Associative learning on imbalanced environments: An empirical study

L. Cleofas-Sánchez[a], J.S. Sánchez[a,], V. García[b], R.M. Valdovinos[c]

[a]Institute of New Imaging Technologies, Department of Computer Languages and Systems,
Universitat Jaume I, Castelló de la Plana, Spain
E-mail: cleofas@uji.es, sanchez@uji.es
[b]División Multidisciplinaria de Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez,
Ciudad Juárez, Chihuahua, Mexico
E-mail: vicente.jimenez@uacj.mx
[c]School of Engineering, Universidad Autónoma del Estado de México, Toluca, Mexico
E-mail: li_rmvr@hotmail.com

Abstract

Associative memories have emerged as a powerful computational neural network model for several pattern classification problems. Like most traditional classifiers, these models assume that the classes share similar prior probabilities. However, in many real-life applications the ratios of prior probabilities between classes are extremely skewed. Although the literature has provided numerous studies that examine the performance degradation of renowned classifiers on different imbalanced scenarios, so far this effect has not been supported by a thorough empirical study in the context of associative memories. In this paper, we fix our attention on the applicability of the associative neural networks to the classification of imbalanced data. The key questions here addressed are whether these models perform better, the same or worse than other popular classifiers, how the level of imbalance affects their performance, and whether distinct resampling strategies produce a different impact on the associative memories. In order to answer these questions and gain further insight into the feasibility and efficiency of the associative memories, a large-scale experimental evaluation with 31 databases, seven classification models and four resampling algorithms is carried out here, along with a non-parametric statistical test to discover any significant differences between each pair of classifiers.

Keywords: Associative memory, Class imbalance, Resampling

1. Introduction

Associative memories have been an active research area in pattern recognition and neuroscience for over 50 years (Palm, 2013). Although typical applications of these connectionist models include image recognition and recovery, data analysis, control, inference and prediction (Danilo et al., 2015; Kareem and Jantan, 2011; Lou and Cui, 2007; Mu et al., 2006; Nazari et al., 2014; Štanclová and Zavoral, 2005), the associative memories have lately emerged as useful classifiers for a large variety of problems in data mining and computational intelligence (Aldape-Pérez et al., 2012, 2015; Sharma et al., 2008; Uriarte-Arcia et al., 2014).

From a practical point of view, classification refers to the assignment of a finite set of samples to predefined classes based on a number of observed variables or attributes. The effective design of classifiers is a complex task where the underlying quality of data becomes critical to further achieve accurate classifications. Besides other factors, the performance of classifiers strongly depends on the adequacy of the data in terms of several intrinsic characteristics such as number of examples, relevance of the attributes, class distribution and completeness of data. In particular, a very common situation in real-life classification problems is to find data sets where the amount of examples available for one class is quite different from that of the other classes. This significant difference in the size of the classes is usually known as class imbalance, and it has been observed that most traditional classifiers perform poorly on the minority class because they assume an even class distribution and equal misclassification costs (Japkowicz and Stephen, 2002).

Examples of applications with skewed class distributions can be found in many different areas ranging from bioinformatics and medicine to finance, network security and text mining. For instance, in credit risk and bankruptcy prediction, the number of observations in the class of defaulters is much smaller than the number of cases belonging to the class of non-defaulters (Kim et al., 2015). Gene expression microarray analysis for cancer classification shows significant differences between the number of cancerous and healthy tissue samples (Blagus and Lusa, 2010). Web spam detection exhibits an asymmetric distribution between legitimate and spam pages (Fdez-Glez et al., 2015).

Within this context, the major research line has traditionally been directed to the development of techniques to tackle the class imbalance at both algorithmic and data levels (He and Garcia, 2009; He and Ma, 2013; López et al., 2013). The methods at the algorithmic level modify the existing learning models for biasing the discrimination process towards the minority class; the data level solutions consist of a preprocessing step to resample the original data set, either by over-

sampling the minority class and/or under-sampling the majority class, until the classes are approximately equally represented. In general, the resampling strategies have been the most widely used because they are independent of the classifier and can be easily implemented for any problem (Cao et al., 2014; García et al., 2012).

Over the last years, however, research on this topic has also put the emphasis on studying the effect of imbalance together with other data complexity characteristics such as overlapping, small disjuncts and noisy data (He et al., 2015; López et al., 2013; Napierala et al., 2010; Prati et al., 2004; Stefanowski, 2013). Another critical subject that has attracted increasing interest in the scientific community is how to assess the performance of a classification model in the presence of imbalanced data sets because most common metrics (e.g., accuracy and error rates) strongly depend on the class distribution and assume equal misclassification costs, which may lead to distorted conclusions (He and Garcia, 2009; Menardi and Torelli, 2014).

In addition to the aforementioned questions, research has also placed its attention on evaluating and comparing the performance of competing classifiers (Brown and Mues, 2012; Kwon and Sim, 2013; López et al., 2013; Liu et al., 2011; Seiffert et al., 2014), but many of these works do not take into consideration the impact of different levels of imbalance on the performance of each particular classification model, neither the implications of using this jointly with one resampling technique or another.

While research in class imbalance has mainly concentrated on well-known classifiers such as support vector machines (Akbani et al., 2004; Hwang et al., 2011; Liu et al., 2011; Maldonado and López, 2014; Yu et al., 2015), kernel methods (Hong et al., 2007; Maratea et al., 2014), k-nearest neighbors (Dubey and Pudi, 2013), decision trees (Kang and Ramamohanarao, 2014) and multiple classifier systems (Díez-Pastor et al., 2015; Galar et al., 2012; Krawczyk et al., 2014; Park and Ghosh, 2014), very few theoretical or empirical analyses have been done so far to thoroughly establish the performance of associative memories when learning from class imbalanced data. Therefore, the present study intends to extend the very preliminary existing works (Cleofas-Sánchez et al., 2013, 2014) by increasing the scope and detail at which a type of associative memory networks (the hybrid associative classifier with translation) performs in the framework of class imbalance. We believe that the extensive experimental analysis here carried out will allow to gain a deeper insight into the feasibility and efficiency of these associative memories and help researchers and practitioners to build effective associative learning models and develop techniques based on associative learning

to handle the class imbalance problem. To sum up, the purpose of this paper is three-fold:

- To explore the performance of the hybrid associative memory with translation and compare this against other popular classification methods of different nature;

- to investigate how the imbalance ratio affects the performance of these associative memories; and

- to analyze the impact of several resampling strategies on the performance of the associative neural network.

The rest of this paper is organized as follows. Section 2 introduces the bases of the associative memory neural networks and in particular, of the hybrid associative classifier with translation. Section 3 briefly describes four resampling methods to deal with the imbalance problem, which will be further used in the experiments. The experimental set-up and databases are presented in Section 4, while the results and statistical tests are discussed in Section 5. Finally, Section 6 remarks the main conclusions and outlines some avenues for future research.

2. Associative memories

The associative memory is an early type of artificial neural network that takes the form of a matrix $\mathbf{M}$ generated from a finite set of p previously known associations, which is called fundamental set $\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \ldots, p\}$, where $\mathbf{x}^\mu$ are the fundamental input patterns of dimension n and $\mathbf{y}^\mu$ are the the fundamental output patterns of dimension m. Then, x_j^μ and y_j^μ denote the j-th component of an input pattern $\mathbf{x}^\mu$ and of an output pattern $\mathbf{y}^\mu$, respectively.

Functionality of the associative memories is accomplished in two phases: learning and recall. The learning process consists of building a matrix $\mathbf{M}$ with a value for each association $(\mathbf{x}^k, \mathbf{y}^k)$. In the recall phase, an output vector $\mathbf{y}$ is obtained from the associative memory; this vector is the most similar to the input vector $\mathbf{x}$. These memories have the capability of storing huge amounts of data that allow to recover the input examples with low computational efforts.

