Journal of Engineering and Applied Sciences 12 (4): 920-928, 2017
ISSN: 1816-949X

A Data Mining Technique for Prediction of Chest Pain using Medical Laboratory Dataset

[1]A. Lourdu Caroline and [2]S. Manikandan
[1]Department of Computer Science, Mother Teresa Women's University, Kodaikanal, India
[2]Department of Computer Scienceand Eng., Sriram Engineering College, Chennai, India

Abstract: Data mining techniques have been used in medical research for many years and have been known to be effective. In order to solve such problems as long-waiting time, congestion and delayed patient care, faced by emergency departments, this study concentrates on building a hybrid methodology, combining data mining techniques such as association rules and classification trees. The methodology is applied to real-world emergency data collected from a hospital and is evaluated by comparing with other techniques. The methodology is expected to help physicians to make a faster and more accurate classification of chest pain diseases.

Key words: Data mining, medical decisions, medical domain knowledge, chest pain, combining

INTRODUCTION

Chest pain may be caused by poor blood flow to the heart leading to angina or by a sudden blockage in the coronary arteries resulting in a heart attack. Other causes of chest pain include indigestion, reflux, muscle strain, inflammation in the rib joints near the breastbone and herpes zoster or shingles. Chest pain may also be caused by problems in your lungs, esophagus, muscles, ribs, or nerves, for example. Some of these conditions are serious and life threatening. Others are not. If you have unexplained chest pain, the only way to confirm its cause is to have a doctor evaluate you.

Chest pain refers to pain felt anywhere in the chest area from the level of your shoulders to the bottom of your ribs. It is a common symptom. There are many causes of chest pain. It can often be difficult to diagnose the exact cause of chest pain without carrying out some tests and investigations.

Common causes of chest pain: There are various possible causes of chest pain. Below is a brief overview of some of the more common causes.

Angina: Like any other muscle in the body, the heart (coronary) muscle needs a good blood supply. The coronary arteries supply blood to the heart muscle. Angina is a pain that comes from the heart. It is usually caused by narrowing of the coronary arteries. This narrowing causes a reduced blood supply to part or parts, of the heart muscle. The narrowing is caused by fatty patches or plaques (atheroma) which develop within the inside lining of arteries. Plaques of atheroma can form gradually over years. They may be in one or more places in the coronary arteries.

The chest pain caused by angina may feel like an ache, discomfort or tightness across the front of your chest when you exert yourself. Pain may also (or sometimes just) be felt in your arms, neck, jaw or stomach. Angina pain usually eases within 10 min when you rest.

Heart attack (myocardial infarction): During a heart attack, a coronary artery or one of its smaller branches is suddenly blocked. This means that the part of the heart muscle supplied by this artery loses its blood and oxygen supply. Unless the blockage is quickly removed, this part of the heart muscle is at risk of dying. A heart attack is often referred to as a myocardial infarction. When part of the heart muscle is damaged, it is said to be infarcted. The term myocardial infarction means damaged heart muscle. The blockage of the coronary artery during a heart attack is usually caused by a blood clot. A blood clot may form if there is some atheroma within the lining of the artery. A crack can develop in the patch of atheroma and this can trigger the clotting mechanism in the blood to form a blood clot over the patch of atheroma. Treatment with clot busting medication or a procedure called angioplasty can break up the blood clot. This means that blood flow

Corresponding Author: A. Lourdu Caroline, Department of Computer Science, Mother Teresa Women's University, Kodaikanal, India

through the artery can be restored. If this treatment is given quickly, it can prevent damage to the heart muscle or limit the extent of the damage.

Gastro-oesophageal reflux disease: This is a general term which describes a range of situations including: Acid reflux when acid leaks up (refluxes) from the stomach into the gullet (oesophagus). Oesophagitis when there is inflammation of the lining of the oesophagus. This inflammation is due to irritation of the lining caused by the reflux of stomach acid. When we eat, food passes down the oesophagus into the stomach. Usually a band of muscle (a sphincter) at the bottom of the oesophagus prevents acid reflux from the stomach back up into the oesophagus. If this sphincter is not working well then acid reflux can occur.

