

Enhancement of the Heuristic Optimization Based on Extended Space Forests using Classifier Ensembles

Zeynep Kilimci^{1,3} and Sevinç Omurca²

¹Department of Computer Engineering, Dogus University, Turkey

²Department of Computer Engineering, Kocaeli University, Turkey

³Department of Information Systems Engineering, Kocaeli University, Turkey

Abstract: *Extended space forests are a matter of common knowledge for ensuring improvements on classification problems. They provide richer feature space and present better performance than the original feature space-based forests. Most of the contemporary studies employ original features as well as various combinations of them as input vectors for extended space forest approach. In this study, we seek to boost the performance of classifier ensembles by integrating them with heuristic optimization-based features. The contributions of this paper are fivefold. First, richer feature space is developed by using random combinations of input vectors and features picked out with ant colony optimization method which have high importance and not have been associated before. Second, we propose widely used classification algorithm which is utilized baseline classifier. Third, three ensemble strategies, namely bagging, random subspace, and random forests are proposed to ensure diversity. Fourth, a wide range of comparative experiments are conducted on widely used biomedicine datasets gathered from the University of California Irvine (UCI) machine learning repository to contribute to the advancement of proposed study. Finally, extended space forest approach with the proposed technique turns out remarkable experimental results compared to the original version and various extended versions of recent state-of-art studies.*

Keywords: *Classifier ensembles, extended space forests, ant colony optimization, decision tree.*

Received November 11, 2017; accepted March 11, 2018

<https://doi.org/10.34028/iajit/17/2/6>

1. Introduction

Ensemble learning is a very popular research area in the literature and has also known as committees of learners, mixture of experts, ensemble of classifiers, ensemble algorithms [1, 4, 14, 16, 35]. The idea behind of this principle is to make use of more than one classifier. Thus, the classification task is performed more robustly and accurately [10, 24, 26, 27, 28]. Supervised machine learning techniques such as naïve Bayes classifiers, decision trees, Support Vector Machines (SVM), Artificial Neural Networks (ANN), k-Nearest Neighborhood (k-NN) are commonly used for the ensemble strategies. Especially, decision tree is more widely implemented in the literature for ensemble classifiers compared to the other classification methods [1, 2, 3, 4, 5, 11]. Utilization of more than one decision tree emerges decision forests for the classifier ensembles.

As much as the selection of classifier, the individual success and diversity of base learners are also determinative factors of the ensemble success. As the diversity of base learner increases, the classification success of system becomes better. The usage of different or same base learners is requisite in order to provide diversity. Diversity is maintained with several conventional ensemble algorithms such as bagging, random subspaces, random forests, and rotation forest for the same base learners. For different base learners,

it is already achieved by blending different learning algorithms with various decision making techniques such as majority voting, stacking, cascading. In this work, we focus on the same base learners to provide diversity is called homogeneous classifier ensembles.

In this paper, we propose to contribute to the classification performance of ensemble system using heuristic optimization based technique namely, Ant Colony Optimization (ACO) for biomedicine datasets. For this purpose, extended space forest employs the association of randomly chosen features and ACO based significant features to construct extended feature space. Then, decision tree construction is carried out on training phase according to the ensemble algorithms namely, bagging, random subspace, and random forest. For demonstrating the contribution of proposed technique, we conduct experiments on datasets available from the University of California Irvine (UCI) machine learning repository [8].

The rest of the paper is organized as follows: section 2 gives related researches on the use of ensemble systems and ant colony optimization. In section 3, we give proposed feature extension technique employed in the experiments. Experiment setup and results are demonstrated in sections 4 and 5. Section 6 concludes the paper with a discussion and conclusions.

2. Related Work

Ensemble learning is the collection of methods that builds a set of classifiers and combines their classification predictions by using majority voting [1, 4, 10, 13, 23, 30]. Previous studies [20, 26, 29, 30, 32] have indicated that the final decision of ensemble system is more accurate and robust than any of the individual classifiers in the ensemble. In an interesting study [34] on ensemble of feature sets and classification algorithms, authors focus on the effectiveness of ensemble methods for sentiment classification. Experimental results show that the usage of ensemble methods for both combining different feature sets and classification algorithms is efficient to boost classification performance. Another important study [17] on ensemble learning proposes the novel SVM technique namely, the Ensemble of SVM (EnSVM), to excel the classification success of SVM on imbalanced data. The extensive experiments demonstrate that their proposed technique is competitive, effective and superior in proportion to several methods with various data sampling techniques. The other study also [25] proposes to investigate the effectiveness of the ensemble classifiers for offline cursive character recognition. Their experiment results point out that the accuracy of recognition success using ensemble algorithms is better than the traditional recognition accuracies for offline cursive character recognition.

