

Combining Overall and Local Class Accuracies in an Oracle-based Method for Dynamic Ensemble Selection

Leila M. Vriesmann*, Alceu S. Britto Jr[†], Luiz S. Oliveira[‡], Alessandro L. Koerich[§] and Robert Sabourin[§]

*Department of Informatics
State University of Ponta Grossa (UEPG)
Ponta Grossa, Parana, Brazil
Email: lmvriesmann@uepg.br

[†]Postgraduate Program in Informatics
Pontifical Catholic University of Parana (PUCPR)
Curitiba, Parana, Brazil
Email: alceu@ppgia.pucpr.br

[‡]Department of Informatics
Federal University of Parana (UFPR)
Curitiba, Parana, Brazil
Email: lesoliveira@inf.ufpr.br

[§]École de Technologie Supérieure (ÉTS)
University of Quebec
Montreal, Quebec, Canada
Email: {alessandro.koerich, robert.sabourin}@etsmtl.ca

Abstract—This paper presents a k -nearest oracle-based dynamic ensemble selection method in which overall local accuracy (OLA) and local class accuracy (LCA) are combined into a two-step selection scheme. The OLA and LCA are computed on the neighborhood of the test pattern in a validation set to filter out the classifiers selected by the k -nearest oracles. The complementary information of OLA and LCA has shown to be an interesting alternative to approximate the classification performance to that estimated for the oracle of the initial pool of classifiers. The results were compared with the recognition rates of the majority voting of all classifiers in the initial pool, and also with the recognition rates of related classifier and ensemble selection methods which have inspired the proposed method and its variants. The proposed method achieved the best results on 5 out of 8 experiments using small and large datasets of different applications.

I. INTRODUCTION

Multiple Classifier Systems (MCSs) have been an interesting alternative to deal with the wide variability usually observed in most pattern recognition problems [1]. The rationale behind that is take advantage of the possible complementarity of diverse classifiers. It is expected that different classifiers are experts on different regions of the problem feature space.

In such a context, the dynamic selection of classifier(s) (DSC) has shown to be a promising strategy. During the testing phase of a DSC-based method, one classifier or an ensemble of them is chosen from a pool composed of diverse

classifiers for each test instance. The rationale behind this is to select the most appropriate classifier(s) for each sample since they represent different classification difficulties. In addition, selecting an ensemble rather than a single classifier is justified by current practices. With an ensemble, we can avoid the risk related to the choice of a single classifier over the rest for each test instance [13].

In a DSC approach, if the classifier(s) that gives the correct label for each instance is always selected, we can say that the oracle was reached [4]. Thus, the DSC methods crave for having their classification accuracy close to the oracle performance. Based on that, the concept of oracle is usually present in the evaluation of the proposed methods. This means that the proposed method is compared against the upper limit in terms of performance of a given pool of classifiers. The oracle performance is estimated by considering that if at least one classifier can correctly classify a particular test sample, then the pool can also do so as well.

An interesting review of dynamic classifier selection methods is presented in [12]. The authors propose a taxonomy, in which the main contributions to DSC are categorized according to how they measure the competence of the classifiers to be selected. What is mostly commonly seen with this approach is the use of a partitioning scheme based on the NN-rule to define the neighborhood of a test instance in the feature space during the testing phase. In this case, the competence

of each classifier is defined on a local region on the feature space represented by the training or validation dataset. One may find measures of competence based on pure accuracy (overall local accuracy or local class accuracy) [22], ranking of classifiers [18], probabilistic information [6], [15], classifier behavior calculated on output profiles [7], [17], [2], and oracle information [14], [13]. Moreover, we may find measures that consider interactions among classifiers, such as diversity [20], [19], [16], ambiguity [10], [5], [2] or other grouping approaches [23][24].

Among those different approaches, three simple and performing competence measures are the inspiration to the proposed method. Two of them were proposed by Woods et al [22] for its dynamic classifier selection approach based on accuracy, named *DS-LA*. The first measure calculates the overall local accuracy (OLA) of the classifiers in the local region of the feature space close to the unknown pattern. The OLA of each classifier is computed as the percentage of the correct recognition of the samples in the local region. In the second one, they calculate the local class accuracy (LCA), *i.e.*, the percentage of correct classifications within the local region, but considering only those examples where the classifier has given the same class as the one it gives for the unknown pattern. In both versions of the *DS-LA* method, OLA and LCA-based, the partitioning of the feature space is defined based on the k -nearest neighbors of the unknown pattern in the training dataset during the testing phase. Moreover, in both algorithms a single classifier is selected from the pool. Thanks to its simplicity and good performance, the *DS-LA* has been frequently compared with new DCS approaches.

