

# A Reinforcement-based Mechanism to Select Features for Classifiers in Ensemble Systems

Anne M P Canuto<sup>1</sup>, Karliane M O Vale<sup>2</sup> and Antonino Feitosa<sup>1</sup>

<sup>1</sup> Informatics and Applied Mathematics Department  
Federal University of Rio Grande do Norte (UFRN) Natal, RN - BRAZIL, 59072-970  
*anne@dimap.ufrn.br and antonino\_feitosa@yahoo.com.br*

<sup>2</sup> Campus Angicos  
Federal Rural University of Semi-Árido (UFERSA) Angicos, RN - BRAZIL, 59515-000  
*karlianev@gmail.com*

**Abstract:** Classifier ensemble are systems composed of a set of individual classifiers structured in a parallel way and a combination module, which is responsible for providing the final output of the system. One way of increasing diversity in classifier ensembles is to use feature selection methods in order to select subsets of attributes for the individual classifiers. In this paper, it is investigated the use of a simple reinforcement-based method, called ReinSel, in ensemble systems. This method is inserted into the filter approach of feature selection methods and it chooses only the attributes that are important only for a specific class through the use of a reinforcement procedure.

## I. Introduction

In the search for efficient pattern recognition systems, the idea of combining different classification methods has emerged as potentially very promising [3,16]. The main example of this idea is the ensemble systems (or committees), which exploit the idea that a pool of different classifiers can offer complementary information about patterns to be classified, improving the effectiveness of the overall recognition process. These systems are composed of a set of individual classifiers, organized in a parallel way, that receive the input patterns and send their output to a combination method which is responsible for providing the final output of the system.

In the context of ensemble, diversity is one aspect that has been acknowledged as very important [4,7,16]. For example, there is clearly no accuracy gain in a system that is composed of a set of identical base classifiers. In fact, the ideal situation is a set of individual classifiers with uncorrelated errors — they are combined in such a way as to minimize the effect of these failures. In other words, the individual classifiers should be diverse among themselves. Diversity in ensemble systems can be reached when the individual classifiers are built under different circumstances, such as: parameter setting, classification type and datasets. In the case of different datasets, the use of feature selection or data distribution methods in ensemble systems usually

increases the diversity of the members of an ensemble. This is because the individual classifiers will perform the same task but they were built using different basis (subsets of features).

Although diversity is an important aspect when designing ensemble systems, it is also important to search for accuracy. In this case, it is needed to have diverse and accurate classifiers in ensemble systems. Unlike traditional methods, feature selection methods for ensemble systems must take into consideration these two criteria. Based on this, class-based selection methods seem to be a good choice for ensembles. In these methods, each individual classifier will be responsible for one class of the problem. Along with robustness and practicality, the class-based feature selection methods need to choose features which have a high discriminative power, which are able to distinguish a class of a problem from the others.

The main aim of this paper is to present the ReinSel (Reinforcement Selection) method, which is a simple class-based feature selection method that uses a reinforcement procedure (reward/punishment parameter) in order to rank the attributes of a dataset. In using a reinforcing procedure, it is aimed to increase the discriminative power of the feature subsets, yet keeping their robustness and practicality. This method tries to choose attributes that are good only for the corresponding class. In order to do that, a two-step procedure is performed. In this sense, the ensemble systems will be composed by classifiers which are expert in answering about the belongingness of an input pattern to a specific class. In this sense, there will be, at least, one classifier per class in the ensemble systems. In order to analyze the feasibility of the ReinSel method, it will be compared with a standard filter-based method, which also uses a class-based procedure and they will be applied to two different sets of datasets (artificial and real).

This paper is divided into seven sections and it is organized as follows. Section 2 describes the research works related to the subject of this paper, while a brief

description of ensemble systems is illustrated in Section 3. In Section 4, the use of feature selection methods in ensemble systems is described, focusing on the definition of the ReinSel method. Section 5 presents the methodology used in the experimental work of this paper, while an analysis of the results provided by the empirical analysis is shown in Section 6. Finally, Section 7 presents the final remarks of this paper.

## II. Related Works

Recently, several authors have investigated the use of feature selection methods in classification systems, such as in [1,6,8,11,12,13,14,19,20,21,23,24,25,27,30,31,32,36]. In the context of ensembles, for instance, several feature selection methods have been proposed in the literature, which can be broadly divided into the two main approaches, which are:

- **Filter:** In this approach, as it can be found in [1,21,26,27,28], there is no need for a classification method to be used during the feature selection process. In other words, the feature selection process is independent from the classification method. In [28], for instance, the authors used different criteria to rank the attributes and to distribute them among the classifiers of different structures of ensembles;
- **Wrapper:** In this approach, as it can be found in [1,6,10,11,17,24,25,30,31], the feature selection process is dependent from the classification method. The feature subset is chosen based on the classification method used. Two different classification methods lead to different feature subset chosen. In [24], for instance, the authors presented an ensemble feature selection approach based on a hierarchical multi-objective genetic algorithm, where the first level generates several classifiers by conducting feature selection and the second one chooses the best ensemble among such classifiers. The fitness functions for the first level were accuracy and the number of features, while accuracy and diversity are the fitness functions in the second level.

In traditional feature selection (single classification methods), the wrapper approach involves the computational overhead of evaluating candidate feature subsets by executing a given learning algorithm on the database using each feature subset under consideration. The filter approach is generally computationally more efficient than the wrapper approach. However, the major drawback of the filter approach is that an optimal selection of features may not be independent of the inductive and representational biases of the learning algorithm that is used to construct the classifier. In addition, it is important to emphasize that the robustness of the traditional feature selection method (for single classifiers) is the most relevant characteristic when evaluating these methods.

In the context of ensembles, as already mentioned, there are other issues which are very important and must be taken into consideration when choosing a feature selection method and the main one is diversity. As the ensemble system used a two-step decision making process (the individual classifier level and the combination method level), the dependency of the chosen subset with the

classification methods can be smoothed out, since the classifier accuracy is not the only parameter to define the accuracy of the ensemble system. Furthermore, when using heterogeneous ensembles, different types of classification methods are used and a poor performance of one individual classifier (bad choice of the subset of attributes) can be overcome by a good performance of a different classifier. In this case, the diversity plays an important role in the efficiency of these systems.

Furthermore, in analyzing the performance of the feature selection approach when using ensembles, it is noted that the major drawback of the wrapper approach is emphasized, since the fitness function usually has to take into consideration the accuracy of the whole ensemble system, increasing even further the complexity of this function. In some works, as in [24,25], the fitness function is based on the accuracy of individual classifiers and only after the choice of a reduced set of attributes is done, these attributes are distributed over the individual classifiers. However, these works do not consider the accuracy of an ensemble and, most of the times, the used fitness functions do not reflect the real situation of the ensemble systems. Alternatively, the work in [1] has used both filter and wrapper approaches, in which the filter approach was used in the first phase, while a genetic search is employed in the second phase. It does smooth out the problem, but it still has a reasonable complexity involving its processing.

In summary, a feature selection method has to be efficient, robust and simple. Based on this, the use of the filter approach to select different subsets of attributes for the individual classifiers in an ensemble has become an interesting (and efficient) choice.

## III. Ensemble Systems

As previously mentioned, the goal of using ensembles is to improve the performance of a pattern recognition system in terms of better generalization and/or in terms of increased efficiency and clearer design [4,30]. There are two main issues in the design of an ensemble, which are: the ensemble components and the combination methods that will be used. In relation to the first issue, the members of an ensemble are chosen and implemented. The correct choice of the set of individual classifiers is fundamental to the overall performance of an ensemble. The ideal situation would be a set of base classifiers with uncorrelated errors - they would be combined in such a way as to minimize the effect of these failures. In other words, the individual classifiers should be diverse among themselves (for more details about diversity, see section 3.1). According to its structure, an ensemble can be divided in two main approaches: heterogeneous and homogeneous. The first approach combines different types of classification algorithms as individual classifiers. On the other hand, the second approach combines classification algorithms of the same type.