Costochondritis: The rib cage is a bony structure that protects the lungs inside. Bones are hard and solid and so don't tend to bend or move. However, the rib cage needs to move as our lungs expand when we take in a breath during breathing. Cartilage is a softer and more flexible material that is found in joints around the body. Cartilage attaches the ribs to the breastbone (sternum) and the sternum to the collar bones (clavicles). This means that the rib cage is able to move during breathing. The joints between each rib and the cartilages are called the costochondral joints. The joints between the cartilages and the sternum are called the costosternal joints. The joints between the sternum and the clavicles are called the costoclavicular joints. Costochondritis causes chest pain, felt at the front of the chest. The pain is typically sharp and stabbing and is worse with movement, exertion and deep breathing.

Strained chest wall muscle: There are various muscles that run around and between the ribs to help the rib cage to move during breathing. These muscles can sometimes be strained and can lead to chest pain in that area. If a muscle is strained, there has been stretching or tearing of muscle fibres, often because the muscle has been stretched beyond its limits. For example, a strained chest wall muscle may sometimes develop after heavy lifting, stretching, sudden movement or lengthy (prolonged) coughing. The chest pain is usually worse on movement and on breathing in.

Anxiety: Anxiety is quite a common cause of chest pain. As well as feeling fearful, worried and tense, anxiety can sometimes cause physical symptoms including chest pain.

In some people, the chest pain can be so severe that it is mistaken for angina. Chest pain due to anxiety is known as Da Costa's syndrome. Da costa's syndrome may be more common in people who have recently had relatives or friends diagnosed with heart problems or in people who themselves have recently had a heart attack. Investigations show that the coronary arteries are normal with no narrowing

Pleurisy: Pleurisy is most often caused by a viral infection. It can cause a 'pleuritic' chest pain which is a sharp, stabbing pain. The pain can be felt anywhere in the chest, depending on the site of the inflammation. The pain is typically made worse by breathing in or by coughing, as this causes the two parts of the inflamed pleura to rub over each other. There are other more serious causes of pleuritic pain but these are much less common than viral pleurisy. Anything that causes inflammation or damage at the edge of the lung next to the pleura can cause pleuritic pain. For example, Pneumonia. A blood clot in the lung (pulmonary embolism or PE). A collapsed lung (pneumothorax)

Peptic ulcer: A common symptom of a peptic ulcer is pain in the upper tummy (abdomen) just below the breastbone (sternum). The pain usually comes and goes and can sometimes be felt as chest pain. Sometimes food makes the pain worse. The pain may wake you from your sleep. Bloating, retching and feeling sick are other symptoms. You may also feel particularly 'full' after a meal. Complications of peptic ulcers can occur in some cases and can be serious. Complications include: bleeding of the ulcer. Perforation the ulcer goes right through (perforates) the wall of the stomach. Food and acid in the stomach then leak into the abdominal cavity. This usually causes severe pain and is a medical emergency.

Shingles: The virus usually affects one nerve only, on one side of the body. Symptoms occur in the area of skin that the nerve supplies. The usual symptoms are pain and a rash. If a nerve supplying the skin on the chest is affected, shingles can cause chest pain. The pain is a localised band of pain and can range from mild to severe.

Pulmonary Embolism (PE): APE usually causes sharp chest pain felt when breathing in (pleuritic chest pain). In a large PE, chest pain can be felt in the centre of the chest behind the sternum. Often you feel like you cannot breathe deeply. You can also feel breathless and the degree of breathlessness will depend on the size and position of the PE. Coughing up blood (haemoptysis), a mildly high temperature (fever) and a fast heart beat rate are other symptoms. You may also feel faint or even collapse because a large blood clot can cause the blood pressure to drop significantly. There may also be symptoms of a DVT such as pain in the back of the calf in the leg, tenderness of the calf muscles or swelling of a leg or foot.

Pneumothorax: A pneumothorax is air that is trapped between a lung and the chest wall. The air gets there either: From the lungs. The common type of pneumothorax is a primary spontaneous pneumothorax. The pneumothorax develops for no apparent reason in an otherwise healthy person. It is thought to be due to a tiny tear in an outer part of the lung. Air then escapes from the lung and gets trapped between the lung and the chest wall. A secondary spontaneous pneumothorax is also possible. The pneumothorax develops as a complication of an existing lung disease which can make the lung more liable to tear. For example, Chronic Obstructive Pulmonary Disease (COPD) from outside the body. For example, an injury to the chest from a stab wound or a car crash can cause a pneumothorax.

