

# MULTI-PERSPECTIVE CREATION OF DIVERSITY FOR IMAGE CLASSIFICATION IN ENSEMBLE LEARNING CONTEXT

HAN LIU<sup>1</sup>, SHYI-MING CHEN<sup>2</sup>

<sup>1</sup> School of Computer Science and Informatics, Cardiff University, Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom

<sup>2</sup>Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

E-MAIL: liuh48@cardiff.ac.uk, smchen@mail.ntust.edu.tw

## Abstract:

Image classification is a special type of classification tasks in the setting of supervised machine learning. In general, in order to achieve good performance of image classification, it is important to select high quality features for training classifiers. However, different instances of images would usually present very diverse features even if the instances belong to the same class. In other words, one types of features may better describe some instances, whereas other instances present more other types of features. The above description can indicate that the same learning algorithm may be capable of learning from some parts of a data set but show weaker ability to learn from other parts of a data set, given that different algorithms usually show different suitability for learning from instances that show various characteristics. On the other hand, image features are typically in the form of continuous attributes which can be handled by decision tree learning algorithms in various ways, leading to diverse classifiers being trained. In this paper, we investigate diversified adoption of the C4.5 and KNN algorithms from different perspectives, such as diversified use of instances and various ways of handling continuous attributes. In particular, we propose a multi-perspective approach of diversity creation for image classification in the setting of ensemble learning. We compare the proposed approach with those popular algorithms that are used to train classifiers on either a full set of original features or a subset of selected features for image classification. The experimental results show that the performance of image classification is encouraging through the adoption of our proposed approach of ensemble creation.

## Keywords:

Machine learning; Image classification; Feature selection; Ensemble learning; Diversity creation

## 1. Introduction

Image classification has become a popular application of machine learning, due to its high demand in real-life pattern recognition, e.g., handwritten digits recognition, letter

recognition and gesture recognition. In the setting of machine learning, if each image involves only one target object to be recognized, then the image can be treated as an instance, which belongs to a specific class. For example, in the context of handwritten digits recognition, each image involves a specific one of the 10 digits (0-9) to be recognized, so each of the 10 digits is defined as a class.

Since it is the essence of machine learning to learn from features, it is necessary to make sure that good quality features are selected and algorithms that are suitable to learn from the selected features are employed, towards production of high performance classifiers. However, image instances usually present very diverse features, even if the instances belong to the same class, which indicates the importance of dealing with feature diversity among image instances [1].

Traditional machine learning algorithms popularly used for image classification include C4.5 [2], Naïve Bayes (NB) [3], K Nearest Neighbors (KNN) [4] and Multi-layer Perceptron (MLP) [5]. However, due to the above-mentioned case of feature diversity, it is very necessary to investigate diversified use of each of the algorithms to produce classifiers that are complementary to each other leading to the increase of the number of correct classifications. In this paper, we investigate diversified adoption of the C4.5 and KNN algorithms from different perspectives, i.e., we propose a multi-perspective approach of diversity creation for image classification in the setting of ensemble learning.

The rest of this paper is organized as follows. Section 2 provides a review of popular approaches that can be used to create diversity among classifiers in an ensemble. In Section 3, we describe the procedure of our proposed approach of multi-perspective creation of diversity in the setting of ensemble learning. The setup of our experiments is described in Section 4 in details alongside the discussions of the results. The paper is concluded in Section 5 by highlighting the contributions and suggesting further directions.

## 2. Related work

Ensemble learning is aimed to train multiple classifiers that are combined (fused) to jointly classify new instances. As pointed out in [6], in order to achieve effectively that the classification performance is improved by fusion of multiple classifiers, it is crucial to ensure that the inclusion of a very badly performing classifier in the ensemble is avoided and that the ensemble needs to consist of diverse classifiers that can be complementary to each other.

The term ‘diversity’ generally means that different classifiers make incorrect classifications on different instances, i.e., it would be ideal to have each instance classified correctly by at least of one of the classifiers in the ensemble. In this case, the most optimistic outcome, which is referred to as ‘Oracle’ [7], is that all the instances can be classified correctly through fusion of the classifiers.

In machine learning literature, there have been some commonly adopted ways of diversity creation involved in popular ensemble learning approaches, such as Bagging [8], Random Subspace [9] and Boosting [10].

