

RF Ensemble Novelty Based on Optimized & Backpropagated NNs

Hasan Koyuncu and Rahime Ceylan

Abstract—This paper presents a classifier model based on Rotation Forest (RF) ensemble structure for biomedical data classification. Classifiers based on RF are generally implemented by using Decision Trees. In this study, optimized Neural Network (NN) is preferred as being the base classifier in RF so as to achieve higher classification performance. Two optimization techniques, Artificial Bee Colony Optimization (ABC) and Particle Swarm Optimization (PSO), are utilized to improve the performance of NN for escaping from local minima. In this way, PSO-NN and ABC-NN based RF structures are designed, and they are called as RF (PSO-NN) and RF (ABC-NN), respectively. In these classifiers, initial weights of NNs are found by using PSO or ABC algorithms. The implemented classifiers based on RF are applied to biomedical datasets (Wisconsin Breast Cancer and Pima Indian Diabetes) that are taken from UCI Machine Learning Repository. Furthermore, fourteen different ensemble structures are generated using these algorithms to prove the superiority of the proposed method. When the results are examined using several performance metrics, it is seen that RF (ABC-NN) classifier achieves to more reliable and better results than other classifiers.

Index Terms—Rotation forest, particle swarm optimization, artificial bee colony optimization, neural networks, biomedical data classification.

I. INTRODUCTION

Biomedical data classification is the most important part of Computer Aided Diagnosis systems (CADs), and it determines the decision in diagnosis & treatment. Because of this reason, classification methods and their combinations play an important role in machine learning. In the literature, there are many studies about classifiers which can detect different classes in biomedical data as better. One of these structures is ensemble classifiers which enable the use of several classifiers that can be the same kind or different kinds.

Bagging, Boosting, AdaBoost and Rotation Forest (RF) are well-known ensemble classifiers. As being one of these classifiers, RF is designed on feature extraction method, and it was firstly proposed by Rodriguez *et al.* [1]. Several studies about RF were made which demonstrate its excellence to other ensemble classifier methods in classification problems. Liu and Huang [2] implemented a cancer classification system by RF. They showed that RF with Independent Component Analysis (ICA) obtains higher accuracy than RF with

Principle Component Analysis (PCA). Kotsiantis [3] compared several ensemble classifier methods with each other on benchmark classification problems. His study represented that Boosting and RF algorithms are stronger than other ensemble classifier methods. Han *et al.* [4] presented Radial Basis Neural Network (RBNN) based RF for remote sensing image classification.

RF technique is often used with Decision Tree (DT), however it can be used with different classifiers like Neural Network (NN), Support Vector Machine (SVM), etc. For this purpose, different methods performing the task of classification can be used in the base classifier part of RF. In the literature, there exist lots of simple & hybrid classifiers used on Wisconsin Breast Cancer (WBC) and Pima Indian Diabetes (PID) datasets. Karabatak and Ince [5] used Association Rules (AR) with NN structure on WBC. Wang [6] examined different classifiers using WEKA for pattern classification process on WBC. In [6]; Naïve Bayes, BP Neural Network (BPNN), J48 Decision Tree (C4.5), Support Vector Machine, JRip (RIPPER rule learner) methods were used for finding the optimum classifier. Chang and Liou [7] compared Logistic Regression (LR), Decision Tree (DT), Genetic Algorithm (GA) and Artificial Neural Network (ANN) on WBC. In another study, Rule Extraction was made by Kamruzzaman and Islam [8] by using ANN algorithm (REANN). Jeatrakul and Wong [9] presented a comparison between five techniques on PID. A novel ANN structure was designed by Pradhan and Sahu [10]. In their technique, GA and Backpropagation algorithms were used for training of the novel NN algorithm. They compared this technique with Functional Link ANN (FLANN) and with several classification systems on PID. Hosseinpour *et al.* [11] used WEKA for comparing different classifier systems on PID, and they obtained the best classification results with the use of Bagging.

However, Backpropagation method which is used for training of NN can trap in a local minimum of error function. Herein, this algorithm includes local minimization methods that have no mechanism allowing them to escape from the influence of a local minimum [12]. At this point, optimization algorithms are added to NN for prevention of this problem.

PSO and ABC are generally used in training of NNs to adjust the initial parameters of the classifier system. At this juncture, the aim is to optimize the initial parameters of classifier for achieving the best classification results. Korürek and Doğan [13] utilized PSO to improve the classification accuracy of RBFNN in ECG beat classification problem. Melgani and Bazi [14] used PSO and Support Vector Machine (SVM) for classification of Electrocardiogram (ECG) signals. Yusong and Naval [15] used Multiobjective

Manuscript received May 22, 2017; revised August 5, 2017. This work was supported by the Coordinatorship of Selçuk University's Scientific Research Projects.

The authors are with the Electrical & Electronics Engineering Department, Selçuk University, Konya, Turkey (e-mail: hasankoyuncu@selcuk.edu.tr, rpektatli@selcuk.edu.tr).

PSO for adjusting the network architecture, weights and biases. Zhao *et al.* [16] trained NN ensembles with different stochastic approaches including PSO and ABC algorithms. In summary, there are lots of studies in which PSO or ABC algorithms are utilized to optimize the parameters of classifiers in the literature.

This study focuses on RF structure in which optimized NNs are formed as base classifiers for biomedical data classification. In this way, a new design based on RF ensemble is proposed, and novel RF ensemble structures based on optimized NNs (RF (PSO-NN), RF (ABC-NN)) are designed. Herein, PSO and ABC techniques are adopted to adjust the first weight values in NN for avoiding from local minimum by enhancing performance, accuracy and efficiency of NN. The performance evaluation of proposed structures is realized on two medical datasets (WBC and PID). These structures are shown in Fig. 1.

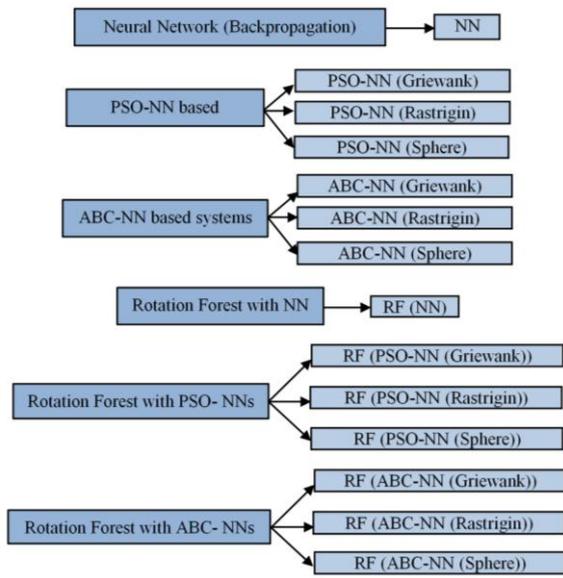


Fig. 1. Block representation for fourteen different classifiers.