There are also several studies on classifier ensembles with extended space. The influential study [4] proposes the extended feature space by choosing new features randomly and adding them to original feature space. They observe that all extended versions outperform original versions for all ensemble algorithms. To get higher classification performance of ensemble system, they suggest utilizing the extended space methods. The recent studies [1, 2] on extended space decision trees propose to increase the ensemble accuracy by suggesting another approach. Instead of randomly producing, new features with high classification capacity are generated by computing the gain ratio of each different candidate features. Thus, they combine newly generated features and existing features in order to extend feature space. Authors conclude that the extended space forest is an effective method to increase prediction accuracy but it can be improved by using significant features instead of selecting randomly. Another recent study [32] investigates the effectiveness of enhanced Random Subspace Method Based on Part-of-Speech Technique (POS-RS), for sentiment categorization field. They conclude that POS-RS is preferable method to excel the success of classification and applied to the other text classification problems.

3. Proposed Technique and Ensemble Strategies

As noted in the previous works [1, 2, 4] the enrichment of feature space ensures significant contribution to the classification performance. The studies so far on extended space forests utilize either randomly chosen features [4] or the specific feature selection method such as gain ratio [1, 2] to determine new candidate features to be consolidated to the original feature space. In the proposed approach on extended space forests, we take the approach one step further in extended space forests by employing heuristic optimization technique namely, ACO as a feature selection method in addition to the random combinations of input vectors.

The ant colony optimization is an optimization technique that can be also employed for feature selection on various domains. It is based on finding the shortest paths from the nest to food source by means of pheromone trails, which is an odorous substance and is excreted by ants. Therefore, the deposition of pheromone is the fundamental factor in order to discover the shortest paths over a certain period of time. The way excreted pheromone is used by more ants and pheromone trails probabilistically enforce to choose the previously marked path for each isolated ant. On less preferred paths, pheromone evaporates over time and the shortest path is discovered by means of the higher ratio of ant traversals. For this reason, there is a transition probabilistic rule for each ant to determine the probability of being selected corresponding path. Hence, ACO technique is attractive for feature selection process that can direct search to optimal subset every time [9]. Easy implementation and superior performance of ACO encourages us to enrich feature space. When 100 base learners are thought to be used for ensemble strategy, it is expected to use random features for each base learner (for each isolated ant) at first. Furthermore, the pheromone intensity is updated on the route drawn for each ant to avoid stagnation and then the isolated ants can choose different routes (features). Therefore, the extension of feature space can be provided with different features for each base learner. The probabilistic transition rule, expressing the probability of an ant at feature i choosing to travel to feature j at time t :

$$p_{ij}^k(t) = \begin{cases} \frac{\sum_{l \in J_k^i} [\tau_{il}(t)]^\alpha [\eta_{ij}^\beta]}{[\tau_{il}(t)]^\alpha [\eta_{ij}^\beta]} & \text{if } j \in J_k^i, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where k is the number of ants, η_{ij} is the heuristic desirability of selecting feature j when at feature i , J_k^i is the set of ant k 's unvisited features, and $\tau_{ij}(t)$ is the amount of virtual pheromone on edge (i, j) , α provides global information and determine the relative importance of the pheromone value, β is the heuristic information and present local information. Producing a

number of k ants is the first step for ACO feature selection process. In this study, the number of ants is set to the number of features within dataset. Thus, each ant begins with one random feature and they travel edges probabilistically until stopping gauge is fulfilled. The subsets are congregated and then evaluated. Once the algorithm has performed a certain number of times or an optimal subset has attained, the overall feature selection process terminates by obtaining the best feature output. If neither condition holds, it is inevitable to update the intensity of pheromone, then new ants are produced and the feature selection process reiterates once more.

Ultimately, the extended feature space comprises original, randomly chosen features and significant ones selected with ACO. In addition to the d number of original features, the d number of space extension parameter is adjusted to extend feature space due to its superior performance as stated in [4]. While the first half of new features is composed from significant features chosen with ACO, the remaining half is generated as the randomly combined pairs of the original features. After constructing the enriched feature space, the extended space forest is evaluated with the well-known classification algorithm, namely decision tree by using three types of ensemble algorithms such as bagging, random subspaces, and random forests. Figure 1 illustrates the process of extended space forest with proposed technique.

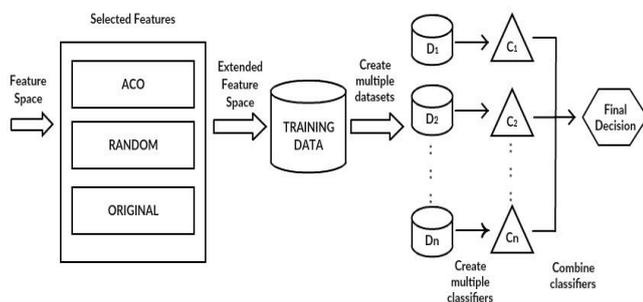


Figure 1. The process of extended space forests with proposed technique.

