

Research Article

A Contribution to the Study of Ensemble of Self-Organizing Maps

Leandro Antonio Pasa,¹ José Alfredo Ferreira Costa,² and Marcial Guerra de Medeiros²

¹Department of Electrical Engineering, Federal University of Technology-Paraná, 85884-000 Medianeira, PR, Brazil

²Department of Electrical Engineering, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

Correspondence should be addressed to Leandro Antonio Pasa; pasa@utfpr.edu.br

Received 20 September 2015; Accepted 25 November 2015

Academic Editor: Costas Panagiotakis

Copyright © 2015 Leandro Antonio Pasa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study presents a factorial experiment to investigate the ensemble of Kohonen Self-Organizing Maps. Clusters Validity Indexes and the Mean Square Quantization Error were used as a criterion for fusing Kohonen Maps, through three different equations and four approaches. Computational simulations were performed with traditional dataset, including those with high dimensionality, not linearly separable classes, Gaussian mixtures, almost touching clusters, and unbalanced classes, from the UCI Machine Learning Repository and from Fundamental Clustering Problems Suite, with variations in map size, number of ensemble components, and the percentage of dataset bagging. The proposed method achieves a better classification than a single Kohonen Map and we applied the Wilcoxon Signed Rank Test to evidence its effectiveness.

1. Introduction

An ensemble is a collection of individual classifiers with different parameters which may lead to a higher generalization than when it is working separately, as well as a decrease in variance model and higher noise tolerance when compared to a single component [1]. Each classifier operates independently and generates a solution that is combined by the ensemble producing a single output, as illustrated in Figure 1.

For a successful outcome of the ensemble, it is necessary that the components generalize differently and the errors introduced by each component must be uncorrelated because there is no point combining models that adopt the same procedures and assumptions [2]. Furthermore, with a greater diversity among the components, the ensemble's performance is increased [3].

For the Kohonen Self-Organizing Maps (SOM) [4], uncorrelated errors can be obtained, for example, using different sets of training or using variations of training parameters for each neural network. The combination of the outputs of the SOM maps into a single output can be made by the fusion of the neurons of the maps that compose

the machine committee. These neurons must represent the same input space of the region; namely, the weight vectors of neurons to be fused should be very close.

The interest on ensemble methods has grown rapidly and we can find its applications in a variety of knowledge areas. In addition to ensemble or machine committee, there are several denominations in the literature, such as classifiers committees, combined classifiers, classifier ensemble, and multiple classifier systems.

There are many ensemble applications to solve problems in different areas. In [5] they compared merged maps with the traditional SOM for document organization and retrieval. As a criterion for combining maps, the Euclidean distance between neurons was used in order to select the neurons that were aligned, working with two maps each time, until all maps were fused into one. In [6] the proposed method outperforms the performance of the SOM in MSQE (Mean Squared Quantization Error) and topology preservation, by effectively locating the prototypes and relating the neighbor nodes. In [7] authors aimed to preserve map's topology in order to obtain the most truthful visualization of datasets. In [8] they investigate the use of relative validity indexes in cluster ensemble selection. These indexes select

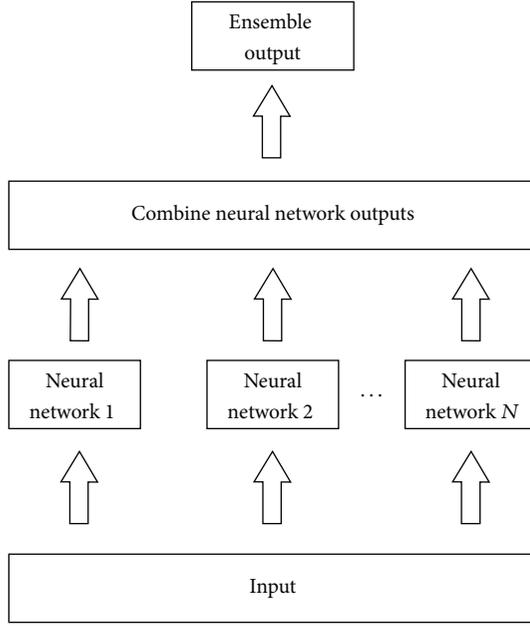


FIGURE 1: Block diagram of an ensemble.

the components with the highest quality to participate in the ensemble.

Different applications can be mentioned such as the mechanical fault diagnosis for high voltage circuit breakers [9], audio classification and segmentation [10], biomedical data [11], image segmentation [12], robotic [13], identification and characterization of computer attacks [14], unsupervised analysis of outliers on astronomical data [15], financial distress model [16], credit risk assessment [17], computer vision quality inspection of components [18], performance optimization of a classifier ensemble employed for target tracking in video sequences [19], disease prediction [20], and prediction of the duration of bus trips several days ahead [21]. Interesting reviews about this subject can be found in [22–25].

This work presents a contribution to the study of ensemble of Self-Organizing Maps with equal sizes. The goal is to improve the classification accuracy when compared to a single map. The paper is organized as follows: Section 2 presents some important concepts for this work, Section 3 shows the methodology proposed for the maps fusion, Section 4 presents the results obtained from the simulations, and Section 5 presents the conclusions and proposals for future works.

2. Background

2.1. Self-Organizing Maps. The Self-Organizing Maps (SOM) is a neural network model, known as a method for dimensionality reduction, data visualization, and data classification. It has competitive and unsupervised learning, performing a nonlinear projection of the input space R_e^p , with $p \gg 2$, in a grid of neurons arranged in two-dimensional array. This neural network has only two layers: an input and an output layer [4].