In Section II, a detailed information about PSO, ABC, RF and the proposed classifier are given. In Section III, the used datasets and application results are explained by graphics and tables. Herein, results are investigated using several performance evaluation metrics. In Section IV, these application results are discussed by evaluating the performance of proposed classifier vs. other similar studies in the literature. Finally, in Section V, concluding remark is given.

II. METHODS

In this study, actualization of 4 aims is targeted: (1) To emphasize the importance of optimization for base classifiers. (2) To show the importance of function selection and its effects on optimization techniques and base classifier. (3) To examine the effects of complex base classifiers to the results of RF. For seeing it obviously, we compare RF (complex classifiers) with RF (NN). (4) To find the optimal classifier for biomedical data classification from inside of the used classifier models.

For realizing these purposes, fourteen different structures (Fig. 1) are implemented by using NN, PSO, ABC and RF

techniques. Classification performances of structures are compared on medical datasets.

In this section, the used methods are explained to understand the theoretical basis for the proposed classifier. Then, the design based on RF & optimization algorithms & NNs is explained in a comprehensive manner.

A. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a stochastic optimization method simulating the social behaviours of birds [17]. PSO includes remarkable advantages, and one of the most important ones is to have a few adjustable parameters. Another advantage is to own *social* and *individual* parameters which exactly reflect the swarm behaviour. Because of these reasons, PSO is more practical, more stable and faster than many others.

In PSO, every bird is called as a particle (*i*) searching space for finding food. In PSO, every loop means for birds to get closer to the food source. For this purpose, PSO uses Eq.(1) and Eq.(2) [17].

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 (X_{pbest(i)}(t) - X_i(t)) + c_2 r_2 (X_{gbest}(t) - X_i(t)) \quad (1)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (2)$$

In Eq.(1) and Eq.(2); X_i and V_i are the position and velocity of *i*.th particle, respectively. X_{gbest} is the global best position value obtained through all particles, and this parameter exposes the social behaviour of swarm. $X_{pbest(i)}$ is the individual best position value of *i*.th particle, and this parameter indicates the individual behaviour of particles in swarm. c_1 & c_2 are acceleration constants adjusted according to Eq.(3), and they are chosen in the interval of [0, 4]. The selection of c_1 & c_2 at low values can provide sensitive movements to particles. However, too low values may lead to undesired consequences, and the particle's arrival to optimum value can take long time. The selection of c_1 & c_2 at high values can provide particles to move faster to optimum value. But too high values can lead the particle to disregard the optimum value. In other words, acceleration constants play an important role on particles for obtainment of global points. In our study, we fixed c_1 & c_2 parameters as $c_1=c_2=2$, since this notification was preferred by virtue of exposing good performance in many applications on PSO.

$$c_1 + c_2 \leq 4 \quad (3)$$

r_1 & r_2 parameters are randomly generated values within the range [0, 1]. ω is inertia weight which has a similar feature with the momentum coefficient in NN, and it's changed in the interval of [0.5, 1]. In our study, we adjust inertia weight as equal to 0.9 at the beginning of iteration, and its value is 0.4 at the end. Thus, a parabolic descent movement is utilized for the change of inertia weight, and it is very efficient on many applications in the literature. According to Eq.(1) and Eq.(2), position of particle is updated after the update of velocity.

B. Artificial Bee Colony Optimization

Artificial Bee Colony Optimization (ABC) is an effective optimization method proposed by Karaboga [18]. ABC

simulates the foraging behaviour of honey bee swarms. In ABC, bees are divided into three groups: (1) Employed bees, (2) Onlooker bees and (3) Scout bees. Employed bees are the ones which find food source, carry food to hive, and do waggle dance in dancing area. Onlooker bees are the ones waiting in dancing area for choosing a food source based on a probability choice. Scout bees are the ones which start a random search for finding food source.

In ABC algorithm, the first half of colony consists of employed bees, and the second half constitutes the onlookers. The number of employed bees is equal to the number of food sources. The employed bee whose food source is exhausted becomes a scout bee [18].

ABC is formed by four sections which constitute the frame of structure: (1) Initialization section, (2) Employed bee section, (3) Onlooker bee section, and (4) Scout bee section. Also these sections are connected to each other as being nested. The pseudocode of ABC algorithm is as follows:

1) Initialization Section

Randomly generate the food sources (x_m) according to Eq. (4). l_i is the lower and u_i is the upper limit of space.

$$x_m = l_i + rand(0,1) * (u_i - l_i) \quad (4)$$

2) Employed Bee Section

Generate the neighbor food sources (v_{mi}) according to Eq. (5) for all food sources. v_{mi} vectors are going to be compared with x_m vectors.

$$v_{mi} = x_{mi} + \phi_{mi} (x_{mi} - x_{ki}) \quad (5)$$

In Eq.(5); x_{ki} is a randomly chosen food source, and ϕ is a random number within the range [-1, 1]. i is a randomly chosen index within the range [1, dimension number]. Herein, k cannot own to the same value with parameter m .

Calculate the fitness function according to Eq. (6)

$$fit_m(x_m) = \begin{cases} 1 / (1 + f_m(x_m)), & f_m(x_m) \geq 0 \\ 1 + |f_m(x_m)|, & x < 0 \end{cases} \quad (6)$$

For $m=1$: population size

If ($fit_m(v_{mi})$ is better than $fit_m(x_m)$) then $x_m=v_{mi}$,

Else $x_m=x_m$ and $count(m)=count(m)+1$

End

3) Onlooker Bee Section

Calculate the profitabilities for all food sources (x_m vectors) according to Eq. (7)

$$P_m = \frac{fit_m(x_m)}{\sum_{m=1}^{SN} fit_m(x_m)} \quad (7)$$

$z=0$; $m=1$;

While ($z < population\ size$)

If ($P_m > rand(0, 1)$)

Generate the new neighbour food source (y_{mi}) according to Eq. (5)

Calculate its fitness according to Eq. (6)

If ($fit_m(y_{mi})$ is better than $fit_m(x_m)$) then $x_m=y_{mi}$,

Else $x_m=x_m$ and $count(m)=count(m)+1$ and $z=z+1$

End

Else $m=m+1$ (if m exceeds the population size, then $m=1$)

End

End

4) Scout Bee Section

For $m=1$:population size

If ($count(m) > user\ defined\ limit$), then use Eq. (4) for reproducing x_m

End

Limit parameter usually varies within the range of [0, 100]. But Karaboga and Akay [19] suggest a notification about *limit* in their study, and this notification is expressed in Eq.(8).

$$limit = SN * D \quad (8)$$

In Eq.(8); SN is the number of employed bees (food sources number) and D is the dimension. We use this notification in our study for the parameter *limit*.

