International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 8 (2015) pp. 19077-19095

http://www.ripublication.com

An Efficient Predictive and Diagnosis Model Using Bayes Shared Information Criterion Based on Associative Classifier

K.Gayathri[1] and Dr. M.Chitra[2]

[1]Ph.D Research Scholar,
Department of Computer Science,
Research and Development centre,
Bharathiar University and
Assistant Professor, Valluvar College of Science and Management, Karur. India.
[2]Professor, Department of IT,
Sona College of Technology, Salem, India.
[1]kstgayathri@gmail.com.[2]chitra_slm@yahoo.com

ABSTRACT

Feature Selection-based models have proven to be an effective solution for reducing the space dimension as well as other problems for effective diagnosis of stroke disease. This paper focuses on developing a Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC) resulting in huge classification rules that makes prediction in efficient manner by pruning classification rules. Most design and implementations of Feature Selection-based models do not efficiently utilize multiple related disease class patterns. Instead of independently computing the importance of features for each disease class patterns, our framework, (BSIC-AC) applies Shared Information for multiple related disease patterns, thus, increasing the feature selection performance. Besides, considering the fact that dealing with large scale dynamic dataset is time consuming with the increase in disease class patterns, we propose shared information between the feature and disease class patterns using Associative Classifier. The Associative Classifier for large number of classification rules makes effective diagnosis of stroke along with its causes as a measure for feature selection to reduce the space dimension from medical datasets. Instead of generating huge number of feature subsets in BSIC-AC framework, while reducing space dimension, missing of important features are avoided by applying greedy approach thus reducing the false positive rate. Extensive experiments on medical dataset named Echocardiogram Data Set from UCI repository demonstrate that our

framework outperforms other state-of-the-art feature selection-based models for early detection of stroke and alarming with the possibility of disease in terms of feature selection performance, minimizing space dimension at reduced interval of time and classification accuracy.

Keywords: Feature Selection, Shared Information Criterion, Associative Classifier, Space Dimension, Greedy Approach

Introduction

Detecting and controlling the most infectious diseases at an early stage has been the major concerns health care industry. Several researchers have focused on these aspects and provided remedy accordingly. Disease Detection using Markov Switching Models (DD-MSM) [1] focused on identifying the infectious disease using Markov models using statistical surveillance method. However, the feature Selection using Markov models do not obtain multiple related disease class patterns. Group Incremental Approach using Rough Set Technique (GIA-RST) [2] applied rough set technique for obtain the features with the aid of decision table at a much shorter time. However, rules extracted at dynamic period did not update with respect to time.

Another feature selection method introduced in [3] used shared knowledge discovery across multiple tasks and improved the average precision value. However, novel classes with respect to multiple tasks remained unaddressed. This issue was solved in [4] with the application of novel class detection framework called, concept-drifting that improved the classification results with the aid of decision tree model.

Real world scenarios including, pattern matching, machine learning approach not only minimizes the dimensionality problem of large disease patterns being generated, but also helps in effective classification of disease. Fuzzy Rough Set (FRS) [5] was designed with the motive of generating rule-based classifier to reduce the redundant attributes based on rule induction. However, randomness with respect to data perturbation remained unsolved.

In [6], filter-based data partitioning method was designed to arrive at the statistical characteristics of training partitions based on Clustering, Declustering and Selection (CDS). However, data subsets with overlaps were not considered. Clinical studies were conducted in [7] for measuring the overlaps for identifying cerebrovascular disease. In [8], descriptive statistics was developed with the objective of stroke patients using Barthel index.

Early detection of etiologic diagnosis helps in stroke acute management and treatment. In [9], statistical analysis were developed using Receiver Operating Characteristic (ROC) curves to measure the demographic and vascular risk factors related to stroke improving the rate of sensitivity. However, prediction was not made in an efficient manner. Classification and Regression Tree (CART), Iterative Dichotomized 3 (ID3), and Decision Tree (DT) was applied in [10] for measuring the early prediction of heart disease. But, significance of the features was not considered.

In [11], Link based Associative Classifier (LAC) was used for obtaining higher accuracy and discovering associations between the compounds and diseases.

Though associations were established in an efficient manner, feature selection was not made in a discriminatory way. Multi Filtration Feature Selection (MFFS) [12] with the aid of variance coverage to improve the classification accuracy.

However, rapid growth of health industry and need for personalized care for the diseased patients at an early stage require us to discover a new framework to address the above said issues. Therefore, the contributions of our work include the following:

- To reduce the space dimension from medical data sets using Bayes Shared Information Criterion by obtaining multiple related disease patterns.

- To improve feature selection performance by reducing the irrelevant and redundant disease patterns by applying features selection matrix.

- To enhance the classification accuracy by applying optimized associative classifier model by applying classification rules where class labels are fetched (i.e.,) filtered out with respect to set of attributes based on pruning object.

- To reduce the false positive rate based on total compatible length using greedy algorithm that effectively move down the value of attributes and select the attributes that is more compatible with the current selection of disease patterns.

The remaining parts of this paper are structured as follows. In Section 2, the related works on predicting and classifying disease patterns are reviewed. We introduce the framework of the Bayes Shared Information Criterion based on Associative Classifier (BSIC-AC) in Section 3. The experiments are conducted in Section 4 and corresponding results are analyzed in Section 5. Finally, we summarize our concluding work in Section 6.

1. Related works

Early diagnosis of Alzheimer disease was focused on [13] to significantly control the computation cost with the aid of non-invasive intelligent models. The diagnosis was performed using Fractal Dimension (FD) and linear parameters with the motive of enhancing the performance. However, linear parameters with multiple rating scales remained unaddressed. In [14], a hybrid intelligent model was designed combining rough set, genetic approach and Bayesian network with the objective of improving the classification precision.