Bagging, which stands for bootstrap sampling, involves diversity creation through random sampling of training instances with replacement. In particular, there are  $n$  different training samples ( $D_1, D_2, \dots, D_n$ ) drawn from the original training data  $D$ . Since replacement is acceptable for the sampling, the new sample  $D_i$  may contain some duplicates and some instances in  $D$  may never appear in  $D_i$ , leading to 63.2% of the instances in  $D$  being selected into each new sample  $D_i$  on average. In terms of ensemble creation, the Bagging approach is designed to employ a specific learning algorithm and train a base classifier  $h_i$  on each newly drawn sample  $D_i$ . Since the  $n$  training samples ( $D_1, D_2, \dots, D_n$ ) represent  $n$  different subsets of  $D$ , the  $n$  base classifiers trained on  $n$  different training samples ( $D_1, D_2, \dots, D_n$ ) are likely to be diverse and complementary to each other.

The Random Subspace method is similar to the Bagging approach in terms of diversity creation. However, instead of randomly sampling of training instances, the Random Subspace method aims to randomly select  $n$  feature subsets, i.e., random projection of feature subspace, such that a base classifier  $h_i$  is learned from each of the  $n$  randomly selected feature subsets by using the same learning algorithm. Since the  $n$  base classifiers are trained on  $n$  different feature subsets, it leads to a high likelihood that the base classifiers are diverse and complementary to each other.

In practice, Bagging and Random Subspace have been jointly used to create decision tree ensembles and the method of ensemble creation is referred to as Random Forest [11], which has shown its great capability of advancing the performance of decision tree learning.

In contrast to Bagging and Random Subspace, which can enable the training of multiple base classifiers in parallel, Boosting is designed to train base classifiers sequentially. In particular, there are  $n$  iterations and at each iteration  $i$  a classifier  $h_i$  is trained, where the training of  $h_i$  is done by giving higher weights to instances incorrectly classified by  $h_{i-1}$  [12]. In this case, it is likely to obtain  $n$  classifiers that are diverse and complementary to each other, due to the focus on learning from different incorrectly classified instances at different iterations. In practice, the Boosting approach has also been used successfully to create decision tree ensembles and a popular method of such decision tree ensemble creation is referred to as Gradient Boosted Tree [13].

The ways of diversity creation through the above-mentioned ensemble learning approaches are general, which can be used to create ensembles based on any specific learning algorithms. Also, for all of the three ensemble learning approaches, the base classifiers must be trained using the same learning algorithm. However, it is very usual that the learning strategies of different algorithms are diverse, and some of the learning algorithms can be parametric, which means that different parameters setting for the same learning algorithm could result in very diverse classifiers being produced. Therefore, we design the ways of diversity creation from multiple perspectives in Section 3 for achieving more effective ensemble learning, in addition to the above-mentioned ways involved in the Bagging, Random Subspace and Boosting approaches, respectively.

## 3. The proposed multi-perspective approach of diversity creation

The proposed approach involves creation of diversity from multiple perspectives in the setting of ensemble learning. The proposed approach is shown in Fig.1 to illustrate its procedure of creating primary and secondary ensembles. All secondary ensembles, which are created separately following the procedure shown in Fig.1, are fused to make up the final ensemble. Therefore, the whole procedure essentially involves multiple levels of diversity creation.

At the sampling stage, the original data set  $D$  is taken to draw  $n$  training samples  $D_1, D_2, \dots, D_n$  through adopting the bootstrap sampling approach, which leads to primary creation of diversity. On each sample, we adopt the same learning algorithm with different setting of parameters for classifiers training, which enables the secondary creation of diversity, resulting from diverse parameters setting involved in the same algorithm. While various learning algorithms are used, multiple secondary ensembles can be created through adopting the Bagging approach, which enables the further creation of diversity between different secondary ensembles.

