Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task

Nesma Settouti, Mohammed El Amine Bechar and Mohammed Amine Chikh

Biomedical Engineering Department, University of Tlemcen, Algeria

***Abstract* — This work is builds on the study of the 10 top data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) community in December 2006. We address the same study, but with the application of statistical tests to establish, a more appropriate and justified ranking classifier for classification tasks. Current studies and practices on theoretical and empirical comparison of several methods, approaches, advocated tests that are more appropriate. Thereby, recent studies recommend a set of simple and robust non-parametric tests for statistical comparisons classifiers. In this paper, we propose to perform non-parametric statistical tests by the Friedman test with post-hoc tests corresponding to the comparison of several classifiers on multiple data sets. The tests provide a better judge for the relevance of these algorithms.**

***Keywords* — The Top 10 Data Mining Algorithms, Classification, Statistical Comparisons Of Classifiers, Non-Parametric Test, Friedman Test, Post-Hoc Procedures.**

I. Introduction

Today, in the field of pattern recognition exists a large number of classifiers and feature selection methods. It is clear that no single model exists for all pattern recognition problems and no single technique is applicable to all problems. Rather, what we have is a bag of tools and a bag of problems [1]. Despite the numerous work in the field, that did not allow to highlight the indisputable superiority of one method of classification to another or a feature selection on another.

The identification of the top 10 algorithms by Wu et al. [2], [3] inspire us to explore these 10 algorithms, including their impact and new research issues. These 10 algorithms cover classification, clustering, statistical learning, association analysis, and link mining, which are all among the most important topics in data mining research and development, as well as for curriculum design for related data mining, machine learning, and artificial intelligence courses.

The initiative of identifying the top 10 data mining algorithms comes from a nomination and voting process. As the first step in the identification process, in September 2006 Wu et al. invited ACM KDD[1] Innovation Award and IEEE ICDM[2] Research Contributions Award winners to each nominate up to 10 best-known algorithms in data mining.

After the nominations, the step 2, concerns the verification of each nomination form its citations, and removing those nominations that did not have at least 50 citations. All remaining (18) nominations were then organized in 10 topics: association analysis, classification, clustering, statistical learning, bagging and boosting, sequential patterns, integrated mining, rough sets, link mining, and graph mining.

In the third step of the identification process, a wider involvement of the research community was necessary. Divers Program Committee members (KDD, ICDM, SDM[3], and ACM KDD) were invited to vote for the 10 well-known algorithms for the 18-algorithms candidate list. The voting results of this step were presented at the ICDM06 panel on Top 10 Algorithms in Data Mining.

In classification task, the choice of classifier in this long list of methods is primordial for a better recognition and this especially on the medical field. Modern medicine needs computer assistance for detection, diagnosis and classification of certain diseases in a very short time hence the need for a classification system. The use of intelligent methods to perform this classification is becoming more frequent. Although the decision of the doctor is the most critical factor in the diagnosis, a diagnostic to medical aid has developed and gained popularity, these systems are even considered as essential in many medical disciplines. In practice, there are already many applications using automatic learning that allows assisting clinicians in their diagnostic procedures. These approaches can provide a more accurate diagnosis and reduce the maximum errors due to fatigue and doubts of the doctor.

Therefore, in order to have an effective classification / regression systems from a set of representative examples of a population dataset. We must make the best choice of classifier. In this condition gives rise to a series of questions:

- How to know which is the most suitable classifier for a specific dataset?
- Are there cases to identify a classifier as a “logic” choice?
- What are the principles of selecting a classifier?

A recommended approach is to test several different classifiers as well as different parameter sets within each algorithm, and then to select the most effective using the non-parametric statistical tests. We talk about non-parametric tests when we make no assumptions about the distribution of variables. Also known as, free distribution tests, i.e. the quality of the results do not depend, a priori, on the underlying data distribution. An extensive study is performed by the Friedman test with post-hoc tests corresponding to the comparison of several classifiers on multiple data sets, in order to validate the most appropriate classifier structure, in terms of the correct classification rate and the generalization ability.