Another method that used hybrid feature selection algorithm was introduced in [15] to obtain better classification accuracy using Support Vector Machine (SVM) and Sequential Backward Floating Search (SBFS). However, the time factor was not taken into consideration. Adaptive Fuzzy Neuro Inference System (AFNIS) was constructed in [16] with the objective of accuracy rate of prediction of heart disease at

an early stage.

The application of Data mining in the health care industry is mainly used for efficient prediction and diagnosis of heart disease with minimized number of attributes. In [17], Naïve Bayes, classification by clustering and decision tree was applied to minimize the number of attributes and also identifying the relationships between predictions. Though inconsistency and missing values were resolved, intensity of the disease was highly unpredictable.

Gene path analysis [18] was constructed for generating accurate diagnostic model and prevention of autism spectrum disorder (ASP). Another method was introduced in [19], using classification algorithms for efficient prediction of patterns at an early stage with the aid of decision tree model. Multilayer perceptron and fuzzy technique was applied in [20], with the objective of improving the features being generated for obtaining efficient classified results for acute leukemia cells.

In this work, we construct a Bayes Shared Information Criterion based on Associative Classifier with the main aim of reducing the space dimension and improve the classification accuracy for predicting the stroke disease at an early stage.

2. Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC)

This section describes the proposed framework, BSIC-AC which stands for Bayes Shared Information Criterion based on Associative Classifier. The idea is to create a Shared Information Criterion using Bayes model that improves the features being selected and minimizes the space dimension improving the classification accuracy.

2.1 Design of Bayes Shared Information Criterion

Classifier performance is highly related to many elements like the type of classifier being selected the input features opted for, the expected output and so on. This work inspects the role played by feature selection and classification on feature selection improvement for effective diagnosis of stroke disease. Proper feature selection significantly minimizes the size of classifier and therefore improves the classifier performance by minimizing the impact of irrelevant disease patterns and redundant generation of patterns. Irrelevant disease patterns cause false relationship between the disease class patterns and the classifier output. Redundant generation of patterns increases the classifier complexity reducing the diagnosis of stroke disease.

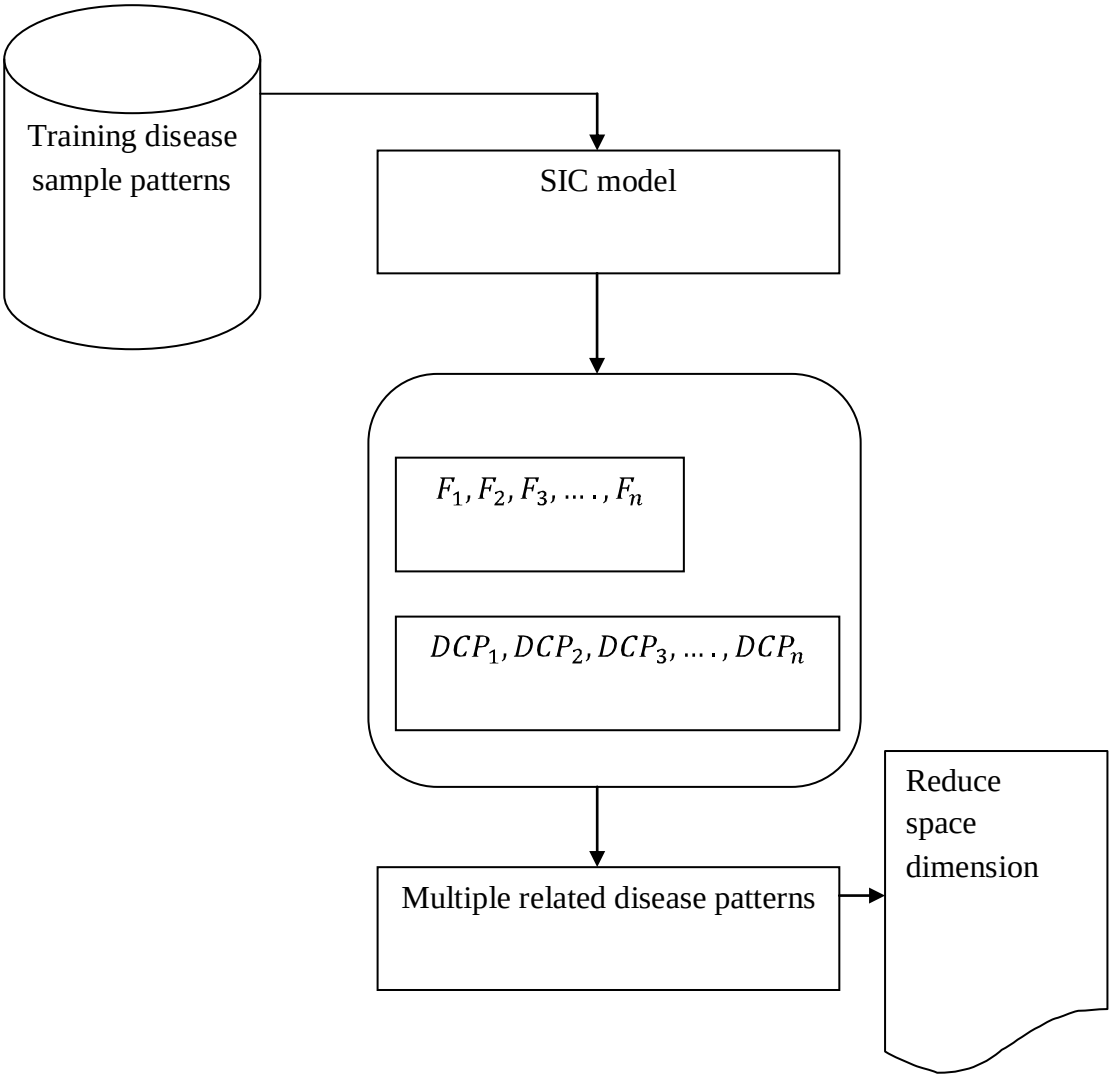

Figure 1 Construction of Bayes Shared Information Criterion

Figure 1 shows the construction of Bayes Shared Information Criterion model. The input to the Bayes SIC model consists of diseased sample patterns extracted through training database. The objective of Bayes Shared Information Criterion (BSIC) model in BSIC-AC framework is to increase the performance of features being selected for diagnosis of stroke disease. The SIC model in BSIC-AC framework extracts multiple feature selection functions of different disease patterns of stroke disease in an integrated fashion. As a result, the framework shared knowledge between multiple tasks results in efficient decision making while diagnosing the stroke disease.