The network inputs x correspond to the p -dimensional vector space. The neuron i of the output layer is connected to all the inputs of the network, being represented by a vector of synaptic weights, also in p -dimensional space. Neurons in output layer are connected to the adjacent neurons through a neighborhood relationship that describes the topological structure of the map.

During the training phase, following a random sequence, input patterns x are compared to the neurons of the output layer. Through the Euclidean distance criterion a winning neuron, called BMU (Best Match Unit), is chosen and will represent the weight vector with the smallest distance to the input pattern; that is, the BMU will be the most similar to the input x . Assigning the winner neuron index by c , the BMU can be formally defined as a neuron according to

$$\|x - x_c\| = \arg \min_i \|x - w_i\|. \quad (1)$$

Equation (2) adjust the BMU weights and the neighboring neurons. Here t indicates the iteration of the training process, $x(t)$ is the input pattern, and $h_{ci}(t)$ is the nucleus of neighborhood around the winner neuron c :

$$w_i(t+1) = w_i(t) + \alpha(t) \cdot h_{ci}(t) \cdot [x(t) - w_i(t)]. \quad (2)$$

There are two ways of training a Kohonen map: sequential and batch training. The sequential training process of Self-Organizing Maps consists of three stages:

- (1) Competition: the input pattern is presented to the network one by one and the output layer neurons will compete against each other. The smallest Euclidean distance between the synaptic weights of neurons and the input pattern determines the only winner (BMU).
- (2) Cooperation: chosen the winner neuron, this becomes the center of a topological neighborhood of activated neurons.
- (3) Adaptation: neurons of certain topological neighborhood adjusted their synaptic weights so that nearby neurons to BMU suffer greater adjustments than the more distant neurons.

Figure 2 shows the process of weights updating of winner neuron and its neighborhood. The solid lines correspond to the situation before the updating and the dashed lines show the change due to the approach of BMU and their input pattern neighbors x . The bright and dark circles, respectively, represent the neurons before and after adjustment of the weights.

The Gaussian function is used to describe the influence of the BMU on its neighbors (update the synaptic weights). The activated neuron tends to excite more intensively neurons in its immediate vicinity and decays smoothly with the lateral distance. Equation (3) shows how the neuron j excitation is calculated in relation to the winning neuron i where d_{ij} is the distance between the neurons i and j .

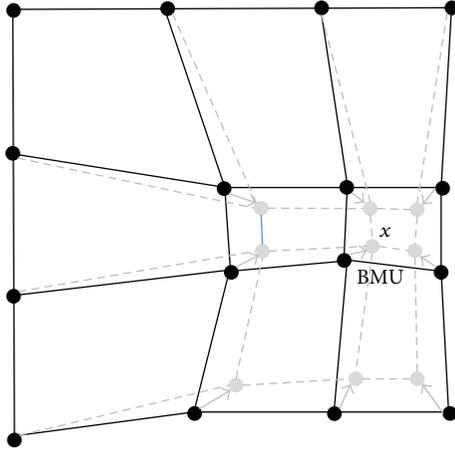


FIGURE 2: Weights updating [38].

At the end of the learning process, the output layer neurons represent a nonlinear approximation of the input patterns:

$$h_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right). \quad (3)$$

In batch training process all input patterns are presented to the network at once. Thus, the learning rate parameter $\alpha(t)$ is not necessary in (2) and randomness is not required in the presentation of the input vector to the network.

2.2. Cluster Validity Index. Cluster validity indexes (CVI) are used to evaluate the clustering algorithm results. Therefore, it is possible to check if an algorithm found the groups that best fit the data. The literature presents several CVI and most of them have high computational complexity, which can be a complicating factor in applications involving large data volumes. To overcome this problem, a modification in the CVI calculations was proposed in [16] using a vector quantization produced by Kohonen Map. The synaptic weight vectors (prototypes) are used instead of the original data and this leads to the decrease of the data amount, so the computational complexity for calculating the CVI decreases too. In the author's proposal, the hits should be part of the equation and not just prototypes, to avoid possible differences between the values calculated with all data and only the prototypes.

The CVI calculation always involves a measure of distance. The modification described in [26] changes the way these processes are performed. We explain this by comparing the two equations, the first for calculating the distance between two clusters C_i and C_j in the traditional way (4) and the other equation with the modifications for use with the SOM (5). One has

$$\delta_{i,j} = \frac{1}{|C_i||C_j|} \sum_{\substack{z \in C_i \\ y \in C_j}} d(x, y). \quad (4)$$

In (4), $d(x, y)$ is a distance measure, and $|C_i|$ and $|C_j|$ refer to the clusters' amount of points C_i and C_j , respectively. When the amount of those points is high, the computational complexity is also high.

Equation (5) shows the modification considering the hits [26]:

$$\delta_{i,j} = \frac{1}{|C_i||C_j|} \sum_{\substack{w_i \in W_i \\ w_j \in W_j}} h(w_i) \cdot h(w_j) \cdot d(w_i, w_j), \quad (5)$$

where W_i and W_j are the SOM prototype sets that represent the clusters C_i and C_j , respectively, $d(w_i, w_j)$ is the same distance measure type of (4), $h(w_i)$ is the prototype's hits w_i belonging to W_i , and $h(w_j)$ is the prototype's hits w_j belonging to W_j .

Equation (5) presents a lower computational cost since the quantities involved, w_i and w_j , are lower than C_i and C_j . The inclusion of the prototypes' hits leads to error minimization caused by the vector quantization that Kohonen Map produces, since it introduces in the calculation the approach for the points density in the input space, here represented by prototypes.

