(8): 1054-1059.
18. Mahmood, M. and M.R. Kuppa, 2010. Early detection of clinical parameters in heart disease by improved decision tree algorithm, 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems, Warangal, IEEE, 2010.