GNN-DES: A New End-to-End Dynamic Ensemble Selection Method Based on Multi-label Graph Neural Network

Mariana de Araujo Souza[1](✉), Robert Sabourin[1],
George Darmiton da Cunha Cavalcanti[2], and Rafael Menelau Oliveira e Cruz[1]

[1] École de Technologie Supérieure, Université du Québec, Montréal QC, Canada
mariana.araujo.souza@gmail.com,
{robert.sabourin,rafael.menelau-cruz}@etsmtl.ca

[2] Centro de Informática, Universidade Federal de Pernambuco, Recife,
Pernambuco, Brazil
gdcc@cin.ufpe.br

Abstract. Most dynamic ensemble selection (DES) techniques rely solely on local information to single out the most competent classifiers. However, data sparsity and class overlap may hinder the region definition step, yielding an unreliable local context for performing the selection task. Thus, we propose in this work a DES technique that uses both the local information and classifiers' interactions to learn the ensemble combination rule. To that end, we encode the local information into a graph structure and the classifiers' information into multiple meta-labels, and learn the DES technique end-to-end using a multi-label graph neural network (GNN). Experimental results over 35 high-dimensional problems show the proposed method outperforms most evaluated DES techniques as well as the static baseline, suggesting its suitability for dealing with sparse overlapped data.

Keywords: Dynamic ensemble selection · Graph neural networks · Meta-learning · Data sparsity

1 Introduction

Dynamic ensemble selection (DES) techniques assume the classifiers in an ensemble make distinct mistakes in different areas of the feature space. Thus, they

The authors would like to thank the Canadian agencies FRQ (Fonds de Recherche du Québec) and NSERC (Natural Sciences and Engineering Research Council of Canada), and the Brazilian agencies CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-42795-4_6.

M. Vento et al. (Eds.): GbRPR 2023, LNCS 14121, pp. 59–69, 2023.
https://doi.org/10.1007/978-3-031-42795-4_6

attempt to choose a subset of the models according to their perceived competence for classifying each sample in particular, often resulting in superior performances compared to static selection schemes, which label all test instances with the same set of classifiers [5]. Most DES techniques rely on the locality assumption to solve the dynamic selection task, in the sense that similar samples should be correctly labeled by a similar set of classifiers. These techniques require delimiting a region called the Region of Competence (RoC), via clustering [22], nearest neighbors rule [2,12,23], distance-based potential function [29], recursive partitioning [25], and/or fuzzy hyperboxes [6], in which the classifiers competences are estimated according to some criteria, such as local accuracy [12], classifier behavior [2], ensemble diversity [22], and meta-learning [3], among others.

The local region can thus have a large impact on the performance of these techniques [5], and so several methods attempt to directly improve its distribution. Filtering out the samples from the RoC is done in [17,18,24] based on the Item Response Theory (IRT) discrimination index, class distribution, and instance characterization, respectively. The RoC is characterized in [14] using a must link and a cannot link graph that are then used together with the classifiers' local accuracy to estimate their competence in the region. These techniques attempt to characterize and improve the local distribution for the classifier estimation step but they still rely solely on the locality assumption to compute a handcrafted competence estimation rule over an already defined region. While these approaches work generally well over a vast array of problems, such as class imbalanced distributions [17,24], local methods are known to struggle over high dimensionality and class ambiguity [26,31] and can present a strong sensitivity to overlap and data sparsity [21], with the latter being often associated with an increased class boundary complexity [10,15]. Such challenging scenarios may affect the local region definition and weaken the locality assumption, which in turn may limit the application of the dynamic selection techniques over real-world problems that present these characteristics, such as medical imaging data [7] and DNA microarray data [15] used for disease detection.

We also find in the literature a few dynamic selection techniques that do not rely on the locality assumption to perform the dynamic selection task [16,19]. Instead, they define the task as a multi-label meta-problem and learn the selection rule based on the classifiers' inter-dependencies, thus the meta-learner yields the ensemble combination rule for each input query instance without defining the RoC or explicitly estimating the classifiers' competences. While this approach could be interesting over the scenarios where the local context does not favor the dynamic classifier selection task, these techniques completely disregard the local context and can perform poorly against simple local accuracy-based techniques [19]. Both techniques also present a high computational cost due to the use of a meta-learner ensemble [19] and Monte Carlo sampling [16].