In this study, PSO and ABC are utilized for avoiding from local minimum in NN. In this manner, PSO-NN and ABC-NN structures are implemented to improve the performance of NN.

C. Rotation Forest

An ensemble classifier includes several base classifiers. Because of this reason, the use of these techniques is more convenient than the use of one classifier for achieving a good performance on several areas like signal processing, pattern recognition, image classification, etc.

Rotation Forest (RF) is an ensemble classifier method proposed by Rodriguez *et al.* [1], and it's used on various technical areas in the literature. RF ensembles tend to generate base classifiers which are more accurate than those formed by *Adaboost* and by *Random Forest*, and more diverse than those formed by *Bagging* [20]. Moreover, RF uses bootstrap method which aims to produce smaller datasets by replacing and choosing patterns from the original dataset. Thus, it obtains diversity as in *Bagging* method.

RF ensemble structure consists of base classifiers which are trained and tested with improved and diverse data. For this purpose, the first step is the change of feature vector's places and separation of the data into subsets. Then diversity is obtained by bootstrap process [20]. After this process, feature vectors are handled by Principle Component Analysis (PCA) and eigenvector coefficients are obtained [20]. Then, the improved data is formed by multiplying the original data with eigenvector coefficients obtained by PCA. Thus the data is obtained to be used in base classifier. *At this point, it must be specified that all base classifiers in ensemble are trained and tested with different data.* For this purpose, the whole process (change of feature vectors' places, separation of data into subsets, bootstrap and PCA) is done for all classifiers, individually [20]. In this way, several improved data are formed to be joined to base classifiers. Rokach presents RF algorithm as follows [20]:

1) Training phase

For $i = 1$ to T do

Split the feature set into K subsets: $F_{i,j}$ (for $j=1,\dots,K$)

For $j = 1$ to K **do**

Let $S_{i,j}$ be the data set S for the features in $F_{i,j}$

Eliminate from $S_{i,j}$ a random subset of classes

Select a bootstrap sample from $S_{i,j}$ of size 75% of the number of objects in $S_{i,j}$.

Denote the new set by $S_{i,j}'$

Apply PCA on $S_{i,j}'$ to obtain the coefficients in a matrix $C_{i,j}$

End For

Arrange the $C_{i,j}$ for $j= 1$ to K in a rotation matrix R_i as in Eq.(9)

$$R_i = \begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(2)}, \dots, a_{i,1}^{(M_i)} & [0] & \dots & [0] \\ [0] & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \dots, a_{i,2}^{(M_i)} & \dots & [0] \\ \dots & \dots & \dots & \dots \\ [0] & [0] & \dots & a_{i,k}^{(1)}, a_{i,k}^{(2)}, \dots, a_{i,k}^{(M_i)} \end{bmatrix} \quad (9)$$

Construct R_i^a by rearranging the columns of R_i so as to match the order of features in F

End For

2) Output: Build classifier M_i using (SR_i^a, X) as training set.

In training and outputs phases; I is a base inducer, S is the original training set, T is the number of iterations, and K is the number of subsets.

D. Proposed Algorithms

Rotation Forest usually uses PCA for feature extraction. However ICA (Independent Component Analysis) and other feature extraction techniques can be used. In this study, PCA is used for feature extraction as in the original version of RF.

Rotation Forest is usually used with DT base classifiers. In this study, we use NN structure as being the base classifier in terms of its productivity and efficiency on pattern classification. Furthermore, we intend to improve the output of ensemble classifier by changing NN with optimization based NNs. Therefore, NN, PSO-NN (with three different fitness functions) and ABC-NN (with three different fitness functions) are used as base classifiers. Process scheme of Rotation Forest, the implemented structures and the learning algorithm for PSO-NN & ABC-NN can be seen in Fig. 2. Herein, the used NN is chosen as multilayer perceptron which has three layers.

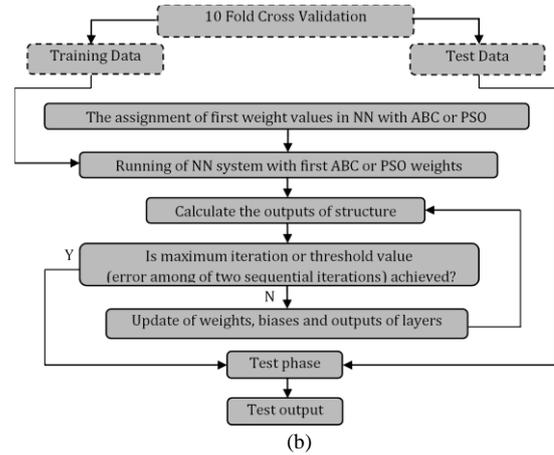
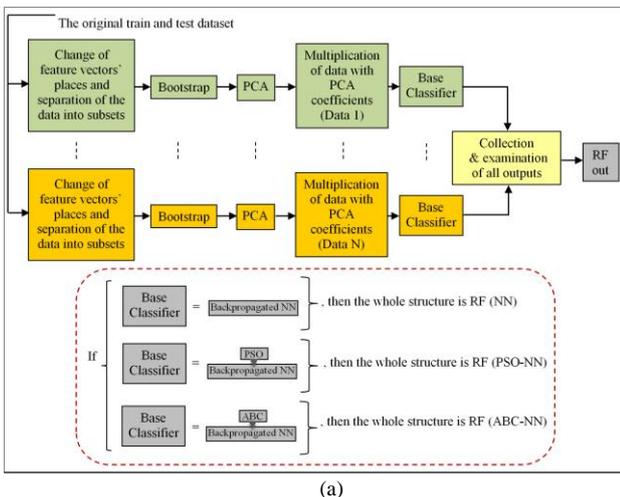


Fig. 2. (a) Process scheme of RF and implemented structures; (b) Learning algorithm for PSO-NN and ABC-NN structures.

