

A Framework for Accurate Disease Diagnosis using Cover Data Mining Rule on Homogenous Data

¹K. Gayathri and ²M. Chitra

¹Department of Computer Science, Research and Development Centre, Valluvar College of Science and Management, Bharathiar University, Karur, India

²Department of IT, Sona College of Technology, Salem, India

Key words: Sequential class covering, disease diagnosis, first order rule, homogenous data classifier, boosting, dimensionality, machine learning

Abstract: Knowledgeable data is the fundamental step for discovering different types of patterns from large database. The pattern to be discovered from vast amount of data employs classification technique. Classification (i.e., classifier) builds a model with the relationship between the attribute set, class set and input data. However, most of the classification techniques do not fit with a good starting point on classifying multiple data sources class patterns. Even if it works on multiple data sources class patterns, it produces both the best and worst cases of result set. On occurrence of worst case result, patterns are not nested properly resulting in the tradeoff while fetching high class accuracy result. These drawbacks in the current work are overcome in our research work by working with sample of large quantities of information about patients and their medical conditions. In this research, an efficient framework for accurate disease diagnosis, Sequential Class Covering Rule based Homogeneous Data Classifier (SCCR-HDC) is proposed. Initially, SCCR-HDC framework uses the classifier tree to analyze medical information about patients from different dimensional level. For analyzing this classifier tree, a modern boosting based machine learning concept is introduced. The analyzed results of the tree are used for rule formation in the second step for efficient diagnosis of the disease patterns. The rule formed is applied on the training and test sample homogenous data to easily diagnosis the disease class accuracy. A sequential class covering rule is formed to extract the best result patterns in sequential manner from the current set of training data instances. Similarly, to diagnosis the normal, abnormal, critical disease patterns from the test samples, a searching process called, first order rule based general to precise searching process is performed in SCCR-HDC

Corresponding Author:

K. Gayathri

Department of Computer Science, Research and Development Centre, Valluvar College of Science and Management, Bharathiar University, Karur, India

Page No.: 8-16

Volume: 12, Issue 2, 2018

ISSN: 1815-9354

Research Journal of Agronomy

Copy Right: Medwell Publications

framework. Experiment is conducted on the factors such as class accuracy rate on disease diagnosis,

classifier tree based time rate on predicting disease pattern and precision rate on categorizing disease patterns.

INTRODUCTION

Stroke related heart diseases represent a large group of neurological disorders with homogeneous nature of data and pathological expression affecting the heart for unknown reasons and progress in a relentless manner. Knowledge Transfer with Low Quality data (KT-LQ) (Quanz *et al.*, 2002) addressed the issues related to feature extraction with the aid of sparse coding which resulted in improvement in the level of prediction. However, it did not fit with a good starting point on classifying multiple data sources class patterns. Mining Iterative Generators (MI-Gen) (Lo *et al.*, 2011) though introduced the concept of representative rules with closed iterative patterns but was not effective in scalability. Ahmadzadeh *et al.* (2014) addressed the issues related to scalability with respect to Childhood Arterial Ischemic Stroke (AIS). Effective diagnosis of stroke in children was also addressed by Ciccone *et al.* (2011) using clinical and physical examination. But, neonatal stroke was not included.

Brain sufferings due to Small Vessel Infarct (SVI) or “lacunar” stroke results in 20-25% of all ischemic strokes. A broader view of diagnosis and management was presented by Behrouz *et al.* (2012) for managing lacunar stroke. Context based electronic health record by Hsu *et al.* (2012) included a medical decision making process to address issues related with large amount of information. However, complexity increased with the increased in information. To reduce the time complexity, back propagation based neural network algorithm was applied by Popoola *et al.* (2013) to increase the overall prediction accuracy. A stroke detection system based on ischemic was conducted on tomography images to improve the sensitivity of disease diagnosis with the aid of unsupervised region growing algorithm (Tyan *et al.*, 2014). Another supervised learning method was presented by Wang *et al.* (2014) called probabilistic disease progression model. However, the model limits with amount of training data by using only disease progression models.

In this study, we consider the problem of developing an accurate disease diagnosis framework that classifies multiple data sources class patterns from different dimensional level and to efficiently diagnosis the disease patterns for minimizing classifier tree based time on predicting disease pattern. This is a challenging problem because of the analysis of medical information about patients from different dimensional level experience different stage of disease with different diagnosis patterns ranging from normal to critical stage. Even though some statistics may be estimated or predicted, applying

traditional disease diagnosis techniques such as representative rules (Quanz *et al.*, 2002) and sparse coding (Lo *et al.*, 2011) to obtain optimal solution can be time consuming and hence, it is not applicable for homogenous data in practice.