Thus, we propose in this work a dynamic selection technique that learns from the instances' relationships and classifiers' interactions jointly to better deal with high dimensional overlapped data. To that end, we model the data into a graph structure that can represent the samples' local and class inter-relations. We also model the classifiers' interactions as the multi-labels of the dynamic selection meta-problem. We then train a multi-label graph neural network (GNN) to yield the dynamic classifier combination rule in an end-to-end manner, without resorting to handcrafted meta-features or explicit RoC definition.

Graph neural networks operate directly on graph-structured data and are able to produce high-level representations of nodes and graphs [30]. The first GNNs were proposed for transductive learning and were unable to yield embeddings for unseen nodes, such as the Graph Convolutional Network (GCN) [11] which first generalized the convolution operation to the vertex domain. However, several models have been since proposed that work in inductive scenarios. The GraphSAGE model [9], seen as an extension of the GCN for inductive learning, learns a set of functions that aggregate the features from sampled neighboring nodes to produce the node embeddings. The Graph Attention Network (GAT) [28], which also works for inductive problems, presents a self-attention mechanism that allows the assignment of different weights to the neighbors in order to increase the model's capacity and to naturally deal with graphs that present variable node degrees.

Thus, by using a multi-label GNN as our meta-classifier, we leverage both the classifiers' inter-dependencies, represented in the meta-labels, and the samples' local interactions, represented in the graph, so that internally the network may learn an embedded space where the locality assumption for the dynamic selection task is stronger. We then contribute to the dynamic ensemble selection research area by (a) proposing an end-to-end technique that combines the information from the local data and the classifiers' interactions to better deal with sparse and overlapped data, and (b) evaluating the proposed method and ten other techniques over 35 high dimensional small sample sized (HDSSS) problems to assess whether learning from the two sources of information help overcome the limitations the current dynamic selection techniques present.

This work is organized as follows. The proposed method is introduced in Sect. 2. The experiments are reported in Sect. 3. Lastly, we summarize our conclusions in Sect. 4.

2 Graph Neural Network Dynamic Ensemble Selection Technique

We propose in this work the Graph Neural Network Dynamic Ensemble Selection (GNN-DES) technique, which attempts to better deal with locally complex scenarios in sparse overlapped data by combining the information from the samples' local context and the classifiers' interactions. To that end, we model the former using a graph structure, which is capable of representing the samples' local and class relationships, and model the latter by learning the dynamic selection task as a multi-label meta-problem.

Figure 1 describes the general steps of the GNN-DES technique. In memorization, the training set $\mathcal{T}$ and the pool of classifiers C are used to assign the samples' meta-labels U and construct the known graph $G_{\mathcal{T}}$, which are both then used to train the multi-label meta-learner GNN. In generalization, the query

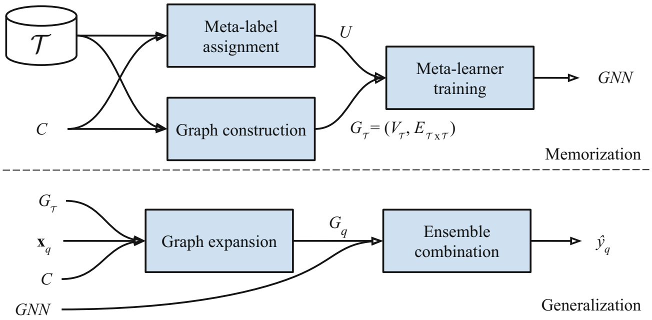

Fig. 1. Description of the GNN-DES technique. $\mathcal{T} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$ is the training set and $U = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N\}$ their corresponding meta-labels, $C = \{c_1, c_2, ..., c_{|C|}\}$ is the ensemble of classifiers. $G_\mathcal{T}$ is the known graph, composed of the set of training vertexes ($V_\mathcal{T}$) and edges ($E_{\mathcal{T}\times\mathcal{T}}$), G_q is the evaluation graph, composed of the $G_\mathcal{T}$ in addition to the query vertex (v_q) and its edges ($E_{q\times\mathcal{T}}$), and $\mathbf{x}_q$ and $\hat{y}_q$ are the query instance and its predicted label, respectively.

instance $\mathbf{x}_q$ is added to the known graph to produce the evaluation graph G_q, which is input to the meta-learner and used to produce the dynamic ensemble combination and then the output prediction $\hat{y}_q$.