The third interesting approach, named KNORA (k -nearest-oracles), is a dynamic ensemble selection method that has shown very promising results by considering the neighborhood of the test pattern in a validation set as “oracles” which offer their advice about the most promising classifiers for a given unknown pattern [13]. In fact, the classifiers that recognize the k -nearest neighbors are selected according to four different strategies. Some interesting results were reported on different pattern recognition problems.

In this paper we describe an alternative strategy to measure the competence of the classifiers in the pool in order to select the most promising ones for a particular test instance. Based on the taxonomy proposed in [12], it may be considered as an oracle-based approach in which overall local accuracy (OLA), local class accuracy (LCA) and ambiguity among classifiers are combined to filter the pre-selection of classifiers done by the oracles. The main contribution is to show that complementary information based on OLA, LCA and ambiguity may improve the performance of a k -nearest oracle-based dynamic ensemble selection. Besides, by considering only the pre-selected classifiers based on the suggestion of the k -nearest oracles of the test instance, we can reduce the effort necessary to calculate the OLA and LCA information.

This paper is organized in 4 sections. Section II describes the proposed method and its variants, while Section III presents the experiments undertaken to evaluate the proposed method and the observed results. Finally, in Section IV, we present the conclusion and further work.

II. THE PROPOSED METHOD

An ensemble with high overall and local class accuracies has more chance to recognize a test instance. However, the computation of OLA and LCA is quite expensive since it needs to be done for all classifiers in the pool. With this in mind, we propose a two-step method to select the classifiers to compose the ensemble for a given test instance. This process, named DESCOLA (Dynamic Ensemble Selection using Class and Overall Local Accuracies) is illustrated in Figure 1.

In the first step, inspired in the oracle-based approach proposed in [13], we find the k -nearest neighbors of the test sample t in a special validation set representing a meta-space in which we know the classifiers that correctly recognize each sample. Afterwards, the OLA of each classifier suggested by the oracles is calculated over the same k -nearest neighbors of the test instance. At this point, it is important to notice that by considering only the classifiers suggested by the oracles, we avoid the calculation of the OLA information for all classifiers in the pool. According to the OLA value, which is computed as denoted in Equation 1, the classifiers are pre-selected to compose the ensemble.

$$OLA_{j,k}(t) = \frac{N_j}{k}, \quad (1)$$

where k is the number of neighbors of the test sample t , j is the identity of the classifier, and N_j is the number of neighbors that the classifier j recognizes.

Given the ensemble provided by the k -nearest oracles of the unknown pattern, the OLA value is used in a first filtering process. Two different strategies are considered, as follows:

- DESCOLA-Eliminate (DE): It is selected only the classifiers that have the best OLA value, as shown in the left side of the Figure 2a. The circle represents the validation instances. The hexagon with X is the test pattern. The painted circles are the neighbors. In this case, there are 5 neighbors ($k = 5$). Following the flow, we have the classifiers (ellipses C_i, C_k, C_m, C_n and C_p) that recognize at least one neighbor. After this, we have the selected classifiers (C_i, C_k and C_m), with $OLA = 1$. The other classifiers (C_n and C_p) recognize only one neighbor, and then their $OLA = \frac{1}{5}$. Then, in the right superior part of the Figure 2a, the filtered ensemble is presented. Different from the KNORA-Eliminate (KE), here the OLA information is used to filter the classifiers suggested by the k -nearest oracles of the test instance in the validation set.
- DESCOLA-Union (DU): It is selected all the classifiers that have OLA value greater than zero. In other words, it is selected the classifiers that recognize at least one neighbor, as shown in the left side of the Figure 2b. In that figure, the circle represents the validation instances. The hexagon with X is the test instance. The painted circles are the neighbors. In this case, there are 5 neighbors ($k = 5$). Following the arrow, we have the classifiers (ellipses C_i, C_k, C_m, C_n and C_p) that recognize at least one neighbor. The 5 classifiers are select to compose an ensemble, as