III. EXPERIMENTS AND DISCUSSIONS

In this study, a novel RF ensemble based on optimized NNs is presented. Six different classification techniques are evaluated, and fourteen different classifier structures are compared on WBC and PID datasets via 10-fold cross validation test method. These classification techniques are as follows: NN, PSO-NN, ABC-NN, RF (NN), RF (PSO-NN) and RF (ABC-NN).

Fourteen different classifier structures are obtained by using different objective functions in aforesaid techniques (Fig. 1). In this study, we use eight performance metrics to evaluate the effectiveness of fourteen classifiers: Sensitivity, Specificity, Precision, F-Measure, Gmean, Confusion Matrix, Accuracy Rate of Classification (ARC or CA) and area under ROC curve (AUC).

All these metrics give a statistical evaluation for the performance of system by using true positive, true negative, false positive and false negative [21]. For medical pattern classification; true positive (TP) symbolizes the number of sick patterns properly defined as sick. False positive (FP) indicates the number of normal patterns improperly defined as sick. False negative (FN) symbolizes the number of sick patterns improperly defined as healthy. True negative (TN) indicates the number of normal patterns properly defined as healthy [21].

Sensitivity (recall), Specificity, Precision, F-measure, G-mean [22], [23] and CA are presented in Eq. (10-15) [21].

$$\text{Sensitivity} = TP / (TP + FN) \quad (10)$$

$$\text{Specificity} = TN / (TN + FP) \quad (11)$$

$$\text{Precision} = TP / (TP + FP) \quad (12)$$

$$F - \text{measure} = 2 \times [(\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})] \quad (13)$$

$$G\text{mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \quad (14)$$

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

The aims of study were specified in Section II. In this section, proposed classifier structures are applied to WBC and PID datasets for actualizing these aims.

A. Parameter Adjustment

The targets of proposed structures are specified as “0” and “1”. Because of this reason, logarithmic sigmoid function is used in NN. Furthermore, minimum and maximum position boundaries are defined as “0” and “1” in PSO and ABC algorithms. In addition to position limitations, PSO has velocity limitation too. The velocity boundary is specified within the range [0, 1], because the outputs of structures are “0” and “1”.

PSO and ABC techniques are utilized to optimize the first weights in PSO-NN, ABC-NN, RF (PSO-NN) and RF (ABC-NN) structures. After that, Backpropagation is performed for update of NN. Furthermore, fourteen classifiers are formed by using three different fitness functions (Griewank, Rastrigin and Sphere) for examining the effects of fitness functions on optimization and classification.

Learning rate, hidden node number and maximum iteration number (kept constant at 500) are the basic parameters in NN, and their optimum values are specified experimentally. These parameters are investigated as follows:

- Learning rate is changed as 0.1, 0.3, 0.5, 0.7, 0.85, 0.9
- Number of hidden nodes is evaluated for 2, 4, 6, 8, 10, 12
- Number of base classifier is changed as 3, 6 and 9.

Optimum parameter values, at which the best performances are obtained, can be seen in Table I on WBC and PID datasets. At this point, optimum operation conditions settled in Table I is obtained by using CA metric, since it’s the most preferred criterion in pattern classification area.

TABLE I: OPTIMUM PARAMETER VALUES FOUND ON WBC AND PID DATASETS FOR ALL CLASSIFICATION STRUCTURES

| Structure | On WBC Dataset | | | On PID Dataset | | |
|------------------------|----------------|------|----------|----------------|------|-----|
| | HN | lr | BCN | HN | lr | BCN |
| Neural Networks | 10 | 0.85 | --- | 2 | 0.85 | --- |
| PSO-NN(Griewank) | 10 | 0.85 | --- | 2 | 0.85 | --- |
| PSO-NN(Rastrigin) | 10 | 0.5 | --- | 2 | 0.7 | --- |
| PSO-NN(Sphere) | 10 | 0.85 | --- | 2 | 0.85 | --- |
| ABC-NN(Griewank) | 6 | 0.9 | --- | 4 | 0.3 | --- |
| ABC-NN(Rastrigin) | 6 | 0.7 | --- | 4 | 0.3 | --- |
| ABC-NN(Sphere) | 6 | 0.9 | --- | 4 | 0.1 | --- |
| RF (NN) | 10 | 0.85 | 6 | 4 | 0.1 | 6 |
| RF (PSO-NN(Griewank)) | 10 | 0.85 | 3 | 2 | 0.85 | 9 |
| RF (PSO-NN(Rastrigin)) | 10 | 0.85 | 3 | 2 | 0.85 | 3 |
| RF (PSO-NN(Sphere)) | 10 | 0.85 | 9 | 2 | 0.85 | 9 |
| RF (ABC-NN(Griewank)) | 6 | 0.85 | 3 | 4 | 0.1 | 9 |
| | | | and 9 | | | |
| RF (ABC-NN(Rastrigin)) | 6 | 0.85 | 9 | 4 | 0.1 | 6 |
| RF (ABC-NN(Sphere)) | 6 | 0.85 | 6 | 4 | 0.1 | 9 |
| | | | and 9 | | | |

(HN: Hidden Node Number, lr: Learning Rate, BCN: Base Classifier Number)

In the rest of this paper, comparisons are made by fixing the optimum operation conditions in Table I. Other tables and figures show the best results or the obtainment of optimum operation conditions.

Base Classifier Number (BCN) of RF based methods are presented as being visual. On account of being the second popular metric, AUC based comparisons are presented by using graphics to reveal the priorities of techniques with a closer look. Also different statistical metrics (except CA and AUC) are used for evaluating the performance in a comprehensive manner.

B. Results on WBC Dataset

WBC dataset has missing values. For preventing its negative effects on consequences, the number of patterns in dataset was reduced from 699 to 683 which consist of 444 benign and 239 malignant patterns. WBC includes 9 features in each of 683 patterns. These features are clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses.

This study focuses to evaluate the effects of using different fitness functions and to determine the optimum classifier structures for biomedical data classification. However, in addition to these assessments, classifier structures based on RF are evaluated according to BCN. For proposed classifiers, evaluation graphics on WBC are presented in Fig. 3.

As seen in Fig. 3, the best CA rates are obtained by RF (PSO-NN (Sphere)) and by RF (ABC-NN) with Sphere and Rastrigin functions, when BCN is chosen as “9”.