In this paper, we focus on the problem of determining the most suitable classifier to solve a given problem of classification. The choice of the classifier is already guided by operational constraints, but beyond these constraints, and after that the classifier is configured through a learning basis, the rate of generalization of the classifier (or Accuracy) which is the criterion characterizing its performance. This rate, usually unknown, is estimated using a generalized basis. This estimate, therefore, depends on the classification problem studied, the classifier uses the learning base and widespread basis. These dependencies are studied theoretically and experimentally over a dozen different classifiers. The problem of the validity of the comparison of two or more classifiers by estimates of their generalization rate is also studied by using non-

1. Association for Computing Machinery Knowledge Discovery and Data Mining
2. IEEE International Conference on Data Mining
3 SIAM International Conference on Data Mining

DOI: 10.9781/ijimai.2016.419

parametric tests. A ranking of classifiers goal provides in this work by testing this top 10 algorithms on different databases.

This paper is organized as follows: first, in section 2, the definitions of non-parametric tests are exposed. We present briefly in section 3, the 10 classifiers candidates in the standings. In Section 4, the step of experimentations and results, we discuss and analyze the results performed on 10 medical databases of different distributions and sizes. We conclude with a synthesis of this approach and a ranking of the 10 best methods for classification. (Section 5).

II. The Non-Parametric Tests

A non-parametric test is a test where the model does not specify the conditions that must fulfill the parameters of the population, which the sample was extracted. However, certain conditions of application should be checked. The samples must be considered random (when all people have the same probability to be a part of the sample) and single (all individuals who should form the sample has taken independently of each other) [4], and possibly independent from each other (use of random number tables). The random variables considered are generally assumed continuous.

Instead of entering in a debate "for or against" the nonparametric tests by opposing their parametric counterparts based on the normal data distribution. We try to characterize the situations where it is more (or less) advantageous to use them.

a) Advantages:

1. Their use is justified when the conditions for application of other methods are not satisfied even after possible variable transformations.
2. The probability of the results for most non-parametric tests are exact probabilities regardless of the distribution and the population shape with the sample is drawn.
3. For samples of very small size to N = 6, the only possibility is using a non-parametric test, unless the exact nature of the distribution of the population is precisely known. This allows a reduction in the cost or time needed to collect the information.
4. There are non-parametric tests for processing composite samples based on observations from different populations. Such data may only be processed by the parametric tests without making unrealistic assumptions.
5. Only non-parametric tests exist that allow the treatment of qualitative data either in rows or expressed more or less (ordinal scale) or nominal.
6. Non-parametric tests are easier to learn and apply than the parametric tests. Their relative simplicity often results from the replacement of the values observed either by alternative variables indicating membership in one or the other class observation or by the rows, i.e. the number order of observed values arranged in ascending order. Thus, the median is generally preferred to the mean, as seating position.

b) Disadvantages:

1. The parametric tests, when their conditions are satisfied, are more potent than the non-parametric tests.
2. A second disadvantage is the difficulty in finding the description of the tests and their tables of significant values. Fortunately, the standard statistical software gives the significance levels.

We selected the appropriate tests depending on the type of measurement, the shape of the frequency distribution and the number of samples that are available (see diagram Fig.1). Therefore, the Friedman test with the post-hoc approach applies to the static comparison studied in this work.

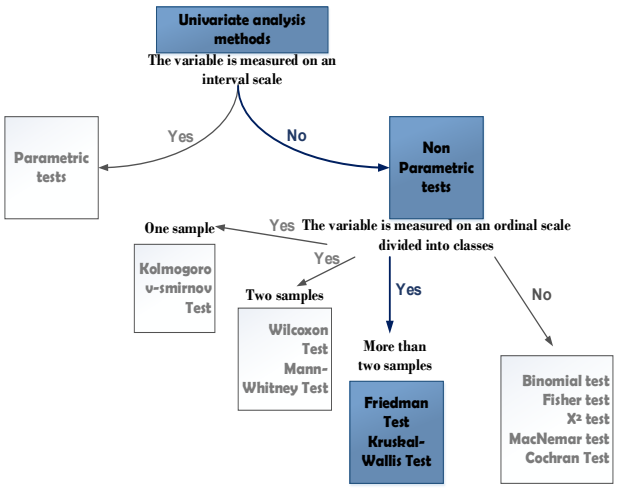

Fig 1. Univariate analysis methods diagram.

A. Friedman Test for Multiple Comparisons

In order to better assess the results obtained for each algorithm, we adopt in this study the post-hoc Friedman test methodology proposed by Demsar [5] for the comparison of several algorithms over multiple datasets.