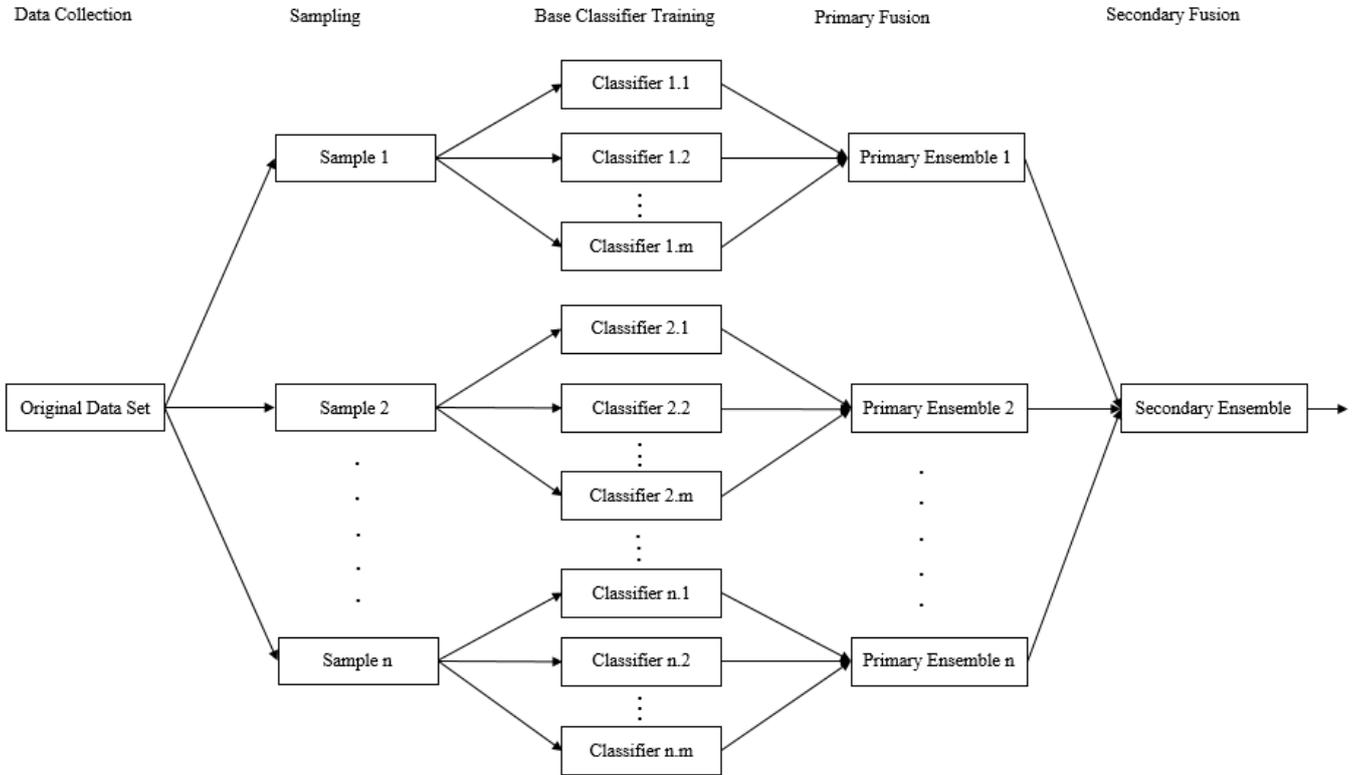


FIGURE 1. The procedure of the proposed approach.

In terms of ensemble creation,  $m$  base classifiers are trained on each of the  $n$  training samples drawn from the original training set, while  $m$  different ways of parameters setting are taken for the employed learning algorithm to train  $m$  base classifiers on sample  $D_i$ .

Furthermore, the  $n$  primary ensembles  $E_{j_1}, E_{j_2}, \dots, E_{j_n}$  created on the  $n$  training samples  $D_1, D_2, \dots, D_n$  are fused to make up a secondary ensemble  $E_j$ . While  $q$  different algorithms are employed to create secondary ensembles following the procedure shown in Fig. 2, there would be totally  $q$  secondary ensembles created.

For the primary ensembles, the members of the same ensemble are fused through one of the algebraic rules of fusion and the most popular one is the mean rule [7]. However, for fusion of the members of each secondary ensemble, majority vote is adopted to achieve the secondary fusion following the commonly used strategy involved in the Bagging approach [8] for providing an output.

The final ensemble is created through simple fusion of all the  $q$  secondary ensembles, where an algebraic rule of fusion is taken for finally classifying each new instance.