Let us consider two types of patients, considered to be normal (features) and patients suffering from stroke (diseased class patterns). Then the BSIC states that there exists certain relationship between features and diseased classed patterns or

certain information being shared, which is formalized as

$$Prob(F_1, F_2, F_3, \ldots., F_n \,|Features) \quad (1)$$

$$Prob(DCP_1, DCP_2, DCP_3, \ldots., DCP_n \,|Diseased\ Class\ Patterns) \quad (2)$$

From (1) and (2), the probability of the features with respect to diseased class patterns is shown. Suppose the SIC model in SIC-AC framework selects 'n' features for 't' tasks, and then the Shared Information Criterion using Bayesian model for feature selection is formalized as follows.

$$FS = [Cons(FSM_a) + (\propto * \beta\ (FSM_a))] \quad (3)$$

Where $FSM_a$is the Feature Selection Matrix for 'a^{th}' task, the consistency between the features and diseased classed patterns are evaluated using $Cons(FSM_a)$ with '$\propto$' denoting the shared information parameter and $\beta\ (FSM_a)$ is the shared information function. Then, the posterior probability for Shared Information Criterion to identify the relationship between features and diseased classed patterns are given as below.

$$Prob(F|F_1, F_2, F_3, \ldots., F_n) = \frac{Prob(FS(F_1,F_2,F_3,\ldots.,F_n))}{Prob\ (F_1,F_2,F_3,\ldots.,F_n)} \quad (4)$$

$$Prob(DCP|DCP_1, DCP_2, DCP_3, \ldots., DCP_n) = 1 - Prob(F|F_1, F_2, F_3, \ldots., F_n) \quad (5)$$

From (4) and (5), with the Shared Information Criterion based on the Bayesian probability model, multiple features selection for different tasks (i.e., identify disease class patterns) is learnt in an integrated manner, improving the feature selection performance and reducing the space dimension.

2.2 Construction of Optimized Associative Classifier model (improves classification accuracy)

With the increase in stroke disease class patterns, shared information between the feature and disease class patterns are performed with the aid of Associative Classifier. Though, associative classifiers improve and enhance the classification accuracy rate of diseased patterns being identified, however, the integration of several rules are difficult to correlate. So Optimized Associative Classifier (OAC) model for effective associative classification in BSIC-AC framework are considered. The OAC model prune the classification rules into a subset with effective feature selection made through Bayes SIC model.

The Optimized Associative Classifier (OAC) model in the proposed framework, BSIC-AC works with the motive of choosing maximal-relevancy minimal-redundancy (i.e., attaining optimality) for stroke disease and then compile into an easy and optimized interpretable classifier. Consider the associative classifier with the objective of maximal-relevancy and minimal-redundancy with the view of

pruning classification rule. Let $A = \{A_1, \dots, A_n\}$ denote a set of attributes from Echocardiogram Data Set, and the values of the attributes be denoted as pruning object $VA = \{va_1, .., va_n\}$, i.e. $va_i \in A_i$.

Let $CL = \{c_1, .., c_n\}$denotes class labels for stroke disease, then Optimized Associative Classifier (OAC) model for effective associative classification in SIC-AC framework represents the pruning of classification rules 'CL' from values of the attributes to class labels as given below.

$$CL: \{A_1, \dots, A_n\} \rightarrow CL \quad (6)$$

From (6), given a set of values of attributes$= \{va_1, .., va_n\}$, the Optimized Associative Classifier (OAC) model returns class label $CL \in C$.