The Friedman test [6] is a non-parametric test (free distribution) used to compare observations repeated on the same subjects. This is also called a non-parametric randomized black analysis of variance. The test statistic for the Friedman's test is a Chi-square with a-1 degrees of freedom, where a is the number of repeated measures. When the p−value for this test is small (usually < 0.05) you have evidence to reject the null hypothesis. The goal of this test is to determine whether there are significant differences among the algorithms considered over given sets of data. The test determines the ranks of the algorithms for each individual data set.

Garcia et al. [7] and Derrac et al. [8] considered non-parametric tests for multiple comparison as well as post-hoc procedures for NxN comparisons, for classification tasks. The studies illustrate that first the Friedman test should be conducted in order to detect whether statistically significant differences occur among the examined algorithms. Moreover, these tests rank the algorithms from the best performing one to the poorest one. If statistical significance is revealed, then the researcher may proceed to accomplish post-hoc procedures to point out which pair of algorithms differ significantly.

B. Post-hoc procedures for NxN comparisons

The Friedman can only detect significant differences over the whole multiple comparison, although they are not in a position to establish interrelations between the algorithms under consideration. If the null hypothesis of equivalence of rankings is rejected by these tests, the researcher may proceed with post-hoc procedures. Fig. 2 present the different procedures for multiple comparison NxN.

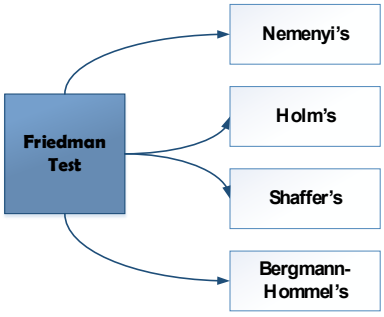

Fig 2. Non-parametric tests and post-hoc procedures for NxN comparisons.

TABLE I.
POST-HOC PROCEDURES FORNXNCOMPARISONS

PROCEDURE	DESCRIPTION	APV FORMULA
Nemenyi	Calculates the adjusted value of α in a single step by dividing it by the number of comparisons accomplished, i.e., k(k 1)/2.	min {v; 1}, where v =k(k- 1)pi/2.
Holm	Step-down method, it rejects H1 to Hi-1 if i is the smallest integer such that pi >α (k(k-1)/2-i+1).	min {v; 1}, where v = Max{(k(k-1)/2-j+ 1)pj : 1≤j≤i}
Shaffer	Following Holms step-down method, at stage j, instead of discarding Hi if pi ≤α(k(k-1)/2-i+ 1), discards Hi if pi≤α/ti, where ti is the maximum number of hypotheses which can be true given that any (i,…,1)hypotheses are false.	min {v; 1}, where v = max{tj pj : 1≤j≤i}
Bergmann and Hommel	Rejects all Hj with j ∉ A, where the acceptance set A, given as A= ∪ { I : I exhaustive, min{Pi: i ∈ g} > α/||I||}, is the index set of null hypotheses which are retained.	Min{v; 1}, where v = Max{||I|| min{pj, j ∈ I} : I exhaustive;i ∈ I}

In Table I a set of post-hoc procedures is presented for NxN comparisons. Trawinski et al. [9] summarized for each procedure a brief outline of its scheme and the formula for computation of the Adjusted P-Value (APV). The notation used in Table I is as follows:

- Indexes *i* and *j* apply to a given comparison or hypothesis in the family of hypotheses. Index i always concerns the hypothesis whose APV is being determined and index j refers to another hypothesis in the family;
- p_j is the p-value calculated for the j-th hypothesis;
- k is the number of predictors being compared.

III. The 10 Classifiers Candidates

There have been a large number of data mining algorithms rooted in these fields to perform different data analysis tasks. The 10 algorithms identified by the IEEE International Conference on Data Mining (ICDM) and presented in this article are among the most influential algorithms for classification, clustering, statistical learning and association analysis. We focus only on the Classification task of evaluating experiments with the listed algorithms in Table II over a set of selected medical databases.