Meta-label Assignment. In the meta-label assignment step, we characterize the competences of the classifiers by assigning to the samples meta-labels associated with their correct classification. This allows the meta-classifier to exploit the diverse behavior of the classifiers through learning the inter-dependecies between the meta-labels. Thus, to obtain the meta-labels, we evaluate the training set over the ensemble and we assign to each sample $(\mathbf{x}_i, y_i) \in \mathcal{T}$ a meta-label vector $\mathbf{u}_i$ of size $|C|$ so that $u_{i,k} = 1$ if the classifier c_k correctly labels $\mathbf{x}_i$, otherwise, $u_{i,k} = 0$.

Graph Construction. In the graph construction step, we aim to characterize in the known graph $G_\mathcal{T}$ the local context of the data that may be useful for the dynamic selection task. More specifically, we wish to embed the information of how reliable a sample is to indicate a good set of competent classifiers for another sample according to the locality assumption and the class relations. Thus, we link the samples that have a similar output response from the classifiers, as that may indicate they share a subset of competent classifiers. However, if the two samples belong to the same class we build a *strong* link, where the closer the samples the larger the edge weight as we expect the locality assumption to be stronger. Samples from different classes, on the other hand, are assigned a *weak* link, where the closer the samples the smaller the edge weight as the class ambiguity may indicate a weaker locality assumption.

Thus, to build the known graph $G_\mathcal{T}$, we project the training samples into the *decision* space, in which the axes represent the responses of each classifier in the pool. Then, we link each sample so that it has at least one strong link, to its nearest neighbor from the same class, and calculate its maximum margin for connection as a function of this link. Then, all samples within an instance's maximum margin are connected and their weights are set according to Eq. (1), where $d_{i,j}$ is the normalized *L1* distance between the samples $(\mathbf{x}_i, y_i), (\mathbf{x}_j, y_j) \in \mathcal{T}$ projected into the decision space, d_i^{max} is the maximum margin for connection of $(\mathbf{x}_i, y_i)$, and τ is a preset threshold.

$$\begin{aligned} e_{i,j} &= \begin{cases} 1 - d_{i,j}, \mathit{if}\ (d_{i,j} \leq d_i^{max} \vee d_{i,j} \leq d_j^{max}) \wedge y_i = y_j, \\ d_{i,j}^2, \mathit{if}\ (d_{i,j} \leq d_i^{max} \vee d_{i,j} \leq d_j^{max}) \wedge y_i \neq y_j, \\ 0, \mathit{otherwise}, \end{cases} \\ d_i^{max} &= \ \mathit{min}(d_{i,k}, \forall \mathbf{x}_k \in \mathcal{T} | y_k = y_i) + \tau \end{aligned} \tag{1}$$

Meta-learner Training. Using the known graph $G_{\mathcal{T}}$ and the meta-labels U we fit the meta-learner in a supervised manner in the final step of the proposed method in memorization. We use a graph neural network core to learn and produce the node embeddings and a dense layer of size $|C|$ with sigmoid activation as the output layer of the network so that each output node of the network represents a classifiers' weight in the dynamic ensemble combination rule. We use a GNN core that is capable of inductive learning, and we fit the model using the binary cross-entropy loss, weighted so that the harder to classify the sample, the higher its weight, so as to encourage the model to focus on the more difficult samples. We measure the instance hardness as the number of classifiers in the pool that can label it correctly.

Graph Expansion. In generalization, we first expand the known graph to include the query instance in the data structure as to provide the meta-learner with its local context to obtain its ensemble combination rule. Thus, we project the query $\mathbf{x}_q$ into the decision space using the ensemble C and connect it to its nearest neighbor. Based on that, its maximum margin for connection is calculated and the edge weights between the query and the instances that fall within the margin are calculated as shown in (2), where $d_{q,j}$ is the normalized *L1* distance between the samples $\mathbf{x}_q$ and $(\mathbf{x}_j, y_j) \in \mathcal{T}$ projected into the decision space, d_q^{max} is the query's maximum margin for connection, and τ is the preset threshold used in the graph construction step. The evaluation graph G_q is then built as the union between the known graph $G_{\mathcal{T}}$, the query vertex v_q, and the set of all its edges $E_q = \{e_{q,j}, \forall \mathbf{x}_j \in \mathcal{T}\}$.

