

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/310896571>

Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches

Article in Knowledge-Based Systems · November 2016

DOI: 10.1016/j.knosys.2016.11.017

CITATIONS

93

READS

1,763

4 authors, including:



Borja Seijo-Pardo

University of A Coruña

9 PUBLICATIONS 188 CITATIONS

[SEE PROFILE](#)



Verónica Bolón-Canedo

University of A Coruña

114 PUBLICATIONS 2,782 CITATIONS

[SEE PROFILE](#)



Amparo Alonso-Betanzos

University of A Coruña

254 PUBLICATIONS 4,383 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



SMARTEES - local social innovation [View project](#)



SMARTEES: Social Innovation Modelling Approaches to Realizing Transition to Energy Efficiency and Sustainability [View project](#)

Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches

B. Seijo-Pardo*, I. Porto-Díaz*, V. Bolón-Canedo*, A. Alonso-Betanzos*

*Department of Computer Science - University of A Coruña
Campus de Elviña s/n 15071 - A Coruña, Spain*

Abstract

In the last decade, ensemble learning has become a prolific discipline in pattern recognition, based on the assumption that the combination of the output of several models obtains better results than the output of any individual model. On the basis that the same principle can be applied to feature selection, we describe two approaches: (i) homogeneous, i.e., using the same feature selection method with different training data and distributing the dataset over several nodes; and (ii) heterogeneous, i.e., using different feature selection methods with the same training data. Both approaches are based on combining rankings of features that contain all the ordered features. The results of the base selectors are combined using different combination methods, also called aggregators, and a practical subset is selected according to several different threshold values (traditional values based on fixed percentages, and more novel automatic methods based on data complexity measures). In testing using a Support Vector Machine as a classifier, ensemble results for seven datasets demonstrate performance that is at least comparable and often better than the performance of individual feature selection methods.

Keywords: Ensemble Learning, Feature Selection, Ranking Aggregation, Classification, SVM-Rank, Data Complexity Measures

1. Introduction

In recent years the datasets used in machine learning have increased considerably in size. Because feature selection can eliminate irrelevant and redundant information it has begun to play a very important role in tackling large datasets, with the added benefits of saving on storage and improving computation time for machine learning algorithms.

*Corresponding author

Email addresses: `borja.seijo@udc.es` (B. Seijo-Pardo), `iporto@udc.es` (I. Porto-Díaz), `vbolon@udc.es` (V. Bolón-Canedo), `ciamparo@udc.es` (A. Alonso-Betanzos)

In the past, machine learning methods used a single learning model. However, it has been observed that multiple prediction models can be used to solve the same problem. This approach, called *ensemble learning*, builds on the assumption that combining the output of multiple experts is better than using the output of any single expert [1, 2]. Ensemble learning has been successfully applied to classification problems, but is also a means for improving other machine learning functions such as feature selection.

In feature selection, the individual selectors in an ensemble are known as *base selectors*. If the base selectors are all of the same kind, the ensemble is known as *homogeneous*; otherwise the ensemble is *heterogeneous*. Ensembles can be formed in several ways [3]. In this paper we explore two in particular: (i) N selections using the same feature selection algorithm, with each selection using different training data; and (ii) N selections using different feature selection algorithms that use the same training data. The first approach improves computation time by processing data in parallel nodes, whereas the second approach ensures stable and robust feature selection that achieves competitive results irrespective of the scenario.

There are two different approaches [4] to evaluating the features of a dataset: *individual evaluation* and *subset evaluation*. In individual evaluation, a ranking of features is returned by assigning a level of relevance to each feature. In subset evaluation, successive subsets of features—generated according to a predefined search strategy—are evaluated iteratively according to an optimality criterion until a final subset of selected features is obtained. The individual evaluation approach is incapable of removing redundant features because these are likely to have similar rankings. In contrast, the subset evaluation approach can handle both feature redundancy and feature relevance but may be computationally inefficient because of having to search through all the feature subsets. To avoid computational problems, in our research for this article, we focused on the individual evaluation approach to feature selection.

Feature selection algorithms—and therefore base selectors—can be classified in three different categories according to structure, namely, as wrappers, filters, or embedded algorithms [5]. Wrappers work hand in hand with a classification algorithm, in such a way that error in each iteration by the classifier yields a score for the wrapper evaluation function. Filters, which require fewer computational resources than wrappers, use a fast evaluation function and are independent of the classifier. Lastly, embedded methods can be viewed as in an intermediate position between wrappers and filters. They work with a classifier, but they do not rely on error (like wrappers), but on some intermediate result, e.g., weights. They are computationally more costly than filters, but less so than wrappers.

In our research we used an ensemble of filters and embedded methods rather than the typical single method. The goal was to introduce diversity and increase the regularity of the feature selection process, so as to take advantage of the strengths of the individual selectors and overcome their weaknesses. According to how the data was distributed and which feature selection approaches were applied, we used two approaches that we tested using a *Support Vector Machine*

(*SVM*) [6] classifier. Experimental validation of the methodology using seven different datasets shows the suitability of the proposed ensembles. In a real-world scenario, the benefit is that the user does not have to decide which feature selection algorithm might be the most appropriate for a given problem.

The remainder of this article is organized as follows. Section 2 describes and discusses previous related works. Section 3 describes the proposed method and its algorithm, along with individual rankers and the *SVM-Rank* [7] method used to combine individual rankers. Section 4 describes the datasets, experimental design, and experimental results. Section 5 discusses the results, and finally, Section 6 contains our concluding remarks and proposals for future research.

2. Background

Feature selection has been applied to many machine learning and data mining problems, with the aim of selecting a subset of features that minimizes the prediction error of a given classifier [8]. Guyon and Elisseeff [8] and Hall and Holmes [9] describe different approaches to feature selection, including feature construction, feature ranking, and multivariate feature selection, and also describe efficient search methods and feature validity assessment methods.

It has been observed that better results could be obtained by combining different machine learning methods so as to solve the same problem, an approach called *ensemble learning*. Since combining classifiers would intuitively appear to be a natural step forward once a critical mass of knowledge regarding individual classifiers has been accumulated, this approach has rapidly attracted the interest of the pattern recognition and machine learning communities [2].

The most widely used ensemble learning methods applied to classification are *bagging* [10] and *boosting* [11], due to their theoretical performance guarantees and strong experimental results. Bagging creates an ensemble by training individual classifiers on bootstrapped samples from the training set, generated by random selection with replacement of n instances, where n is the size of the training set. As a result, each classifier is trained on an average of 63.2% of training instances, and the classifiers' predictions are combined using simple voting. In the boosting approach, sampling is proportional to the weight of an instance.

Recent works have proposed improving feature selection algorithm robustness using multiple feature selection evaluation criteria. Of several studies performed in order to achieve more accurate classification, one [12] conducted on 21 UCI datasets [13] analyzed and compared five measures of diversity for possible use in ensemble feature selection; four search strategies for ensemble feature selection were considered with simple random subspacing, namely, genetic search, hill-climbing, and ensemble forward and backward sequential selection. Based on the idea of multiple feature selection evaluation criteria, many feature selection ensembles have been proposed. In one study, five different filters were employed, each of which selected a different subset of features to train and test five classifiers whose outputs were combined by simple voting [14]. In another study,

the *Multicriterion Fusion-based Recursive Feature Elimination (MCF-RFE)* algorithm was developed with the goal of improving both classification performance and the stability of the feature selection results [15]. Yet another study proposed a feature ranking scheme for *Multilayer Perceptron (MLP)* ensembles [16], used with a stopping criterion based on the *Out-of-Bootstrap (OOB)* estimate [17]; the versatility of this base classifier in removing irrelevant features was demonstrated experimentally using benchmark data. In a scenario as challenging as DNA microarray classification, an ensemble of filters rather than a single filter was used with both synthetic and real data [18].

Finally, in a number of works all the feature selection methods in the final ensemble were ranker methods. Diversity can be achieved by combining various rankers in order to yield more stable and robust results. Three commonly used filter-based feature ranking techniques for text classification problems were used by Olsson and Oard [19], where the combining methods used were lowest, highest, and average rank. Outstanding work in this area has been done in two especially interesting studies by Wang et al. [20] [21], the first one examining ensembles of six commonly used filter-based rankers, and the second one studying 17 ensembles of feature ranking techniques with six commonly-used rankers, a signal-to-noise filter technique (*S2N*) [22], and 11 threshold-based rankers. In the second of the Wang et al. studies cited above [21], the ensembles were composed of two to 18 individual feature selection methods. Other studies describe different methods for combining individually generated rankings, with the aim of obtaining a final ensemble. The combination of individual rankings covers from simple methods—based on computing the mean, median, minimum, etc.—to more complex methods like *Complete Linear Aggregation* [23] (*CLA*), *Robust Ensemble Feature Selection (Rob-EFS)* [24], and *SVM-Rank* [25].

As stated before, the variability of results for different datasets is one of the problems of choosing a feature selection technique. The aim of our research was to develop a method to reduce the variability induced by using individual feature selection methods and so take advantage of their strengths and overcome their weaknesses. We combined the many outputs of as many individual rankers were implemented using a combination method. We then tested different combination methods to obtain a final ranking of features, from which we extracted a smaller subset of features by applying a threshold. Finally, we used this subset to train a *Support Vector Machine with Radial-Basis-Function (SVM-RBF)* [26] in order to test the proposed ensembles for classification accuracy.

3. Proposed Method

One of the main problems of applying feature selection is selecting an appropriate method for a given problem. Each feature selection method has its strengths and weaknesses, and performance depends on the dataset type, but also on restrictions related to the scenario (accuracy, time, cost, etc). However, despite the availability of growing numbers of methods, researchers generally agree that no ideal feature selection method exists. Some knowledge of existing algorithms is generally required in order to be able to choose a method that is

appropriate to the problem. One possible solution to this problem is to use an ensemble of methods.

3.1. Ensemble Approaches

Although there are several ways to design an ensemble [27], we focus on two in particular:

- *Homogeneous distributed ensemble.* N models are generated using the same feature selection method but different training data (Figure 1). An important issue with ensemble methods is the computation time required in comparison to individual methods. One solution is to parallelize the training task by distributing the training data among a number of nodes. The same method is then executed on each node, and the rankings obtained are combined using a combination method. In our research the number of nodes was set to $N = 10$. The pseudo-code for this approach is given in Algorithm 1.

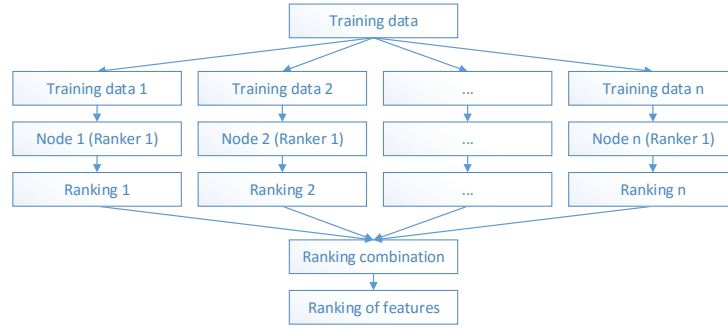


Figure 1: Homogeneous distributed ensemble.

- *Heterogeneous centralized ensemble.* N models are generated using different feature selection methods, but the same training data (Figure 2). This approach takes account of the strengths and weaknesses of the individual methods. The several different methods are trained using the same training data, and the output is then combined using a combination method. The pseudo-code of this approach is given in Algorithm 2.

In both approaches, the A_n outputs obtained were combined using a ranking combination method to obtain a single ranking. Since the individual methods

Algorithm 1: Pseudo-code of the homogeneous distributed ensemble

Data: N — number of different nodes

Data: T — threshold of the number of features to be selected

Result: P — classification prediction

- 1 Separate the training data in the N nodes.
 - 2 **for** each n from 1 to N **do**
 - 3 Obtaining ranking A_n using the same feature selection method on the node n
 - 4 **end**
 - 5 A = combining rankings A_n with a ranking combination method
 - 6 A_t = Select T top attributes from A
 - 7 Build classifier SVM-RBF with the selected attributes A_t
 - 8 Obtain prediction P
-

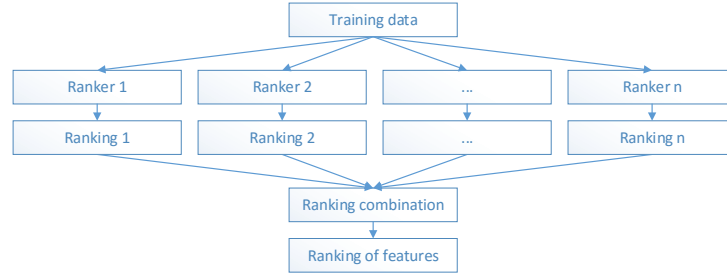


Figure 2: Heterogeneous centralized ensemble.

Algorithm 2: Pseudo-code of the heterogeneous centralized ensemble

Data: N — number of ranker methods

Data: T — threshold of the number of features to be selected

Result: P — classification prediction

- 1 **for** each n from 1 to N **do**
 - 2 Obtaining ranking A_n using feature selection method n
 - 3 **end**
 - 4 A = combining rankings A_n with a ranking combination method
 - 5 A_t = Select T top attributes from A
 - 6 Build classifier SVM-RBF with the selected attributes A_t
 - 7 Obtain prediction P
-

used for feature selection are rankers, it was necessary to establish a threshold T in order to obtain a practical subset of features, A_t . An *SVM-RBF* [26] was then used to check the suitability of the proposed ensemble in terms of classification error.

3.2. Feature Selection Methods

Of the broad suite of feature selection methods available, three filters and two embedded methods were chosen for this study:

- **Information Gain (InfoGain)** [28]. This filter is one of the most common univariate methods for evaluating features, which it does according to the information gain, considering a single feature at a time.
- **Minimum Redundancy Maximum Relevance (mRMR)** [29]. This filter uses mutual information to select the most relevant and minimally redundant features for the target class, i.e., it selects features that are maximally dissimilar to each other.
- **ReliefF** [30]. This filter is an extension of the original Relief algorithm [31] that works by randomly sampling an instance from the dataset and then locating its nearest neighbor from the same and opposite class. The values of the nearest neighbor attributes are compared to the sampled instance and used to update relevance scores for each attribute. The rationale is that a useful attribute should differentiate between instances from different classes, and have the same value for instances from the same class. Compared to Relief, ReliefF is more robust, better handles multiclass problems and incomplete and noisy data, can be applied in all situations, has low bias, allows interaction among features, and may capture local dependencies which other methods miss.
- **Recursive Feature Elimination for Support Vector Machines (SVM-RFE)** [32]. This embedded method uses the current set of features to train an *SVM* classifier iteratively by removing the least important features, as indicated by the weights in the *SVM* solution.
- **Feature Selection-Perceptron (FS-P)** [33]. This embedded method is based on a perceptron, a type of artificial neural network that can be seen as the simplest kind of feedforward neural network, namely, a linear classifier. This method is based on training a perceptron in a supervised learning context. Interconnection weights are used to indicate the most relevant features and provide a ranking.

The above five feature selection methods are rankers, which means that they do not select a subset of features, but sort all the features. This particular set of rankers was selected because: (i) they are based on different metrics and so ensure great diversity in the final ensemble; and (ii) they are widely used by feature selection researchers.

We conducted a small diversity study using two datasets (*Spambase* and *Isolet*). Final rankings obtained by the five rankers listed above were compared using Spearman’s rank correlation coefficient [34] (Table 1) and Kendall’s rank correlation coefficient [35] (Table 2). The ρ value in the range $[-1, 1]$ reflects the relationship between rankings, with 1 indicating that the compared rankings were equal.

Dataset	Ranker	InfoGain	mRMR	ReliefF	SVM-RFE	FS-P
Spambase	InfoGain	1.0000	0.2011	0.0714	-0.2040	-0.1736
	mRMR	0.2011	1.0000	-0.0811	0.1313	0.0838
	ReliefF	0.0714	-0.0811	1.0000	-0.0672	0.0380
	SVM-RFE	-0.2040	0.1313	-0.0672	1.0000	0.0565
	FS-P	-0.1736	0.0838	0.0380	0.0565	1.0000
Isolet	InfoGain	1.0000	0.0971	-0.0677	-0.0320	-0.0521
	mRMR	0.0971	1.0000	0.0295	0.0534	0.0062
	ReliefF	-0.0677	0.0295	1.0000	0.0115	-0.0291
	SVM-RFE	-0.0320	0.0534	0.0115	1.0000	0.0331
	FS-P	-0.0521	0.0062	-0.0291	0.0331	1.0000

Table 1: ρ value of Spearman’s rank correlation coefficient.

