

# CHISC-AC: COMPACT HIGHEST SUBSET CONFIDENCE-BASED ASSOCIATIVE CLASSIFICATION

*S P Syed Ibrahim*<sup>1\*</sup>, *K R Chandran*<sup>2</sup>, and *C J Kabila Kanthasamy*<sup>3</sup>

<sup>1</sup>*School of Computing Science and Engineering, VIT University - Chennai Campus, Tamilnadu, India*  
E-mail: [syedibrahim.sp@vit.ac.in](mailto:syedibrahim.sp@vit.ac.in)

<sup>2</sup>*Department of Information Technology, PSG College of Technology, Coimbatore, Tamilnadu, India*  
E-mail: [chandran\\_k\\_r@yahoo.co.in](mailto:chandran_k_r@yahoo.co.in)

<sup>3</sup>*Department of Information Technology, Avinashilingam University for Women, Coimbatore, Tamilnadu, India*  
E-mail: [cjkapila@gmail.com](mailto:cjkapila@gmail.com)

## ABSTRACT

*The associative classification method integrates association rule mining and classification. Constructing an efficient classifier with a small set of high quality rules is a highly important but indeed a challenging task. The lazy learning associative classification method successfully removes the need for a classifier but suffers from high computation costs. This paper proposes a Compact Highest Subset Confidence-Based Associative Classification scheme that generates compact subsets based on information gain and classifies the new samples without constructing classifiers. Experimental results show that the proposed system out performs both the traditional and the existing lazy learning associative classification methods.*

**Keywords:** Associative classification, Lazy learning, Information gain, Confidence

## 1 INTRODUCTION

Data mining deals principally with extracting knowledge from data. In the present world where data is all around us, there is a pressing need to extract knowledge or interesting information that is hidden in the available data. Association rule mining and classification are data mining functionalities. Association rule mining is concerned with extracting a set of highly correlated features shared among a large number of records in a given database. It uses unsupervised learning where no class attribute is involved in finding the association rule. On the other hand, classification uses supervised learning where class attribute is involved in the construction of a classifier to predict the new instance. Both association and classification are significant and efficient data mining techniques.

Associative classification is a recent and rewarding technique that applies the methodology of association rule mining to classification and achieves high classification accuracy. Associative classifier construction has two steps. In the first, association rule mining is applied to discover class association rules. The important element in controlling the number of rules generated in associative rule mining is the support threshold. If the support value is high, then the number of rules generated is very small, but many high-confidence rules may get eliminated. On the other hand, if the support value is set to a minimum, large numbers of rules are generated. In the second step, the generated rules are ranked. After rule ranking, only the high-ranking rules are chosen to build a classifier while the rest are pruned.

Generating, ranking, and selecting a small subset of high-quality rules with which to construct the classifier is a challenging task indeed. Merschmann and Plastino (2007, 2010) proposed a lazy learning method based on the probabilistic analysis of patterns to classify datasets. This method delays the processing of the data until a new sample needs to be classified. Syed et al. (2011) proposed a lazy learning method based on support and confidence measures. These two methods improve the classification accuracy but lead to high computation cost. Chen et al. (2006) and Zhang et al. (2011) proposed a new approach for associative classification based on information gain where more informative attributes are chosen for rule generation.

The above proposals motivated us to investigate a new computational technique for lazy learning associative classification. In these methods, the rule generation phase in associative classification is a difficult step that requires a large amount of computation. To reduce the computation cost, the proposed method uses an information gain-based lazy learning associative classification method that computes subset confidence only for

the information gain attribute. This method reduces the number of subsets generated. Our proposed method does not build a generalized classifier from training data for classification of new samples but directly predicts classification using compact subset confidence.

The rest of the paper is organized as follows: Section 2 discusses related past work in this field, and Section 3 explains the process of the proposed system. Section 4 shows the working principle of the proposed system using an example. The final section presents the experimental results and observations followed by the conclusion.