The associative memories can be of two types (Aldape-Pérez et al., 2012): heteroassociative (e.g., linear associator) and autoassociative (e.g., Hopfield network). The heteroassociative memories relates input patterns with output patterns

of distinct nature and formats ($\mathbf{x}^\mu \neq \mathbf{y}^\mu$), while the autoassociative memories are a particular case where $\mathbf{x}^\mu = \mathbf{y}^\mu$ and $n = m$.

Some of the most widely-studied models of associative memories are lernmatrix (Steinbuch, 1961), the linear associator (Anderson, 1972; Kohonen, 1972), the Hopfield network (Hopfield, 1982), the bidirectional associative memory (Kosko, 1987), and the morphological associative memory (Ritter et al., 1998). The popularity of these models comes from their capability of storing huge amounts of data that allow to properly recover the most similar patterns given an input vector with low computational efforts, but they have some difficulties to discriminate between classes.

2.1. The hybrid associative classifier with translation

With the aim of extending the applicability of classical associative memories to classification tasks, Santiago (2003) introduced the hybrid associative classifier with translation (HACT), which is based on the learning phase of the linear associator and the recall phase of the lernmatrix. The learning phase of the linear associator comprises two steps:

1. For each association $(\mathbf{x}^\mu, \mathbf{y}^\mu)$, compute the matrix $(\mathbf{y}^\mu) \cdot (\mathbf{x}^\mu)^T$, where $(\mathbf{x}^\mu)^T$ is the transpose input vector.

2. Sum the p matrices to obtain the memory $\mathbf{M} = \sum_{\mu=1}^{p} (\mathbf{y}^\mu) \cdot (\mathbf{x}^\mu)^T$, being $m_{i,j} = \sum_{\mu=1}^{p} (y_i^\mu)(x_j^\mu)^T$ its (i,j)-th component.

On the other hand, the recall phase of the lernmatrix consists of finding the class of an input pattern $\mathbf{x}^\mu$, that is, to find the components of the vector $\mathbf{y}^\mu$ associated to $\mathbf{x}^\mu$, whose i-th component is calculated according to the following expression:

$$
y_i^\mu = \begin{cases} 1 & \text{if} \quad \sum_{j=i}^{n} m_{i,j} x_j^\mu = \max_{h=1}^{p} \left[\sum_{j=i}^{n} m_{h,j} x_j^\mu \right] \\ 0 & \text{otherwise} \end{cases} \tag{1}
$$

Note that if $\mathbf{x}^\mu$ belongs to class c, this expression leads to an m-dimensional vector with all components equal to zero except for the c-th component whose value is equal to 1.

In addition, the HACT model incorporates the translation of the axes to a new origin located at the centroid of the fundamental input patterns. Let $A = \{\mathbf{x}^\mu \mid \mu = 1, 2, \ldots, p\}$ be a set of n-dimensional fundamental input patterns that belong to

Algorithm 1 *HACT – learning phase*

$s \leftarrow 0$
for all $\mathbf{x}^\mu \in A$ **do**
  $s \leftarrow s + \mathbf{x}^\mu$
end for
$\overline{\mathbf{x}} \leftarrow s/p$ $\{\overline{x}$ is the new origin of the coordinates axes$\}$
for all $\mathbf{x}^\mu \in A$ **do**
  $\hat{\mathbf{x}}^\mu \leftarrow \mathbf{x}^\mu - \overline{\mathbf{x}}$ $\{$Translate the input patterns to the new coordinates axes$\}$
  $c \leftarrow 1$
  $\{$Assign an output vector of size m to patterns that belong to class $c\}$
  while $c \leq m$ **do**
    if $class(\hat{\mathbf{x}}^\mu) = c$ **then**
      $y_c^\mu = 1$
    else
      $y_c^\mu = 0$
    end if
  end while
end for
$\mathbf{M} \leftarrow 0$ $\{$Apply the learning phase of the linear associator$\}$
for all $(\hat{\mathbf{x}}^\mu, \mathbf{y}^\mu)$ **do**
  $\mathbf{M} \leftarrow \mathbf{M} + (\mathbf{y}^\mu) \cdot (\hat{\mathbf{x}}^\mu)^T$
end for

---

m classes, then the learning phase of HACT to construct the matrix $\mathbf{M}$ consists in the steps of Algorithm 1.

Once the matrix $\mathbf{M}$ has been generated, the classification of a new input pattern $\mathbf{x}$ consists of applying the recovery phase of lernmatrix according to expression 1.

3. Resampling methods to handle class imbalance

Much research has been conducted to tackle the class imbalance problem by applying different resampling strategies that change the class distribution of the data set (Yang et al., 2011). This can be done by either over-sampling the minority class or under-sampling the majority class until both classes are approximately equally represented. Despite their advantages and popularity, several shortcomings remain associated with both these strategies because they artificially alter the prior class probabilities (Barandela et al., 2004): whilst under-sampling may result in throwing away potentially useful information about the majority class and perturbing the a priori probability of the training set (Dal Pozzolo et al., 2015),

over-sampling worsens the computational burden of most classifiers, may increase the likelihood of overfitting and introduce noise that could result in a loss of performance.

3.1. Over-sampling

The simplest method to increase the size of the minority class corresponds to random over-sampling (ROS), which balances the class distribution through the random replication of minority class examples. Although this method appears to be quite effective without adding new information into the data set, it may increase the likelihood of overfitting because it makes exact copies of the minority class examples.

In order to avoid overfitting, Chawla et al. (2002) proposed the Synthetic Minority Oversampling TEchnique (SMOTE) to up-size the minority class. Instead of merely replicating examples, this algorithm generates artificial examples of the minority class by interpolating existing instances that lie close together. It first finds the k positive nearest neighbors for each minority class example and then, the synthetic examples are generated in the direction of some or all of those nearest neighbors. SMOTE allows the classifier to build larger decision regions that contain nearby examples of the minority class. Depending upon the amount of over-sampling required, a certain number of examples from the k nearest neighbors are randomly chosen.

3.2. Under-sampling

Random under-sampling (RUS) aims at balancing the data set through the random removal of examples that belong to the over-sized class. Despite important information can be lost when examples are discarded at random, it has empirically been shown to be one of the most effective resampling methods.

Unlike the random approach, other under-sampling methods are based on a deterministic selection of the examples to be eliminated. For instance, Laurikkala (2001) introduced the Neighborhood CLeaning rule (NCL), which uses the Wilson's editing algorithm (Wilson, 1972) to remove only the majority class cases that border the minority class examples, assuming that they are noise.

4. Experimental set-up

With the aim of evaluating the performance of the HACT model on imbalanced domains, we have carried out a series of experiments on a large collection of imbalanced data sets using four resampling algorithms (ROS, SMOTE, RUS,

and NCL) to preprocess the original data sets and seven classifiers implemented with the default parameter settings given by the WEKA toolkit (Hall et al., 2009): the Bayes network (BNet), a multilayer perceptron network with backpropagation (MLP), a normalized Gaussian radial basis function (RBF), a support vector machine with a linear kernel(SVM), a pruned decision tree (C4.5), the nearest neighbor with generalization (NNge), and the voted perceptron (VP). Note that we have chosen classifiers of different nature and complexity in order to draw meaningful conclusions, including four neural networks (BNet, MLP, RBF, VP), a linear classifier (SVM), a decision tree (C4.5), and an instance-based learner (NNge).

4.1. Data sets

The experimental set-up has been designed following the common practice (Akbani et al., 2004; Barua et al., 2014; Díez-Pastor et al., 2015; Galar et al., 2013; López et al., 2012, 2013; Yu et al., 2013, 2015), which consists of generating skewed data sets with different levels of class imbalance. In particular, the scheme proposed in the literature is to transform multi-class data sets by combining several original classes to shape the majority and minority classes. For instance, Glass0123_vs_456 is an imbalanced version of Glass database where the classes 0, 1, 2 and 3 have been combined to form a unique minority class and the original classes 4, 5 and 6 have been joined to represent the majority class. Similarly, Glass6 has been created by leaving the original class 6 as the minority class and merging the rest of the classes into a single majority class.