Dataset	Ranker	InfoGain	mRMR	ReliefF	SVM-RFE	FS-P
Spambase	InfoGain	1.0000	0.1278	0.0476	-0.1466	-0.1266
	mRMR	0.1278	1.0000	-0.0602	0.0940	0.0464
	ReliefF	0.0476	-0.0602	1.0000	-0.0489	0.0288
	SVM-RFE	-0.1466	0.0940	-0.0489	1.0000	0.0351
	FS-P	-0.1266	0.0464	0.0288	0.0351	1.0000
Isolet	InfoGain	1.0000	0.0652	-0.0449	-0.0212	-0.0337
	mRMR	0.0652	1.0000	0.0216	0.0373	0.0053
	ReliefF	-0.0449	0.0216	1.0000	0.0084	-0.0168
	SVM-RFE	-0.0212	0.0373	0.0084	1.0000	0.0213
	FS-P	-0.0337	0.0053	-0.0168	0.0213	1.0000

Table 2: ρ value of Kendall’s rank correlation coefficient.

It can be seen that most of the ρ values are far from 1, indicating great differences between the paired rankings (obviously, when the same ranker method rankings were compared, the ρ value was 1, as can be seen in the table diagonals). This small experiment demonstrated that the set of feature selection rankers chosen for this study ensured enough diversity in their behaviors.

3.3. Ranking Combination

We describe two different kinds of ensembles that obtain a final ranking of features by combining outputs (rankings) from individual rankers using a com-

bination method—also known as an aggregator—to produce a unique final output. Several different combination methods are available, ranging from simple calculation measures—minimum, maximum, mean, etc.—to more sophisticated measures like *SVM-Rank*.

3.3.1. SVM-Rank

SVM-Rank [7] is an *SVM*-based method that can be trained to learn ranking functions. The *SVM-Rank* algorithm considers a training set S of size n containing feature selection methods q with their rankings r according to (1):

$$(q_1, r_1), (q_2, r_2), \dots, (q_n, r_n) \quad (1)$$

The algorithm selects a ranking function f that maximizes (2):

$$\tau_S(f) = \frac{1}{n} \sum_{i=1}^n \tau(r_{f(q_i)}, r_i) \quad (2)$$

The function f must maximize (2) and must generalize beyond the training data. Consider the class of linear ranking functions (3) defined as:

$$(c_i, c_j) \in f_{\vec{w}}(q) \Leftrightarrow \vec{w} \Phi(q, c_i) > \vec{w} \Phi(q, c_j), \quad (3)$$

where \vec{w} is a weight vector that is adjusted by learning. $\Phi(q, c)$ is a mapping between method q and feature c . For any weight vector \vec{w} , the points are ordered by their projection onto \vec{w} . Maximizing (2) is equivalent to finding the weight vector so that the maximum number of the following inequalities is satisfied (4):

$$\forall (c_i, c_j) \in r_k : \vec{w} \Phi(q_k, c_i) > \vec{w} \Phi(q_k, c_j) \mid k = 1, \dots, n \quad (4)$$

The solution to this problem is approximated, analogously to *SVM* classification, by introducing slack variables $\xi_{i,j,k}$ and minimizing the upper bound $\sum_{i,j,k}$. This renders the problem equivalent to an *SVM* classification on pairwise difference vectors $\Phi(q_k, c_i) - \Phi(q_k, c_j)$.

3.3.2. Other Combination Methods

We used the other combination methods listed in Table 3, all belonging to the *RobustRankAggreg* package implemented in the R and Matlab languages [36]. These combination methods consider a set of n feature selection ranker methods, where $\mathcal{Q} = \{q_i, i = 1, \dots, n\}$, and where each q_i is associated with a list of m objects that represent the relevance of features in the range $[0, 1]$. Once the relevance of each feature in the individual ranking method is obtained, one of the reduction functions shown in Table 3 is applied. The result is a reduced final ranking that is ordered according to the calculated relevance factor. The final relevance values of the features will be in the range $[0, 1]$, where higher and lower values reflect more important and less important features in the dataset, respectively.

Function	Formula	Description
min	$\min\{q_1(d_j), q_2(d_j) \dots q_n(d_j)\}$	Reduction function based on simple arithmetic operations. It selects the minimum of the relevance values yielded by the rankings [37].
median	$\text{median}\{q_1(d_j), q_2(d_j) \dots q_n(d_j)\}$	Reduction function based on simple arithmetic operations. It selects the median of the relevance values yielded by the rankings [37].
mean	$\frac{1}{n} \sum_{i=1}^n q_i(d_j)$	Reduction function based on simple arithmetic operations. It selects the average of the relevance values yielded by the rankings [37].
geomMean	$(\prod_{i=1}^n q_i(d_j))^{1/n}$	Reduction function based on simple arithmetic operations. It selects the geometric average of the relevance values yielded by the rankings [37].
Stuart	$Pq[X \leq \rho] = 1 - Pq[\hat{q}_1 \leq 1 - \mathcal{B}_{n,1}^{-1}(\rho), \dots, \hat{q}_n \leq 1 - \mathcal{B}_{n,1}^{-1}(\rho)]$	Reduction function based on statistical sorting distributions. It uses the <i>Beta</i> distribution to obtain the ρ value [38].
RRA	$\min_{i=1, \dots, n} \mathcal{B}_{k,n}(r), \mathcal{B}_{k,n}(q) = Pr[\hat{q}_k \leq q_k]$	Reduction function based on statistical sorting distributions. Based on the <i>Stuart</i> function, it improves the efficiency-accuracy connection through the use of <i>Bonferroni</i> correction when calculating the ρ value [36].

Table 3: Reduction functions for feature rankings.

3.4. Threshold Values

Since the feature selection methods sort all the features, it was necessary to set a threshold in order to obtain a practical subset of features. Most works in the literature use several thresholds that retain different percentages of features [39, 40]. Since thresholds are dependent on the particular dataset being studied, several attempts have been made to derive a general automatic threshold [41, 42]. For our research we used five different threshold values to delimit data dimensionality, two of which were automatic thresholds: one based on a data complexity measure, the *Fisher discriminant ratio* [43], that was developed by us in previous research [44, 45], and the $\log_2(n)$ threshold. The five different threshold values used were as follows:

- ***Fisher discriminant ratio* (F)** [43]. This is defined for a multidimensional problem as:

$$F = \frac{\sum_{i=1, j=1, i \neq j}^c p_i p_j (\mu_i - \mu_j)^2}{\sum_{i=1}^c p_i \sigma_i^2}, \quad (5)$$

where μ_i , σ_i^2 , and p_i are the mean, variance, and proportion of the i th class c , respectively. *Fisher discriminant ratio* values are calculated individually for each feature of the dataset. In practice, it is preferable to use the *Fisher*

discriminant ratio inverse ($1/F$) to establish the threshold and obtain the final subset, as the smaller value renders the problem more tractable. Therefore, the final formula for calculating the complexity value e of each feature is defined as:

$$e = \alpha \times 1/F + (1 - \alpha) \times \rho \quad (6)$$

where α , with a value in the interval $[0, 1]$, is a parameter that controls both the relative emphasis on the number of features retained and the weight given to the complexity measure (empirically established as $\alpha = 0.75$ for this research), ρ is the percentage of features retained (ranging from 1 to the total number of features in the dataset) and $1/F$ is the inverse of the *Fisher discriminant ratio*.

- **$\log_2(n)$.** This threshold, where n is the number of features in a given dataset, was chosen for our research, given that it has been recommended to select $\log_2(n)$ metrics for software quality prediction [21], and that it has been shown to be suitable for use with Weka [46] in building random forest learners [41].
- **10%.** This threshold selects the top 10% of the most relevant features of the final ordered ranking.
- **25%.** This threshold selects the top 25% of the most relevant features of the final ordered ranking.
- **50%.** This threshold selects the top 50% of the most relevant features of the final ordered ranking.

3.5. Classification Method

Of the classification methods available in the literature, we chose to use an *SVM* [5] algorithm for this study. The *SVM*, a classification method that has received much attention in recent years [47, 48], is based on the idea of *Structural Risk Minimization (SRM)* [49]. *SVM* methods have performed extremely well in a wide range of applications—even comparably with techniques like neural networks [47]—and have been used as powerful tools in solving classification problems.

The *SVM* algorithm learns the decision surface of two distinct classes of input points. From the description given by the support vectors of the data, it is able to form a decision border around the training data domain with little or no knowledge of the data outside this border. First, the data are mapped—through a Gaussian or other type of kernel—to a feature space in a larger dimension space (for instance, if the input points are in R^2 then they are mapped to R^3), with the aim being maximum separation between classes. The algorithm then locates a separating hyperplane that maximizes the margin m between classes in this space, as shown in Figure 3. This border function can also separate the two data types back into the input space. Maximizing the margin m is a quadratic

programming (QP) problem that can be solved through its dual problem by introducing Lagrange multipliers. The solution for the optimal hyperplane can be written as a combination of a few input points (i.e., the support vectors).

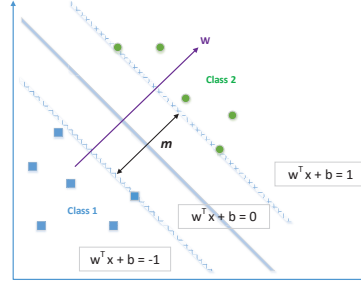


Figure 3: Border decisions in SVM classification methods.

4. Experimental Study

The proposed ensembles were tested on seven different datasets (Table 4), conforming an interesting suite against which to check suitability. The number of samples and features ranged from 1 484 to 67 557 and from 8 to 10 000, respectively, and the datasets represented both binary and multiclass problems.

Dataset	Samples	Features	Classes	Download
Yeast	1 484	8	10	UCI repository [13]
Spambase	4 601	57	2	UCI repository [13]
Madelon	2 400	500	2	UCI repository [13]
Connect4	67 557	42	3	UCI repository [13]
Isolet	7 797	617	26	UCI repository [13]
USPS	9 298	256	10	FS repository [50]
Pixraw10P	100	10 000	10	FS repository [50]

Table 4: Datasets used in the experimental study.

4.1. Experimental Procedure

The experimental procedure we applied was as follows:

1. We partitioned each of the seven datasets (Table 4) according to a ten-fold cross-validation scheme.

2. We applied the feature selection process as indicated by the ensemble approach used (see Section 3.1).
3. We combined the individual rankings using different combination methods so as to obtain a final ranking (see Section 3.3).
4. We obtained a practical subset of features according to different thresholds (see Section 3.4).
5. We checked the suitability of the ensemble approach against the individual methods, using an *SVM* as a classifier to measure test error (see Section 3.5). The *SVM* classifier used a Gaussian *Radial Basis Function (RBF)* with values $C = 1$ and $\gamma = 0.01$ (the default values for both parameters in Weka).
6. To the ten different results obtained after ten-fold cross-validation, we applied the Kruskal-Wallis test to check if there were significant differences between individual and ensemble strategies (for a significance level of $\alpha = 0.05$) and Tukey’s multiple comparison procedure [51] to identify results that were not significantly worse than the best individual result.

4.2. Results for the Homogeneous Distributed Ensemble

This experiment consisted of comparing the individual and homogeneous distributed ensemble approaches using the same rankers, different ranking combination methods, and different training data. Recall that the aim with the homogeneous distributed ensemble was to reduce training time while maintaining classification accuracy. Average training times, average test errors and standard deviations are shown in the tables that follow.

Table 5 shows the average training times in seconds for the five feature selection methods applied to the seven datasets. Individual strategies whose average times were not significantly worse than those of the ensemble strategies using the same feature selection method are labeled with a superscript dagger (see Section 4.1). As one example, for the *SVM-RFE* method applied to the *Yeast* dataset, the performance of the individual strategy was not significantly worse than that of the ensemble strategy in terms of average training time (the individual and ensemble results are labeled with a superscript dagger), whereas, in contrast, for this method applied to the *Madelon* dataset, the individual strategy was significantly worse than the ensemble strategy (only the ensemble result is labeled with a superscript dagger). Note that the time spent on the ranking combination can be considered negligible. Overall it can be seen how the ensemble strategy considerably improved training times.

The next five tables (Tables 6, 7, 8, 9, and 10) show average test errors and standard deviations for the five thresholds. The ensemble approaches whose average test results were not significantly worse than for the individual strategy using the same feature selection method are labeled with a superscript dagger (see Section 4.1).

It can be seen that test error rates for individual and ensemble strategies were comparable. Table 6 shows that, for 215 of 245 experiments performed with the *Fisher discriminant ratio* threshold, the average test errors for the ensemble strategy were not significantly worse than those for the individual methods;

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	Ensem.	$0.01^{\dagger} \pm 0.01$	$0.01^{\dagger} \pm 0.01$	$0.10^{\dagger} \pm 0.25$	$0.02^{\dagger} \pm 0.01$	$0.06^{\dagger} \pm 0.01$	$0.03^{\dagger} \pm 0.01$	$0.18^{\dagger} \pm 0.30$
	Indiv.	0.01 ± 0.01	$0.02^{\dagger} \pm 0.02$	$0.02^{\dagger} \pm 0.01$	0.05 ± 0.01	0.18 ± 0.01	0.05 ± 0.01	$0.08^{\dagger} \pm 0.01$
mRMR	Ensem.	$0.01^{\dagger} \pm 0.00$	$8.67^{\dagger} \pm 2.58$	$218.35^{\dagger} \pm 3.91$	$0.23^{\dagger} \pm 0.01$	$25.49^{\dagger} \pm 0.14$	$3.06^{\dagger} \pm 0.21$	$737.83^{\dagger} \pm 30.31$
	Indiv.	0.01 ± 0.01	13.54 ± 3.89	510.90 ± 25.37	1.38 ± 0.07	59.64 ± 0.38	25.86 ± 3.02	1265.63 ± 87.41
ReliefF	Ensem.	$0.01^{\dagger} \pm 0.01$	$0.01^{\dagger} \pm 0.00$	$0.12^{\dagger} \pm 0.23$	$0.48^{\dagger} \pm 0.03$	$0.20^{\dagger} \pm 0.01$	$0.13^{\dagger} \pm 0.07$	$0.63^{\dagger} \pm 0.21$
	Indiv.	$0.02^{\dagger} \pm 0.01$	0.20 ± 0.04	0.69 ± 0.04	37.57 ± 3.86	8.35 ± 0.44	5.88 ± 0.90	17.27 ± 2.38
SVM-RFE	Ensem.	$0.03^{\dagger} \pm 0.03$	$0.01^{\dagger} \pm 0.01$	$6.51^{\dagger} \pm 13.51$	$7.01^{\dagger} \pm 0.66$	$37.82^{\dagger} \pm 9.57$	$13.5^{\dagger} \pm 0.11$	$32.61^{\dagger} \pm 8.51$
	Indiv.	$0.05^{\dagger} \pm 0.03$	0.12 ± 0.06	1744.28 ± 218.17	691.62 ± 90.16	2662.18 ± 249.78	1082.89 ± 121.71	3167.53 ± 190.01
FS-P	Ensem.	$0.03^{\dagger} \pm 0.02$	$0.06^{\dagger} \pm 0.06$	$0.54^{\dagger} \pm 0.04$	$0.84^{\dagger} \pm 0.17$	$18.63^{\dagger} \pm 0.35$	$1.59^{\dagger} \pm 0.31$	$18.79^{\dagger} \pm 3.14$
	Indiv.	0.30 ± 0.09	0.73 ± 0.12	4.91 ± 0.18	13.60 ± 2.43	179.35 ± 16.19	17.62 ± 0.49	227.91 ± 12.84

Table 5: Average training time in seconds taken by the homogeneous distributed ensemble and the individual rankers. The superscript dagger indicates times for individual rankers that were not significantly worse than for the ensembles.

furthermore, in nine of these experiments the ensemble method achieved significantly better results. Table 7 shows that, for the $\log_2(n)$ threshold, the ensemble methods obtained favorable results in 230 of the 245 experiments. For the three percentage thresholds, results were broadly similar. Thus, for the 10% threshold, in 217 of the 245 experiments the test errors returned by the ensemble were not significantly worse than those returned by the individual method (Table 8), and a further six of these experiments yielded significantly better test error percentages than the individual method. Increasing the percentage of features retained did not imply better results; when the thresholds were 25% (Table 9) and 50% (Table 10), the ensemble strategies obtained favorable results in 218 and in 216 of the 245 experiments, respectively.

