




An Ensemble System with Random Projection and Dynamic Ensemble Selection

Manh Truong Dang¹, Anh Vu Luong², Tuyet-Trinh Vu¹,
Quoc Viet Hung Nguyen³, Tien Thanh Nguyen^{2,3} ,
and Bela Stantic³

¹ School of Information and Communication Technology,
Hanoi University of Science and Technology, Hanoi, Vietnam

² School of Applied Mathematics and Informatics,

Hanoi University of Science and Technology, Hanoi, Vietnam

³ School of Information and Communication Technology, Griffith University,
Gold Coast, Australia

thanh.nguyen3@griffithuni.edu.au

Abstract. In this paper, we propose using dynamic ensemble selection (DES) method on ensemble generated based on random projection. We first construct the homogeneous ensemble in which a set of base classifier is obtained by a learning algorithm on different training schemes generated by projecting the original training set to lower dimensional down spaces. We then develop a DES method on those base classifiers so that a subset of base classifiers is selected to predict label for each test sample. Here competence of a classifier is evaluated based on its prediction results on the test sample's k -nearest neighbors obtaining from the projected data of validation set. Our proposed method, therefore, gains the benefits not only from the random projection in dimensionality reduction and diverse training schemes generation but also from DES method in choosing an appropriate subset of base classifiers for each test sample. The experiments conducted on some datasets selected from four different sources indicate that our framework is better than many state-of-the-art DES methods concerning to classification accuracy.

Keywords: Ensemble method · Random projection
Multiple classifiers system · Dynamic ensemble selection

1 Introduction

In designing of an ensemble, there are three phases to be considered namely generation, selection, and combination. In the first phase, the learning algorithm(s) learn on the training set(s) to obtain base classifiers. In the second phase, a single classifier of a subset of the best classifier is selected. In the last phase, the decisions made by classifiers of the ensemble are combined to obtain the final one [1].

Homogeneous ensemble methods like Bagging [2] and Random Subspace [3] focus on the generation phase in which these methods concentrate on generating new training schemes from the original training set. In 1984, Johnson and Lindenstrauss (JL)

introduced an extending of Lipschitz continuous maps from metric spaces to Euclidean spaces as well as the JL Lemma [4]. The lemma begins with a linear transformation (known as a random projection) from a p -dimensional space \mathbb{R}^p (called up space) to a q -dimensional space \mathbb{R}^q (called down space). Due to the unstable property, random projections have used to construct the homogeneous ensemble [5].

In this study, we first employ random projection to generate the homogeneous ensemble system to solve the classification tasks. In detail, the original training set is projected to many down spaces to generate new training schemes. Due to the unstable property of random projection in which the generated training scheme is different to original training set as well as the other schemes, a learning algorithm can learn on these schemes to obtain the diverse base classifiers. We then consider the selection phase by selecting a subset of classifiers (also called ensemble of classifiers or EoC) associated with some random projections to predict class label. Here we propose using a DES method [1, 6] to the random projection-based ensemble in which instead of using all base classifiers for the prediction, only a subset of them is selected to predict the class label for a specific test sample. The selection is based on the neighborhood of the test sample belonging to the validation set in the local region of the projected feature space. The merits of our work lie in the following: to the best of our knowledge, this is the first approach to dynamically select EoC associated with random projections to predict class label for each sample.

The paper is organized as follows. In Sect. 2, random projections and dynamic classifier/ensemble selection are introduced. In Sect. 3, the proposed method based on the combination of random projection and DES is proposed. Experimental results are presented in Sect. 4 in which the results of the proposed method are compared with those produced by some benchmark algorithms on 15 selected datasets. Finally, the conclusions are presented in Sect. 5.

2 Related Methods

2.1 Random Projection

Given a finite set of p -dimension data $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$, we consider a linear transformation $T: \mathbb{R}^p \rightarrow \mathbb{R}^q: \mathbf{Z} = T[\mathcal{D}] = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\} \subset \mathbb{R}^q$ and $\mathbf{z}_i = T(\mathbf{x}_i)$. If the linear transformation T can be represented in the form of matrix \mathbf{R} ($\mathbf{z}_i = T(\mathbf{x}_i) = \mathbf{R}\mathbf{x}_i$) so that if each element of the matrix is generated according to a specified random distribution, T is known as a random projection. In practice, the random projection is simply obtained by using a random matrix $\mathbf{R} = 1/\sqrt{q}\{r_{ij}\}$ of size $(p \times q)$, where r_{ij} are random variables such that $E(r_{ij}) = 0$ and $\text{Var}(r_{ij}) = 1$. Several forms of \mathbf{R} are summarized in [7] in which Plus-minus-one and Gaussian are the most popular random projections.