TABLE II
DESCRIPTION OF THE 10 CLASSIFIERS CANDIDATES

METHODS	DESCRIPTION	REF
AdaBoost	Adaptive Boosting Negative Correlation Learning Extension with C4.5 Decision Tree as Base Classifier.	[10]
Apriori	Association rule mining using the Apriori algorithm.	[11]
Bagging	Multi-classifier learning approach with C4.5 as baseline algorithm.	[12], [13]
C4.5	Generate classifier expressed as decision trees	[14]
CART	Classification and Regression Tree.	[15]
EM	Expectation-Maximization algorithm	[16]
K-means	K means Classifier.	[17]
KNN	K-Nearest Neighbors Classifier.	[18]
NB	Nave-Bayes.	[19], [20]
SVM	Support vector networks.	[21]

IV. Experimentations and Results

A. Tools for the experimentations

All experiments were conducted using KEEL (Knowledge Extraction based on Evolutionary Learning) [22], [23], an open source Java software tool that can be used for a large number of different knowledge data discovery tasks. It contains a wide variety of algorithms for creating, learning, optimizing and evaluating various models ranging from soft computing ones to support vector machines, decision trees for regression, and linear regression. KEEL algorithms are employed to carry out.

The experiments listed in Table II, where references to source papers are shown. Details of the algorithms can also be found on the KEEL web site http://www.keel.es/.

B. Benchmark data sets

Twelve Medical and Biological datasets mainly selected from the UCI Machine Learning Repository [24], and ASU feature selection Repository [25]. These are used to evaluate the performance of the top 10 algorithms; their characteristics are described in Table III.

TABLE III
DESCRIPTION OF EXPERIMENTAL MEDICAL DATABASES

DATA SETS	# INSTANCES	# FEATURES	# LABELS
Appendicitis	106	9	2
Breast cancer	699	9	2
Dermatology	358	34	6
Diabetes	672	8	2
Heart	270	13	2
Heberman	306	3	2
Hepatitis	155	20	2
Liver disorder	345	7	2
Lymphoma	96	19	4
Mammographic	830	5	2
New-thyroid	215	5	3
Post-operative	87	8	3

C. Results

The top 10 machine learning algorithms were run in KEEL individually for 12 data sets using 10-fold cross validation (10cv) and the prediction error was measured with the Root Mean Square Error (RMSE) in Table IV.

Firstly, we used the non-parametric Friedman test to evaluate the rejection of the hypothesis that all the classifiers perform equally well for a given level. It ranks the algorithms for each dataset separately, the best performing algorithm getting the higher rank, for example, in the case with 4 classifiers, the best rank its equals to 1, the second best rank 2 etc.

TABLE IV
THE RMSE OF THE TOP 10 COMPARED ALGORITHMS OVER TWELVE MEDICAL AND BIOLOGICAL DATA SETS

DATA SETS	ADABOOST	APRIORI	BAGGING	C4.5	CART	EM	K-MEANS	KNN	NB	SVM
Appendicitis	0.197	0.182	0.153	0.153	0.137	0.181	0.176	0.164	0.198	0.127
Breast cancer	0.275	0.311	0.236	0.236	0.291	0.322	0.315	0.319	0.292	0.250
Dermatology	0.178	0.503	0.038	0.260	0.299	0.300	0.153	0.304	0.489	0.365
Diabetes	0.178	0.503	0.038	0.260	0.299	0.300	0.153	0.304	0.489	0.365
Heart	0.310	0.340	0.236	0.236	0.269	0.244	0.531	0.242	0.444	0.370
Heberman	0.121	0.218	0.434	0.270	0.300	0.330	0.510	0.335	0.435	0.294
Hepatitis	0.125	0.422	0.182	0.182	0.125	0.187	0.449	0.193	0.138	0.182
Liver disorder	0.181	0.245	0.139	0.330	0.333	0.389	0.256	0.389	0.255	0.368
Lymphoma	0.207	0.499	0.263	0.239	0.228	0.250	0.543	0.240	0.485	0.165
Mammographic	0.162	0.146	0.054	0.163	0.175	0.260	0.412	0.256	0.302	0.160
New-thyroid	0.452	0.319	0.260	0.059	0.067	0.027	0.506	0.025	0.348	0.060
Post operative	0.377	0.260	0.301	0.301	0.342	0.479	0.458	0.496	0.284	0.302