To sum up, the ensemble strategy considerably reduced training times compared to the individual approaches, and there were no significant differences in test errors between the two strategies in 1096 of the 1225 experiments performed. In other words, in 89.47% of the cases the performance of the ensemble strategy was not significantly worse than the performance of the individual methods. Note also that, in 30 of the 1225 experiments, the significant differences were even in favor of the homogeneous distributed ensemble.

4.3. Results for the Heterogeneous Centralized Ensemble

This experiment consisted of comparing the individual and ensemble approaches using different ranker and ranking combination methods and the same training data. Recall that the main aim with the heterogeneous centralized ensemble was to maintain or improve classification performance while freeing the user from having to decide on the most appropriate feature selection method for any given situation. Average test errors and standard deviations for the five thresholds are shown in the tables that follow. The algorithms whose average test error results were not significantly worse than the best result are labeled with a superscript dagger (see Section 4.1).

The experimental results demonstrated the suitability of the proposed ensemble, since they matched or improved on the results achieved by the individual feature selection methods. The ensemble errors were not significantly different

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	E-SVMRank	58.22 [†] ±3.95	15.28 [†] ±2.98	34.75 [†] ±3.36	30.27 [†] ±1.73	65.99 ±6.16	54.50 [†] ±2.10	82.00 [†] ±13.77
	E-Min	57.75 [†] ±4.28	14.85 [†] ±2.86	33.46 [†] ±2.84	28.83 [†] ±2.28	56.88 [†] ±7.01	56.69 [†] ±1.76	87.00 [†] ±12.42
	E-Median	58.22 [†] ±3.95	16.65 [†] ±3.55	33.63 [†] ±3.07	30.76 [†] ±0.71	62.13 [†] ±6.58	55.83 [†] ±2.15	81.00 [†] ±15.84
	E-Mean	58.22 [†] ±3.95	15.13 [†] ±2.51	34.75 [†] ±3.51	30.76 [†] ±0.71	65.59 ±4.91	55.11 [†] ±2.18	87.00 [†] ±12.01
	E-GeomMean	58.22 [†] ±3.95	16.61 [†] ±2.60	34.00 [†] ±2.55	30.76 [†] ±0.71	64.15 [†] ±5.61	57.98 [†] ±2.24	84.00 [†] ±15.09
	E-Stuart	58.22 [†] ±3.95	13.71 [†] ±1.99	34.21 [†] ±2.96	30.25 [†] ±1.56	65.82 ±6.44	56.51 [†] ±1.95	88.00 [†] ±15.26
	E-RRR	58.22 [†] ±3.95	14.37 [†] ±3.63	34.25 [†] ±3.02	29.80 [†] ±1.68	65.26 [†] ±4.41	56.29 [†] ±1.59	81.00 [†] ±15.47
	Individual	57.48 [†] ±4.40	17.93 [†] ±2.84	34.04 [†] ±3.48	30.75 [†] ±0.52	55.05 [†] ±5.43	56.00 [†] ±1.64	84.00 [†] ±14.30
mRMR	E-SVMRank	56.81 [†] ±3.53	18.71 ±3.80	52.00 ±3.34	32.22 [†] ±2.71	46.90 [†] ±2.99	18.11 [†] ±1.98	22.03 [†] ±18.13
	E-Min	58.22 [†] ±3.95	15.78 [†] ±1.54	37.42 [†] ±3.88	30.77 [†] ±0.71	47.56 [†] ±2.38	16.73 [†] ±0.87	22.04 [†] ±17.40
	E-Median	56.47 [†] ±4.12	14.78 [†] ±1.89	49.42 [†] ±5.97	32.90 [†] ±0.89	46.06 [†] ±2.22	15.10 [†] ±1.87	23.27 [†] ±19.75
	E-Mean	57.42 [†] ±3.78	19.78 ±3.66	52.75 ±3.68	32.39 [†] ±1.86	47.20 [†] ±2.81	16.07 [†] ±0.86	23.17 [†] ±19.50
	E-GeomMean	56.61 [†] ±3.84	19.17 [†] ±6.03	49.67 [†] ±5.56	31.98 [†] ±0.75	46.19 [†] ±2.78	15.41 [†] ±1.23	22.57 [†] ±18.20
	E-Stuart	57.42 [†] ±3.78	20.17 ±2.94	50.58 [†] ±5.55	33.20 [†] ±1.11	47.83 ±3.50	19.45 ±1.75	21.51 [†] ±18.48
	E-RRR	57.55 [†] ±3.79	20.47 ±2.15	50.38 [†] ±5.24	33.37 [†] ±1.01	49.61 ±2.54	14.43 [†] ±2.06	23.24 [†] ±18.34
	Individual	53.44 [†] ±3.19	13.35 [†] ±1.69	41.92 [†] ±1.61	32.27 [†] ±0.55	43.93 [†] ±1.08	16.48 [†] ±1.47	22.00 [†] ±18.74
ReliefF	E-SVMRank	57.48 [†] ±3.31	16.41 [†] ±4.03	33.33 [†] ±3.07	27.31 ±1.90	61.05 ±1.96	25.02 [†] ±3.80	82.00 [†] ±19.54
	E-Min	57.96 [†] ±4.18	16.24 [†] ±2.01	33.59 [†] ±2.99	28.21 [†] ±2.33	61.58 ±2.08	25.36 [†] ±3.95	86.00 [†] ±20.04
	E-Median	57.35 [†] ±3.95	16.85 [†] ±3.54	33.21 [†] ±2.89	27.29 ±2.01	61.50 ±1.98	26.61 [†] ±2.87	87.00 [†] ±17.51
	E-Mean	57.82 [†] ±3.33	14.41 [†] ±2.49	33.42 [†] ±3.05	27.72 [†] ±2.12	61.50 ±1.98	25.86 [†] ±2.22	82.00 [†] ±18.57
	E-GeomMean	57.82 [†] ±3.33	14.84 [†] ±3.09	33.50 [†] ±2.93	27.72 [†] ±2.12	61.54 ±1.93	29.54 ±1.21	81.00 [†] ±21.00
	E-Stuart	56.81 [†] ±5.78	14.74 [†] ±2.28	33.42 [†] ±3.05	27.27 ±1.99	61.24 ±2.47	26.58 [†] ±2.81	82.00 [†] ±17.17
	E-RRR	56.81 [†] ±5.78	15.28 [†] ±2.46	33.38 [†] ±2.91	27.62 ±2.16	61.24 ±2.47	25.10 [†] ±3.18	83.00 [†] ±22.57
	Individual	57.48 [†] ±4.40	15.84 [†] ±2.15	33.33 [†] ±4.14	30.76 [†] ±0.72	56.46 [†] ±1.86	25.75 [†] ±2.70	86.00 [†] ±19.55
SVM-RFE	E-SVMRank	49.60 [†] ±3.72	11.93 [†] ±1.24	34.00 [†] ±2.61	33.58 [†] ±1.76	55.01 [†] ±2.54	24.38 [†] ±4.04	26.00 [†] ±21.34
	E-Min	55.13 [†] ±4.88	11.69 [†] ±1.42	33.50 [†] ±2.46	31.79 [†] ±3.25	49.99 [†] ±5.40	23.93 [†] ±3.46	27.00 [†] ±22.72
	E-Median	51.42 [†] ±4.92	11.48 [†] ±0.73	33.79 [†] ±2.90	33.25 [†] ±1.51	53.55 [†] ±4.28	24.47 [†] ±3.97	26.00 [†] ±20.55
	E-Mean	50.07 [†] ±4.32	12.06 [†] ±1.21	33.88 [†] ±2.95	33.17 [†] ±1.77	54.32 [†] ±2.93	25.42 [†] ±4.24	28.00 [†] ±22.88
	E-GeomMean	50.95 [†] ±4.86	11.32 [†] ±0.83	33.33 [†] ±3.15	33.13 [†] ±1.22	54.73 [†] ±2.42	24.75 [†] ±4.35	26.00 [†] ±20.43
	E-Stuart	50.07 [†] ±4.32	11.43 [†] ±0.86	33.50 [†] ±3.37	33.85 [†] ±1.26	55.70 [†] ±2.53	25.77 [†] ±3.79	27.00 [†] ±22.62
	E-RRR	50.07 [†] ±4.32	11.15 [†] ±1.05	33.96 [†] ±3.47	32.14 [†] ±2.73	54.05 [†] ±2.57	25.28 [†] ±3.73	29.00 [†] ±21.98
	Individual	54.45 [†] ±4.56	12.82 [†] ±2.39	35.75 [†] ±4.36	33.52 [†] ±0.92	51.53 [†] ±6.24	16.55 [†] ±7.92	20.00 [†] ±24.94
FS-P	E-SVMRank	55.73 [†] ±3.20	12.26 [†] ±1.56	34.63 [†] ±3.60	33.75 [†] ±1.23	56.89 ±2.28	12.64 [†] ±9.62	75.00 [†] ±19.00
	E-Min	56.75 [†] ±4.10	11.85 [†] ±0.70	40.58 [†] ±7.45	33.56 [†] ±1.81	55.80 ±3.32	12.51 [†] ±9.70	80.00 [†] ±15.63
	E-Median	57.89 [†] ±3.93	11.19 [†] ±0.90	33.38 [†] ±2.75	33.63 [†] ±1.86	57.04 [†] ±1.73	12.51 [†] ±9.70	82.00 [†] ±13.17
	E-Mean	56.74 [†] ±3.49	12.00 [†] ±1.60	34.88 [†] ±3.50	34.13 [†] ±0.90	56.71 ±1.71	12.51 [†] ±9.70	80.00 [†] ±14.91
	E-GeomMean	56.07 [†] ±3.23	11.39 [†] ±1.08	33.58 [†] ±3.20	33.48 [†] ±1.72	56.77 ±1.70	12.51 [†] ±9.70	81.00 [†] ±12.87
	E-Stuart	55.93 [†] ±3.38	11.54 [†] ±1.15	34.00 [†] ±2.83	33.56 [†] ±1.94	56.30 ±1.31	12.51 [†] ±9.70	82.00 [†] ±13.98
	E-RRR	55.93 [†] ±3.38	12.65 [†] ±2.05	34.96 [†] ±2.90	34.02 [†] ±1.05	56.51 ±2.47	12.51 [†] ±9.70	78.00 [†] ±15.49
	Individual	54.18 [†] ±3.28	12.19 [†] ±1.74	34.71 [†] ±3.32	33.43 [†] ±2.31	60.25 [†] ±2.23	8.74 [†] ±3.95	78.00 [†] ±15.49

Table 6: Homogeneous distributed ensemble with a Fisher discriminant ratio threshold: average estimated percentage test errors. The superscript dagger indicates ensemble results that were not significantly different than those for individual rankers.

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	E-SVMRank	57.62 [†] ±5.05	13.28 [†] ±0.66	34.58 [†] ±2.14	30.77 [†] ±0.53	67.62 ±1.59	57.88 [†] ±2.09	80.00 [†] ±22.62
	E-Min	57.29 [†] ±5.03	12.93 [†] ±1.63	33.54 [†] ±2.07	30.85 [†] ±0.52	61.38 [†] ±5.53	62.43 ±2.91	85.00 [†] ±20.68
	E-Median	57.62 [†] ±5.05	13.11 [†] ±1.06	33.67 [†] ±2.02	30.75 [†] ±0.52	64.10 [†] ±4.25	53.76 [†] ±2.92	86.00 [†] ±22.82
	E-Mean	57.62 [†] ±5.05	12.80 [†] ±1.38	37.50 [†] ±6.87	30.75 [†] ±0.52	67.14 ±1.47	58.48 [†] ±1.31	84.00 [†] ±20.12
	E-GeomMean	57.62 [†] ±5.05	13.17 [†] ±0.62	34.08 [†] ±2.26	30.75 [†] ±0.52	66.94 [†] ±1.97	56.79 [†] ±2.94	85.00 [†] ±21.10
	E-Stuart	57.62 [†] ±5.05	12.93 [†] ±1.40	34.21 [†] ±2.10	30.78 [†] ±0.56	67.30 ±1.68	53.58 [†] ±2.91	84.00 [†] ±20.18
	E-RRR	57.62 [†] ±5.05	12.87 [†] ±1.68	34.46 [†] ±1.60	30.81 [†] ±0.55	66.13 [†] ±2.41	54.67 [†] ±1.97	82.00 [†] ±20.38
	Individual	58.76 [†] ±3.92	13.41 [†] ±1.87	33.83 [†] ±4.70	30.77 [†] ±0.52	58.23 [†] ±6.83	56.00 [†] ±1.64	82.00 [†] ±21.50
mRMR	E-SVMRank	55.06 [†] ±4.67	21.84 [†] ±1.55	50.79 ±2.49	32.98 [†] ±0.65	47.52 [†] ±1.75	25.79 [†] ±1.95	24.00 [†] ±19.66
	E-Min	57.62 [†] ±5.05	29.32 [†] ±2.31	36.92 [†] ±4.27	30.76 [†] ±0.53	47.02 [†] ±3.56	26.56 [†] ±1.34	26.00 [†] ±17.14
	E-Median	55.47 [†] ±5.26	26.97 [†] ±2.53	48.46 [†] ±5.09	32.91 [†] ±0.45	45.86 [†] ±2.58	26.83 [†] ±1.58	25.00 [†] ±19.02
	E-Mean	56.89 [†] ±5.45	21.21 [†] ±1.68	50.38 ±2.90	33.08 [†] ±0.41	47.75 ±1.66	25.18 [†] ±1.22	24.00 [†] ±19.01
	E-GeomMean	56.34 [†] ±4.72	24.89 [†] ±2.09	49.71 [†] ±4.32	32.23 [†] ±0.65	46.57 [†] ±3.03	29.55 ±1.75	27.00 [†] ±17.52
	E-Stuart	56.89 [†] ±5.45	22.78 [†] ±1.75	49.75 [†] ±4.46	33.08 [†] ±0.41	48.38 ±1.97	23.59 [†] ±1.25	26.00 [†] ±18.27
	E-RRR	56.28 [†] ±5.69	22.41 [†] ±1.94	50.54 ±5.50	33.09 [†] ±0.41	49.96 ±1.78	24.03 [†] ±1.50	25.00 [†] ±16.30
	Individual	53.91 [†] ±3.52	22.82 [†] ±1.99	42.17 [†] ±2.77	32.22 [†] ±0.35	43.49 [†] ±1.92	24.79 [†] ±1.14	24.00 [†] ±17.76
ReliefF	E-SVMRank	56.42 [†] ±6.38	16.54 [†] ±2.85	33.46 [†] ±2.13	30.09 [†] ±0.69	61.23 [†] ±1.87	29.21 [†] ±1.01	87.00 [†] ±18.91
	E-Min	56.08 [†] ±6.29	14.80 [†] ±2.64	33.50 [†] ±2.11	30.46 [†] ±0.77	61.45 [†] ±2.74	36.12 ±1.33	85.00 [†] ±21.65
	E-Median	57.62 [†] ±5.05	17.08 [†] ±2.70	33.42 [†] ±1.98	30.03 [†] ±0.73	61.78 [†] ±1.92	32.11 [†] ±1.16	87.00 [†] ±18.91
	E-Mean	57.02 [†] ±6.38	14.80 [†] ±3.21	33.54 [†] ±2.13	30.06 [†] ±0.73	61.78 [†] ±1.92	32.73 [†] ±1.79	87.00 [†] ±18.91
	E-GeomMean	57.02 [†] ±6.38	14.93 [†] ±3.59	33.58 [†] ±2.21	30.06 [†] ±0.73	61.87 [†] ±1.81	29.51 [†] ±1.31	87.00 [†] ±18.91
	E-Stuart	56.75 [†] ±6.45	15.08 [†] ±2.79	33.54 [†] ±2.13	30.15 [†] ±0.63	61.58 [†] ±2.30	38.27 ±1.52	86.00 [†] ±18.60
	E-RRR	57.35 [†] ±5.17	14.37 [†] ±1.70	33.50 [†] ±2.05	30.02 [†] ±0.69	61.58 [†] ±2.30	30.87 [†] ±1.16	84.00 [†] ±21.30
	Individual	58.76 [†] ±3.92	21.28 [†] ±3.10	33.21 [†] ±4.24	30.74 [†] ±0.49	59.47 [†] ±1.71	31.10 [†] ±1.78	86.00 [†] ±19.55
SVM-RFE	E-SVMRank	49.60 [†] ±3.56	11.67 [†] ±1.88	33.83 [†] ±1.43	33.95 [†] ±0.83	53.60 [†] ±2.47	33.46 ±2.66	50.00 [†] ±23.49
	E-Min	54.05 [†] ±3.90	12.04 [†] ±1.12	33.71 [†] ±2.10	34.14 [†] ±0.51	51.11 [†] ±5.41	26.13 [†] ±3.29	36.00 [†] ±23.73
	E-Median	50.27 [†] ±3.51	11.39 [†] ±1.30	33.96 [†] ±1.43	33.53 [†] ±0.94	50.44 [†] ±2.14	26.13 [†] ±3.29	34.00 [†] ±23.95
	E-Mean	49.60 [†] ±3.56	11.80 [†] ±2.34	33.83 [†] ±1.51	33.87 [†] ±0.76	53.25 [†] ±2.19	29.12 [†] ±4.10	35.00 [†] ±24.66
	E-GeomMean	50.27 [†] ±3.51	11.37 [†] ±1.30	33.92 [†] ±1.79	34.09 [†] ±0.55	51.44 [†] ±2.24	27.21 [†] ±3.36	34.00 [†] ±23.95
	E-Stuart	49.60 [†] ±3.56	11.17 [†] ±1.83	33.46 [†] ±2.04	34.16 [†] ±0.49	53.94 ±2.32	28.75 [†] ±4.32	32.00 [†] ±23.19
	E-RRR	49.60 [†] ±3.56	11.06 [†] ±1.48	33.71 [†] ±2.01	34.18 [†] ±0.48	53.23 [†] ±1.68	29.68 [†] ±2.86	37.00 [†] ±26.77
	Individual	55.86 [†] ±4.44	12.54 [†] ±2.58	33.50 [†] ±4.35	33.90 [†] ±0.44	48.12 [†] ±5.98	26.00 [†] ±3.33	32.00 [†] ±25.73
FS-P	E-SVMRank	55.47 [†] ±4.96	11.84 [†] ±1.90	33.88 [†] ±1.64	34.13 [†] ±0.49	56.25 ±2.44	53.81 [†] ±4.67	75.00 [†] ±19.00
	E-Min	56.01 [†] ±5.71	11.48 [†] ±1.88	43.17 [†] ±8.19	34.12 [†] ±0.47	55.43 ±2.81	53.81 [†] ±4.67	87.00 [†] ±13.37
	E-Median	58.09 [†] ±5.11	11.63 [†] ±1.81	34.00 [†] ±1.78	34.15 [†] ±0.50	56.80 ±2.02	53.81 [†] ±4.67	86.00 [†] ±10.75
	E-Mean	56.82 [†] ±6.62	12.74 [†] ±2.38	33.83 [†] ±1.86	34.13 [†] ±0.50	55.86 ±2.27	53.81 [†] ±4.67	87.00 [†] ±11.60
	E-GeomMean	55.81 [†] ±5.93	11.35 [†] ±1.39	34.17 [†] ±1.75	34.14 [†] ±0.52	56.54 ±2.24	53.81 [†] ±4.67	84.00 [†] ±11.74
	E-Stuart	56.88 [†] ±4.70	11.50 [†] ±1.85	33.58 [†] ±1.68	34.14 [†] ±0.49	55.84 ±2.01	53.81 [†] ±4.67	84.00 [†] ±11.74
	E-RRR	56.88 [†] ±4.70	12.00 [†] ±2.00	33.58 [†] ±1.71	34.09 [†] ±0.54	56.71 ±2.23	53.81 [†] ±4.67	82.00 [†] ±12.29
	Individual	54.38 [†] ±3.11	12.00 [†] ±1.90	33.50 [†] ±4.42	34.16 [†] ±0.41	61.86 [†] ±1.33	57.18 [†] ±7.10	81.00 [†] ±8.76

