OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Heart Disease Prediction Using Associative Relational Classification Technique (Acar) With Som Neural Network

Amit Sahu[1], Mayank saxena[2]
Department of Computer Science & Engineering PCST, Bhopal, India

ABSTRACT: Mining of medical diagnoses of data is very difficult task in current data mining approach. The heart disease data is collective information of blood pressure, Cholesterol problem, diabetes and another complex disease. The relational of one disease to another is rare so classification task is very difficult. So prediction of heart disease is very critical. in the process of data mining rule based classification technique used for prediction. The rule based classification technique based on association rule mining. The better rule mining technique the better classification and predication of heart disease. in this paper proposed a association based ensemble classification method for heart disease prediction. the association ensemble classifier based on association rule mining and self –organized map network model. For the association rule used Apriori-like algorithm. This algorithm generates numbers of rules for all combination of factor of heart disease and divided into different level such as high level , middle level and low level, the all level ensemble through SOM network and generate optimized set of heart disease prediction.

Keywords: Heart Disease, Associative Classification, Ensemble and SOM Network

I. INTRODUCTION

The diversity of lifestyle invites much disease in our body. In all disease heart disease which is also called cardiovascular disease is considered as one of the leading cause of death in the world with high prevalence in the Asia subcontinent [1, 2]. There are several risk factors which account for the heart disease such as age, sex, smoking etc. Patients with Hereditary risk factors (such as: high blood pressure, diabetes) have more chances of heart disease. Some risk factors are controllable. While having so many risk factors, it is a complicated task to analyze heart disease on the basis of patient's report [4]. Particularly, doctors take decision on their intuition and experience rather than on the knowledge-rich data hidden in the database. In healthcare transactions, data is too complex and huge to be processed and analyzed by traditional methods. It requires high skills and experiences for correct decisions. Classification based on association rules, also called associative classification, is a technique that uses association rules to build classifier. Generally it contains two steps: first it finds all the class association rules (CARs) whose right-hand side is a class label, and then selects strong rules from the CARs to build a classifier [5]. In this way, associative classification can generate rules with higher confidence and better understandability comparing with traditional approaches. Thus associative classification has been studied widely in both academic world and industrial world, and several effective algorithms [6] have been proposed successively. However, all the above algorithms only focus on processing data organized in a single relational table. In practical application, data is often stored dispersedly in multiple tables in a relational database. Simply converting multi-relational data into a single flat table may lead to the high time and space cost, moreover, some essential semantic information carried by the multi-relational data may be lost. Thus the existing associative classification algorithms cannot be applied in a relational database directly [7]. We propose a novel algorithm, ACAR, for associative classification which can be applied in multi-relational data environment. The main idea of ACAR is to mine relevant features of each class label in each table respectively, and generate strong classification rules[8]. The ensemble of different rules in different level used SOM based ensemble classifier. The SOM based ensemble classifier classified the data very accurately. The self-organizing map (SOM) is one of the most popular algorithms in the classification of data with a good performance regarding rate of classification[9]. The SOM is a widely used unsupervised neural network for clustering high dimensional input data and mapping these data into a two-dimensional representation space. Self-organizing map is one of the most fascinating topics in the neural network field. The SOM introduced by Kohonen (1982), is a neural network that maps signals from a high-dimensional space to a one- or two-dimensional discrete lattice of neuron units. Each neuron stores a weight. The SOM organizes unknown data into groups of similar patterns, according to a similarity criterion. Such networks can learn to detect regularities and correlations in

their input and adapt their future responses to that input accordingly. An important feature of this neural network is its ability to process noisy data[13]. The map preserves topological relationships between inputs in a way that neighboring inputs in the input space are mapped to neighboring neurons in the map space. The rest of paper is organized as follows. In Section II discuss related work of associative classification. The Section III proposed method for classification. The section IV discusses experimental result and finally followed section V conclusion and future scope

II. RELATED WORK

In this section discuss the related work in the field of medical data classification using associative classification using neural network and other optimization technique. The neural network is important area of research in the field of data mining classification. The neural network optimized the level of classification and improved the ratio of classification.