TABLE VI
ADJUSTED P-VALUES FOR N X N COMPARISONS OF THE TOP 10 ALGORITHMS OVER 12 DATA SETS

i	Hypothesis	Unadjusted p	pNeme	pHolm	pShaf	pBerg
1	C4.5 vs .K-means	***0.000209***	***0.009397***	***0.009397***	***0.009397***	***0.009397***
2	Bagging vs .K-means	***0.000455***	0.020483	0.020027	0.016386	0.012845
3	C4.5 vs .NB	***0.003012***	0.135554	0.129529	0.108443	0.098299
4	AdaBoost vs .K-means	***0.00336***	0.151179	0.1411	0.120943	0.012596
5	CART vs .K-means	***0.00336***	0.151179	0.1411	0.120943	0.012596
6	AdaBoost vs .SVM	***0.003743***	0.168428	0.149714	0.134742	0.129529
7	Bagging vs .NB	***0.005706***	0.256766	0.222531	0.205413	0.188684
8	C4.5 vs .EM	0.02391	1.075958	0.908587	0.860766	0.739714
9	AdaBoost vs .NB	0.028441	1.279844	1.052316	1.023875	0.962429
10	CART vs .NB	0.028441	1.279844	1.052316	1.023875	0.962429
11	NB vs .SVM	0.030971	1.393714	1.084	1.023875	1.01455
12	Bagging vs .EM	0.039753	1.788871	1.351591	1.152828	1.123586
13	Apriori vs .C4.5	0.043114	1.94015	1.422777	1.250319	1.205413
14	C4.5 vs .KNN	0.046713	2.102103	1.494828	1.354688	1.333108
15	Apriori vs .Bagging	0.068707	3.091828	2.129926	1.992512	1.71346
16	Bagging vs .KNN	0.073997	3.329882	2.219921	2.145924	2.091892
17	K-means vs .KNN	0.085576	3.850929	2.48171	2.48171	2.397533
18	Apriori vs .K-means	0.091892	4.135149	2.572982	2.572982	2.572982
19	AdaBoost vs .EM	0.138011	6.210483	3.72629	3.312258	3.238016
20	CART vs .EM	0.138011	6.210483	3.72629	3.312258	3.238016
21	EM vs .SVM	0.14719	6.623537	3.72629	3.532553	3.277384
22	EM vs .K-means	0.14719	6.623537	3.72629	3.532553	3.277456
23	AdaBoost vs .Apriori	0.212299	9.553437	4.882868	4.670569	4.312559
24	Apriori vs .CART	0.212299	9.553437	4.882868	4.670569	4.312559
25	AdaBoost vs .KNN	0.224916	10.121215	4.882868	4.723234	4.599059
26	Apriori vs .SVM	0.224916	10.121215	4.882868	4.723234	4.599059
27	CART vs .KNN	0.224916	10.121215	4.882868	4.723234	4.599059
28	KNN vs .SVM	0.23806	10.712699	4.882868	4.723234	4.599059
29	KNN vs .NB	0.328277	14.772476	5.580713	5.580713	5.580713
30	Apriori vs .NB	0.345231	15.535398	5.580713	5.580713	5.580713
31	C4.5 vs .SVM	0.418492	18.832151	6.277384	6.277384	6.277384
32	C4.5 vs .CART	0.438145	19.716515	6.277384	6.277384	6.277384
33	AdaBoost vs .C4.5	0.438145	19.716515	6.277384	6.277384	6.277384
34	K-means vs .NB	0.458318	20.624296	6.277384	6.277384	6.277384
35	EM vs .NB	0.479001	21.555056	6.277384	6.277384	6.277384
36	Bagging vs .SVM	0.543997	24.479865	6.277384	6.277384	6.277384
37	Bagging vs .CART	0.566597	25.496882	6.277384	6.277384	6.277384
38	AdaBoost vs .Bagging	0.566597	25.496882	6.277384	6.277384	6.277384
39	EM vs .KNN	0.787406	35.433292	6.277384	6.277384	6.277384
40	Apriori vs .EM	0.813456	36.605519	6.277384	6.277384	6.277384
41	Bagging vs .C4.5	0.839714	37.787108	6.277384	6.277384	6.277384
42	AdaBoost vs .SVM	0.973108	43.789878	6.277384	6.277384	6.277384
43	Apriori vs .KNN	0.973108	43.789878	6.277384	6.277384	6.277384
44	CART vs .SVM	0.973108	43.789878	6.277384	6.277384	6.277384
45	AdaBoost vs .CART	1	45	6.277384	6.277384	6.277384

Then, the Friedman test compares the average ranks of the algorithms and calculates the Friedman statistic. If a statistically significant difference in the performance is detected, which means that some of the hypotheses in the experimentation have different distribution from one another, therefore, our next step will be to try to find out which pairs of our algorithms are significantly different then each other. We proceed with a post-hoc test.