The results are based on 31 imbalanced data sets taken from the KEEL Data Set Repository (Alcalá-Fdez et al., 2011), whose main characteristics are summarized in Table 1. These data sets have been divided into two groups according to their imbalance ratio (IR), that is, the ratio of the majority class size to the minority class size. A first group comprises data sets with a low/moderate imbalance ratio (IR < 10), while the second category gathers the strongly imbalanced databases (IR ≥ 10).

The five-fold cross-validation method has been adopted for the experimental design because it seems to be the best estimator of classification performance compared to other methods, such as bootstrap with a high computational cost or re-substitution with a biased behavior. Each original data set has randomly been divided into five stratified parts of of size $n/5$ (where n denotes the total number of samples in the data set); for each fold, four blocks have been pooled as the training set, and the remaining portion has been used as an independent test set.

Table 1: Data sets used in the experiments

Data set	#Samples	#Attributes	IR	Data set	#Samples	#Attributes	IR
Wisconsin	683	9	1.86	Ecoli0347_vs_56	257	7	9.28
Yeast1	1484	8	2.46	Yeast05679_vs_4	528	8	9.35
Haberman	306	3	2.78	Ecoli067_vs_5	220	6	10.00
Vehicle1	846	18	2.90	Led7digit02456789_vs_1	443	7	10.97
Vehicle3	846	18	2.99	Ecoli01_vs_5	240	6	11.00
Glass0123_vs_456	214	9	3.20	Yeast1_vs_7	459	7	14.30
Ecoli2	336	7	5.46	Ecoli4	336	7	15.80
Ecoli1	336	7	3.36	Glass4	214	9	15.47
Segment0	2308	19	6.02	Yeast1458_vs_7	693	8	22.10
Glass6	214	9	6.38	Glass5	214	9	22.78
Yeast3	1484	8	8.10	Yeast2_vs_8	482	8	23.10
Ecoli3	336	7	8.60	Yeast4	1484	8	28.10
Ecoli034_vs_5	200	7	9.00	Yeast1289_vs_7	947	8	30.57
Yeast0256_vs_3789	1004	8	9.14	Ecoli0137_vs_26	281	7	39.14
Ecoli046_vs_5	203	6	9.15	Yeast6	1484	8	41.40
Ecoli0346_vs_5	205	7	9.25				

The classification models have been applied to the sets preprocessed by the resampling strategies and also to each original training set, which can be deemed as a baseline for comparison. The results from classifying the test samples have been averaged across the five runs and then evaluated for significant differences using the non-parametric Wilcoxon signed-rank test (Demšar, 2006).

4.2. Evaluation criteria

Although the accuracy (and its counterpart, the error rate) is the most commonly used metric to evaluate the performance of classification systems, it has been demonstrated not to be appropriate when the prior class probabilities are very different because it does not consider misclassification costs, is strongly biased to favor the majority class, and is very sensitive to class skews (Daskalaki et al., 2006; Japkowicz, 2013). Therefore, the effectiveness of classifiers in imbalanced domains has traditionally been evaluated by using other criteria, such as precision, area under the ROC curve, F-measure, and geometric mean of accuracies.

In this paper, the geometric mean of accuracies has been chosen because it represents a good trade-off between the rates measured separately on each class.

$$Gmean = \sqrt{TPrate \cdot TNrate} \qquad (2)$$

where $TPrate$ and $TNrate$ denote the accuracy on the minority and majority classes, respectively.

5. Experimental results and discussion

The purpose of these experiments is three-fold: (i) to evaluate the performance of the HACT model in comparison to that of other classifiers when the data set is imbalanced; (ii) to find out whether there exist differences in its behavior depending on the imbalance ratio; and (iii) to determine how the application of resampling strategies affects the effectiveness of the HACT classifier. At this point, we believe that the degree of difficulty of classification of imbalanced data sets depends on their class imbalance ratio. This hypothesis can be extended to the resampling techniques and to the different classifiers. This is the reason why the experiments and analysis have been performed according to two levels of imbalance.

The $Gmean$ results have been tested for statistically significant differences by means of non-parametric tests, which are generally preferred over the parametric methods because the usual assumptions of independence, normality and homogeneity of variance are often violated due to the non-parametric nature of the problems (Demšar, 2006). Both pairwise and multiple comparisons have been used in this paper. First, we have computed the Friedman's ranking of the classifiers for each resampling strategy and each data set independently according to the $Gmean$ results: as there are eight competing classifiers, the ranks for each data set go from 1 (best) to 8 (worst); in case of ties, average ranks are assigned. Then the average rank of each model across all data sets has been calculated.

Afterwards, the Wilcoxon's paired signed-rank test has been used to find out statistically significant differences between each pair of classifiers for two levels of significance ($\alpha = 0.10$ and $\alpha = 0.05$). This statistic ranks the differences in performances of two algorithms for each data set, ignoring the signs, and compares the ranks for the positive and the negative differences.

5.1. Results on databases with low/moderate imbalance

The detailed results over each problem are given in the Appendix. Here the Friedman's average ranks for the eight classification models have been plotted in Figure 1. For the original training sets, the MLP and NNge classifiers arise as the algorithms with the lowest rankings, that is, the highest performance in average; when using the under-sampling strategies to balance the data sets, it seems that the best classifiers depend on the particular algorithm used. With the over-sampling techniques, SVM yields the lowest rankings. Despite the HACT model has never obtained the lowest average ranks, it is worth noting that their rankings are not too far from those of the best classifiers.
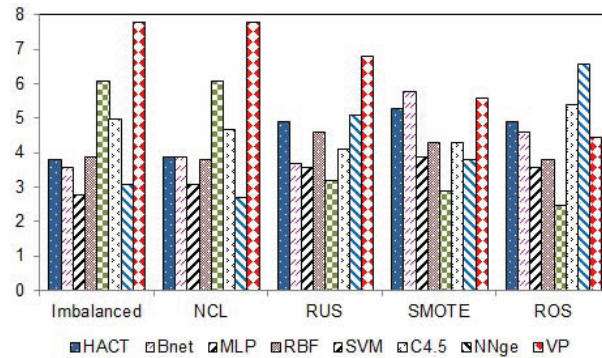
Figure 1: Friedman's average ranks for databases with low/moderate imbalance

Tables 2–4 report the Wilcoxon's test for the data sets with a low/moderate imbalance ratio. The upper diagonal half of each table corresponds to a significance level of $\alpha = 0.10$ (10% or less chance), whereas the lower diagonal half is for $\alpha = 0.05$. The symbol "•" indicates that the classifier in the row significantly outperforms the classifier in the column, and the symbol "○" means that the classifier in the column performs significantly better than the classifier in the row.

Table 2: Wilcoxon's statistic for the original training sets with low/moderate imbalance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–				•			•
(2) BNet		–	○		•			•
(3) MLP			–	•	•	•		•
(4) RBF			○	–	•			•
(5) SVM	○	○	○	○	–	○	○	•
(6) C4.5			○		•	–	○	•
(7) NNge					•	•	–	•
(8) VP	○	○	○	○	○	○	○	–

Results in Table 2 show that the HACT model performs significantly better than the SVM and VP classifiers at both significance levels when the imbalanced training set has not been preprocessed. In this scenario, the performance of the MLP neural network is significantly better than that of the RBF, SVM, C4.5 and VP algorithms, while NNge outperforms SVM, C4.5 and VP. Although both MLP and NNge appears to be the best alternatives when data sets present a low or moderate imbalance ratio, it has to be noted that the differences between these

11

two classifiers and HACT are not statistically significant and their computational cost is high when the data set size is large.

In Tables 3–4, one can observe that the behavior of HACT varies considerably depending on whether the data sets are under- or over-sampled: while the removal of majority class examples seems not to change its performance significantly, the over-sampling algorithms degrade the effectiveness of the HACT classifier, especially with the SMOTE method (RBF, SVM and NNge perform significantly better than HACT for a significance level of $\alpha = 0.10$).