Literature review: Many experiments are being carried out for evaluating the performance of Naive Bayes and Decision Tree algorithm. The results observed so far indicate that Naïve Bayes outperforms and sometimes Decision Tree. In addition to that an optimization process using genetic algorithm is also being planned in order to reduce the number of attributes without sacrificing accuracy and efficiency for diagnosing the heart disease. There are many possible algorithms for the diagnosis of heart disease which are: A. Naïve Bayes A Naive Bayes classifier predicts that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature (Soni *et al*., 2011).

This classifier is very simple, efficient and is having a good performance. Sometimes it often outperforms more sophisticated classifiers even when the assumption of independent predictors is far. This advantage is especially pronounced when the number of predictors is very large. One of the most important disadvantages of Naive Bayes is that it has strong feature independence assumptions.

Decision trees: Decision Trees (DTs) are a non-parametric supervised learning method used for classification. The main aim is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The structure of decision tree is in the form of a tree. Decision trees classify instances by starting at the root of the tree and moving through it until a leaf node. Decision trees are commonly used in operations research, mainly in decision analysis. Some of the advantages are they can be easily understand and interpret, robust, perform well with large datasets, able to handle both numerical and categorical data. Decision-tree learners can create over-complex trees that do not generalize well from the training data is one the limitation.

Clustering: Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes, called clusters. It helps users to understand the natural grouping or structure in a data set. Clustering is an unsupervised classification and has no predefined classes. They are used either as a stand-alone tool to get insight into data distribution or as a pre-processing step for other algorithms. Moreover, they are used for data compression, outlier detection, understand human concept formation. Some of the applications are Image processing, spatial data analysis and pattern recognition. Classification via clustering is not performing well when compared to other two algorithms.

All these algorithms are implemented with the help of WEKA tool for the diagnosis of chest pain. Data set of 354 records with 13 attributes. These algorithms have been used for analyzing the chest pain dataset. The classification accuracy should be compared for this algorithm. After the comparison attributes are to be reduced for further purpose.

Anbarasi *et al*. (2010) used decision trees to extract clinical reasoning in the form of medical expert's actions that are inherent in a large number of electronic medical records. The extracted data could be used to teach students of oral medicine a number of orderly processes for dealing with patients with different problems depending on time.

Ilayaraja and Meyyappan (2013) utilized a C4.5 algorithm to build a decision tree in order to discover the critical causes of type II diabetes. She has learned about the illness regularity from diabetes data and has generated a set of rules for diabetes diagnosis and prediction.

Bhatla (2012) adopted an association algorithm to find the relationship between diagnosis and prescription. They stated that purchases and medical bills have much in common. Therefore, the Apriori algorithm was useful to figure out large item sets and to generate association rules in medical billing data.

Soni *et al*. (2011) used the Apriori algorithm to mine the rules for the compatibility of drugs from prescriptions to cure arrhythmia in the traditional Chinese medicine database. The experimental results showed that the drug compatibility obtained by the Apriori algorithm is generally consistent with the traditional Chinese medicine for that disease.

Koh and Tan (2011) discovered 'treatment pathways' through mining medical treatment procedures in the emergency department. They found that the workload in the emergency department varies depending on the number of presented patients and is not affected by the type of procedure carried out.

Peter and Somasundaram (2012) has presented a complementary perspective on the activities of the emergency department for specific patient groups: over 75 year old and under 75 year old patients. She thought once validated, these views would be used as decision support tools for delivering better care to this population.

Manikantan and Latha (2013) found a way to raise the accuracy of triage through mining abnormal diagnostic practices in the triage. A two-stage cluster analysis (Ward's method, K-means) and a decision tree analysis were performed on 501 abnormal diagnoses done in an emergency department.

MATERIALS AND METHODS

There are two main issues that affect the performance of Decision Trees; the data discretization method used and the type of Decision Tree used. Reduced error pruning is shown to further improve decision tree performance. The proposed methodology involves systematically testing different discretization techniques, multiple classifiers voting technique and different Decision Trees type in the diagnosis of Chest pain patients. Different combinations of discretization methods, decision tree types and voting are tested to identify which combination will provide the best performance in diagnosing heart disease patients. A test was implemented using MATLAB.

Data source: A total of 603 records with 76 medical attributes (factors) all attributes are numeric-valued have obtained from the Medical Laboratory. While the databases have 76 raw attributes, only 14 of them are actually used below we have lists the attributes.