2.3. Bagging. There are three steps to obtain the components of an ensemble: generation, selection, and combination of individual components. In the component generation step it is necessary diversity among the components to ensure that each network generalizes differently. It can be achieved with random initialization of the weights for each component, the variation in the architecture of the neural network, the variation of the training algorithm or data resampling, among others.

In data resampling case, the Bagging [27] is a widely used technique. It consists in generating different training subsets from a single set via resampling with replacement. The samples selection from the original set has uniform probability distribution. Therefore, there will be some samples selected more than once, because of the replacement. There is the possibility that components do not generalize satisfactorily, but the ensemble of these models can result in a better generalization than each one of the individual components.

3. Materials and Methods

The method consists in the fusion of Kohonen maps formed from fractions of dataset created by bagging algorithm. The process overview is shown in Figure 3. The fusion occurs between two neurons that have the minimum Euclidean distance between them, indicating that they represent the same region of the input space.

The bagging method was used for generating subsets (resampling) from the training set, with uniform probability and replacement. The SOM was applied to each subset produced by bagging and all maps were segmented by k -means algorithm. Since not all the components should be used, because it can affect the final ensemble performance [28], the way to select these candidates will be through cluster validity index, modified for use with SOM, proposed by [26].

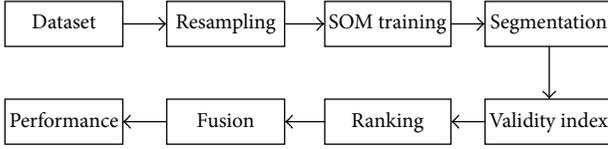


FIGURE 3: Equal sizes maps fusion process overview.

In this work the cluster validity indexes CDbw [29], Calinski-Harabasz [30], generalized Dunn [31], PBM [32], and Davies-Bouldin [33] were used as parameters in fusion decision process (consensus function).

The ranking-based selection method can be described in this way: the candidates to compose the ensemble are ranked based on clusters validity indexes values, which will measure the individual performance of each component.

3.1. Fusion Equations. We tested three different equations for neurons fusion. The first, equation (6), is an arithmetic average between the weight vectors of the neurons to be fused:

$$w_c = \frac{w_i + w_j}{2}. \quad (6)$$

The purpose of second equation was to avoid the influence of neurons without hits, as shown in

$$w_c = \frac{w_i \cdot h_i + w_j \cdot h_j}{h_i + h_j}. \quad (7)$$

Through (8), the third equation tested, we investigated whether the weighted fusion by CVI could result in an improvement in classification accuracy:

$$w_c = \frac{w_i \cdot h_i \cdot CV_i + w_j \cdot h_j \cdot CV_j}{h_i \cdot CV_i + h_j \cdot CV_j}, \quad (8)$$

where w_c is the fusion result of neurons w_i and w_j , considering h_i and h_j the prototype's hits and CV_i and CV_j the CVI from maps to be fused.

3.2. Factorial Experiment. We defined a factorial experiment to validate our approach. The maps size was varied in 10×10 , 15×15 , and 20×20 . Through the bagging 10 sets were generated containing the following quantities of subsets: 10, 20, and 30 up to 100 subsets. These subsets were created with different data rates: 50%, 70%, and 90% of the training set; that is, for each map size, it has ten different numbers of subsets and three different values of bagging percentage. This combination results in 900 maps. Thus, the first experiment map has the size equal to 10×10 , with 10 subsets generated by bagging with a percentage of 50%, as show in Figure 4.

3.3. Approaches. Four distinct approaches were defined to maps fusion, as a combination between ranked maps and maps fusion criteria. The ranking was based in five CVI and MSQE (Mean Square Quantization Error), defined in [6], and the maps fusion was based in five CVI and MSQE improvement criterion:

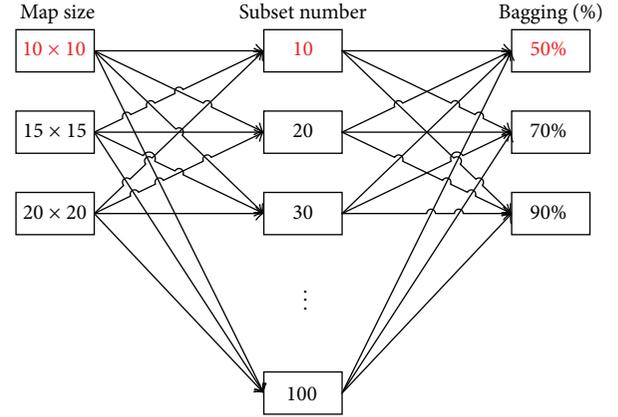


FIGURE 4: Structure of the factorial experiment.

- (A) Maps ranked by CVI and fused by MSQE improvement criterion.
- (B) Maps ranked by MSQE and fused by CVI improvement criterion.
- (C) Maps ranked by CVI and fused by CVI improvement criterion.
- (D) Maps ranked by MSQE and fused by MSQE improvement criterion.

The MSQE indicates how well the units of the map approximate the data on the dataset [7]; that is, it can be employed to evaluate the adaptation quality to the data [6].

In the first approach, the maps were ranked by five CVI, but the fusion process was controlled by MSQE improvement of the fused map. In the second approach, the maps were ranked by MSQE and the fusion process was validated by each one CVI improvement. In third approach, maps were ranked by each CVI and the fusion was controlled by CVI increase; that is, if the maps were ranked according to the Davies-Bouldin cluster validity index, the fusion will be controlled by the improvement of this index. At last, in the fourth approach the maps were ranked by MSQE and fused by MSQE improvement. Figure 5 shows the fusion process.