Once a set of individual classifiers has been created, the next step is to choose an effective way of combining their outputs. As already mentioned, the choice of the best combination method for an ensemble needs the execution of exhaustive testing. In other words, the choice of the combination method of an ensemble is very important and difficult to do. There are a great number of combination

methods reported in the literature [3,16,34]. According to their functioning, there are three main strategies of combination methods: fusion-based, selection-based, and hybrid methods.

- Fusion or combination based Methods: In this class, it is assumed that all classifiers are equally experienced in the whole feature space and the decisions of all of classifiers are taken into account for any input pattern. There are a vast number of combination-based methods reported in the literature. Examples of this class are: Sum, Majority Voting, Naïve Bayes, neural networks, fuzzy neural networks, fuzzy connectives among others;
- Selection-based Methods: In this class, only one classifier is needed to correctly classify the input pattern in selection-based methods. In order to do so, it is important to define a process to choose a member of the ensemble to make the decision, which is usually based on the input pattern to be classified. The choice is typically based on the certainty of the current decision. Preference is given to more accurate classifiers. One of the main methods in classifier selection is Dynamic Classifier Selection (DCS), proposed in [34].
- Hybrid Methods: They are the ones in which selection and fusion techniques are used in order to provide the most suitable output to classify the input pattern. Usually, there is a criterion process to decide whether to use the selection or combination method. The main idea is to use selection only and if only the best classifier is really good to classify the testing pattern. Otherwise, a combination method is used. Two main examples of hybrid methods are: Dynamic Classifier Selection based on multiple classifier behavior (Dcs-MCS) and Dynamic classifier selection using also Decision Templates (Dcs-DT) [16].

In the investigation of this paper, five fusion-based combination methods are in the ensemble systems, which are the following ones: Majority Voting, Sum, Weighted Sum, Multi-layer Perceptron and Naïve Bayesian.

#### A. Diversity in Ensembles

As already mentioned, there is no gain in ensembles that are composed of a set of identical classifiers. The ideal situation, in terms of combining classifiers, would be a set of classifiers that present uncorrelated errors. In other words, the ensemble must show diversity among the members in order to improve the performance of the individual classifiers [9].

Based on this, one can think that the more diverse the set of classifiers are, the more chance an aggregation of these classifiers has to reach the best classifier. Nevertheless, up to date, trying to measure diversity and use it explicitly in the process of building the ensemble does not share the success of the implicit methodologies. This issue has been addressed by some authors, such as in [16,18]. On the other hand, some authors have been working in the idea that diversity is not the only tool to obtain accurate ensembles [5,15]. In this paper, for instance, the role of diversity will be investigated in the context of feature selection methods, not using any explicit diversity measures.

Diversity in ensemble systems can be reached when the individual classifiers are built under different circumstances, such as in the following ways:

- Different parameter settings of the classifiers: In this approach, diversity can be reached through the use of different initial parameters setting of the classification methods. In a neural network, for instance, this would mean varying the weights and topology of a neural network model.
- Different classifier training datasets: In this approach, diversity can be reached through the use of learning strategies such as Bagging and Boosting or the use of feature distribution methods.
- Different classifier types: In this approach, diversity can be reached through the use of different types of classifiers. For instance, usually an ensemble which is composed of neural network and decision tree is more diverse than ensembles composed of only neural networks or decision trees.

In this paper, variations of the diversity are captured using different types of classifiers and different training datasets. Also, the use of ensembles with various levels of diversity is used in order to analyze the influence of the performance of presented feature selection method when varying diversity in the ensembles.

#### IV. Feature Selection in Ensembles

Feature selection methods try to reduce the dimensionality of the attributes of a dataset, spotting the best ones. The attribute subset selection can be defined as the process that chooses the best attributes subset according to a certain criterion, excluding the irrelevant or redundant attributes. In using feature selection methods, it is aimed to improve the quality of the obtained results. In the context of ensembles, for instance, feature selection methods provide different subsets of attributes for the individual classifiers, aiming to reduce redundancy among the attributes of a pattern and to increase the diversity in such systems. In this paper, feature selection methods will be used in ensemble systems. Hence, hereafter, the term feature selection will be used in the context of ensembles.

As already mentioned, there are several feature selection methods that can be used for ensembles, which can be broadly divided into two main approaches, which are: filter and wrapper. In the filter approach, usually, the attributes are ranked based on a certain criterion and the top N attributes are picked. In other words, a general ranking procedure is performed in which attributes are assessed for all classes of the problem, called general ranking process. This means that the diversity of the individual classifiers might be deteriorated, since the chosen features must have an efficient performance for all classes of the problem. In addition, it is well known that different classes of a problem can have different particularities and levels of difficulty. When using a general ranking procedure (for all classes), the difficulties of one class are being distributed among all other classes. In this sense, classes which are not very difficult to be classified may become more difficult. Moreover, different classes of one problem might need a different number of attributes to be classified. For instance, an attribute can be very important to one class and not very

important to other classes. Because of this, the idea of using class-based ranking has emerged.

In using class-based ranking, each classifier of an ensemble is associated with one class, increasing the processing time of the recognition process. However, it is believed that this will be compensated by the decrease in the number of features for each class and by the increase in the accuracy of the classification method. There are some works in the literature which use class-based feature selection, such as the favorite class method [26]. It is important to emphasize that although the choice of the subsets of attributes are based on a specific class, the corresponding classifiers will be built using all patterns of the whole training dataset, which includes patterns of all classes to be recognized.

In some real cases, however, an attribute can have a similar importance for two or more classes. In this case, when using a class-based feature selection, this attribute will probably be chosen for the different classifiers. As the classifiers of an ensemble will be responsible for one class of the application, the choice of this attribute for both classifiers may affect the accuracy of these classifiers, making them confuse patterns of both classes. In other words, it is important that the chosen features must show a high discriminative power, in which they will be able to distinguish one class from the others.

### B. The ReinSel Method

Aiming to add discriminative power in the subset of chosen features and still keeping their robustness, a feature selection method called ReinSel (Reinforcement Selection) is proposed in this paper. This method was originally proposed in [33] and this work further investigates the performance of this method. It is a filter method which will search for attributes that are important for one class and not very important for other classes, making the classification method more secure about the class to be classified (attribute which will not make confusion in the classification method). The main idea is that one classifier, which will be responsible for one class of the problem, will base its decision on attributes which are important only for this class.

The idea of rewarding and penalizing features is to give more discriminative power to the subsets of features and, as a consequence, to build diverse classifiers. In the context of ensembles, it is possible to have classifiers which have high expertise in some parts of the problems (individual classes) and this can be joined together to reach a final decision. When penalizing features which are important to more than one class, it is aimed to select features which are important for just for one class. In this sense, these selected features will be part of classifiers who will be expert in that specific class. As an ensemble is composed of a set of classifiers, the built ensemble can have classifiers which are experts for one class, covering all classes of the problems. In doing this, it is trying to solve two of the main problems in ensemble systems, which are: diversity and dimensionality reduction. In order to choose the attributes, the ReinSel method uses two ranking procedures (two-step).