- Age: It will take age in years as input.
- Sex: It will take two values as input, i.e., value 1: Male and value 0: Female
- Chest pain type: It will take four value as input which shows the chest pain type as value 1: typical type-1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic
- Trest blood pressure: resting blood pressure (in mm Hg on admission to the hospital)
- Chol: serum cholesterol in mg dL^{-1}
- fasting blood sugar: it will take two values as input, i.e., value 1 for fbs > 120 mg dL^{-1} and value 0: for fbs <120 mg dL^{-1}
- Restecg: Resting electrographic results will take 3 value as input, i.e., value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy
- Thalach: Maximum heart rate achieved by the patient
- Exang: Exercise induced angina will take two values as input, i.e., value 1: yes and value 0: no
- Old peak: ST depression induced by exercise relative to rest
- Smoking: will take two values as input, i.e., value 1: yes and value 0: no
- Gastro-oesophageal reflux disease:will take two input values : 0: Heartburn 1: Normal Value
- Costochondritis: will take three values as input: 0: Mild 1:Severe 2: Normal
- Strained chest wall muscle: will take two values as input: 0: tenderness 1: Normal
- Anxiety: will take two values as input 0: Normal and 1: yes
- Pleurisy: will take two values as 0: No 1: yes
- Peptic Ulcer: 0: gastric obstruction 1: perforation 2: Normal Stomach
- Shingles: Contain live virus 0: Normal Value 1: Shingles affected in the Skin
- Pulmonary Embolism (PE): Blood Test: D-Dimer test results take 2 input values, i.e, 0: Normal (No D-Dimer protein clot) 1: Blood Clot abnormalities
- Pneumothorax: X-Ray results will take 2 input values, i.e., 0: Normal Value 1: Collapsed Lung

Classification in data mining: Classification which can be described as analyzing of data in order to obtain models that are used to characterize data classes is the most usual task in data mining. This task focuses on predicting the value of the decision class for an input object within set of classes which are predefined. There are many different classification approaches in the literature. Each of them are developed and proposed by various researchers. The most known techniques are decision tree based classification, neural network based classification, statistical classification, rough set based and genetic algorithm classifiers. We can divide data classification task into two phases. The first phase is called as the learning step and the second phase is called as testing step. In the learning step, a model which defines predetermined set of classes will be constructed. This operation is made by analyzing a set of training data. For this data, each set of elements are assumed to belong a specific, predefined class. In the testing step which is the second leg of classification, the constructed model is tested by using different set of data. In this phase, the accuracy of the classification is estimated by using one of the several proposed approaches. If the estimation of accuracy shows an adequate result, then the generated model can be used for classification of new input sets whose class labels are unknown. Before applying the generated classification model, some data preprocessing techniques can be executed in order to obtain a better accuracy and efficiency for the classification model (Peter and Somasundaram, 2012).

Data cleaning: The removal of noisy data and filling of missing data is considered in this step. There are many different approaches designed by researches in the literature for data cleaning issue (Manikantan and Latha, 2013).

Feature selection: In the initial data set, there may be some attributes which are not related or not important for the learning step of model. The removal of inappropriate and unnecessary attributes from the data set is applied on this step of classification. After this process, the reduced data set is used to generate the classification model. For feature selection task, numerous applications are implemented by various researchers (Palaniappan and Awang, 2008).

Data discretization: The data set which will be used by classification algorithms may have some attributes that cannot be handled by the algorithm itself without applying some transformations. Such as, numerical scaled values are needed to be converted into nominal or discrete values in order to make some of the algorithms work correctly. This conversion step is considered in the data discretization part of classification. More detailed discussion about discretization techniques can be found by Peter and Somasundaram (2012).

Experimental results: The basic phenomenon used to classify the heart disease classification using classifier is its performance and accuracy. The performance of a chosen classifier is validated based on error rate and computation time. The classification accuracy is predicted

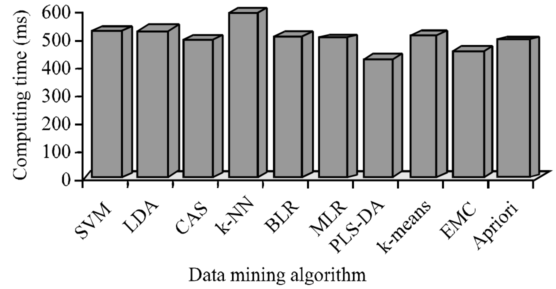

Fig. 1: Performance of computing time

Table 1: Confusion matrix

Predicted	Classified as healthy (0)	Classified as not healthy (1)
Actual healthy (0)	TP	FN
Actual as not healthy (1)	FP	FN

in terms of Sensitivity and Specificity. The computation time is noted for each classifier is taken in to account. Classification matrix displays the frequency of correct and incorrect predictions. It compares the actual values in the test dataset with the predicted values in the trained model. In this example, the test dataset contained 208 patients with heart disease and 246 patients without heart disease (Table 1).