Random projections are useful in dimension reduction since the dimension of the down space can be chosen to be lower than that of up space, i.e., $q < p$. Comparing to Principle Component Analysis (PCA), the directions of random projection are independent of the data while those of PCA are data-dependent and generating the principle components is computationally expensive compare to generating the random matrix in

random projection [8]. Furthermore, Fern and Brodley [5] indicated that random projections are very unstable since the dataset schemes generated from an original data source based on random matrices are quite different. This property is important since other sampling methods like bootstrapping only generate slightly different dataset schemes. Thus an ensemble system based on a set of random projections offers a potential for increased diversity. Until now, random projection has been extensively studied and applied to many areas, for example dimensionality reduction in analyzing noisy and noiseless images, and information retrieval in text documents [8], sparse random projection to approximate the \mathcal{X}^2 kernel [9], in supervised online machine learning [7, 10], and in analyzing clusters [11].

2.2 Dynamic Ensemble/Classifier Selection

In the selection phase of multiple classifier systems, a single classifier or an EoC can be obtained via static or dynamic approach. While in static approach, the selection is conducted during the training process and then the selected classifier or EoC is used to predict the label of all test sample, the dynamic approach works on the classification process by selecting a different classifier or different EoC for each test sample. We distinguish dynamic classifier selection (DCS) and DES term in which DCS techniques select only one classifier while DES techniques select an EoC for each test sample. Recent research on dynamic selection approaches shows its advantages for classification problems [12].

In dynamic selection approach, we first need to evaluate the competence of each base classifier from the pool of classifier and then select only the most competent or ensemble containing the most competent classifiers to classify each specific test sample. Here the competence is computed according to some criteria on the samples in the local region of feature space which can be defined by k -nearest neighbor techniques (in MCB [13], MLA [14], KNOP [15], META-DES [6], KNORA-Union [16], DES-FA [17]), clustering techniques [18], and potential functions (in DES-RRC [19], DES-KL [20], DES-P [20]). The selection criteria includes the accuracy of base classifiers in the local region [16], or meta-learning [6], or probabilistic-based models by considering posterior probability of the classifier on the neighbors of each test sample [19, 20].

3 Proposed Method

In this paper, we propose an ensemble system for label prediction using KNORA Union [1, 16] method and random projections. The survey in [1] shows that simple DES method like KNORA Union is competitive to many more complex methods. Meanwhile, the random project is advantageous in the homogeneous ensemble generation. In detail, in the training process, K random matrices of size $(p \times q)$ denoted by \mathbf{R}_j ($j = 1, \dots, K$) are generated. The new K training schemes \mathbf{Z}_j of size $(N \times q)$ (N is the number of training observations) and then are generated from the original training set \mathcal{D} of size $(N \times p)$ though the projection $\mathcal{D} \xrightarrow{\mathbf{R}_j} \mathbf{Z}_j$ given by:

$$\mathbf{Z}_j = (\mathbf{DR}_j) / \sqrt{q} \tag{1}$$

The ensemble of classifiers BC_j ($j = 1, \dots, K$) is constructed by a learning algorithm \mathcal{K} on training schemes \mathbf{Z}_j . As random projection often generates significantly different training schemes from original training set [5, 7, 10], the system diversity is ensured. In DES, each test sample is predicted by selected EoC; and the EoCs for two different test samples may be different. In general, we define the credit of a base classifier on a sample.

Definition 1: The credit of a base classifier BC_j on a sample \mathbf{x} denoted by $w_j(\mathbf{x})$ is the number of times the prediction of BC_j on \mathbf{x} used in the combining algorithm.