## 2 RELATED WORK

Associative classification was first introduced by Liu et al. (1998). It integrates two well known data mining tasks, association rule discovery and classification. The integration is focused on a special subset of association rules whose right hand side is restricted to a class attribute; for example, consider a rule  $R: X \rightarrow Y$ ,  $Y$  must be a class label. Generally there are two stages in associative classification. In the first stage, association rule generation methods such as a priori candidate generation (Agarwal & Srikant, 1994) or Frequent Pattern (FP) growth algorithms (Han et al., 2000) are used to generate class association rules. For example, the Classification based on Association (CBA) method (Liu et al. 1998) employs a priori candidate generation and other associative methods such as Classification based on Predictive Association Rules (CPAR) (Yin & Han, 2003) and Classification based on Multiple Association Rules (CMAR) (Li et al., 2001) while lazy associative classification (Baralis et al., 2002, 2004, 2008) uses the FP growth algorithm for rule generation.

The rule generation step can generate a very large number of rules and is a difficult step requiring a large amount of computation. Experimental results reported by Baralis et al. (2008) have shown that the CBA method, which uses the a priori association rule mining algorithm, generates more than 80,000 rules for some datasets. This leads to memory size issues and other severe problems, such as overfitting. If all the generated rules are used in the classifier, the accuracy of the classifier is high, but the process of classification is slow and time-consuming (Syed, Chandran, & Jabez, 2011).

In the next step, generated rules are ranked. After rule ranking, only the high-ranking rules are chosen to build a classifier with the remaining rules being pruned. The rule ranking can be based on several different types of parameters and interestingness measures, such as confidence, support, lexicographical order of items, etc.

In the CBA method, the rules are arranged based on their confidence value. If two rules have the same confidence measure values, then the rules are sorted based on their support. If both confident and support values are the same for two rules, then rules are sorted based on their rule length. Even after considering confidence, support, and cardinality measures, if some rules exist with the same values for all three measures, the rules are sorted based on their lexicographic order as in the lazy pruning method (Baralis et al., 2008). After rule ranking, the CBA method uses a database coverage method to construct the classifier. CMAR (Han & Pei, 2001) applies the chi-square test (Snedecor & Cochran, 1989), which gives positively correlated rules that can be used in the classifier. CPAR (Yin & Han, 2003) uses the Laplace accuracy measure to estimate the expected error rate for each rule, and the best rules are selected to construct the classifier.

A recent approach for rule pruning is lazy pruning (Baralis et al., 2008), where a rule is pruned only if it misclassifies the data. The entire ruleset is segregated into three sets, namely, useful rules, harmful rules, and spare rules. A rule that classifies at least one data item correctly is said to be a useful rule, and that which misclassifies a data item is a harmful rule. The leftovers are spare rules, which are not pruned but used when needed. Though the lazy pruning strategy works well for small datasets, in the case of large datasets, there are constraints in memory space and ruleset quality. An evolutionary based associative classification method (Syed, Chandran, & Jabez, 2011) was proposed recently. This approach randomly uses a subset of rules to construct the classifier. The richness of the ruleset improves over generations.

Syed, Chandran, and Abinaya (2011) proposed a compact weighted associative classification scheme based on information gain, wherein the class association rule generation algorithm chooses non-class information gain attributes from the dataset, and all the items are generated based only on that attribute. Thus this algorithm reduces the number of itemsets generated. The algorithm calculates the weighted support and weighted confidence for each item and determines whether the item is frequent or not. All of the above recent associative