The step 5 consists of values of different classification. According to these values the accuracy was calculated. From Fig. 1-3 represents the resultant values of above classified dataset using data mining supervised classification algorithms and it shows the highest accuracy and lowest computing among the three. It is logical from chart that compared on basis of performance and computing time, precision value, error rate (10 fold cross validation, bootstrap validation) and finally the highest accuracy and again lowest computing time. PLS-DA algorithm shows the superior performance compared to other algorithms.

These metrics can be derived from the confusion matrix and can be easily converted to True-Positive (TP) and False-Positive (FP) metrics:

Precision = TP/(TP+FP) Recall = TP/(TP+FN)

- True Positive (TP): Total percentage of members classified as class A belongs to Class A
- False Positive (FP): Total percentage of members of Class A but does not belong to Class A.
- False Negative (FN): Total percentage of members of Class A incorrectly classified as not belonging to Class A
- True Negative (TN): Total percentage of members which do not belong to Class A are classified not a part of Class A. It can also be given as (100%-FP)

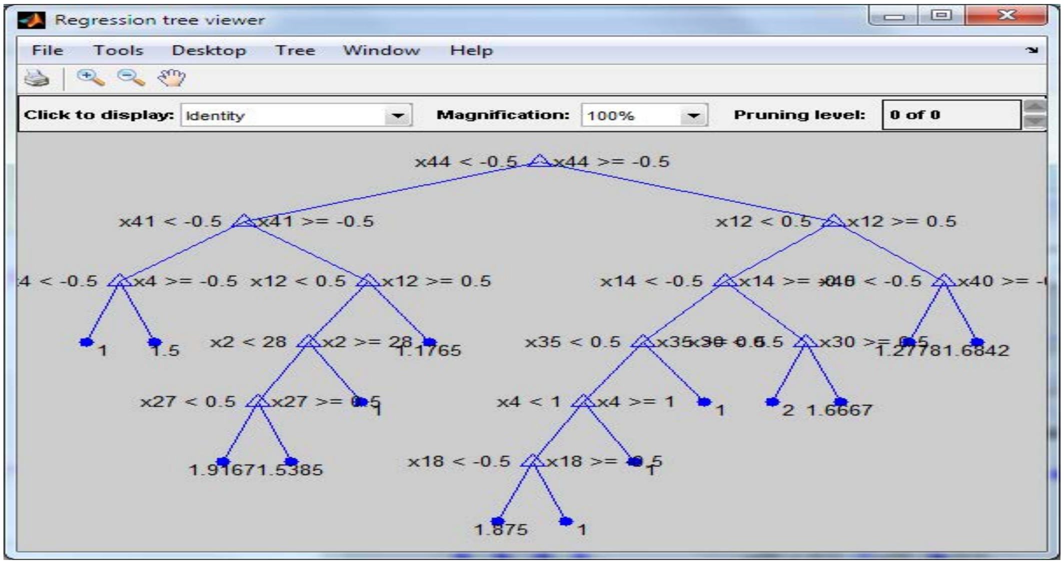

Fig. 2: Proposed modern decision tree classifier

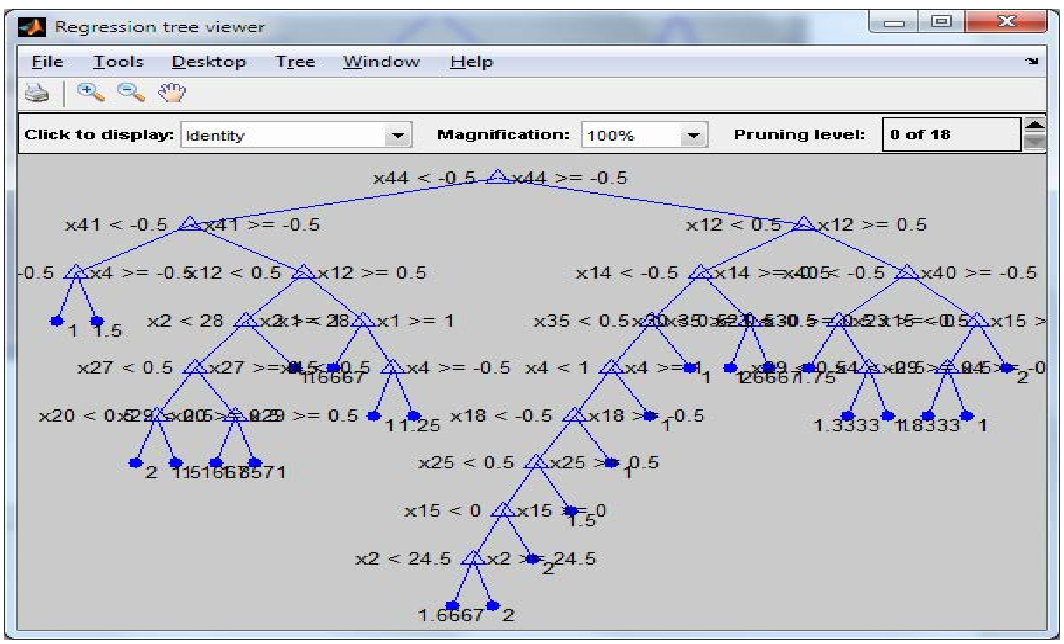

Fig. 3: Decision tree classifier for our algorithm

RESULTS AND DISCUSSION

Our proposed algorithm runs efficiently on large databases and has the capability of handling thousands of input variables. It generates the generalization error as the effective method for estimating missing data and maintains accuracy when large proportion of the data are missing. Our proposed that has been generated can be saved in order to make comparative study about the features of the attributes. To measure the effectiveness of the approach experiments have been conducted.

Meanwhile, decision trees are constructed in a top-down recursive divide-and-conquer manner and the compatibility of decision trees degrades because the output is limited to one attribute. Trees created from the numeric datasets seems to be more complex and also when the database is large the complexity of the tree increases. In comparison with the 16 algorithms the time complexity of Decision trees increases exponentially with the tree height. Hence shallow trees tend to have large number of leaves and high error rates.

