

# Optimizing voting classification using cluster analysis on medical diagnosis data

Androniki Tamvakis  
University of the Aegean  
Department of Marine Sciences  
University Hill, Mytilene, Greece  
Phone: +302251036811  
atamvaki@mar.aegean.gr

Christos-Nikolaos  
Anagnostopoulos  
University of the Aegean  
Department of Cultural Technology  
and Communication  
University Hill, Mytilene, Greece  
Phone: +302251036624  
[canag@aegean.gr](mailto:canag@aegean.gr)

George Tsekouras  
University of the Aegean  
Department of Cultural Technology  
and Communication  
University Hill, Mytilene, Greece  
Phone: +302251036631  
gtsek@ct.aegean.gr

George Anastassopoulos  
Democritus University of Thrace  
Medical School, 8100,  
Alexandroupolis, Greece  
Phone: +302551030503  
[anasta@med.duth.gr](mailto:anasta@med.duth.gr)

## ABSTRACT

Voting ensemble method combines results of single classifiers aiming to offer improved classification performance. However, it is intuitively accepted that the combined classifiers during voting should be both diverse and accurate. In this study, we used the unsupervised method of cluster analysis in four datasets related to medical diagnosis in order to differentiate the single classifiers according to their individual results. Using this information we selected the most accurate among similar classifiers proposing the optimal classifier combination for each dataset. The results show that the estimated combination was actually the best performing during voting training for two of the datasets while in the other two it was one of those that outperformed single classifiers. The proposed methodology is a quick and easy tool for estimating classifier combinations that outperforms the single classifiers during voting.

## Categories and Subject Descriptors

- Computing methodologies~Ensemble methods
- Computing methodologies~Search methodologies

## Keywords

Voting ensemble method, cluster analysis, medical diagnosis, classification.

## 1. INTRODUCTION

Machine learning (ML) is a branch of artificial intelligence containing various techniques that learn and make prediction based on collected data [1]. Classification is one of the most popular ML applications, whereby an object (instance) needs to be

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

*16th EANN workshops*, September 25 - 28, 2015, Rhodes Island, Greece  
© 2015 ACM. ISBN 978-1-4503-3580-5/15/09...\$15.00  
DOI: <http://dx.doi.org/10.1145/2797143.2797156>

assigned into a group (with a specific label) based on a number of observed attributes related to that object (input variables) [2]. Classification tasks assessed by ML are commonly used in health sciences especially for successful medical diagnosis [3], disease prognosis [4], factor aetiology [5] and medical image analysis [6].

The vital role of classification has been led in the development of various algorithms (classifiers) differentiated in supervised or unsupervised depending on whether the instances of the dataset have a predefined label or not [7]. The most popular supervised classifiers includes rules, trees, instance based learning, multilayer perceptron and support vector machines while the most common unsupervised method is cluster analysis [8,9]. Recent research focuses on coupling existing ML techniques in order to produce more accurate schemas called ensemble methods (EM) [10]. The simplest and most popular way to combine classifiers is voting EM which spans numerous applications on classification tasks related to medical science [11-13].

Voting combines multiple supervised classifiers by collecting their predicted labels of each instance and predicting the label having the highest value (expressed as higher frequency or probability) [14]. However, the selection of the optimal classifier combination is currently an open issue. Previous studies show that the performance of the voting EM enhances the performance of the single classifiers and depends on both the accuracy and the diversity of the classifier results consisting the voting schema [10,15,16]. Thus, a combination consisting by classifiers that offer similar results does not contribute to the improvement of voting performance but instead increases both time requirements and system complexity [17]. On the other hand, just diverse results cannot ensure the accuracy improvement especially when the classifiers are poor performing [14].

The unsupervised technique of cluster analysis includes different algorithms that separate the instances of a dataset into homogenous parts (clusters) [18]. Cluster analysis is a significantly popular ML technique and its application specifically in medical sciences has shown advanced predictive power[19,20]. Moreover, there are promising EMs that take advantage of both supervised and unsupervised learning often by trying to improve the clustering accuracy incorporating the comprehension of

supervised classifiers [21,22]. However, the converse assessment of implementing cluster analysis in supervised learning is restricted.