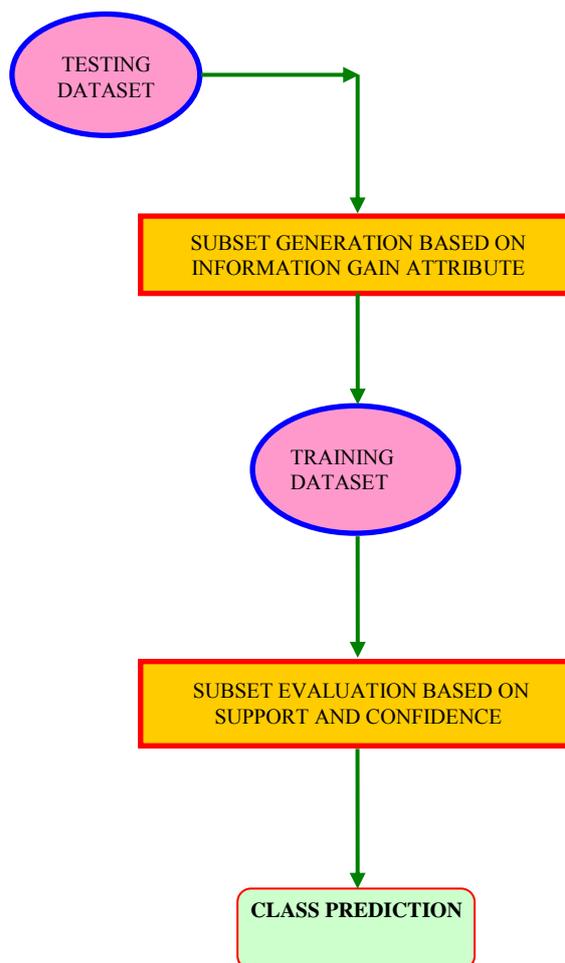
classification methods construct a generalized model (classifier) to classify the new data sample, but our proposed work eliminates the classifier construction and uses information gain to directly choose the best attribute, predicting the class based on the highest subset confidence value. The following section discusses our proposed method.

### 3 PROPOSED METHOD

#### 3.1 Attribute selection based on information gain

Generally, an a priori based rule generation algorithm generates  $2^k$  rules for a dataset with  $K$  items. To reduce the number of rules generated, an information gain attribute can be used. Information gain is a measure used in information theory to quantify the 'information content' of messages (Han & Kamber, 2001). In the ID3 decision tree algorithm (Quinlan, 1986) information gain is used to choose the best split attribute.

In this paper, information gain is used to generate class association rules. In the process of generating these rules, instead of considering all the attributes, the information gain measure selects the best splitting attribute. In this way, our work generates a limited number of good quality rules. Figure 1 shows the system flow of the proposed system, which is called the Compact Highest Subset Confidence-Based Associative Classification (CHISC-AC) method.



**Figure 1.** Proposed system

Suppose an attribute A has n distinct values that partition the dataset D into subsets  $T_1, T_2, \dots, T_n$ . For a dataset,  $\text{freq}(C_k, D) / |D|$  represents the probability that a tuple in D belongs to class  $C_k$ . Then  $\text{info}(D)$  is defined as follows to measure the average amount of information needed to identify the class of a transaction in D:

$$\text{Info}(D) = - \sum_{k=1}^g \frac{\text{freq}(C_k, D)}{|D|} \times \log_2 \left( \frac{\text{freq}(C_k, D)}{|D|} \right), \quad \text{Eq. (1)}$$

where  $|D|$  is the number of transactions in database D and g is the number of classes.

After the dataset D is partitioned into n values of attribute A, the expected information requirement is defined as:

$$\text{Info}_A(D) = \sum_{i=1}^n \frac{|D_i|}{|D|} \times \text{Info}(D_i). \quad \text{Eq. (2)}$$

The information gained by partitioning D according to attribute A is defined as:

$$\text{Gain}(A) = \text{info}(D) - \text{info}_A(D) \quad \text{Eq. (3)}$$

The best split attribute is the one that maximizes the information gain in data set D. These best attributes are then used to generate the subset.

### 3.2 Subset evaluation

After the information gain attribute is identified, the subsets are generated based on this attribute. For each generated subset, confidence values are calculated.

Let  $T_x$  be the set of transactions present in the dataset. Supposing that there are m classes,  $\{C_1, C_2, \dots, C_m\}$ . Class value X is assigned to class  $C_k$ ,  $1 \leq k \leq m$  if and only if

$$P(C_i/X) = \text{Max}(\text{Confidence}(C_i/X), \text{For all } X \subseteq T_x \neq \phi) \quad \text{Eq. (4)}$$

$$\text{Confidence}(X \rightarrow C) = \frac{\text{Occurrence}(X \rightarrow C)}{\text{Occurrence}(X)}, \quad \text{Eq. (5)}$$

where X is the subset and C is the class.