We propose a practical yet provably-efficient modern boosting algorithm “MBoost” to solve this problem. Our algorithm is a meta-heuristic machine learning system that does not require any prior knowledge of the system statistics or any prediction on future disease patterns. Moreover, it is computationally efficient and easy to implement in large practical systems with any disease patterns. MBoost constructs and solves an accurate prediction rule based on the test samples, ‘n’ sample attributes and learning algorithm with rule to offer class accuracy in terms of normal, abnormal or critical disease patterns on homogeneous data.

To complement the analysis, we conduct a simulation study to evaluate our algorithm. Our results show that: MBoost effectively increases the precision rate on categorizing disease patterns by opportunistically categorizing the disease using classifier tree with an appropriate rule formation, MBoost achieves higher rate of class accuracy using first order rule by applying sequential class covering rule, MBoost achieves much higher fairness in classification accuracy while incurring only a marginal increase in computational complexity and while reducing the classifier tree based time rate on predicting disease pattern.

With respect to the algorithm proposed by Lo *et al.* (2011), we make several improvements in this study. The most significant one is that we incorporate a first order rule based general to precise searching, so that our new algorithm is able to take its classification decisions on diagnosing the class pattern as normal, abnormal and critical conditions. As the results show, this ability leads to an improvement in class accuracy rate on disease diagnosis and reduces the false positive rate on disease diagnosing.

Literature review: Approximately 5% of Emergency Department (ED) patients present with neurological symptoms resulting in increased disease diagnostic error. To minimize the diagnostic error, measures for misdiagnosis was handled by Pope and Edlow (2012). Visualization techniques were employed by Borkin *et al.* (2011) to minimize the diagnostic mistakes. To improve the prediction accuracy of disease diagnosis, data mining mechanism was introduced by Masethe and Masethe (2014) using pruned tree. Artificial neural networks was designed by Ghwanmeh *et al.* (2013) with the objective of

improving the system performance and accuracy by addressing three main heart diseases, mitral stenosis, aortic stenosis and ventricular septal defect.

Current state-of-the-art measure handled for disease diagnosis of Alzheimer's Disease (AD) is proven to be expensive and also involves a time-consuming process. Cost effective tools was designed by Laske *et al.* (2015) to improve the system performance. Another method based on hybrid feature (Gallego-Jutgla *et al.*, 2015) was designed for early diagnosis of Alzheimer's disease based on the features being selected. However, the size of dataset for diagnosis of Alzheimer's disease was limited. A linear support vector machine was introduced by Zhang *et al.* (2015) to improve the diagnostic potential. K Nearest Neighbor (KNN) was applied by Shouman *et al.* (2012) for accurate diagnosis of heart disease. However, the dataset used was limited in size.

The most common measure for investigating acute stroke is Echocardiography (Pepi *et al.*, 2010). The role of Echocardiography was discussed using neurological and cardiac evaluation. Another form of acute stroke called as noncontrast CT head was designed by Mainali *et al.* (2014) with the objective of improving the detection of disease using stroke windows. A prediction model by Zecca *et al.* (2014) served as a measure to identify the early diagnosis of ischemic stroke using plasma terminal.

Compared with the existing research, MBoost does not require any of the system statistics or any prediction on future disease patterns. As a result, MBoost is an

efficient solution that improves the class accuracy in terms of normal, abnormal or critical disease patterns that minimizes false positive rate on disease diagnosing on homogeneous data.

Sequential class covering rule based homogeneous data classifier framework: To make easy perception of the ongoing deliberation, we first explain how the framework is modeled with the definitions of key concepts. Based on this consideration, the problems to be solved are determined with SCCR-HDC framework and give our solution framework. The construction of an efficient and accurate classifier for diagnosing accurate class pattern of disease is the main objective in our research paper. The construction of the classifier uses homogenous multiple data sources for predicting the diseases. In this research, a framework called Sequential Class Covering Rule based Homogenous Data Classifier (SCCR-HDC) is designed that diagnose the disease accurately with normal, abnormal and critical class patterns. The rule based disease diagnosis with the test and training homogenous data is shown in Fig. 1.

As shown in Fig. 1, rule based disease diagnosis is performed with the aid of test and training samples of disease which helps in the effective prediction of disease. As shown above, the prediction of disease is performed using a combined classifier method that uses learning algorithm with rule and matches with the testing samples and results in the disease sample patterns ranging from normal, abnormal to critical disease patterns.