As the tree size increases, training error decreases. However, as the tree size increases, testing error decreases at first since we expect the test data to be similar to the training data but at a certain point, the training algorithm starts training to the noise in the data, becoming less accurate on the testing data. At this point we are no longer fitting the data and instead fitting the noise in the data. This is called over fitting to the data, in which the tree is fitted to spurious data. As the tree grows in size, it will fit the training data perfectly and not be of practical use for other data such as the testing set.

Performance measures: In this approach, the classification accuracy rates for the datasets were measured. For example, in the classification problem with two-classes, positive and negative, a single prediction has four possibility. The True Positive rate (TP) and True Negative rate (TN) are correct classifications. A False Positive (FP) occurs when the outcome is incorrectly predicted as positive when it is actually negative. A False Negative (FN) occurs when the outcome is incorrectly predicted as negative when it is actually positive (Table 2).

Accuracy: It refers to the total number of records that are correctly classified by the classifier.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}} \tag{2}$$

Classification error: This refers to the misclassified datasets from the correctly classified records.

True Positive Rate (TP): It corresponds to the number of positive examples that have been correctly predicted by the classification model

False Positive Rate (FP): It corresponds to the number of negative examples that have been wrongly predicted by the classification model

Kappa statistics: A measure of the degree of nonrandom agreement between observers or measurements of the same categorical variable.

Table 2: Confusion table

	Disease	
Prediction	+	-
Test		
+	True Positive (TP)	False Negative (FP)
-	False Negative (FN)	True Negative (TN)

Precision: It is the fraction of retrieved instances that are relevant:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{3}$$

Recall: It is the fraction of relevant instances that are retrieved:

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{4}$$

Root-mean-squared-error: It is a statistical measure of the magnitude of a varying quantity. It can be calculated for a series of discrete values or for a continuously varying function. Since the class label prediction is of multi-class, the result on the test set will be displayed as a two dimensional confusion matrix with a row and a column for each class. Each matrix element shows the number of test cases for which the actual class is the row and the predicted class is the column. Finally, the error rate is one minus this.