Ensemble algorithms used in this work are briefly mentioned. Bagging [4, 6, 12, 18, 21, 31, 32, 33] generates new bootstrap samples utilizing substitution from the original dataset. Then, training is done by constituting on each of these samples and associated them with majority voting ensemble strategy. Random Subspace [4, 11, 12, 15, 19, 21, 22, 31, 32] exploits fairly simple randomness approach for the feature selection. Training is done with a subset of the original feature space instead of including all features for each base learner in the ensemble. Then, the classifier is constructed on different feature subsets illustrated randomly from the original feature set and associated by applying the majority voting. Random Forest [2, 4, 7] combines two approaches namely, Bagging and Random Subspace algorithms. The proposed approach

is also described in details below. In proposed approach, the parameters are given as: $E = \{x_p, y_p\}_{p=1 \dots N} = [X \ Y]$ where X is an $N \times d$ matrix including the training set and Y is an N dimensional column vector covering the class labels. N is the number of training samples, d is the number of features, T is the number of base learners, BL_i is the base learner, E_i is the enhanced training set for BL_i , R_i consists of feature pair indices used in generation of E_i , EA is an ensemble algorithm. At initialization step, ensemble size T , the base learner model BL_i , and the ensemble algorithm EA are chosen.

Algorithm 1: The Proposed Approach

```

#training
for (i=1 to T)
#create new features (EXi) by using feature selection
techniques and randomly paired original features.
#generate d/2 random permutations of the original feature
indices and store in Ri.
#generate d/2 significant features chosen with ACO and
store in Si.
j=1
for (z=1 to d/2)
{
#create jth new feature applying difference operator to Ri(z)th
and Ri(z+1)th features of X matrix.
j=j+1
}
#construct the new training set (Ei) by concatenating the
matrix X (original features), Ri and Si (the new features) as
Ei=[X Si Ci Y].
#train BLi with Ei according to EA.

#testing
for (i=1 to T)
#enhance the feature space of the test sample.
#classify the enhanced test sample with BLi.
#combine the base learners' decisions by the combination
rule of the chosen ensemble algorithm EA.
  
```

4. Experiment Setup

We utilize datasets with different sizes and properties that are available from the UCI Machine Learning Repository [8] in experiments. Characteristics of the datasets are presented in Table 1 including the number of features ($|F|$) and the number of classes ($|C|$). Experiments are carried out by modifying the training set levels and utilizing 10%, 30%, 50%, and 80% percentages as a training data. The accuracy percentage levels are abbreviated with "ts" affix to head a commotion off. The algorithms are launched at each training set levels by partitioning 10 parts randomly and stratified sampling is exploited at this step. Statistical significance tests are also employed in several places by evaluating Student's t-test which is beneficial as the closeness of accuracy results of different techniques is observed. Significance level is set to 0.05 and the difference is accounted as statistically significant when the association of probability and Student's t-Test is lower. The number of base learners is adjusted to 100 as represented in [2,

4]. As we mentioned before feature extension parameter is set to d number of features for all datasets for comparing experiment results with impressive work [4]. To combine the decisions of base learners, majority voting is employed for all ensembles.

Table 1. Characteristics of the datasets.

| Dataset ID | Dataset | F | C | Dataset ID | Dataset | F | C | Dataset ID | Dataset | F | C |
|------------|---------------|----|----|------------|---------------|-----|----|------------|---------------|-----|----|
| 1 | Abalone | 10 | 19 | 13 | Diabetes | 8 | 2 | 25 | primary-tumor | 23 | 11 |
| 2 | Anneal | 62 | 4 | 14 | Glass | 9 | 5 | 26 | ringnorm | 20 | 2 |
| 3 | audiology | 69 | 5 | 15 | heart-statlog | 13 | 2 | 27 | segment | 18 | 7 |
| 4 | Autos | 71 | 5 | 16 | hepatitis | 19 | 2 | 28 | sick | 31 | 2 |
| 5 | balance-scale | 4 | 3 | 17 | Hypothyroid | 31 | 3 | 29 | sonar | 60 | 2 |
| 6 | breast-cancer | 38 | 2 | 18 | Ionosphere | 33 | 2 | 30 | soybean | 83 | 18 |
| 7 | breast-w | 9 | 2 | 19 | iris | 4 | 3 | 31 | splice | 287 | 3 |
| 8 | col10 | 7 | 10 | 20 | kr-vs-kp | 39 | 2 | 32 | vehicle | 18 | 4 |
| 9 | Colic | 60 | 2 | 21 | labor | 26 | 2 | 33 | vote | 16 | 2 |
| 10 | credit-a | 42 | 2 | 22 | letter | 16 | 26 | 34 | vowel | 11 | 11 |
| 11 | credit-g | 59 | 2 | 23 | lymph | 37 | 2 | 35 | waveform | 40 | 3 |
| 12 | d159 | 32 | 2 | 24 | mushroom | 112 | 2 | 36 | zoo | 16 | 4 |

Moreover, it is necessary to specify some parameters for ACO feature selection process. Firstly, the number of ants is equal to the number of features for each dataset. Because of this, the number of ants varies according to the dataset. Then, the algorithm has carried out a certain number of times is the same as the number of base learners, i.e., 100 times. After the algorithm has executed 100 times, the pheromone density is updated and a new set of ants are composed and the process iterates once more. The initial pheromone density of each feature is set to 1 at first, two important information, local and global, about the traversal of ants is determined with the parameters α and β . The choice of α , β is specified experimentally and set to 1 and 0.1, respectively. The pheromone trail evaporation coefficient ($\rho=0.2$) is a parameter to update pheromone trails and located in the range between 0 and 1.