Table 3: Wilcoxon's statistic for the under-sampled sets with low/moderate imbalance

	NCL								RUS							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–							●	–				○			●
(2) BNet		–	○		●			●		–					●	●
(3) MLP			–		●	●		●			–	●			●	●
(4) RBF				–	●	●		●			○	–	○			●
(5) SVM		○	○	○	–	○	○	●	●			●	–		●	●
(6) C4.5			○		●	–	○	●						–		●
(7) NNge					●	●	–	●			○		○		–	
(8) VP	○	○	○	○	○	○	○	–	○	○	○	○	○	○	○	–

Table 4: Wilcoxon's statistic for the over-sampled sets with low/moderate imbalance

	SMOTE								ROS							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–			○	○		○		–				○		●	
(2) BNet		–	○	○	○	○	○			–	○	○	○		●	
(3) MLP		●	–					●	●		–			●	●	
(4) RBF		●		–				●	●			–	○	●	●	
(5) SVM	●	●			–	●		●	●	●		●	–	●	●	●
(6) C4.5		●				–		●			○	○	○	–	●	
(7) NNge		●					–	●	○	○	○	○	○	○	–	
(8) VP				○	○		○	–					○			–

The three values in the cells of Table 5 show how many times the method of the column has been significantly-better/same/significantly-worse, with a significance level of $\alpha = 0.05$, for the databases with a low/moderate imbalance rate. As can be observed, the MLP neural network performs the best with both the imbalanced sets and the under-sampled sets, while the SVM appears to outperform the other

classifiers when the data sets have been over-sized. The HACT model does not present significant differences with the best approaches when using the original training set and the sets under-sampled with the NCL algorithm and therefore, it can considered as one of the best alternatives. However, the use of SMOTE and ROS seems to produce some degradation on the performance of the HACT classifier.

Table 5: Summary of the Wilcoxon's statistic for the sets with low/moderate imbalance

	HACT	Bnet	MLP	RBF	SVM	C4.5	NNge	VP
Original	2/5/0	2/5/0	4/3/0	2/4/1	1/0/6	2/3/2	3/4/0	0/0/7
NCL	1/6/0	2/5/0	3/4/0	2/5/0	1/1/5	2/3/2	3/4/0	0/0/7
RUS	1/5/1	1/6/0	3/4/0	1/4/2	4/3/0	1/5/0	1/4/2	0/0/7
SMOTE	0/6/1	0/2/5	1/6/0	2/5/0	3/4/0	1/6/0	2/5/0	0/4/3
ROS	1/5/1	1/3/3	3/4/0	3/3/1	6/1/0	1/4/3	0/1/6	0/6/1

5.2. Results on databases with high imbalance

Figure 2 shows the Friedman's average ranks for the eight classification models under a high imbalance ratio. For the imbalanced sets, the lowest rankings correspond to the MLP, Bnet and HACT neural networks. Although there is not a best performing classifier with the resampling algorithms, it seems that the HACT model is one of the approaches with lower rankings independently of the technique used.
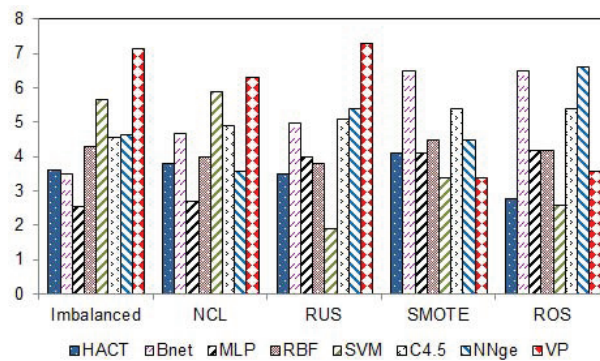

Figure 2: Friedman's average ranks for strongly imbalanced databases

13

The Wilcoxon's test reported in Table 6 reveals that the performance of HACT is significantly better than that of SVM, NNge and VP at both significance levels when the original training sets are strongly imbalanced. In this sense, its behavior seems to be similar to that of MLP since this classifier performs significantly better than RBF, SVM, C4.5, NNge and VP.

Table 6: Wilcoxon's statistic for the original training sets with high imbalance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–				•	•	•	•
(2) BNet		–			•			•
(3) MLP			–	•	•	•	•	•
(4) RBF			○	–	•			•
(5) SVM	○	○	○	○	–	○	○	•
(6) C4.5			○			–		•
(7) NNge	○		○		•		–	•
(8) VP	○	○	○	○	○	○	○	–

When the data sets with a high imbalance rate are under-sampled with the NCL algorithm, Table 7 shows that the HACT model achieves significantly better results than SVM and VP at a significance level of 0.05. The MLP neural network performs significantly better than BNet, RBF, SVM, NNge and VP, but we do not find statistically significant differences between the performance of MLP and that of HACT. In the case of random under-sampling, the SVM appears to be the best performing classifier since its $Gmean$ results are significantly better than those of the remaining algorithms for $\alpha = 0.05$. Nevertheless, the HACT neural network is still better than VP and is not significantly worse than the other classifiers.

Table 7: Wilcoxon's statistic for the under-sampled sets with high imbalance

	NCL								RUS							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–				•		•	•	–	•			○			•
(2) BNet		–	○				○			–			○			•
(3) MLP		•	–	•	•	•	•	•			–		○			•
(4) RBF			○	–	•			•				–	○		•	•
(5) SVM	○		○	○	–		○		•	•	•	•	–	•	•	•
(6) C4.5						–		•					○	–		•
(7) NNge			○		•		–	•				○	○		–	•
(8) VP	○		○	○			○	–	○		○	○	○	○	○	–

On the other hand, when the training sets are over-sampled by means of the SMOTE method (see Table 8), it seems that no classifier performs significantly the best for any value of α. In this case, the BNet approach clearly corresponds to the worst model since its performance is significantly lower than that of any other classifier. With random over-sampling, the conclusions are similar to those drawn for the RUS algorithm, that is, the SVM seems to be the best performing method. However, in this scenario, the performance of HACT is statistically the same as that of the SVM and better than that of BNet and NNge.

Table 8: Wilcoxon's statistic for the over-sampled sets with high imbalance

| | SMOTE | | | | | | | | ROS | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) HACT	–	●			○	●			–	●				●	●	
(2) BNet	○	–	○	○	○	○	○		○	–	○	○	○	○		○
(3) MLP		●	–			●				●	–		○		●	
(4) RBF		●		–						●		–	○		●	
(5) SVM		●			–	●			●	●	●		–	●	●	
(6) C4.5		●	○			–				●			○	–	●	
(7) NNge							–		○		○	○	○	○	–	○
(8) VP		●						–		●					●	–

Table 9 summarizes how many times the method of the column has been significantly-better/same/significantly-worse, with a significance level of $\alpha = 0.05$, for the strongly imbalanced databases. In this case, the HACT model is among the best performing methods, especially with the original training set and the over-sampled sets. With the RUS algorithm, the SVM is clearly the best alternative because it has been significantly better than all the remaining classifiers. On the other hand, VP and Bnet appear to be the worst models: the former with the under-sampling techniques and the latter with the over-sampling algorithms.

Table 9: Summary of the Wilcoxon's statistic for the sets with high imbalance

	HACT	Bnet	MLP	RBF	SVM	C4.5	NNge	VP
Original	3/4/0	2/5/0	5/2/0	2/4/1	1/1/5	1/5/1	2/3/2	0/0/7
NCL	2/5/0	0/6/1	5/2/0	2/4/1	0/3/4	0/7/0	2/4/1	0/3/4
RUS	1/5/1	0/6/1	1/5/1	2/4/1	7/0/0	1/5/1	1/4/2	0/1/6
SMOTE	1/6/0	0/1/6	2/5/0	1/6/0	1/6/0	1/5/1	0/7/0	1/6/0
ROS	2/5/0	0/1/6	2/4/1	2/4/1	5/2/0	2/4/1	0/1/6	1/5/1

6. Conclusions and future work

This paper intends to cover some questions on the application of the HACT neural network, which is a type of autoassociative memory, to problems characterized by skewed class distributions. Although earlier related works have given some insights into the effects of the class imbalance problem on this model, various interesting issues have not been studied in depth. Taking this into account, the main contributions can be summarized as follows: (i) we have empirically analyzed the performance of the HACT model in comparison to that of several standard classifiers; (ii) we have discussed the behavior of these classifiers as a function of the imbalance ratio; and (iii) we have studied how the use of different resampling techniques may affect the performance of the associative memory.