We propose using KNORA Union to our system to find the credit of base classifiers on each test sample. The idea of the method is based on the prediction results of base classifiers on the neighbors of each test sample. In this study, instead of getting the neighbors from the validation set, we consider the neighbors in the projected schemes of validation set. Specifically, the validation set \mathcal{V} is projected to the down spaces as:

$$\mathbf{V}_j = (\mathcal{V}\mathbf{R}_j) / \sqrt{q} \tag{2}$$

Denote $kNN_j(\mathbf{x}^u)$ as the k -nearest neighbors of an test sample \mathbf{x}^u in \mathbf{V}_j . We select base classifiers for \mathbf{x}^u based on their prediction results on $kNN_j(\mathbf{x}_u)$ as if BC_j gives a corrected prediction on each observation in $kNN_j(\mathbf{x}^u)$, its credit $w_j(\mathbf{x}^u)$ will increase by 1. Based on prediction results on all observations belonging to $kNN_j(\mathbf{x}^u)$, we obtain all $w_j(\mathbf{x}^u)$ $j = 1, \dots, K$. It is noted that there is an exception in which all base classifiers misclassify all observations in $kNN_j(\mathbf{x}^u)$ so that base classifiers contribute nothing to the combination. In this case, we simply set $w_j(\mathbf{x}^u) = 1 \forall j = 1, \dots, K$ which means that all base classifiers contribute equally to the prediction for \mathbf{x}^u .

The output of the base classifiers on \mathbf{x}^u are combined to obtain the predicted class label. Let $\{y_m\}_{m=1, \dots, M}$ denotes the set of M labels, $P_j(y_m|\mathbf{x}^u)$ is the probability that \mathbf{x}^u belongs to the class with label y_m given by the BC_j . There are two output types for \mathbf{x}^u namely *Crisp Label* (returns only class label, i.e. $P_j(y_m|\mathbf{x}^u) \in \{0, 1\}$ and $\sum_m P_j(y_m|\mathbf{x}^u) = 1$) and *Soft Label* (returns posterior probabilities that \mathbf{x}^u belongs to a class, i.e. $P_j(y_m|\mathbf{x}^u) \in [0, 1]$ and $\sum_m P_j(y_m|\mathbf{x}^u) = 1$) [21–24]. In this paper, we propose using fixed combining rules [25–27] to combine the output of base classifiers. As the fixed combining rules apply directly to the output of base classifiers to give the prediction, they are simpler and fast to build and run. Several popular fixed combining methods are Sum, Product, Majority Vote, Max, Min, and Median [25, 26]. In this study because base classifiers set different credits on each test sample, the forms of fixed combining rules applied to the outputs of base classifiers are given by:

$$\text{Sum Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \sum_{j=1}^K w_j(\mathbf{x}^u) P_j(y_m | \mathbf{x}^u) \quad (3)$$

$$\text{Product Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \prod_{k=1}^K w_j(\mathbf{x}^u) P_j(y_m | \mathbf{x}^u) \quad (4)$$

$$\text{Max Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \max_{j=1, \dots, K} w_j(\mathbf{x}^u) P_j(y_m | \mathbf{x}^u) \quad (5)$$

$$\text{Min Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \min_{j=1, \dots, K} w_j(\mathbf{x}^u) P_j(y_m | \mathbf{x}^u) \quad (6)$$

$$\text{Median Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \text{median}_{j=1, \dots, K} w_j(\mathbf{x}^u) P_j(y_m | \mathbf{x}^u) \quad (7)$$

$$\text{Majority Vote Rule: } \mathbf{x}^u \in y_t \text{ if } t = \arg \max_{m=1, \dots, M} \sum_{j=1}^K w_j(\mathbf{x}^u) \Delta_{jm} \quad (8)$$

$$\Delta_{jt} = \begin{cases} 1 & \text{if } t = \arg \max_{m=1, \dots, M} P_j(y_m | \mathbf{x}^u) \\ 0 & \text{otherwise} \end{cases}$$