In this paper we propose an integrated methodology for the estimation of the optimal combination of single classifiers that maximizes the classification performance of the voting algorithm. To this aim, the objectives of the study are (a) to assess the efficiency of base classifiers in different medical classification tasks, (b) to determine and analyze the classifier combinations with high voting performance (c) to identify the combination consisting of classifiers with the most diverse and accurate results based on cluster analysis (d) to test whether such combination has a corresponding high performance during voting training.

## 2. METHODS

### 2.1 Datasets

Four datasets (three available from UCI plus one related to the authors' interest [23]) were used for the training of the single classifiers and voting EM (Table 1). All datasets are related to medical science and contain variables expressing potential diagnosis factors of different human maladies. The datasets have different structural characteristics i.e. they differ in number of instances, number and type of input variables and number of labels of the target class. The diversity among the characteristics of the datasets is expected to result in different classification performances of the single classifiers that consists the voting EM. This will help the testing of the newly proposed methodology for the determination of successful classifier combinations.

**Table 1. List of datasets used along with their structure characteristics i.e. number of instances and input variables, type of input variables (c for categorical, i for integer, r for real) and number of the labels of the target class.**

Dataset name	N# of instances	N# and type of input variables	N# of labels of the target class
Arrhythmia	452	279 (c, i, r)	16
Breast cancer	286	9 (c)	2
Diabetes	768	8 (i, r)	2
Abdominal pain	516	16 (c, i, r)	7

### 2.2 Classifier training

Ten single classifiers representing all main machine learning (ML) categories (i.e. rules, trees, instance based learning, perceptron, sequential minimal optimization, regression and Bayes) were trained in order to compare and contrast their classification results (Table 2). The voting EM combines the results of the single classifiers by collecting the votes (i.e. predicted labels of the target class) from multiple classifiers and predicts the label of the target class yielding the highest value (expressed as number of votes or probability) [14]. In the present study, voting EM was trained by combining the results of the 10 single classifiers at all possible triads (i.e. 120 different combinations). Classifier triads were used because 3 is the minimum odd number for algorithm combinations and therefore it helps avoiding potential ties (when half classifiers votes a class label and half another) while it simplifies the voting training (saves computing time and complexity) [24]. The performance of voting EM according to the different classifier triads was explored in order to assess voting efficiency against the performance of single classifiers.

The classification of all single classifiers and voting EM was performed using WEKA machine learning package (version 3.6.9) [25]. The efficiency of the classifiers was evaluated using 10-fold

cross validation using default values (Table 2) and the performance was measured with the percentage of the correctly classified instances (CCI%). The objective of the newly proposed method is the identification of the classifier triad that optimizes the voting performance and thus it is not necessary to tune the base algorithms. Indeed, algorithm tuning will result to the maximization of the voting performance but the determination of the optimal triad will still be the crucial task under consideration.

### 2.3 Cluster analysis

Hierarchical cluster analysis (HCA) is a data mining method which gradually merges different observations in order to provide distinct homogenous sets (clusters) [26]. In this study, the 10 single classifiers (Table 2) are the observations wished to be merged according to their classification results for each dataset. Therefore, HCA quantifies the similarity of the single classifiers and merges them to provide clusters of classifiers presented in a dendrogram.

For the HCA implementation, the successful classification of an instance by a classifier was assigned a "1" score, whereas misclassification was assigned a "0" score. Using this binary assessment, the HCA merges the single classifiers using a specific metric of dissimilarity wishing to measure the distance between classifiers, as follows:

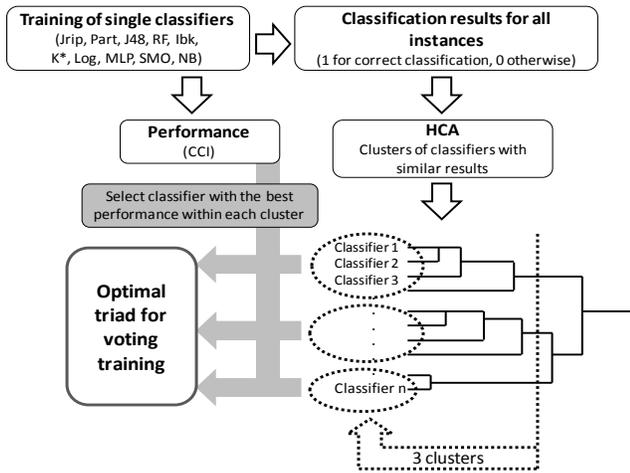
$$distance = \frac{N^{10} \cdot N^{01}}{n^2}$$

where  $N^{10}$  and  $N^{01}$  represent the number of instances correctly classified by one classifier but not by the other and vice versa (diagonal cells of a fourfold table), and  $n$  is the total number of instances. The above distance is calculated for all classifier couples (45 couples derived from the 10 selected classifiers) and using this information the classifiers are gradually merged to clusters using the classical between groups linkage method. The specific metric ranges from 0 to 1 and was selected due to the emphasis it offers to the pattern differences between classifiers [27]. The HCA was performed using IBM SPSS Statistical package (Version 20).

### 2.4 Optimal triad estimation

The proposed estimation of the optimal classifier triad participating voting EM has four steps (Fig 1). Firstly, the 10 single classifiers are trained individually in order to determine their binary classification results (i.e. 1 for every correctly classified instance and 0 for the misclassified one) and their overall performance (i.e. CCI). Secondly the binary classification results are used by the HCA method and the dendrogram containing the similar classifier clusters is constructed. Subsequently, using the dendrogram the three main clusters are defined. Finally, from each of the three clusters we select the classifier with the best overall performance and the optimal classifier triad is then been estimated.

The proposed optimal triad always contains the best performed classifier along with other two classifiers that are as dissimilar as possible with the best one, but with better performances between other similar classifiers (situated in the same cluster). This is a fast and easy way to select optimal classifier combinations for voting EM by satisfying both criteria of containing accurate as long as diverse classifiers. Thus, cluster analysis differentiates the classifiers based on their individual results (diversity) and using this information the best performing classifiers are selected (accuracy).



**Figure 1. Schematic diagram of the methodology followed for the selection of the optimal combination triad for voting training.**

The proposed triad was checked whether outperforms (a) the performances of the three classifiers that consists the triad and (b) the performances of all single classifiers. Furthermore, the proposed triad was checked for its actual voting performance by the determination of the position it holds between the set of the 120 classifier combination in terms of CCI%. Finally, the proposed methodology was measured for the time it requires in order to be completed and this time was compared with the exhaustive search of the optimal voting triad.

### 3. RESULTS

#### 3.1 Results of the single classifiers and voting EM

The classification performances of the single classifiers along with the corresponding performance of the best performing triad during voting for all datasets are presented in Table 2. There was neither a single classifier nor a classifier category that outperformed for all datasets. The performance of the single classifiers show low values and high variation for arrhythmia dataset while the inverse behavior recorded for dataset related with the abdominal pain.

The best performing triad combined by the voting EM has shown higher classification performance compared to the performance of individual classifiers for all of the databases (Table 2). Although the best triad of each dataset contained the best performing single classifier (i.e. J48 for breast cancer and abdominal pain, JRip for arrhythmia and SMO for diabetes), it did not consist by the three best performing single classifiers (e.g. RF for breast cancer and NB for abdominal pain were the 7<sup>th</sup> performed classifier among the 10 tested). Testing the performance of voting EM against single classifiers, it has been shown that a relative small number of triads (out of the 120 possible combinations) exceeded the performance of the single classifiers consisted the triad (Table 3). Moreover, a very small number of triads have shown better performance than the best single classifier for all datasets. Indeed, the percentage of triads that outperformed the single classifiers during voting was under 8% for three datasets and under 20% for diabetes dataset.