3.4. Proposed Fusion Algorithm. The proposed algorithm was applied in each one of the combinations shown in Figure 4 and according to approaches in Figure 5. It can be described as follows.

Algorithm 1 (overall fusion process).

Input. Dataset, map size, subset number, and bagging percentage.

- (1) Generate, through the bagging, subsets from the training set.
- (2) Apply the SOM to the subsets generated, resulting in N_i ($i = 1, 2, \dots$) maps with the same dimension.
- (3) Segment the maps, with k -means algorithm.

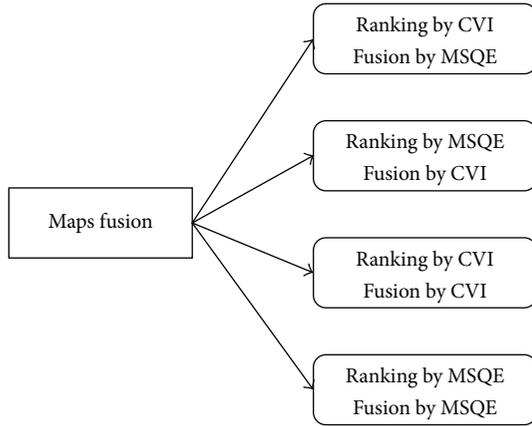


FIGURE 5: The four approaches to maps fusion.

- (4) Calculate the CVI and MSQE for each N_i ($i = 1, 2, \dots$) map and rank them according to this value, from the best index to the worst.
- (5) The base map (one with the best CVI or MSQE value) is fused with the next best map, according to the sorted CVI or MSQE and (6), (7), and (8).
- (6) Calculate the CVI or MSQE of fused map.
- (7) If there is an improvement in the CVI or MSQE value, the map resulting from the fusion becomes the base map, then return to step (5) until all maps are fused in this way. Otherwise, discard the fusion between these two maps and return to step (5) until all maps are fused.

Output. Fused maps.

The main parameters of SOM were hexagonal lattice, Gaussian neighborhood, initial training radius equal to 4, and final training radius equal to 1 and 2000 training epochs.

Due to random initialization, k -means algorithm is repeated 10 times and the result with the smallest sum of squared errors is selected.

4. Results and Discussion

4.1. Datasets. The proposed method was tested through datasets with different characteristics from the UCI Repository [34] and from Fundamental Clustering Problems Suite (FCPS) [35], as shown in Table 1. Lines with missing values were removed from the datasets.

4.2. Results. The aim of the simulation was to verify in which map size the number of subsets and bagging percentage configuration best accuracy value could be obtained. Each experiment was run 10 times and mean values for a 95% confidential interval were obtained. The results were compared with a single SOM map, to verify the improvement in classification accuracy.

TABLE 1: Datasets.

Dataset	Repository	Instances	Attributes	Classes
Chainlink	FCPS	1000	3	3
Engytime	FCPS	4096	2	2
Lsun	FCPS	400	2	3
Tetra	FCPS	400	3	4
TwoDiamonds	FCPS	800	2	2
Wingnut	FCPS	1070	2	2
BC Wisconsin	UCI	699	9	2
Column	UCI	310	6	2
Heart	UCI	303	75	2
Ionosphere	UCI	351	34	2
Iris	UCI	150	4	3
PimaIndians	UCI	768	8	2
Seeds	UCI	210	7	3
Wine	UCI	178	33	3

4.2.1. General Result. Table 2 shows the best experimental results for each dataset; that is, which approach and which CVI resulted in the best accuracy value for each dataset.

The column “Equation” shows the best fusion equation for each dataset. Equation (6) is the simplest of the three tested equations, being just a simple average between two neurons weight vectors. Equation (7) is a weighted average of the hits neurons. Equation (8) makes a weighted average with hits and by CVI, the most complex equation. The results showed an equilibrium between (6) and (7) and (8) achieved the best accuracy for only two bases.

The column “Approach” refers to the way the maps were ranked and fused, as specified in Section 3.3. The best accuracy values were achieved with approaches B (maps ranked by MSQE and fused by CVI improvement criterion) and C (maps ranked by CVI and fused by CVI improvement criterion).

The column “CVI” shows which CVI was used to achieve the best fusion accuracy in relation to the single map, for each dataset (DB means Davies-Bouldin index and CH means Calinski and Harabasz index). The Davies-Bouldin index was the one that achieved the best fusion results over other CVI for these datasets and PBM index did not lead to good results.

The column “Map size” refers to the map size that has achieved the best accuracy value. The 15×15 map did not get the best results for these datasets.

The column “Bagging percentage” shows that, for the majority of dataset, using 90% of the samples for bagging produces better results. A high percentage of the bagging value has a wider range of data for training, leading to better fusion accuracy result. These results show that it is possible to get good accuracies without using the entire training set.

The column “Subsets number” shows the number of subsets that were evaluated to find the best accuracy. For most datasets used in this experiment the best accuracies were obtained with large amounts of subsets. This is an important conclusion: the number of maps available for fusing influences the accuracy that can be achieved.

TABLE 2: Best accuracies results.