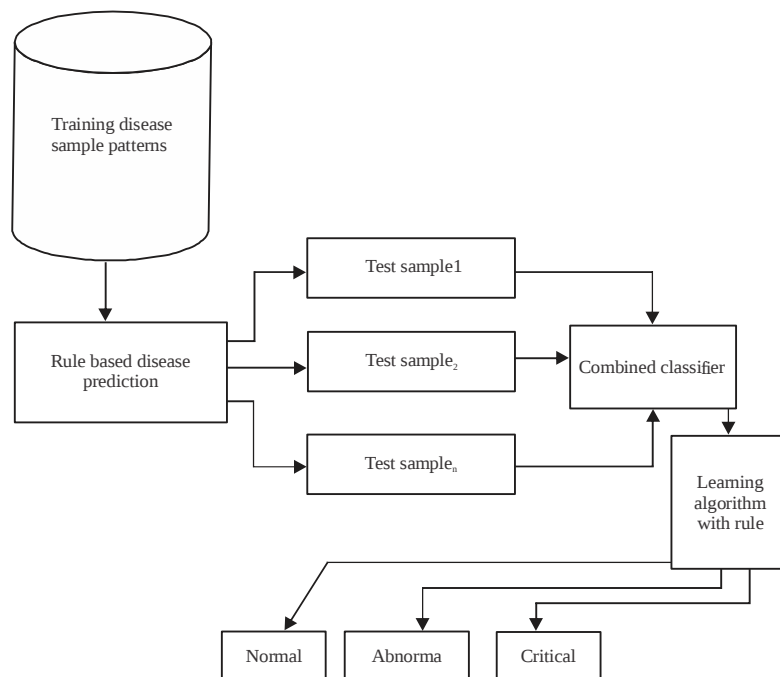


Fig. 1: Rule based disease diagnosis with test and training sample

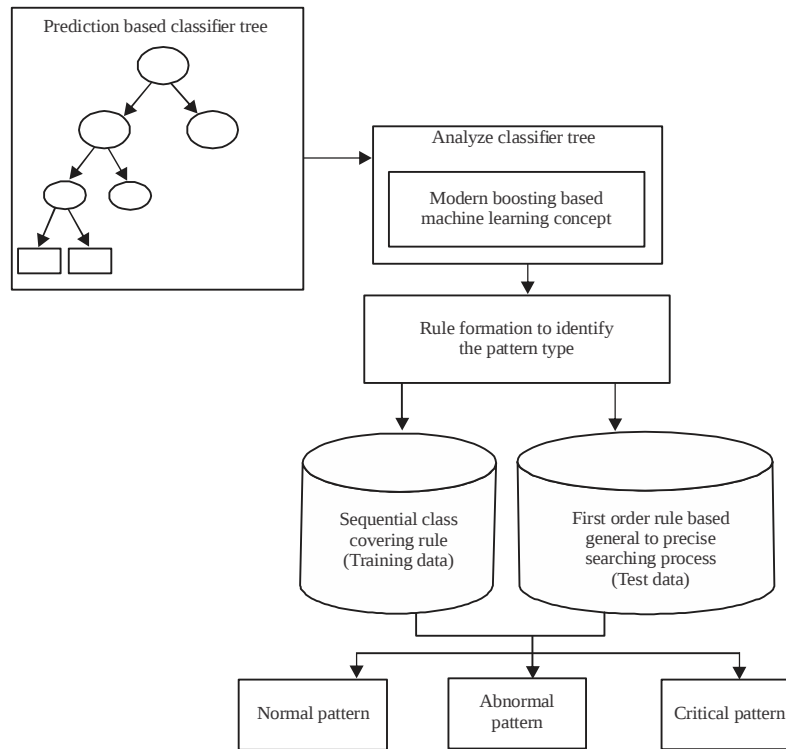


Fig. 2: Overall structural diagram of SCCR-HDC framework

The Class Tree analysis is a type of the machine learning framework used on classifying the predictive disease patterns. The homogenous data samples in SCCR-HDC framework cooperate with each test samples request from the users and perform the processing on the single system. Homogenous system is introduced in SCCR-HDC framework as an efficient means for designing and managing the system. The test samples are compared with the predicted result and machine learning concepts to easily diagnose the disease. The overall structural diagram of SCCR-HDC framework is shown in Fig. 2.

As shown in Fig. 2, prediction based classifier tree is taken on to analyze the disease report. To start with, the Modern Boosting based Machine learning concept is used to efficiently analyze the classifier tree level. The Modern Boosting concept is a meta-heuristic machine learning system which is used to improve the performance level of analyzing the disease patterns.

Next, the rule formation identifies the pattern type on the training and tests samples of homogenous data. The training data uses the sequential class covering rule to diagnose the class of diseases. The test data uses the first order rule based general to precise searching process to easily predict the normal, abnormal and critical patterns specifically. First Order rule based general to precise searching helps on predicting the correct disease from the

general to more specific format. This helps the end users to accurately diagnose the disease patterns using SCCR-HDC framework.

Analyzing classifier tree pattern: The first step involved in the design of SCCR-HDC framework forms the analyzing of classifier tree pattern. The attributes are first classified in SCCR-HDC framework and performs the analyzing process with the modern boosting concept. The attribute is first fed into the root of the classifier tree and attribute of the test samples are checked against the existing training sample tree, for efficient diagnosis of disease. The process is continued repeatedly in SCCR-HDC framework, until every attributes are checked and reaches the sub branches of the classifier tree. These attribute based analyzing of patterns helps in accurate diagnose of disease class patterns.