Table 7: Homogeneous distributed ensemble with a $\log_2(n)$ threshold: average estimated percentage test errors. The superscript dagger indicates ensemble results that were not significantly different than those for individual rankers.

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	E-SVMRank	59.16 [†] ± 4.31	13.06 [†] ± 2.00	34.38 [†] ± 3.66	30.80 [†] ± 0.79	49.17 [†] ± 1.64	47.88 [†] ± 2.09	83.00 [†] ± 20.14
	E-Min	55.12 [†] ± 2.57	13.11 [†] ± 1.97	34.42 [†] ± 3.28	30.73 [†] ± 0.81	48.34 [†] ± 1.91	50.43 ± 2.91	79.00 [†] ± 23.39
	E-Median	58.62 [†] ± 4.46	13.15 [†] ± 1.65	34.38 [†] ± 3.23	30.76 [†] ± 0.76	49.31 [†] ± 2.18	43.76 [†] ± 2.92	81.00 [†] ± 23.73
	E-Mean	60.65 [†] ± 3.78	13.06 [†] ± 1.96	34.58 [†] ± 3.71	30.76 [†] ± 0.76	48.90 [†] ± 1.68	50.48 ± 1.31	83.00 [†] ± 22.71
	E-GeomMean	59.50 [†] ± 3.88	13.06 [†] ± 1.87	34.29 [†] ± 3.40	30.76 [†] ± 0.76	48.92 [†] ± 1.67	46.79 [†] ± 2.94	83.00 [†] ± 23.03
	E-Stuart	61.73 ± 3.30	13.00 [†] ± 1.99	34.42 [†] ± 3.58	30.72 [†] ± 0.73	48.76 [†] ± 1.53	43.58 [†] ± 2.91	84.00 [†] ± 22.97
	E-RRR	61.79 ± 3.39	12.80 [†] ± 1.71	34.38 [†] ± 3.45	30.71 [†] ± 0.72	48.70 [†] ± 1.19	44.67 [†] ± 1.97	82.00 [†] ± 21.56
	Individual	55.13 [†] ± 4.99	13.39 [†] ± 1.24	33.62 [†] ± 3.50	30.76 [†] ± 0.54	48.62 [†] ± 2.30	45.38 [†] ± 1.68	80.00 [†] ± 21.08
mRMR	E-SVMRank	55.12 [†] ± 2.57	22.02 [†] ± 2.90	51.25 [†] ± 3.09	32.97 [†] ± 0.82	50.01 [†] ± 1.87	14.29 [†] ± 1.45	28.00 [†] ± 21.52
	E-Min	55.12 [†] ± 2.57	29.99 ± 2.14	38.59 [†] ± 6.73	30.82 [†] ± 0.87	49.44 [†] ± 1.43	13.77 [†] ± 0.84	29.00 [†] ± 19.76
	E-Median	55.12 [†] ± 2.57	27.52 ± 1.91	49.75 [†] ± 3.53	32.92 [†] ± 0.71	49.99 [†] ± 2.36	12.26 [†] ± 1.08	30.00 [†] ± 21.44
	E-Mean	55.12 [†] ± 2.57	21.28 [†] ± 2.64	50.54 [†] ± 3.14	33.08 [†] ± 0.81	49.94 [†] ± 1.98	14.80 [†] ± 0.72	27.00 [†] ± 19.73
	E-GeomMean	55.12 [†] ± 2.57	24.69 [†] ± 1.50	47.83 [†] ± 3.12	32.14 [†] ± 0.71	50.12 [†] ± 1.83	11.13 [†] ± 1.25	25.00 [†] ± 21.78
	E-Stuart	55.12 [†] ± 2.57	23.06 [†] ± 2.33	47.75 [†] ± 3.72	33.08 [†] ± 0.81	50.07 [†] ± 2.62	12.75 [†] ± 0.75	25.00 [†] ± 21.78
	E-RRR	55.12 [†] ± 2.57	22.52 [†] ± 3.08	47.33 [†] ± 3.71	33.08 [†] ± 0.81	49.52 [†] ± 2.79	12.52 [†] ± 1.00	26.00 [†] ± 19.58
	Individual	55.13 [†] ± 4.99	22.78 [†] ± 2.24	46.42 [†] ± 3.46	32.29 [†] ± 0.49	47.15 [†] ± 1.71	12.94 [†] ± 0.96	26.00 [†] ± 20.66
ReliefF	E-SVMRank	55.12 [†] ± 2.57	16.32 ± 3.79	33.67 [†] ± 3.87	30.10 [†] ± 0.64	57.23 [†] ± 2.68	17.50 [†] ± 1.91	69.00 [†] ± 20.34
	E-Min	55.12 [†] ± 2.57	14.65 ± 2.91	33.46 [†] ± 3.99	30.30 [†] ± 0.77	56.51 [†] ± 2.65	19.69 [†] ± 1.28	72.00 [†] ± 19.04
	E-Median	55.12 [†] ± 2.57	17.56 [†] ± 4.39	33.71 [†] ± 3.88	30.09 [†] ± 0.59	57.40 [†] ± 2.71	18.83 [†] ± 1.75	74.00 [†] ± 19.64
	E-Mean	55.12 [†] ± 2.57	15.30 ± 3.87	34.25 [†] ± 3.88	30.08 [†] ± 0.66	57.29 [†] ± 2.64	18.83 [†] ± 1.75	75.00 [†] ± 22.17
	E-GeomMean	55.12 [†] ± 2.57	15.61 ± 4.09	33.83 [†] ± 4.42	30.08 [†] ± 0.66	57.09 [†] ± 2.65	21.98 ± 1.38	70.00 [†] ± 21.11
	E-Stuart	55.12 [†] ± 2.57	15.67 ± 3.68	34.08 [†] ± 3.40	30.04 [†] ± 0.65	56.89 [†] ± 2.77	19.51 [†] ± 1.56	72.00 [†] ± 19.04
	E-RRR	55.12 [†] ± 2.57	14.52 ± 3.03	34.04 [†] ± 3.85	30.05 [†] ± 0.71	56.83 [†] ± 2.78	19.29 [†] ± 1.07	72.00 [†] ± 19.04
	Individual	55.13 [†] ± 4.99	20.08 [†] ± 3.14	33.17 [†] ± 3.13	30.70 [†] ± 0.50	58.38 [†] ± 2.23	18.79 [†] ± 1.36	71.00 [†] ± 20.79
SVM-RFE	E-SVMRank	51.22 [†] ± 3.36	12.06 [†] ± 2.07	34.46 [†] ± 3.37	33.89 [†] ± 0.97	64.74 ± 3.77	10.23 [†] ± 1.58	20.00 [†] ± 19.05
	E-Min	55.47 [†] ± 5.81	12.12 [†] ± 1.96	34.00 [†] ± 3.89	34.16 [†] ± 0.72	52.73 [†] ± 3.23	11.56 ± 1.42	18.00 [†] ± 19.05
	E-Median	51.42 [†] ± 3.33	11.67 [†] ± 1.99	34.67 [†] ± 3.28	33.54 [†] ± 0.95	62.25 ± 3.04	10.06 [†] ± 1.27	20.00 [†] ± 20.81
	E-Mean	51.29 [†] ± 3.37	12.19 [†] ± 2.07	34.58 [†] ± 3.32	33.84 [†] ± 1.05	65.56 ± 3.24	11.55 ± 0.98	21.00 [†] ± 20.81
	E-GeomMean	51.29 [†] ± 3.37	11.54 [†] ± 1.98	34.63 [†] ± 3.46	33.99 [†] ± 0.84	61.79 ± 3.75	10.10 [†] ± 0.93	19.00 [†] ± 19.86
	E-Stuart	51.42 [†] ± 3.33	11.80 [†] ± 2.17	34.46 [†] ± 3.30	34.17 [†] ± 0.69	63.61 ± 2.26	10.87 ± 0.94	20.00 [†] ± 20.62
	E-RRR	50.88 [†] ± 3.65	11.56 [†] ± 1.85	34.83 [†] ± 3.15	34.17 [†] ± 0.68	64.23 ± 3.72	11.34 ± 0.80	21.00 [†] ± 20.81
	Individual	54.31 [†] ± 6.50	12.50 [†] ± 1.41	31.71 [†] ± 2.56	33.92 [†] ± 0.59	51.58 [†] ± 3.33	8.32 [†] ± 0.70	15.00 [†] ± 18.41
FS-P	E-SVMRank	54.59 [†] ± 4.54	11.43 [†] ± 1.44	34.13 [†] ± 3.61	34.13 [†] ± 0.69	72.44 ± 3.27	15.58 [†] ± 3.51	84.00 [†] ± 17.13
	E-Min	54.99 [†] ± 4.13	11.78 [†] ± 2.78	35.63 [†] ± 5.15	34.13 [†] ± 0.70	71.69 [†] ± 3.64	16.03 [†] ± 3.41	60.00 [†] ± 20.00
	E-Median	55.80 [†] ± 5.14	11.63 [†] ± 2.00	34.04 [†] ± 4.30	34.12 [†] ± 0.67	72.78 ± 3.37	16.03 [†] ± 3.41	60.00 [†] ± 20.00
	E-Mean	55.06 [†] ± 3.52	12.50 [†] ± 2.20	34.79 [†] ± 3.06	34.15 [†] ± 0.72	72.50 ± 3.30	16.03 [†] ± 3.41	60.00 [†] ± 20.00
	E-GeomMean	55.12 [†] ± 3.54	11.34 [†] ± 2.09	35.21 [†] ± 4.29	34.12 [†] ± 0.71	72.67 ± 3.22	16.03 [†] ± 3.41	61.00 [†] ± 21.32
	E-Stuart	55.06 [†] ± 3.52	11.43 [†] ± 1.74	34.33 [†] ± 3.31	34.14 [†] ± 0.71	72.36 ± 3.10	16.03 [†] ± 3.41	60.00 [†] ± 23.09
	E-RRR	55.12 [†] ± 2.57	11.95 [†] ± 1.70	33.96 [†] ± 4.19	34.09 [†] ± 0.69	71.71 [†] ± 2.88	16.03 [†] ± 3.41	60.00 [†] ± 23.09
	Individual	54.66 [†] ± 4.33	12.17 [†] ± 1.52	33.96 [†] ± 2.94	34.18 [†] ± 0.60	64.95 [†] ± 4.53	16.80 [†] ± 2.22	63.00 [†] ± 24.97

Table 8: Homogeneous distributed ensemble with a 10% threshold: average estimated percentage test errors. The superscript dagger indicates ensemble results that were not significantly different than those for individual rankers.

