Journal of King Saud University – Computer and Information Sciences 31 (2019) 235–251

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

A Novel Rotation Forest Modality Based on Hybrid NNs: RF (ScPSO-NN)

Rahime Ceylan, Hasan Koyuncu *

Electrical & Electronics Engineering Department, Selcuk University, Konya, Turkey

ARTICLE INFO

Article history:
Received 25 July 2017
Revised 30 October 2017
Accepted 31 October 2017
Available online 4 November 2017

Keywords:
Rotation Forest
Particle Swarm Optimization
Artificial Bee Colony Optimization
Scout Particle Swarm Optimization
Hybrid classifiers

ABSTRACT

Neural Network (NN), hybrid NN methods and Rotation Forest (RF) ensemble classifier are preferred in pattern analysis owing to their ability for finding efficient solutions on different problems. NN architecture usually includes backpropagation type algorithms in which error is exposed to fluctuations. Hybrid NN methods are generally designed to improve the classification performance of NN. Scout Particle Swarm Optimization (ScPSO) is one of these optimization algorithms including the effective parts of Particle Swarm Optimization (PSO) and Artificial Bee Colony Optimization (ABC). Moreover, RF algorithm usually indicates the same performance as in hybrid NN methods, although it is comprised of Decision Tree (DT) classifiers. At this point, our paper investigates whether RF using the hybrid NNs can outperform other ensemble classifiers in binary-medical pattern classification, or not. With this intention, PSO, ABC and ScPSO are placed in NN algorithms instead of back propagation, and hybrid methods (PSO-NN, ABC-NN and ScPSO-NN) are realized. As a result, RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) architectures are obtained. Classification Accuracy (CA), Area Under Curve (AUC), Sensitivity, Specificity, F-measure, Gmean and Precision metrics are used for a statistical performance comparison, and a test based on 2-fold cross validation method was realized on five medical datasets.

1. Introduction

Breast Cancer (BC) continues to be the most common tumor-based illness in females (Li et al., 2014). For this reason, diagnosis of breast cancer must be detected during its early stages in order to prevent the patient deaths (Shieh et al., 2014). Parkinson disease (PD) is the second most pervasive-neurodegenerative illness in elderly people. According to the diagnosis at an early pre-clinical phase, progression of PD can be taken under control (Trible and Riedere, 2008). The harm originating from diabetes increases slowly during the preliminary stages of the disease, but can cause complications if early diagnosis does not occur (Castro-Rios et al., 2014). Liver-based disorders cause the mortality to increase on worldwide. Herein, alcohol consumption is fully related to liver cirrhosis and liver cancer (Rehm et al., 2013). The death rate connected to alcohol-based liver disorder can be decreased if early detection of the disease takes place. Besides, heart disease gives rise to heart attacks that result in death.

The information mentioned above presents only one important solution for saving lives. That solution is the early diagnosis of disease. Herein, Computer Aided Diagnosis systems (CADs), that reveal the disease, establish an important place in medical decision-making stage. As a result, software experts try to offer new algorithms and systems to diagnosis of various diseases.

Biomedical pattern classification has an important role in CAD systems. In biomedical pattern classification, a sufficient classification process can be carried out via hybrid architectures. There are many skillful classifier systems found in literature, but higher classification performance is generally obtained by utilizing from complex or hybrid classification structures.

Je and Je (2011) generated Gene Expression Programming (GEP) based ensemble classifiers (GEP-A and GEP-B) and tested the techniques on WBC and PID datasets. GEP-B generally preceded others on various trials that achieved 97.21% (WBC) and 78.12% (PID) classification accuracies. Kim et al. (2011) generated a weight-adjusted voting algorithm for classifier ensembles. In their study, the whole architecture was named as WAVE which was comprised of a weight vector of classifiers and a weight vector of

* Corresponding author at: Selçuk University, Engineering Faculty, Electrical & Electronics Engineering Department, 42250 Konya, Turkey.