Modern boosting: Modern Boosting (MBoost) is a meta-heuristic machine learning system in our proposed work where the devised classifier tree helps to attain higher and accurate prediction rule from rough set rule for diagnosing the disease of the test samples (Algorithm 1).

Algorithm 1; MBoost:

Mboost (PR_c)
 Input: Classifier tree, attributes
 Samples, prediction rule
 begin

```

1: for each  $T_i$ 
2:   Train rough rule to high prediction rule set
3:   for each sample  $S_k$  and  $PR_a$ 
4:     if  $n_T > 0$  then
5:       Compute Modern Boosting Distribution ( $D_{(1, 2, \dots, n)}$ )
6:     end if
7:     if  $T_i \neq PR_a$  then
8:        $e^{-n_T} (< 1)$ 
10:    Final Prediction Rule ( $PR_a$ ) =  $P\text{sign}(\sum_T \alpha_T PR_a(T_i))$ 
11:    end if
12:    if  $T_i = PR_a$  then
13:       $e^{n_T} (> 1)$ 
14:    Final Prediction Rule ( $PR_a$ ) =  $P\text{sign}(\sum_T \alpha_T PR_a(T_i))$ 
15:    end if
16:  end for
17: end for
End
Return final Prediction Rule ( $PR_a$ )

```

Modern boosting procedure: The high prediction rule achievement using the modern boosting meta-heuristic learning framework in SCCR-HDC takes the classifier tree ' T_i ' with ' n ' sample attributes ' Att_n ' where α_T include the wholesome attributes in ' $Training_{DB}$ '. The attributes included are patient's information for classifying the tree pattern accurately with high prediction rule ' PR_a '. Boosting distribution with irregular disease set are turned into high precise prediction set by computing ' $D_{(1, 2, \dots, n)}$ ', if the ' n ' sample attributes ' Att_n ' in the classifier tree ' T_i ' does not match with homogenous test data samples, then the result obtained is < 1 . If the test attributes matches with the training samples, then the result sample result obtained is > 1 . By effective analyzing using SCCR-HDC framework, ' T_i ' is the classifier tree which attain high accurate prediction rule result.

The weak classifier tree (i.e., negative sign values) is discarded by the same modern boosting procedure in SCCR-HDC framework. Finally, high accurate prediction tree is grouped together with the summation of positive sign values by providing accurate disease prediction rate. To accurately diagnose the disease level of the patients and to classify the class pattern as normal, abnormal or critical condition, a new type of rule is introduced in our proposed work. The rule based classification is briefly explained in the study.

Rule formation to identify the pattern type: On accurate classification of the rule set, the class patterns are diagnosed from the set of the test and training samples of homogenous data. Data class samples are accurately predicted with the covering rule using if-then procedure. The training database ' $Training_{DB}$ ' contains the homogenous ' x ' data samples whereas the training homogenous data samples holds ' n ' attributes of processed information. Each sample is uniquely stored with separate row ID 'RowID' in SCCR-HDC framework. With the help of the attributed of homogeneous data samples, the result of processed information denotes the level of disease class pattern such as normal, abnormal and critical condition. Figure 3 represents the step by step process.

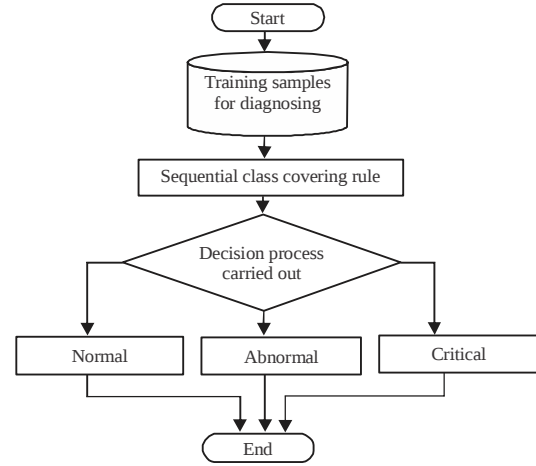


Fig. 3: Flow diagram of disease diagnosing and class pattern

Figure 3 illustrates the flow diagram and step by step procedure of disease diagnosing and obtaining the class patterns. Initially, the training samples are taken and the sequential class covering rule is employed on the input set of the classifier tree. This tree information helps in the proposed framework to sort out the disease type based on the rule. The rule is introduced with the if-then cover procedure in order to increase the results of decision process.