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	E-SVMRank	59.16 [†] ±4.31	18.06 [†] ±1.32	36.71 [†] ±3.16	26.43 [†] ±0.66	43.18 [†] ±2.47	21.61 [†] ±1.52	73.00 [†] ±25.34
	E-Min	55.12 [†] ±2.57	18.06 [†] ±1.48	36.75 [†] ±3.58	26.41 [†] ±0.67	42.72 [†] ±2.62	21.91 [†] ±2.20	67.00 [†] ±26.71
	E-Median	58.62 [†] ±4.46	18.21 [†] ±1.30	37.29 [†] ±3.36	26.41 [†] ±0.67	43.38 [†] ±2.58	19.75 [†] ±2.42	72.00 [†] ±24.54
	E-Mean	60.65 ±3.78	18.06 [†] ±1.32	36.96 [†] ±3.70	26.41 [†] ±0.67	43.16 [†] ±2.49	18.44 [†] ±1.44	72.00 [†] ±24.54
	E-GeomMean	59.50 [†] ±3.88	18.06 [†] ±1.32	37.21 [†] ±4.22	26.41 [†] ±0.67	43.21 [†] ±2.51	19.63 [†] ±1.23	72.00 [†] ±24.54
	E-Stuart	61.73 ±3.30	18.19 [†] ±1.74	36.33 [†] ±3.70	26.41 [†] ±0.67	43.54 [†] ±2.54	19.63 [†] ±1.23	68.00 [†] ±26.68
	E-RRR	61.79 ±3.39	18.37 [†] ±1.72	36.96 [†] ±3.49	26.41 [†] ±0.67	43.32 [†] ±2.76	20.37 [†] ±1.63	69.00 [†] ±24.42
	Individual	53.17 [†] ±6.89	17.67 [†] ±1.90	36.42 [†] ±3.59	26.34 [†] ±0.72	43.68 [†] ±1.94	19.48 [†] ±2.44	69.00 [†] ±26.44
mRMR	E-SVMRank	55.12 [†] ±2.57	16.54 [†] ±1.50	52.17 [†] ±2.63	31.50 [†] ±0.85	48.92 [†] ±3.01	9.53 ±0.61	35.00 [†] ±20.95
	E-Min	55.12 [†] ±2.57	18.34 ±1.62	40.17 [†] ±7.64	29.07 ±1.05	49.34 [†] ±3.02	9.14 [†] ±1.48	35.00 [†] ±20.95
	E-Median	55.12 [†] ±2.57	16.61 [†] ±1.61	51.25 [†] ±2.63	31.58 [†] ±1.14	48.97 [†] ±3.08	9.71 ±1.00	33.00 [†] ±22.91
	E-Mean	55.12 [†] ±2.57	16.74 [†] ±1.73	52.83 [†] ±2.19	31.75 [†] ±0.87	48.51 [†] ±3.04	9.67 ±0.95	31.00 [†] ±21.85
	E-GeomMean	55.12 [†] ±2.57	16.76 [†] ±1.68	48.00 [†] ±2.53	30.46 [†] ±1.36	48.61 [†] ±3.09	8.00 [†] ±1.20	31.00 [†] ±19.67
	E-Stuart	55.12 [†] ±2.57	16.80 [†] ±1.73	48.92 [†] ±2.79	31.90 [†] ±1.11	48.07 [†] ±3.24	9.89 ±1.21	35.00 [†] ±20.95
	E-RRR	55.12 [†] ±2.57	16.56 [†] ±1.04	48.96 [†] ±2.92	33.53 [†] ±0.66	47.99 [†] ±3.17	7.59 [†] ±1.23	30.00 [†] ±19.23
	Individual	53.17 [†] ±6.89	15.65 [†] ±1.95	49.17 [†] ±1.66	32.01 [†] ±0.72	46.72 [†] ±2.71	7.24 [†] ±1.00	32.00 [†] ±20.98
ReliefF	E-SVMRank	55.12 [†] ±2.57	17.11 [†] ±2.26	35.67 [†] ±3.71	25.56 [†] ±0.69	56.91 [†] ±2.43	27.88 [†] ±0.90	53.00 [†] ±20.20
	E-Min	55.12 [†] ±2.57	15.06 ±2.81	36.54 [†] ±3.09	25.56 [†] ±0.69	54.84 [†] ±2.29	25.59 [†] ±0.86	51.00 [†] ±21.80
	E-Median	55.12 [†] ±2.57	17.54 [†] ±2.17	36.21 [†] ±3.71	25.56 [†] ±0.69	55.71 [†] ±2.63	27.63 [†] ±0.65	52.00 [†] ±21.66
	E-Mean	55.12 [†] ±2.57	16.84 [†] ±1.97	35.79 [†] ±4.32	25.56 [†] ±0.69	57.41 [†] ±2.33	27.07 [†] ±0.16	52.00 [†] ±21.66
	E-GeomMean	55.12 [†] ±2.57	16.82 [†] ±1.85	36.83 [†] ±2.90	25.56 [†] ±0.69	56.87 [†] ±2.34	28.89 [†] ±0.26	55.00 [†] ±21.15
	E-Stuart	55.12 [†] ±2.57	16.80 [†] ±1.97	36.58 [†] ±3.26	25.57 [†] ±0.68	56.38 [†] ±2.38	27.59 [†] ±0.34	52.00 [†] ±21.66
	E-RRR	55.12 [†] ±2.57	16.93 [†] ±1.98	35.92 [†] ±3.03	25.57 [†] ±0.68	55.78 [†] ±2.70	28.20 ±0.78	55.00 ±21.15
	Individual	53.17 [†] ±6.89	19.34 [†] ±3.69	36.58 [†] ±3.47	25.62 [†] ±0.53	54.08 [†] ±3.64	25.62 [†] ±0.53	50.00 ±24.50
SVM-RFE	E-SVMRank	51.22 [†] ±3.36	11.56 [†] ±4.21	38.04 [†] ±2.41	33.11 [†] ±0.91	57.84 ±3.71	7.51 [†] ±0.88	23.00 [†] ±21.53
	E-Min	55.47 [†] ±5.81	12.54 [†] ±4.33	35.42 ±3.89	33.99 [†] ±0.74	45.94 [†] ±3.45	7.68 [†] ±1.23	23.00 [†] ±22.33
	E-Median	51.42 [†] ±3.33	11.71 [†] ±4.63	36.38 ±3.02	33.11 [†] ±1.06	54.71 ±3.52	7.13 [†] ±1.43	25.00 [†] ±21.16
	E-Mean	51.29 [†] ±3.37	10.89 ±3.96	37.25 [†] ±2.84	32.65 [†] ±1.27	58.59 ±3.04	9.99 ±1.46	25.00 [†] ±21.16
	E-GeomMean	51.29 [†] ±3.37	12.02 [†] ±4.62	36.54 ±2.93	33.24 [†] ±1.09	54.98 ±2.45	9.24 ±0.81	23.00 [†] ±22.33
	E-Stuart	51.42 [†] ±3.33	12.54 [†] ±4.51	36.58 ±2.98	33.60 [†] ±1.07	58.65 ±3.23	9.67 ±1.21	23.00 [†] ±22.33
	E-RRR	50.88 [†] ±3.65	12.54 [†] ±4.54	36.25 ±2.56	33.56 [†] ±1.25	59.93 ±3.33	9.52 ±1.78	25.00 [†] ±22.46
	Individual	53.56 [†] ±6.45	18.04 [†] ±3.99	41.83 [†] ±2.82	33.16 [†] ±1.18	42.20 [†] ±1.77	5.56 [†] ±0.69	22.00 [†] ±22.01
FS-P	E-SVMRank	54.59 [†] ±4.53	10.13 [†] ±3.05	36.58 [†] ±2.65	34.07 [†] ±0.72	75.94 ±4.45	6.63 [†] ±1.16	69.00 [†] ±14.49
	E-Min	54.99 [†] ±4.13	12.41 [†] ±5.15	36.29 [†] ±3.74	34.17 [†] ±0.69	69.44 [†] ±4.90	6.63 [†] ±1.16	56.00 [†] ±21.19
	E-Median	55.80 [†] ±5.14	11.67 [†] ±3.91	37.25 [†] ±2.17	34.17 [†] ±0.69	74.45 ±4.03	6.63 [†] ±1.16	55.00 [†] ±21.21
	E-Mean	55.06 [†] ±3.52	10.15 [†] ±3.13	36.96 [†] ±3.01	34.16 [†] ±0.70	75.43 ±4.39	6.63 [†] ±1.16	56.00 [†] ±21.19
	E-GeomMean	55.12 [†] ±3.54	10.80 [†] ±3.83	36.58 [†] ±2.41	34.17 [†] ±0.69	75.00 ±4.17	6.63 [†] ±1.16	56.00 [†] ±21.19
	E-Stuart	55.06 [†] ±3.52	10.00 [†] ±3.07	36.08 [†] ±2.06	34.07 [†] ±0.73	74.86 ±4.08	6.63 [†] ±1.16	55.00 [†] ±20.68
	E-RRR	55.12 [†] ±2.57	9.67 [†] ±3.00	35.75 ±2.84	33.82 [†] ±0.64	75.07 ±3.76	6.63 [†] ±1.16	55.00 [†] ±20.68
	Individual	53.03 [†] ±4.51	14.45 [†] ±5.53	39.17 [†] ±1.94	34.03 [†] ±0.54	61.45 [†] ±4.34	7.27 [†] ±1.15	59.00 [†] ±20.25

Table 9: Homogeneous distributed ensemble with a 25% threshold: average estimated percentage test errors. The superscript dagger indicates ensemble results that were not significantly different than those for individual rankers.

Ranker		Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
InfoGain	E-SVMRank	59.16 [†] ±4.31	16.89 [†] ±1.20	38.50 [†] ±3.72	24.78 [†] ±0.55	37.98 [†] ±3.34	14.93 ±1.13	57.00 [†] ±28.55
	E-Min	55.12 [†] ±2.57	16.76 [†] ±1.35	39.29 [†] ±4.11	24.97 [†] ±0.69	36.73 [†] ±2.67	12.92 [†] ±1.02	59.00 [†] ±26.13
	E-Median	58.62 [†] ±4.46	16.87 [†] ±1.14	38.50 [†] ±3.56	24.76 [†] ±0.70	38.46 [†] ±2.93	11.37 [†] ±1.68	57.00 [†] ±26.27
	E-Mean	60.65 [†] ±3.78	16.89 [†] ±1.21	39.04 [†] ±3.94	24.79 [†] ±0.52	38.05 [†] ±3.40	12.33 [†] ±0.54	60.00 [†] ±27.63
	E-GeomMean	59.50 [†] ±3.88	16.93 [†] ±1.20	39.38 [†] ±3.53	24.80 [†] ±0.45	38.01 [†] ±3.12	10.43 [†] ±1.35	59.00 [†] ±28.12
	E-Stuart	61.73 ±3.30	16.93 [†] ±1.17	38.13 [†] ±3.98	25.09 [†] ±0.89	38.34 [†] ±2.52	13.62 [†] ±1.12	60.00 [†] ±27.63
	E-RRR	61.79 ±3.39	17.02 [†] ±1.16	38.04 [†] ±3.40	25.22 [†] ±0.94	38.17 [†] ±2.40	13.51 [†] ±0.82	60.00 [†] ±27.63
	Individual	54.05 [†] ±4.31	16.95 [†] ±1.28	38.54 [†] ±3.49	24.90 [†] ±0.79	37.98 [†] ±2.36	11.79 [†] ±1.10	58.00 [†] ±28.21
mRMR	E-SVMRank	55.12 [†] ±2.57	14.13 [†] ±1.17	39.96 [†] ±2.96	29.75 [†] ±0.80	43.50 [†] ±3.18	7.33 ±0.55	34.00 [†] ±23.89
	E-Min	55.12 [†] ±2.57	15.37 [†] ±1.21	40.08 [†] ±2.97	28.27 ±1.25	42.02 [†] ±3.69	6.27 [†] ±1.41	37.00 [†] ±24.67
	E-Median	55.12 [†] ±2.57	14.08 [†] ±1.05	39.67 [†] ±3.02	30.55 [†] ±0.87	42.81 [†] ±3.73	5.27 [†] ±1.25	35.00 [†] ±25.02
	E-Mean	55.12 [†] ±2.57	14.11 [†] ±1.11	40.17 [†] ±3.33	29.82 [†] ±0.76	42.66 [†] ±3.49	5.79 [†] ±1.53	37.00 [†] ±24.67
	E-GeomMean	55.12 [†] ±2.57	14.06 [†] ±1.06	39.83 [†] ±2.81	28.99 ±0.66	42.62 [†] ±3.79	5.46 [†] ±1.38	37.00 [†] ±24.67
	E-Stuart	55.12 [†] ±2.57	14.24 [†] ±1.02	40.21 [†] ±4.11	30.20 [†] ±0.79	43.71 [†] ±2.91	7.84 ±0.77	33.00 [†] ±25.37
	E-RRR	55.12 [†] ±2.57	16.13 ±2.07	40.25 [†] ±3.50	31.05 [†] ±1.02	43.91 [†] ±3.26	6.32 [†] ±1.41	35.00 [†] ±25.02
	Individual	54.05 [†] ±4.31	13.28 [†] ±1.79	39.21 [†] ±3.48	31.01 [†] ±0.51	46.75 [†] ±3.36	5.36 [†] ±0.69	36.00 [†] ±24.59
ReliefF	E-SVMRank	55.12 [†] ±2.57	17.84 [†] ±0.85	38.92 [†] ±1.86	23.39 [†] ±0.66	47.02 [†] ±3.95	7.10 ±1.04	37.00 [†] ±23.07
	E-Min	55.12 [†] ±2.57	17.34 [†] ±1.23	39.09 [†] ±2.54	23.39 [†] ±0.59	45.52 [†] ±4.28	6.16 [†] ±1.14	35.00 [†] ±22.36
	E-Median	55.12 [†] ±2.57	17.76 [†] ±0.91	40.33 [†] ±3.73	23.23 [†] ±0.64	46.11 [†] ±3.79	7.74 ±1.06	36.00 [†] ±23.15
	E-Mean	55.12 [†] ±2.57	17.82 [†] ±0.96	38.79 [†] ±1.74	23.22 [†] ±0.64	47.04 [†] ±3.92	5.00 [†] ±0.84	38.00 [†] ±22.75
	E-GeomMean	55.12 [†] ±2.57	17.61 [†] ±0.90	39.58 [†] ±2.32	23.22 [†] ±0.64	46.36 [†] ±4.08	6.38 [†] ±0.57	35.00 [†] ±22.36
	E-Stuart	55.12 [†] ±2.57	17.91 ±0.80	38.71 [†] ±2.16	23.48 [†] ±0.67	46.30 [†] ±4.25	6.27 [†] ±0.85	38.00 [†] ±23.07
	E-RRR	55.12 [†] ±2.57	18.08 ±0.69	38.71 [†] ±2.77	23.39 [†] ±0.76	45.91 [†] ±4.12	6.38 [†] ±0.94	35.00 [†] ±22.36
	Individual	54.05 [†] ±4.31	16.17 [†] ±1.15	39.92 [†] ±2.89	23.51 [†] ±0.43	50.33 [†] ±3.45	5.56 [†] ±0.69	37.00 [†] ±23.12
SVM-RFE	E-SVMRank	51.22 [†] ±3.36	18.02 [†] ±1.11	39.46 [†] ±3.51	31.34 [†] ±2.03	48.71 ±4.14	5.62 [†] ±0.95	46.00 [†] ±22.47
	E-Min	55.47 [†] ±5.81	17.67 [†] ±1.16	39.00 [†] ±3.05	31.37 [†] ±1.86	39.98 [†] ±1.62	7.70 ±1.16	33.00 [†] ±21.47
	E-Median	51.42 [†] ±3.33	18.06 [†] ±1.11	39.75 [†] ±3.25	31.79 [†] ±2.45	48.08 ±4.18	5.35 [†] ±1.27	35.00 [†] ±23.55
	E-Mean	51.29 [†] ±3.37	17.97 [†] ±1.11	39.21 [†] ±3.01	31.37 [†] ±2.02	48.92 ±3.63	7.12 ±0.85	34.00 [†] ±22.91
	E-GeomMean	51.29 [†] ±3.37	17.82 [†] ±1.29	39.71 [†] ±2.70	30.51 [†] ±1.64	44.57 [†] ±3.29	5.70 [†] ±1.16	34.00 [†] ±22.91
	E-Stuart	51.42 [†] ±3.33	17.82 [†] ±1.00	39.38 [†] ±2.95	31.49 [†] ±1.60	48.36 ±4.29	6.35 [†] ±0.91	33.00 [†] ±22.27
	E-RRR	50.88 [†] ±3.65	18.04 [†] ±1.13	39.17 [†] ±2.92	31.85 [†] ±1.48	50.31 [†] ±4.73	7.83 ±1.34	36.00 [†] ±23.17
	Individual	52.57 [†] ±6.65	18.11 [†] ±1.66	42.25 [†] ±3.23	31.80 [†] ±2.35	37.70 [†] ±3.12	4.98 [†] ±0.66	33.00 [†] ±23.12
FS-P	E-SVMRank	54.59 [†] ±4.54	17.15 [†] ±3.23	39.63 [†] ±2.33	34.07 ±0.64	65.99 ±6.19	5.19 [†] ±0.85	61.00 [†] ±20.79
	E-Min	54.99 [†] ±4.13	17.26 [†] ±2.65	38.75 [†] ±2.28	34.17 ±0.69	60.06 [†] ±6.32	5.19 [†] ±0.85	49.00 [†] ±23.78
	E-Median	55.80 [†] ±5.14	18.08 [†] ±1.06	39.21 [†] ±3.28	34.06 ±0.67	65.24 ±6.32	5.19 [†] ±0.85	48.00 [†] ±22.01
	E-Mean	55.06 [†] ±3.52	18.06 [†] ±1.15	39.17 [†] ±2.43	34.08 ±0.65	66.14 ±6.26	5.19 [†] ±0.85	48.00 [†] ±22.01
	E-GeomMean	55.12 [†] ±3.54	18.24 [†] ±1.12	38.54 [†] ±2.90	34.12 ±0.69	64.77 ±6.65	5.19 [†] ±0.85	49.00 [†] ±23.78
	E-Stuart	55.06 [†] ±3.52	18.34 [†] ±1.07	39.00 [†] ±2.07	34.04 [†] ±0.64	65.51 ±6.61	5.19 [†] ±0.85	48.00 [†] ±22.01
	E-RRR	55.12 [†] ±2.57	18.15 [†] ±1.23	39.71 [†] ±3.61	33.82 [†] ±1.12	65.33 ±6.24	5.19 [†] ±0.85	48.00 [†] ±22.01
	Individual	54.80 [†] ±5.87	16.06 [†] ±4.18	40.67 [†] ±3.84	32.55 [†] ±0.96	48.39 [†] ±4.95	5.22 [†] ±0.63	53.00 [†] ±23.12

Table 10: Homogeneous distributed ensemble with a 50% threshold: average estimated percentage test errors. The superscript dagger indicates ensemble results that were not significantly different than those for individual rankers.