Deciding which class will be assigned to the instance X is based on an analysis of the subsets of attribute values associated with higher confidence.

All confidence values are sorted, and a lower limit is set to limit the number of possibilities.

The lower limit can be calculated as

$$\text{LowerLimit} = \frac{\text{Highest Confidence}}{\text{Sqrt}(M)}, \quad \text{Eq. (6)}$$

where M is the number of classes.

Maximum class confidence is assigned to the testing samples.

## 4 SAMPLE COMPUTATION

Let us consider the sample training and testing dataset of medical observations (artificially constructed) as given in Tables 1 and 2, respectively, where the training dataset contains 14 transactions and 2 class values, and the task is to predict the class label for the test (new) data instance. Among the four attributes, attribute 'CD4 Cell Count' is selected as the information gain attribute as it has the maximum information gain value.

The CHISC-AC algorithm calculates the confidence value for each subset of the testing dataset, and class labels are assigned based on high subset confidence.

**Table 1.** Sample dataset

CD4 Cell Count	Sweating at Night	Tuberculosis (TB)	Temperature	AIDS
>500	High	No	Normal	No
>500	High	No	High	No
<200	High	No	Normal	Yes
200 ... 500	Medium	No	Normal	Yes
200 ... 500	Nil	Yes	Normal	Yes
200 ... 500	Nil	Yes	High	No
<200	Nil	Yes	High	Yes
200 ... 500	Medium	No	High	No
>500	Medium	No	Normal	No
>500	Nil	Yes	Normal	Yes
200 ... 500	Medium	Yes	Normal	Yes
>500	Medium	Yes	High	Yes
<200	Medium	No	High	Yes
<200	High	Yes	Normal	Yes

**Table 2.** Test dataset

200 ... 500	Medium	No	Normal	?
-------------	--------	----	--------	---

Consider that the information attribute is CD4 Cell Count.

For the test dataset, the following subset can be created. The generated rules using information gain-based associative classification are:

{CD4 Cell Count =200 ... 500}  
 {CD4 Cell Count =200 ... 500, Sweating at Night = Medium}  
 {CD4 Cell Count =200 ... 500, Sweating at Night = Medium, Tuberculosis (TB) = No}  
 {CD4 Cell Count =200 ... 500, Sweating at Night = Medium, Tuberculosis (TB) = No, Temperature = Normal}

The other itemsets that are commonly generated by the apriori based association rule mining procedure are eliminated. The excluded rules are:

{Sweating at Night == High}, {Tuberculosis == No}, {Temperature = Normal}  
 {Sweating at Night == High, Tuberculosis == No}  
 {Sweating at Night == High, Temperature == Normal}  
 {Tuberculosis == No, Temperature == Normal}  
 {CD4 Cell Count =200 ... 500, Tuberculosis == No}  
 {CD4 Cell Count =200 ... 500, Temperature == Normal}  
 {CD4 Cell Count =200 ... 500, Sweating at Night == High, Temperature == Normal}  
 {CD4 Cell Count =200 ... 500, Tuberculosis == No, Temperature == Normal}  
 {Sweating at Night == High, Tuberculosis == No, Temperature == Normal}

The proposed system generates only 4 rules whereas the existing CBA algorithm generates 13 rules. This clearly shows that the proposed algorithm will generate a minimal number of rules even for huge datasets with many attributes.

**Table 3.** Sample computation

Rules	Class (AIDS)	Confidence with Class
{CD4 Cell Count =200 ... 500}	Yes	3
	No	2
{CD4 Cell Count =200 ... 500, Sweating at Night = Medium}	Yes	2
	No	1
{CD4 Cell Count =200... 500, Sweating at Night = Medium, Tuberculosis (TB) = No}	Yes	1
	No	1
{CD4 Cell Count =200 ... 500, Sweating at Night = Medium, Tuberculosis (TB) = No, Temperature = Normal}	Yes	1
	No	0

$$\begin{aligned} LowerLimit &= \frac{Highest\ Confidence}{Sqrt(No\ of\ Class)} \\ &= \text{floor}(3/1.414) = 2 \end{aligned} \quad \text{Eq. (7)}$$

We can see that two *yes* classes and one *no* class transaction contain more than the lower limit. Therefore, the *yes* class is assigned to the test data.