[1] In this paper author proposed an ensemble method and classification for heart diseases and to improve the decision of the classifiers for heart disease diagnosis. Homogeneous ensemble is applied for heart disease classification and finally results are optimized by using Genetic algorithm. Data is evaluated by using IO-fold cross validation and performance of the system is evaluated by classifiers accuracy, sensitivity and specificity to check the feasibility of our system. Comparison of our methodology with existing ensemble technique has shown considerable improvements in terms of classification accuracy. The focused on the optimized heart disease classification problem. Genetic Algorithm has been found a very good technique for optimization and searching for quality solution. The proposed framework of SVM classifier ensemble and optimization of results using Genetic Algorithm technique improved the classification accuracy as compared to existing work.

[2] In this paper author proposed a genetic algorithm based feature selection for the heart diseases and the details are, we presented a genetic algorithm (GA) based feature-selection method to find informative features that play a significant role in discrimination of samples. Selected subsets from multiple GA runs were used to build a classifier. The proposed approach can be combined with various classifiers to improve classification performance and selection of the most discriminative features. Starting with a set of pre-selected features by using a filter (1000 features), we used a GA combined with Fisher's linear discriminate analysis (LDA) to explore the space of feature subsets. In fact, our proposed approach employs GA and uses the LDA classifier to evaluate the fitness of a given candidate feature subset. An external test set was chosen by using Kohonen self-organizing maps (SOMs) to evaluate the performance of feature selection at the final stage. The proposed method can be used to diagnose CHD in patients without using any angiographic techniques, which may have a high risk of death for the individuals.

[3] In this paper author discussed on data mining method for heart disease, the details are data classification is based on supervised machine learning algorithms which result in accuracy, time taken to build the algorithm. Tanagra tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared. The selection of algorithms is based on their performance, but not around the test dataset itself, and also comprising the predictions of the classification models on the test instance. Training data are produced by recording the predictions of each algorithm, using the full training data both for training and for testing. Performance is determined by running 10-fold cross-validations and averaging the evaluations for each training dataset. Several approaches have been proposed for the characterization of learning domain. the performance of each algorithm on the data attribute is recorded. The algorithms are ranked according to their performance of the error rate. Author also deals with the results in the field of data classification obtained with Naive Bayes algorithm, Decision list algorithm and k-nn algorithm, and on the whole performance made known Naive Bayes Algorithm when tested on heart disease datasets. Naive Bayes algorithm is the best compact time for processing dataset and shows better performance in accuracy prediction.

[4] In this paper author described a heart beat classification system using optimization techniques such as particle swarm optimization, to proposes a novel system to classify three types of electrocardiogram beats, namely normal beats and two manifestations of heart arrhythmia. This system includes three main modules: a feature extraction module, a classifier module, and an optimization module. In the feature extraction module, a proper set combining the shape features and timing features is proposed as the efficient characteristic of the patterns. In the classifier module, a multi-class support vector machine (SVM)-based classifier is proposed. For the optimization module, a particle swarm optimization algorithm is proposed to search for the best value of the SVM parameters and upstream by looking for the best subset of features that feed the classifier. Simulation results show that the proposed algorithm has very high recognition accuracy. This high efficiency is achieved with only little features, which have been selected using particle swarm optimizer.

[5] In this paper author presented a Back propagation neural network and genetic algorithm for medical diseases diagnosis classification, by using the Three-Term Back propagation (TBP) network based on the Elitist

Multiobjective Genetic Algorithm (MOGA). One of the recent MOGAs is a Non-dominated Sorting Genetic Algorithm II (NSGA-II), which is used to reduce or optimize the error rate and network structure of TBP simultaneously to achieve more accurate classification results. In addition accuracy, sensitivity, specificity and 10-fold cross validation are used as performance evaluation indicators to evaluate the outcome of the proposed method.