The algorithm of the ReinSel method is presented in Figure 1, while Figure 2 shows the general functioning of

the ReinSel method, applying to a problem with three classes. The first ranking process (step 1) is similar to the ones used in classical filter approaches, in which one criterion  $t$  is used to rank, in a decreasing order, the attributes. Once the first ranking is done, a second ranking process (step 2) is used for the final choice of the subsets. In this second ranking, the position of the attribute is defined based on the ranking of this attribute in the analyzed class, along with the position of this attribute in the other classes.

#### Feature Section Algorithm

**Input:** Data  $D = f(X_n, Y_n)$ ,

where  $X_i = \{x_{1i}, x_{2i}, \dots, x_{mi}\}$  and  $Y_i = \{1, \dots, L\}$ ;

$c$  = number of features and  $L$  = number of

labels (classes)

1. For  $r = 1$  to  $L$

1.1. Calculate  $V_r^1 = \text{Ranking}(X_n, t)$  % first ranking, based on criterion  $t$

2. For  $r = 1$  to  $L$

2.1. For  $i = 1$  to  $c$

2.1.1. Calculate  $Rew_{ir}$  and  $Pun_{ir}$

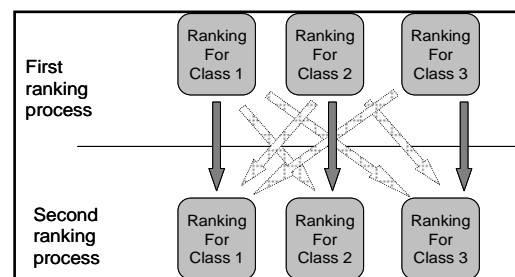
2.1.2. Apply the RP parameter to attribute  $i$

2.2. Calculate  $V_r^2 = \text{Ranking}(X, RP)$  % second ranking

2.3. Select the first  $N$  to be part

**Output:**

**Figure 1:** Pseudo-code of the ReinSel Algorithm



**Figure 2:** the process of a class-based feature selection method

The main idea of the ReinSel method is that if an attribute is highly (in the top positions) ranked in an analyzed class, this means that this attribute is important for the classification process of this class and this will be positively counted for this attribute (high values for  $Rew_{ir}$ , step 2.1.1). In addition, if an attribute is also highly ranked on the other classes, this means that this attribute is important for more than one class and the choice of this attribute can affect the accuracy of the classifiers, making it not very discriminative. In this case, this will be negatively counted for the attribute (high values for  $Pun_{ir}$ , step 2.1.1).

In order to take into consideration the positive and negative points of an attribute, the second ranking procedure is based on a RP (reward/punishment) parameter, in which an attribute is rewarded for this position in the ranking of the analyzed class (the magnitude of this reward depends on the position of the attribute in the ranking) and it is punished by its highest position in the ranking of the other classes (the magnitude of this punishment depends on the position of the attribute in the ranking). Then, the RP parameter can be described as follows:

$$RP_{ir} = \text{Re } w_{ir} - \text{Pun}_{ir} \quad (1)$$

Where:

$$\text{Re } w_{ir} = v_{ir}^1 + \frac{NA}{NA + R_{ir}} \quad (2)$$

And

$$\text{Pun}_{ir} = \max_{s=1, \dots, L / s \neq r} \left( v_{is}^1 + \frac{NA}{NA + R_{is}} \right) \quad (3)$$

Where:

- $v_{ir}^1$  is the result of the first ranking of attribute  $i$  for class  $r$ ;
- $R_{ir}$  is the ranking of attribute  $i$  in class  $r$  and
- NA is the total number of attributes used in the dataset.

The first ranking is performed and its result is calculated for all attributes of all classes. For instance, in analyzing the  $i$ -th attribute for class  $r$ , its result for the corresponding class ( $v_{ir}^1$ ) will be considered for the reward calculation,

while the result of this attributes for the other classes ( $v_{is}^1$ ,  $s=1, \dots, L$ ,  $s \neq r$ ) will be considered for the punishment calculation. Along with the result of the first ranking, the reward/punishment calculation also takes into consideration the position (ranking) of the attribute  $i$  in class  $c$ . In other words, a good position in the ranking means a high reward or punishment.

The idea of the RP parameter is that the more an attribute is in the top position of the analyzed class, the higher is the reward. In equation 2, if an attribute  $i$  is the first one in the ranking of the analyzed class ( $R_{ir} = 0$ ), the result of the first ranking ( $v_{ir}^1$ ) is added 1 (the maximum possible reward). On the other hand, if this attribute is in the last position ( $R_{ir} = NA-1$ ), the result of the first ranking ( $v_{ir}^1$ ) is added  $\approx 0.5$  (the minimum possible reward). The same idea is used for the punishment part, in which the highest ranking in the other classes an attribute is, the higher is its punishment. For instance, if an attribute is the top attribute in the ranking of one class (it is not the analyzed class), then its position is 0 ( $R_{jr}$ ) and its result of the first ranking is 1 (the maximum possible punishment). Also, its result of the first ranking will be added  $\approx 0.5$  (the minimum possible punishment). The punishments of an attribute is chosen by the maximum punishment possible (max operator) of its position in the other classes. The idea is to choose the worst possible scenario.

In using the RP parameter in the illustration of the ReinSel method, figure 2, the solid arrows indicates the reward influence (eq. 2) and dashed arrows indicates the punish influences (eq. 3). After the use of the RP parameter in all parameters of all classes, the attributes are ranked in a decreasing order and the first N attributes are picked (step 2.3). The value of N might be fixed or not, in which a

threshold might be used in order to define whether an attribute must be picked or not. In this paper, an analysis of the ReinSel method using both situations (fixed and varying values of N) is done.

## V. Experimental Work

### A. Datasets

All ensemble systems are applied to six different datasets, distributed in two sets, which are called: real and artificial datasets. The artificial datasets are described as follows are composed of two datasets, called Gaussian3 and Simulated6: These datasets are synthetic databases that simulate microarray data and were created to test the ML algorithms in the gene expression analysis [22]. These two synthetic databases have attributes related to each of their classes. Thus, there is a set of 200 attributes in Gaussian3 that determine each class. This relation between attribute and class is exclusive. The 200 attributes that determine Class 2, for example, do not determine Class 0 or Class 1. Similarly, Simulated6 presents sets of 50 attributes exclusively related to each class. Simulated6 has 6 classes with 50 exclusively related attributes, totaling 300 attributes. As Simulated6 actually has 600 attributes, the remaining 300 are noisy ones [22].

The real datasets are described as follows:

- St Jude Leukemia is a database obtained from gene expression data of leukemia cells [22,35]. Its data is represented by a matrix  $M(i, j)_{n \times m}$ , in which its  $n$  lines (examples, objects, instances) are samples of cells with leukemia and its  $m$  columns (attributes) are genes. In this specific dataset,  $n=248$  and  $m=985$ . Each element  $(i, j)$  of the matrix is a number that measures the expression level of gene  $j$  found in cell sample  $i$ . The attribute values of the databases present quite distinct scales. Therefore, in this paper, all attributes were normalized to the interval  $[0, 1]$ .
- Splice dataset contains a primate splice-junction gene sequences (DNA) with associated imperfect domain theory, obtained from the UCI repository [2]. A total of 3190 Instances using 60 attributes were used. These attributes describe sequences of DNA used in the process of creation of proteins.
- Images. This dataset was also taken from the UCI repository (segmentation dataset) [2]. The 2.310 instances were drawn randomly from a dataset of 7 outdoor images. The images were hand-segmented to create a classification for every instance, where each instance is a 3x3 region. Eighteen attributes were extracted from each region.
- Proteins. This dataset represents a hierarchical classification, manually detailed, of known structures of proteins. They are organized according to their evolutionary and structural relationship. The main protein classes are all- $\alpha$ , all- $\beta$ ,  $\alpha/\beta$ ,  $\alpha+\beta$  and small. It is an unbalanced dataset, which has a total of 582 patterns, in which 111 patterns belong to class all- $\alpha$ , 177 patterns to class all- $\beta$ , 203 to  $\alpha/\beta$ , 46 to class  $\alpha+\beta$  and 45 to class small.