We use the Nemenyi, Holm, Shaffer, Bergmann, and Hommel tests, find out which of the tested methods are distinctive among an NxN comparison. The post-hoc procedure is based on a specific value on the significance level α. Additionally, the obtained pvalue should be examined in order to check how different given two algorithms are. We fix the significance level α= 0.10 for all comparisons. Average rankings of the 10 algorithms over 12 medical and biological data sets for produced by the Friedman test are shown in Table V.

TABLE V
AVERAGE RANKINGS OF THE ALGORITHMS

ALGORITHM	RANKING
AdaBoost	4.5417
Apriori	6.0833
Bagging	3.8333
C4.5	3.5833
CART	4.5417
EM	6.375
K-means	8.1667
KNN	6.0417
NB	7.25
SVM	4.5833

D. Discussion

The results achieved in post-hoc comparisons for α = 0.10 are depicted in Table VI. The unadjusted values and adjusted p-values for Nemenyi, Holm, Shaffer, and Bergmann-Hommel tests for NxN comparisons for all possible 45 pairs of algorithms are placed in Table VI. The pvalues below 0.10 indicate that respective algorithms differ significantly in prediction errors; they were marked with an italic font.

Among the NxN procedures, the Bergmann-Hommels procedure is the most powerful one, but it requires intensive computation in comparisons comprising a bigger number of predictors. Thus, the Shaffers static routine or the Holms step down method is recommended. It should be noted that with 45 hypotheses Holm, Nemenyis, Shaffer and Bergmann-Hommel ones discard only four methods. C4.5 and Bagging revealed significantly better performance than most of the 10 algorithm, thus propelling it to the top spot of classification algorithms.

However, for multiple comparisons the more data sets used in tests the bigger the number of null-hypotheses rejected. Our investigation proved the usefulness and strength of multiple comparison, statistical procedures to analyses and select machine learning algorithms. The ranking reveals that C4.5, Bagging are the most influential for classification tasks. In the third place, we have CART and Adaboost with an equal score, after that SVM, Apriori, KNN, NB and finally the unsupervised classifier K-means.

V. Conclusion

In contemporary machine learning, one cannot say that a given algorithm is superior over another one, without the use of statistical analysis. Experimental results must be accompanied by a thorough statistical analysis, to prove that the reported differences between analyzed models are significant.

In this paper, we studied the application of non-parametric statistical tests and post-hoc procedures devised to perform multiple comparisons of classification algorithms over medical and biological benchmark data sets. We conducted experiments on statistical procedures designed especially for multiple NxN comparisons with the top 10 algorithms in data mining. The tests provide a ranking of the top 10 algorithms, revealing the C4.5, Bagging, Adaboost, CART and SVM for the five most relevant classification algorithms.

References

[1] L. Kanal, “Patterns in pattern recognition: 1968-1974,” Information Theory, IEEE Transactions on, vol. 20, no. 6, pp. 697–722, Nov 1974.

[2] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu,

[3] Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg, “Top 10 algorithms in data mining,” Knowl. Inf. Syst., vol. 14, no. 1, pp. 1–37, Dec. 2007.

[4] Xindong Wu and Vipin Kumar, The Top Ten Algorithms in Data Mining, Chapman & Hall/CRC, 1st edition, 2009.

[5] R. A. Fisher, Statistical Methods for Research Workers, Cosmo study guides. Cosmo Publications, 1925.

[6] Janez Demsar, “Statistical comparisons of classifiers over multiple data˘ sets,” J. Mach. Learn. Res., vol. 7, pp. 1–30, Dec. 2006.

[7] M. Friedman, “The use of ranks to avoid the assumption of normality implicit in the analysis of variance,” Journal of the American Statistical Association, vol. 32, no. 200, pp. 675–701, 1937.

[8] Salvador Garcıa, Alberto Fernandez, Julian Luengo, and Francisco Herrera, “Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power,” Inf. Sci., vol. 180, no. 10, pp. 2044–2064, May 2010.