5. Experiment Results

Ensemble Accuracy (EA) is utilized as an evaluation metric to demonstrate the contribution of our work. Abbreviations are employed for the ensemble algorithms and feature selection techniques as follows: BG: Bagging, RS: Random subspaces, RF: Random forests, X_0 : Original version of dataset for X ensemble algorithm, X_{RD} : Extended space forest by adding randomly selected features for X ensemble algorithm, $X_{ACO + RD}$: Extended space forest by adding randomly chosen and ACO based features for X ensemble algorithm.

The classification success is ordered as $RF_{ACO + RD} > RS_{ACO + RD} > BG_{ACO + RD} > RF_{RD} > RS_{RD} > BG_{RD} > RF_0 > RS_0 > BG_0$ at ts80 in terms of ensemble accuracies as seen in Table 2. Except ts30, the best classification

performance is performed by $RF_{ACO + RD}$ at all training set sizes. Thus, it can be asserted that the enhanced space forests based upon ACO and random features contribute to the classification performance significantly for 36 datasets. Except ts30, the success order of original versions of ensemble algorithms is $RF > RS > BG$. At smaller training set levels like ts30, the performance order of original versions of ensemble algorithms is different but not enough to claim statistically considerable because of the closeness of accuracy results. Moreover, RF_{ACO+RD} outperforms others at all training set percentages except ts30 level. At ts30, RS_{ACO+RD} is competitive and surpasses other techniques by at 1%. The combination of random subspace as an ensemble algorithm and ACO based and randomly selected features as a feature selection technique yields by far the highest accuracies at this training set level.

At the last of the Table 2, average accuracy results are presented. Average accuracy results for bagging algorithm demonstrate that the extended space forests by adding randomly selected and ACO based features is the winner with 87.9% accuracy value at ts80 compared to others. The success order of original and extended versions of bagging algorithm is given as follows: $BG_{ACO+RD} > BG_{RD} > BG_0$. Similarly, the extended version with ACO+Randomly selected features (RD) outperforms others for the random subspace and the random forest algorithms with 88.1% and 88.4% accuracy results, respectively. Like bagging algorithm, the performance order of random subspace and random forest is the same. Hence, our proposed method based extended space forest (ACO+RD) is the best model to enhance the classification accuracy for each ensemble algorithm in terms of average accuracy results.

For the original space forests, random forest is the best ensemble algorithm with 87.0% classification success and followed by random subspace with 86.9% and bagging with 86.2% accuracy result, respectively. The classification performance of ACO+RD based extended space forests is ordered in a similar way as: $RF > RS > BG$ and the classification performance is consistent with the state-of-art results [4]. This order is also valid for the extended space forests with randomly chosen features but the classification accuracies are different from each other in that 87.8% RF, 87.7% RS, 87.3% BG.

It is significant to perceive that the classification results of random forest and random subspace algorithms are close to each other but yet, random forest algorithm generally surpasses the others at ts80 by evaluating Table 2. As a feature selection technique the consolidation of random and ACO based features is an optimal in order to enrich feature space.

Table 2. The Classification accuracies of extended and original versions of ensemble algorithms at ts80.