Dataset	Equation	Approach	CVI	Map size	Bagging (%)	Subsets number	Single SOM accuracy	Proposed algorithm accuracy
BC Wisconsin	(6)	B	DB	20 × 20	90	80	0.9634	0.9661
Chainlink	(8)	B	DB	10 × 10	70	80	0.6654	0.7685
Column	(7)	C	DB	10 × 10	70	100	0.6674	0.7219
Engytime	(6)	A	CH	20 × 20	70	10	0.9538	0.9587
Heart	(6)	B	Dunn	20 × 20	90	30	0.7879	0.8037
Ionosphere	(7)	C	DB	10 × 10	90	20	0.7037	0.7211
Iris	(6)	B	Dunn	10 × 10	50	70	0.9200	0.9260
Lsun	(8)	C	DB	10 × 10	90	100	0.7500	0.9850
PimaIndians	(7)	C	CH	20 × 20	50	60	0.6458	0.6939
Seeds	(7)	C	DB	20 × 20	70	60	0.8809	0.9243
Tetra	(6)	B	Dunn	20 × 20	90	90	1	1
TwoDiamonds	(6)	C	Dunn	10 × 10	90	50	1	1
Wine	(6)	B	CDbw	10 × 10	90	80	0.9382	0.9635
Wingnut	(7)	C	CDbw	10 × 10	50	100	0.7539	0.8983

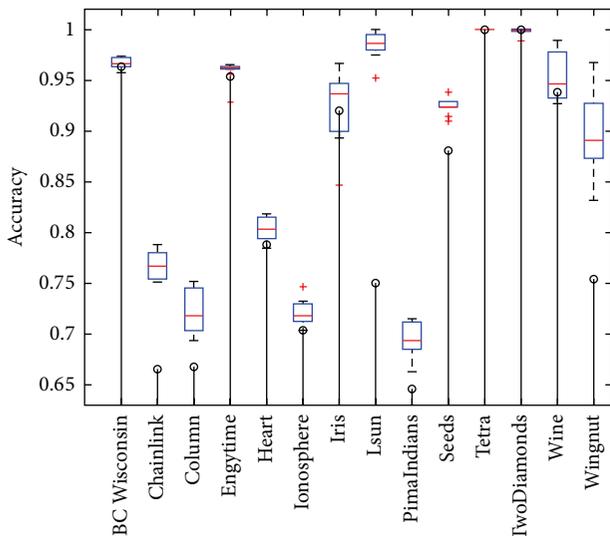


FIGURE 6: Classification accuracy for each dataset.

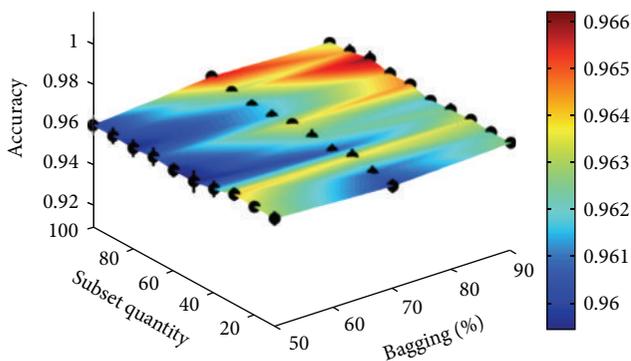


FIGURE 7: BC Wisconsin accuracy.

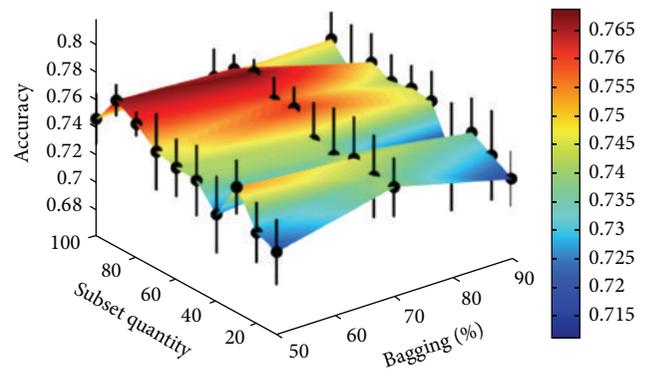


FIGURE 8: Chain-link accuracy.

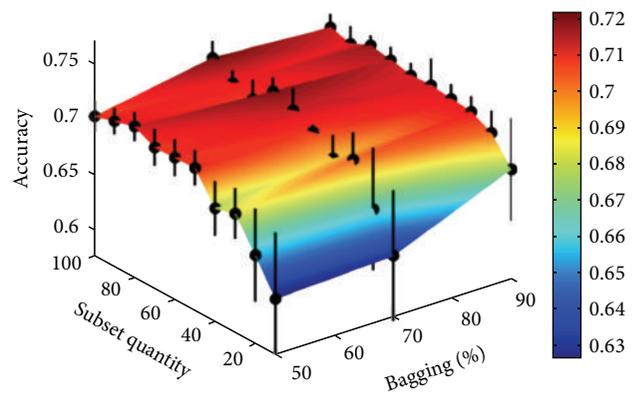


FIGURE 9: Column accuracy.

Figure 6 shows the classification accuracy obtained for each dataset. Black vertical lines show the single SOM accuracy value. As we can see, in all situations the accuracies achieved by the proposed method were higher than the classification accuracies for a single map, except for Tetra

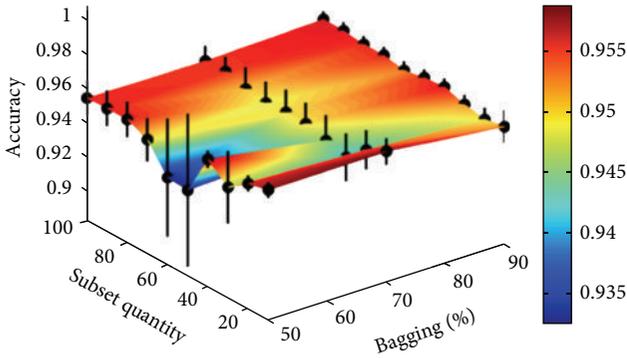


FIGURE 10: Engtime accuracy.