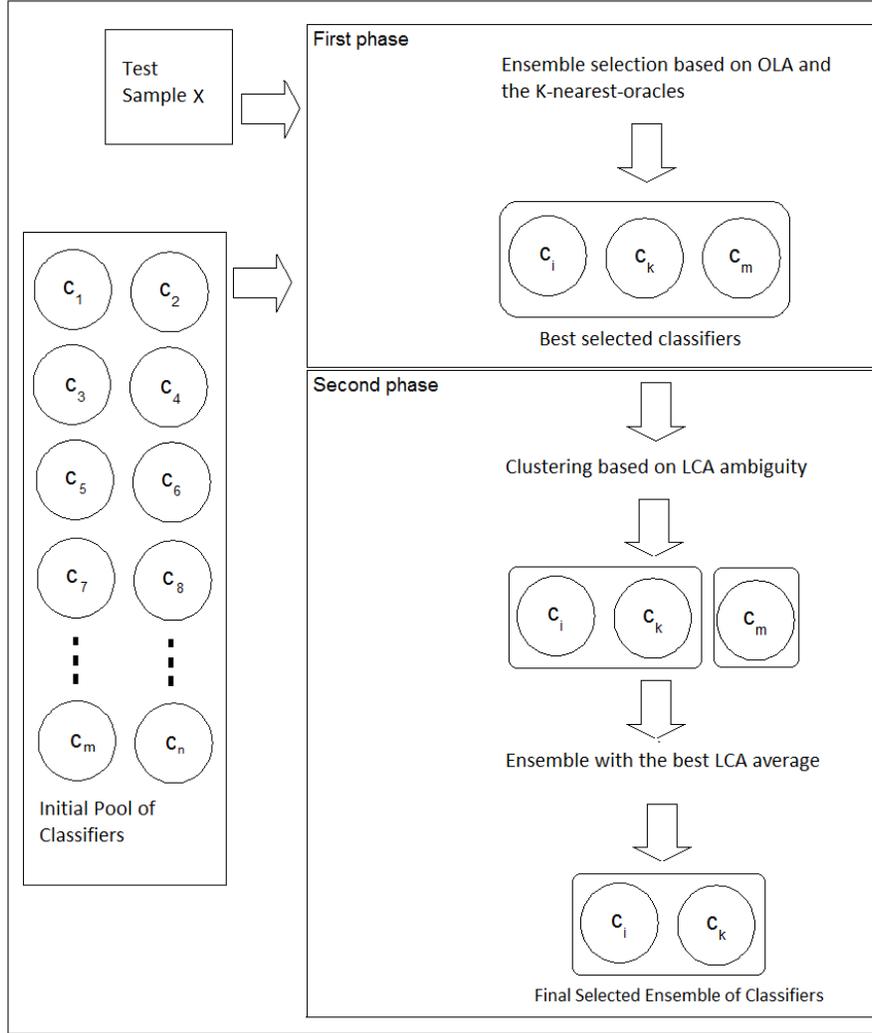


Fig. 1: General Overview of the Proposed Method

output in the first step. Different from the KNORA-Union (KU), here the classifiers with OLA greater than one are considered just once.

In the second step (“LCA” step), the pre-selected classifiers are firstly organized into groups by considering the ambiguity among them. For this purpose, all pre-selected classifiers assign a class to the test instance t . The ambiguity between the classifier C_j and the ensemble E_i with respect to t can be computed as described by the Equation 2.

$$\alpha_{ij}(t) = \begin{cases} 0 & \text{if } C_j(t) = \bar{E}_i(t) \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

where $C_j(t)$ and $\bar{E}_i(t)$ are the outputs of the classifier C_j and the ensemble E_i , respectively.

The ambiguity of the ensemble E_i with respect to the test instance t can be defined as the average of the ambiguity of its classifiers as denoted in Equation 3:

$$\bar{\alpha}_i(t) = \frac{\sum_{j=1}^{N_i} \alpha_{ij}(t)}{N_i}, \quad (3)$$

where N_i is the amount of classifiers in the ensemble E_i and $\alpha_{ij}(t)$ is the ambiguity value of classifier j .

The objective is to organize the pre-selected classifiers in ensembles where the ambiguity is zero. It means a full consensus among the classifiers inside each ensemble. In other words, all classifiers predict the same class to the test instance.

Afterwards, given the classifiers organized into ensembles by consensus, we need to select the most promising ensemble to provide the final decision for the test instance. At this point, we compute the LCA of each ensemble with respect of the class it has predicted.

As described in [4], let us suppose that the classifier C_j assigns the class w_p to the test pattern t , that is, $C_j(t) = w_p$. Then, LCA of C_j can be estimated as described by the Equation 4.