## 5 EXPERIMENTAL RESULTS AND DISCUSSION

The proposed CHISC-AC system was tested using benchmark datasets from the University of California at Irvine Repository (UCI Repository) (Blake & Merz, 1998). A brief description of the datasets is presented in Table 4. The experiments were carried out on a PC with Intel Core 2 Duo CPU with a clock rate of 2.00 Ghz and 4 GB of main memory. The 10 cross fold validation approach (Han & Kamber, 2001) was used where 90% of the data was randomly chosen from the dataset and used as a training dataset, and the remaining 10% was used as the testing dataset. The training dataset was used to construct a model for classification. After constructing the classifier, the test dataset was used to estimate classifier performance.

### 5.1 Accuracy Computation

Accuracy measures the ability of the classifier to correctly classify unlabeled data. It is the ratio of the number of correctly classified data over the total number of given transactions in the test dataset.

$$Accuracy = \frac{Number\ of\ Objects\ Correctly\ Classified}{Total\ Number\ of\ Objects\ in\ the\ Testset} \quad \text{Eq. (8)}$$

**Table 4.** Dataset description

Dataset	Transactions	Number of Attributes	Classes
Balance Scale	625	5	3
Breast Cancer	286	10	2
Breast -w	699	10	2
Car	1728	7	4
Credit - a	690	16	2
Diabetes	768	9	2
Ecoli	336	8	8
Flare	1389	13	6
Glass	214	10	6
Ionosphere	351	35	2
Iris	150	5	3
Mushroom	8124	23	2

**Table 5.** Accuracy comparison

DATA-SET	Info Gain Attribute	Traditional Associative Classification	Existing System	Proposed Lazy Associative Classification
		CBA	HiSP – GC	CHiSC
Balance Scale	2	73.64	75.80	71.27
Breast	6	74.13	70.85	85.22
Breast –w	2	93.70	89.71	90.63
Car	6	67.33	69.53	70.12
Credit – a	9	76.48	73.34	83.33
Diabetes	2	69.10	67.10	76.75
Ecoli	6	73.53	70.60	80.00
Flare	1	85.19	87.05	84.71
Glass	8	57.94	71.42	60.91
Ionosphere	5	82.29	86.90	96.29
Iris	4	93.33	94.33	87.22
Mushroom	5	91.89	87.65	98.52
<b>Average</b>		<b>78.21</b>	<b>78.69</b>	<b>82.08</b>

**Table 6.** Average time taken to predict a single instance

Dataset	HiSP-GC	CHiSC
Balance Scale	0.0040	0.0014
Breast	0.0576	0.0026
Breast –w	0.0177	0.0027
Car	0.0406	0.0024
Credit – a	0.0346	0.0046
Diabetes	0.0049	0.0041
Ecoli	0.0174	0.0012
Flare	0.0253	0.0143
Glass	0.0293	0.0026
Ionosphere	0.0123	0.0124
Iris	0.0014	0.0092
Mushroom	0.5910	0.0592

Table 4 shows the description of various datasets. Table 5 shows the accuracy comparison. The first column gives the dataset name; the next column describes the highest information gain attribute that will be used in the proposed system. The third column shows the accuracy for various datasets using the traditional associative classification method (Liu et al., 1998). The fourth column shows the accuracy for HiSP-GC, a lazy learning associative classification method proposed by Merschmann and Plastino (2010). The last column shows accuracy values for various datasets using the proposed CHISC-AC system. Classification of the test dataset

was done ten times, and the overall accuracy was obtained by calculating the average of the accuracy values obtained from the ten different runs.