**Table 2. CCI of the single classifiers plus the corresponding percentage of the best triad during voting for all databases.**

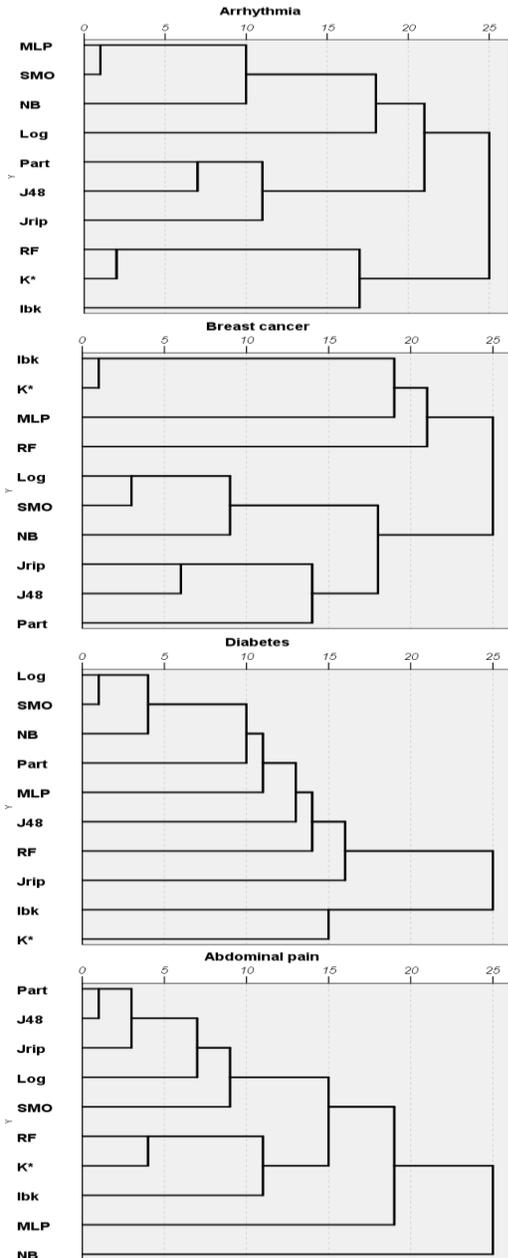
Classifiers	Description	Default WEKA values	Arrhythmia	Breast cancer	Diabetes	Abdominal pain
<b>Rules</b>						
JRip	Ripper algorithm for rule induction	Minimum weight of the instances in a rule = 2	<b>70.35</b>	70.99	76.04	85.05
Part	Rules from partial decision trees	Minimum N# of instances per rule = 2	63.94	71.33	75.27	80.62
<b>Trees</b>						
J48	C4.5 decision tree learner	Minimum N# of instances per leaf = 2	64.38	<b>75.52</b>	73.83	<b>85.67</b>
RF	Forest of random trees	N# of trees = 10	65.93	69.58	73.96	82.95
<b>Lazy</b>						
IBk	Nearest neighbor learner	$k = 1$	52.88	72.38	70.17	79.46
K*	Instance based learner with entropic distance function	Parameter of global blending = 20	55.97	73.43	69.14	80.04
<b>Functions</b>						
Log	Multinomial logistic regression	-	55.09	68.88	77.22	82.17
MLP	Multilayer Perceptron trained with backpropagation	N# of neurons = mean of the N# of input variables and the N# of labels of the target class	67.26	64.69	75.40	82.36
SMO	Sequential minimal optimization for support vector classification	Complexity parameter = 1	70.13	69.58	<b>77.34</b>	82.59
<b>Bayes</b>						
NB	The Naïve Bayes classifier using estimator classes	-	62.39	71.68	76.31	80.81
<b>Meta</b>						
Vote	Algorithm for combining classifier results	Combination rule = average of probabilities	71.25 JRip-RF-SMO	76.63 J48-RF-K*	77.86 J48-SMO-NB	86.62 JRip-J48-NB

**Table 3. Number and percentages (in parenthesis) of the triads with better performance during voting than the classifiers consisting each triad and the best performing single classifier for all datasets.**

	Arrhythmia	Breast cancer	Diabetes	Abdominal pain
Triads with better performance than the consisting classifiers	37 (30.8%)	28 (23.3%)	40 (33.3%)	46 (38.3%)
Triads with better performance than the best single classifier	8 (6.7%)	6 (5.0%)	23 (19.2%)	9 (7.5%)

### 3.2 Estimating and testing the optimal triad for voting training

The dendrograms exported by the HCA for all datasets are presented in Figure 2.