Sequential class covering rule: The sequential class covering rule of SCCR-HDC framework performs the sequential process of interaction with the training samples to accurately store the information in the database. The prediction rule ' PR_a ' of the initial set performs the machine learn one rule procedure. SCCR-HDC framework checks the class sequentially with the set of attributes to diagnose the diseases. If the class pattern is not satisfied with the stored set of rule in the database, then the rule is appended in the row as the new list. These rules are used on effectively identifying the disease class patterns. The Sequential Class Covering (SCC) rule is formularized as:

$$\text{SCC Rule} = \text{Machine Learn One Rule} \quad (1)$$

$$(T, n \rightarrow \text{Add}(\text{RowID}))$$

The sequential cover procedure, with one rule at a time is checked between the high accurate predicted classifier tree based attributes. The new predicted disease attributes are appended on the training sample database for the accurate diagnosing of diseases. Sequential class cover rule in our proposed framework covers a set of rules with the positive examples. On reducing the learning problem, learning set of rules are placed in the training set database for the accurate diagnosing of diseases:

$$\text{Add Rule } (PR_a) = PR_a + \{PR_a + 1\} \quad (2)$$

Add rule into 'Training_{DB}' improves the class pattern diagnosis on homogenous data samples. Data sample objects with undefined rule are added with the new rowed as 'PR_a+1' information. The rule performance before and after adding of the new conjunct (i.e., information) from the database is always maintained constant in the SCCR-HDC framework. On the training database, the homogenous classified class patterns are sorted into normal, abnormal and critical conditions for more accurate result diagnosing with minimal processing time.

First order rule based general to precise searching: On accurate diagnosis of the disease, the SCCR-HDC framework uses the machine learns one rule database for testing process. The test homogenous data samples use the first order rule as it has more representational power on diagnosing the diseases applied on any varying length of attribute information about the patients. The first order rule in the proposed framework accommodates the attributes with the precondition rule. The precondition first order rule and machine learn one rule are matched to fetch the precise result from the generic 'Training_{DB}'. The first order rule on attaining the precise result uses if then rules with the relational assertions that are conveniently used on diagnosing the class pattern as normal, abnormal and critical conditions:

$$\begin{aligned} \text{Precise search} &= \text{if (First order logic rule)} \\ &\text{hen } (n < R1, R2, \dots, Rn) \end{aligned} \quad (3)$$

This rule helps to reduce the generic step into precise step in SCCR-HDC framework for easy diagnosing of diseases. The most general hypotheses disease information attribute structure matches with the every machine learns one rule 'PR_a' to classify accurate class patterns. The relational assertions are also carried out in (Eq. 3) for easy diagnosis of diseases. In our proposed framework, the user input takes the 'Query_i' query information and produce the best diagnosing result on the searching process.

MATERIALS AND METHODS

Experimental evaluation: Sequential Class Covering Rule based Homogenous Data Classifier (SCCR-HDC) framework is implemented in the JAVA platform. JAVA platform for analyzing this proposed framework uses the Echocardiogram data set from UCI repository. The Echocardiogram data set contains the information related to the heart patient which represents the multivariate dataset used for efficient classification of patient's class

patterns. It has 132 instances with the 12 attribute values. The dataset taken for the experiment contain all the information about the patients who suffered from the stroke disease on the heart.

The attributes explained in this system are survival, still alive, age at heart attack, pericardial-effusion, fractional shortening, epss, lvdd, wall motion score, wall motion index, mult, name, group, alive at 1. SCCR-HDC framework is compared against the existing work Knowledge Transfer with Low Quality data (KT-LQ) (Quanz *et al.*, 2002) and Mining Iterative Generators (MI-Gen) (Lo *et al.*, 2011). Experiment is conducted on the factors such as classifier tree based time rate on predicting disease pattern, false positive rate on disease diagnosing and precision rate on categorizing disease patterns.

The classifier tree based time rate on predicting the disease patterns using the framework SCCR-HDC refers to the time taken to build the classifier tree and comparison of test sample to the training samples on homogeneous data. It is measured in terms of milliseconds (msec):

$$\text{Classifier tree}_{\text{time}} = \text{Time (Test}_{\text{DB}} + \text{Training}_{\text{DB}}) \quad (4)$$

The classifier tree based time 'Classifier Tree_{time}' given above measures the time taken to compare test database 'Test_{DB}' with the 'Training_{DB}'. Lower the classification time, more efficient the method is said to be. The precision rate on categorizing the disease patterns using the framework SCCR-HDC is the fraction of retrieved disease patterns that are relevant to the find. Higher the precision rate, more efficient the method is said to be. It is measured in terms of percentage (%). The mathematical evaluation for precision rate on categorizing the disease patterns is given below:

$$P = \frac{(\text{Relevant}_{\text{DP}} * \text{Retrieved}_{\text{DP}})}{\text{Relevant}_{\text{DP}}} \quad (5)$$

The precision rate 'P' given above is the ratio of difference between the relevant and retrieved disease patterns to the relevant disease patterns in homogeneous data. The false positive rate using the framework SCCR-HDC is the proportion of absent disease diagnosis that yield positive test outcomes, i.e., the conditional probability of a positive test result given an absent event. In order to diagnosis the disease in an efficient manner, the false positive rate should be comparatively less. It is measured in terms of percentage (%).

$$\text{FPR} = \frac{(\text{Retrieved}_{\text{DP}} * \text{Relevant}_{\text{DP}})}{\text{Retrieved}_{\text{DP}}} \quad (6)$$

RESULTS AND DISCUSSION

Discussions on Sequential Class Covering Rule based Homogeneous Data Classifier (SCCR-HDC): The Sequential Class Covering Rule based Homogenous Data Classifier (SCCR-HDC) framework is compared against the existing Knowledge Transfer with Low Quality data (KT-LQ) (Quanz *et al.*, 2002) and Mining Iterative Generators (MI-Gen) (Lo *et al.*, 2011). The experimental results using JAVA are compared and analyzed through Table 1 and Fig. 4.