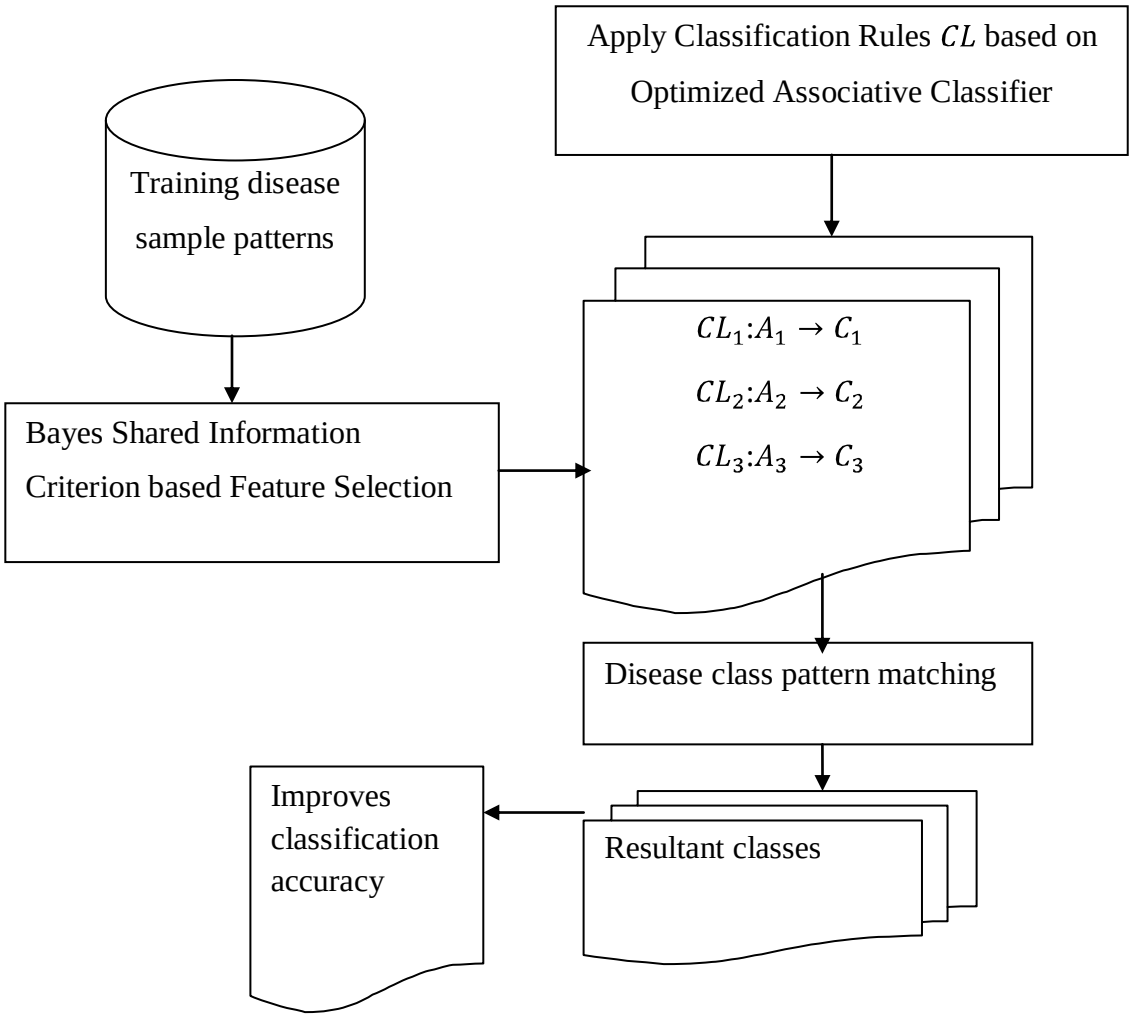

Figure 2 Construction of Optimized Associative Classifier model

Figure 2 shows the construction of Optimized Associative Classifier model. As shown in the figure, to start with, efficient feature selection is made using Bayes Shared Information Criterion where the input is obtained from training disease sample patterns. Followed by this, to prune Classification Rules, Optimized Associative Classifier model is applied to the features being selected. The pruned classified rule using Optimized Associative Classifier model is provided in Table 1.

The table includes the rule ID with the rule generated according to disease pattern. Here the optimized refers to the maximal-relevancy minimal-redundancy for obtaining stroke disease patterns. With this, the classification rules are applied and according to the rules, diseased class pattern matching is performed and resultant classes are obtained.

Table 1 Pruned classified rule using Optimized Associative Classifier model

Classification Rule ID	Diseased Classed Patterns (Relevant)	Pruned classified rule according to disease pattern
1	DCP_1	$1,2: 1 \rightarrow a$
2	DCP_2	$1,2: 2 \rightarrow b$
3	DCP_3	$1,2,3: 1 \rightarrow c$
4	DCP_4	$1,2,3: 2 \rightarrow d$
5	DCP_5	$1,2,3,4: 1 \rightarrow e$
6	DCP_6	$1,2,3,4: 2 \rightarrow f$

As shown in figure 2 and table 1, the application of Optimized Associative Classifier model in SIC-AC framework helps in improving the classification accuracy.

2.3 Greedy Approach (reduces the false positive rate)

Finally, while generating numerous amounts of feature subsets in BSIC-AC framework, missing of important features are minimized through greedy approach thus minimizing the false positive rate. Instead of producing huge feature subsets as in tradition associative classifier model, the proposed framework, BSIC-AC adopts a greedy approach than traditional rule-based associative classifiers to maximize the relevancy and minimize the redundancy of stroke disease patterns. To avoid over fitting, BSIC-AC uses the pruned classified rule from Table 1 for measuring and evaluate each pruned classified rule and uses the best 'n' rules. The framework, BSIC-AC uses the greedy algorithm based on Total Compatible Length (TCL).

The Greedy algorithm using binary tree based on TCL (GTCL) that first indexes the attributes according to the completion time, starting from the earliest to the latest. Then the attributes are greedily selected by moving down the values of the attributes and selecting the attributes that is compatible with the current selection.

In the setting of multiple related disease class patterns, the framework select 'n' features for 't' tasks in a simultaneous manner to minimize redundancy and therefore reducing the false positive rate. At the same time, in order to add '$n+1$' features for '$t+1$' tasks, then the framework BSIC-AC maximize the relevancy of TCL by adding and updating that feature to a subset (FSM_a) of the '$t+1$' tasks.