[6] In This paper author analyses the performance of various classification function techniques in data mining for prediction heart disease from the heart disease data set. The classification algorithms used and tested in work are Logistics, Multi-layer Perception and Sequential Minimal Optimization algorithms. The performance factor used for analyzing the efficiency of algorithm are clustering accuracy and error rate. The result show logistics classification function efficiency is better than multi-layer perception and sequential minimal optimization. Three classification algorithms techniques in data mining are intelligent for predicting heart disease. They are function based Logistic, Multilayer perception and Sequential Minimal Optimization algorithm. By analyzing the experimental results, it is observed that the logistic classification algorithms technique turned out to be best classifier for heart disease prediction because it contains more accuracy and least error rate. In future we tend to improve performance efficiency by applying other data mining techniques and optimization techniques. It is also enhanced by reducing the attributes for the heart disease dataset.

[7] In This paper author describes about a prototype using data mining techniques, namely Naïve Bayes and WAC (weighted associative classifier). It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be established. It can serve a training tool to train nurses and medical students to diagnose patients with heart disease. It is a web based user friendly system and can be used in hospitals if they have a data ware house for their hospital. Presently we are analyzing the performances of the two classification data mining techniques by using various performance measures.

[8] In this paper author discussed on a diseases Using Genetic Algorithm and Ensemble Support Vector Machine, Support vector machine (SVM) is believed to be more efficient than neural network and traditional statistical-based classifiers. an ensemble of SVM classifiers use multiple models to obtain better predictive accuracy and are more stable than models consist of a single model. Genetic algorithm (GA), on the other hand, is able to find optimal solution within an acceptable time, and is faster than dynamic programming with exhaustive searching strategy. By taking the advantage of GA in quickly selecting the salient features and adjusting SVM parameters, it was combined with ensemble SVM to design a clinical decision support system (CDSS) for the diagnosis of patients with severe OSA, and then followed by PSG to further discriminate normal, mild and moderate patients. The results show that ensemble SVM classifiers demonstrate better diagnosing performance than models consisting of a single SVM model and logistic regression analysis.

[9] In this paper The aim of author for this work is to design a GUI based Interface to enter the patient record and predict whether the patient is having Heart disease or not using Weighted Association rule based Classifier. The prediction is performed from mining the patient's historical data or data repository. In Weighted Associative Classifier (WAC), different weights are assigned to different attributes according to their predicting capability. It has already been proved that the Associative Classifiers are performing well than traditional classifiers approaches such as decision tree and rule induction.

[10] In this paper author Enhanced the Prediction of Heart Disease with Feature Subset Selection based on a Genetic Algorithm, Genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient. Thirteen attributes are reduced to 6 attributes using genetic search. Subsequently, three classifiers like Naive Bays, Classification by clustering and Decision Tree are used to predict the diagnosis of patients with the same accuracy as obtained before the reduction of number of attributes. Also, the observations exhibit that the Decision Tree data mining technique outperforms other two data mining techniques after incorporating feature subset selection with relatively high model construction time.

III. PROPOSED METHODOLOGY

In this section discuss the proposed methodology of ensemble associative classification based on SOM network and also discuss the associative classification. the ensemble classification technique improved the classification and prediction of heart disease.

III.A Associative classification (ACAR)

Let D is the dataset. Let I be the set of all items in D and C be the set of class labels. We say that a data case $d_i \in D$ contains $X \subseteq I$, a subset of items, if $X \subseteq d_i$. A class association rule (CAR) is an implication of the form $X \rightarrow c$, where $X \subseteq I$, and $c \in C$. Bing Liu et al. [22] first proposed the AC approach, named classification based on association algorithm (CBA), for building a classifier based on the set of discovered class association rules. The difference between rule discovery in AC and conventional frequent item set mining is that the former

task may carry out multiple frequent item set mining processed for mining rules of different classes simultaneously. Data mining in associative classification (AC) framework usually consists of two steps [10] Generating all the class association rules (CARs) which has the form of Iset＝＞c, where Iset is an item set and c is a class.

Building a classifier based on the generated CARs. Generally, a subset of the association rules was selected to form a classifier and AC approaches are based on the confidence measure to select rules [12].