The numerical values of the classifier tree based time achieved by the methods under study are tabulated in Table 1. Table 1 shows that the framework SCCR-HDC achieved optimal classification time using Echocardiogram data set. The 35 test samples are considered for performing the classification operations and compared with the existing methods. We also choose three variants of still-alive attributes namely a binary variable, dead at end of survival period and still alive. Here we take the user preference still-alive attributes and implement the classification of disease patterns by comparing SCCR-HDC with two other state-of-the-art methods namely KT-LQ and MI-Gen, respectively. Finally, the performances of all three different methods are illustrated in Fig. 4.

The convergence plot for 35 testing samples is depicted in Fig. 4. From Fig. 4, we can note that the proposed SCCR-HDC framework achieved the optimal classification time compared to other methods. The numerical results of the classification time achieved by the methods under study are tabulated in Table 1. We also figure out that in Fig. 4, the proposed Sequential Class Covering Rule based Homogenous Data Classifier (SCCR-HDC) framework algorithm shows a rise in the initiation of the convergence graphs due. However, when 20 test samples were considered, a down trend in classifier tree based time is observed because of the changes in the attributes for different test samples being considered.

The classifier tree based time rate on predicting the disease patterns is reduced by applying classifier tree pattern. With this classifier tree pattern, attribute based analyzing of patterns is performed using the framework SCCR-HDC which helps in minimizing the time rate on predicting disease patterns by 24-47% compared to KT-LQ (Quanz *et al.*, 2002). Besides using the framework SCCR-HDC, the weak classifier tree is removed by modern boosting procedure in SCCR-HDC framework that helps in reducing the classifier tree based time rate on predicting the disease patterns as only the strong classifier tree is used for predicting the disease patterns. This reduces the time taken to classify using SCCR-HDC by 50-77% compared to MI-Gen (Lo *et al.*, 2011).

Table 1: Tabulation for classifier tree based time

No. of samples (S)	Classifier tree based time on predicting the disease patterns (msec)		
	SCCR-HDC	KT-LQ	MI-Gen
5	23.35	34.38	41.35
10	35.85	46.88	58.88
15	39.44	49.47	61.47
20	33.13	44.16	56.16
25	41.32	52.35	64.35
30	38.35	49.38	61.38
35	45.85	56.88	68.88

Table 2: Tabulation for precision rate

No. of samples (S)	Precision rate on categorizing disease patterns (%)		
	SCCR-HDC	KT-LQ	MI-Gen
5	0.135	0.123	0.112
10	0.158	0.146	0.135
15	0.173	0.161	0.150
20	0.165	0.153	0.142
25	0.179	0.167	0.156
30	0.174	0.162	0.151
35	0.182	0.170	0.159

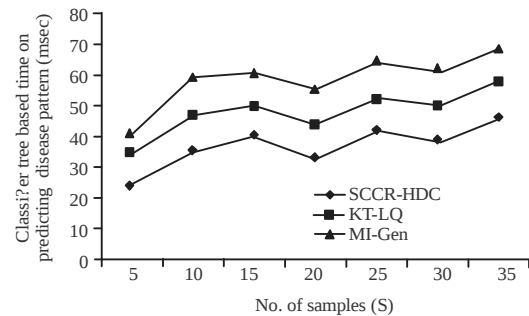


Fig. 4: Performance analysis of classifier tree based time

The proposed Sequential Class Covering Rule based Homogenous Data Classifier (SCCR-HDC) framework is compared with the two existing methods in terms of precision rate. The measures the precision rate obtained from the SCCR-HDC in Table 2. The training samples consists of 132 whereas for experimental purpose, testing samples of size 35 is selected and applied to MBoost algorithm conducted in JAVA using the Echocardiogram data set. The convergence plot using MBoost algorithm with differing attributes of survival of the patients as each patient or samples suffered from the disease at different time periods is depicted in Fig. 5. We can notice that the proposed SCCR-HDC framework had better Precision rate compared to KT-LQ (Quanz *et al.*, 2002) and MI-Gen (Lo *et al.*, 2011), respectively.