Table 5 shows that the accuracy comparison with minimum support of 1% and minimum confidence is greater than 50%. The proposed system uses information gain attribute to generate rules. The proposed algorithm has about a +3.87 percent improvement over the traditional associative classification and about a +3.39 percent improvement over the existing lazy learning (HiSP-GC) method. There are 3 instances where the new method gives poorer results. This is due to class distribution. In these databases one class attribute occurred in very low distribution. For example, the Balance Scale dataset contains 3 class labels namely L, B, and R, distributed as 46.08%, 07.84%, and 46.08%. In these cases the proposed system may not learn about the class B. Therefore, further research is needed to improve the algorithm.

Table 6 shows the average time taken to predict a single instance. For example, the existing system takes 0.591 seconds to predict a single instance for the mushroom dataset because the mushroom dataset consists of 8124 records, 23 attributes, and 2 class labels. Here the number of records and attributes is high. Therefore, the existing system takes a large amount of time to predict this single instance. On the other hand, the proposed system using the information gain attribute takes only 0.0592 seconds to predict a single instance from the mushroom dataset.

### 5.3 Sensitivity and Specificity

An imbalance in class occurs when one of the classes in a dataset is represented by a smaller number of samples than the other classes (Barandela et al., 2003; Weiss, 2004). The accuracy measure is used extensively to compare the performance of classifiers but may not be suitable for evaluating imbalanced datasets (Tan, 2006). The class imbalance issue might skew the prediction accuracy of classification models (Guha & Schurer, 2008; Hsieh et al., 2008), resulting in a weakened performance of machine learning algorithms (Kang & Cho, 2006). Therefore, other metrics such as sensitivity, specificity, precision, and recall are used.

To use these metrics, the following terminology is generally used.

1. The rare class is denoted as a positive class and the majority class is denoted as a negative class.
2. True Positive (TP) corresponds to the number of positive examples correctly predicted by the classifier.
3. False Negative (FN) corresponds to the number of positive examples wrongly predicted as negative class by the classifier.
4. False Positive (FP) corresponds to the number of negative examples wrongly predicted as positive class by the classifier.
5. True Negative (TN) corresponds to the number of negative examples correctly predicted by the classifier.

Sensitivity is defined as the fraction of positive examples correctly predicted as positive by the classifier.

$$\text{Sensitivity (True Positive Rate)} = \text{TP} / (\text{TP} + \text{FN}) \quad \text{Eq. (9)}$$

Specificity is defined as the fraction of negative examples correctly predicted as negative by the classifier.

$$\text{Specificity (True Negative Rate)} = \text{TN} / (\text{TN} + \text{FP}) \quad \text{Eq. (10)}$$

**Table 7.** Sensitivity (True Positive Rate – TPR)

Dataset	CBA	HiSP- GC	CHSC
Breast Cancer	0.33	0.33	<b>0.57</b>
Breast –w	<b>0.67</b>	0.54	0.58
Credit – a	0.68	0.63	<b>0.76</b>
Diabetes	0.48	0.44	<b>0.60</b>
Ionosphere	0.67	0.70	<b>0.89</b>
Mushroom	0.91	0.81	<b>0.98</b>

**Table 8.** Specificity (True Negative Rate-TNR)

Dataset	CBA	HiSP – GC	CHiSC
Breast Cancer	0.89	0.89	<b>0.95</b>
Breast –w	0.95	0.96	<b>0.97</b>
Credit – a	0.79	0.78	<b>0.86</b>
Diabetes	0.83	0.78	<b>0.85</b>
Ionosphere	0.88	0.92	<b>0.96</b>
Mushroom	0.93	0.93	<b>0.98</b>

Table 7 shows the sensitivity and Table 8 shows the specificity for six datasets. The proposed system has a high sensitivity (TPR) and specificity (TNR) for most of the datasets. The only dataset with low sensitivity is Breast-w. This is due to missing values. Further research on replacing missing values may improve the classification accuracy.

## 5.4 Precision and recall

Sensitivity and specificity measures are used in binary class datasets whereas precision and recall measures are used in multiclass problems.