Ranker	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
E-SVMRank	54.78 [†] ± 3.27	11.26 [†] ± 1.52	34.75 [†] ± 5.16	30.06 [†] ± 2.02	52.06 [†] ± 4.82	31.21 ± 5.21	77.00 ± 20.58
E-Min	53.44 [†] ± 3.19	18.93 ± 3.05	34.29 [†] ± 3.33	31.36 [†] ± 1.11	51.21 [†] ± 1.79	18.04 [†] ± 6.13	60.00 [†] ± 17.64
E-Median	56.00 [†] ± 2.95	11.67 [†] ± 1.57	33.21 [†] ± 4.33	29.98 [†] ± 1.81	54.57 ± 4.15	37.70 ± 4.90	85.00 ± 19.00
E-Mean	54.78 [†] ± 3.27	12.82 [†] ± 3.16	34.46 [†] ± 4.50	29.84 [†] ± 2.20	49.72 [†] ± 3.51	30.47 ± 4.70	44.00 [†] ± 24.13
E-GeomMean	53.44 [†] ± 3.19	11.28 [†] ± 1.54	33.67 [†] ± 3.93	30.90 [†] ± 0.94	52.25 [†] ± 3.67	36.01 ± 3.45	60.00 [†] ± 23.09
E-Stuart	54.78 [†] ± 3.27	11.28 [†] ± 1.54	33.63 [†] ± 3.94	30.30 [†] ± 1.94	51.94 [†] ± 3.40	35.74 ± 4.08	50.00 [†] ± 29.06
E-RRA	56.47 [†] ± 5.04	14.13 [†] ± 3.37	33.17 [†] ± 3.95	29.69 [†] ± 2.00	50.30 [†] ± 5.44	28.53 [†] ± 4.06	59.00 [†] ± 21.32
InfoGain	57.48 [†] ± 4.40	17.93 ± 2.84	34.04 [†] ± 3.48	30.75 [†] ± 0.52	55.05 ± 5.43	56.00 ± 1.64	84.00 ± 14.30
mRMR	53.44 [†] ± 3.19	13.35 [†] ± 1.69	41.92 ± 1.61	32.27 [†] ± 0.55	43.93 [†] ± 1.08	16.48 [†] ± 1.47	22.00 [†] ± 18.74
ReliefF	57.48 [†] ± 4.40	15.84 [†] ± 2.15	33.33 [†] ± 4.14	30.76 [†] ± 0.72	56.46 ± 1.86	25.75 [†] ± 2.70	86.00 ± 19.55
SVM-RFE	54.45 [†] ± 4.56	12.82 [†] ± 2.39	35.75 [†] ± 4.36	33.52 ± 0.92	51.53 [†] ± 6.24	16.55 [†] ± 7.92	20.00 [†] ± 24.94
FS-P	54.18 [†] ± 3.28	12.19 [†] ± 1.74	34.71 [†] ± 3.32	33.43 ± 2.31	60.25 ± 2.23	8.74 [†] ± 3.95	78.00 ± 15.49

Table 11: Heterogeneous centralized ensemble with a *Fisher discriminant ratio* threshold: average estimated percentage test errors. The superscript dagger indicates results that were not significantly different from the best result.

from the lowest average error for the individual methods in 40 of the 49 experiments performed with the *Fisher discriminant ratio* threshold (Table 11), or in 41 of the 49 experiments performed with the $\log_2(n)$ threshold (Table 12). Results were similar as the threshold increased to allow the use of more features. The ensemble method obtained favorable results in 42 of the 49 experiments performed with a 10% threshold (Table 13) and with a 25% threshold (Table 14), and in 40 of the 49 experiments performed with a 50% threshold (Table 15).

Ranker	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
E-SVMRank	54.58 [†] ± 3.33	12.22 [†] ± 2.98	33.71 [†] ± 4.82	31.30 [†] ± 0.71	50.58 [†] ± 3.94	31.21 [†] ± 2.48	77.00 ± 20.58
E-Min	53.91 [†] ± 3.52	20.04 ± 1.78	33.46 [†] ± 4.40	32.28 ± 0.60	46.80 [†] ± 3.14	33.82 [†] ± 4.90	65.00 [†] ± 20.68
E-Median	57.62 [†] ± 3.20	11.76 [†] ± 1.09	33.46 [†] ± 4.45	31.48 [†] ± 0.44	55.36 ± 4.76	37.70 ± 4.70	85.00 ± 19.00
E-Mean	54.24 [†] ± 3.22	14.76 [†] ± 4.52	33.50 [†] ± 4.30	31.40 [†] ± 0.68	51.26 [†] ± 4.62	30.47 [†] ± 3.45	44.00 [†] ± 20.66
E-GeomMean	53.91 [†] ± 3.52	12.19 [†] ± 2.92	33.33 [†] ± 4.37	31.25 [†] ± 0.73	51.73 [†] ± 3.26	36.01 ± 6.08	56.00 [†] ± 26.75
E-Stuart	54.65 [†] ± 4.16	11.30 [†] ± 1.37	33.33 [†] ± 4.37	31.22 [†] ± 0.78	50.57 [†] ± 3.99	35.74 ± 4.06	54.00 [†] ± 27.97
E-RRA	56.20 [†] ± 3.94	13.67 [†] ± 3.15	33.50 [†] ± 4.44	30.97 [†] ± 0.78	50.97 [†] ± 4.16	28.53 [†] ± 0.00	70.00 [†] ± 19.44
InfoGain	58.76 [†] ± 3.92	13.41 [†] ± 1.87	33.83 [†] ± 4.70	30.77 [†] ± 0.52	58.23 ± 6.83	56.00 ± 1.64	82.00 ± 21.50
mRMR	53.91 [†] ± 3.52	22.82 ± 1.99	42.17 ± 2.77	32.22 ± 0.35	43.49 [†] ± 1.92	24.79 [†] ± 1.14	24.00 [†] ± 17.76
ReliefF	58.76 [†] ± 3.92	21.28 [†] ± 3.10	33.21 [†] ± 4.24	30.74 [†] ± 0.49	59.47 ± 1.71	31.10 [†] ± 1.78	86.00 ± 19.55
SVM-RFE	55.86 [†] ± 4.44	12.54 [†] ± 2.58	33.50 [†] ± 4.35	33.90 ± 0.44	48.12 [†] ± 5.98	26.00 [†] ± 3.33	32.00 [†] ± 25.73
FS-P	54.38 [†] ± 3.11	12.00 [†] ± 1.90	33.50 [†] ± 4.42	34.16 ± 0.41	61.86 ± 1.33	57.18 ± 7.10	81.00 ± 8.76

Table 12: Heterogeneous centralized ensemble with a $\log_2(n)$ threshold: average estimated percentage test errors. The superscript dagger indicates results that were not significantly different from the best result.

Focusing on the behavior of the individual feature selection rankers (the bottom five rows in each table), it can be observed that not one of the five individual methods significantly outperformed the ensemble approaches for any dataset or threshold combination. Therefore, although an individual method might well perform better than an ensemble method in a given scenario, overall it would appear that the ensemble approach is the most consistent and reliable

Ranker	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
E-SVMRank	54.18 [†] ± 2.72	11.26 [†] ± 1.52	36.29 [†] ± 3.12	31.40 [†] ± 0.87	51.49 [†] ± 1.93	14.09 [†] ± 0.98	84.00 ± 14.30
E-Min	50.27 [†] ± 3.97	19.63 ± 2.46	35.21 [†] ± 2.79	32.09 ± 0.64	58.48 ± 2.27	15.98 ± 2.52	47.00 [†] ± 27.10
E-Median	54.18 [†] ± 2.72	11.67 [†] ± 1.57	36.63 [†] ± 3.86	31.51 [†] ± 0.66	52.79 [†] ± 2.57	13.18 [†] ± 0.70	50.00 [†] ± 21.60
E-Mean	54.18 [†] ± 2.72	13.02 [†] ± 3.70	36.21 [†] ± 3.61	31.36 [†] ± 0.75	51.22 [†] ± 1.81	14.12 [†] ± 1.19	30.00 [†] ± 18.26
E-GeomMean	54.18 [†] ± 2.72	11.28 [†] ± 1.54	37.25 [†] ± 3.82	31.20 [†] ± 0.78	54.89 ± 2.10	14.36 ± 1.60	48.00 [†] ± 25.58
E-Stuart	54.18 [†] ± 2.72	11.28 [†] ± 1.54	37.04 [†] ± 4.26	31.16 [†] ± 0.70	51.65 [†] ± 1.15	13.65 [†] ± 1.56	43.00 [†] ± 24.06
E-RRA	54.18 [†] ± 2.72	14.50 [†] ± 4.12	36.54 [†] ± 3.50	31.01 [†] ± 0.44	50.29 [†] ± 1.36	13.56 [†] ± 1.13	47.00 [†] ± 24.52
InfoGain	55.13 [†] ± 4.99	13.39 [†] ± 1.24	33.62 [†] ± 3.50	30.76 [†] ± 0.54	48.62 [†] ± 2.30	45.38 ± 1.68	80.00 ± 21.08
mRMR	55.13 [†] ± 4.99	22.78 ± 2.24	46.42 ± 3.46	32.29 ± 0.49	47.15 [†] ± 1.71	12.94 [†] ± 0.96	26.00 [†] ± 20.66
ReliefF	55.13 [†] ± 4.99	20.08 ± 3.14	33.17 [†] ± 3.13	30.70 [†] ± 0.50	58.38 ± 2.23	18.79 ± 1.36	71.00 ± 20.79
SVM-RFE	54.31 [†] ± 6.50	12.50 [†] ± 1.41	31.71 [†] ± 2.56	33.92 ± 0.59	51.58 [†] ± 3.33	8.32 [†] ± 0.70	15.00 [†] ± 18.41
FS-P	54.66 [†] ± 4.33	12.17 [†] ± 1.52	33.96 [†] ± 2.94	34.18 ± 0.60	64.95 ± 4.53	16.80 ± 2.22	63.00 ± 24.97

Table 13: Heterogeneous centralized ensemble with a 10% threshold: average estimated percentage test errors. The superscript dagger indicates results that were not significantly different from the best result.

Ranker	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
E-SVMRank	54.18 [†] ± 2.72	16.45 [†] ± 2.73	38.88 [†] ± 2.85	28.26 [†] ± 1.01	51.33 [†] ± 3.06	8.77 [†] ± 1.57	75.00 ± 17.16
E-Min	50.27 [†] ± 3.97	19.52 [†] ± 1.83	38.33 [†] ± 2.50	30.32 ± 1.05	53.23 ± 1.59	7.47 [†] ± 0.81	50.00 [†] ± 27.08
E-Median	54.18 [†] ± 2.72	17.13 [†] ± 3.16	39.08 [†] ± 3.12	29.25 ± 1.62	51.62 [†] ± 1.64	8.10 [†] ± 1.61	38.00 [†] ± 27.00
E-Mean	54.18 [†] ± 2.72	16.50 [†] ± 2.66	38.71 [†] ± 3.39	28.01 [†] ± 1.29	50.76 [†] ± 2.95	9.41 ± 1.30	30.00 [†] ± 26.25
E-GeomMean	54.18 [†] ± 2.72	18.87 [†] ± 2.13	39.92 [†] ± 2.52	27.77 [†] ± 0.97	54.83 ± 2.15	7.72 [†] ± 1.40	41.00 [†] ± 26.01
E-Stuart	54.18 [†] ± 2.72	16.84 [†] ± 2.82	38.92 [†] ± 2.77	27.95 [†] ± 0.91	52.62 ± 1.71	8.77 [†] ± 1.72	41.00 [†] ± 26.01
E-RRA	54.18 [†] ± 2.72	17.04 [†] ± 2.65	39.75 [†] ± 2.91	27.81 [†] ± 0.76	50.07 [†] ± 3.12	8.89 [†] ± 1.95	41.00 [†] ± 26.01
InfoGain	53.17 [†] ± 6.89	17.67 [†] ± 1.90	36.42 [†] ± 3.59	26.34 [†] ± 0.72	43.68 [†] ± 1.94	19.48 ± 2.44	69.00 ± 26.44
mRMR	53.17 [†] ± 6.89	15.65 [†] ± 1.95	49.17 ± 1.66	32.01 ± 0.72	46.72 [†] ± 2.71	7.24 [†] ± 1.00	32.00 [†] ± 20.98
ReliefF	53.17 [†] ± 6.89	19.34 [†] ± 3.69	36.58 [†] ± 3.47	25.62 [†] ± 0.53	54.08 ± 3.64	11.69 ± 0.92	50.00 [†] ± 24.50
SVM-RFE	53.56 [†] ± 6.45	18.04 [†] ± 3.99	41.83 ± 2.82	33.16 ± 1.18	42.20 [†] ± 1.77	5.56 [†] ± 0.69	22.00 [†] ± 22.01
FS-P	53.03 [†] ± 4.51	14.45 [†] ± 5.53	39.17 [†] ± 1.94	34.03 ± 0.54	61.45 ± 4.34	7.27 [†] ± 1.15	59.00 [†] ± 20.25

Table 14: Heterogeneous centralized ensemble with a 25% threshold: average estimated percentage test errors. The superscript dagger indicates results that were not significantly different from the best result.