E-mail addresses: rpektatli@selcuk.edu.tr (R. Ceylan), hasankoyuncu@selcuk.edu.tr (H. Koyuncu).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2017.10.011
1319-1578/

instances. WAVE achieved 96.24% and 82.71% classification accuracies on WBC and HS datasets, respectively. Li et al. (2011) proposed a fuzzy-based non-linear transformation method (Proposed Method → PM) and placed it into Support Vector Machine (SVM) architecture. PM-SVM (Gauss) achieved 75.36% and 70.85% classification accuracies on PID and BLD, respectively. Ozcift and Gulten (2011) used different classifiers as the base classifiers of Rotation Forest algorithm. By using WEKA, 30 different classifiers (NN, logistic, Naïve Bayes etc.) were stated in RF. On average, RF ensembles succeeded 74.47%, 80.49% and 87.13% classification accuracies for PID, HS and PD datasets, respectively. Aldape-Pérez et al. (2012) generated Associative Memory-based Classifiers (AMBC) that include a learning phase, a learning reinforcement phase and a classification phase. In their study, WEKA simulation program was used for comparing the proposed method with others. By using a %50–% 50 training-test split, AMBC obtained classification accuracies of 65.40% (BLD), 70.57% (PID) and 83.33% (HS). Couellan and Wang (2015) proposed a bi-level stochastic gradient algorithm for the training of SVMs. In their study, Bi-level SVM was compared to cross validation SVM on PID and WBC datasets by using 5-fold cross validation. Bi-level SVM obtained 96.50% (WBC) and 75.07% (PID) classification accuracy rates. Hsieh et al. (2014) designed a PSO based Fuzzy Hyper-Rectangular Composite Neural Network (PFHRCNN) that applies PSO to adjust the rules formed by trained HRCNN during stable recognition performance. PFHRCNN obtained 82.4% classification accuracy on PD, and 96% on WBC. Tan et al. (2014) used Modified Micro Genetic Algorithm (MmGA) as the optimizer of ensembles, and designed the MmGA-based ensemble optimizer. This method achieved 97.21% and 83.68% classification accuracies on WBC and HS, respectively. Zhang and Suganthan (2014) modified the Random Forest algorithm with an ensemble of feature spaces. Consequently, proposed structure gained more diversity for better classification performance. On WBC dataset, Random Forest ensemble obtained 97.16% classification accuracy. Li and Leng (2015) produced Support Alternating Multi-Conlitron Algorithm (SAMA) originating from SVMs. For better linear classification, alternating multi-conlitrons were used in SAMA. In their study, Li and Leng tested SAMA on HS, PID and PD datasets by using 10-fold cross validation. As a result, classification accuracies were acquired at 66.96% (HS), 81.84% (PD) and 68.57% (PID). Shi et al. (2015) handled Kernel Entropy Component Analysis (KECA) and improved KECA by using a fuzzy set theory. Consequently, Fuzzy Robust KECA (FR-KECA) was obtained, with which classification accuracies were 70.44% on BLD and 78.12% on PID. Tao et al. (2015) designed Minimum Class Spread Constrained SVM (MCSSVM) that is more efficient than SVM. MCSSVM was tested on BLD and PID datasets which obtained classification accuracies of 69.52% and 76.11%, respectively. Xiang et al. (2015) formed a hybrid system based on Gravitational Search Algorithms and k-Nearest Neighbour (k-NN) for feature selection. In their study, classification accuracies were obtained as 77.9% (BIGSA-KNN), 76.3% (BGSA-KNN), 76.1% (QBPSO-KNN), 75.5% (BPSO-KNN) and 74.9% (GA-KNN) on PID dataset.

As is known, Rotation Forest ensemble structures generally consist of DT base classifiers. In this study, we handle RF architecture and modified it with hybrid NNs (PSO-NN, ABC-NN, ScPSO-NN) in order to obtain more robust ensemble classifiers (RF (PSO-NN), RF (ABC-NN), RF (ScPSO-NN)). At this point, two objectives are aimed as follows:

- Achievement of prior performance with RF (hybrid NNs) *vs.* hybrid NNs, RF (NN) and other studies (especially ensemble based ones) in literature.
- Examination of the subject that ScPSO is more compatible than PSO and ABC algorithms in order to work with complex classification architectures or not.

2. Methods

2.1. Particle swarm optimization

Particle Swarm Optimization is a well-known heuristic algorithm inspired by the foraging behaviors of bird swarms (Kennedy and Eberhart, 1995). In PSO, every bird constitutes a solution in feature space, and these individuals are named as 'particles'. Particles attempt to proceed towards the global minimum or maximum point(s) and towards food source(s) iteratively. For this purpose, velocity and position operators are utilized in update of particles. Eqs. (1) and (2) respectively shows the velocity and position phenomenon.