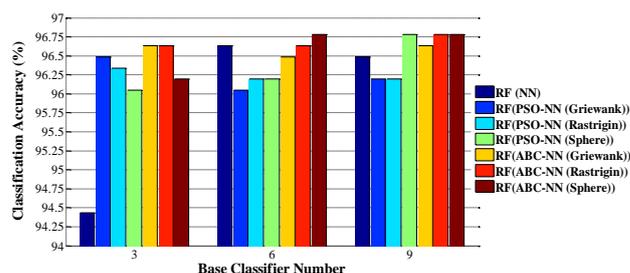


Fig. 3. Evaluation graphic for classifier structures based on RF (WBC dataset).

The whole comparison of structures can be seen in Table II on WBC dataset. Besides, AUC values given in Table II are computed in Fig. 4.

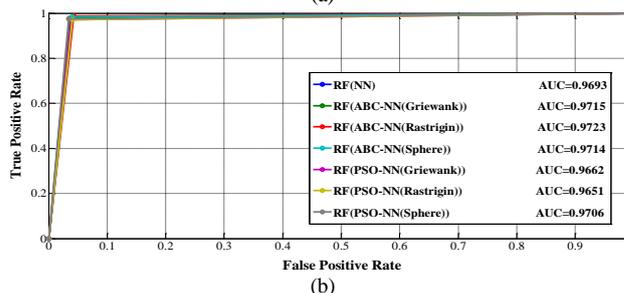
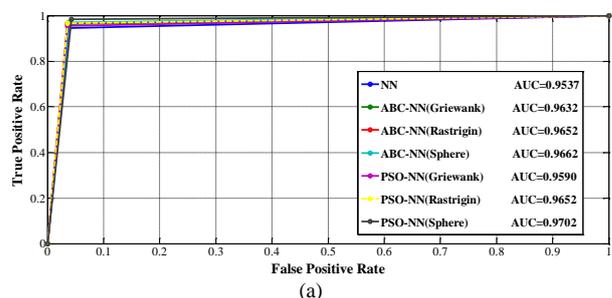


Fig. 4. (a) ROC curves for classifier structures without RF (WBC dataset); (b) ROC curves for classifier structures with RF (WBC dataset).

According to Table II, it can be clearly seen that the best rates are generally obtained by proposed classifiers based on RF. Also CA rates of RF (ABC-NN) structures are generally better than the other choices. Herein, the unique exception is RF(PSO-NN(Sphere)) owing to the best CA rate with RF(ABC-NN(Griewank)). On the other hand, RF(ABC-NN(Rastrigin)) owns to the best AUC score. RF(ABC-NN(Griewank)) achieves the best f-measure and CA scores on trials with WBC. RF(PSO-NN(Sphere)) achieves to the best specificity, precision and classification accuracy. Furthermore, it has the fourth AUC score (97.06%)

which is not much lower than the score of RF(ABC-NN(Rastrigin)). As seen in Fig. 4, the best AUC score (97.23%) is obtained by RF (ABC-NN(Rastrigin)), and the second AUC score (97.15%) is achieved by RF (ABC-NN(Griewank)).

If we make a comparison based on statistics (according to figures and tables), we can infer that RF(PSO-NN(Sphere)) is more suitable to use than others on WBC dataset in terms of obtaining better scores on three of seven metrics. Also its CA rate is the best beside of the obtained close AUC score to the best AUC rate.

TABLE II: COMPARISON OF RESULTS OBTAINED ON WBC DATASET

| Classification Structures | Sensitivity | Specificity | Precision | f-measure | G-mean | CA (%) | AUC |
|---------------------------|-------------|-------------|-----------|-----------|--------|--------|--------|
| Neural Networks | 0.9372 | 0.9595 | 0.9256 | 0.9314 | 0.9483 | 95.17 | 0.9537 |
| PSO-NN(Griewank) | 0.9623 | 0.9595 | 0.9274 | 0.9445 | 0.9609 | 96.05 | 0.9590 |
| PSO-NN(Rastrigin) | 0.9791 | 0.9572 | 0.9249 | 0.9512 | 0.9681 | 96.49 | 0.9652 |
| PSO-NN(Sphere) | 0.9791 | 0.9595 | 0.9286 | 0.9532 | 0.9693 | 96.63 | 0.9702 |
| ABC-NN(Griewank) | 0.9623 | 0.9640 | 0.9350 | 0.9485 | 0.9632 | 96.34 | 0.9632 |
| ABC-NN(Rastrigin) | 0.9707 | 0.9617 | 0.9317 | 0.9508 | 0.9662 | 96.49 | 0.9652 |
| ABC-NN(Sphere) | 0.9875 | 0.9527 | 0.9183 | 0.9516 | 0.9699 | 96.49 | 0.9662 |
| RF (NN) | 0.9791 | 0.9595 | 0.9286 | 0.9532 | 0.9693 | 96.63 | 0.9693 |
| RF (PSO-NN (Griewank)) | 0.9665 | 0.9640 | 0.9352 | 0.9506 | 0.9653 | 96.49 | 0.9662 |
| RF (PSO-NN(Rastrigin)) | 0.9791 | 0.9550 | 0.9213 | 0.9493 | 0.9670 | 96.34 | 0.9651 |
| RF (PSO-NN(Sphere)) | 0.9749 | 0.9662 | 0.9395 | 0.9569 | 0.9705 | 96.93 | 0.9706 |
| RF (ABC-NN (Griewank)) | 0.9791 | 0.9640 | 0.9360 | 0.9571 | 0.9715 | 96.93 | 0.9715 |
| RF (ABC-NN(Rastrigin)) | 0.9833 | 0.9595 | 0.9289 | 0.9553 | 0.9713 | 96.78 | 0.9723 |
| RF (ABC-NN(Sphere)) | 0.9875 | 0.9572 | 0.9255 | 0.9555 | 0.9722 | 96.78 | 0.9714 |