In the setting of the whole framework of ensemble

creation, each of the employed learning algorithms would normally involve various ways of parameters setting leading to the production of diverse classifiers that make up an ensemble. For example, a decision tree learning algorithm can be used to train diverse classifiers by involving different ways of handling continuous attributes and various ways of simplifying decision trees through different pruning methods. In general, slight changes of the parameters of a decision tree learning algorithm could result in very different trees being produced. From this point of view, diverse decision trees can be trained through the above-mentioned ways of parameters setting. Also, the KNN algorithm involves the setting of the K value (the number of nearest neighbors) and the way of assigning weights to different neighbors. Since similarity based classification (based on equally weighted nearest neighbors) is essentially to classify each new instance to the class assigned to the majority of the nearest neighbors, the setting of different values of K could result in different classification results for some of the new instances, i.e., the class assigned to the majority of the nearest neighbors may be different, while different values of K are used. Moreover, different ways of assigning weights to nearest neighbours can

further result in different outcomes regarding the mostly voted class. For example, while there are five nearest neighbors and three of them belong to the positive class, the mostly voted class would be the positive one if the five nearest neighbors are equally weighted. However, if the other two nearest neighbors that belong to the negative class have higher weights than the other three nearest neighbors, then the mostly voted class would be the negative one. Therefore, while multiple classifiers are trained using the same algorithm, the different setting of parameters for the algorithm can really lead to effective creation of the diversity among the trained classifiers.

In the creation of secondary ensembles, since each primary ensemble is created on a specific training data sample drawn randomly from the original training set, it is effectively achievable to encourage the creation of diverse primary ensembles on various training samples using the same learning algorithm, given that the Bagging approach has been popularly used for creating ensembles to advance the performance of most standard learning algorithms [8]. Therefore, the  $n$  primary ensembles created on the  $n$  training samples are also likely to be diverse and complementary to each other.

In the creation of the final ensemble, since the  $q$  secondary ensembles are created using  $q$  different learning algorithms, it is likely to encourage the diversity among the members of the final ensemble, given that different algorithms usually involve diverse strategies of learning from the same data. For example, each of the decision tree learning algorithms is essentially aimed at recursive evaluation of multiple candidate attributes and selection of the best candidate to generate a node of the decision tree being trained in the setting of inductive learning, whereas the KNN algorithm essentially involves measuring the similarity of each of the training instances to a new instance (to be classified) in the setting of instance-based learning. From this point view, it is obvious that diversity can be effectively created inside the final ensemble.

#### 4. Experimental results

Our experiments are conducted in this section using 4 UCI data sets [14] on image classification. Table 1 is presented to show the characteristics of each data set that contains a fairly small number of numeric features but a relatively large number of classes.

The experiments on all the data sets are conducted on the Weka platform using hold-out validation, i.e., each data set is partitioned by randomly selecting 70% of the instances for training and the rest (30%) for testing. The random partitioning on each data set is repeated 10 times and the

average accuracy obtained over 10 runs is taken for comparison of different methods.

TABLE 1. Characteristics of data sets

Data sets	Number of attributes	Number of instances	Number of classes
Letter	16	20000	26
Optdigits	64	5620	10
Pendigits	16	10992	10
Segment	19	2310	7

For results comparison, we compare the classification accuracy obtained using our proposed approach with the one obtained using various popular learning algorithms (C4.5, NB, KNN and MLP) to train classifiers on both the full set of original features and the subset of selected features resulting from using the correlation-based feature subset selection method (CFS) with the ‘BestFirst’ search strategy [15]. In terms of parameters setting for some parametric learning algorithms, the K value is set to 5 for KNN and all the 5 nearest neighbors to a new instance are considered equally weighted for classifying the new instance. The Euclidean distance function [4] is used to calculate the distance between the new instance and each of the training instances. The parameters of MLP are set as follows: hidden layers= (the number of attributes + the number of classes)/2, learning rate = 0.3, momentum = 0.2, batch size = 100, the number of epochs = 500. In addition, C4.5 is used to train unpruned decision trees on continuous attributes.