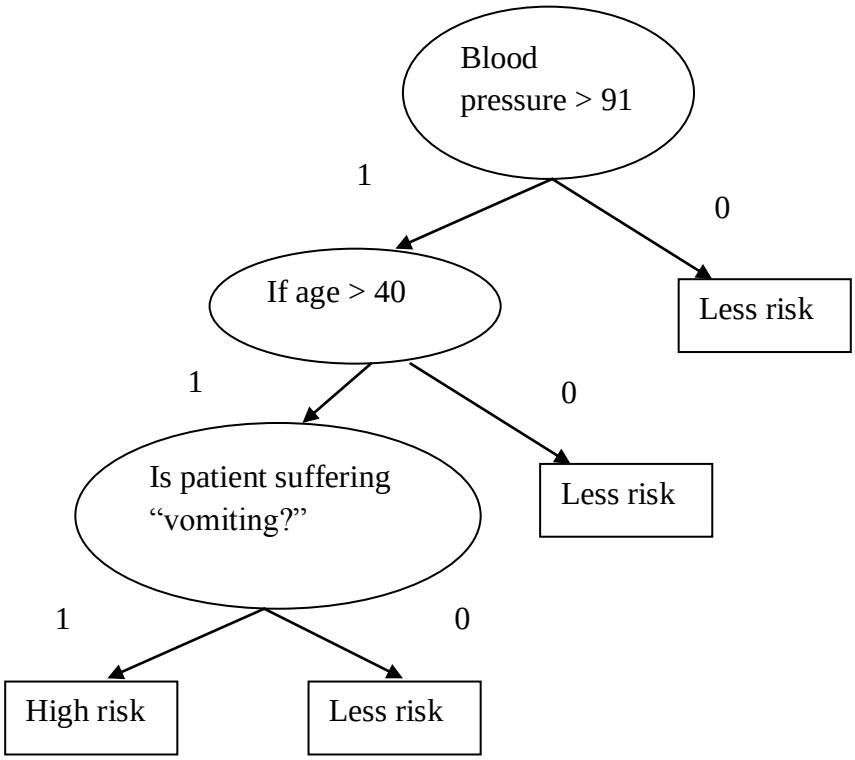

Figure 3 Flow of greedy approach using binary tree

The figure 3 shows the flow of greedy approach using binary tree for multiple related disease class patterns. The greedy approach using binary tree for obtaining an optimal prefix-free binary code (i.e., either 0 or 1), uses the following concept. For each,$1 \leq i \leq n$, a subset feature va_i corresponding to F_i is created. The above process is repeated until minimum subset feature is generated.

Input: Features $F_i = \{F_1, F_2, F_3, \ldots, F_n\}$,

Diseased Class Patterns $DCP_i = (DCP_1, DCP_2, DCP_3, \ldots, DCP_n)$,

Class Labels $CL = \{c_1, .., c_n\}$,

Attributes $A = \{A_1, \ldots, A_n\}$, n features, 't' tasks

For each F_i

    For each DCP_i

        Apply Bayes Shared Information Criterion using (3)

        Generate multiple related disease patterns using (4) and (5)

    End for

End for

For each A_i

    Generate classification rules using (6)

End for

For each n features

    For each t tasks

        Construct binary tree

    End for

End for

Algorithm 1 – Greedy Approach using Binary Tree (GABT)

The algorithm Greedy Approach using Binary Tree (GABT) is discussed above. To start With, for each feature and diseased class patterns, perform Bayes Shared Information Criterion and generate multiple related disease patterns. Upon successful generation of multiple related disease patterns, classification rules are constructed using Optimized Associative Classifier (OAC) model. Based on the pruning classification rules, for each attributes and tasks, binary tree is constructed, for early diagnosis of stroke disease and possibility of disease are identified.

3. Experimental Evaluation

Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC) is implemented in the JAVA platform. The framework uses JAVA platform with Echocardiogram Data Set from UCI repository for measuring and analyzing the proposed framework. BSIC uses Echocardiogram Data Set that consists of information related to the heart patient for efficient classification of patients' disease 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. For experimental purpose, we took 35 samples.

The BSIC-AC framework is compared against the existing Disease Detection using Markov Switching Models (DD-MSM) [1] and Group Incremental Approach using Rough Set Technique (GIA-RST) [2]. Experiment is conducted on the factors such as, feature selection performance, space dimension, classification accuracy and false positive rate on categorizing disease class patterns.

4. Discussion

The Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC) framework is compared with the existing Disease Detection using Markov Switching Models (DD-MSM) [1] and Group Incremental Approach using Rough Set Technique (GIA-RST) [2] using Echocardiogram Data Set for measuring the feature selection performance, space dimension, classification accuracy and false positive rate for categorizing disease class patterns with the objective of early detection of stroke disease.

5.1 Impact of feature selection performance

To support transient performance, in Table 1 we apply an efficient Greedy approach using Binary Tree to obtain the feature selection performance and comparison is made with two other existing methods, GIA-RST and DD-MSM. The feature selection performance using BSIC-AC framework is the rate at which the features are efficiently selected for correct classification of disease patterns. It is the ratio of percent difference between the retrieved and relevant disease patterns for efficient detection of stroke at an early stage. It is measured in terms of percentage (%).

$$FS = [(Relevant\ disease\ patterns) - (Retrieved\ disease\ patterns) * 10] \quad (7)$$

Feature selection (Using BSIC-AC) = [(5-1)*10] = 40

Feature selection (Using DD-MSM) = [(5-2)*10] = 30

Feature selection (Using GIA-RST) = [(5-3)*10] = 20

Table 2 Tabulation for feature selection

Relevant disease patterns (n)	Feature selection (No. of features)		
	BSIC-AC	DD-MSM	GIA-RST
5	40	30	20
10	45	35	25
15	53	40	28
20	60	48	35
25	65	52	38
30	70	55	42
35	75	60	45