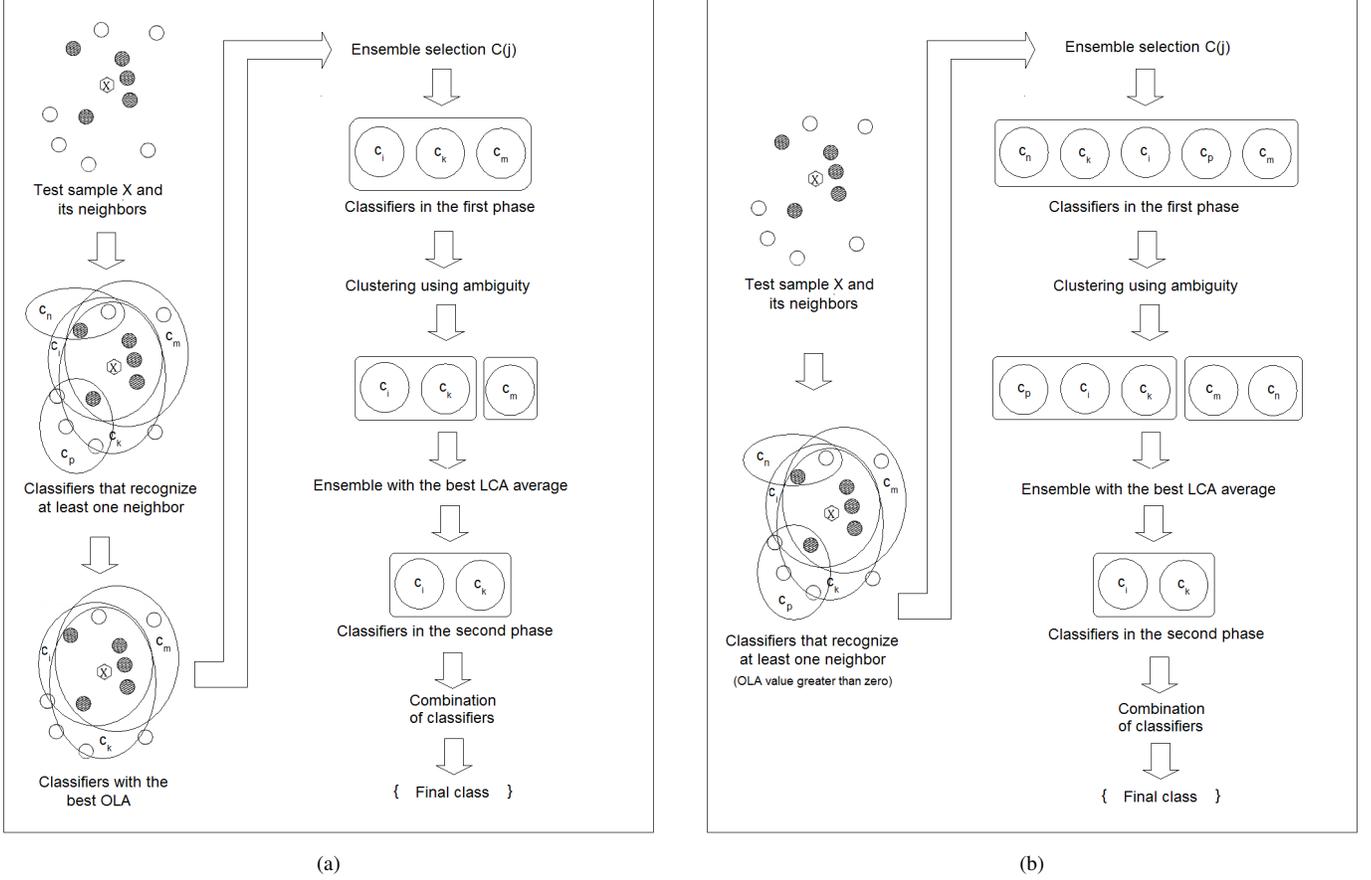


Fig. 2: Proposed Method: (a) DESCOLA-Eliminate (DE), (b) DESCOLA-Union (DU)

$$LCA_{j,v}(t) = \frac{N_{pp}}{\sum_{i=1}^M N_{ip}}, \quad (4)$$

where v is the number of samples in the neighborhood, N_{pp} is the number of neighbors of t that is correctly assigned by C_j to class w_p , and $\sum_{i=1}^M N_{ip}$ is the number of neighbors that is assigned by C_j to class w_p . As one may see the value of the neighborhood size, named here as v , can be different from that used in the first step of the proposed method. The possibility of using different neighborhood sizes for the first and second step allows that the pre-selected classifiers may be analyzed using a different neighborhood.

The LCA of each group of classifiers is defined as the average of the LCA of its elements, as denoted in Equation 5.

$$\overline{LCA}_v^g(t) = \frac{\sum_{i=1}^{N_g} LCA_{i,v}^g(t)}{N_g}, \quad (5)$$

where N_g is the amount of classifiers in the ensemble g , and $LCA_{i,v}^g(t)$ is the LCA value of the classifier i of the ensemble g , which is computed using a neighborhood with size equal to v .