In terms of the setting of our proposed approach, the Bagging approach is adopted leading to the production of 10 training samples randomly drawn from the original training set. On each sample, a primary ensemble is created using either the C4.5 algorithm or the KNN algorithm, with different setting of parameters, while the mean rule is used for algebraic fusion of the base classifiers in each primary ensemble. In particular, 4 base classifiers are trained through parameters setting shown in Table 2 for each primary ensemble created using C4.5. Also, 3 base classifiers are trained through parameters setting shown in Table 3 for each primary ensemble created using KNN.

TABLE 2 Creation of decision tree ensembles

Base classifier number	Discretization of continuous attribute?	Pruning of decision trees?
1	No	No
2	Yes	No
3	No	Yes
4	Yes	Yes

For training the base classifiers shown in Table 2 to create a decision tree ensemble, the Gain Ratio heuristic function [2] is used to select attributes for generating nodes of a decision tree. The multi-level discretization method [16]

is used to handle continuous attributes for training base classifiers 2 and 4. In terms of pruning, the reduced error pruning (REP) algorithm [17] is adopted to simplify decision trees for training base classifiers 3 and 4.

**TABLE 3** Creation of KNN ensembles

Base classifier number	Weighting by $1/distance$ ?	Weighting by $1-distance$ ?
1	No	No
2	Yes	No
3	No	Yes

For training the base classifiers shown in Table 3 to create a KNN ensemble, the K value is set to 5 and the Euclidean distance function [4] is used to calculate the value of *distance*, in order to measure the similarity between a new instance (to be classified) and each of the training instances. In terms of distance weighting, each of the 5 nearest neighbors is weighted by  $1/distance$  for training base classifier 2 and by  $1-distance$  for training base classifier 3.

All the primary ensembles created using C4.5 are fused to create a secondary ensemble, whereas all the other primary ensembles created using KNN are fused to create another secondary ensemble. The two secondary ensembles are fused further to make up the final ensemble for classifying each new instance at the end.

According to the above settings of ensemble creation through the proposed approach, there are totally 70 base classifiers trained to make up the final ensemble. In order to generally increase the computational efficiency, the CFS method is adopted to reduce the dimensionality (the number of attributes) of the ‘Letter’, ‘Optdigits’ and ‘Pendigits’ data sets, due to the relatively large data size (the number of attributes  $\times$  the number of instances).

**TABLE 4** Classification accuracy

Methods	Letter	Optdigits	Pendigits	Segment
NB	0.641	0.908	0.857	0.803
NB+CFS	0.655	0.913	0.836	0.871
MLP	0.819	0.982	0.944	0.962
MLP+CFS	0.777	0.973	0.940	0.944
KNN	0.946	<b>0.985</b>	<b>0.990</b>	0.948
KNN+CFS	0.942	<b>0.985</b>	0.989	0.954
C4.5	0.864	0.895	0.961	0.962
C4.5+CFS	0.868	0.896	0.960	0.962
MPDC	<b>0.954</b>	0.984	0.989	<b>0.969</b>

The results are shown in Table 4 for comparing the proposed approach with the other algorithms that have been popularly used for image classification, where the acronym ‘MPDC’ stands for multi-perspective diversity creation and represents our proposed approach of diversity creation in ensembles. Also, the acronym NB represents that the NB algorithm is adopted for training classifiers on the full set of

original features and the acronym NB+CFS means that the NB algorithm is adopted for training classifiers on a subset of selected features resulting from the use of the CFS method. The same way of acronym definitions also applies to the other three learning algorithms (MLP, KNN and C4.5).

The results show that using the proposed MPDC approach leads to a slight improvement of the performance comparing with the use of the other methods on two out of the four data sets. For the ‘Optdigits’ and ‘Pendigits’ data sets, since the performance obtained using KNN and KNN+CFS has been good enough (accuracy higher 0.985), leading to very little space for improvements of the performance, the adoption of the proposed approach leads to a marginal drop in the classification accuracy. For the other 2 data sets, while the performance obtained using other methods shows a chance for improving the performance further, the adoption of the proposed approach has successfully led to slight advances in the classification accuracy.

Overall, it can be observed from Table 4 that through adopting the proposed MPDC approach it can be achieved effectively to keep the performance more stable and around the peak among the performance obtained using various methods over different data sets, while the performance of each of the other methods is varied on different data sets, e.g., among the other methods, the best performing one on the ‘Letter’, ‘Optdigits’ and ‘Pendigits’ data sets is KNN, but C4.5, C4.5+CFS and MLP perform the best on the ‘Segment’ data set. This phenomenon indicates that it is fairly difficult to achieve that the same learning method is the best performing one on all the data sets. Therefore, it is necessary to adopt the proposed approach of multi-perspective creation of diversity among multiple classifiers in the setting of ensemble learning, in order to achieve more stable performance over different data sets.