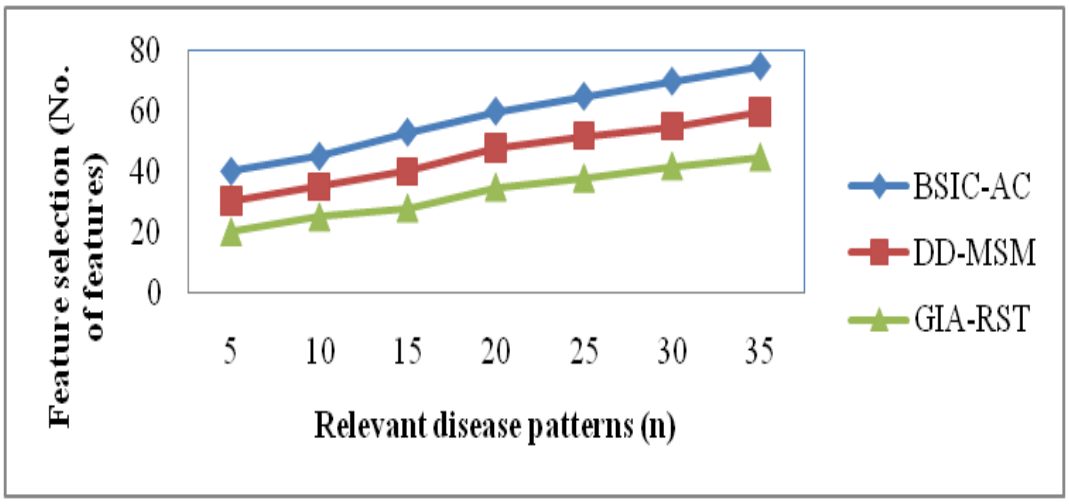

Figure 4 Relevant disease patterns with respect to Feature Selection

Figure 4 show that the proposed Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC) framework provides higher feature selection when compared to DD-MSM [1] and GIA-RST [2]. This is because of the application of Shared Information for multiple related disease patterns that efficiently identifies the most relevant features and makes effective diagnosis of stroke at different time period. The Shared Information Criterion using Bayesian model for feature selection using the probability of the features with respect to diseased class patterns which maximizes the feature selection performance by 20 – 25 % than DD-MSM [1]. Moreover, Shared Information Criterion, relationship between features and diseased classed patterns or certain information being shared which results in the increase in the feature selection performance by 40 – 50 % when compared to GIA-RST [2].

5.2 Impact of classification accuracy

Classification accuracy using BSIC-AC framework is the measure of predictive model that reflects the total number of times that the framework is correctly classified for early detection of stroke. It is measured in terms of percentage (%).

$$CA = \left({}^{No.\ of\ correctly\ classified\ samples}/_{Total\ number\ of\ samples} \right) * 100 \quad (8)$$

CA (Using BSIC-AC) = (4/5) * 100 = 80

CA (Using DD-MSM) = (3/5) * 100 = 60

CA (Using GIA-RST) = (2/5) * 100 = 40

Table 3 Tabulation for classification accuracy

No. of samples	Classification accuracy (%)		
	BSIC-AC	DD-MSM	GIA-RST
5	80.2	60.35	40.15
10	72.5	57.2	33.2
15	70.35	55.28	31.18
20	68.25	52.14	28.04
25	72.25	64.13	30.9
30	74.55	65.25	41.15
35	80.35	70.30	45.20

The comparison of classification accuracy efficiency is presented in table 2 with respect to the varying number of samples in the range of 5 – 35. With increase in the number of samples, the classification accuracy is also increased.

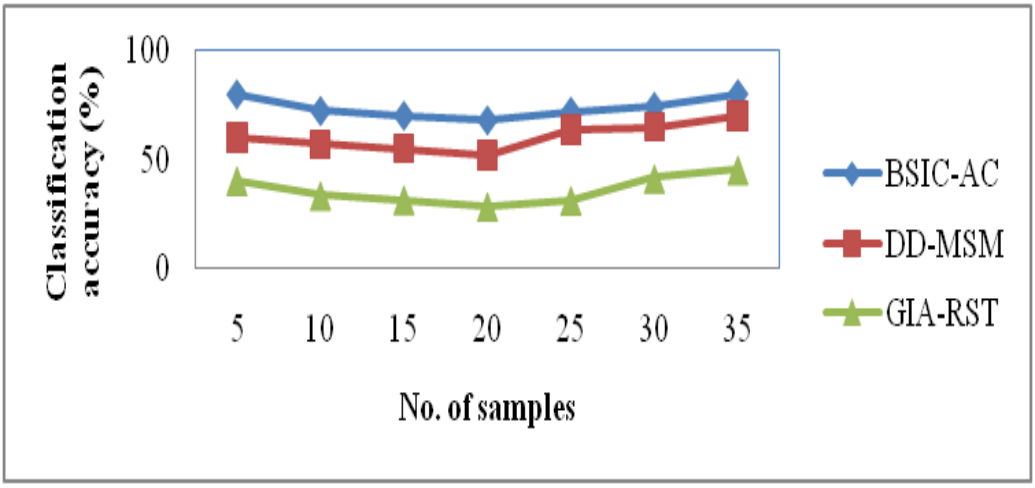

Figure 5 No. of samples with respect to Classification Accuracy

To measure the performance of the classification accuracy, comparison is made with two other existing methods, Disease Detection using Markov Switching Models (DD-MSM) [1] and Group Incremental Approach using Rough Set Technique (GIA-RST) [2] using Echocardiogram Data Set. In Figure 6, the number of samples is varied between 5 and 35. From the figure it is illustrative that the classification accuracy is improved using the proposed Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC) framework when compared to the two other existing methods. This is because with the aid of Optimized Associative Classifier (OAC) model, classification accuracy is attained by applying the classification rules by for stroke disease that efficiently classifies the stroke disease at early stage. As a result, the classification accuracy is improved in BSIC-AC framework by 11 – 34 % and 43 – 58 % than compared to DD-MSM and GIA-RST respectively. In addition, the pruning of classification rules performed with the table that include pruned classified rule according to disease pattern therefore improve the efficiency of classification accuracy.

5.3 Impact of False Positive Rate

False positive rate in BSIC-AC framework is the ratio of presence of wrongly identified disease class patterns that yield positive test outcomes. In order to effectively diagnosis stroke disease, the false positive rate should be comparatively less. It is measured in terms of percentage (%).

$$FPR = \frac{No.of\ samples*(Retrieved_{DCP} - Relevant_{DCP})}{Retrieved_{DCP}} \quad (9)$$

FPR (Using BSIC-AC) = 5 * (50 - 35) / 50 = 1.5

FPR (Using DD-MSM) = 5 * (50 - 25) / 50 = 2.5

FPR (Using GIA-RST) = 5 * (50 - 15) / 50 = 3.5

Table 4 Tabulation for false positive rate

No. of samples	False positive rate (%)		
	BSIC-AC	DD-MSM	GIA-RST
5	1.5	2.5	3.5
10	2	3	3.5
15	2.4	3.5	4.25
20	2.2	3	4
25	3	4	4.75
30	3.5	4.25	5.25
35	4	4.5	5.55

The false positive rate for identifying the disease class patterns using BSIC-AC framework is elaborated in table 3. The framework is considered with varying number of samples at different time period for experimental purpose using JAVA platform.