**Figure 2. Dendrograms of classifiers consisted by the HCA for each dataset.**

The first two dendrograms (i.e. of arrhythmia and breast cancer datasets) show clear diversified classifier clusters while the other two (i.e. of diabetes and abdominal pain datasets) show clusters created with the successive addition of one classifier to the already merged clusters.

The three clusters derived from each HCA dendrogram along with the proposed optimal classifier triad for voting training are presented in Table 4. For two datasets (i.e. arrhythmia and breast cancer) the proposed optimal triad was actually the best performed triad during voting (last column of Table 4). Moreover, for diabetes dataset the proposed triad was actually the 2<sup>nd</sup> best during voting but its performance has negligible difference from the performance of the best one (77.86% for the best and 77.85% for the 2<sup>nd</sup>). As for the abdominal pain dataset, the proposed triad achieved the 6<sup>th</sup> better performance of voting EM (CCI=86.04%). However, even for abdominal pain dataset the proposed methodology resulted in the determination of a triad (one out of the nine possible) (Table 3) that had higher performance than the best performing single classifier (i.e. J48 with CCI=85.67%). The latter is definitely happening also for the other three datasets since the proposed triads had the best voting performance.

**Table 4. The three classifier clusters based on HCA for all datasets along with the determination, the corresponding rank and the performance (CCI %) of the proposed classifier triad.**

Classifier clusters based on HCA	Proposed optimal classifier triad	Voting rank of the proposed classifier triad
<b>Arrhythmia</b>		
MLP,SMO,NB,Log	➔ SMO	
Part,J48,Jrip	➔ Jrip	1 <sup>st</sup>
RF,K*,Ibk	➔ RF	
<b>Breast cancer</b>		
Ibk,K*,MLP	➔ K*	
RF	➔ RF	1 <sup>st</sup>
Log,SMO,NB,Jrip,J48,Part	➔ J48	
<b>Diabetes</b>		
Log,SMO,NB,Part,MLP,J48,RF	➔ SMO	
Jrip	➔ Jrip	2 <sup>nd</sup> (77.85%)
Ibk,K*	➔ Ibk	
<b>Abdominal pain</b>		
Part,J48,Jrip,Log,S	➔ J48	
MO,RF,K*,Ibk	➔ MLP	6 <sup>th</sup> (86.04%)
MLP	➔ MLP	
NB	➔ NB	

The training time of the base classifiers differ along the datasets depending on its size and structure (Table 5). As a result, the required time of the exhaustive search of the optimal triad and the processed time of the newly proposed searching method are correspondently being changed. Thus, for arrhythmia the classifier

training, the exhaustive search and the proposed method needed the most time to be completed compared with the other datasets. That is due to the big number of input variables (i.e. 279) consisting the arrhythmia dataset. Moreover, the proposed method needed much less time to indicate the possible optimal triad for all datasets. This time is less than the 3% of the time that the exhaustive search needs to be completed.

**Table 5. Training time (in CPU sec) of the single classifiers plus the required time of the exhaustive search of the optimal triad and the proposed method.**

	Arrhythmia	Breast cancer	Diabetes	Abdominal pain
<i>Classifiers</i>				
Jrip	20	1	1	1
PART	25	2	2	2
J48	10	1	1	1
RF	18	3	8	2
IBk	5	1	1	1
K*	65	3	8	3
Log	5900	2	2	5
MLP	6000	90	20	65
SMO	9	2	2	3
NB	5	1	1	1
<i>Cluster analysis</i>				
HCA	2	2	2	2
<i>Method to identify the optimal triad</i>				
Exhaustive search	440,172	3,672	1,656	2,952
Proposed method	12,229	104	48	84