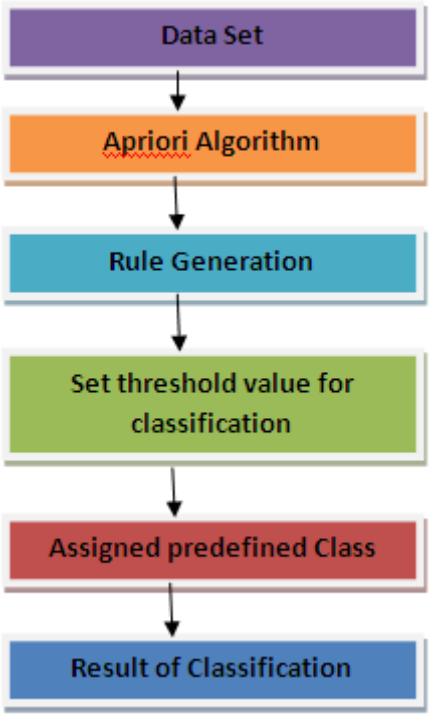

Figure .1 Association Classifications

III.B PROPOSED METHOD

Proposed models are creating for data training for minority and majority class data sample for processing of associative classification level of rules. The associative classification process a data input for training phase for SMOTE and CMTNN sampling technique for classifier. While single-layer SOM networks can potentially learn virtually any input output relationship, SOM networks with single layers might learn complex relationships more quickly[12]. The function SOM creates wiener and successor matrix. For example, a ensemble layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The ensemble -layer network also has connections from the input to all cascaded layers. The additional connections might improve the speed at which the network learns the desired relationship. SOM artificial intelligence model is similar to feed-forward back-propagation neural network in using the back-propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons. Tan-sigmoid transfer function, log - sigmoid transfer function and pure linear threshold functions were used to reach the optimized status.

1. Data are passes through ACR
2. ACR makes a multi-level rule set using rule mining algorithm
3. level of rules going to SOM ensemble process
4. The training phase data are passes through SMOTE AND CMTNN sampler
5. The sampling of data passes through SOM AND balanced the data for minority and majority ratio of class
6. The sampled data assigned to k-type binary class
7. Binary class data are coded in bit form
8. if code bit value is single assigned the class value
9. Else data goes to training phase
10. . Balanced part of training is updated
11. Find accuracy and relative mean Error
12. Exit

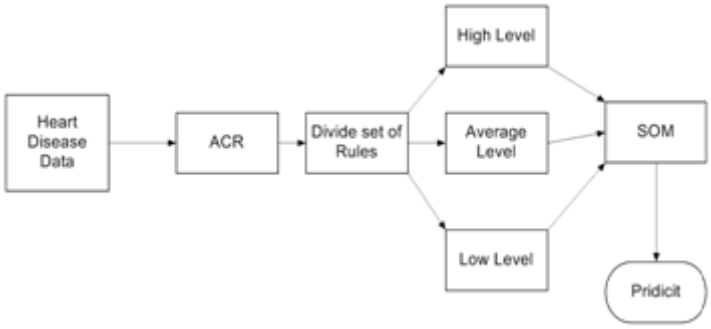

Figure 2 proposed model for ensemble based associative classification with SOM network

IV. EXPERIMENTAL RESULT ANALYSIS

It is simulating on mat lab 7.8.0 and for this work we use Intel 1.4 GHz Machine. MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation Matlab is a software program that allows you to do data manipulation and visualization, calculations, math and programming. It can be used to do very simple as well as very sophisticated tasks. Three datasets (Cleveland, SPECT and Statlog) are obtained from UCI machine learning repository and other is indian dataset [10]. IO-folds cross validation is applied in all experiments. Training and testing sets are generated randomly form the dataset. In table I is the comparison of performance between ACAR and the proposed scheme. The figure clearly shows that classifiers perform best on Cleveland dataset. Considerable performance has also been achieved on other datasets by using ensemble based optimizing technique.