[9] Joaqun Derrac, Salvador Garca, Daniel Molina, and Francisco Herrera, “A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms,” Swarm and Evolutionary Computation, vol. 1, no. 1, pp. 3 – 18, 2011.

[10] Bogdan Trawinski, Magdalena Smetek, Zbigniew Telec, and Tadeusz Lasota, “Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms.,” Applied Mathematics and Computer Science, vol. 22, no. 4, pp. 867–881, 2012.

[11] S. Wang and X. Yao, “Multiclass imbalance problems: Analysis and potential solutions,” IEEE Transactions on Systems and Man and Cybernetics Part B, vol. 42, no. 4, pp. 1119–1130, 2012.

[12] R. Srikant and R. Agrawal, “Mining quantitative association rules in large relational tables,” in ACM SIGMOD International Conference on Management of Data, 1996, pp. 1–12.

[13] Leo Breiman, “Bagging predictors,” Mach. Learn., vol. 24, no. 2, pp. 123–140, Aug. 1996.

[14] Mikel Galar, Alberto Fernandez, Edurne Barrenechea,´ Humberto Bustince, and Francisco Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” Pattern Recogn., vol. 44, no. 8, pp. 1761–1776, Aug. 2011.

[15] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kauffman, 1993.

[16] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, Classification and Regression Trees, Chapman and Hall (Wadsworth and Inc.), 1984.

[17] Tapio Schneider, “Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values,” Journal of Climate, vol. 14, no. 5, Mar. 2001.

[18] J.B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[19] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inf. Theor., vol. 13, no. 1, pp. 21–27, Sept. 2006.

[20] M.E. Maron, “Automatic indexing: An experimental inquiry,” Journal of the ACM (JACM), vol. 8, no. 3, pp. 404–417, 1961.
[21] P. Domingos and M. Pazzani, “On the optimality of the simple bayesian classifier under zero-one loss,” Machine Learning, vol. 29, pp. 103–137, 1997.
[22] C. Cortes and V. Vapnik, “Support vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995.
[23] J. Alcala Fdez, L. Sunchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernández, and F. Herrera, “Keel: A software tool to assess evolutionary algorithms for data mining problems,” Soft Comput., vol. 13, no. 3, pp. 307–318, Oct. 2008.
[24] Jesus Alcala-Fdez, Alberto Fernandez, Julian Luengo, Joaquın Derrac, and Salvador Garcıa, “KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework,” Multiple-Valued Logic and Soft Computing, vol. 17, no. 2-3, pp. 255– 287, 2011.
[25] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, “Uci repository of machine learning databases,” 1998.
[26] Reza Zafarani and Huan Liu, “Asu repository of social computing databases,” 1998.

Nesma Settouti is an Assistant professor at Tlemcen University, Algeria. She received her Engineer degree in Electrical Biomedical from the Tlemcen University in 2009. In 2011, she obtains a Magisterial degree in the same option. She is currently pursuing her PhD Thesis in the Biomedical Engineering Laboratory and LIMOS of Aubière (France), her research interests are in computer assisted medical decision support systems, ensemble methods, neural networks, clustering methods, optimization, classification and artificial intelligence. She had published a great deal of research studies published at national and international journals, conference proceedings as well as chapter book in computer assisted medical decision support systems.

Mohammed El Amine Bechar got his master degree in Biomedical Engineering with an emphasis in medical images processing in 2013, from the University of Tlemcen, Algeria. Currently, he is working towards his Ph.D. In Biomedical engineering laboratory at the University of Tlemcen, Algeria. His main interests are in the areas of computer assisted medical decision support systems, classification, artificial intelligence and image processing.

Mohammed Amine Chikh is a Professor at the Tlemcen University. He is graduated from The Electrical Engineering Institut (INELEC) of Boumerdes –Algeria in 1985 with engineering degree in Computer science and in 1992 with a Magister of Electronic from Tlemcen University. He also received a Ph.D in electrical engineering from the University of Tlemcen (Algeria) and INSA of Rennes (France) in 2005. Actually, he is currently Professor at Tlemcen University, Algeria, and head of CREDOM research team at Biomedical Engineering Laboratory. He conducted post-doctoral teaching and research at the University of Tlemcen. Pr Chikh has published over 90 journal and conference papers to date and is involved in a variety of funded research projects related to biomedical engineering. His is a member of several scientific conferences. His research interests have been in artificial intelligence, machine learning, and medical data.