Experimental results on a large collection of imbalanced data sets using four resampling techniques and seven different classifiers allow to remark some important findings that were not discussed in the few preliminary works:

- When the imbalanced training set is not preprocessed by some resampling algorithm, the HACT model performs as well as the best classifier in most situations;

- the superiority of HACT with respect to other classifiers is more significant in the case of strongly imbalanced data sets, both on the original (imbalanced) training sets and on the balanced sets;

- the resampling strategies affect differently the performance of HACT depending on the imbalance ratio. Thus, when the imbalance is low or moderate, the over-sampling methods degrade the effectiveness of the HACT model considerably. However, with strongly imbalanced data, both under- and over-sampling have given rise to very similar results.

Despite the promising results in terms of the geometric mean of acuracies exhibited by the HACT model and supported by non-parametric statistical tests, there are still several interesting open research questions to work on. For instance, the present paper has focused on only the class imbalance problem, but a thorough and extensive analysis of the behavior of these neural networks when data sets are affected by multiple complexities deserves further future consideration. Another direction for extending this study is to investigate the applicability of the associative memories on imbalanced big data and data streams in a time-varying environment (concept change and concept drift), which are currently two hot topics in the literature related to expert and intelligent information systems.

16

Acknowledgment

Appendix

This appendix provides five tables with the $Gmean$ results for the experimental analysis carried out in the present work. Table 10 contains the values for all the databases and classifiers achieved when using the imbalanced training sets, Table 11–14 shows the results with the sets preprocessed by NCL, random under-sampling, SMOTE, and random over-sampling.

References

Akbani, R., Kwek, S., Japkowicz, N., 2004. Applying support vector machines to imbalanced datasets. In: 15th European Conference on Machine Learning. Pisa, Italy, pp. 39–50.

Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F., 2011. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing 17, 255–287.

Aldape-Pérez, M., Yañez Márquez, C., Camacho-Nieto, O., Argüelles-Cruz, A. J., 2012. An associative memory approach to medical decision support systems. Computer Methods and Programs in Biomedicine 106 (3), 287–307.

Aldape-Pérez, M., Yañez Márquez, C., Camacho-Nieto, O., López-Yáñez, I., Argüelles-Cruz, A. J., 2015. Collaborative learning based on associative models: Application to pattern classification in medical datasets. Computer in Human Behavior 51 (Part B), 771–779.

Anderson, J. A., 1972. A simple neural network generating an interactive memory. Mathematical Bioscience 14, 197–220.

Table 10: Geometric mean for the original data sets

	HACT	BNet	MLP	RBF	SVM	C4.5	NNge	VP
wisconsin	97.70	97.38	95.60	96.35	96.98	94.84	96.81	91.56
yeast1	66.20	64.10	62.83	50.69	41.97	63.64	68.58	41.39
haberman	62.65	40.43	49.82	38.81	0.00	48.31	52.11	24,46
vehicle1	63.39	67.57	75.80	63.94	27.97	63.01	70.91	0.00
vehicle3	64.93	67.51	72.66	59.81	0.00	63.51	67.23	0.00
glass-0-1-2-3_vs_4-5-6	92.68	87.89	91.93	89.27	86.51	91.49	70.64	0.00
ecoli1	87.04	84.99	85.34	88.31	80.60	85.78	86.20	72.14
ecoli2	81.13	85.65	89.00	90.65	74.83	85.57	87.18	47.99
segment0	71.86	98.78	99.39	97.71	98.89	98.38	99.00	65.22
glass6	89.12	91.06	83.93	87.00	86.85	79.73	84.16	0.00
yeast3	76.38	84.50	85.07	86.51	67.02	84.99	87.45	24.89
ecoli3	80.54	83.91	76.16	58.37	0.00	68.65	64.31	0.00
ecoli-0-3-4_vs_5	77.46	83.20	88.18	91.42	83.43	79.03	87.94	39.86
yeast-0-2-5-6_vs_3-7-8-9	69.46	72.16	69.36	60.48	31.80	57.87	70.16	46.84
ecoli-0-4-6_vs_5	77.33	88.71	88.47	85.90	80.40	79.53	88.23	22.18
ecoli-0-3-4-6_vs_5	77.51	82.07	88.23	91.70	83.44	79.97	87.49	31.19
ecoli-0-3-4-7_vs_5-6	77.22	69.13	88.47	83.19	71.95	76.78	83.94	34.57
yeast-0-5-6-7-9_vs_4	74.07	41.53	68.70	28.10	0.00	62.54	62.18	20.00
ecoli-0-6-7_vs_5	78.28	80.42	85.73	86.39	67.08	73.60	88.77	0.00
led7digit-0-2-4-5-6-7-8-9_vs_1	79.58	87.67	88.92	81.65	90.17	87.34	56.32	82.62
ecoli-0-1_vs_5	74.47	86.20	89.03	89.03	83.67	79.70	82.72	38.73
yeast-1_vs_7	64.25	32.57	51.25	31.47	0.00	44.43	40.21	0.00
glass4	82.23	56.71	86.69	86.02	25.81	76.88	87.64	0.00
ecoli4	79.39	80.49	88.73	88.59	59.16	79.72	83.13	31.62
yeast-1-4-5-8_vs_7	59.24	0.00	18.22	0.00	0.00	0.00	0.00	0.00
glass5	87.23	91.33	89.00	82.84	0.00	89.23	70.36	0.00
yeast-2_vs_8	76.98	74.08	73.83	77.29	74.08	0.00	70.48	67.01
yeast4	71.36	53.13	54.09	0.00	0.00	44.35	52.21	13.49
yeast-1-2-8-9_vs_7	63.28	40.69	36.45	18.28	0.00	48.26	36.10	0.00
ecoli-0-1-3-7_vs_2-6	78.54	83.36	83.51	83.36	83.51	70.58	62.90	29.94
yeast6	72.33	82.45	69.39	0.00	0.00	75.25	65.09	0.00