The definition of precision and recall is given below.

$$\text{Precision (P)} = \frac{TP}{TP + FP} \quad \text{Eq. (11)}$$

$$\text{Recall(R)} = \frac{TP}{TP + FN} \quad \text{Eq. (12)}$$

Precision and recall measures are useful when the dataset consists of multiple classes. Thus we compute precision and recall for the following datasets: Balance Scale, Car, Ecoli, Flare, and Glass.

**Table 9.** Precision

Dataset	CBA	HiSP – GC	CHiSC
Balance Scale	0.64	0.64	<b>0.72</b>
Car	0.43	0.55	0.55
Ecoli	0.67	0.73	<b>0.80</b>
Flare	0.63	0.66	<b>0.73</b>
Glass	0.50	0.67	<b>0.75</b>

**Table 10.** Recall

Dataset	CBA	HiSP – GC	CHiSC
Balance Scale	0.67	0.64	0.62
Car	0.43	0.48	<b>0.52</b>
Ecoli	0.67	0.61	<b>0.75</b>
Flare	0.88	<b>0.93</b>	0.84
Glass	0.67	0.73	<b>0.75</b>

High precision implies that most of the predicted classes are correct. But high precision at low recall levels indicates that the classifier might not have predicted the rare class correctly. Tables 9 and 10 depict the precision and recall of the existing and proposed associative classifiers for multiclass datasets. It should be noted that the Car dataset has equal precision with the existing system. There is less recall for both the Balance Scale and Flare datasets due to smaller numbers in the particular class in the class distribution.

## 6 ACCURACY PERFORMANCE VALIDATION

To validate the statistical significance of the proposed associative classification algorithms, a cross validated paired t-test (Dietterich, 1998) was performed. The test was performed at level  $p = 0.05$  for all the datasets. The CHiSC algorithm was compared with CBA and the existing HiSP – GC associative classifiers using 12 datasets. Table 11 reports the significant evaluation results.

**Table 11.** Paired t - test

Method		Paired t-test, $p=0.05$	
		Existing Methods (Y)	
		CBA	HiSP – GC
Proposed Method (X)	CHiSC	9/2	6/4

*X/Y indicates that Proposed Method is significantly better X times / Worst Y times than the Existing Method.*

Table 11 shows paired t –test values for an average class accuracy of 10-cross fold validation. The test shows that average classification values of 10 different runs is significantly better for our proposed CHiSC algorithm for 9 and 6 different datasets (worse for 2 and 4 datasets) when compared with CBA and HiSP – GC methods respectively. The results in Table 11 indicate that CHiSC gives statistically significant better performance in these cases.

## 7 CONCLUSION

The CHISC-AC method is a computationally efficient classifier. This method predicts the class for each sample in a database based on an evaluation of its subsets of attribute values. This method uses the information gain attribute to generate a compact subset and calculates the *posteriori* confidence for each subset. Then it predicts the class based on that knowledge. CHISC-AC was tested with thirteen datasets from the UCI Machine Learning Repository. When compared with the traditional associative classifier and the existing lazy learning associative classification method, the experimental results show that the proposed system functions better than the traditional and existing methods most of the time and takes an average of 0.0133 seconds to predict an instance, less time than the other two methods use.

## 8 REFERENCES

- Agrawal, R. & Srikant, R. (1994) Fast algorithms for mining association rule. *Proceedings of the 20th International Conference on Very Large Data Bases*, pp 487-499.
- Barandela, R., Sanchez, J.S., Garcia, V., & Rangel, E. (2003) Strategies for learning in class imbalance problems. *Journal of Pattern Recognition* 36, pp 849–851.
- Baralis, E. & Torino, P. (2002) A lazy approach to pruning classification rules. *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02)*. Maebashi City, Japan, p 35-42.
- Baralis, E., Chiusano, S., & Graza, P. (2004) On support thresholds in associative classification. *Proceedings of the 2004 ACM Symposium on Applied Computing*. Nicosia, Cyprus: ACM Press, pp 553–558,.
- Baralis, E., Chiusano, S., Garza, P. (2008) A Lazy Approach to Associative Classification. *IEEE Transactions on Knowledge and Data Engineering* 20 (2), pp 156-171.
- Blake, C.L. & Merz, C.J. (1998) *UCI Repository of machine learning databases*.
- Chen, G., Liu, H., Yu, L., Wei, Q., & Zhang, X. (2006) A new approach to classification based on association rule mining. *Science Direct, Decision Support Systems* 42, pp 674– 689.