The group that presents the highest LCA average value is selected as the ensemble to decide the final class of the

test instance. In case of a tie, or if the average of the LCA in all committee is null, it is selected the group with highest number of classifiers. In case of a tie in both aspects, a group is randomly selected. The DESCOLA method is described in the Algorithm 1.

Another interesting point is that the second step can be used independently, i.e., without the first step. This process was named DESLCA (Dynamic Ensemble Selection using Local Class Accuracy) and it uses all classifiers in the initial pool. First, for a test instance, it computes the class to each classifier. Then, it groups the classifiers according to the ambiguity and computes the LCA average of its elements. The ensemble with the highest LCA average value is selected to assign the final class to the test instance.

In this work, the final class is obtained through the majority voting of the classifiers that compose the selected ensemble. The next section presents the experimental results of the proposed method and its variants.

III. EXPERIMENTAL RESULTS

This section describes the experimental protocol, the used datasets, the learning technique used to generate the initial pool of classifiers, and, finally, the observed results.

Algorithm 1 DESCOLA method

Input the pool of classifiers P ; the validation set Va , the testing set Te ; the neighborhood size for the first step k ; the neighborhood size for the second step v ; Option = (ELIMINATE or UNION);
Output EoC_t^* ; an ensemble of classifiers for each unknown sample t in Te ;
Construct mVa ; a meta-space where each Va instance is linked to the classifiers of P that correctly recognize it;
for each testing sample t in Te **do**
 Find the k nearest neighbors of t in the mVa ;
 $EoC_{oracles}$ = the classifiers suggested by the oracles in mVa ;
 if (Option == ELIMINATE) **then**
 EoC_{ola} = the classifier(s) with the best OLA value in $EoC_{oracles}$;
 else
 EoC_{ola} = the classifier(s) with OLA value greater than zero in $EoC_{oracles}$;
 end if
 Organize the classifiers in EoC_{ola} into groups according to ambiguity using Equation 3;
 Find the v nearest neighbors of t in the meta-space mVa ;
 for each group g of classifiers organized from EoC_{ola} **do**
 Compute the average $\overline{LCA}_v^g(t)$ using Equations 4 and 5 ;
 end for
 Select the best ensemble for t as $EoC_t^* = \arg \max_g \{ \overline{LCA}_v^g(t) \}$;
 Use the ensemble EoC_t^* to classify t ;
end for

TABLE I: Database Characteristics

| Data set | # Classes | # Samples | # Feat. | RS Card. |
|----------|-----------|-----------|---------|----------|
| Letter | 26 | 20000 | 16 | 12 |
| LD | 2 | 345 | 6 | 4 |
| PD | 2 | 768 | 8 | 4 |
| IS | 7 | 2310 | 19 | 4 |
| Wbc | 2 | 569 | 30 | 5 |
| Wine | 3 | 178 | 13 | 6 |
| NIST | 10 | 75089 | 132 | 32 |
| LMD | 10 | 1300 | 110 | 56 |

A. Datasets and Pool of Weak Classifiers

The approach was evaluated on 6 pattern recognition problems (*Letter*, *Liver Disorder (LD)*, *Pima Diabets (PD)*, *Image Segmentation (IS)*, *Wincosin breast cancer (Wbc)*, *Wine*) taken from the UCI machine learning repository [25], and also on other two interesting applications, the recognition of handwritten digits available on the *NIST SD19* [26]; and the classification of music genres using the Latin Music Database (*LMD*) [21]).

The method used for creating the initial pool with diverse k -NN classifiers (where $k = 1$) is based on random subspaces [9]. The use of NN classifiers allows to create pools composed of weak classifiers in order to gain a better understanding of the impact of their combination and also the use of dynamic selection. Table I summarizes the main characteristics of each dataset (number of classes, samples, features, and the random subspace cardinality). All datasets were normalized, except the *NIST SD19*, for which the configurations used in [13] were preserved.

The experimental protocol is based on 10-folds cross-validation, where 50% of the 9 training folds were used for the validation set necessary in the proposed method. The pool of initial classifiers is composed of 10 elements. This protocol was used in almost all experiments, except for the

Image Segmentation (*IS* dataset), *NIST SD19* and *LMD*. For the *IS* dataset, we have used the training (210 samples) and the testing (2,100 samples) datasets suggested on the UCI data repository. For the *NIST SD19* and the *LMD* datasets, we have used the same protocol already used in [13] and [3], respectively. The experiments using the *NIST SD19* are based on three datasets: the training set with 5,000 samples (hsf_{0-3}), the validation set containing 10,000 samples (hsf_{0-3}) and the test set containing 60,089 samples (hsf_{7}). For *NIST* dataset, 100 classifiers were generated with feature subsets having only 32 features out of 132 (the same pool of weak classifiers proposed in [13]).