	Method	Support	Confidence	Accuracy (%)	Runtime (sec)
Cleveland	ACAR	0.3	0.5	84.57	7.4375
	AC-EN	0.3	0.5	90.1	6.879
SPECT	ACAR	0.3	0.5	89.36	8.451
	AC-EN	0.3	0.5	92.34	9.678
Statlog	ACAR	0.3	0.5	90.31	5.7812
	AC-EN	0.3	0.5	91.54	4.561
Indian dataset	ACAR	0.3	0.5	81.23	7.891
	AC-EN	0.3	0.5	86.23	5.671

Table I. Maximum accuracy of the ACAR and ensemble ACAR

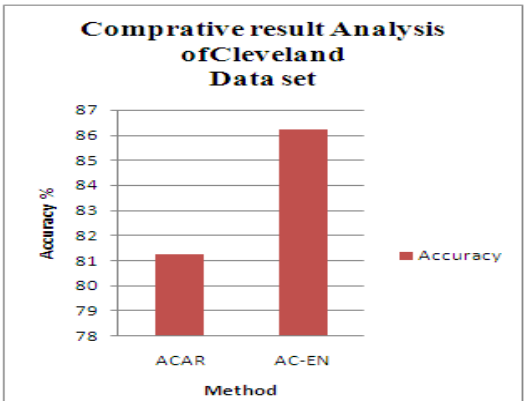

Figure 3 shows that comparative result of Cleveland data set this data set is collection of heart disease. the proposed algorithm shows that better prediction of ACAR algorithm.

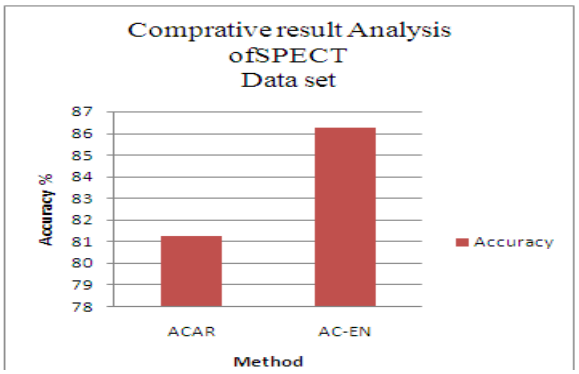

Figure 4 shows that comparative result of SPECT data set this data set is collection of heart disease. The proposed algorithm shows that better prediction of ACAR algorithm**.**

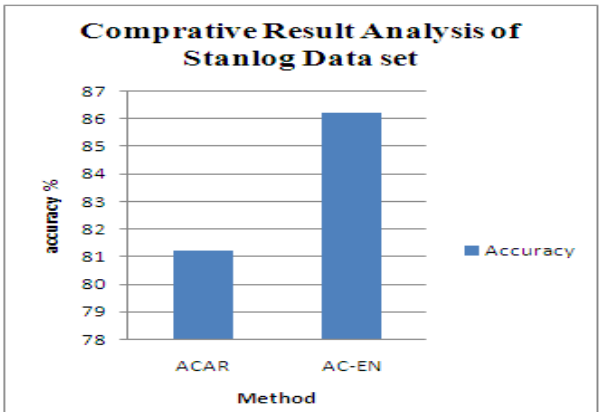

Figure 5 shows that comparative result of STANLOG data set this data set is collection of heart disease. The proposed algorithm shows that better prediction of ACAR algorithm.

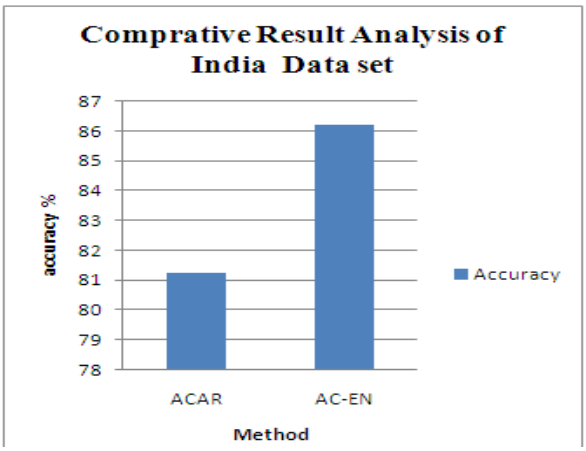

Figure 6 shows that comparative result of Indian data set this data set is collection of heart disease. The proposed algorithm shows that better prediction of ACAR algorithm.