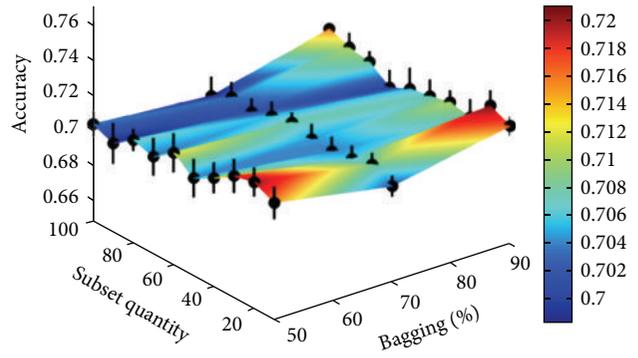


FIGURE 12: Ionosphere accuracy.

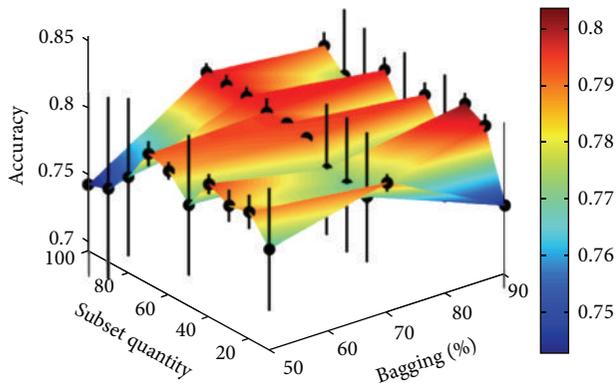


FIGURE 11: Heart accuracy.

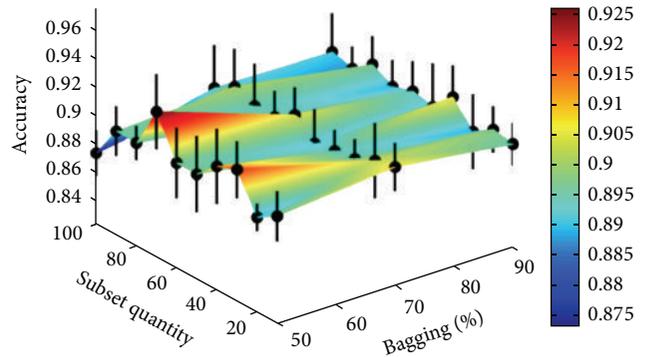


FIGURE 13: Iris accuracy.

and TwoDiamonds datasets. Thus, we show that through Self-Organizing Map fusion we can raise the value of classification accuracy.

In [36, 37] authors present ways to compare the performance of classifiers over multiple datasets. Among the tests presented in those papers, we chose to use the nonparametric Wilcoxon Signed Rank Test, at the significance level 0.05. This test was performed to evaluate whether the result obtained by Kohonen maps committee was statistically different from the accuracies obtained by a single SOM. The null hypothesis H_0 was that there is no difference between the accuracies obtained by the single map and the maps committee.

The Wilcoxon test rejected the null hypothesis, with p value equal to 0.0004882; that is, the proposed method obtained accurate results that are different and better than the accuracy of a single map.

4.2.2. Accuracy Results according to Bagging Percentage and Subset Quantity. Figures 7–20 show the accuracy (z-axis) obtained for each dataset according to bagging percentage (x-axis) and subsets quantity (y-axis) variation. Black lines represent the confidence interval for each experiment. In the 3D figures, hot regions represent the highest accuracy obtained and cold regions show lower accuracies obtained by the experiment. For example, Figure 7 shows that higher accuracies for BC Wisconsin dataset were achieved with

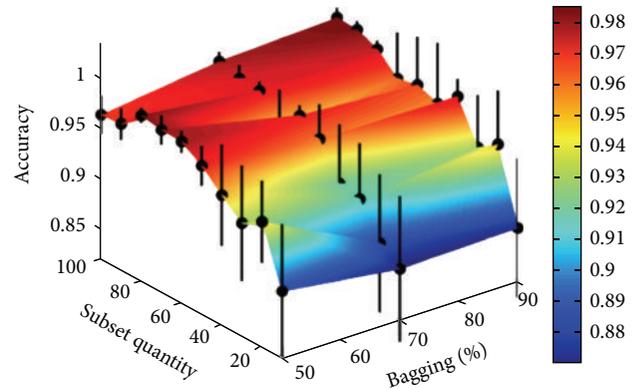


FIGURE 14: Lsun accuracy.

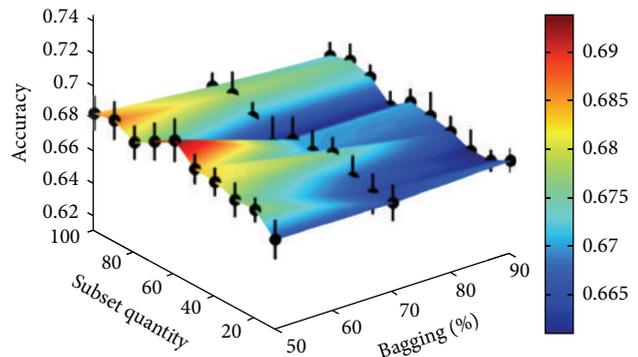


FIGURE 15: PimaIndians accuracy.