## 4. DISCUSSION

In the present study, 10 single classifiers corresponding to various ML categories were trained using four substantially different datasets related to medical science. The results show that there is no single classifier that is always most accurate for all datasets, in accordance with previous studies [28]. Furthermore we combined the results of the single classifiers in triads using the simplest but most popular EM method i.e. voting. The results indicate that there are classifier triads that outperform the performance of all single classifiers in agreement with the general acceptance combining multiple classifiers may obtain to a substantial accuracy improvement of classification tasks [29]. However, the number of classifier triads that offer better performance than single classifiers is rather small for all datasets (average percentage less than 10% of the total number of triads). Analyzing the voting optimal triad of each dataset we found out that they did not consisted by the three more accurate single classifiers in agreement with previous studies [14,30]. These findings show that the selection of the classifier triads that exceed single classifier performance and especially the determination of the optimal one are rather difficult tasks.

Two famous methods of supervised and unsupervised learning (i.e. ML classifiers and cluster analysis) are coupled by using cluster analysis in order to identify the optimal classifier combination for voting training. Moreover, the popular consensus that the two crucial factors that improve accuracy of EM are the diversity and the individual performance of classifier combinations [31] was also taken into account in the present study. Thus, the proposed methodology integrates diversity by the use of HCA and performance by the selection of the most accurate single classifiers among others with similar results. The

determined optimal triad of each dataset achieved improved performance during voting compared to the corresponding performance of the single classifiers. Therefore, the determined triads were found among the rather small number of triads that outperforms single classifiers for each dataset (i.e. 8, 6, 23 and 9 triads). Subsequently, the determined triad for three of the datasets (i.e. arrhythmia, breast cancer and diabetes) achieved the higher performance among all possible triads that trained the voting EM while for the last one (i.e. abdominal pain dataset) was among the 5% of best performed (6th among 120 triads). These findings show that the proposed methodology is efficient in determining a classifier triad with improved classification results.

The time that the proposed methodology needs to be completed is much lower than the corresponding of the exhaustive search of the optimal classifier triad. The required time can be extensive and even prohibitive when more time consuming classifiers are used (e.g. multiple neural networks with different architectures) or the training database has large number of instances or input variables. On the other hand, the proposed method needs only the initial training of the base classifiers and the application of the cluster analysis which is known as a computationally fast method dealing with a relative small number of input variables (i.e. number of selected classifiers) and binary data (i.e. correct classification or not) [32].

Considering that (a) voting EM can offer substantially improvement of classification performance than single classifiers, (b) the selection of the appropriate classifiers to participate in the voting EM is a difficult and open issue and (c) the exhaustive search of the optimal classifier combination is complex and time consuming especially for medical databases that are big in size, the newly proposed methodology is an efficient tool for the identification of the base classifiers that should be combined using voting EM. The method is recommended to single ML users from any scientific area wishing to achieve improved classification performance using their familiar classifiers. The users will find the method easy, quick, demanding no extra complex training (except for the HCA which can be implemented to any statistical or machine learning package) and flexible as they can use any type of classifiers (e.g. same or not).

## 5. ACKNOWLEDGMENTS

Part of this study was financially supported by the research project (code number 2522) "Synergy for the sustainable development and safe use of the Greek tourist beaches – Beachtour", which is implemented within the framework of the Action "Cooperation 2011 - Partnerships of Production and Research Institutions in Focused Research and Technology" of the Operational Programme "Competitiveness and Entrepreneurship" (OPCE II), (Action's Beneficiary: General Secretariat for Research and Technology- MIA-RTDI), and is co-financed by the European Regional Development Fund (ERDF) and the Greek State.