$$\begin{aligned} e_{q,j} &= \begin{cases} 1 - d_{q,j}, & \mathit{if}\ d_{q,j} \leq d_q^{max}, \\ 0, & \mathit{otherwise}, \end{cases} \\ d_q^{max} &= \min(d_{q,k}, \forall \mathbf{x}_k \in \mathcal{T}) + \tau \end{aligned} \tag{2}$$

Ensemble Combination. We then induce the meta-learner GNN with the evaluation graph G_q to produce the network's outputs $\{o_{q,k}, \forall k \in |C|\}$, which represent the weighted support of each classifier when aggregating their responses. The class with the largest support is output as the query's predicted label $\hat{y}_q$.

3 Experiments

We evaluate in the experiments how well the DES techniques perform against a static selection baseline, and whether the proposed method is able to outperform them over the HDSSS problems. We describe the experimental protocol and present the results next.

Ensemble Methods. We use as our baseline an AdaBoost (ADA) [8] ensemble composed of 100 Decision Stumps, and we also evaluate 10 dynamic ensemble methods, namely: the K-Nearest Oracles Union (KNU) [12], the K-Nearest Oracles Eliminate (KNE) [12], the Dynamic Ensemble Selection-KNN (DKNN) [22], the K-Nearest Output Profiles (KNOP) [2], the META-DES [3], the Randomized Reference Classifier (RRC) [29], the Chained Dynamic Ensemble (CHADE) [19], the Online Local Pool technique (OLP) [23], the OLP++ [25] and the Forest of Local Trees (FLT) [1]. Except for CHADE, all of them are local-based techniques, though they may define the RoC using distinct methods or in different spaces, and the OLP, OLP++ and FLT are not evaluated using the AdaBoost ensemble as they produce their own pool. We also include the performance of the Oracle [13], an abstract model that always selects the correct classifier if it exists, to provide an upper limit to the performance of the DES techniques.

Hyperparameters. The techniques' hyperparameters were set as recommended in their papers if no implementation is available in the DESLib [4] library, or to their default value otherwise. The GNN-DES threshold was set to $\tau = 0.05$, and the meta-learner contained two GraphSAGE [9] layers of size 512 units, as in [20], and one dense output layer, as in [9]. We use the attentional aggregation function from [28] in the convolutional layers as the local samples may have distinct importances for the DES task. To cope with the small sample-sized problems, we sampled only 5 samples in each convolutional layer and applied L_2 regularization with $\lambda = 0.01$, which was empirically observed to help the training according to the validation loss curves. Moreover, 20% of the training set was used for validation/early stopping, and the validation nodes were connected to the known graph $G_{\mathcal{T}}$ as if unknown samples (2). The GNN was trained over 150 epochs, with a patience of 30 epochs, a batch size of 300, and the adaptive learning rate is initially set to 0.005, as in [28]. We also performed a sweep on the drop-out rate in the set $\{0.0, 0.2, 0.5\}$ as in [20], and the model with the best micro-averaged multi-label precision in validation was chosen.

Datasets and Evaluation. We use the datasets shown in Table 1, which are the same set of problems used in [25] with the exception of four datasets over which the ensemble method generated fewer than the set amount of classifiers in the pool. The testbed contains two-class HDSSS datasets (with at least 100 features) taken from the OpenML repository [27]. The columns N, F, and IR indicate the problems' number of instances, number of features and imbalance ratio, respectively, while the ratio F/N conveys the problems' sparsity and is associated with a higher data complexity [10,15].

We evaluate the datasets using a 10-fold cross-validation procedure using the folds available at the repository for reproducibility, and as in [25] we use the training set as the dynamic selection set (DSEL), a labeled dataset used for RoC definition [5], due to the limited number of instances in several datasets. Also due to the varying imbalance ratios in the testbed, we use the macro-averaged recall, or balanced accuracy rate, as the performance measure, to account for the class disproportion without focusing on one of the classes.

Table 1. Characteristics of the datasets used in the experiments.