| Dataset ID | BG ₀ | BG _{RD} | BG _{ACO+RD} | RS ₀ | RS _{RD} | RS _{ACO+RD} | RF ₀ | RF _{RD} | RF _{ACO+RD} | Dataset ID | BG ₀ | BG _{RD} | BG _{ACO+RD} | RS ₀ | RS _{RD} | RS _{ACO+RD} | RF ₀ | RF _{RD} | RF _{ACO+RD} |
|------------|-----------------|------------------|----------------------|-----------------|------------------|----------------------|-----------------|------------------|----------------------|---------------|-----------------|------------------|----------------------|-----------------|------------------|----------------------|-----------------|------------------|----------------------|
| 1 | 27.3 | 28.4 | 29.2 | 27.2 | 27.2 | 27.5 | 28.1 | 27.8 | 27.8 | 19 | 97.1 | 96.8 | 97.0 | 96.5 | 97.3 | 97.5 | 96.9 | 96.8 | 96.5 |
| 2 | 99.1 | 99.3 | 99.5 | 99.2 | 99.0 | 99.4 | 99.6 | 99.6 | 99.6 | 20 | 99.1 | 99.3 | 99.5 | 98.4 | 99.5 | 99.0 | 98.8 | 99.9 | 99.9 |
| 3 | 89.3 | 89.5 | 90.4 | 87.4 | 90.6 | 92.1 | 87.2 | 87.3 | 87.2 | 21 | 90.7 | 89.2 | 90.4 | 93.9 | 92.4 | 92.8 | 92.2 | 96.1 | 98.3 |
| 4 | 73.1 | 75.0 | 76.5 | 72.6 | 74.7 | 75.9 | 72.1 | 75.3 | 76.2 | 22 | 93.7 | 97.1 | 98.5 | 96.0 | 97.0 | 97.5 | 96.1 | 96.8 | 97.1 |
| 5 | 86.1 | 97.2 | 98.7 | 87.0 | 93.2 | 95.2 | 88.1 | 98.6 | 99.1 | 23 | 85.7 | 87.8 | 88.3 | 86.8 | 86.6 | 86.4 | 86.2 | 85.8 | 86.0 |
| 6 | 73.4 | 74.1 | 74.9 | 75.3 | 75.1 | 75.0 | 75.4 | 73.1 | 74.4 | 24 | 98.7 | 98.9 | 98.7 | 99.2 | 99.7 | 99.9 | 99.7 | 99.9 | 99.9 |
| 7 | 97.3 | 97.5 | 97.5 | 97.8 | 97.3 | 97.1 | 97.8 | 97.9 | 97.5 | 25 | 51.3 | 50.7 | 51.6 | 51.8 | 51.5 | 51.0 | 51.7 | 53.6 | 55.4 |
| 8 | 81.6 | 81.0 | 81.4 | 81.7 | 81.2 | 81.4 | 81.7 | 81.5 | 81.7 | 26 | 95.7 | 96.5 | 97.6 | 97.8 | 97.6 | 97.5 | 96.4 | 97.1 | 97.8 |
| 9 | 85.3 | 86.5 | 87.9 | 85.9 | 87.2 | 87.8 | 84.2 | 86.8 | 89.5 | 27 | 97.3 | 97.7 | 97.7 | 97.6 | 98.2 | 98.7 | 98.2 | 98.0 | 98.5 |
| 10 | 88.1 | 88.8 | 89.1 | 89.2 | 89.0 | 89.2 | 88.6 | 87.7 | 88.3 | 28 | 99.1 | 99.0 | 99.9 | 98.3 | 99.2 | 99.6 | 98.7 | 98.6 | 98.3 |
| 11 | 77.9 | 78.8 | 79.3 | 77.6 | 78.7 | 79.3 | 78.1 | 78.6 | 79.0 | 29 | 79.4 | 79.1 | 79.5 | 80.8 | 80.7 | 80.5 | 81.7 | 82.4 | 83.7 |
| 12 | 99.1 | 99.7 | 99.7 | 99.0 | 99.6 | 99.9 | 99.9 | 99.9 | 99.9 | 30 | 93.2 | 92.8 | 93.0 | 93.5 | 93.2 | 93.0 | 92.5 | 93.4 | 94.6 |
| 13 | 76.9 | 76.6 | 76.3 | 76.2 | 76.9 | 77.1 | 77.1 | 78.0 | 78.7 | 31 | 96.0 | 99.9 | 99.9 | 97.1 | 96.5 | 96.8 | 96.5 | 97.6 | 97.6 |
| 14 | 74.1 | 75.8 | 76.2 | 75.1 | 77.3 | 77.9 | 74.3 | 74.0 | 74.5 | 32 | 76.0 | 77.2 | 77.7 | 76.4 | 79.8 | 80.9 | 77.3 | 80.2 | 84.1 |
| 15 | 82.2 | 82.3 | 82.1 | 83.6 | 83.2 | 83.0 | 84.2 | 82.1 | 83.7 | 33 | 97.1 | 98.0 | 98.6 | 97.3 | 98.2 | 98.6 | 97.8 | 98.2 | 98.4 |
| 16 | 82.9 | 85.7 | 86.7 | 85.7 | 87.4 | 88.0 | 86.0 | 86.9 | 87.4 | 34 | 83.4 | 87.6 | 89.5 | 88.1 | 91.0 | 92.5 | 88.5 | 90.4 | 91.3 |
| 17 | 99.6 | 99.5 | 99.4 | 97.7 | 99.9 | 99.9 | 99.7 | 99.7 | 99.5 | 35 | 86.1 | 87.7 | 88.2 | 87.5 | 88.4 | 88.9 | 88.2 | 88.7 | 88.0 |
| 18 | 93.9 | 94.9 | 95.6 | 94.9 | 95.1 | 95.4 | 94.0 | 94.3 | 94.0 | 36 | 96.5 | 97.1 | 97.8 | 99.1 | 99.0 | 99.0 | 99.7 | 99.7 | 99.5 |
| | | | | | | | | | | Mean Accuracy | 86.2 | 87.3 | 87.9 | 86.9 | 87.7 | 88.1 | 87.0 | 87.8 | 88.4 |

As we pointed out above, if we compare all versions of the space forests according to the ensemble algorithms, we can attain the classification success order as: $RF_{ACO+RD} > RS_{ACO+RD} > BG_{ACO+RD} > RF_{RD} > RS_{RD} > BG_{RD} > RF_0 > RS_0 > BG_0$. In this study, we try to acquire better classification performance compared to the previous studies [1, 2, 4] and implement training procedure on the extended feature space with an appropriate feature selection model and then associate with various ensemble algorithms.