The convergence plot for measuring precision using Echocardiogram data set is shown in Fig. 5. From Fig. 5, we can notice that the proposed SCCR-HDC framework converge higher precision rate than KT-LQ (Quanz *et al.*, 2002) and MI-Gen (Lo *et al.*, 2011), respectively by yielding higher precision rate which increases the

Table 3: Tabulation for false positive rate

No. of samples (S)	False positive rate on disease diagnosis (%)		
	SCCR-HDC	KT-LQ	MI-Gen
5	28.33	39.34	45.36
10	31.35	42.36	48.38
15	25.45	36.46	42.48
20	33.39	44.39	50.41
25	29.78	40.79	46.81
30	33.45	44.46	50.48
35	32.18	43.19	49.21

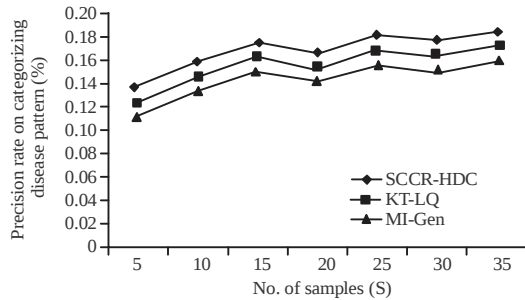


Fig. 5: Performance analysis of precision rate

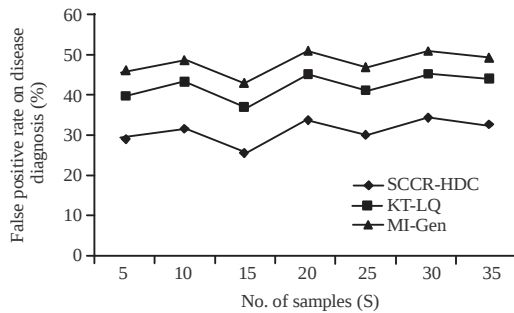


Fig. 6: Performance analysis of false positive rate

performance measure. Moreover, SCCR-HDC framework had obtained the best precision rate of 0.182% followed by KT-LW, MI-Gen with precision rate of 0.170 and 0.159%, respectively.

The prediction rate on categorizing the disease patterns is improved by applying Modern Boosting (MBoost) which is a type of meta-heuristic machine learning system that helps in achieving higher and accurate prediction rate on categorizing the disease patterns using SCCR-HDC by 6-8% compared to KT-LW (Quanz *et al.*, 2002). In addition, with the application of final prediction rule used in the framework SCCR-HDC evaluates the boosting distribution in an efficient manner for homogeneous data resulting in the increased precision rate on categorizing disease patterns by 12-17% when compared to MI-Gen (Lo *et al.*, 2011), respectively.

The measure of false positive rate efficiency is clearly tabulated in Table 3. Table 3 shows the comparison of false positive rate efficiency obtained with JAVA for different patient information varying the name,

survival, age-at-heart-attack and so on. The table reveals that SCCR-HDC is a better alternative compared to two other methods KT-LQ and MI-Gen, respectively.

Convergence characteristics of measure of false positive rate efficiency for thirty five sample images with varying name, survival, age-at-heart-attack of different age groups are considered and compared with two other methods and are shown in Fig. 6. It is observed that the false positive rate on disease diagnosis employing MBoost algorithm for SCCR-HDC makes the false positive rate of heart reach its lowest value of 32.17% when compared to the other two methods which registered 43.18 and 49.21%, respectively.

The false positive rate on disease diagnosis is reduced in the framework SCCR-HDC using the sequential class covering rule. By applying sequential class covering rule, the accurate classification of disease patterns is performed using the first order rule set using the if-then procedure. As a result, the false positive rate on disease diagnosis is reduced drastically when compared to the existing Knowledge Transfer with Low Quality data (KT-LQ) (Quanz *et al.*, 2002) that only used low quality data for feature extraction resulting in the increased false positive rate on disease diagnosing by 32-43%. Besides with the aid of first order rule based general to precise searching, the framework SCCR-HDC has more representational power than the existing Mining Iterative Generators (MI-Gen) (Lo *et al.*, 2011) that was carried out only on the part of software design whereas in the proposed framework, an in-depth insights were provided for efficient classification of disease patterns resulting in the minimization of false positive rate by 52-66%.

CONCLUSION

The conventional disease diagnosis model used to classify various medical images usually designed for providing high quality diagnosis of disease which may not give satisfactory result for multiple data sources class patterns. To improve the accuracy of disease diagnosis and reduce the false positive rate on classification of diseases, Sequential Class Covering Rule based Homogeneous Data Classifier (SCCR-HDC) framework is designed based on MBoost algorithm is designed. With the application of modern boosting based machine learning the classification time for classifying the disease patterns according to normal, abnormal or critical using the Echocardiogram data set from UCI repository is well performed in an organized manner. Analyzing the classifier tree pattern using Modern Boosting (MBoost) is applied to achieve these results. The two step model, sequential process of interaction using sequential class covering rule and efficient matching using first order rule introduced in the framework SCCR-HDC, resulted in

significant improvement over the state-of-the-art methods in terms of classification time, precision rate and false positive rate on disease diagnosis. Experimental results show that the proposed framework SCCR-HDC is better than the two other state-of-the-art methods.