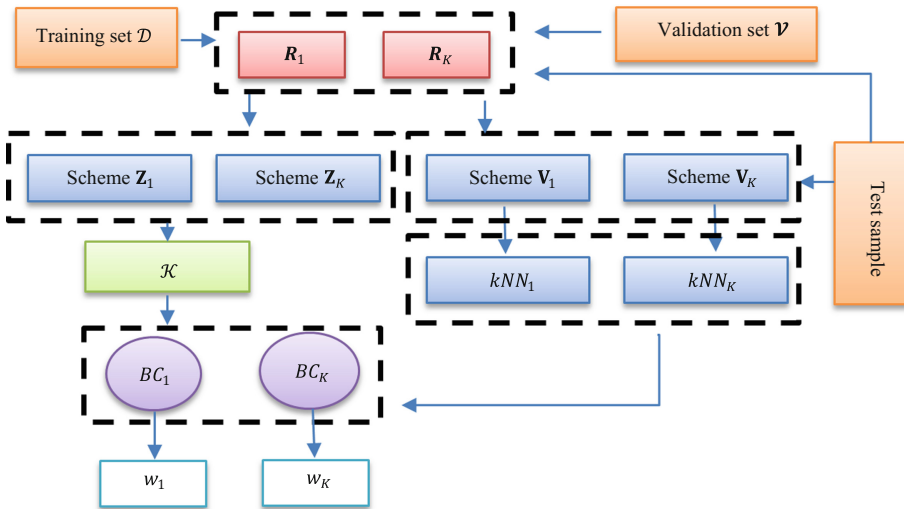


Fig. 1. The process to find the credits of base classifiers

Algorithm: DES with Random Projections-based ensemble	
<hr/>	
Training Process	
<hr/>	
Input:	Training set: \mathcal{D} , Validation set: \mathcal{V} , Dimension of down space: q , number of projections: K , learning algorithm: \mathcal{K}
Output:	Base classifiers: BC_j , Validation schemes \mathbf{V}_j , and random matrices: \mathbf{R}_j , ($j = 1, \dots, K$)
<hr/>	
	For $j=1$ to K
	Generate random matrix \mathbf{R}_j ;
	Get schemes \mathbf{Z}_j by (1)
	$BC_j = \text{Learn}(\mathcal{K}, \mathbf{Z}_j)$;
	Get validation scheme \mathbf{V}_j by (2)
	End
	Return $\{BC_j\}$, $\{\mathbf{V}_j\}$, and $\{\mathbf{R}_j\}$ ($j = 1, \dots, K$)
<hr/>	
Classification process	
<hr/>	
Input	Test sample \mathbf{x}^u , $\{BC_j\}$, $\{\mathbf{R}_j\}$, and $\{\mathbf{V}_j\}$ ($j = 1, \dots, K$)
Output	Predicted class label for \mathbf{x}^u
<hr/>	
	For $j=1$ to K
	$\mathbf{z}_j^u = \mathbf{x}^u \mathbf{R}_j / \sqrt{q}$; $w_j(\mathbf{x}) = 0$;
	End For
<hr/>	
	For $j=1$ to K
	Find $kNN_j(\mathbf{x}^u)$ in \mathbf{V}_j
	For each v in kNN_j
	For $j=1$ to K
	Predict label \hat{y} for v by BC_j ;
	If ($\hat{y} = y_v$)
	$w_j(\mathbf{x}^u) = w_j(\mathbf{x}^u) + 1$;
	End
	End
	End
	End
	If ($w_j(\mathbf{x}) = 0 \ \forall j = 1, \dots, K$)
	$w_j(\mathbf{x}) = 1 \ j = 1, \dots, K$;
	End
	Predict using a fixed combining rule (3)-(8);
<hr/>	

4 Experimental Studies

4.1 Datasets and Settings

We evaluated the proposed method on 15 datasets from UCI [28], STATLOG project [29], Knowledge Extraction based on Evolutionary Learning (KEEL) [30] and Ludmila Kuncheva Collection of real medical data (denoted by LKC) [31]. Information about the datasets is summarized in Table 1.

Table 1. Information of datasets in evaluation

Datasets	# of features	# of observations	# of classes	Source
Pima	8	768	2	UCI
Bupa	6	345	2	UCI
Wdbc	30	568	2	UCI
Blood	4	748	2	UCI
Sonar	60	208	2	UCI
Ionosphere	34	315	2	UCI
Vertebral	6	310	2	UCI
Waveform without noise (WwtN)	21	5000	3	UCI
Ecoli	7	336	8	UCI
Glass	9	214	6	UCI
Thyroid	3	215	3	LKC
Satimage	19	6435	7	STATLOG
Phoneme	6	5404	2	KEEL
Monk2	6	4322	2	KEEL
Mammographic	5	961	2	KEEL