ROC curves depict the performance of a classifier without regard to class distribution or error costs. They plot the number of positives included in the sample on the vertical axis, expressed as a percentage of the total number of positives, against the number of negatives included in the sample, expressed as a percentage of the total number of negatives, on the horizontal axis. Information retrieval researches define parameters called recall and precision:

$$\text{Recall} = \frac{\text{No. of documents retrieved that are relevant}}{\text{Total No. of documents that are relevant}}$$

$$\text{Precision} = \frac{\text{No. of documents retrieved that are relevant}}{\text{No. of documents retrieved that are retrived}}$$

F-measure is another information retrieval measure that is calculated from TP, FP, FN or recall or precision values:

$$\text{f\_measure} = \frac{2\times\text{recall}\times\text{Precision}}{\text{recall}+\text{Precision}}$$

$$f\_measure = \frac{2\times TP}{2\times TP + FP + FN}$$

Decision tree classifier:

- Accuracy = 50.68
- Error = 0.49
- Specificity = 0.9753
- Sensitivity = 0.2527
- Positive predictive value = 0.5374
- Negative predictive value = 0.9200
- Positive likelihood = 1.3052
- Negative likelihood = 0.0977
- Elapsed time is 0.592824 sec

Proposed fast boost decision tree classifier:

- Accuracy = 68.8
- Error = 0.31
- Specificity = 0.5185
- Sensitivity = 0.6593
- Positive predictive value = 0.5753
- Negative predictive value = 0.6061
- Positive likelihood = 1.5221
- Negative likelihood = 0.7302
- Elapsed time is 0.481823 sec

In the Fig. 4 and 5, accuracy of our proposed algorithms is better than the other sixteen data mining techniques. The classification error also is lesser than other sixteen data mining technique. Compare with J48, decision tree data mining technique is lesser accuracy and lesser classification error. Here in this work decision tree technique has same value of accuracy and classification error. The Kstar and simple Logistics classifier has same accuracy (Table 3).

Table 3: Performance analysis of various classifiers

Classifier	Accuracy (%)	True positive rate	False positive rate	Precision (%)	Recall (%)	Classification error (%)	Kappa statistics	RMS error
Decision tree	50.68	0.507	0.230	0.478	0.507	49.32	0.211	0.404
Random forest	63.34	0.633	0.254	0.570	0.633	36.66	0.354	0.313
J48	64.45	0.500	0.544	0.521	0.500	35.55	0.344	0.310
PRISM	63.45	0.750	0.350	0.825	0.750	36.55	0.635	0.718
IBK	54.50	0.871	0.594	0.571	0.871	45.50	0.484	0.539
Naïve bayes	53.75	0.571	0.594	0.528	0.571	46.25	0.484	0.539
SMO	54.00	0.643	0.465	0.629	0.643	46.00	0.589	0.632
Bayes net	52.50	0.681	0.502	0.625	0.681	47.50	0.585	0.536
Simple logisitics	49.80	0.547	0.450	0.520	0.547	50.20	0.580	0.598
KStar	50.25	0.564	0.459	0.561	0.564	49.75	0.480	0.654
NNge	51.20	0.655	0.500	0.540	0.655	48.80	0.490	0.655
PART	49.99	0.652	0.550	0.650	0.652	50.01	0.500	0.654
ZeroR	52.25	0.584	0.546	0.643	0.584	47.75	0.480	0.680
AD tree	61.18	0.500	0.640	0.684	0.500	38.82	0.465	0.500
Simple cart	60.16	0.600	0.490	0.682	0.600	39.84	0.470	0.654
Multi layer perception	61.58	0.546	0.500	0.855	0.546	38.42	0.495	0.356
Proposed	69.51	0.656	0.215	0.414	0.674	30.25	0.321	0.204