## 6. REFERENCES

- [1] Tamvakis, A., Trygonis, V., Miritzis, J., Tsirtsis, G., and Spatharis, S. 2014. Optimizing biodiversity prediction from abiotic parameters. *Environmental Modelling & Software* 53 (Mar. 2014), 112-120. DOI=<http://dx.doi.org/10.1016/j.envsoft.2013.12.001>.
- [2] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F., and Nielsen, H. 2000. Assessing the accuracy of prediction

- algorithms for classification: an overview. *Bioinformatics* 16 (Feb. 2000), 412-424.
- [3] Kononenko, I. 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 23 (Aug. 2001), 89-109. . DOI= [http://dx.doi.org/10.1016/S0933-3657\(01\)00077-X](http://dx.doi.org/10.1016/S0933-3657(01)00077-X).
- [4] Cruz, J. A. and Wishart, D. S. 2006. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics* 2, 59-77.
- [5] McKinney, B. A., Reif, D. M., Ritchie, M. D., and Moore, J. H. 2006. Machine learning for detecting gene-gene interactions: a review. *Applied Bioinformatics* 5 (Dec. 2006), 77-88. DOI= <http://dx.doi.org/10.2165/00822942-200605020-00002>.
- [6] Nanni, L., Lumini, A., and Brahnam, S. 2010. Local binary patterns variants as texture descriptors for medical image analysis. *Artificial Intelligence in Medicine* 49 (Jun. 2010), 117-125. DOI= <http://dx.doi.org/10.1016/j.artmed.2010.02.006>.
- [7] Sathya, R. and Abraham, A. 2013. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence* 2, 34-38. DOI= <http://dx.doi.org/10.14569/IJARAI.2013.020206>.
- [8] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. 2006. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review* 26 (Nov 2006), 159-190. DOI= <http://dx.doi.org/10.1007/s10462-007-9052-3>.
- [9] Tsekouras, G. E., Anagnostopoulos, C., Gavalas, D., and Dafni, E. 2007. Classification of Web documents using fuzzy logic categorical data clustering. In *Artificial Intelligence and Innovations: from Theory to Applications*, C. Boukis, A. Pnevmatikakis, L. Polymenakos, Eds. Springer US, 93-100. DOI= [http://dx.doi.org/10.1007/978-0-387-74161-1\\_11](http://dx.doi.org/10.1007/978-0-387-74161-1_11).
- [10] Dietterich, T. 2000. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*, Springer Berlin Heidelberg, 1-15.
- [11] Tanwani, A. K., Afridi, J., Shafiq, M. Z., and Farooq, M. 2009. Guidelines to select machine learning scheme for classification of biomedical datasets. In *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, C. Pizzuti, M. Ritchie, M. Giacobini, Eds. Springer Berlin Heidelberg, 128-139. DOI= [http://dx.doi.org/10.1007/978-3-642-01184-9\\_12](http://dx.doi.org/10.1007/978-3-642-01184-9_12).
- [12] Peng, C. R., Liu, L., Niu, B., Lv, Y. L., Li, M. J., Yuan, Y. L., Zhu, Y. B., Lu, W. C., and Cai, Y. D. 2011. Prediction of RNA-binding proteins by voting systems. *Journal of Biomedicine and Biotechnology* 2011, 506205, DOI= <http://dx.doi.org/10.1155/2011/506205>.
- [13] Huang, C. H., Peng, H. S., and Ng, K. L. 2015. Prediction of cancer proteins by integrating protein interaction, domain frequency and domain interaction data using machine learning algorithms. *BioMed Research International* 2015, 312047, DOI= <http://dx.doi.org/10.1155/2015/312047>.
- [14] Ruta, D. and Gabrys, B. 2005. Classifier selection for majority voting. *Information Fusion* 6 (Mar. 2005), 63-81.
- [15] Tan, A. C. and Gilbert, D. 2003. Ensemble machine learning on gene expression data for cancer classification. *Applied Bioinformatics* 2, S75-S83.
- [16] Kuncheva, L. I. and Hadjitodorov, S.T. 2004. Using diversity in cluster ensembles. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 1214-1219. DOI= <http://dx.doi.org/10.1109/ICSMC.2004.1399790>.