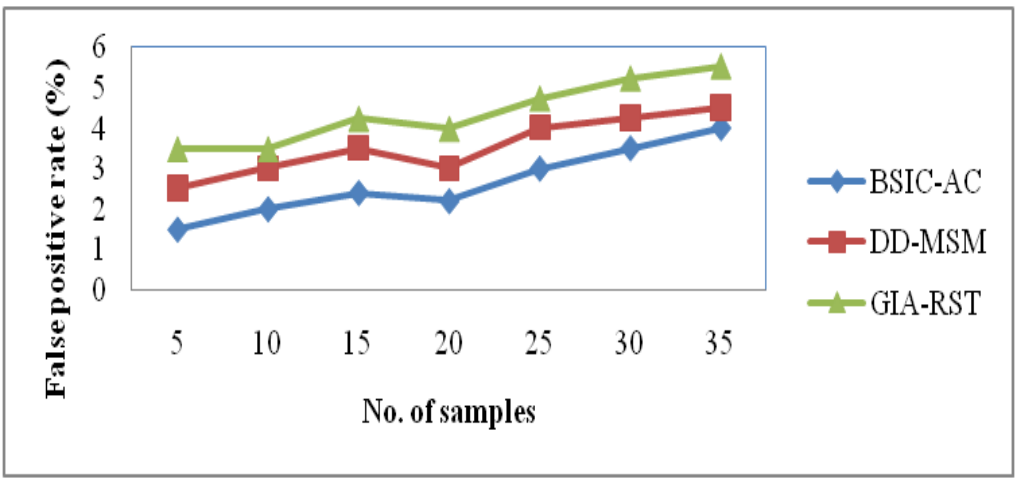

Figure 6 No. of samples with respect to false positive rate

In figure 6, we depict the false positive rate for maximizing the relevancy and minimizing the redundancy of stroke disease patterns using different sample sets with ranges between 5 and 35 for experimental purpose and applied in JAVA. From the figure, the value of false positive rate time achieved using the proposed BSIC-AC framework is lower when compared to two other existing methods namely, DD-MSM [1] and GIA-RST [2]. Besides we can also observe that by increasing the number of samples, though the false positive rate is also increased, but comparatively the false positive rate is lesser using the proposed BSIC-AC when compared to DD-MSM and GIA-RST respectively. This is because with the application of greedy approach than when compared to the tradition rule-based associative classifiers that produces huge feature subsets, the proposed framework, BSIC-AC adopts a greedy approach to maximize the relevancy and minimize the redundancy of stroke disease patterns. As a result, the false positive rate is reduced by 12 – 66 % compared to DD-MSM. Also, by applying binary tree based on TCL using Greedy algorithm, selecting the attributes that is compatible with the current selection which further reduces the false positive rate by 38 – 81 compared to GIA-RST [2].

5.4 Impact of Space Dimension

Finally, table 5 provides the space dimension of BSIC-AC framework using seven different samples that is measured in terms of percentage (%).

Table 5 Tabulation for Space Dimension

Method	Space Dimension (%)
BSIC-AC	65.35
DD-MSM	71.35
GIA-RST	75.83

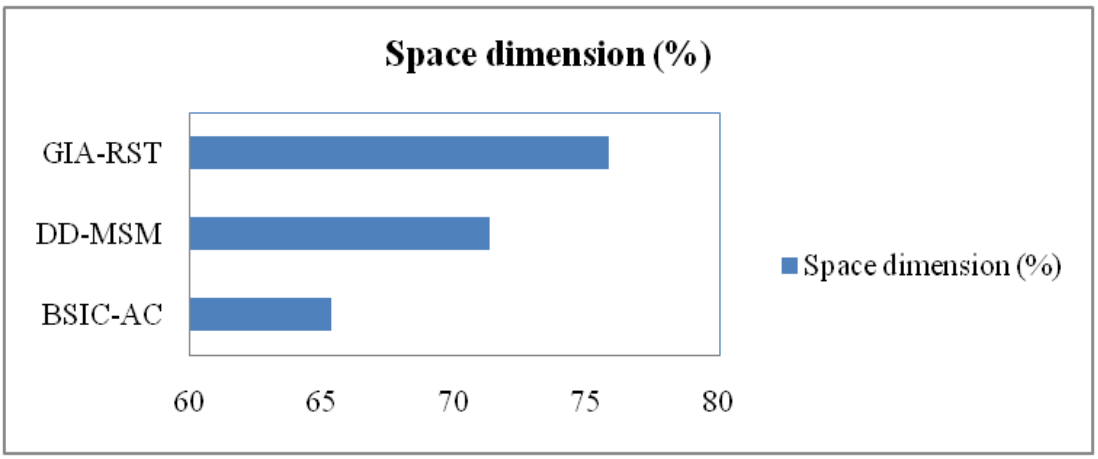

Figure 7 Measure of Space Dimension

Lastly the space dimension used using three different methods is shown in figure 7. From the figure 7 it is illustrative that the proposed BSIC-AC framework potentially yields better results than existing DD-MSM [1] and GIA-RST [2]. The significant results achieved using the BSIC-AC framework is because the application of Bayes Shared Information Criterion offers feature selection effectiveness with most powerful shared criterion for diagnosis of stroke disease and minimizes the space dimension by 9.81 % compared to DD-MSM. With the application of feature selection matrix, the consistency between the features and diseased class patterns are evaluated and the features are further reduced to a coarser construction. As a result, multiple features selection for different tasks is reduced up to 6.27 % when compared with the GIA-RST.