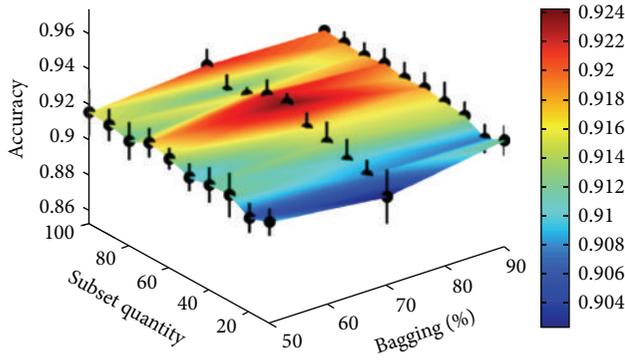


FIGURE 16: Seeds accuracy.

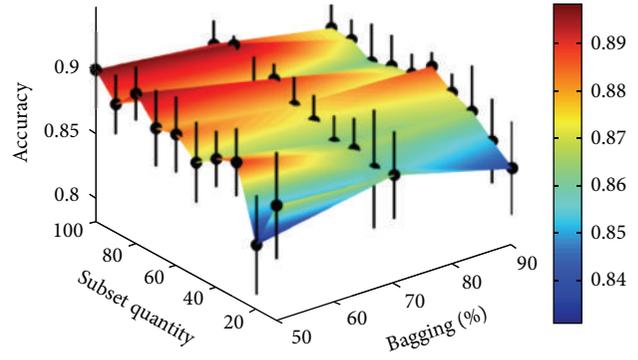


FIGURE 20: Wingnut accuracy.

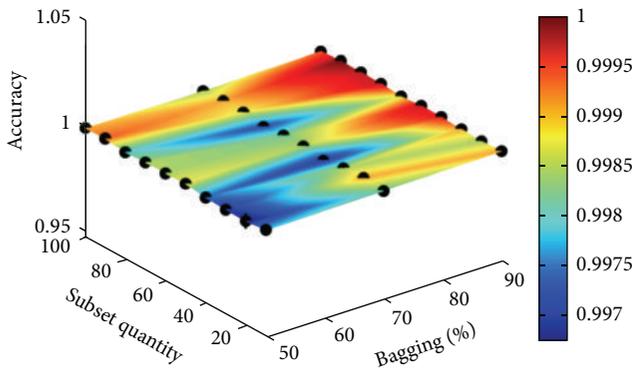


FIGURE 17: Tetra accuracy.

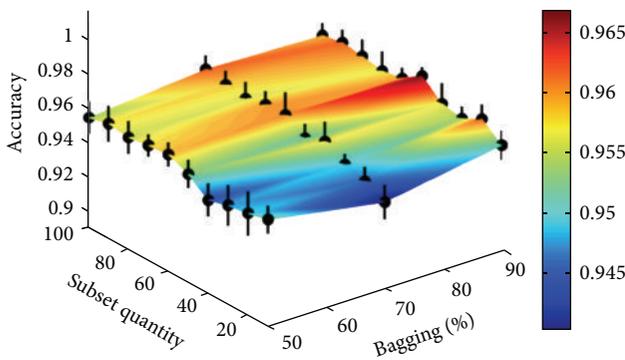


FIGURE 18: Wine accuracy.

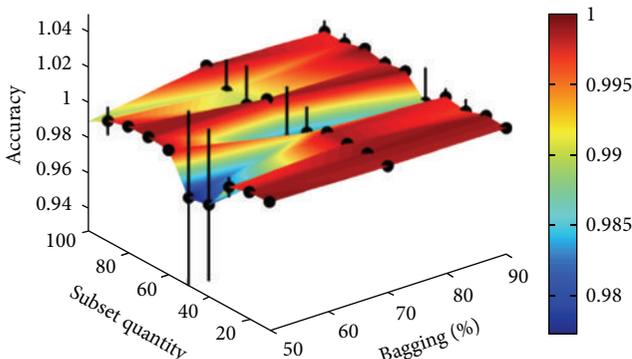


FIGURE 19: TwoDiamonds accuracy.

a combination of higher amounts of subsets and a higher percentage of bagging.

5. Conclusions

This work investigated the possibility of fusing Kohonen maps ranked by cluster validity indexes and MSQE. Computer simulations evaluated different setting for map size, subsets number, and different amounts (percentage) of the training set. Datasets with distinct characteristics, including unbalanced classes, were selected to test the proposed method.

The Kohonen Map fusion method achieves values similar or superior to the ones observed in regular SOM (single map). Using a percentage of the training data (not all data), the proposed model achieved satisfactory results compared to SOM, which used all the training set.

The best results were obtained with high amounts of subsets and with high bagging percentage. Among the five CVI used in this study, only the PBM index did not lead to good results. Good results were achieved by ranking the maps by MSQE and fusing by CVI criterion and by ranking the maps by CVI and fusing by CVI criterion.

An important contribution of this paper is to show that we can use CVI and MSQE as a function consensus in ensemble of Self-Organizing Maps.

Future work may explore the influence of certain parameters such as the number of maps, the adjustments in SOM training, and the segmentation method and evaluate strategies to enhance visualization of the fused maps.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings*, vol. 1857 of *Lecture Notes in Computer Science*, pp. 1–15, Springer, Berlin, Germany, 2000.