Experiment results exhibit that the combination of random and ACO based features and random forest as an ensemble algorithm has an exceptional classification performance.

Unlike smaller training set percentage levels, random forest algorithm with ACO+RD reaches the maximum value at ts80. All versions of random subspace algorithm have the following best classification performances. So long as training set percentages increase, the successes of all enhanced space forest versions also rise up and vice versa.

Table 3. Comparison between pairs of algorithms: “win (significant win)/ loss (significant loss)” row vs. column.

| | BG _{ACO+RD} | BG _{RD} | BG ₀ | RS _{ACO+RD} | RS _{RD} | RS ₀ | RF _{ACO+RD} | RF _{RD} | RF ₀ |
|----------------------|----------------------|------------------|-----------------|----------------------|------------------|-----------------|----------------------|------------------|-----------------|
| BG _{ACO+RD} | 0/0 | 28(6)/8(0) | 28(15)/8(0) | 15(5)/21(11) | 17(7)/19(6) | 24(12)/12(3) | 16(8)/20(9) | 18(11)/18(6) | 21(9)/15(5) |
| BG _{RD} | 8(0)/28(6) | 0/0 | 27(9)/9(1) | 6(2)/30(12) | 11(3)/25(8) | 21(7)/15(5) | 11(4)/25(14) | 13(6)/23(9) | 17(6)/19(5) |
| BG ₀ | 8(0)/28(15) | 9(1)/27(9) | 0/0 | 5(0)/31(14) | 6(0)/30(15) | 10(2)/26(10) | 4(1)/32(17) | 8(1)/28(14) | 6(2)/30(10) |
| RS _{ACO+RD} | 21(11)/15(5) | 30(12)/6(2) | 31(14)/5(0) | 0/0 | 25(3)/11(0) | 23(12)/13(1) | 17(4)/19(8) | 21(7)/15(4) | 25(10)/11(1) |
| RS _{RD} | 19(6)/17(7) | 25(8)/11(3) | 30(15)/6(0) | 11(0)/25(3) | 0/0 | 20(9)/16(1) | 10(2)/26(9) | 17(4)/19(5) | 22(8)/14(0) |
| RS ₀ | 12(3)/24(12) | 15(5)/21(7) | 26(10)/10(2) | 13(1)/23(12) | 16(1)/20(9) | 0/0 | 10(0)/26(14) | 10(3)/26(8) | 14(2)/22(4) |
| RF _{ACO+RD} | 20(9)/16(8) | 25(14)/11(4) | 32(17)/4(1) | 19(8)/17(4) | 26(9)/10(2) | 26(14)/10(0) | 0/0 | 24(9)/12(0) | 20(12)/16(1) |
| RF _{RD} | 18(6)/18(11) | 23(9)/13(6) | 28(14)/8(1) | 15(4)/21(7) | 19(5)/17(4) | 26(8)/10(3) | 12(0)/24(9) | 0/0 | 23(7)/13(2) |
| RF ₀ | 15(5)/21(9) | 19(5)/17(6) | 30(10)/6(2) | 11(1)/25(10) | 14(0)/22(8) | 22(4)/14(2) | 16(1)/20(12) | 13(2)/23(7) | 0/0 |

Original versions of all ensemble algorithms present the lowest classification accuracies at higher training set percentages. At these training set levels, the choice of original versions of ensemble algorithms will not be a good preference for the classification problems. The following consequences can be drawn from Table 3: BG_{ACO+RD} has higher accuracy than BG₀ over 28 datasets out of 36, and has 15 significant wins.

RS_{ACO+RD} has higher accuracy than BG_{ACO+RD} over 21 datasets out of 36, and has 11 significant wins. RF₀ has higher accuracy than RS₀ over 22 datasets out of 36, and 4 significant wins. RF_{RD} has higher accuracy than RS_{RD} over 19 datasets out of 36, and 5 significant wins.

6. Discussion and Conclusions

The superiority of ensemble systems is a widely accepted assumption in machine learning domain as mentioned before. Owing to this approach, it is recommended to produce more accurate and robust models. In this work, we propose to investigate the contribution of enhanced space forests to the classification performance by utilizing ensemble algorithms. For this purpose, we take the concept one step further in extended space forests by employing ACO as a feature selection technique which have not been combined before. Moreover, this is the first research for the extended space forests with classifier ensembles in terms of utilizing heuristic optimization technique, ant colony optimization, as a feature selection technique. Features chosen with ACO and random combinations of input vectors are blended with the original features to constitute a new extended feature space. Then, the enriched feature space is carried out on three popular ensemble algorithms (Bagging, Random Subspaces, and Random Forest) by utilizing decision tree as a base learner. Finally, the extended space forests developed by the proposed method ensure noteworthy enhancement to the classification performance in comparison to the original version and various extended versions of recent state-of-art studies. Considering the overall classification performances, space forests with the original feature space have the lowest accuracy results at all training set levels and this is an indicator that the space forests having original feature space tend to development.