In [3], the audio signal was converted into spectrograms from which features based on their visual representation were extracted. The entire feature set extracted from each music has 885 features. From the original feature set, it was selected only 110 features by means the CFS (*Correlation-based Feature Selection*) method [8], with the default input parameters of the Weka tool. It provides 3-fold cross-validation already used in the experiments described in [3]: 600 samples for training, 400 for validation and 300 for testing. Ten classifiers were generated with feature subsets having only 56 features out of 110.

B. Results and Discussions

Table II contains the results of related works which are considered as benchmark for the proposed method and its variants. This table reports for each dataset the average recognition rate with standard deviation obtained by the oracle of the corresponding pool of classifiers, the combination of all classifiers in the pool using the majority voting rule (All), the KNORA-Eliminate (KE) and corresponding k value, the KNORA-Union (KU) and corresponding k value, the results of the two versions of the original DS-LA, OLA-based (OLA) and LCA-based (LCA).

Table III presents the results of the proposed method and

TABLE II: Performance of related works. Best average recognition rates (RR), the corresponding standard deviations (s), and the corresponding neighborhood size $[k]$. Pools are composed of 10 classifiers, except for the NIST dataset were 100 classifiers were generated

| Data | Oracle | All | KNORA-Eliminate (KE) | KNORA-Union (KU) | DS-LA (OLA-based) | DS-LA (LCA-based) |
|--------|--------------|--------------|----------------------|------------------|-------------------|-------------------------|
| Letter | 98.90 (0.19) | 95.18 (0.49) | 95.48 (0.34)[1] | 95.60 (0.39)[3] | 94.06 (0.73)[30] | 95.94 (0.46)[2] |
| LD | 97.69 (2.65) | 65.82 (3.29) | 65.49 (4.80)[10] | 68.12 (3.54)[7] | 65.86 (8.52)[11] | 66.40 (7.16)[2] |
| PD | 98.44 (1.19) | 73.44 (3.69) | 72.13 (3.74)[2] | 74.35 (3.61)[6] | 69.80 (6.66)[16] | 74.60 (4.47)[17] |
| IS | 99.52 (-) | 88.33 (-) | 88.33 (-)[1] | 88.33(-)[1] | 79.14 (-)[1] | 92.43 (-)[1] |
| Wbc | 99.82 (0.55) | 96.13 (1.84) | 96.48 (2.05)[1] | 96.48 (2.05)[1] | 93.50 (3.42)[4] | 96.13 (2.32)[14] |
| Wine | 100 (-) | 94.41 (4.54) | 97.75 (2.91)[2] | 96.67 (4.68)[2] | 95.52 (5.11) [12] | 97.19 (3.96)[22] |
| NIST | 99.95 (-) | 96.28 (-) | 97.52 (-)[7] | 97.25 (-)[1] | 94.11 (-)[30] | 97.55 (-)[3] |
| LMD | 90.22 (1.50) | 69.22 (1.39) | 68.89 (1.95)[2] | 70.00 (1.00)[28] | 66.33 (1.53) [28] | 70.22 (0.38)[5] |

TABLE III: Performance of the proposed method and its variants. Best average recognition rates (RR), the corresponding standard deviations (s), and the corresponding neighborhood sizes $[k, v]$ in case of DE and DU, or just $[k]$ when using DESLCA. Pools are composed of 10 classifiers, except for the NIST dataset were 100 classifiers were generated

| Data | DESCOLA-Eliminate (DE) | DESCOLA-Union (DU) | DESLCA |
|--------|---------------------------|-----------------------|------------------------|
| Letter | 96.16 (0.25)[2,2] | 95.98 (0.73)[1,10] | 95.98 (0.37)[1] |
| LD | 68.15 (6.41)[15,1] | 67.85 (6.80)[2,1] | 68.69 (7.70)[2] |
| PD | 73.70 (4.31)[4,1] | 74.73 (5.78)[4,2] | 73.82 (3.60)[29] |
| IS | 92.62 (-)[1,1] | 92.62 (-)[1,1] | 92.62 (-)[1] |
| Wbc | 97.01 (2.49)[2,2] | 96.13 (2.59)[1,1] | 96.13 (2.59)[1] |
| Wine | 98.30 (3.79) [2,2] | 97.75 (3.92)[2,1] | 97.19 (3.96)[2] |
| NIST | 97.67 (-)[4,6] | 97.46 (-)[10,3] | 97.46 (-)[9] |
| LMD | 70.00 (1.86) [2;2] | 69.78 (1.84)[3,4] | 69.67 (1.33)[2] |