- [2] M. P. Perrone and L. N. Cooper, "When networks disagree: ensemble methods for hybrid neural networks," in *Neural Networks for Speech and Image Processing*, pp. 126–142, Chapman and Hall, 1993.
- [3] J. F. Díez-Pastor, J. J. Rodríguez, C. I. García-Osorio, and L. I. Kuncheva, "Diversity techniques improve the performance of the best imbalance learning ensembles," *Information Sciences*, vol. 325, pp. 98–117, 2015.
- [4] T. Kohonen, *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*, Springer, Berlin, Germany, 2nd edition, 1997.
- [5] A. Georgakis, H. Li, and M. Gordan, "An ensemble of SOM networks for document organization and retrieval," in *Proceedings of the International Conference on Adaptive Knowledge Representation and Reasoning (AKRR '05)*, Espoo, Finland, 2005.
- [6] C. Sandoval, R. Salas, S. Moreno, and H. Allende, "Fusion of self-organizing maps," in *Computational and Ambient Intelligence*, vol. 4507 of *Lecture Notes in Computer Science*, pp. 227–234, Springer, Berlin, Germany, 2007.
- [7] E. Corchado and B. Baruque, "WeVoS-ViSOM: an ensemble summarization algorithm for enhanced data visualization," *Neurocomputing*, vol. 75, pp. 171–184, 2012.
- [8] M. C. Naldi, A. C. P. L. F. Carvalho, and R. J. G. B. Campello, "Cluster ensemble selection based on relative validity indexes," *Data Mining and Knowledge Discovery*, vol. 27, no. 2, pp. 259–289, 2013.
- [9] J. Zhang, M. Liu, K. Wang, and L. Sun, "Mechanical fault diagnosis for HV circuit breakers based on ensemble empirical mode decomposition energy entropy and support vector machine," *Mathematical Problems in Engineering*, vol. 2015, Article ID 101757, 6 pages, 2015.
- [10] S. Zahid, F. Hussain, M. Rashid, M. H. Yousaf, and H. A. Habib, "Optimized audio classification and segmentation algorithm by using ensemble methods," *Mathematical Problems in Engineering*, vol. 2015, Article ID 209814, 11 pages, 2015.
- [11] S. J. Fodeh, C. Brandt, T. B. Luong et al., "Complementary ensemble clustering of biomedical data," *Journal of Biomedical Informatics*, vol. 46, no. 3, pp. 436–443, 2013.
- [12] Y. Jiang and Z.-H. Zhou, "SOM ensemble-based image segmentation," *Neural Processing Letters*, vol. 20, no. 3, pp. 171–178, 2004.
- [13] K. H. Low, W. K. Leow, and M. H. Ang Jr., "An ensemble of cooperative extended kohonen maps for complex robot motion tasks," *Neural Computation*, vol. 17, no. 6, pp. 1411–1445, 2005.
- [14] L. L. DeLooze, "Attack characterization and intrusion detection using an ensemble of self-organizing maps," in *Proceedings of the IEEE Information Assurance Workshop*, pp. 108–115, IEEE, West Point, NY, USA, June 2006.
- [15] D. Fustes, C. Dafonte, B. Arcay et al., "SOM ensemble for unsupervised outlier analysis: application to outlier identification in the Gaia astronomical survey," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1530–1541, 2013.
- [16] C.-F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress," *Information Fusion*, vol. 16, pp. 46–58, 2014.
- [17] R. F. Lopez and J. M. R. Jeronimo, "Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5737–5753, 2015.
- [18] S. K. Goumas, I. N. Dimou, and M. E. Zervakis, "Combination of multiple classifiers for post-placement quality inspection of components: a comparative study," *Information Fusion*, vol. 11, no. 2, pp. 149–162, 2010.
- [19] I. Visentini, L. Snidaro, and G. L. Foresti, "Diversity-aware classifier ensemble selection via f-score," *Information Fusion*, vol. 28, pp. 24–43, 2016.
- [20] C. H. Weng, T. C. Huang, and R. P. Han, "Disease prediction with different types of neural network classifiers," *Telematics and Informatics*, vol. 33, no. 2, pp. 277–292, 2016.
- [21] J. M. Moreira, A. M. Jorge, J. F. Sousa, and C. Soares, "Improving the accuracy of long-term travel time prediction using heterogeneous ensembles," *Neurocomputing*, no. 150, pp. 428–439, 2015.
- [22] N. C. Oza and K. Tumer, "Classifier ensembles: select real-world applications," *Information Fusion*, vol. 9, no. 1, pp. 4–20, 2008.
- [23] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. 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 Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.
- [24] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 3, pp. 337–372, 2011.
- [25] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, no. 1, pp. 3–17, 2014.
- [26] M. L. Gonçalves, M. L. De Andrade, J. A. Ferreira Costa, and J. Zullo, "Data clustering using self-organizing maps segmented by mathematic morphology and simplified cluster validity indexes: an application in remotely sensed images," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '06)*, pp. 4421–4428, IEEE, Vancouver, Canada, July 2006.
- [27] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [28] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial Intelligence*, vol. 137, no. 1-2, pp. 239–263, 2002.
- [29] M. Halkidi and M. Vazirgiannis, "A density-based cluster validity approach using multi-representatives," *Pattern Recognition Letters*, vol. 29, no. 6, pp. 773–786, 2008.
- [30] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, 1985.
- [31] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 3, pp. 301–315, 1998.
- [32] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognition*, vol. 37, no. 3, pp. 487–501, 2004.
- [33] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224–227, 1979.
- [34] K. Bache and M. Lichman, *Machine Learning Repository*, School of Information and Computer Sciences, University of California, Irvine, Irvine, Calif, USA, 2013, <http://archive.ics.uci.edu/ml>.
- [35] A. Ultsch, "Clustering with SOM. U*C," in *Proceedings of the Workshop on Self-Organizing Maps (WSOM '05)*, pp. 75–82, Paris, France, 2005.
- [36] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

- [37] S. García and F. Herrera, “An extension on ‘statistical comparisons of classifiers over multiple data sets’ for all pairwise comparisons,” *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.
- [38] J. Vesanto, *Data exploration process based on the self-organizing map [Ph.D. thesis]*, Helsinki University of Technology, 2002.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