- [17] Zhou, Z. H., Wu, J., and Tang, W. 2002. Ensembling neural networks: Many could be better than all. *Artificial Intelligence* 137 (May 2002), 239-263.
- [18] Kaufman, L. and Rousseeuw, P.J. 1990. *Finding groups in data: An introduction to cluster analysis*. Wiley, New York.
- [19] Ma, L. Y., Chan, P., Gu, Z. Q., Li, F. F., and Feng, T. 2015. Heterogeneity among patients with Parkinson's disease: Cluster analysis and genetic association. *Journal of the Neurological Sciences* 351 (Apr 2015), 41-45. DOI= <http://dx.doi.org/10.1016/j.jns.2015.02.029>.
- [20] Chen, L., Lin, Z. X., Lin, G. S., Zhou, C. F., Chen, Y. P., Wang, X. F., and Zheng, Z. Q. 2015. Classification of microvascular patterns via cluster analysis reveals their prognostic significance in glioblastoma. *Human Pathology* 46 (Jan 2015), 120-128. DOI= <http://dx.doi.org/10.1016/j.humpath.2014.10.002>.
- [21] Dimitriadou, E., Weingessel, A., and Hornik, K. 2001. Voting-Merging: an ensemble method for clustering. In *Artificial Neural Networks - ICANN 2001*, G. Dorffner, H. Bischof, K. Hornik, Eds. Springer Berlin Heidelberg, 217-224.
- [22] Dudoit, S. and Fridlyand, J. 2003. Bagging to improve the accuracy of a clustering procedure. *Bioinformatics* 19, 1090-1099. DOI= <http://dx.doi.org/10.1093/bioinformatics/btg038>.
- [23] Iliou, T., Anagnostopoulos, C. N., Stephanakis, I., and Anastassopoulos, G. 2013. Combined classification of risk factors for appendicitis prediction in childhood. In *Engineering Applications of Neural Networks*, L. Iliadis, H. Papadopoulos, C. Jayne, Eds. Springer Berlin Heidelberg, 203-211. DOI= [http://dx.doi.org/10.1007/978-3-642-41016-1\\_22](http://dx.doi.org/10.1007/978-3-642-41016-1_22).
- [24] Kuncheva, L. I. 2004. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons Inc., Hoboken, New Jersey.
- [25] Hall, M., Frank, E., Holmes, G., Pfahringer B., Reutemann P., and Witten I.H. 2009. The WEKA Data Mining Software: an update. *ACM SIGKDD Explorations* 11 (Jun 2009), 10-18. DOI= <http://doi.acm.org/10.1145/1656274.1656278>.
- [26] Michaud, P. 1997. Clustering techniques. *Future Generation Computer Systems* 13 (Nov 1997), 135-147. DOI= [http://dx.doi.org/10.1016/S0167-739X\(97\)00017-4](http://dx.doi.org/10.1016/S0167-739X(97)00017-4).
- [27] IBM Corp. Released. 2011. *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY, IBM Corp.
- [28] Kittler, J., Hatef, M., Duin, R. W. D., and Matas, J. 1998. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (Mar 1998), 226-239. DOI= <http://dx.doi.org/10.1109/34.667881>.
- [29] Canuto, A. M. P., Abreu, M. C. C., de Melo Oliveira, L., Xavier, J., and Santos, A. d. 2007. Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for

- ensembles. *Pattern Recognition Letters* 28 (Mar 2007), 472-486. DOI = <http://dx.doi.org/10.1016/j.patrec.2006.09.001>.
- [30] Chou, J. S., Tsai, C. F., Pham, A. D., and Lu, Y. H. 2014. Machine learning in concrete strength simulations: Multination data analytics. *Construction and Building Materials* 73 (Dec 2014), 771-780. DOI=<http://dx.doi.org/10.1016/j.conbuildmat.2014.09.054>.
- [31] Shipp, C. A., and Kuncheva, L. I. 2002. Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion* 3 (Jun 2002), 135-148. DOI=[http://dx.doi.org/10.1016/S1566-2535\(02\)00051-9](http://dx.doi.org/10.1016/S1566-2535(02)00051-9).
- [32] Fraley, C., and Raftery, A. E. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 41, 578-588. DOI=<http://dx.doi.org/10.1093/comjnl/41.8.578>.