TABLE III: COMPARISON OF RESULTS OBTAINED ON PID DATASET

| Classification Structures | Sensitivity | Specificity | Precision | f-measure | G-mean | CA (%) | AUC |
|---------------------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|
| Neural Networks | 0.6605 | 0.798 | 0.6367 | 0.6484 | 0.7260 | 75 | 0.7327 |
| PSO-NN(Griewank) | 0.6828 | 0.816 | 0.6655 | 0.6740 | 0.7464 | 76.95 | 0.7494 |
| PSO-NN(Rastrigin) | 0.6567 | 0.826 | 0.6692 | 0.6629 | 0.7365 | 76.69 | 0.7414 |
| PSO-NN(Sphere) | 0.6828 | 0.816 | 0.6655 | 0.6740 | 0.7464 | 76.95 | 0.7494 |
| ABC-NN(Griewank) | 0.6605 | 0.822 | 0.6654 | 0.6629 | 0.7368 | 76.56 | 0.7412 |
| ABC-NN(Rastrigin) | 0.6828 | 0.816 | 0.6655 | 0.6740 | 0.7464 | 76.95 | 0.7468 |
| ABC-NN(Sphere) | 0.6157 | 0.842 | 0.6762 | 0.6446 | 0.7200 | 76.30 | 0.7288 |
| RF (NN) | 0.6605 | 0.84 | 0.6887 | 0.6743 | 0.7449 | 77.73 | 0.7520 |
| RF (PSO-NN (Griewank)) | 0.6754 | 0.808 | 0.6582 | 0.6667 | 0.7387 | 76.17 | 0.7417 |
| RF (PSO-NN(Rastrigin)) | 0.6754 | 0.81 | 0.6558 | 0.6655 | 0.7397 | 76.30 | 0.7427 |
| RF (PSO-NN(Sphere)) | 0.6828 | 0.808 | 0.6559 | 0.6691 | 0.7428 | 76.43 | 0.7454 |
| RF (ABC-NN (Griewank)) | 0.6455 | 0.848 | 0.6948 | 0.6692 | 0.7399 | 77.73 | 0.7476 |
| RF (ABC-NN(Rastrigin)) | 0.6493 | 0.84 | 0.6850 | 0.6667 | 0.7385 | 77.34 | 0.7446 |
| RF (ABC-NN(Sphere)) | 0.6567 | 0.848 | 0.6984 | 0.6769 | 0.7463 | 78.13 | 0.7498 |

C. Results on WBC Dataset

PID dataset includes 8 features in each of 768 patterns. There are 500 benign and 268 malignant patterns in dataset. In particular, all patients are females at least 21 years old of Pima Indian heritage. Attribute information of PID dataset are number of times pregnant, plasma glucose concentration a 2 hours in oral glucose tolerance test, diastolic blood pressure, triceps skin fold thickness, 2-hour serum insulin, body mass index, diabetes pedigree function and age.

As in WBC trials, fourteen classifiers are investigated for the obtainment of best one in PID based experiments.

Evaluation graphic can be seen in Fig. 5.

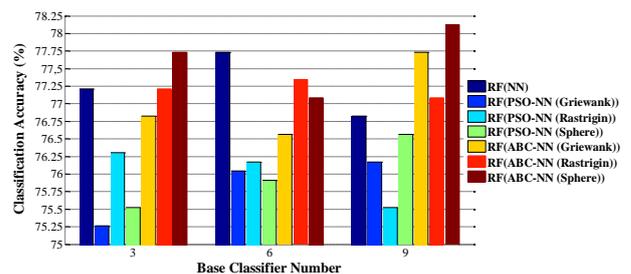


Fig. 5. Evaluation graphic for classifier structures based on RF (WBC dataset).

As seen in Fig. 5, the best CA rate is obtained by RF (PSO-NN (Sphere)) when BCN is chosen as “9”. The obtained results for all classification structures are presented in Table III.

If Table III is examined, it can be seen that the best CA (78.13%) is achieved by RF(ABC-NN(Sphere)). Also RF(ABC-NN(Sphere)) owns to the best scores on four metrics (specificity, precision, f-measure, CA). Besides, the second best AUC score is obtained by RF(ABC-NN(Sphere)) as seen in Fig. 6 and Table III.

If we make a comparison according to evaluation metrics, RF(NN) could only achieve the best score on AUC. In other words, it cannot make any progress on other metrics. Contrary to RF(NN), RF(ABC-NN(Sphere)) achieves to the best scores on four different evaluation metrics. Besides, it has the second best AUC score after RF(NN). Therefore, we can infer that RF (ABC-NN(Sphere)) structure is more convenient to use than other classification structures on PID dataset.

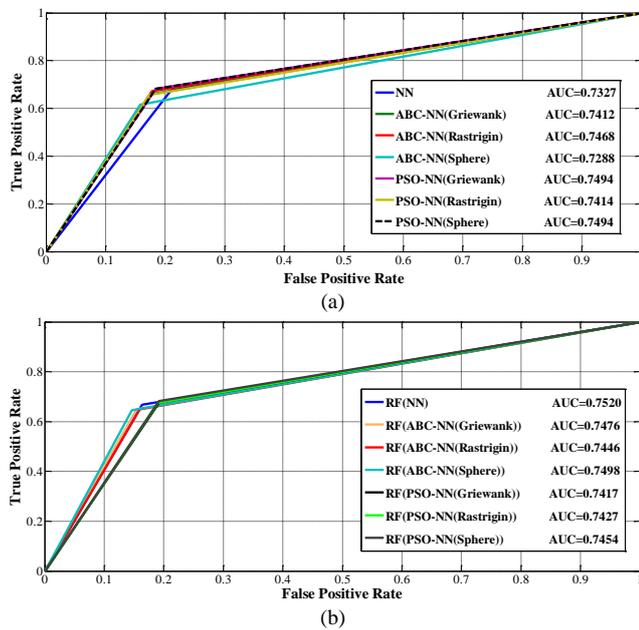


Fig. 6. (a) ROC curves for classifier structures without RF (PID dataset); (b) ROC curves for classifier structures with RF (PID dataset).

IV. DISCUSSIONS

In this study, six different classification structures (NN, PSO-NN, ABC-NN, RF (NN), RF (PSO-NN), RF (ABC-NN)) are compared and tested on WBC & PID datasets by 10-fold cross validation method. Several performance evaluation metrics are used for evaluating these structures. As a result, evaluation of the best structure is given in Table IV according to each criterion.

According to Table IV, RF(PSO-NN(Sphere)) structure achieves to the best scores on three different metrics (specificity, precision, classification accuracy) on WBC based trials. For PID dataset, RF(ABC-NN(Sphere)) structure has the best scores according to four different metrics (specificity, precision, f-measure, classification accuracy).