Ranker	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
E-SVMRank	54.18 [†] ± 2.72	17.08 ± 1.91	41.92 [†] ± 2.84	26.15 [†] ± 0.79	45.34 [†] ± 4.29	5.54 [†] ± 0.57	63.00 [†] ± 23.59
E-Min	50.27 [†] ± 3.97	17.37 ± 2.03	41.29 [†] ± 2.36	26.58 ± 0.92	46.45 ± 4.76	5.71 [†] ± 0.68	49.00 [†] ± 25.14
E-Median	54.18 [†] ± 2.72	16.93 [†] ± 2.00	40.96 [†] ± 2.98	26.55 [†] ± 1.27	45.47 [†] ± 4.33	5.26 [†] ± 0.65	41.00 [†] ± 22.34
E-Mean	54.18 [†] ± 2.72	16.93 [†] ± 1.99	41.63 [†] ± 2.80	26.19 [†] ± 0.73	45.30 [†] ± 4.13	5.54 [†] ± 0.53	38.00 [†] ± 24.40
E-GeomMean	54.18 [†] ± 2.72	16.71 [†] ± 1.86	41.04 [†] ± 2.99	26.43 ± 0.71	45.62 [†] ± 4.26	5.59 [†] ± 0.68	44.00 [†] ± 22.71
E-Stuart	54.18 [†] ± 2.72	17.04 ± 2.10	40.42 [†] ± 3.04	26.52 ± 0.84	44.14 [†] ± 3.95	5.71 [†] ± 0.67	41.00 [†] ± 23.31
E-RRA	54.18 [†] ± 2.72	17.00 [†] ± 2.84	41.29 [†] ± 3.11	26.26 [†] ± 0.74	42.41 [†] ± 3.28	5.70 [†] ± 0.67	41.00 [†] ± 31.31
InfoGain	54.05 [†] ± 4.31	16.95 ± 1.28	38.54 [†] ± 3.49	24.90 [†] ± 0.79	37.98 [†] ± 2.36	11.79 ± 1.10	58.00 [†] ± 28.21
mRMR	54.05 [†] ± 4.31	13.28 [†] ± 1.79	39.21 [†] ± 3.48	31.01 ± 0.51	46.75 ± 3.36	5.36 [†] ± 0.69	36.00 [†] ± 24.59
ReliefF	54.05 [†] ± 4.31	16.17 [†] ± 1.15	39.92 [†] ± 2.89	23.51 [†] ± 0.43	50.33 ± 3.45	5.56 [†] ± 0.69	37.00 [†] ± 23.12
SVM-RFE	52.57 [†] ± 6.65	18.11 ± 1.66	42.25 [†] ± 3.23	31.80 ± 2.35	37.70 [†] ± 3.12	4.98 [†] ± 0.66	33.00 [†] ± 23.12
FS-P	54.80 [†] ± 5.87	16.06 ± 4.18	40.67 [†] ± 3.84	32.55 ± 0.96	48.39 ± 4.95	5.22 [†] ± 0.63	53.00 [†] ± 23.12

Table 15: Heterogeneous centralized ensemble with a 50% threshold: average estimated percentage test errors. The superscript dagger indicates results that were not significantly different from the best result.

approach to a feature selection process.

5. Discussion

Below we analyze and discuss the average training time, estimated test error and standard deviation results, as presented above, for our two different ensemble approaches, in terms of assessing their relative performance, illustrating the benefits of the ensemble approach and explaining certain experimental results.

5.1. Homogeneous Distributed Ensemble

Figures 4 and 5 show the average time gains for the homogeneous distributed ensemble versus the individual approaches. As can be observed, the homogeneous ensembles compared to the individual approaches improved times by a factor of 100 on average in the best case. The feature selection method whose average training times most improved in the distribution process was the embedded *SVM-RFE*. The fact that the *InfoGain* filter yielded the poorest improvement is not surprising, since it is an univariate and fast method – so even attempts at parallelization produced no improvement [52]. Figure 5, referring to the datasets, shows that the best time improvement occurred with the *Connect4* dataset; this was because it had the largest number of samples of all the datasets and so was able to take most advantage of the distribution process. For datasets with greater dimensionality and smaller sample sizes, e.g., *Madelon*, *Isolet*, *USPS* and *Pixraw10P*, the time improvement with the *SVM-RFE* method was also significant, mainly due to the relatively small number of samples used in each iterative training run by the *SVM* classifier used by this method.

The most important advantage of the homogeneous distributed ensemble approach was clearly the great reduction in training times while classification performance held at reasonable—and sometimes even improved—levels. This outcome reflects the notion of divide-and-conquer since, in some cases, the result obtained by a feature selection method may be more accurate when the focus is on a local region of the data.

5.2. Heterogeneous Centralized Ensemble

The experimental results demonstrate the suitability of the heterogeneous centralized ensemble approach, since they match or improve on the estimated test error achieved by the individual feature selection methods. Figure 6 shows the number of cases for which the results obtained by the individual and the heterogeneous centralized ensemble approaches were not significantly different than the best result (in other words, the number of times that results were comparable with the best result). As can be observed, the *E-RRA* ensemble approach obtained results that were not significantly different from the best result in all 35 experiments, compared to 28 out of 35 experiments for the best performing individual feature selection method (*SVM-RFE*). Five of the remaining six ensemble methods (*E-SVMRank*, *E-Median*, *E-Mean*, *E-GeomMean* and *E-Stuart*)

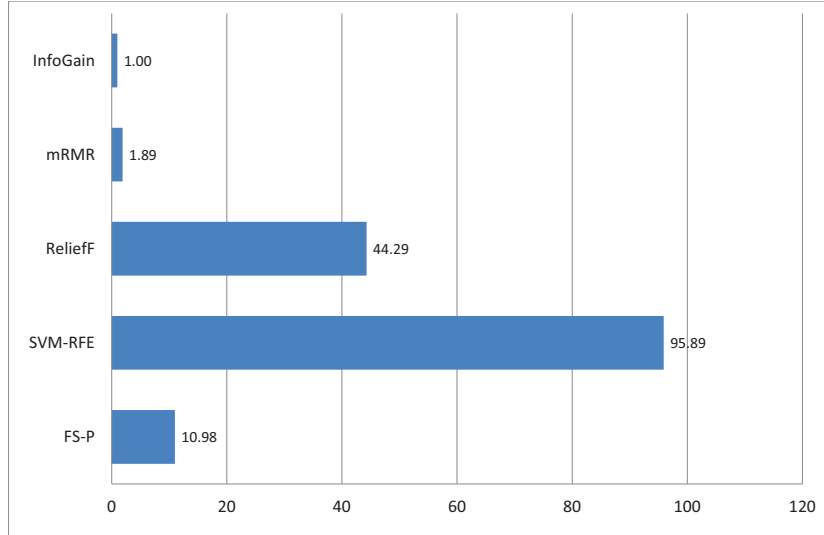


Figure 4: Average speedup for homogeneous distributed ensembles versus individual approaches.

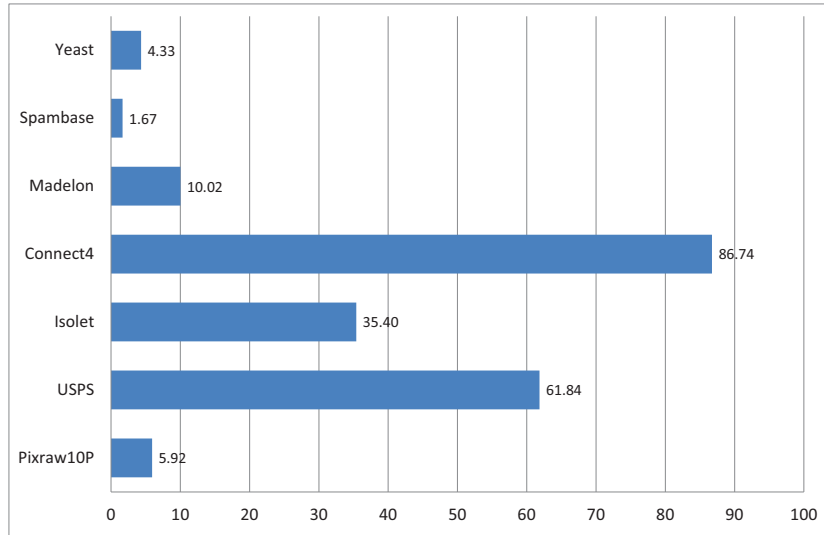


Figure 5: Average speedup for each dataset using the homogeneous distributed ensembles versus individual approaches.

matched or (mostly) improved on the results obtained by *SVM-RFE*, obtaining results that were not significantly different in 28-33 of the 35 experiments. Overall, an ensemble approach would seem to be the most reliable approach to feature selection, although in some specific cases, an individual method (not always the same one) might well perform better than the ensemble.

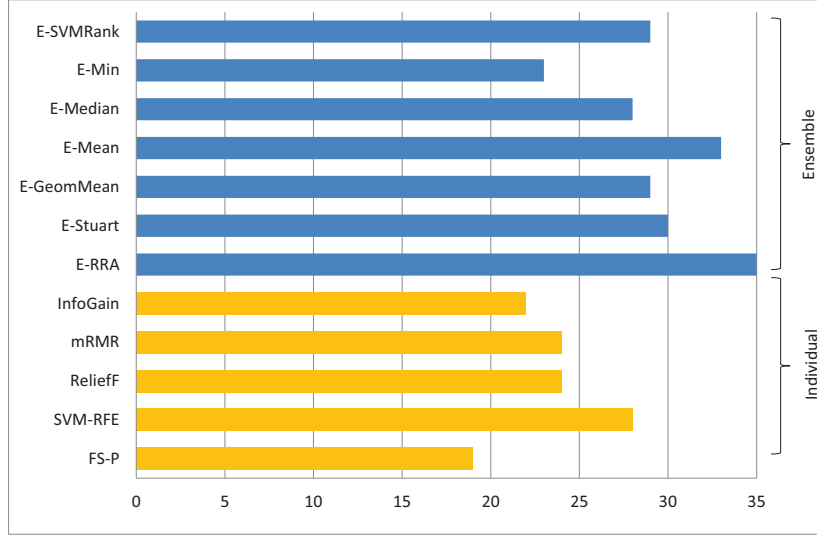


Figure 6: Number of cases when the results obtained by the individual and the heterogeneous centralized ensemble approaches were comparable with the best result.

5.3. Combination Methods and Threshold Values

Comparative analyses were performed in terms of average percentage test error for the different combination methods (Figure 7) and for the different threshold values (Figure 8). In both figures the first six groups of bars reflect the six standard datasets (larger sample size than dimension) and the last group is a microarray dataset.

The analysis for Figure 7 is divided into two parts according to the studied dataset. Thus:

- *Standard datasets.* It can be seen that the different combination methods obtained similar results except the ensemble that used the *Min* reduction function (second bar in each set of bars). This combination method produced irregular results, in that it obtained the best average test error for the *Yeast*, *Madelon* and *USPS* datasets, but the worst average test error for the *Spambase* and *Connect4* datasets. Thus, with the exception of *Min*, one combination method cannot be recommended over another, given the results obtained.

- *Microarray dataset.* In this case, accuracy varied greatly depending on the combination method, with *E-SVMRank* achieving the worst result, and *E-Mean* obtaining the best result. These results can be explained by the nature of this dataset, radically different from the other six datasets. It can be concluded that the choice of combination method can influence final results for microarray datasets – a conclusion consistent with that reported elsewhere regarding an extensive study of microarray datasets [45].

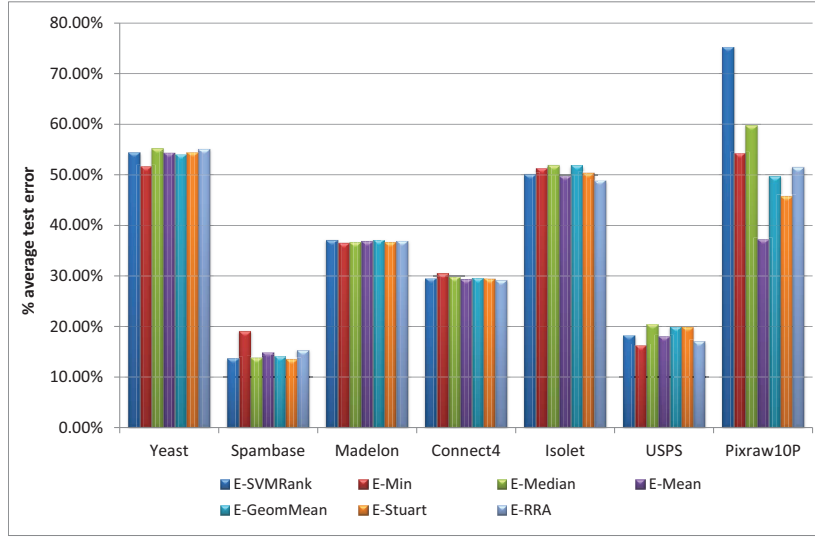


Figure 7: Comparison of average estimated percentage test errors for the different combination methods.

Figure 8 shows, as expected, that the optimal threshold value depends on the dataset used. Selecting a larger or smaller number of features may be beneficial or detrimental to the average test error, depending on the nature of the dataset studied. The general pattern is that there are datasets that benefit from selecting a lower number of features, and datasets that are penalized when dimensionality is excessively reduced. This is because some datasets obtain great advantage from implementing a feature selection process to remove noise before classification, whereas other datasets do not benefit from this kind of preprocessing. In this study, it can be seen that a feature selection step significantly improved results for the *Madelon* and *Spambase* datasets, but not so for the *Connect4*, *Isolet*, *USPS* and *Pixraw10P* datasets. For these datasets, the greatest benefit provided by the feature selection process was the improved training and processing times in the subsequent classification process.

In general, the *Fisher discriminant ratio* and the $\log_2(n)$ threshold values, compared to the other threshold values, considerably reduced dataset dimensionality (see Table 16), while holding classification performance at reasonable

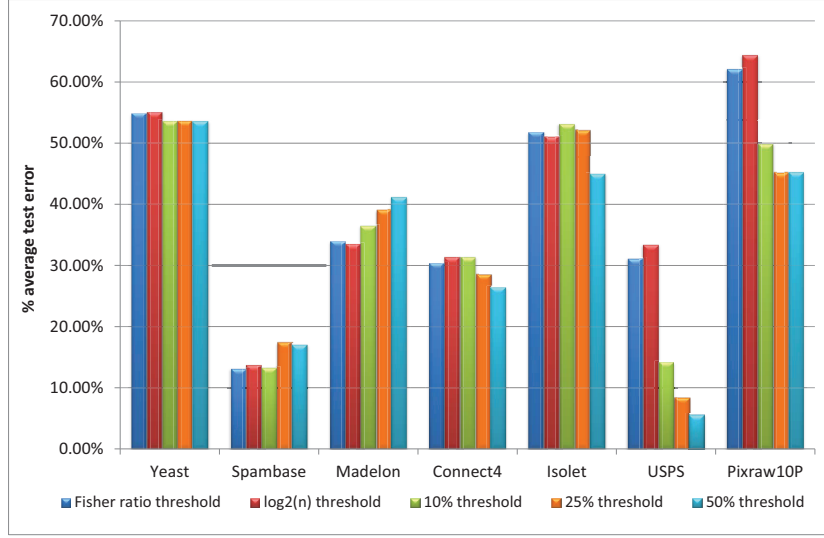


Figure 8: Comparison of average estimated percentage test errors for the different threshold values.

levels, and so can be rated as two satisfactory alternative approaches to limiting the final ranking. Although satisfactory average test error results were obtained for the 50% threshold, it might not be a very suitable threshold for large dimensionality datasets.

Dataset	Total	Fisher ratio	$\log_2(n)$	10%	25%	50%
Yeast	8	3	3	1	2	4
Spambase	57	6	6	6	15	29
Madelon	500	9	9	50	125	250
Connect4	42	10	5	5	11	21
Isolet	617	9	9	62	155	309
USPS	256	16	8	26	64	128
Pixraw10P	10 000	26	13	1 000	2 500	5 000

Table 16: Average number of features selected for the five threshold values.

5.4. Comparison with Subset Evaluation Methods

We performed a comparative analysis between the *E-RRA* heterogeneous centralized ensemble method and the *CFS* [53] subset evaluation method (Table 17). Different threshold values were selected to obtain the final practical subset for the ensemble approach, and several search methods were used for subset evaluation.