its variants. The DESCOLA-Eliminate (DE), the DESCOLA-Union (DU) and the DESLCA method. For DE and DU, the corresponding k and v values are presented. The variable k in this paper is related to the size of neighborhood used to define the local region of the test instance in the validation set. Similarly, the variable v represents the number of neighbors to estimate the ‘‘LCA’’ in the second step of the proposed DESCOLA method. In addition, the last column in Table III presents the DESLCA variant, when it was applied directly in the pool of weak classifiers, aiming to select ensemble without consider the first step of the DESCOLA method. This means without the OLA information.

All parameters of the proposed method were defined using the validation dataset, and then the final performance was evaluated on the test set. The k and v values were considered from 1 to 30, and the settings that provide the best performance were used in the test dataset. It is worth noting that for some datasets the best size of the neighborhood defined on the validation set for the first step of the proposed method differs from that used for the second one. For instance, this is the case of the NIST dataset, for which the best results were achieved using $k = 4$ and $v = 6$ for the DE strategy, and $k = 10$ and $v = 3$ for the DU strategy. This can be explained by the use of different information to measure the competence of the classifiers at each step (OLA and LCA). In addition, the results have shown that the value of these parameters are problem dependent. The best result for each dataset is highlighted (bold face) in Tables II and III. These tables show that, in most cases, the DE approach has provided the best recognition rates.

As one may see, the performance of the proposed DE surpasses the combination of all classifiers in the pool (All) for all datasets. In addition, in 5 over 8 experiments, it surpasses

all the related dynamic selection methods (KE, KU and the DS-LA methods). We can observe that even the DU and the DESLCA methods also can be better than combining all classifiers using the majority voting rule, in (6 over 8) and (7 over 8) experiments, respectively.

To evaluate the significance of the reported results, the Friedman’s test was performed [11]. The p value obtained was 0.000177. It represents significant difference using a confidence level of 95%. Despite the oracle value was not reached, the DESCOLA and DESLCA allowed to increase the recognition rates on the most of the datasets.

IV. CONCLUSION

An alternative k -Nearest Oracle based method was proposed, in which the dynamic selection of the classifiers to compose the ensemble is done in two steps that combines OLA and LCA information. The proposed method was compared to the original KNORA and DS-LA methods and to the combination of all classifiers in the pool on 8 datasets. According to the experiments, the best recognition rates of DESCOLA variations, named DESCOLA-Eliminate and DESCOLA-Union, in most of cases, outperformed the KNORA and the DS-LA recognition rates. In addition, the DESCOLA also surpassed the results obtained by the majority vote fusion of the classifiers in the pool, with the advantage of having less classifiers used in the classification process. A variant of the proposed method presenting only the ‘‘LCA’’ step (DESLCA) has provided promising results. However, it is worth noting that in this case all classifiers in the pool are used to calculate the LCA value.

The experimental results have shown that the oracle performance was not reached, and further work is still necessary.

However, DESCOLA and DESLCA methods proved to be a good alternative when working with dynamic ensemble selection of classifiers. As future work, we plan to evaluate bagging and boosting as an alternative to create the initial pool of classifiers, and also the use of other weak classifiers instead of k -NNs.

ACKNOWLEDGMENT

The authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) and the Fundação Araucária that supports scientific and technological development in the State of Parana (Brazil).