Table 11: Geometric mean for the data sets preporcessed by NCL

	HACT	BNet	MLP	RBF	SVM	C4.5	NNge	VP
wisconsin	97.58	97.99	96.43	96.68	96.98	94.07	97.29	91.65
yeast1	65.96	59.52	63.01	57.48	44.90	61.07	67.46	39.13
haberman	66.27	42.81	43.68	49.11	31.16	48.37	50.26	11.22
vehicle1	64.28	66.09	66.93	60.80	38.74	62.83	57.43	0.00
vehicle3	64.52	63.38	55.96	48.60	0.00	51.41	58.30	0.00
glass-0-1-2-3_vs_4-5-6	93.97	86.04	89.67	90.43	83.96	88.06	89.126	0.00
ecoli1	85.40	85.86	82.24	88.26	82.19	86.12	86.04	80.95
ecoli2	79.66	81.69	91.30	92.28	77.14	86.74	86.77	59.06
segment0	71.76	98.90	99.28	97.66	98.89	98.30	99.31	53.20
glass6	88.83	87.33	86.62	78.31	86.85	88.33	82.77	0.00
yeast3	7.41	79.81	85.14	85.24	67.00	82.37	87.52	36.79
ecoli3	76.21	62.93	74.46	76.15	0.00	66.83	79.02	16.91
ecoli-0-3-4_vs_5	77.46	85.87	88.18	88.69	83.43	79.94	91.94	44.34
yeast-0-2-5-6_vs_3-7-8-9	68.22	74.43	68.89	68.89	51.09	60.60	70.15	39.05
ecoli-0-4-6_vs_5	78.28	88.96	91.70	83.44	83.44	76.40	83.44	22.18
ecoli-0-3-4-6_vs_5	77.46	88.47	88.72	83.21	83.67	82.76	86.13	0.00
ecoli-0-3-4-7_vs_5-6	75.87	68.98	88.86	84.30	71.95	79.31	79.82	40.00
yeast-0-5-6-7-9_vs_4	71.20	23.70	58.35	57.05	0.00	53.70	59.22	0.00
ecoli-0-6-7_vs_5	77.06	80.42	85.51	80.22	63.25	77.07	83.46	0.00
led7digit-0-2-4-5-6-7-8-9_vs_1	77.22	87.67	91.12	89.04	90.03	83.72	90.51	87.44
ecoli-0-1_vs_5	75.98	86.20	89.03	88.82	83.67	82.90	86.01	38.64
yeast-1_vs_7	62.99	0.00	36.35	8.15	0.00	25.72	25.78	0.00
glass4	81.69	53.37	62.63	44.50	25.81	89.91	54.62	0.00
ecoli4	78.58	80.62	91.90	86.32	54.77	79.84	83.67	0.00
yeast-1-4-5-8_vs_7	56.27	0.00	0.00	0.00	0.00	0.00	0.00	40.00
glass5	69.56	0.00	100.00	44.72	0.00	83.26	44.72	40.00
yeast-2_vs_8	74.08	74.08	74.08	74.00	74.08	0.00	74.08	74.08
yeast4	68.88	0.00	37.10	19.53	0.00	14.14	19.93	14.14
yeast-1-2-8-9_vs_7	56.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ecoli-0-1-3-7_vs_2-6	77.12	83.20	83.20	83.51	83.20	70.58	83.51	44.72
yeast6	70.74	79.67	62.98	73.55	0.00	71.59	56.04	0.00

Table 12: Geometric mean for the data sets preporcessed by RUS

	HACT	BNet	MLP	RBF	SVM	C4.5	NNge	VP
wisconsin	95.88	97.48	96.20	96.45	97.36	95.23	95.96	89.27
yeast1	67.26	67.60	70.06	66.11	70.00	70.90	67.40	68.43
haberman	60.39	52.25	60.68	52.42	50.90	63.70	56.54	48.87
vehicle1	62.02	66.03	77.01	69.71	74.52	71.66	64.94	62.91
vehicle3	65.00	68.06	77.33	71.32	74.02	73.48	67.38	64.59
glass-0-1-2-3_vs_4-5-6	84.49	89.09	91.77	89.36	85.86	89.78	91.13	42.11
ecoli1	86.99	85.09	86.69	88.17	86.88	88.90	87.12	87.15
ecoli2	82.78	89.63	88.57	87.11	88.63	85.62	86.94	80.02
segment0	75.68	99.01	99.02	97.10	98.11	98.29	99.34	89.12
glass6	87.81	92.06	85.77	84.77	90.54	85.16	87.32	37.50
yeast3	85.30	90.53	90.48	88.71	89.39	91.57	92.14	86.37
ecoli3	88.07	87.15	82.76	83.86	86.36	85.21	80.92	81.04
ecoli-0-3-4_vs_5	88.24	93.30	87.19	85.63	88.54	83.67	82.20	81.85
yeast-0-2-5-6_vs_3-7-8-9	76.37	75.75	74.91	77.51	78.36	79.38	72.60	77.01
ecoli-0-4-6_vs_5	86.49	84.20	88.47	83.68	88.62	79.81	81.77	76.68
ecoli-0-3-4-6_vs_5	85.99	91.21	81.30	88.38	87.07	83.85	85.74	71.68
ecoli-0-3-4-7_vs_5-6	86.49	80.05	92.98	85.08	89.90	82.83	82.07	80.06
yeast-0-5-6-7-9_vs_4	79.60	78.87	75.40	73.75	78.38	72.80	73.00	78.37
ecoli-0-6-7_vs_5	85.25	85.91	87.46	79.37	86.24	84.85	86.24	69.61
led7digit-0-2-4-5-6-7-8-9_vs_1	88.54	88.45	85.83	85.41	89.47	89.24	73.80	85.96
ecoli-0-1_vs_5	89.67	88.78	85.71	86.55	88.34	85.49	82.68	72.37
yeast-1_vs_7	67.11	45.40	69.07	71.47	77.73	59.43	67.12	58.43
glass4	86.43	86.17	88.79	87.10	86.21	78.72	85.90	39.37
ecoli4	92.09	90.60	93.55	90.09	95.48	84.90	93.20	52.84
yeast-1-4-5-8_vs_7	54.00	0.00	53.36	57.36	59.90	53.14	57.30	41.94
glass5	56.57	90.52	91.59	96.27	90.79	88.66	84.84	18.49
yeast-2_vs_8	74.09	74.08	65.45	70.82	72.53	69.68	63.44	59.48
yeast4	81.39	81.13	77.50	79.96	81.70	83.97	78.58	76.96
yeast-1-2-8-9_vs_7	64.62	42.03	68.96	64.48	73.89	66.67	62.47	46.07
ecoli-0-1-3-7_vs_2-6	80.86	60.95	70.74	80.24	80.87	76.51	78.86	38.14
yeast6	87.81	87.40	83.03	88.67	88.69	82.10	80.03	86.58

Table 13: Geometric mean for the data sets preporcessed by SMOTE

	HACT	BNet	MLP	RBF	SVM	C4.5	NNge	VP
wisconsin	97.70	97.38	95.60	96.35	96.98	94.84	96.81	91.56
yeast1	66.21	71.13	71.30	69.84	70.42	70.58	69.13	68.19
haberman	62.48	63.43	56.13	54.32	47.78	65.78	58.15	45.40
vehicle1	63.04	68.51	74.99	69.04	72.01	69.77	72.90	52.65
vehicle3	64.85	68.01	77.21	72.25	71.75	69.63	68.27	62.30
glass-0-1-2-3_vs_4-5-6	86.55	90.53	91.87	95.23	88.10	89.09	87.92	0.00
ecoli1	87.41	85.25	89.31	88.37	90.40	85.43	89.18	87.53
ecoli2	89.18	84.02	91.34	89.35	90.17	85.57	88.94	89.55
segment0	75.27	97.98	99.86	97.20	99.16	99.06	99.00	98.66
glass6	88.45	93.31	88.82	88.33	89.31	89.29	90.19	58.89
yeast3	86.85	46.94	90.71	89.16	88.80	91.88	92.83	89.59
ecoli3	87.47	78.85	85.16	88.81	89.56	77.40	85.73	87.64
ecoli-0-3-4_vs_5	89.07	87.93	88.18	90.12	89.32	89.59	88.19	91.33
yeast-0-2-5-6_vs_3-7-8-9	76.81	68.67	75.31	75.30	79.39	75.36	75.80	78.45
ecoli-0-4-6_vs_5	87.84	87.73	86.50	88.72	89.15	89.14	90.93	86.48
ecoli-0-3-4-6_vs_5	88.12	82.30	89.67	90.94	89.67	89.16	90.18	83.30
ecoli-0-3-4-7_vs_5-6	86.44	83.54	86.11	88.82	90.10	88.82	89.46	89.06
yeast-0-5-6-7-9_vs_4	81.06	70.67	74.60	75.80	78.68	77.41	74.62	77.30
ecoli-0-6-7_vs_5	83.49	81.47	87.46	84.62	86.49	87.46	88.43	85.38
led7digit-0-2-4-5-6-7-8-9_vs_1	88.83	84.41	88.05	83.72	88.10	90.28	56.18	88.44
ecoli-0-1_vs_5	89.43	88.82	88.23	85.61	89.86	82.32	85.42	90.67
yeast-1_vs_7	69.29	52.75	60.27	65.15	76.59	62.23	71.62	66.80
glass4	86.73	90.14	91.99	89.91	90.03	83.69	90.98	59.55
ecoli4	93.79	86.05	91.31	95.29	90.87	76.48	94.11	94.00
yeast-1-4-5-8_vs_7	64.59	17.98	51.54	56.97	64.46	33.45	55.34	62.41
glass5	49.73	89.00	94.17	83.26	60.57	99.03	70.54	50.18
yeast-2_vs_8	74.08	58.91	75.08	74.39	74.08	78.86	77.06	76.95
yeast4	82.88	33.54	80.95	81.68	81.98	76.14	79.18	82.56
yeast-1-2-8-9_vs_7	69.13	18.22	64.97	59.90	74.12	54.66	59.03	72.10
ecoli-0-1-3-7_vs_2-6	81.58	70.32	87.62	83.20	84.91	70.06	70.19	85.76
yeast6	87.34	33.74	83.55	87.51	87.86	78.99	84.18	88.26