- Guha, R. & Schurer, S.C. (2008) Utilizing high throughput screening data for predictive toxicology models: protocols and application to MLSCN assays. *Journal of Computer Aided Modelling*, pp 367–384.
- Han, J., Pei, J., & Yin, Y. (2000) Mining frequent patterns without candidate generation. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. Dallas, TX: ACM Press, pp 1–12.
- Han, J. & Kamber, M. (2008) *Data Mining: Concepts and Techniques*. New York: Morgan Kaufmann Publishers.
- Hsieh, J.H., Wang, X.S., & Teotico, D. (2008) Differentiation of AmpC beta-lactamase binders vs. decoys using classification kNN QSAR modeling and application of the QSAR classifier to virtual screening. *Journal of Computer Aided Modeling*, pp 593–609.
- Li, W., Han, J., & Pei, J. (2001) CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. *Proceedings of IEEE International Conference on Data Mining (ICDM '01)*.
- Liu, B., Hsu, W., & Ma, Y. (1998) Integrating Classification and Association Rule Mining. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp 80–86.
- Merschmann, L. & Plastino, A. (2007) A lazy data mining approach for protein classification. *IEEE Transactions on Nanobioscience* 6(1), pp 36-42.
- Merschmann, L. & Plastino, A. (2010) HiSP-GC: A Classification Method Based on Probabilistic Analysis of Patterns. *Journal of Information and Data Management* 1(3), pp 423–438.
- Quinlan, J. (1986) Induction of decision trees. *Machine Learning*, pp 81–106.
- Snedecor, W. & Cochran, W. (1989) *Statistical Methods. Eighth Edition*. Iowa State University Press.
- Syed Ibrahim, S.P. & Chandran, K. R. (2011) Efficient Associative Classification Using Genetic Network Programming. *International Journal of Computer Applications* 29(6), pp 1-8. ISSN 0975-8887.
- Syed Ibrahim, S. P. & Chandran, K.R. (2011) Compact Weighted Class Association Rule Mining using Information Gain. *International Journal of Data Mining and Knowledge Management Process (IJDMP)* 1(6), pp 1-13. ISSN 2230 – 9608.
- Syed Ibrahim, S.P., Chandran, K.R., & Abinaya, M.S. (2011) Compact Weighted Associative Classification. *IEEE International Conference on Recent Trends in Information Technology (ICRTIT 2011)*. MIT, Anna University, Chennai, June 3-5, pp 1099 – 1104.
- Syed Ibrahim, S.P., Chandran, K. R., & Christopher, J. (2011) An Evolutionary Approach for Ruleset Selection in a Class Based Associative Classifier. *European Journal of Scientific Research*, pp 422-429. ISSN 1450-216X.50(3).
- Syed Ibrahim, S.P., Chandran, K.R., & Nataraj, R.V. (2011) LLAC: Lazy Learning in Associative Classification. *Springer Lecture Series in Communications in Computer and Information Science (CCIS), Advances in Communication and Computers* 190, Part I, pp 631 – 638.
- Tan, P., Kumar, V., & Steinbach, M. (2006) *Introduction to Data Mining*, Pearson Education.
- Weiss, G. M. (2004) Mining with rarity: a unifying framework. *SIGKDD Explorer Newsletter* 6, pp 7–19.
- Yin, X. & Han, J. (2003) CPAR: Classification Based on Predictive Association Rules. *Proceedings of Third SIAM International Conference on Data Mining (SDM '03)*.
- Zhang, X., Chen, G., & Wei, Q. (2011) Building a highly-compact and accurate associative classifier. *Applied Intelligence* 34(1), pp 74-86.

(Article history: Received 2 July 2014, Accepted 19 September, Available online 6 October 2014)