REFERENCES

- Ahmadzadeh, K.L., V. Bhardwaj, S.A. Johnson and K.E. Kane, 2014. Pediatric stroke presenting as a seizure. *Case Rep. Emerg. Med.*, Vol. 2014, 10.1155/2014/838537.
- Behrouz, R., A.R. Malek and M.T. Torbey, 2012. Small vessel cerebrovascular disease: The past, present and future. *Stroke Res. Treat.*, Vol. 2012, 10.1155/2012/839151.
- Borkin, M., K. Gajos, A. Peters, D. Mitsouras and S. Melchionna *et al.*, 2011. Evaluation of artery visualizations for heart disease diagnosis. *IEEE Trans. Visualiz. Comput. Graph.*, 17: 2479-2488.
- Ciccone, S., M. Cappella and C. Borgna-Pignatti, 2011. Ischemic stroke in infants and children: Practical management in emergency. *Stroke Res. Treat.*, Vol. 2011, 10.4061/2011/736965.
- Gallego-Jutgla, E., J. Sole-Casals, F.B. Vialatte, M. Elgendi, A. Cichocki and J. Dauwels, 2015. A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. *J. Neural Eng.*, Vol. 12, No. 1. 10.1088/1741-2560/12/1/016018.
- Ghwanmeh, S., A. Mohammad and A. Al-Ibrahim, 2013. Innovative artificial neural networks-based decision support system for heart diseases diagnosis. *J. Intell. Learn. Syst. Applic.*, 5: 176-183.
- Hsu, W., R.K. Taira, S. El-Saden, H. Kangarloo and A.A.T. Bui, 2012. Context-based electronic health record: Toward patient specific healthcare. *IEEE Trans. Inform. Technol. Biomed.*, 16: 228-234.
- Laske, C., H.R. Sohrabi, S.M. Frost, K. Lopez-de-IPina and P. Garrard *et al.*, 2015. Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's Dementia*, 11: 561-578.
- Lo, D., J. Li, I. Wong and S.C. Khoo, 2011. Mining iterative generators and representative rules for software specification discovery. *IEEE Trans. Knowledge Data Eng.*, 23: 282-296.
- Mainali, S., M. Wahba and L. Elijevich, 2014. Detection of early ischemic changes in noncontrast CT head improved with ISRN *Neurosci.*, Vol. 2014, 10.1155/2014/654980.
- Masethe, H.D. and M.A. Masethe, 2014. Prediction of heart disease using classification algorithms. *Proceedings of the World Congress on Engineering and Computer Science*, Volume 2, October 22-24, 2014, San Francisco, USA., pp: 1-4.
- Pepi, M., A. Evangelista, P. Nihoyannopoulos, F.A. Flachskampf and G. Athanassopoulos *et al.*, 2010. Recommendations for echocardiography use in the diagnosis and management of cardiac sources of embolism *European Association of Echocardiography (EAE)* (a registered branch of the ESC). *Eur. J. Echocardiogr.*, 11: 461-476.
- Pope, J.V. and J.A. Edlow, 2012. Avoiding misdiagnosis in patients with neurological emergencies. *Emerg. Med. Int.*, Vol. 2012, 10.1155/2012/949275
- Popoola, L.T., G. Babagana and A.A. Susu, 2013. Thrombo-embolic stroke prediction and diagnosis using artificial neural network and genetic algorithm. *Int. J. Res. Rev. Applied Sci.*, 14: 655-661.
- Quanz, B., J. Huan and M. Mishra, 2012. Knowledge transfer with low-quality data: A feature extraction issue. *IEEE Trans. Knowledge Data Eng.*, 24: 1789-1802.
- Shouman, M., T. Turner and R. Stocker, 2012. Applying k-nearest neighbour in diagnosing heart disease patients. *Int. J. Inform. Educ. Technol.*, 2: 220-223.
- Tyan, Y.S., M.C. Wu, C.L. Chin, Y.L. Kuo, M.S. Lee and H.Y. Chang, 2014. Ischemic stroke detection system with a computer-aided diagnostic ability using an unsupervised feature perception enhancement method. *Int. J. Biomed. Imaging.*, Vol. 2014, 10.1155/2014/947539.
- Wang, X., D. Sontag and F. Wang, 2014. Unsupervised learning of disease progression models. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 24-27, 2014, New York, USA., pp: 85-94.
- Zecca, B., C. Mandelli, A. Maino, C. Casiraghi and G. Bolla *et al.*, 2014. A bioclinical pattern for the early diagnosis of cardioembolic stroke. *Emerg. Med. Int.*, Vol. 2014, 10.1155/2014/242171.
- Zhang, W., X. Yang, S. Lui, Y. Meng and L. Yao *et al.*, 2015. Diagnostic prediction for social anxiety disorder via multivariate pattern analysis of the regional homogeneity. *BioMed Res. Int.*, Vol. 2015.