Table 14: Geometric mean for the data sets preporcessed by ROS

	HACT	BNet	MLP	RBF	SVM	C4.5	NNge	VP
wisconsin	95.88	97.80	96.19	96.46	97.28	95.34	96.58	92.09
yeast1	66.79	71.77	69.94	66.73	70.63	69.26	65.33	70.79
haberman	59.60	61.39	55.67	57.59	55.95	54.00	47.10	61.65
vehicle1	62.94	67.94	81.29	70.36	76.00	70.23	58.01	65.60
vehicle3	64.84	67.99	81.35	72.46	73.23	70.89	52.90	65.62
glass-0-1-2-3_vs_4-5-6	84.86	89.48	91.77	91.68	88.68	86.95	82.32	44.02
ecoli1	87.58	86.30	85.70	88.33	89.78	89.31	85.31	88.33
ecoli2	89.45	86.84	91.13	89.01	90.96	86.38	83.97	90.79
segment0	76.02	93.17	99.70	96.91	99.16	99.02	97.54	99.06
glass6	88.07	89.19	85.77	87.57	88.32	86.99	74.00	7.35
yeast3	88.03	89.59	91.32	89.70	88.76	88.96	89.59	88.91
ecoli3	84.12	82.25	86.12	86.45	85.95	78.33	74.33	85.77
ecoli-0-3-4_vs_5	87.46	88.94	87.43	88.44	89.32	92.73	85.14	88.03
yeast-0-2-5-6_vs_3-7-8-9	86.41	75.09	74.35	78.14	79.41	73.61	68.25	79.12
ecoli-0-4-6_vs_5	87.07	76.40	83.26	86.13	88.89	78.17	88.23	84.06
ecoli-0-3-4-6_vs_5	87.07	84.71	88.72	88.23	89.42	81.84	87.49	89.80
ecoli-0-3-4-7_vs_5-6	88.12	81.21	89.45	86.34	90.31	85.48	83.75	86.01
yeast-0-5-6-7-9_vs_4	81.19	66.06	72.85	77.09	78.68	63.91	53.70	77.22
ecoli-0-6-7_vs_5	84.30	85.73	90.80	87.94	86.49	85.29	89.21	86.69
led7digit-0-2-4-5-6-7-8-9_vs_1	88.54	88.31	87.48	87.03	87.97	86.31	89.01	87.96
ecoli-0-1_vs_5	89.39	83.29	88.02	91.35	89.00	76.39	82.72	90.24
yeast-1_vs_7	70.11	54.28	74.28	70.61	77.38	55.87	44.24	71.37
glass4	82.99	47.93	89.22	82.19	90.27	84.13	44.36	54.51
ecoli4	93.81	83.13	91.16	90.14	90.58	82.73	77.21	93.54
yeast-1-4-5-8_vs_7	65.65	41.71	33.80	55.48	64.61	46.83	25.75	64.53
glass5	64.19	44.72	89.23	70.19	93.69	94.17	44.72	48.39
yeast-2_vs_8	74.09	52.90	72.70	69.70	74.08	79.56	70.63	73.35
yeast4	82.47	57.13	74.85	75.88	80.45	63.42	36.74	82.45
yeast-1-2-8-9_vs_7	67.27	49.08	48.14	65.04	73.29	59.55	47.68	70.44
ecoli-0-1-3-7_vs_2-6	85.14	44.64	81.50	83.05	85.09	69.94	70.58	79.13
yeast6	87.90	74.56	79.62	86.38	87.73	78.31	50.62	86.44

Barandela, R., Valdovinos, R. M., Sánchez, J. S., Ferri, F. J., 2004. The imbalanced training sample problem: Under or over sampling? In: Fred, A., Caelli, T. M., Duin, R. P. W., Campilho, A. C., de Ridder, D. (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. Vol. 3138 of Lecture Notes in Computer Science. Springer, pp. 806–814.

Barua, S., Islam, M., Yao, X., Murase, K., 2014. MWMOTE–Majority weighted minority oversampling technique for imbalanced data set learning. IEEE Trans. on Knowledge and Data Engineering 26 (2), 405–425.

Blagus, R., Lusa, L., 2010. Class prediction for high dimensaional class-imbalanced data. BMC Bioinformatics 11 (523), 1–17.

Brown, I., Mues, C., 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications 39 (3), 3446–3453.

Cao, P., Zhao, D., Zaiane, O., 2014. Hybrid probabilistic sampling with random subspace for imbalanced data learning. Intelligent Data Analysis 18 (6), 1089–1108.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artifial Intelligence Research 16 (1), 321–357.

Cleofas-Sánchez, L., Camacho-Nieto, O., Sánchez, J. S., Valdovinos-Rosas, R. M., 2014. Equilibrating the recognition of the minority class in the imbalance context. Applied Mathematics and Information Sciences 8 (1), 27–36.

Cleofas-Sánchez, L., García, V., Martín-Félez, R., Valdovinos, R. M., Sánchez, J. S., Camacho-Nieto, O., 2013. Hybrid associative memories for imbalanced data classification: An experimental study. In: Carrasco-Ochoa, J. A., Martínez-Trinidad, J. F., Rodríguez, J. S., di Baja, G. S. (Eds.), Pattern Recognition. Vol. 7914 of Lecture Notes in Computer Science. Springer, pp. 325–334.

Dal Pozzolo, A., Caelen, O., Bontempi, G., 2015. When is undersampling effective in unbalanced classification tasks? In: Appice, A., Rodrigues, P. P., Santos Costa, V., Soares, C., Gama, J., Jorge, A. (Eds.), Machine Learning and Knowledge Discovery in Databases. Vol. 9284 of Lecture Notes in Computer Science. Springer, pp. 200–215.

Danilo, R., Jarollahi, H., Gripon, V., Coussy, P., Conde-Canencia, L., Gross, W., 2015. Algorithm and implementation of an associative memory for oriented edge detection using improved clustered neural networks. In: IEEE International Symposium on Circuits and Systems. Lisbon, Portugal, pp. 2501–2504.

Daskalaki, S., Kopanas, I., Avouris, N., 2006. Evaluation of classifiers for an uneven class distribution problem. Applied Artificial Intelligence 20 (5), 381–417.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30.

Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C., Kuncheva, L. I., 2015. Random Balance: Ensembles of variable priors classifiers for imbalanced data. Knowledge-Based Systems 85, 96–111.

Dubey, H., Pudi, V., 2013. Class based weighted k-nearest neighbor over imbalance dataset. In: Pei, J., Tseng, V., Cao, L., Motoda, H., Xu, G. (Eds.), Advances in Knowledge Discovery and Data Mining. Vol. 7819 of Lecture Notes in Computer Science. Springer, pp. 305–316.

Fdez-Glez, J., Ruano-Ordás, D., Fdez-Riverola, F., Méndez, J., Pavón, R., Laza, R., 2015. Analyzing the impact of unbalanced data on web spam classification. In: Omatu, S., Malluhi, Q. M., Gonzalez, S. R., Bocewicz, G., Bucciarelli, E., Giulioni, G., Iqba, F. (Eds.), 12th International Conference on Distributed Computing and Artificial Intelligence. Vol. 373 of Advances in Intelligent Systems and Computing. Springer, pp. 243–250.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F., 2012. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews 42 (4), 463–484.