On WBC trials, RF(ABC-NN(Sphere)) owns to the best scores on two metrics (sensitivity and gmean), while RF(PSO-NN(Sphere)) obtains one best score (sensitivity) on PID. In other words, the use of optimized NNs based RF is

convenient, but it can change from one data to another. Furthermore, it is seen that the fitness function of PSO and ABC is “Sphere” in these (best) classification structures. So it can be inferred that *sphere* fitness function is superior to other functions for the usage in optimized NN based RF.

TABLE IV: EVALUATION OF THE BEST STRUCTURE
A) RESULTS FOR WBC DATASET

| Measurements | Structure achieving the best score | Best Score |
|--------------|---|------------|
| Sensitivity | RF(ABC-NN(Sphere))/ABC-NN(Sphere) | 98.75 |
| Specificity | RF (PSO-NN(Sphere)) | 96.62 |
| Precision | RF (PSO-NN(Sphere)) | 93.95 |
| F_measure | RF (ABC-NN(Griewank)) | 95.71 |
| Gmean | RF (ABC-NN(Sphere)) | 97.22 |
| AUC | RF (ABC-NN(Rastrigin)) | 97.23 |
| CA | RF(PSO-NN(Sphere))/RF(ABC-NN(Griewank)) | 96.93 |

B) RESULTS FOR PID DATASET

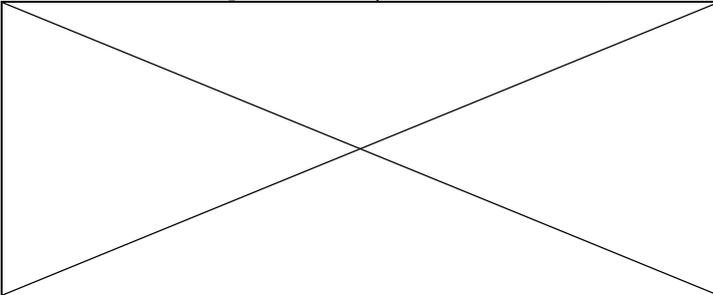
| Measurements | Structure achieving the best score | Best Score |
|--------------|---|------------|
| Sensitivity | PSO-NN(Griewank) / PSO-NN(Sphere) ABC-NN(Rastrigin)/RF(PSO-NN(Sphere)) | 68.28 |
| Specificity | RF(ABC-NN(Griewank))/RF(ABC-NN(Sphere)) | 84.80 |
| Precision | RF (ABC-NN(Sphere)) | 69.84 |
| F_measure | RF (ABC-NN(Sphere)) | 67.69 |
| Gmean | PSO-NN(Griewank) / PSO-NN(Sphere) / ABC-NN(Rastrigin) | 74.64 |
| AUC | RF (NN) | 75.20 |
| CA | RF (ABC-NN(Sphere)) | 78.13 |

In this study, the assignment of initial weights is performed by optimization algorithm in the optimized NN based RF. But, other updates are realized by backpropagation algorithm. Because of this reason, these structures (PSO-NN and ABC-NN) are cascade (not hybrid) structures.

As defined before, the most used and preferred metric is CA in pattern classification area. According to Table II and Table III, RF (NN) achieves to CA rates as 96.63% and 77.73% on WBC and PID datasets, respectively. Besides, RF (PSO-NN(Sphere)) obtains CA rates as 96.93% and 76.43%, while RF (ABC-NN (Sphere)) achieves to the accuracy rates of 96.78% and 78.13%, respectively. According to these results, it can be inferred that the use of RF (ABC-NN (Sphere)) is more reliable than RF (PSO-NN(Sphere)), since its CA rates are never below the rates of RF(NN), optimized NN or NN on all experiments. Concerning this, it can be said that the use of optimization algorithm in the assignment of initial weights takes an important role on performance. In relation to that result, the use of optimized & backpropagated NN is more convenient than the unique the use of backpropagated NN. Also it’s obvious that the use of ensemble classifier (RF (NN) and RF (ABC-NN (Sphere))) is more reliable than the use of one classifier (NN) or optimized classifiers (PSO-NN and ABC-NN). In Table V, the proposed classifiers are compared with various studies in the literature.

If Table V is examined, it can be seen that the best classification results are obtained by optimized NN based RFs on WBC & PID datasets. In other words, the use of NN can be a better choice than the use of DT based classifiers (like as J48). Besides, the use of proposed methods is more convenient than the use of Bagging or Boosting as an ensemble classifier.

TABLE V: COMPARISON OF PROPOSED TECHNIQUES WITH OTHER STUDIES

| Structures (tested on WBC) | CA (%) | Structures (tested on PID) | CA (%) |
|---|----------------------|-----------------------------|----------------------|
| Novel NN [10] | 95.02 (10xCV) | FLANN [13] | 59.80 (10xCV) |
| BPNN [9] | 95.16 (10xCV) | NN [13] | 65.10 (10xCV) |
| NN [8] | 95.20 (3xCV) | MFS1 [13] | 67.50 (10xCV) |
| AR2-NN [8] | 95.60 (3xCV) | BSS [13] | 67.70 (10xCV) |
| JRip [9] | 95.60 (10xCV) | kNN [13] | 69.70 (10xCV) |
| Decision Tree (C4.5) [9] | 95.75 (10xCV) | MFS2 [13] | 70.50 (10xCV) |
| REANN [11] | 96.28 (10xCV) | Novel ANN [13] | 73.4 (10xCV) |
| Naïve Bayes [9] | 96.48 (10xCV) | Random Forest [11] | 73.56 (10xCV) |
| NN with Multiobjective PSO [3] | 96.61 (10xCV) | GRNN [12] | 75.26 (10xCV) |
| RF (PSO-NN (Sphere)) | 96.93 (10xCV) | PNN [12] | 75.26 (10xCV) |
|  | | RBFNN [11] | 75.39 (10xCV) |
| | | MLP [11] | 75.39 (10xCV) |
| | | BPNN [12] | 76.17 (10xCV) |
| | | CMTNN [12] | 76.49 (10xCV) |
| | | REANN [11] | 76.56 (10xCV) |
| | | RBFNN [12] | 76.56 (10xCV) |
| | | AdaBoost [11] | 77.21 (10xCV) |
| | | Multiclass Classifier [11] | 77.21 (10xCV) |
| | | Bagging [11] | 77.47 (10xCV) |
| | | RF (ABC-NN (Sphere)) | 78.13 (10xCV) |

(10xCV:10-fold cross validation)

V. CONCLUSIONS

In this study, a new RF structure based on optimized NNs is proposed. This classifier model is realized as two structures designed by PSO and ABC. In these structures, PSO and ABC algorithms are used for finding the initial weights of NN. Fourteen different structures are formed by using these classifiers. Thus, not only the performances of NN, PSO-NN and ABC-NN based RFs are examined, but also the outputs of classifiers are analysed according to different fitness functions in optimization algorithms.