V. CONCLUSION AND FUTURE WORK

Currently, ACAR uses a support-confidence framework to discover frequent item sets and generate classification rules. It may discover more relevant features of each class label by using related measures extending current framework. Also the current algorithm could be improved in terms of efficiency by using the optimization technique. Multiple relational classification algorithm modified by SOM so improved rate of classification in comparison of ACAR. In the process of SOM the calculation complexity are increases, the complexity of time are also increases. Our proposed algorithm test heart disease data set. In this data set the rate of classification is 92%.We also use another data set Indian heart disease and estimate some little bit difference of rate of classification is 91%. The rate of classification increases in previous method on the consideration of

time complexity. In future we minimize the complexity of time and also increase the rate of classification using Meta heuristic function such as ant colony optimization, power of swarm (pos) and dendrites cell algorithm

REFERENCES

[1]. [1] Benish Fida, Muhammad Nazir, Nawazish Naveed, Sheeraz Akram "Heart Disease Classification Ensemble Optimization Using Genetic Algorithm" Ieee 2011. Pp 19-25.

[2]. [2] Mahdi Vasighi, Ali Zahraei, Saeed Bagheri, Jamshid Vafaeimanesh "Diagnosis Of Coronary Heart Disease Based On Hnmr Spectra Of Human Blood Plasma Using Genetic Algorithm-Based Feature Selection" Wiley Online Library 2013. 318-322.

[3]. [3] Asha Rajkumar, G.Sophia Reena" Diagnosis Of Heart Disease Using Data Mining Algorithm" Global Journal Of Computer Science And Technology ,Vol-10 2010. Pp 38-44.

[4]. [4] Ali Khazaee " Heart Beat Classification Using Particle Swarm Optimization" Intelligent Systems And Applications, 2013. Pp 25-33.

[5]. [5] Ashraf Osman Ibrahim, Siti Mariyam Shamsuddin, Nor Bahiah Ahmad, Sultan Noman Qasem "Three-Term Backpropagation Network Based On Elitist Multiobjective Genetic Algorithm For Medical Diseases Diagnosis Classification" Life Science Journal 2013. Pp 1815-1823.

[6]. [6] Pramod Kumar Yadav K.L.Jaiswal, Shamsher Bahadur Patel, D. P.Shukla "Intelligent Heart Disease Prediction Model Using Classification Algorithms" International Journal Of Computer Science And Mobile Computing Vol-2, 2013. Pp 102-107.

[7]. [7] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra "Performance Analysis Of Classification Data Mining Techniques Over Heart Disease Data Base" International Journal Of Engineering Science & Advanced Technology Vol-2, 2012. Pp 470-478.

[8]. [8] Liang-Wen Hang, Hsuan-Hung Lin, Yi-Wu Chiang Hsiang-Ling Wang, Yung-Fu Chen " Diagnosis Of Severe Obstructive Sleep Apnea With Model Designed Using Genetic Algorithm And Ensemble Support Vector Machine" Applied Mathematics & Information Sciences, 2013. Pp 227-237.

[9]. [9] Jyoti Soni, Uzma Ansari, Dipesh Sharma, Sunita Soni "Intelligent And Effective Heart Disease Prediction System Using Weighted Associative Classifiers" International Journal On Computer Science And Engineering Vol-3, 2011. Pp 285-2397.

[10]. [10] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar "Enhanced Prediction Of Heart Disease With Feature Subset Selection Using Genetic Algorithm" International Journal Of Engineering Science And Technology Vol-2. 2010, Pp 5370-5376.

[11]. [11] J. H. Eom, S. C. Kim and B.T. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction", Expert Systems with Applications 34, 2008, Pp 2465-2479.

[12]. [12] R. Das, I. Turkoglu, A. Sengur "Effective diagnosis of heart disease through neural networks ensembles"Expert Systems with Applications, 2009, Volume 36, Issue 4, Pp 7675-7680.

[13]. [13] A. Ebrahimzadeh, A. Khazaee"Detection of premature ventricular contractions using MLP neural networks: A comparative study Measurement" 2010. Pp 103–112.

[14]. [14] S. Osowski, T. Markiewicz, L. T. Hoai "Recognition and classification system of arrhythmia using ensemble of neural networks Measurement" vol. 41, 2008, Pp 610–617