The ensemble size is defined by the number of classes of the used dataset. In this sense, ensembles of size 3 are used for Gaussian and Splice datasets, ensembles of size 5 are used for St. Jude Leukemia and Proteins datasets, ensembles of size 6 are used for Simulated6 dataset and ensembles of size 7 are used for Image dataset.

### B. Methods and Material

In order to investigate the performance of the ReinSel method, an empirical analysis is performed. Therefore, the proposed method along with the original filter method (only the first ranking) is analyzed in this investigation. The ensemble systems using feature selection are also compared with ensemble systems with no feature selection. For the ensemble systems using feature selections (original and ReinSel), one criterion was used as basis (first ranking) for the ranking of the attributes, which is.

- Spearman Correlation: it is a non-parametric measure of correlation. In other words, it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables [24]. When using Spearman, the idea is to see the correlation between one attribute and the class label. However, the class label is transformed into a binary attribute, in which 1 means patterns of the analyzed class and 0 means patterns of other classes. The main idea here is to select features which are able to distinguish one class from the others.

Four types of classification methods are used as individual classifiers for the ensemble systems, which are: k-NN (nearest neighbor), C4.5 (decision tree), NB (Naïve Bayesian Learning) and MLP (multi-layer Perceptron) neural network. The choice of the aforementioned classifiers was due to the diversity in the classification criteria used by each method chosen. In addition, these individual classifiers are combined using five different combination methods, which are: Majority Voting, Sum, Weighted Sum, MLP and Naïve Bayesian. (for more details, see Section 5.2).

For each system, several different configurations are used, which are homogeneous (HO) and Heterogeneous (HE) structures of ensembles. It is important to emphasize that the heterogeneous structures include ensembles with 2, 3 and 4 different types of classifiers were taken into consideration. As there are several possibilities for each structure, this paper presents the average of the accuracy delivered by all possibilities of the corresponding structure. For instance, the homogeneous ensembles (NH) represent the average values for NB, k-NN, MLP and DT ensembles.

The individual classifiers and the ensemble systems were built by using the  $k$ -fold cross-validation methodology. In addition, to compare the impact of the ReinSel method, the accuracies of the ensembles when using this method were compared with the ensemble systems using the original feature selection method and with systems with no feature selection. To do this comparison, a statistical test was applied, which is called hypothesis test (t-test) [16]. It is a test which involves testing two learned hypotheses on identical test sets. In order to perform the test, a set of

samples (classifier results) from both algorithms should be used to calculate error mean and standard deviation. Based on the information provided, along with the number of samples, the significance of the difference between the two sets of samples, based on a degree of freedom ( $\alpha$ ), is defined. In this paper, the confidence level is 95% ( $\alpha = 0.05$ ).

For the feature selection procedure, the choice of the number of attributes to be part of a subset is done under two main approaches, fixed and variable. In the fixed approach, a fixed number of attributes is chosen for all classes. In this investigation, this number was chosen in such a way that all classifiers will use the same number of attributes of the whole dataset or half of them. For the Gaussian dataset, for instance, there are 600 attributes and three classes. Then, the numbers of attributes per classifier were 100 ( $100 \times 3 = 300 \rightarrow$  half of attributes) and 200 ( $200 \times 3 = 600 \rightarrow$  the same number of attributes of the whole dataset). In the variable approach, a threshold was defined and all attributes whose values are higher than the threshold are chosen. This means that the number of attributes seen by the classifiers is variable. Also, this threshold was defined by each dataset. Table 1 shows the average number of attributes used in this experiment. It is important to observe that the subsets used by the original feature selection are bigger than the ones used by the ReinSel method. These values were intentionally chosen in order to analyze the performance of the ReinSel method in extreme situations (having to choose a small number of attributes).

**Table 1:** The average number of attributes used when applying the variable strategies, for original and ReinSel methods

	Original	ReinSel
Gaussian3	72	45
Simulated6	51	48.16
St Jude	207	74.83
Splice	43.33	10.66
Proteins	46.2	11.2
Image	5	3.42

### C. Combination Methods

As mentioned previously, this investigation was conducted using five combination methods, which are: Majority voting (MV), Sum, Weighted Sum (W-Sum), Neural Network (MLP) and Naïve Bayes (NB). All the combination methods used in this study were obtained from the WEKA machine learning visual package (<http://www.cs.waikato.ac.nz/~ml/weka>). Majority Voting and Sum methods have no parameters to be set and they are used as it is in Weka. For the Weighted-Sum, a very simple way of calculating weights was used, which was based on the recognition rate of the classifiers in a validation set. The configurations of the other two combination methods were chosen through exhaustive tests in each method using each data set. As six datasets were used, it is presented below a description of the parameters used in the combination methods for the majority of the datasets.

- Multilayer Perceptron (MLP): A 3-layer neural network was used and the termination condition

was number of iterations. Learning rate = 0.01, momentum = 0.5 and the number of hidden neurons lies in the [3,30] interval. All the other parameter were set to the WEKA default;

- Naïve Bayesian (NB): The use of the kernel estimator parameter was set for true. All the other parameter were set to their default;

Three of the combination methods are trainable methods (MLP, NB and W-Sum). In this sense, the datasets are divided into 11 folds (subsets) of equal size (keeping the distributions of the classes in each fold). From these sets, 10 of them are used in the 10-fold cross validation procedure and the remaining one works as the validation set to train the combination methods.

## VI. Results and Discussion

Table 2 shows the accuracy and standard deviation of the ensemble systems when using no feature selection, for all six datasets. In other words, all individual classifiers of the systems use all attributes of the corresponding dataset. The average accuracies of the individual classifiers are also shown in Table 2 (first column). In all datasets, ensembles using all five combination methods are also shown in Table 2. As it can be observed from Table 2, for all datasets, both homogeneous (HO) and Heterogeneous (HE) structures of ensembles are analyzed.

The accuracy of the ensemble systems in Table 2 increased when increased the number of different types of classifiers, since the accuracy of the heterogeneous ensembles were higher than the homogeneous one, for most of the analyzed cases. It is important to notice that the accuracies of the individual classifiers are very similar, in which the homogeneous individual classifiers had a slightly higher accuracies than the heterogeneous ones. Despite that, the heterogeneous ensembles provided the highest accuracies, for all combination methods analyzed. This result shows that the combination of different classifiers tend to provide more ensembles [4]. The combination method that provided the highest increase in accuracy of the heterogeneous ensembles is Weighted Sum (5.11%, average increase), followed by NB (4.19), MLP (3.69), MV (3.06) and Sum (2.31).

### A. The Artificial Datasets

In this section, the results of the ensemble systems when applied to the artificial datasets (Gaussian and Simulated) are shown. Table 3 shows the results of the ensemble systems for these datasets. The number of attributes for each subset was fixed and variable. For the variable number of attributes, a threshold is chosen and the attributes whose values are higher than this threshold is chosen to be part of the subset. Finally, two different ensemble structures are analyzed, which are: homogeneous and heterogeneous.