We performed extensive comparison study with several well-known algorithms to validate our approach. In this study, we compared with several well-known DES methods namely MCB [13], MLA [14], KNOP [15], META-DES [6], DES-FA [17], DES-RRC [19], DES-KL [20], and DES-P [20]. The experiments concerning to those methods and the proposed method are conducted the same as experiments in [6, 12] (the value of k is set to 7). For the proposed method, we used C4.5 learning algorithm as the learning algorithm on 200 new training schemes to construct 200 base classifiers [7, 10, 21]. The new training sets were generated by using Gaussian-based random projections [7, 10] in which q was set as $q = 2 \times \log_2(p)$. We used Sum Rule to combine the results of EoC on each test sample.

We used Friedman test [32] to assess the statistical significance of the classification results of multiple methods on multiple datasets. Here we test the null hypothesis that “all methods perform equally” on the test datasets. If the null hypothesis is rejected, a post-hoc test is then conducted. In this paper, we used Shaffer’s procedure for all pairwise comparisons [32]. The difference in the performance of two methods is treated as statistically significant if the p-value computed from the post-hoc test statistic is smaller than an adjusted value of confident level computed from Shaffer’s procedure [32]. We set the confident level α to 0.05.

4.2 Comparing to Benchmark Algorithms

The experimental results of the benchmark algorithms and the proposed method are shown in Table 2. The proposed method obtains the best classification result in 10 datasets. On some datasets, the accuracy of the proposed method is significantly better

Table 2. Mean of accuracy of benchmark algorithms and the proposed method

Database	DES-FA	MLA	MCB	KNOP	DES-RRC	META-DES	DES-KL	DES-P	Proposed method
Pima	73.95	77.08	76.56	73.42	77.64	79.03	77.97	76.87	77.87
Bupa	61.62	58	58	65.23	68.01	70.08	67.11	67.46	68.53
Wdbc	97.88	95.77	97.18	95.42	96.94	97.4	97.13	96.78	97.61
Blood	73.4	76.06	73.4	77.54	78.02	79.14	78.83	77.72	79.82
Sonar	78.52	76.91	76.56	75.72	80.77	80.55	78.15	79.49	88.50
Ionosphere	88.63	81.81	87.5	85.71	88.8	89.94	88.42	88.42	96.30
Vertebral	82.05	77.94	84.61	86.98	86.76	86.89	84.19	86.76	86.93
WwtN	84.01	79.95	78.75	84.21	84.63	84.56	84.61	84.59	86.88
Ecoli	75.29	76.47	76.47	80	80.66	77.25	79.95	79.83	82.12
Glass	55.32	57.6	67.92	62.45	66.04	66.87	63.32	63.13	67.93
Thyroid	95.37	94.79	95.95	95.95	97.61	96.78	97.04	96.98	97.89
Satimage	93	93.28	95.86	95.86	96.38	96.21	96.2	96.22	96.52
Phoneme	79.06	64.94	73.37	78.92	74.65	80.35	77.13	81.64	96.22
Monk2	75.92	75.92	74.07	80.55	80.98	83.24	80.85	79.93	87.44
Mammographic	80.28	75.55	81.25	82.21	85	84.82	84.12	84.98	84.89

*The best results for each dataset are highlighted in bold

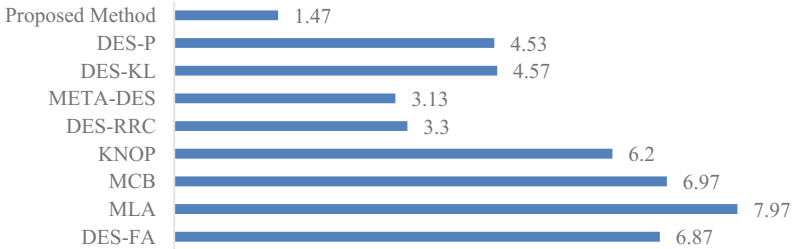


Fig. 2. Average ranking of all methods

than the best result of all benchmark algorithms, for example on Sonar (88.50 vs. 80.77 of DES-RRC), Ionosphere (96.30 vs. 89.94 of META-DES), and Phoneme (96.22 vs. 81.64). For the remaining five datasets, the difference between the accuracy of the proposed method and the best results are not significant except for two datasets, namely Pima (77.87 vs. 79.03 of META-DES) and Bupa (68.53 vs. 70.08 of META-DES).