Dataset	N	F	IR	F/N	Dataset	N	F	IR	F/N
tumors_C	60	7129	1.86	118.82	OVA_Endometrium	1545	10935	24.33	7.08
leukemia	72	7129	1.88	99.01	OVA_Uterus	1545	10935	11.46	7.08
AP_Endometrium_Lung	187	10935	2.07	58.48	OVA_Ovary	1545	10935	6.80	7.08
AP_Omentum_Uterus	201	10935	1.61	54.40	OVA_Breast	1545	10935	3.49	7.08
AP_Omentum_Lung	203	10935	1.64	53.87	fri_c4_100_100	100	100	1.13	1.00
AP_Lung_Uterus	250	10935	1.02	43.74	tecator	240	124	1.35	0.52
AP_Omentum_Ovary	275	10935	2.57	39.76	fri_c4_250_100	250	100	1.27	0.40
AP_Ovary_Uterus	322	10935	1.60	33.96	gina_agnostic	3468	970	1.03	0.28
AP_Omentum_Kidney	337	10935	3.38	32.45	gina_prior	3468	784	1.03	0.23
AP_Colon_Prostate	355	10935	4.14	30.80	fri_c4_500_100	500	100	1.30	0.20
AP_Colon_Omentum	363	10935	3.71	30.12	spectrometer	531	101	8.65	0.19
AP_Uterus_Kidney	384	10935	2.10	28.48	scene	2407	299	4.58	0.12
AP_Endometrium_Breast	405	10935	5.64	27.00	mfeat-pixel	2000	240	9.00	0.12
AP_Breast_Prostate	413	10935	4.99	26.48	mfeat-factors	2000	216	9.00	0.11
AP_Breast_Omentum	421	10935	4.47	25.97	fri_c4_1000_100	1000	100	1.29	0.10
AP_Colon_Ovary	484	10935	1.44	22.59	yeast_ml8	2417	116	70.09	0.05
AP_Colon_Kidney	546	10935	1.10	20.03	sylva_prior	14395	108	15.25	0.01
AP_Breast_Kidney	604	10935	1.32	18.10					

Results. Table 2 summarizes the performances of the Oracle, the static selection baseline (ADA) and the other 10 dynamic ensemble methods besides the proposed GNN-DES. The average performances per dataset are available in the supplementary material. We can see that the GNN-DES yielded the highest average balanced accuracy rate and the highest average rank among all techniques. Moreover, the GNN-DES obtained a higher average performance over at least half of the datasets compared to all techniques except for the META-DES, another local-based meta-learning technique.

Performing the non-parametric Wilcoxon signed-rank test over the pairs of techniques, we obtain the p-values shown in Table 3. First, we observe that the GNN-DES statistically outperformed with significance $\alpha = 0.05$ all evaluated techniques except the KNOP and the META-DES. As these three best-performing and statistically similar techniques are the only ones to rely on the local information in the decision space (in addition to the feature space, in the case of the META-DES), the results suggest that this approach may be better indicated for dynamic classifier selection on HDSSS problems.

We can also observe in Table 3 that the GNN-DES was the only dynamic ensemble method to statistically outperform the static selection baseline (ADA) with $\alpha = 0.05$ over the HDSSS datasets. This suggests that not only do the DES techniques generally struggle over these sparse datasets, as could be reasonably expected, but also that GNN-DES might behave somewhat differently from the classical local-based approaches, possibly due to the inclusion of the other source of information relative to the classifiers' interactions. However, how exactly the learned embedded space may affect the behavior of the GNN-DES and in which situations this information is valuable are questions to be analyzed in the future. All in all, we believe these promising results over the HDSSS problems warrant further investigation into the proposed approach.

Table 2. Mean balanced accuracy rate and rank, averaged over all datasets. The *Win-tie-loss* row refers to the number of datasets the GNN-DES obtained a higher, equal, or lower average performance to the column-wise technique.

	Oracle	ADA	KNU	KNE	DKNN	KNOP	META-DES	RRC	CHADE	FLT	OLP	OLP++	GNN-DES
Mean	99.97	88.14	87.12	84.08	84.73	88.88	88.90	85.30	83.93	84.88	81.35	86.22	**89.03**
Mean rank	n/a	5.31	5.36	8.39	8.01	4.90	4.29	6.33	7.69	5.90	10.59	7.11	**4.13**
Win-tie-loss	n/a	23-0-12	21-3-11	31-0-4	32-0-3	18-1-16	15-0-20	21-3-11	28-0-7	22-1-12	33-0-2	27-1-7	n/a

Table 3. Resulting p-values of the Wilcoxon signed-rank test between average balanced accuracy rates of all pairs of techniques, rounded to the second decimal point. Values below $\alpha = 0.05$ are in bold, rounded values below 0.01 are underlined, and the symbols ± indicate whether the column-wise technique statistically outperformed or not the row-wise technique.