In a general perspective, it can be noticed from Table 3 that the ensemble systems provided accuracies which are similar to the ones obtained by the ensemble systems with no feature selection (Table 2), mainly when using subsets

with a fixed number of features. When using the ReinSel method, the accuracy of the ensemble systems had improvements in all analyzed cases, when compared with the ensembles using the original feature selection method. In addition, it can be seen in Table 3 that the standard deviation of the ensemble systems are very high, reaching values around 15%. It is important to emphasize that these two datasets have a small number of patterns (60 patterns the whole dataset, having 6 or 5 patterns in each fold). In this case, one pattern which is wrongly classified by the ensemble system has a strong impact in the accuracy level of this system, making the standard deviation have high values.

In relation to the strategy to choose the size of the feature subsets, the use of fixed size has provided ensembles with higher accuracies. However, it is important to emphasize that the average sizes of the subsets were smaller than the first fixed subset (100 for Gaussian and 50 for simulated). This was intentionally chosen. It is important to notice that even having produced smaller subsets, the ensembles which used the ReinSel method have provided the highest accuracies. In comparing the increase of accuracy of the ReinSel method over the original one under the different choice of the subset sizes, ensembles using the first fixed subset have provided the highest increase (11.14 percentage points of average increase), followed by the second fixed subset and variable.

In analyzing Table 3 based on the ensemble structures, it can be seen that the use of heterogeneous structures have provided ensembles with higher accuracies, when compared with the homogeneous ones. In addition, the highest increase in accuracy of the ReinSel method was provided by the heterogeneous structure (5.45, average increase), followed by homogeneous (4.34).

In analyzing Table 3 based on the combination methods, the ensembles which reached the highest accuracies were the ones combined by MLP, followed by NB, W-Sum, Sum and MV. However, the highest increase in accuracy, when comparing ReinSel with the original one, was reached with ensembles combined by MV (9.58 average increase), followed by MLP (4.71), NB (4.58), Sum (4.33) and W-Sum (1.33). The negative point is the weak performance of the ensembles combined by the weighted sum, for both feature selection methods. This is an indication that the use of weights is important, but the search for the optimal weights is not an easy task.

In the statistical test, the comparison of the ReinSel method with ensembles using the original methods showed that the improvement of the ReinSel method was statistically significant in around 39% of the analyzed cases (23 out of 60). In these 23 cases, 56% were for the Simulated dataset (13 statistically significant improvements), while 44% were for Gaussian dataset (10). When analyzing the combination method, the method which provided the highest number of statistically significant improvement was Majority voting, which had 26% of the statistically significant improvements (5), followed by Sum and Weighted Sum with 22% each of them (5), by NB with 17% and MLP with 13% (3).

In relation to the ensemble structures, the number of statistically significant improvement was basically the same for both structures, with 12 for heterogeneous and 11 for

homogeneous. It is important to mention that the sample used by the heterogeneous is higher than the ones used by the homogeneous ones. In this experiment, values show the average accuracy of all possible combinations. In the homogeneous structures, this means four combinations (ensembles composed of NB, DT, k-NN and MLP) are used and each of them was built under 10-fold cross validation. In this case, the homogeneous structures have, in total, a sample composed of 40 members. This value is used in the statistical test as the number of groups of each case. In contrast, in the heterogeneous structures, the number of combination is very high, leading to a high number of members. This makes the statistical test to detect more easily statistical differences when using large samples.

The last analysis with the artificial datasets is related to the amount of noisy attributes chosen by the feature selection methods during their processing. As already mentioned, the simulated dataset are composed of 600 attributes in which 300 are noisy ones. It is important to analyze whether the feature selection methods were capable of avoiding to choose these attributes. In order to do this analysis, Table 4 shows the average amount of noisy attributes that were pick for ReinSel and the original methods.

From Table 4, it is possible to observe that the use of Spearman correlation was very effective to detect noisy attributes, independently of the feature selection method. It is an expected result since a noisy attributes can have high variability in any particular class. Although they have high variability, they do not have high correlation, making easy for Spearman to detect these attributes. The other important point from Table 4 is that the ReinSel method did not selected any noisy attributes in two cases (first fixed and variable) and less than one noisy attribute in the remaining case. This is an indication that the reinforcement process is helping the feature selection to detect the important features, discarding the noisy ones.

### B. The Real Datasets

In this section, the results of the ensemble systems when applied to the real datasets (Splice, St Jude, Proteins and Image) are shown. Table 5 shows the results of the ensemble systems for these datasets. As in the previous Section, the number of attributes for each subset was fixed and variable.

In a general perspective, it can be noticed from Table 5 that the ensemble systems provided accuracies which are similar to the ones obtained by the ensemble systems with no feature selection (Table 2), mainly when using subsets with a fixed number of features. The only exception is the Image dataset, in which the accuracies of the systems with feature selections had a lower performance than the ones without feature selections. It is believed that as this dataset has a small number of attributes used by the feature subsets deteriorated the accuracies of the ensemble systems. In addition, ensembles using the ReinSel method, the accuracy of the ensemble systems had improvements in all analyzed cases, when compared with the ensembles using the original feature selection method. However, once again, the use of ReinSel in the Image dataset did not increase the accuracy of the ensemble systems, when compared with the original

feature selection method. This fact happened mainly for variable feature subset. The possible reason for this performance is the small number of attributes and the high number of classes that might have affected negatively the ReinSel method.

In relation to the strategy to choose the size of the feature subsets, the use of fixed size has provided ensembles with higher accuracies in almost all cases. Once again, for variable subsets, the ensembles which used the ReinSel method have provided the highest accuracies, even having smaller feature subsets. In comparing the increase of accuracy of the ReinSel method over the original one under the different choice of the subset sizes, ensembles using the first fixed subset have provided the highest increase (4.9 of average increase), followed by the second fixed subset (4.3) and variable (0.90).

In relation to the ensemble structures, Table 5 shows a similar behavior of the last section. In other words, the use of heterogeneous structures has provided ensembles with higher accuracies, when compared with the homogeneous ones. In addition, the highest increase in accuracy of the ReinSel method was provided by the heterogeneous structure (3.59, average increase), followed by homogeneous (3.13).

Of the combination methods, the ensembles which reached the highest accuracies were the ones combined by MLP, followed by NB, Sum, W-Sum and MV. However, the highest increase in accuracy, when comparing ReinSel with the original one, was once again reached with ensembles combined by MV (4.98 average increase), followed by W-Sum (4.15), Sum (3.16), MLP (2.59) and NB (2.55). Although the ensembles combined by MLP had provided the highest accuracies, they have provided one of the lowest increases (ReinSel against Original). MLP is itself a more elaborated method which can overcome some weakness of the input patterns. In this sense, the use of less representative features was overcome by the MLP processing. Therefore, the increase in accuracy was not as high as in more simplified models like Sum and Majority Voting. Despite that, an average increase of 2.59 percentage points is not that dreadful.

In the statistical test, the comparison of the ReinSel method with ensembles using the original methods showed that the improvement of the ReinSel method was much higher than the artificial datasets, in which there are statistically significant improvements in 75% of the analyzed cases (89 out of 120). In these 89 cases, 32% of them were found to be for the Image dataset (28 out of 89), followed by Splice (26), proteins (25) and St Jude (10). It is important to emphasize the excellent result obtained by Image dataset, since the ReinSel method was statistically better than in more than 93% of the cases for these datasets (with 28 out of 30). Even in the case of Image dataset, the improvement of the ReinSel methods occurred in 33,3% of the analyzed cases for this dataset.

In relation to the ensemble structures, as in the previous section, the number of statistically were higher for the homogeneous structure, where 52% of the statistically significant improvements was obtained (46), while 48% were obtained by the heterogeneous ones (43). Finally, of the combination method, the Weighted Sum method has provided the highest number of statistically significant



improvements, with 23% of the cases for each combination method (21), followed by Majority voting and NB with 20% (18), Sum with 19% (17) and MLP with 18% (16).