Figure 2 shows the average rankings of the benchmark algorithms and the proposed method. It can be seen that the proposed method is ranked first (1.47), followed by META-DES (3.13) and DES-RRC (3.3). We conducted the Friedman test base on the rankings of the top five performing algorithms, i.e., DES-RRC, META-DES, DES-KL, DES-P, and the proposed method. In this case, the p-value computed by Friedman test is 1.44E-5. We rejected the null hypothesis of Friedman test and conducted the post-hoc test for all pairwise comparisons among those methods. From the Shaffer’s test results shown in Table 3, the proposed method is better than all four benchmark

Table 3. Statistical test result with Shaffer's procedure

Methods	P-value	Reject hypothesis	Methods	P-value	Reject hypothesis
DES-P vs. Proposed method	3.86E-6	Y	DES-RRC vs. DES-P	0.05675	N
DES-KL vs. Proposed method	8.76E-6	Y	META-DES vs. DES-KL	0.07349	N
DES-RRC vs. Proposed method	0.00666	Y	DES-RRC vs. DES-KL	0.08326	N
META-DES vs. Proposed method	0.00791	Y	DES-KL vs. DES-P	0.86249	N
META-DES vs. DES-P	0.04964	N	DES-RRC vs. META-DES	0.95396	N

**Y or N means that the performance of two methods is different or no different respectively*

algorithms. It shows the advantages of combining random projection and DES in a building a high-performance ensemble method.

5 Conclusion

We have introduced a novel ensemble by using two techniques DES and random projection to generate a single system. At first, original training set is projected to K down spaces to generate K training schemes. A learning algorithm will learn on these schemes to obtain K associated base classifiers. Validation set is also projected to the K down spaces so as to be used for DES in classification process. In classification process, a test sample is first projected to each of the down spaces. We determine how a base classifier be selected based on its prediction outcomes on the neighbors of each projected schemes of validation set. The experiments conducted on 15 datasets show that our framework is better than many of the state-of-the-art dynamic classifier/ensemble selection methods concerning to classification accuracy. In the future, the model can be extended to incrementally deal with stream data [33].

References

1. Britto, A.S., Sabourin, R., Oliveira, L.E.S.: Dynamic selection of classifiers—a comprehensive review. *Pattern Recog.* **47**(11), 3665–3680 (2014)
2. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996)
3. Ho, T.K.: The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 832–844 (1998)
4. Johnson, W., Lindenstrauss, J.: Extensions of Lipschitz mapping into Hilbert space. In: *Conference in Modern Analysis and Probability*, vol. 26, pp. 189–206 (1984). Contemporary Mathematics, American Mathematical Society
5. Fern, X.Z., Brodley, C.E.: Random projection for high dimensional data clustering: a cluster ensemble approach. In: *ICML*, pp. 186–193 (2003)

6. Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C., Ren, T.I.: META-DES: a dynamic ensemble selection framework using meta-learning. *Pattern Recog.* **48**(5), 1925–1935 (2015)
7. Nguyen, T.T., Nguyen, T.T.T., Pham, X.C., Liew, A.W.-C., Bezdek, J.C.: An ensemble-based online learning algorithm for streaming data. *CoRR abs/1704.07938* (2017)
8. Bingham, E., Mannila, H.: Random projection in dimensionality reduction: applications to image and text data. In: *ACM SIGKDD*, pp. 245–250 (2001)
9. Wang, Z., Yuan, X.-T., Liu, Q.: Sparse random projection for χ^2 kernel linearization: algorithm and applications to image classification. *Neurocomputing* **151**(1, 3), 327–332 (2015)
10. Pham, X.C., Dang, M.T., Dinh, V.S., Hoang, S., Nguyen, T.T., Liew, A.W.-C.: Learning from data stream based on random projection and Hoeffding tree classifier. In: *DICTA 2017* (in press)
11. Rathore, P., Bezdek, J.C., Erfani, S.M., Rajasegarar, S., Palaniswami, M.: Ensemble fuzzy clustering using cumulative aggregation on random projections. *IEEE Trans. Fuzzy Syst.* (2017, in press). <https://doi.org/10.1109/TFUZZ.2017.2729501>
12. Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C.: Dynamic classifier selection: recent advances and perspectives. *Inf. Fusion.* **41**, 195–216 (2018)
13. Giacinto, G., Roli, F.: Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recog.* **34**, 1879–1881 (2001)
14. Smits, P.C.: Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Trans. Geosci. Remote Sens.* **40**(4), 801–813 (2002)
15. Cavalin, P.R., Sabourin, R., Suen, C.Y.: Dynamic selection approaches for multiple classifier systems. *Neural Comput. Appl.* **22**(3–4), 673–688 (2013)
16. Ko, A.H.R., Sabourin, R., Britto, A.S.: From dynamic classifier selection to dynamic ensemble selection. *Pattern Recog.* **41**, 1735–1748 (2008)
17. Cruz, R.M.O., Cavalcanti, G.D.C., Ren, T.I.: A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. In: *IJCNN*, pp. 1126–1133 (2011)
18. Soares, R.G.F., Santana, A., Canuto, A.M.P., de Souto, M.C.P.: Using accuracy and diversity to select classifiers to build ensembles. In: *IJCNN*, pp. 1310–1316 (2006)
19. Woloszynski, T., Kurzynski, M.: A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recog.* **44**, 2656–2668 (2011)
20. Woloszynski, T., Kurzynski, M., Podsiadlo, P., Stachowiak, G.W.: A measure of competence based on random classification for dynamic ensemble selection. *Inf. Fusion* **13**(3), 207–213 (2012)
21. Nguyen, T.T., Nguyen, T.T.T., Pham, X.C., Liew, A.W.-C.: A novel combining classifier method based on variational inference. *Pattern Recog.* **49**, 198–212 (2016)
22. Nguyen, T.T., Nguyen, M.P., Pham, X.C., Liew, A.W.-C.: A hybrid classification system with fuzzy rule and classifier ensemble. *Inf. Sci.* **422**, 144–160 (2018)
23. Nguyen, T.T., Liew, A.W.-C., Pham, X.C., Nguyen, M.P.: A novel 2-stage combining classifier model with stacking and genetic algorithm based feature selection. In: Huang, D.-S., Jo, K.-H., Wang, L. (eds.) *ICIC 2014. LNCS (LNAI)*, vol. 8589, pp. 33–43. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09339-0_4
24. Nguyen, T.T., Liew, A.W.-C., Tran, M.T., Nguyen, M.P.: Combining multi classifiers based on a genetic algorithm – a Gaussian mixture model framework. In: Huang, D.-S., Jo, K.-H., Wang, L. (eds.) *ICIC 2014. LNCS (LNAI)*, vol. 8589, pp. 56–67. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09339-0_6

25. Nguyen, T.T., Liew, A.W.-C., Tran, M.T., Pham, X.C., Nguyen, M.P.: A novel genetic algorithm approach for simultaneous feature and classifier selection in multi classifier system. In: CEC, pp. 1698–1705 (2014)
26. Nguyen, T.T., Pham, X.C., Liew, A.W.-C., Pedrycz, W.: Aggregation of classifiers: a justifiable information granularity approach. CoRR abs/1703.05411 (2017)
27. Nguyen, T.T., Pham, X.C., Liew, A.W.-C., Nguyen, M.P.: Optimization of ensemble classifier system based on multiple objectives genetic algorithm. In: ICMLC, vol. 1, pp. 46–51 (2014)
28. Bache, K., Lichman, M.: UCI Machine Learning Repository (2013)
29. King, R.D., Feng, C., Sutherland, A.: STATLOG: comparison of classification algorithms on large real-world problems. *Appl. Artif. Intell. Int. J.* **9**(3), 289–333 (1995)
30. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Mult. Val. Log. Soft Comput.* **17**(2–3), 255–287 (2011)
31. Kuncheva, L.: Ludmila Kuncheva collection LKC (2004)
32. Garcia, S., Herrera, F.: An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *J. Mach. Learn. Res.* **9**, 2579–2596 (2008)
33. Nguyen, T.T., Weidlich, M., Duong, C.T., Yin, H., Nguyen, Q.V.H.: Retaining data from streams of social platforms with minimal regret. In: IJCAI, pp. 2850–2856 (2017)