REFERENCES

- [1] R. Ranawana, and V. Palade, "Multi-Classifer Systems: Review and a roadmap for developers," *Int. J. Hybrid Intell. Syst.*, 3, 1, 35–61, 2006.
- [2] P. Cavalin, R. Sabourin, and C. Suen, "Dynamic selection approaches for multiple classifier systems," *Neural Comp. and Appl.*, vol. 22, no. 3-4, pp. 673–688, 2013.
- [3] Y. Costa, L. Oliveira, A. Koerich, and F. Gouyon, "Music genre recognition using spectrograms," in *Int. Conf. on Systems, Signals and Image Proc.*, Sarajevo, Bosnia and Herzegovina, 2011.
- [4] L. Didaci, G. Giacinto, F. Roli, and G. L. Marcialis, "Rapid and brief communication: A study on the performances of dynamic classifier selection based on local accuracy estimation," *Pattern Recogn.*, vol. 38, no. 11, pp. 2188–2191, 2005.
- [5] E. M. dos Santos, R. Sabourin, and P. Maupin, "A dynamic overproduce-and-choose strategy for the selection of classifier ensembles," *Pattern Recogn.*, vol. 41, no. 10, pp. 2993–3009, 2008.
- [6] G. Giacinto and F. Roli, "Methods for dynamic classifier selection," in *Image Analysis and Processing. Proc., 10th Int. Conf. on*, 1999, pp. 659–664.
- [7] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behaviour," *Pattern Recogn.*, vol. 34, pp. 1879–1881, 2001.
- [8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.
- [9] T. K. Ho, "The random space method for constructing decision forests," *IEEE Trans. on Pattern Analysis and Machine Intell.*, vol. 20, no. 8, pp. 832–844, 1998.
- [10] T. Ho, J. Hull, and S. Srihari, "Decision combination in multiple classifier systems," *Pattern Analysis and Machine Intell., IEEE Trans. on*, vol. 16, no. 1, pp. 66–75, 1994.
- [11] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, 2nd ed. Wiley-Interscience, Jan. 1999.
- [12] A. S. Britto Jr., R. Sabourin, and L. E. Oliveira, "Dynamic selection of classifiers - a comprehensive review," *Pattern Recogn.*, vol. 47, no. 11, pp. 3665 – 3680, 2014.
- [13] A. Ko, R. Sabourin, and A. Britto Jr., "From dynamic classifier selection to dynamic ensemble selection," *Pattern Recogn.*, vol. 41, no. 5, pp. 1718–1731, 2008.
- [14] L. Kuncheva and J. Rodriguez, "Classifier ensembles with a random linear oracle," *Knowledge and Data Eng., IEEE Trans. on*, vol. 19, no. 4, pp. 500–508, 2007.
- [15] M. Kurzynski, T. Woloszynski, and R. Lysiak, "On two measures of classifier competence for dynamic ensemble selection - experimental comparative analysis," in *Communications and Inf. Technologies (ISCIT), Int. Symposium on*, 2010, pp. 1108–1113.
- [16] R. Lysiak, M. Kurzynski, and T. Woloszynski, "Probabilistic approach to the dynamic ensemble selection using measures of competence and diversity of base classifiers," in *Hybrid Art. Intell. Systems*, ser. Lecture Notes in Computer Science, E. Corchado, M. Kurzyński, and M. Woźniak, Eds. Springer Berlin Heidelberg, 2011, vol. 6679, pp. 229–236.
- [17] A. Nabiha and F. Nadir, "New dynamic ensemble of classifiers selection approach based on confusion matrix for arabic handwritten recognition," in *Multimedia Comp. and Systems (ICMCS), 2012 Int. Conf. on*, 2012, pp. 308–313.
- [18] M. Sabourin, A. Mitiche, D. Thomas, and G. Nagy, "Classifier combination for hand-printed digit recognition," in *Doc. Analysis and Recogn., 1993, Proc. of the Second Int. Conf. on*, 1993, pp. 163–166.
- [19] A. Santana, R. Soares, A. M. P. Canuto, and M. P. de Souto, "A dynamic classifier selection method to build ensembles using accuracy and diversity," in *Neural Networks, 2006. SBRN '06. Ninth Brazilian Symp. on*, 2006, pp. 36–41.
- [20] H. Shin and S. Sohn, "Combining both ensemble and dynamic classifier selection schemes for prediction of mobile internet subscribers," *Expert Systems with Appl.*, vol. 25, no. 1, pp. 63–68, 2003.
- [21] C. N. Silla, A. L. Koerich, and C. A. A. Kaestner, "The latin music database," in *9th Int. Conf. on Music Inf. Retrieval*, 2008, pp. 451–456.
- [22] K. Woods, W. Kegelmeyer Jr, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *Pattern Analysis and Machine Intell., IEEE Trans. on*, vol. 19, no. 4, pp. 405–410, 1997.
- [23] J. Xiao and C. He, "Dynamic classifier ensemble selection based on gmdh," in *Computational Sciences and Optimization, 2009. CSO 2009. Int. Joint Conf. on*, vol. 1, 2009, pp. 731–734.
- [24] M. O. R. Cruz, R. Sabourin, G. D. C. Cavalcanti, and I. R. Tsang, "On Meta-Learning for Dynamic ensemble selection," in *Proc. of the International Conference on Pattern Recognition, ICPR. Stockholm, 2014*, pp. 24–28.
- [25] Bache, K. and Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
- [26] P. J. Grother, "Nist special database 19 - handprinted forms and characters database," in *National Institute of Standards and Technology (NIST), Tech. Rep.*, 1995.
- [27] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *Proc. of the 12th European Conf. on Machine Learning, EMCL '01. London, 2001*, pp. 576–587.