## VII. Final Remarks

This paper proposed ReinSel, which is a feature selection method to be used in ensemble systems. This method applies a reinforcement procedure in which it selects important attributes to a corresponding class and the ensemble systems need to have, at least, one classifier to correctly recognize each class. The main idea of the ReinSel method is to use simple and efficient method to distribute attributes among the individual classifiers of an ensemble.

In order to analyze the feasibility of the ReinSel method, an empirical evaluation was performed. In this analysis, ensemble systems using the ReinSel method were compared with systems using a standard (original) feature selection method, as well as systems with no feature selection methods. All three types of systems were applied to six datasets, using fixed or variable sizes of subsets for the individual classifiers.

Based on the empirical analysis elaborated in this paper, the main points that could be concluded are described as follows.

- The ensemble systems using the ReinSel method had always provided similar accuracies, when compared with systems using no feature selection method. This shows that the idea of choosing attributes that are important only for one class has had a positive effect in the performance of the individual classifiers and, as a consequence, in the performance of the ensemble systems. In addition, the use of ReinSel have led to more accurate ensembles, when compared with the original feature selection method;
- The use of heterogeneous structures of ensembles has led to more accurate ensembles, when compared with the homogeneous ones. In addition, the use of heterogeneous structures was more positive for the ReinSel than for the original one, since it provided a higher increase in accuracy, when compared with the homogeneous ones, for both sets of datasets (artificial and real);
- Of the combination methods, the ensembles which reached the highest accuracies were the ones combined by MLP and the highest increase, when comparing ReinSel with the original method, was reached when using Majority Voting.
- The statistical test illustrated that the highest number of statistically significant improvements was reached by the real datasets, in which 75% of the analyzed cases were statistically significant improvements. It is important to emphasize the excellent performance of the ReinSel methods for Image dataset, having statistically significant improvements in more than 93% of the analyzed cases. In analyzing both sets of datasets, the heterogeneous structures provided higher number of statistically significant improvements, when compared with the homogeneous ones. Finally, it could be noted that the impact of the ReinSel was more positive in

simple combination methods, such as Majority voting, Sum and Weighted sum, when compared with more elaborated ones (MLP and NB). In these simple methods, the number of statistically significant improvements was higher than the more elaborated ones.

- One of the artificial datasets (Simulated6) had noisy attributes and it was possible to analyze the amount of noisy attributes chosen by the feature selection methods. According to this analysis, the ReinSel method managed to avoid choosing noisy attributes. The ReinSel method managed to choose less than 1 attribute, while the original method has chosen an average of 16 noisy attributes.

## References

- [1] M. Bacauskiene, A. Verikas, A. Gelzinisa and D.Valinciusa. A feature selection technique for generation of classification committees and its application to categorization of laryngeal images. *Pattern Recognition* 42(5), 645 – 654, 2009.
- [2] C.L Blake and C.J Merz. “UCI Repository of machine learning databases”. Univ of California, Dept of Information and Computer Science. [<http://www.ics.uci.edu/~mlern/MLRepository.html>]
- [3] A. Canuto.” Combining neural networks and fuzzy logic for applications in character recognition”. PhD thesis, Univ of Kent, 2001.
- [4] A Canuto, M Abreu, L Oliveira, J Xavier Junior and A Santos. Investigating the Influence of the Choice of the Ensemble Members in Accuracy and Diversity of Selection-based and Fusion-based Methods for Ensembles. *Pattern Recognition Letters*, 28(4), pp. 472-486, 2007.
- [5] A M Canuto, A Santos and R Vargas. Ensembles of ARTMAP-Based Neural Networks: An Experimental Study. Accepted to be published in *Applied Intelligence Journal*, 2009.
- [6] D.Caragea, A.Silvescu, and V.Honavar. “Decision tree induction from distributed, heterogeneous, autonomous data sources”. In *Conf on Int Systems Design and App (ISDA)*, 2003.
- [7] J Czyn and M Sadeghi and J Kittler and L Vandendorpe. Decision fusion for face authentication, *Proc First Int Conf on Biometric Authentication*, 686-693, 2004.
- [8] J Derrac, S Garcia and F Herrera. A First Study on the Use of Coevolutionary Algorithms for Instance and Feature Selection. *4th International Conference of Hybrid Artificial Intelligence Systems (HAIS)*, LNCS 5572, 557-564, 2009.
- [9] T. G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems*, Lecture Notes in Computer Science (LNCS 1857), 2000, Springer, pp. 1–15.
- [10] X Fu and L. Wang: Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. *IEEE Trans. System, Man, Cybern, Part B: Cybernetics* 33, 399-409, 2003.
- [11] Gerra-Salcedo, C.,Whitley, D.: Genetic approach to feature selection for ensemble creation. In: *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 236–243, (1999).

- [12] T K Ho. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(8), 832–844 (1998).
- [13] J Hua, W D. Tembe and E R. Dougherty. Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition* 42(3), 409-424, 2009.
- [14] R Jensen and Q Shen. New Approaches to Fuzzy-Rough Feature Selection. *IEEE Transactions on Fuzzy Systems*, 17( 4), 824-838, 2009.
- [15] A H-R Ko, R Sabourin, and AS Britto. Combining Diversity and Classification Accuracy for Ensemble Selection in Random Subspaces. *International Joint Conference on Neural Networks (IJCNN)*, pp. 2144-2151, 2006.
- [16] L I Kuncheva. *Combining Pattern Classifiers. Methods and Algorithms*, Wiley, 2004.
- [17] L I Kuncheva and L C Jain. Designing classifier fusion systems by genetic algorithms. *IEEE Trans. Evol. Comput.* 4(4), 327–336 (2000).
- [18] L Kuncheva and C Whitaker. Measures of diversity in classifier ensembles, *Mach Learning*, 51, 181-207, 2003.
- [19] B Liu, C Wan, and L.P. Wang, An efficient semi-supervised gene selection method via spectral biclustering, *IEEE Trans. Nano-Bioscience*, vol.5, pp.110-114, 2006.
- [20] H Liu.; E Dougherty.; J Dy.; K Torkkola.; E Tuv; H Peng; C Ding; F Long; M Berens; L Parsons; Z Zhao; L Yu.; G Forman; Evolving feature selection, *IEEE Intelligent Systems*, vol.20, pp.64-76, 2005.
- [21] P.J Modi and P.W Tae Kim. “Classification of Examples by multiple Agents with Private Features”. *IEEE/ACM Int Conf on Intelligent Agent Technology*, 223-229, 2005.
- [22] S. Monti, P. Tamayo, J. Mesirov, and T. Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, vol.52:91–118, 2003.
- [23] I-S Oh; J-S Lee and B-R Moon: Hybrid genetic algorithms for feature selection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 26, 1424 – 1437, 2004.
- [24] L S Oliveira, M Morita and R Sabourin. Feature selection for ensembles applied to handwriting recognition. *International Journal of Document Analysis* 8(4): 262–279 (2006).
- [25] D. Opitz, Feature selection for ensembles, in: *Proc. 16th Nat. Conf. on A. Intelligence*, AAAI Press, pp. 379–384, 1999.
- [26] N. Oza. and K Tumer. Input Decimation Ensembles: Decorrelation through Dimensionality Reduction. In *Proc. Of 12nd Int. Workshop on Multiple Classifier Systems (MCS)*, 2001, LNCS 2096, pp.238-247.
- [27] J J Rodriguez, L I Kuncheva and C.J. Alonso, Rotation Forest: A new classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (10), 2006, 1619-1630.
- [28] L Santana; D Oliveira; A Canuto and M Souto. A Comparative Analysis of Feature Selection Methods for Ensembles with Different Combination Methods. In: *International Joint Conference on Neural Networks (IJCNN)*, pp 643-648, 2007.
- [29] C. Spearman, "The proof and measurement of association between two things" *Amer. J. Psychol.* , 15 pp. 72–101, (1904).
- [30] A Tsymbal; M. Pechenizkiy and P.Cunningham. Diversity in search strategies for ensemble feature selection •*Information Fusion*, Vol 6(1), pp. 83-98, 2005.
- [31] A Tsymbal, S Puuronen and D W Patterson. Ensemble feature selection with the simple Bayesian classification. *Inf. Fusion* 4, 87–100 (2003).
- [32] K Tumer and N C Oza. Input decimated ensembles. *Pattern Anal. Appl.* 6, 65–77 (2003).
- [33] K Vale, F Dias, A Canuto and M Souto. A Class-Based Feature Selection Method for Ensemble Systems. In: *8th International Conference on Hybrid Intelligent Systems*, 2008, Barcelona. *IEEE proceedings of HIS - International Conference on Hybrid Intelligent Systems*, 2008. v. 1. p. 596-601.
- [34] K Woods, W Kegelmeyer and K Bowyer. Combination of Multiple Classifiers using Local Accuracy estimates, *IEEE Trans on Patt Analysis and Mach Intelligence*, 19(4), 405-410 (1997).
- [35] E. J. Yeoh et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell.*, 1(2):133–143 (2002).
- [36] T Zhang. On the Consistency of Feature Selection using Greedy Least Squares Regression. *Journal of Machine Learning Research*, 10, 555-568, 2009.