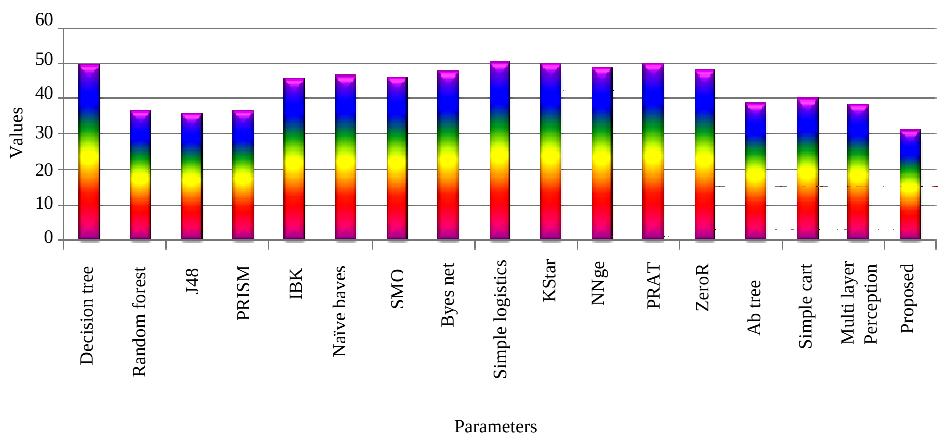

Fig. 4: Performance analysis of various classifiers error of algorithm

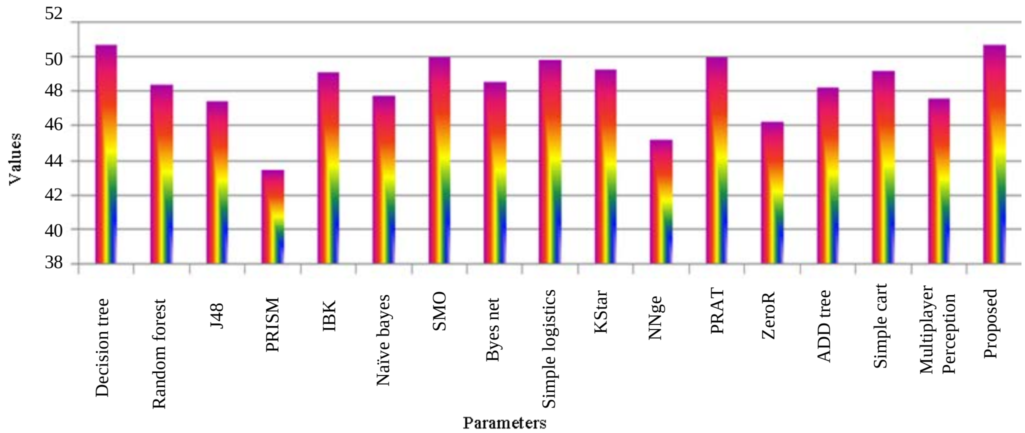

Fig. 5: Performance analysis of various classifiers accurracy of algorithm

CONCLUSION

Health care related data are huge in nature and they arrive from various birthplaces all of them not wholly suitable in structure or quality. These days, the utilization of knowledge and experience of copious specialists and medical screening data of patients collected in a database during the diagnosis process has been widely accepted. The proposed method is applied over a Chest Pain Patient dataset of 354 records of patients suffering from various chest pain related diseases. The prediction results are encouraging and the efficiency of the method in frequent itemset generation is better than existing methods. In this paper presented an efficient approach for fragmenting and extracting substantial forms from the chest pain data warehouses for the efficient prediction of chest pain.

SUGGESTIONS

In the future research, we have planned to design and develop an efficient chest pain prediction system with patient prescription support using the web mining and data warehouse techniques.

REFERENCES

Anbarasi, M., E. Anupriya and N.C.S.N. Iyengar, 2010. Enhanced prediction of heart disease with feature subset selection using genetic algorithm. Int. J. Eng. Sci. Technol., 2: 5370-5376.

Bhatla, K.N., 2012. An analysis of heart disease prediction using different data mining techniques Intl. J. Eng. Res.Technol., Vol. 1.

Ilayaraja, M. and T. Meyyappan, 2013. Mining medical data to identify frequent diseases using Apriori algorithm. Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22, 2013, IEEE, Karaikudi, India, ISBN:978-1-4673-5845-3, pp: 194-199.

Koh, H.C. and G. Tan, 2011. Data mining applications inhealthcare. J. Healthcare Inf. Manage., 19: 65-72.

Manikantan, V. and S. Latha, 2013. Predicting the analysis of heart disease symptoms using medicinal data mining methods. Intl. J. Adv. Comput. Theory Eng., 2: 46-51.

Palaniappan, S. and R. Awang, 2008. Intelligent heart disease prediction system using data mining techniques. Proceedings of the International Conference on Computer Systems and Applications, March 31-April 4, 2008, Doha, pp: 108-115.

Peter, T.J. and K. Somasundaram, 2012. An empirical study on prediction of heart disease using classification data mining techniques. Proceedings of the International Conference on Advances in Engineering, Science and Management, March 30-31, 2012, Nagapattinam, Tamil Nadu, pp: 514-518.

Soni, J., U. Ansari, D. Sharma and S. Soni, 2011. Predictive data mining for medical diagnosis: An overview of heart disease prediction. Int. J. Comput. Appl., 17: 43-48.