5. Conclusion

Minimizing the space dimensions and improving the feature selection for disease diagnosis has become the key for health care industry, to achieve higher classification accuracy effectiveness and improve the search time taken for identifying the disease patterns at early stage. In this work, we investigate the performance effects of predicting and diagnosing model to minimize the space dimensions by proposing a framework, Bayes Shared Information Criterion (BSIC) based on Associative Classifier (AC). The BSIC-AC framework based on Bayes Shared Information Criterion model and Optimized Associative Classifier model provides an efficient means of identifying the stroke disease class patterns at an early stage with minimal search time. First, we study the use of Bayes Shared Information Criterion using the feature selection matrix that provides the consistency between the features and diseased classed patterns using shared information function. Second, classification accuracy is achieved using Optimized Associative Classifier that improves pruning classification rule with the objective of maximal-relevancy and minimal-redundancy with the aid of class labels. Finally, the greedy algorithm using binary tree based on Total Compatible Length (TCL) for multiple related disease class patterns for obtaining an optimal prefix-free binary code reduces the false positive rate. The method was implemented using JAVA and examined the performance of BSIC-AC which shows that BSIC-AC has satisfactory performance in terms of feature selection performance, classification accuracy, false positive rate and space dimension for early detection of stroke disease compared to the state-of-the-art methods.

REFERENCES

[1] Hsin-Min Lu, Daniel Zeng, and Hsinchun Chen," Prospective Infectious Disease Outbreak Detection Using Markov Switching Models", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[2] Jiye Liang, Feng Wang, Chuangyin Dang and Yuhua Qian," A Group Incremental Approach to Feature Selection Applying Rough Set Technique",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014

[3] Sen Wang, XiaojunChang, XueLi, QuanZ.Sheng, WeitongChen," Multi-task support vector machines for feature selection with shared knowledge discovery", Signal Processing, Elsevier, Dec 2014

[4] Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham," Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[5] Suyun Zhao, Eric C.C. Tsang, Degang Chen, and XiZhao Wang," Building a Rule-Based Classifier—A Fuzzy-Rough Set Approach", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 5, MAY 2010

[6] Rozita A. Dara, Masoud Makrehchi, and Mohamed S. Kamel," Filter-Based Data Partitioning for Training Multiple Classifier Systems",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[7] Francesco Brigo, Frediano Tezzon, Raffaele Nardone," Late-onset seizures and risk of subsequent stroke: A systematic review", Epilepsy & Behavior, Elsevier, Nov 2013

[8] Jarin Chindaprasirt, Kittisak Sawanyawisuth, Paiboon Chattakul, Panita Limpawattana, Somsak Tiamkao, Patcharin Aountri, and Verajit Chotmongkol," Age Predicts Functional Outcome in Acute Stroke Patients with rt-PA Treatment", Hindawi Publishing Corporation ISRN Neurology Volume 2013

[9] Bruno Zecca, Clara Mandelli, AlbertoMaino, Chiara Casiraghi, Giovanbattista Bolla, Dario Consonni, Paola Santalucia, and Giuseppe Torgano," A Bioclinical Pattern for the Early Diagnosis of Cardioembolic Stroke", Hindawi Publishing Corporation Emergency Medicine International Volume 2014

[10] Vikas Chaurasia, Saurabh Pal," Early Prediction of Heart Diseases Using Data Mining Techniques", Caribbean Journal of Science and Technology, Oct 2013

[11] Pulan Yu, David J. Wild," Discovering Associations in Biomedical Datasets by Linkbased Associative Classifier (LAC)", PLOS ONE, December 2012 | Volume 7 | Issue 12

[12] S. Sasikala, S. Appavu alias Balamurugan, S. Geetha," Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical data set", Applied Computing and Informatics, Elsevier, Mar 2014

[13] Karmele López-de-Ipinaa, Jordi Solé-Casalsb, Harkaitz Eguiraunc,J.B. Alonsod, C.M. Traviesod, Aitzol Ezeizaa, Nora Barrosoa,Miriam Ecay-Torrese, Pablo Martinez-Lagee, Blanca Beitia," Feature selection for spontaneous speech analysis to aid inAlzheimer's disease diagnosis: A fractal dimension approach", Computer Speech and Language, Elsevier, Aug 2014

[14] Ziming Yin, Yinhong Zhao, Xudong Lu, and Huilong Duan," A Hybrid Intelligent Diagnosis Approach for Quick Screening of Alzheimer's Disease Based on Multiple Neuropsychological Rating Scales", Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2015

[15] Juanying Xie, Jinhu Lei, Weixin Xie, Yong Shi and Xiaohui Liu," Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases", Health Information Science & Systems, Oct 2013

[16] Negar Ziasabounchi, Iman Asekrzade,"ANFIS Based Classification Model for Heart Disease Prediction", International Journal of Electrical & Computer Sciences, Vol.14, No.2, Apr 2014

[17] Shamsher Bahadur Patel, Pramod Kumar Yadav, Dr. D. P.Shukla," Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques", Journal of Agriculture and Veterinary Science, Volume 4, Issue 2, Aug 2013

[18] E Skafidas, R Testa, D Zantomio, G Chana, IP Everall and C Pantelis," Predicting the diagnosis of autism spectrum disorder using gene pathway analysis", Molecular Psychiatry, Sep 2012

[19] Hlaudi Daniel Masethe, Mosima Anna Masethe," Prediction of Heart Disease using Classification Algorithms", World Congress on Engineering and Computer Science, Oct 2014

[20] Aimi Abdul Nasir, Mohd Yusoff Mashor, and Rosline Hassan," Classification of Acute Leukaemia Cells using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks", The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013