## Author Biographies

**Anne M P Canuto** received the BS degree from the Federal University of Rio Grande do Norte, Brazil, in 1992, the MSc degree from the Federal University of Pernambuco, Brazil, in 1995, and the PhD degree from the University of Kent, in 2001. Currently, she is an associate professor in the informatics and applied mathematics department, Federal University of Rio Grande do Norte. She has published over 50 articles in scientific journals and conferences. Her interests include pattern recognition, classifier combination, and multi-agent systems.

**Karliane M O Vale** received the B.S. degree from the Universidade Potiguar, Brazil, in 1999 and the M.Sc. degree from the Federal University of Rio Grande do Norte, Brazil, in 2009. Currently, she is an associate professor in the Federal Rural University of Semi-Árido, Brazil. Her interests include pattern recognition, classifier combination and feature selection.

**Antonino A Feitosa Neto** received the B.S. degree from the Federal University of Rio Grande do Norte, Brazil, in 2010. Currently, he is a M.Sc. student at the Federal University of Rio Grande do Norte, Brazil. His interests include pattern recognition, classifier combination and optimization techniques.

**Table 2:** Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using no feature selection

GAUSSIAN3						
	Ind	MV	Sum	W-Sum	MLP	NB
HO	81.94±8.41	81.66±10.85	81.66±10.85	68.75±12.96	88.75±6.17	83.75±3.07
HE	81.83±8.51	89.44±7.40	84.80±9.78	87.70±7.91	98.40±2.76	97.79±2.72
SIMULATED6						
HO	84.65±10.96	86.66±10.07	87.50±11.14	85.83±11.45	92.91±5.22	92.50±4.78
HE	82.45±11.60	89.51±10.47	89.84±10.77	88.89±10.60	97.96±4.13	96.56±6.31
ST. JUDE LEUKEMIA						
HO	95.15±3.51	96.67±3.19	96.67±3.15	96.47±3.49	96.87±3.13	96.77±3.14
HE	95.14±3.57	97.83±3.07	97.84±3.00	97.92±2.33	98.61±2.75	98.25±2.85
SPLICE						
HO	85.92±1.97	86.93±1.99	87.01±1.90	87.26±1.88	88.36±1.76	86.64±1.92
HE	86.25±1.97	91.02±1.82	92.73±1.86	93.11±1.82	94.50±1.70	93.21±1.91
PROTEINS						
HO	76.13±4.24	77.44±3.83	78.20±3.77	77.48±3.69	81.21±4.08	79.92±3.80
HE	75.83±4.25	80.18±4.23	81.88±3.79	80.51±3.75	83.91±3.48	83.04±3.95
IMAGE						
HO	92.51±1.43	92.44±1.31	93.12±1.39	92.93±1.53	96.16±0.99	95.28±1.14
HE	92.38±1.43	95.11±1.19	95.49±1.14	95.64±1.18	97.42±0.84	96.29±1.10

**Table 3:** Accuracy and Standard Deviation of the ensemble systems using feature selection methods for the artificial datasets. The criterion used for ranking the attributes is Spearman

	Spearman Gaussian3				Spearman Simulated			
	Fixed 1 (F1) - 100 attributes				Fixed 1 (F1) - 50 attributes			
	Original		ReinSel		Original		ReinSel	
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	60.27±12.86	60.41±12.95	62.08±16.09	62.11±16.19	39.72±13.47	39.69±13.52	55.41±13.60	55.36±13.76
MV	77.08±15.77	72.33±15.95	71.25±15.33	72.72±17.11	41.66±15.65	42.12±15.32	54.16±14.40	59.24±15.19
Sum	84.16±12.73	79.16±16.18	85.00±13.05	82.90±15.09	55.00±19.36	57.94±17.82	69.16±14.34	73.29±15.46
W-Sum	86.66±10.45	86.78±13.12	87.08±8.16	89.90±11.10	64.16±15.66	66.39±15.45	68.33±10.68	68.66±12.53
MLP	90.41±6.34	90.74±8.27	90.41±6.77	95.97±6.03	91.66±6.50	93.15±7.05	93.75±7.11	94.57±7.48
NB	83.75±6.17	93.10±7.65	87.08±8.56	97.32±3.58	88.74±8.37	89.33±9.96	89.16±9.12	91.97±9.11
	Fixed 2 (F2) - 200 attributes				Fixed 2 (F2) - 100 attributes			
	Original		ReinSel		Original		ReinSel	
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	63.33±14.35	63.43±14.44	86.66±11.14	86.66±11.16	56.11±14.87	55.94±14.94	71.87±12.77	71.53±12.77
MV	73.33±14.23	72.38±14.47	88.33±7.98	94.37±8.20	69.16±15.03	69.40±16.33	81.24±9.17	81.00±9.53
Sum	86.66±8.98	82.38±13.58	89.58±4.83	93.22±8.81	84.58±13.29	84.64±13.13	86.66±9.17	86.59±9.34
W-Sum	87.08±7.94	87.00±11.34	90.00±6.14	95.84±6.72	89.16±12.42	85.43±12.65	83.75±7.46	83.61±8.54
MLP	90.41±6.80	95.66±4.49	90.83±5.46	98.62±3.87	94.16±6.46	95.58±3.39	97.91±2.19	96.59±2.91
NB	84.58±5.29	92.62±5.64	85.00±6.14	98.12±4.59	91.24±7.19	94.78±5.32	95.41±4.71	95.76±3.50
	Variable				Variable			
	Original		ReinSel		Original		ReinSel	
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	60.00±15.18	60.01±15.21	64.44±14.71	64.59±14.84	35.76±12.92	35.74±12.80	41.38±13.24	41.25±13.21
MV	73.75±15.73	71.19±16.19	77.50±14.75	76.31±16.19	34.58±13.81	35.79±13.77	38.33±13.02	39.86±12.19
Sum	87.08±10.85	82.61±14.30	87.08±10.04	80.21±14.66	57.50±17.02	56.97±16.54	44.58±17.11	44.21±17.90
W-Sum	87.91±8.89	88.43±10.43	84.16±11.19	84.53±12.68	74.58±12.02	67.04±15.01	49.16±12.18	50.61±11.69
MLP	90.41±4.41	96.63±4.24	90.41±5.57	97.85±2.44	91.25±11.24	93.71±7.48	95.00±7.02	94.66±6.97
NB	87.91±7.32	94.85±5.23	88.75±6.58	95.07±5.29	90.00±7.16	88.95±9.75	92.50±8.71	89.80±8.94