	ADA	KNU	KNE	DKNN	KNOP	META-DES	RRC	CHADE	FLT	OLP	OLP++	GNN-DES
ADA	n/a	0.38	**0.01**(−)	**0.01**(−)	0.10	0.10	0.22	**0.01**(−)	0.51	**0.01**(−)	**0.01**(−)	**0.04**(+)
KNU		n/a	**0.01**(−)	**0.01**(−)	0.11	0.08	0.06	**0.01**(−)	0.86	**0.01**(−)	0.37	**0.03**(+)
KNE			n/a	0.26	**0.01**(+)	**0.01**(+)	0.24	0.66	0.15	**0.01**(−)	0.15	**0.01**(+)
DKNN				n/a	**0.01**(+)	**0.01**(+)	0.42	0.95	0.30	**0.01**(−)	0.21	**0.01**(+)
KNOP					n/a	0.12	0.06	**0.01**(−)	0.65	**0.01**(−)	**0.01**(−)	0.62
META-DES						n/a	**0.02**(−)	**0.01**(−)	0.27	**0.01**(−)	**0.01**(−)	0.59
RRC							n/a	0.12	0.67	**0.01**(−)	0.85	**0.01**(+)
CHADE								n/a	0.35	0.09	0.25	**0.01**(+)
FLT									n/a	**0.01**(−)	0.47	0.13
OLP										n/a	**0.01**(+)	**0.01**(+)
OLP++											n/a	**0.01**(+)
GNN-DES												n/a

4 Conclusion

We proposed in this work the GNN-DES technique, which learns the dynamic classifier combination rule from the instances' relationships and classifiers' interactions to deal with sparse overlapped data. We encode the local and class relations between the samples into a graph structure and the ensemble competence information into multiple meta-labels, and then fit our meta-learner, a multi-label GNN model, to perform the DES task in an end-to-end manner.

Experiments over 35 HDSSS datasets showed that the DES techniques in the literature had difficulty in surpassing the static selection baseline, especially the techniques based solely on similarities in the feature space for RoC definition. The locality assumption in the decision space was shown to perform better over the sparse data, and the three techniques that use this approach performed similarly and the best. Moreover, the GNN-DES was the only technique to statistically outperform the baseline in addition to 8 of the 10 evaluated DES techniques, suggesting its suitability for dealing with sparse and overlapped data.

Future work in this line of research may involve evaluating the impact of using different ensemble methods and hyperparameters to analyze the relationship between the graph characteristics and the technique's performance. Furthermore, we may analyze the behavior of the technique in different local contexts and its relation to the learned embedded space to investigate in which scenarios the meta-learner improves the locality assumption for the DES task.

References

1. Armano, G., Tamponi, E.: Building forests of local trees. Pattern Recogn. **76**, 380–390 (2018)
2. Cavalin, P.R., Sabourin, R., Suen, C.Y.: LoGID: an adaptive framework combining local and global incremental learning for dynamic selection of ensembles of HMMs. Pattern Recogn. **45**(9), 3544–3556 (2012)
3. Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C., Ren, T.I.: META-DES: a dynamic ensemble selection framework using meta-learning. Pattern Recogn. **48**(5), 1925–1935 (2015)
4. Cruz, R.M.O., Hafemann, L.G., Sabourin, R., Cavalcanti, G.D.C.: DESlib: a dynamic ensemble selection library in python. J. Mach. Learn. Res. **21**(8), 1–5 (2020)
5. Cruz, R.M., Sabourin, R., Cavalcanti, G.D.: Dynamic classifier selection: recent advances and perspectives. Inf. Fusion **41**, 195–216 (2018)
6. Davtalab, R., Cruz, R.M., Sabourin, R.: Dynamic ensemble selection using fuzzy hyperboxes. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–9 (2022)
7. El-Sappagh, S., et al.: Alzheimer's disease progression detection model based on an early fusion of cost-effective multimodal data. Futur. Gener. Comput. Syst. **115**, 680–699 (2021)
8. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)

9. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, pp. 1024–1034 (2017)
10. Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. IEEE Trans. Pattern Anal. Mach. Intell. **24**(3), 289–300 (2002)
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
12. Ko, A.H.-R., Sabourin, R., de Souza Britto Jr., A.: A new dynamic ensemble selection method for numeral recognition. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 431–439. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72523-7_43
13. Kuncheva, L.I.: A theoretical study on six classifier fusion strategies. IEEE Trans. Pattern Anal. Mach. Intell. **24**(2), 281–286 (2002)
14. Li, D., Wen, G., Li, X., Cai, X.: Graph-based dynamic ensemble pruning for facial expression recognition. Appl. Intell. **49**(9), 3188–3206 (2019)
15. Lorena, A.C., Costa, I.G., Spolaôr, N., De Souto, M.C.: Analysis of complexity indices for classification problems: cancer gene expression data. Neurocomputing **75**(1), 33–42 (2012)
16. Narassiguin, A., Elghazel, H., Aussem, A.: Dynamic ensemble selection with probabilistic classifier chains. In: Ceci, M., Hollmén, J., Todorovski, L., Vens, C., Džeroski, S. (eds.) ECML PKDD 2017. LNCS (LNAI), vol. 10534, pp. 169–186. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71249-9_11
17. Oliveira, D.V., Cavalcanti, G.D., Porpino, T.N., Cruz, R.M., Sabourin, R.: K-nearest oracles borderline dynamic classifier ensemble selection. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2018)
18. Pereira, M., Britto, A., Oliveira, L., Sabourin, R.: Dynamic ensemble selection by K-nearest local Oracles with discrimination index. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence, pp. 765–771. IEEE (2018)
19. Pinto, F., Soares, C., Mendes-Moreira, J.: CHADE: metalearning with classifier chains for dynamic combination of classifiers. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken, J. (eds.) ECML PKDD 2016. LNCS (LNAI), vol. 9851, pp. 410–425. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46128-1_26
20. Salehi, A., Davulcu, H.: Graph attention auto-encoders. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence, pp. 989–996 (2020)
21. Sánchez, J.S., Mollineda, R.A., Sotoca, J.M.: An analysis of how training data complexity affects the nearest neighbor classifiers. Pattern Anal. Appl. **10**(3), 189–201 (2007)
22. Soares, R.G., Santana, A., Canuto, A.M., de Souto, M.C.P.: Using accuracy and diversity to select classifiers to build ensembles. In: The 2006 IEEE International Joint Conference on Neural Network (IJCNN) Proceedings, pp. 1310–1316 (2006)
23. Souza, M.A., Cavalcanti, G.D., Cruz, R.M., Sabourin, R.: Online local pool generation for dynamic classifier selection. Pattern Recogn. **85**, 132–148 (2019)
24. Souza, M.A., Sabourin, R., Cavalcanti, G.D.C., Cruz, R.M.O.: Local overlap reduction procedure for dynamic ensemble selection. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–9 (2022)
25. Souza, M.A., Sabourin, R., Cavalcanti, G.D., Cruz, R.M.: OLP++: an online local classifier for high dimensional data. Inf. Fusion **90**, 120–137 (2023)
26. Vandaele, R., Kang, B., De Bie, T., Saeys, Y.: The curse revisited: when are distances informative for the ground truth in noisy high-dimensional data? In: International Conference on Artificial Intelligence and Statistics, pp. 2158–2172. PMLR (2022)

27. Vanschoren, J., van Rijn, J.N., Bischl, B., Torgo, L.: OpenML: networked science in machine learning. SIGKDD Explor. **15**(2), 49–60 (2013)
28. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018)
29. Woloszynski, T., Kurzynski, M.: A probabilistic model of classifier competence for dynamic ensemble selection. Pattern Recogn. **44**(10), 2656–2668 (2011)
30. Xia, F., et al.: Graph learning: a survey. IEEE Trans. Artif. Intell. **2**(2), 109–127 (2021)
31. Zhang, S.: Challenges in KNN classification. IEEE Trans. Knowl. Data Eng. **34**(10), 4663–4675 (2022)