## 5. Conclusions

In this paper, we have proposed a multi-perspective approach of diversity creation for image classification in the setting of ensemble learning. Also, we have conducted experiments for comparison of the proposed ensemble creation approach with C4.5, NB, KNN and MLP, while both the full set of original features and a reduced set of selected features are provided for the four algorithms to train classifiers. The experimental results show that further advances in the performance have been achieved through adopting the proposed approach, which shows its potential advantage of keeping the classification performance more stable and around the peak among the performance obtained using various methods, over different data sets.

In the future, the adoption of fuzzy set theory will be

investigated to achieve effective setting of fuzzy ensemble learning [18]. It is also worth to explore the use of granular computing techniques [19] for multi-level creation of diversity and various ways of selection of optimal feature sets in the setting of multi-granularity ensemble learning.

### Acknowledgements

This work is supported by the Social Data Science Lab and the School of Computer Science and Informatics at the Cardiff University in the UK.

### References

- [1] W. Ding, X. Wang, H. Liu, and B. Hu, "An empirical study of shape recognition in ensemble learning context", in Proceedings of International Conference on Wavelet Analysis and Pattern Recognition, Chengdu, China, 2018, pp. 256-261.
- [2] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 1993.
- [3] I. Rish, "An empirical study of the Naive Bayes classifier", in Proceedings of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, USA, 2001, Vol. 3, No. 22, pp. 41-46.
- [4] J. Zhang, "Selecting typical instances in instance-based learning", in Proceedings of the Ninth International Workshop on Machine Learning, Aberdeen, Scotland, United Kingdom, 1992, pp. 470-479.
- [5] R. Collobert, and S. Bengio, "Links between perceptrons, MLPs, and SVMs", in Proceedings of the 21st International Conference on Machine learning, Banff, Alberta, Canada, 2004.
- [6] Z. H. Zhou, "Ensemble Methods: Foundations and Algorithms", Chapman and Hall/CRC, 2012.
- [7] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies?", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 2, pp. 281-286, 2002.
- [8] L. Breiman, "Bagging predictors", Machine Learning, Vol. 24, No. 2, pp. 123-140, 1996.
- [9] T. K. Ho, "The random subspace method for constructing decision forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, pp. 832-844, 1998.
- [10] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm", in Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 1996, pp. 148-156.
- [11] L. Breiman, "Random forests", Machine Learning, Vol. 45, No. 1, pp. 5-32, 2001.
- [12] M. P. P. Jr, "Combining classifiers: from the creation of ensembles to the decision fusion", in Proceedings of the 24th SIBGRAPI Conference on Graphics, Patterns, and Image Tutorials, Alagoas, Brazil, 2011, pp. 1-10.
- [13] J. O. Ogutu, H. P. Piepho, and T. Schulz-Streeck, "A comparison of random forests, boosting and support vector machines for genomic selection", BMC Proceedings, Vol. 5, No. 3, 2011.
- [14] M. Lichman, UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml>, 2013.
- [15] M. A. Hall and L. A. Smith, "Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper", in Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference, Orlando, Florida, 1999, pp. 235-239.
- [16] U. M. Fayyad and K. B. Irani, "Multi-level discretization of continuous-valued attributes for classification learning", in Proceedings of the 13th International Joint Conference on Artificial Intelligence, Chambéry, France, 1993, pp. 1022-1027.
- [17] T. Elomaa and M. Kaariainen, "An analysis of reduced error pruning", Journal of Artificial Intelligence Research, Vol. 15, No. 1, pp. 163-187, 2001.
- [18] G. Nakai, T. Nakashima, and H. Ishibuchi, "A fuzzy ensemble learning method for pattern classification", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 15, No. 6, pp. 671-681, 2003.
- [19] X. Wang, W. Ding, H. Liu, and X. Huang, "Shape recognition through multi-level fusion of features and classifiers", Granular Computing, Vol 5, No. 3, 2020.