**Table 4:** The average number of noisy attributes chosen by the features selection methods, for the Simulated dataset, for all three subset sizes, F1(50), F2(100) and variable

	Spearman		
	F1	F2	variable
Original	0.33	46	2.66
ReinSel	0	0.16	0

**Table 5:** Accuracy and Standard Deviation of the ensemble systems using feature selection methods for the real datasets. The criterion used for ranking the attributes is Spearman

	Spearman - SPLICE 20 ATRIBUTOS				Spearman -ST. JUDE 80 ATRIBUTOS			
	Original		ReinSel		Original		ReinSel	
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	73.38±2.49	73.63±2.48	82.94±2.08	83.14±2.07	89.42±5.42	89.41±5.40	92.84±4.52	92.70±4.59
MV	84.41±2.16	85.48±2.11	89.38±1.88	90.18±1.94	96.05±3.66	96.75±3.54	96.05±3.97	96.86±3.37
Sum	85.72±2.23	85.77±2.12	89.51±1.79	91.64±1.73	96.36±3.18	97.18±3.30	96.55±3.49	96.97±3.25
W-Sum	85.76±2.03	87.06±2.01	89.26±1.58	91.92±1.72	95.75±3.17	96.45±3.23	96.05±3.62	96.72±3.22
MLP	89.76±1.74	91.09±1.84	91.95±1.66	94.00±1.59	95.75±4.54	97.06±3.40	97.06±3.33	98.18±2.73
NB	89.03±2.08	90.61±1.82	91.19±1.78	93.44±1.62	95.05±3.66	95.12±3.21	96.07±4.02	97.77±3.13
	SPLICE - Fixed 2 (F2) - 20 Attributes				ST. JUDE - Fixed 2 (F2) - 160 Attributes			
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	73.62±2.67	73.97±2.66	87.96±1.79	88.29±1.79	90.09±4.72	90.10±4.72	94.56±4.10	94.43±4.15
MV	82.71±2.02	84.22±2.05	88.84±1.69	92.49±1.68	95.34±4.13	96.21±3.34	96.24±3.58	97.61±3.05
Sum	83.06±2.36	84.97±1.97	89.24±1.90	94.24±1.68	95.55±3.34	96.41±3.14	96.34±3.57	97.64±3.00
W-Sum	83.94±2.40	86.04±1.98	87.77±1.72	94.22±1.70	94.24±3.11	95.12±3.15	96.55±3.59	97.71±2.91
MLP	87.37±1.94	89.63±1.97	90.28±1.64	95.33±1.48	95.55±4.64	97.92±2.54	97.27±3.17	98.73±2.25
NB	86.37±2.16	89.34±1.95	89.49±1.77	94.91±1.62	94.85±3.42	95.99±2.88	96.86±3.42	98.16±2.96
	SPLICE - Variable				ST. JUDE - Variable			
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	79.58±2.43	79.81±2.42	83.62±2.62	83.65±2.62	90.24±4.75	90.25±4.76	92.39±4.94	92.24±5.01
MV	84.87±2.13	88.69±2.23	88.64±2.76	89.24±2.72	94.44±3.99	95.52±3.59	95.34±4.09	96.42±3.81
Sum	84.09±1.87	89.46±1.99	89.78±2.67	90.19±2.65	94.45±3.67	95.88±3.46	96.55±3.71	97.00±3.36
W-Sum	84.27±1.97	90.40±1.88	89.99±2.49	90.55±2.57	93.73±3.55	94.25±3.60	96.46±3.60	96.89±3.24
MLP	88.64±1.73	93.09±1.81	93.78±2.17	94.00±2.15	95.27±5.17	97.62±3.02	97.16±3.13	98.17±2.92
NB	87.31±1.76	91.76±1.94	92.32±2.43	92.67±2.40	95.26±3.85	96.42±2.86	96.25±4.07	97.32±3.72
	Proteins - Fixed 1 (F1) - 25 Attributes				Image - Fixed 1 (F1) - 3 Attributes			
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	69.30±5.15	69.15±5.19	77.75±4.70	77.79±4.67	76.39±2.45	76.13±2.47	83.63±2.09	83.57±2.09
MV	76.37±4.24	76.94±4.64	81.04±4.72	81.92±4.68	86.43±2.28	87.57±2.32	90.02±1.70	91.94±1.55
Sum	79.58±2.98	79.57±3.90	81.46±3.88	82.79±4.15	92.41±1.64	92.79±1.62	92.28±1.47	93.86±1.34
W-Sum	77.23±3.64	77.80±3.81	80.14±4.07	81.77±4.19	86.66±1.91	87.48±1.96	90.55±1.73	93.16±1.53
MLP	82.80±3.67	83.16±3.70	83.56±4.26	84.64±4.02	95.55±1.21	96.99±1.09	96.65±1.16	97.27±0.88
NB	81.81±3.32	81.79±3.89	82.71±3.96	83.38±4.14	94.78±1.44	94.68±1.13	95.61±0.93	95.87±0.88
	Proteins - Fixed 2 (F2) - 50 Attributes				Image - Fixed 2 (F2) - 6 Attributes			
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	70.43±4.96	70.41±4.84	78.20±4.41	78.27±4.33	82.35±2.00	82.25±2.01	87.91±1.85	87.85±1.87
MV	75.33±4.96	77.16±4.91	79.96±4.19	81.83±4.33	87.14±1.77	89.48±1.81	90.27±1.47	92.63±1.62
Sum	77.18±4.13	79.00±4.17	80.69±3.91	82.51±3.73	90.60±1.90	92.43±1.44	91.61±1.55	93.84±1.33
W-Sum	75.76±3.90	77.55±4.12	80.17±4.16	81.48±3.77	86.52±1.71	90.10±1.68	91.35±1.64	94.00±1.30
MLP	81.76±3.62	82.52±3.49	82.79±3.51	84.98±3.63	94.12±1.25	96.62±0.88	96.79±0.79	97.50±0.77
NB	78.89±4.11	80.31±3.70	82.02±3.87	83.12±3.89	93.99±1.03	94.59±1.05	95.28±0.99	95.97±0.76
	Proteins - Variable				Image - Variable			
	HO	HE	HO	HE	HO	HE	HO	HE
Ind.	74.93±4.92	74.84±4.88	76.34±4.87	76.38±4.85	60.64±2.16	60.47±2.19	69.55±2.08	69.47±2.09
MV	76.79±4.60	78.15±5.00	77.73±4.48	80.37±4.75	80.77±2.22	81.83±2.43	85.88±1.73	87.55±1.54
Sum	78.16±4.95	79.87±4.69	78.42±4.10	81.33±4.34	91.39±1.61	92.54±1.61	88.74±1.53	90.64±1.41
W-Sum	77.78±4.42	78.94±4.60	78.80±4.22	80.68±4.21	91.48±1.62	92.64±1.60	86.75±1.54	89.80±1.31
MLP	80.44±4.60	82.60±3.92	80.52±4.14	82.97±4.09	93.89±1.45	94.41±1.17	91.99±1.47	92.72±1.09
NB	78.55±5.34	81.07±5.33	79.49±4.68	81.59±4.76	91.90±1.42	92.27±1.46	90.16±1.31	91.00±1.31