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

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{2}$$

In Eq. (1), ω is the inertia weight that confines the step size. $V_i(t)$ is the current velocity (step size). r_1 and r_2 coefficients are randomly produced values within the range [0, 1]. c_1 and c_2 coefficients are the acceleration constants leading the particles towards global points. Herein, the values of these coefficients can generally accelerate or slacken the particles where the sum of coefficients is equalized to '4'. $X_i(t)$ is the current position, $X_{pbest(i)}(t)$ symbolizes the best individual position of the *i*.th particle, and $X_{gbest}(t)$ stands for the best position in the whole swarm. In Eq. (2), $V_i(t+1)$ is the newly generated velocity that will form the new position ($X_i(t+1)$) by being summed with the old position ($X_i(t)$).

2.2. Artificial bee colony optimization

Artificial Bee Colony Optimization (ABC) simulates the foraging behaviors of honey bees (Karaboga and Akay, 2009). In ABC, four phases are actively used for the update of positions. At the beginning of iteration, initialization phase operates Eq. (3) in order to generate particles.

$$x_m = l_i + rand(0,1) * (u_i - l_i) \tag{3}$$

In Eq. (3), l_i and u_i become the lower and upper boundaries of the positions, respectively. $rand(0,1)$ is a fixed parameter generated within the range [0,1] as r_1 and r_2 coefficients in PSO.

After initialization, loop part starts with the employed bee phase with which Eqs. (4) and (5) are used for comparing the old positions with new ones as the result of Eq. (4).

$$\upsilon_{mi} = x_{mi} + \phi_{mi}(x_{mi} - x_{ki}) \tag{4}$$

$$fit_m(x_m) = \begin{cases} 1/(1 + f_m(x_m)), & f_m(x_m) > 0 \\ 1 + |f_m(x_m)|, & f_m(x_m) < 0 \end{cases} \tag{5}$$

In Eq. (4), x_{mi} specifies the *m*.th position, ϕ is a constant within the range [−1, 1], and *i* stands for a randomly generated number in [1, dimension number]. υ_{mi} symbolizes the newly generated position to be compared with the old one (x_m). At this point, the important point is that *k* cannot have the same value with *m*, otherwise, $(x_{mi} - x_{ki})$ will be equal to '0'. In Eq. (5), $f_m(x_m)$ holds the result of objective function. Additionally, the fitness value ($fit_m(x_m)$) is obtained by examining the condition of the objective function. As a result, υ_{mi} is replaced with x_m in the case of a better fitness value.

Onlooker bee phase uses Eq. (4) for generating new particles and Eq. (5) for comparing their fitness values. But, initially, Eq. (6) is operated to choose a position value (*m*) to be used in Eq. (4). In Eq. (6), P_m is the probability choice of x_m and P_m is compared to a randomly generated value within the range [0,1] for achieving a diversified selection.

$$P_m = \frac{fit_m(x_m)}{\sum_{m=1}^{SN} fit_m(x_m)} \quad (6)$$

Scout bee phase is activated if progress is unable to be achieved among a user-defined iteration number called *limit* (Karaboga and Akay, 2009). According to Eq. (3), scout bees regenerate new solutions (regeneration of ineffective positions).

2.3. Scout particle swarm optimization

Scout Particle Swarm Optimization is a modified PSO algorithm proposed by Koyuncu and Ceylan (2015). In ScPSO, scout bee phase is added to standard PSO algorithm in order to upgrade the performance.

PSO updates the particles' positions using position and velocity concepts, while ABC uses the equations that are settled in the employed bee and onlooker bee phases. By activating the scout bee phase, ABC regenerates insufficient particles that are blocked from advancing their position towards the global region(s).

PSO does not include a stable or variable parameter that controls the efficiency of updates. In consideration of this information, ScPSO is formed by adding the scout bee phase to the basic PSO algorithm (Koyuncu and Ceylan, 2015).

According to the study of Koyuncu and Ceylan (2015), ScPSO exhibits a challenging performance *vs.* PSO due to better convergence. Fig. 1 shows the flow chart of ScPSO algorithm.