Galar, M., Fernández, A., Barrenechea, E., Herrera, F., 2013. EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. Pattern Recognition 46 (12), 3460–3471.

García, V., Sánchez, J. S., Mollineda, R. A., 2012. On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowledge-Based Systems 25 (1), 13–21.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., 2009. The WEKA data mining software: an update. SIGKDD Explorations Newsletter 11 (1), 10–18.

He, H., Garcia, E. A., 2009. Learning from imbalanced data. IEEE Trans. on Knowledge and Data Engineering 21, 1263–1284.

He, H., Ma, Y., 2013. Imbalanced Learning: Foundations, Algorithms, and Applications. Wiley – IEEE Press, Piscataway, NJ.

He, M., Weir, J. D., Wu, T., Silva, A., Zhao, D.-Y., Qian, W., 2015. K Nearest Gaussian–A model fusion based framework for imbalanced classification with noisy dataset. Artificial Intelligence Research 4 (2), 126–135.

Hong, X., Chen, S., Harris, C., 2007. A kernel-based two-class classifier for imbalanced data sets. IEEE Trans. on Neural Networks 18 (1), 28–41.

Hopfield, J. J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 79 (8), 2554–2558.

Hwang, J. P., Park, S., Kim, E., 2011. A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function. Expert Systems with Applications 38 (7), 8580–8585.

Japkowicz, N., 2013. Assessment metrics for imbalanced learning. In: Imbalanced Learning: Foundations, Algorithms, and Applications. Wiley – IEEE Press, pp. 187–210.

Japkowicz, N., Stephen, S., 2002. The class imbalance problem: A systematic study. Intelligent Data Analysis 6 (5), 429–449.

Kang, S., Ramamohanarao, K., 2014. A robust classifier for imbalanced datasets. In: Tseng, V., Ho, T., Zhou, Z.-H., Chen, A., Kao, H.-Y. (Eds.), Advances in Knowledge Discovery and Data Mining. Vol. 8443 of Lecture Notes in Computer Science. Springer, pp. 212–223.

Kareem, E. I. A., Jantan, A., 2011. An intelligent traffic light monitor system using an adaptive associative memory. International Journal of Information Processing and Management 2 (2), 23–39.

Kim, M.-J., Kang, D.-K., Kim, H. B., 2015. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. Expert Systems with Applications 42 (3), 1074–1082.

Kohonen, T., 1972. Correlation matrix memories. IEEE Trans. on Computers C–21, 353–359.

Kosko, B., 1987. Adaptive bidirectional associative memories. Applied Optics 26 (23), 4947–4960.

Krawczyk, B., Woniak, M., Schaefer, G., 2014. Cost-sensitive decision tree ensembles for effective imbalanced classification. Applied Soft Computing 14, Part C, 554–562.

Kwon, O., Sim, J. M., 2013. Effects of data set features on the performances of classification algorithms. Expert Systems with Applications 40 (5), 1847–1857.

Laurikkala, J., 2001. Improving identification of difficult small classes by balancing class distribution. In: 8th Conference on Artificial Intelligence in Medicine. Cascais, Portugal, pp. 63–66.

Liu, Y., Yu, X., Huang, J. X., An, A., 2011. Combining integrated sampling with SVM ensembles for learning from imbalanced datasets. Information Processing & Management 47 (4), 617–631.

López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 250, 113–141.

López, V., Fernández, A., Moreno-Torres, J. G., Herrera, F., 2012. Analysis of pre-processing vs. cost-sensitive learning for imbalanced classification. open problems on intrinsic data characteristics. Expert Systems with Applications 39 (7), 6585–6608.

Lou, X., Cui, B., 2007. Robust asymptotic stability of uncertain fuzzy BAM neural networks with time-varying delays. Fuzzy Sets and Systems 158 (24), 2746–2756.

Maldonado, S., López, J., 2014. Imbalanced data classification using second-order cone programming support vector machines. Pattern Recognition 47 (5), 2070–2079.

Maratea, A., Petrosino, A., Manzo, M., 2014. Adjusted F-measure and kernel scaling for imbalanced data learning. Information Sciences 257, 331–341.

Menardi, G., Torelli, N., 2014. Training and assessing classification rules with imbalanced data. Data Mining and Knowledge Discovery 28 (1), 92–122.

Mu, X., Artiklar, M., Watta, P., Hassoun, M., 2006. An RCE-based associative memory with application to human face recognition. Neural Processing Letters 23 (3), 257–271.

Napierala, K., Stefanowski, J., Wilk, S., 2010. Learning from imbalanced data in presence of noisy and borderline examples. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (Eds.), Rough Sets and Current Trends in Computing. Vol. 6086 of Lecture Notes in Computer Science. Springer, pp. 158–167.

Nazari, S., Eftekhari-Moghadam, A.-M., Moin, M.-S., 2014. A novel image stenography scheme based on morphological associative memory and permutation schema. Security and Communications Networks 8 (2), 110–121.

Palm, G., 2013. Neural associative memories and sparse coding. Neural Networks 37, 165–171.

Park, Y., Ghosh, J., 2014. Ensembles of α-trees for imbalanced classification problems. IEEE Trans. on Knowledge and Data Engineering 26 (1), 131–143.

Prati, R. C., Batista, G. E., Monard, M. C., 2004. Learning with class skews and small disjuncts. In: Bazzan, A. L., Labidi, S. (Eds.), Advances in Artificial Intelligence. Vol. 3171 of Lecture Notes in Computer Science. Springer, pp. 296–306.

Ritter, G., Sussner, P., Diaz-de Leon, J. L., 1998. Morphological associative memories. IEEE Trans. on Neural Networks 9 (2), 281–293.

Santiago, R., 2003. Clasificador Híbrido de Patrones basado en la Lernmatrix de Steinbuch y el Linear Associator de Anderson-Kohonen. Master's thesis, Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico D.F. (in Spanish).

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., Folleco, A., 2014. An empirical study of the classification performance of learners on imbalanced and noisy software quality data. Information Sciences 259, 571–595.

Sharma, N., Ray, A. K., Sharma, S., Shukla, K. K., Pradhan, S., Aggarwal, L. M., 2008. Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network. Journal of Medical Physics 33 (3), 119–3126.

Stefanowski, J., 2013. Overlapping, rare examples and class decomposition in learning classifiers from imbalanced data. In: Ramanna, S., Jain, L. C., Howlett, R. J. (Eds.), Emerging Paradigms in Machine Learning. Vol. 13 of Smart Innovation, Systems and Technologies. Springer, pp. 277–306.

Steinbuch, K., 1961. Die lernmatrix. Kybernetik 1 (1), 36–45.

Uriarte-Arcia, A. V., López-Yáñez, I., Yáñez Márquez, C., 2014. One-hot vector hybrid associative classifier for medical data classification. PLoS ONE 9 (4), e95715.

Štanclová, J., Zavoral, F., 2005. Hierarchical associative memories: The neural network for prediction in spatial maps. In: Roli, F., Vitulano, S. (Eds.), Image Analysis and Processing. Vol. 3617 of Lecture Notes in Computer Science. Springer, pp. 786–793.

Wilson, D. L., 1972. Asymptotic properties of nearest neighbour rules using edited data. IEEE Trans. on Systems, Man and Cybernetics 2, 408–421.

Yang, J., Yu, X., Xie, Z.-Q., Zhang, J.-P., 2011. A novel virtual sample generation method based on gaussian distribution. Knowledge-Based Systems 24 (6), 740–748.

Yu, H., Mu, C., Sun, C., Yang, W., Yang, X., Zuo, X., 2015. Support vector machine-based optimized decision threshold adjustment strategy for classifying imbalanced data. Knowledge-Based Systems 76, 67–78.

Yu, H., Ni, J., Zhao, J., 2013. ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data. Neurocomputing 101, 309–318.