According to results, the effect of novelty does not sharply improve the classification performance, but slightly changes the output as being positive if a detailed assessment is realized by analysing the output on eight different metrics.

In the literature, RF (NN) is available, but there does not exist a study about optimized NNs based RF. According to the obtained results, the better results are found by classifier in which ABC is used with *Sphere* fitness function.

Consequently, it's revealed that this paper contributes to the literature on the following subjects:

- Efficient ensemble novelties performing the task of pattern classification
- A new viewpoint to RF ensemble algorithm
- Formation of an ensemble classifier with optimized & backpropagated NNs
- A detailed study on biomedical and binary pattern classification
- A comprehensive comparison of base, hybrid and ensemble classifiers

In future work, hybrid PSO-NN and hybrid ABC-NN structures will be examined that owns PSO and ABC algorithms in the update of weight & biases. In other words, backpropagation part will be changed with optimization algorithms in order to achieve better classification results.

REFERENCES

[1] J. J. Rodriguez, L. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE T. Pattern Anal.*, vol. 28, pp. 1619-1630, 2006.

[2] K. H. Liu and D. S. Huang, "Cancer classification using Rotation Forest," *Comput. Biol. Med.*, vol. 38, pp. 601-610, 2008.

[3] S. Kotsiantis, "Combining bagging, boosting, rotation forest and random subspace methods," *J. Artif. Intell. Res.*, vol. 35, pp. 223-240, 2011.

[4] M. Han, X. Zhu, and W. Yao, "Remote sensing image classification based on neural network ensemble algorithm," *Neurocomputing*, vol. 78, pp. 133-138, 2012.

[5] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Syst. Appl.*, vol. 36, pp. 3465-3469, 2000.

[6] J. Y. Wang, "Data Mining Analysis (breast-cancer data)," AI term project, Lunghwa Univ., Taoyuan County, Taiwan, 2003. Register number: D9115007.

[7] W. P. Chang and D. M. Liou, "Comparison of Three Data Mining Techniques with Genetic Algorithm in the Analysis of Breast Cancer Data," *J. Telemed. Telecare*, vol. 9, pp. 1-16, 2008.

[8] S. M. Kamruzzaman and M. D. Islam, "An algorithm to extract rules from artificial neural networks for medical diagnosis problems," *Int. J. Inf. Tech. Decis.*, vol. 12, pp. 41-59, 2006.

[9] P. Jeatrakul and K. W. Wong, "Comparing the performance of different neural networks for binary classification problems," in *Proc. The Eighth International Symposium on Natural Language Processing*, 2009, pp. 111-115.

[10] M. Pradhan and R. K. Sahu, "Predict the onset of diabetes disease using Artificial Neural Network (ANN)," *Int. J. Comput. Sci. Eng. Technol. (IJCSSET)*, vol. 2, pp. 303-311, 2011.

[11] N. Hosseinpour, S. Setayeshi, K. Ansari-asl, and M. Mosleh, "Diabetes Diagnosis by Using Computational Intelligence Algorithms," *Int. J. Adv. Res. Comput. Sci. Softw. Eng. (IJARCSSE)*, vol. 2, pp. 71-77, 2012.

[12] V. P. Plagianakos, G. D. Magoulas, and M. N. Vrahatis, "Supervised Training Using Global Search Methods," *Noncon. Optim. Its. App.*, vol. 54, pp. 421-432, 2001.

[13] M. Korürek and B. Doğan, "ECG beat classification using particle swarm optimization and radial basis function neural network," *Expert Syst. Appl.*, vol. 37, pp. 7563-7569, 2010.

[14] F. Melgani and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *IEEE T. Inf. Technol. B*, vol. 12, pp. 667-677, 2008.

[15] J. P. T. Yusiong and Jr P. C. Naval, "Training neural networks using multiobjective particle swarm optimization," *Lect. Notes Comput. Sci.*, vol. 4421, pp. 879-888, 2006.

[16] Z. S. Zhao, X. Feng, Y. Y. Lin, F. Wei, S. K. Wang, T. L. Xiao, M. Y. Cao, and Z. G. Hou, "Evolved neural network ensemble by multiple heterogeneous swarm intelligence," *Neurocomputing*, vol. 149, pp. 29-38, 2015.

[17] R. Ceylan and H. Koyuncu, "A new breakpoint in hybrid particle swarm-neural network architecture: Individual boundary adjustment," *Int. J. Inf. Tech. Decis.*, vol. 15, pp. 1313-1343, 2016.

- [18] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical Report 06, Erciyes Univ., Kayseri, Turkey, 2005.
- [19] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, pp. 108-132, 2009.
- [20] L. Rokach, *Pattern Classification Using Ensemble Methods. Series in Machine Perception and Artificial Intelligence*, Singapore: World Scientific, 2010.
- [21] D. Sharma, U. B. Yadav, and P. Sharma, "The concept of sensitivity and specificity in relation to two types of errors and its application in medical research," *J. Reliab. Stat. Stud.*, vol. 2, pp. 53-58, 2009.
- [22] C. Ferri, J. Hernandez-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recogn. Lett.*, vol. 30, pp. 27-28, 2009.
- [23] Y. Tang, Y. Q. Zhang, N. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE T. Syst. Man Cy. B*, vol. 39, pp. 281-288, 2009.

research interests are hybrid classifiers, ensemble learning, medical image analysis, pattern recognition and optimization.



Rahime Ceylan graduated from Electrical-Electronics Engineering Department of Selcuk University with B.Sc. degree in 2000 and from Electrical-Electronics Engineering Department of Selcuk University with M.Sc. degree in 2004. She completed her PhD degree at the Electrical-Electronics Engineering Department of Selcuk University. Her current research interests are artificial neural networks, fuzzy clustering, biomedical, pattern recognition and classification, digital signal processing and digital image processing



Hasan Koyuncu graduated from Electrical-Electronics Engineering Department of Selcuk University with B.Sc. degree in 2011 and from Electrical-Electronics Engineering Department of Selcuk University with M.Sc. degree in 2013. He is continuing to his PhD education in the area of medical image analysis at Selcuk University. His current