As well as the classification success of enhanced space forests, execution time analysis is also appraised in terms of testing and training times. More training time is necessitated for the enhanced space forests in proportion to the original ones owing to covering more features and directly proportional to the search time of the features, vice versa. The complexity of base learners which drops a hint about the testing time is proportional to the number of nodes in a tree. Hence, the most complex base learners are constituted by random forest algorithm due to having the biggest trees. It is substantial to emphasize that the enhanced space forests are entailed less testing time because of having smaller trees compared to the original space forests.

It is considerable to compare experiment results with the state-of-art studies [2, 4] on extended space forests to demonstrate the contribution of the proposed technique. Authors employ 36 datasets from the UCI repository [8] and all of them are common with ours. The classification success of randomly enhanced space forests which is proposed by [4] is consistent with our experiment results at ts50 in terms of ensemble algorithms. The classification performance of ensemble algorithms is ordered as: RF>RS>BG.

Furthermore, the classification success of the proposed technique for bagging algorithm (86.1%) outperforms experiment results of their study (85.3%) with approximately 1% improvement. When it is considered on the random forest algorithm, the proposed technique turns out well with the accuracy rate of 86.5% whereas the study of Amasyalı and Ersoy [4] provides accuracy rate with 85.9%. Though the random subspace algorithm of both studies presents close experiment results, the proposed technique of ours boosts the classification performance of whole system compared to the study [4].

Another work [2] also strives to demonstrate the superior classification performance of enhanced space forest by using gain ratio technique for only random forest ensemble algorithm. They make use of ten different datasets from UCI repository and succeed 86.4% classification performance. Finally, the study is concluded that the great contribution of enhanced space forests to the classification success cannot be ignored. The slightest difference between ours (86.5%) and theirs (86.4%) can be arisen from the number of features, the number of datasets, differences in experimental settings and the ambiguity of the training set levels. Thus, it is noteworthy to specify that the combination of the proposed technique and random forest algorithm (RF_{ACO + RD}) predominantly surpasses state-of-art studies.

To sum up, the enhanced space forest approach advances the classification success of system compared to the original versions. over and above, it is observed that the enhancement of extended space forest with classifier ensembles using ant colony optimization as a feature selection technique exhibits better classification performance in proportion to the other enhanced space forest techniques. In future, we plan to carry out different base learners to the classification problems.

References

- [1] Adnan M., Islam M., and Kwan P., "Extended Space Decision Tree," in *Proceedings of Machine Learning and Cybernetics Conference*, Lanzhou, pp. 219-230, 2014.
- [2] Adnan M. and Islam M., "Comprehensive Method for Attribute Space Extension for Random Forest," in *Proceedings of 17th International Conference on Computer and Information Technology*, Dhaka, pp. 25-29, 2014.
- [3] Ahmed A. and Brown G., "Random Projection Random Discretization Ensembles-Ensembles of Linear Multivariate Decision Trees," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1225-1239, 2014.
- [4] Amasyalı M. and Ersoy O., "Classifier Ensembles with the Extended Space Forest," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 549-562, 2014.