As can be seen in Table 17, the ensemble approach obtained results that were not significantly different than those of the subset approach in most of the

Method	Yeast	Spambase	Madelon	Connect4	Isolet	USPS	Pixraw10P
CFS (BestFirst)	49.94 [†] ± 4.53	18.32 [†] ± 2.18	34.00 [†] ± 3.44	31.12 ± 0.57	23.94 [†] ± 1.52	6.03 [†] ± 0.66	26.00 [†] ± 20.66
CFS (Forward)	49.94 [†] ± 4.53	17.86 [†] ± 2.06	34.00 [†] ± 3.44	31.12 ± 0.57	46.39 ± 2.36	8.20 [†] ± 1.13	30.00 [†] ± 23.78
CFS (Greedy-Backward)	49.94 [†] ± 4.53	18.36 [†] ± 2.20	34.00 [†] ± 3.44	31.12 ± 0.57	24.10 [†] ± 1.59	5.95 [†] ± 0.63	23.00 [†] ± 21.66
CFS (Genetic)	49.94 [†] ± 4.53	18.82 ± 1.84	35.96 [†] ± 2.76	30.88 ± 1.25	26.55 [†] ± 1.65	5.44 [†] ± 0.54	27.00 [†] ± 21.12
E-RRA (Fisher Ratio)	56.47 [†] ± 5.04	14.13 [†] ± 3.37	33.17 [†] ± 3.95	29.69 ± 2.00	50.30 ± 5.44	28.53 ± 4.06	59.00 [†] ± 21.32
E-RRA ($\log_2(n)$)	56.20 [†] ± 3.94	13.67 [†] ± 3.15	33.50 [†] ± 4.44	30.97 ± 0.78	50.97 ± 4.16	28.53 ± 0.00	70.00 ± 19.44
E-RRA (10%)	54.18 [†] ± 2.72	14.50 [†] ± 4.12	36.54 [†] ± 3.50	31.01 ± 0.44	50.29 ± 1.36	13.56 ± 1.13	47.00 [†] ± 24.52
E-RRA (25%)	54.18 [†] ± 2.72	17.04 [†] ± 2.65	39.75 [†] ± 2.91	27.81 [†] ± 0.76	50.07 ± 3.12	8.89 [†] ± 1.95	41.00 [†] ± 26.01
E-RRA (50%)	54.18 [†] ± 2.72	17.00 [†] ± 2.84	41.29 ± 3.11	26.26 [†] ± 0.74	42.41 [†] ± 3.28	5.70 [†] ± 0.67	41.00 [†] ± 31.31

Table 17: Traditional feature selection methods: average estimated percentage test errors. The superscript dagger shows results that were not significantly different than the best result.

experiments performed. The ensemble approach even achieved the lowest test error values for the *Spambase*, *Madelon*, and *Connect4* datasets, whereas the subset evaluation approach obtained its worst result for the *Connect4* dataset.

5.5. The Two Ensemble Approaches Compared

Finally, concluding this section is a graphical comparison between the homogeneous distributed and the heterogeneous centralized ensembles (Figure 9), based on using the two best heterogeneous centralized ensembles in terms of average test error (*E-Mean* and *E-RRA*), and the single best homogeneous distributed ensemble in terms of average training times (*SVM-RFE*). Note that the combination methods used by the homogeneous distributed ensemble were also the *Mean* and the *RRA* functions.

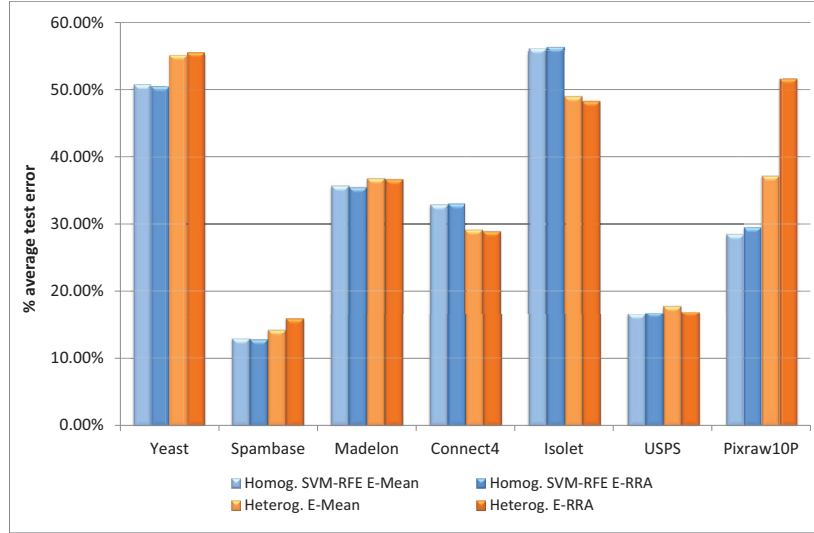


Figure 9: Comparison of average estimated percentage test errors for homogeneous distributed and heterogeneous centralized ensembles.

As can be seen in Figure 9, the homogeneous distributed ensemble obtained significantly better results than the heterogeneous centralized ensemble for the *Yeast* dataset – the smallest of all the datasets in terms of both size and dimension. As dataset size and dimension increased, however, the difference in test error between the two ensemble types diminished to the point of reversing the situation. Thus, the best result for the heterogeneous centralized ensemble was obtained for the *Connect4* and *Isolet* datasets; this is explained by the fact that the homogeneous approach can improve computation time—at the cost of a minimal reduction in accuracy—by distributing the sample over different nodes. Finally, when the *RRA* combination method was used, a very high average test error was returned for the *Pixraw10P* dataset, a result that was improved on slightly when the *Mean* combination method was used. As pointed out in Section 5.3, microarray datasets are especially susceptible to great accuracy variations depending on the combination method used.

To sum up, we propose applying the following rules-of-thumb:

- If the dataset is very large and a reduction in training time is crucial, the homogeneous distributed ensemble is the better option, since it considerably reduces training time while ensuring reasonable classification accuracy.
- If the dataset is reasonably small, or the user is uncertain as to which of the available algorithms to choose, the heterogeneous centralized ensemble is probably the better option, since it does not require the user to decide between feature selection methods, and may even, in some cases, improve classification accuracy.

6. Conclusions and Future Work

In recent years, the dimensions of the datasets used for machine learning have increased to the point where they can be very large, either/both in terms of number of samples/features; they also may contain redundancy, or be noisy, multivariate, or nonlinear. Since not all feature selection methods can tackle these issues, the user is faced with the difficult decision of which method to use in a particular situation.

As a solution to this problem, we have described two different designs of feature selection ensembles that take advantage of the combination of different individual methods. The homogeneous distributed ensemble is generated using the same feature selection method, and the dataset is then distributed over several nodes so as to reduce computational time by parallelizing the training task. The heterogeneous centralized ensemble, which consists of using different feature selection methods for the same training data, tries to take advantage of the strengths and overcome the weaknesses of the individual methods. The latter approach has the added benefit of freeing the user from the task of deciding which method best suits a particular scenario.

The two ensembles work with multiple feature ranking methods, so it is necessary to combine their different rankings, using an aggregator, to produce

a common final output. We experimented with seven datasets representative of medium- to large-sized problems, and were able to demonstrate competitive results for the ensembles—irrespective of the dataset and threshold—without incurring any deterioration in classification accuracy. The homogeneous distributed ensemble considerably improved training times over the individual methods and errors were stable between the two strategies. The heterogeneous centralized ensemble matched—and sometimes even improved on—the results achieved by the individual feature selection methods. We would suggest that the homogeneous distributed ensemble is particularly suitable for large datasets, while the heterogeneous centralized ensemble has the advantage of freeing the user from decision making regarding the best possible feature selection method for a given problem.

In comparing the different combination methods it was observed that choosing one or another method does not influence the final estimated error, at least for the six standard methods studied in this work. We also observed, for five of the seven studied datasets, that the *Fisher discriminant ratio* and the $\log_2(n)$ threshold values—compared to the percentage threshold values—reduced dataset dimensions considerably whilst maintaining reasonable classification accuracy, indicating these to be appropriate values for limiting feature rankings. Furthermore, the *Fisher discriminant ratio* threshold value, which depends on the nature of the dataset studied, obtains a threshold that is automatically tuned for each scenario. Our future research—along these lines—will focus on designs for the automatic selection of thresholds according to the characteristics of particular dataset types.

7. Acknowledgments

This research has been financially supported in part by the Spanish Ministerio de Economía y Competitividad (research project TIN 2015-65069-C2-1-R), by European Union FEDER funds, and by the Consellería de Industria of the Xunta de Galicia (research project GRC2014/035). V. Bolón-Canedo acknowledges Xunta de Galicia postdoctoral funding (ED481B 2014/164-0).

- [1] L. I. Kuncheva, C. J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Machine learning* 51 (2) (2003) 181–207.
- [2] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.
- [3] M. Bramer, *Principles of data mining*, Vol. 180, Springer, 2007.
- [4] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, *The Journal of Machine Learning Research* 5 (2004) 1205–1224.
- [5] I. Guyon, *Feature extraction: foundations and applications*, Vol. 207, Springer Science & Business Media, 2006.

- [6] C. Cortes, V. Vapnik, Support-vector networks, *Machine learning* 20 (3) (1995) 273–297.
- [7] T. Joachims, Optimizing search engines using clickthrough data, in: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2002, pp. 133–142.
- [8] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *The Journal of Machine Learning Research* 3 (2003) 1157–1182.
- [9] M. A. Hall, G. Holmes, Benchmarking attribute selection techniques for discrete class data mining, *Knowledge and Data Engineering, IEEE Transactions on* 15 (6) (2003) 1437–1447.
- [10] L. Breiman, Bagging predictors, *Machine learning* 24 (2) (1996) 123–140.
- [11] R. Schapire, The strength of weak learnability, *Machine learning* 5 (2) (1990) 197–227.
- [12] A. Tsymbal, M. Pechenizkiy, P. Cunningham, Diversity in search strategies for ensemble feature selection, *Information fusion* 6 (1) (2005) 83–98.
- [13] M. Lichman, UCI machine learning repository (2013).
URL <http://archive.ics.uci.edu/ml>
- [14] V. Bolón-Canedo, N. Sánchez-Marroño, A. Alonso-Betanzos, An ensemble of filters and classifiers for microarray data classification, *Pattern Recognition* 45 (1) (2012) 531–539.
- [15] F. Yang, K. Mao, Robust feature selection for microarray data based on multicriterion fusion, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 8 (4) (2011) 1080–1092.
- [16] T. Windeatt, R. Duangsoithong, R. Smith, Embedded feature ranking for ensemble mlp classifiers, *IEEE Transactions on Neural Networks* 22 (6) (2011) 988–994.
- [17] T. Windeatt, M. Prior, Stopping criteria for ensemble-based feature selection, in: *Multiple Classifier Systems*, Springer, 2007, pp. 271–281.
- [18] V. Bolón-Canedo, N. Sánchez-Marroño, A. Alonso-Betanzos, Data classification using an ensemble of filters, *Neurocomputing* 135 (2014) 13–20.
- [19] J. Olsson, D. W. Oard, Combining feature selectors for text classification, in: *Proceedings of the 15th ACM international conference on Information and knowledge management*, ACM, 2006, pp. 798–799.
- [20] H. Wang, T. M. Khoshgoftaar, K. Gao, Ensemble feature selection technique for software quality classification., in: *SEKE*, 2010, pp. 215–220.

- [21] H. Wang, T. M. Khoshgoftaar, A. Napolitano, A comparative study of ensemble feature selection techniques for software defect prediction, in: Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on, IEEE, 2010, pp. 135–140.
- [22] C.-H. Yang, C.-C. Huang, K.-C. Wu, H.-Y. Chang, A novel ga-taguchi-based feature selection method, in: Intelligent Data Engineering and Automated Learning–IDEAL 2008, Springer, 2008, pp. 112–119.
- [23] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, Y. Saeys, Robust biomarker identification for cancer diagnosis with ensemble feature selection methods, *Bioinformatics* 26 (3) (2010) 392–398.
- [24] A. Ben Brahim, M. Limam, Robust ensemble feature selection for high dimensional data sets, in: High Performance Computing and Simulation (HPCS), 2013 International Conference on, IEEE, 2013, pp. 151–157.
- [25] B. Seijo-Pardo, V. Bolón-Canedo, I. Porto-Díaz, A. Alonso-Betanzos, Ensemble feature selection for rankings of features, in: Advances in Computational Intelligence, Springer, 2015, pp. 29–42.
- [26] J. Park, I. W. Sandberg, Universal approximation using radial-basis-function networks, *Neural computation* 3 (2) (1991) 246–257.
- [27] M. Bramer, Data for data mining, in: Principles of data mining, Springer, 2013, pp. 9–19.
- [28] J. Quinlan, Induction of decision trees, *Machine learning* 1 (1) (1986) 81–106.
- [29] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27 (8) (2005) 1226–1238.
- [30] I. Kononenko, Estimating attributes: analysis and extensions of relief, in: Machine Learning: ECML-94, Springer, 1994, pp. 171–182.
- [31] K. Kira, L. A. Rendell, The feature selection problem: Traditional methods and a new algorithm, in: AAAI, 1992, pp. 129–134.
- [32] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine learning* 46 (1-3) (2002) 389–422.
- [33] M. Mejía-Lavalle, E. Sucar, G. Arroyo, Feature selection with a perceptron neural net, in: Proceedings of the international workshop on feature selection for data mining, 2006, pp. 131–135.
- [34] S. B. Lysterly, The average spearman rank correlation coefficient, *Psychometrika* 17 (4) (1952) 421–428.

- [35] H. Abdi, The kendall rank correlation coefficient, *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA (2007) 508–510.
- [36] R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis, *Bioinformatics* 28 (4) (2012) 573–580.
- [37] P. Willett, Combination of similarity rankings using data fusion, *Journal of chemical information and modeling* 53 (1) (2013) 1–10.
- [38] S. Aerts, D. Lambrechts, S. Maity, P. Van Loo, B. Coessens, F. De Smet, L.-C. Tranchevent, B. De Moor, P. Marynen, B. Hassan, et al., Gene prioritization through genomic data fusion, *Nature biotechnology* 24 (5) (2006) 537–544.
- [39] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, Recent advances and emerging challenges of feature selection in the context of big data, *Knowledge-Based Systems* 86 (2015) 33–45.
- [40] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, *Knowledge and information systems* 34 (3) (2013) 483–519.
- [41] T. M. Khoshgoftaar, M. Golawala, J. Van Hulse, An empirical study of learning from imbalanced data using random forest, in: *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*, Vol. 2, IEEE, 2007, pp. 310–317.
- [42] M. Mejía-Lavalle, E. Sucar, G. Arroyo, Feature selection with a perceptron neural net, in: *Proceedings of the international workshop on feature selection for data mining, 2006*, pp. 131–135.
- [43] M. Basu, T. K. Ho, *Data complexity in pattern recognition*, Springer Science & Business Media, 2006.
- [44] L. Morán-Fernández, V. Bolón-Canedo, A. Alonso-Betanzos, A time efficient approach for distributed feature selection partitioning by features, in: *Advances in Artificial Intelligence, Springer, 2015*, pp. 245–254.
- [45] B. Seijo-Pardo, V. Bolón-Canedo, A. Alonso-Betanzos, Using a feature selection ensemble on DNA microarray datasets, in: *Proceeding of 24th European Symposium on Artificial Neural Networks, 2016*, pp. 277–282.
- [46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The weka data mining software: an update, *ACM SIGKDD explorations newsletter* 11 (1) (2009) 10–18.
- [47] C. J. Burges, A tutorial on support vector machines for pattern recognition, *Data mining and knowledge discovery* 2 (2) (1998) 121–167.
- [48] B. Schölkopf, C. J. Burges, A. J. Smola, *Advances in kernel methods: support vector learning*, MIT press, 1999.

- [49] V. Vapnik, The nature of statistical learning theory, Springer Science & Business Media, 2000.
- [50] J. Li, K. Cheng, S. Wang, F. Morstatter, R. Trevino, J. Tang, H. Liu, Feature selection: A data perspective, arXiv preprint arXiv:1601.07996.
- [51] J. W. Tukey, Comparing individual means in the analysis of variance, *Biometrics* (1949) 99–114.
- [52] C. Eiras-Franco, V. Bolón-Canedo, S. Ramos, J. González-Domínguez, A. Alonso-Betanzos, J. Touriño, Paralelización de algoritmos de selección de características en la plataforma weka, in: CAEPIA 2015 (Workshop BigDADE), 2015, pp. 949–958, "<http://simd.albacete.org/actascaepia15>. Last access: 03-06-2016".
- [53] M. A. Hall, Correlation-based feature selection of discrete and numeric class machine learning.