- [5] Breiman L., Friedman J., Olshen R., and Stone C., *Classification and Regression Trees*, Chapman and Hall, 1984.
- [6] Breiman L., "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.
- [7] Breiman L., "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [8] Frank A. and Asuncion A., UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/index.php>, Last Visited, 2010.
- [9] Goodarzi M., Freitas M., and Jensen R., "Ant Colony Optimization As A Feature Selection Method in the QSAR Modeling of Anti-HIV-1 Activities of 3-(3,5-Dimethylbenzyl)Uracil Derivatives Using MLR, PLS and SVM Regressions," *Chemometrics and Intelligent Laboratory Systems*, vol. 98, no. 2, pp. 123-129, 2009.
- [10] Gopika D. and Azhagusundari B., "An Analysis on Ensemble Methods in Classification Tasks," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 7, pp. 7423-7427, 2014.
- [11] Ho T., "The Random Subspace Method for Constructing Decision Forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998.
- [12] Kotsiantis S., "Combining Bagging, Boosting, Rotation Forest and Random Subspace Methods," *Artificial Intelligence Review*, vol. 35, no. 3, pp. 223-240, 2011.
- [13] Koutanaei F., Sajedi H., and Khanbabaei M., "A Hybrid Data Mining Model of Feature Selection Algorithms and Ensemble Learning Classifiers for Credit Scoring," *Journal Retailing and Consumer Services*, vol. 27, pp. 11-23, 2015.
- [14] Kuncheva L. and Whitaker C., "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," *Journal Machine Learning*, vol. 51, no. 2, pp. 181-207, 2003.
- [15] Kuncheva L., Rodríguez J., Plumpton C., Linden D., and Johnston S., "Random Subspace Ensembles for Fmri Classification," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 531-542, 2010.
- [16] Liua Y. and Yaob X., "Ensemble Learning Via Negative Correlation," *Neural Networks*, vol. 12, no. 10, pp. 1399-1404, 1999.
- [17] Liu Y., Yu X., Huang J., and An A., "Combining Integrated Sampling with SVM Ensembles for Learning from Imbalanced Datasets," *Information Processing and Management*, vol. 47, no. 4, pp. 617-631, 2011.
- [18] Martin B., Jirka M., and Franky, "Multi-Label Text Classification via Ensemble Techniques," *International Journal of Computer and Communication Engineering*, vol. 1, no. 1, pp. 62-71, 2012.
- [19] Nanni L. and Lumini A., "Random Subspace for an Improved Biohashing for Face Authentication," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 295-300, 2008.
- [20] Onan A., Korukoglu S., and Bulut H., "Ensemble of Keyword Extraction Methods and Classifiers in Text Classification," *Expert Systems with Application*, vol. 57, pp. 232-247, 2016.
- [21] Panov P. and Dzeroski S., "Combining Bagging and Random Subspaces to Create Better Ensembles," in *Proceedings of International Conference on Intelligent Data Analysis*, Ljubljana, pp. 118-129, 2007.
- [22] Pedrajas N. and Boyer D., "Boosting Random Subspace Method," *Neural Networks*, vol. 21, no. 9, pp. 1344-1362, 2008.
- [23] Peralta B. and Soto A., "Embedded Local Feature Selection within Mixture of Experts," *Information Sciences*, vol. 269, pp. 176-187, 2014.
- [24] Polikar R., "Ensemble Based Systems in Decision Making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, 2006.
- [25] Rahman A. and Verma B., "Effect of Ensemble Classifier Composition on Offline Cursive Character Recognition," *Information Processing and Management*, vol. 49, no. 4, pp. 852-864, 2013.
- [26] Ren Y., Zhang L., and Suganthan P., "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 41-53, 2016.
- [27] Rokach L., "Taxonomy for Characterizing Ensemble Methods in Classification Tasks: A Review and Annotated Bibliography," *Computational Statistics and Data Analysis*, vol. 53, no. 12, pp. 4046-4072, 2009.
- [28] Rokach L., "Ensemble Based Classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1-39, 2010.
- [29] Rooney N., Wang H., and Taylor P., "An Investigation into The Application of Ensemble Learning for Entailment Classification," *Information Processing and Management*, vol. 50, no. 1, pp. 87-103, 2014.
- [30] Singh V. and Pradhan M., "Advanced Methodologies Employed in Ensemble of Classifiers: A Survey," *International Journal of Science and Research*, vol. 3, no. 12, pp. 591-595, 2014.
- [31] Skurichina M. and Duin R., "Bagging, Boosting and the Random Subspace Method for Linear Classifiers," *Pattern Analysis and Applications*, vol. 5, pp. 121-135, 2002.

- [32] Wang G., Zhang Z., Sun J., Yang S., and Larson C., "POS-RS: A Random Subspace Method for Sentiment Classification Based on Part-of-Speech Analysis," *Information Processing and Management*, vol. 51, no. 4, pp. 458-479, 2015.
- [33] Wang X. and Tang X., "Random Sampling for Subspace Face Recognition," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 91-104, 2006.
- [34] Xia R., Zong C., and Li S., "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification," *Information Sciences*, vol. 181, no. 6, pp. 1138-1152, 2011.
- [35] Zaghdoudi R. and Seridi H., "Combination of Multiple Classifiers for Off-Line Handwritten Arabic Word Recognition," *The International Arab Journal of Information Technology*, vol. 14, no. 5, pp. 713-720, 2017.



Zeynep Kilimci received the M.Sc. degree in computer and information science from the Dogus University (DU), Istanbul, Turkey, in 2013. She was a software engineer in datawarehouse department at DenizBank Istanbul, Turkey, between 2009 and 2011. She is a research assistant at DU and Ph.D. student at Computer Engineering Department of Kocaeli University, Kocaeli, Turkey. Her research interests are machine learning, ensemble learning, deep learning, text mining, and artificial intelligence.



Sevinç Omurca was born in Isparta, Turkey in 1979. She received the M.Sc. degree in electronics and communications engineering and Ph.D. degree in computer engineering from the University of Kocaeli, Turkey, in 2004 and 2009, respectively. In 2001, she joined the Department of Computer Engineering, University of Kocaeli as a Research assistant; she became an Assistant Professor in 2010 and an Associate Professor in 2016. Her current research interests include machine learning, natural language processing, and artificial intelligence.