2.4. Rotation forest

The Rotation Forest (RF) is a Decision Tree (DT) based ensemble classifier system proposed by Rodriguez et al. (2006). RF employs a set of classifiers to be trained with improved datasets formed from the original data. Herein, diversity plays an important role in classification performance. With diversity, RF aims to produce the prior performance *vs.* other ensemble architectures. To this end, RF ensures diversity through two different steps (the bootstrap method and the change of feature vectors' places process). In order for best performance to be achieved, a trade-off between diversity-accuracy is needed with the adjustment of the base classifier and subset numbers.

In the first part of RF, places of the feature vectors are changed and the whole matrix is divided into submatrices. Later, bootstrap is applied to all submatrices at the rate of 75%. Then, Principle Component Analysis (PCA) is applied to submatrices in order to obtain eigenvectors. Thereafter, eigenvectors are stated as being diagonal in a matrix called '*rotation matrix*'. Consequently, coefficients are produced that are multiplied with the original data. After the multiplication process, improved and rotated data is obtained.

For RF, Rodriguez et al. (2006) suggest the pseudocode shown below:

Training part

- *X:* the objects in the training dataset (an *Nxn* matrix)
- *Y:* the labels of the training set (an *Nx1* matrix)
- *L:* the number of classifiers in the ensemble
- *K:* the number of subsets
- *{W₁, W₂, ..., W_c}*: the set of class labels

For i = 1:*L*

- Preparation of the rotation matrix R_i^a
  - Splits F (feature set) into K subsets: $F_{i,j}$ (for j = 1:K)
  - For j = 1:K
    1. Let $X_{i,j}$ be the dataset X for the features in $F_{i,j}$
    2. Eliminates from $X_{i,j}$ a random subset of classes
    3. Selects a bootstrap sample from $X_{i,j}$ of size 75% of the object number in $X_{i,j}$. Denotes the new set by $X_{i,j}$'
    4. Applies PCA to $X_{i,j}$' to obtain the coefficients in matrix $C_{i,j}$
  - Arranges the $C_{i,j}$, for j = 1,. . .,K in a rotation matrix R_i as in Eq. (7)

$$R_i = \begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(2)}, \ldots . a_{i,1}^{(M_1)} & [0] & \ldots & [0] \\ [0] & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \ldots, a_{i,2}^{(M_2)} & \ldots & [0] \\ \ldots & \ldots & \ldots & \ldots \\ [0] & [0] & \ldots & a_{i,k}^{(1)}, a_{i,k}^{(2)}, \ldots, a_{i,k}^{(M_k)} \end{bmatrix} \quad (7)$$

  - Constructs R_i^a by rearranging the columns in order to match the order of features in F.
  - Builds classifier D_i using (X R_i^a, Y) as the training set

Classification part

- For a given x, let $d_{i,j}(xR_i^a)$ be the probability assigned by the classifier $D_{i,j}$ to the hypothesis that x comes from class W_j. Calculate the confidence for each class, W_j, by the average combination method (8).

$$\mu_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(xR_i^a), j = 1, \ldots, c \quad (8)$$

- Assign x to the class with the largest confidence

The bootstrap generates a smaller dataset by replacing and by choosing patterns from the original dataset. The change of the feature vectors' places implies that all eigenvectors are multiplied with a different part of data according to the submatrices in rotation matrix. It should not be forgotten that all classifiers are trained with improved datasets (different training datasets). However in tests, diversity (the bootstrap and change of the feature vectors' places) is not needed, and all classifiers are tested with the same data (test data).

2.5. Proposed method

RF consists of DT classifiers in its classifier unit. In this study, we modify the RF algorithm in which the hybrid NNs are used as base classifiers. For this purpose, the fully optimized NNs are chosen due to their ability to achieve a better classification performance than simple base classifiers (NNs, DTs, SVM, etc.). At this point, optimization algorithms are directly settled in the update part of NNs in lieu of the backpropagation algorithm. As a result, a robust RF structure is generated for binary-medical pattern classification that challenges the DT based RFs and most of ensemble classifier methods advised in literature. Fig. 2 shows the operation of hybrid NNs.

In Fig. 2;

- The first process stands for the generation of weight and bias values of NN. At this point, particle values (values of weight & bias vectors) can be restricted in a boundary specified by user.
- At the second step, training of NN is realized in order to produce the output of system, and Mean Squared Error (MSE) is obtained by comparing the obtained output with the target.

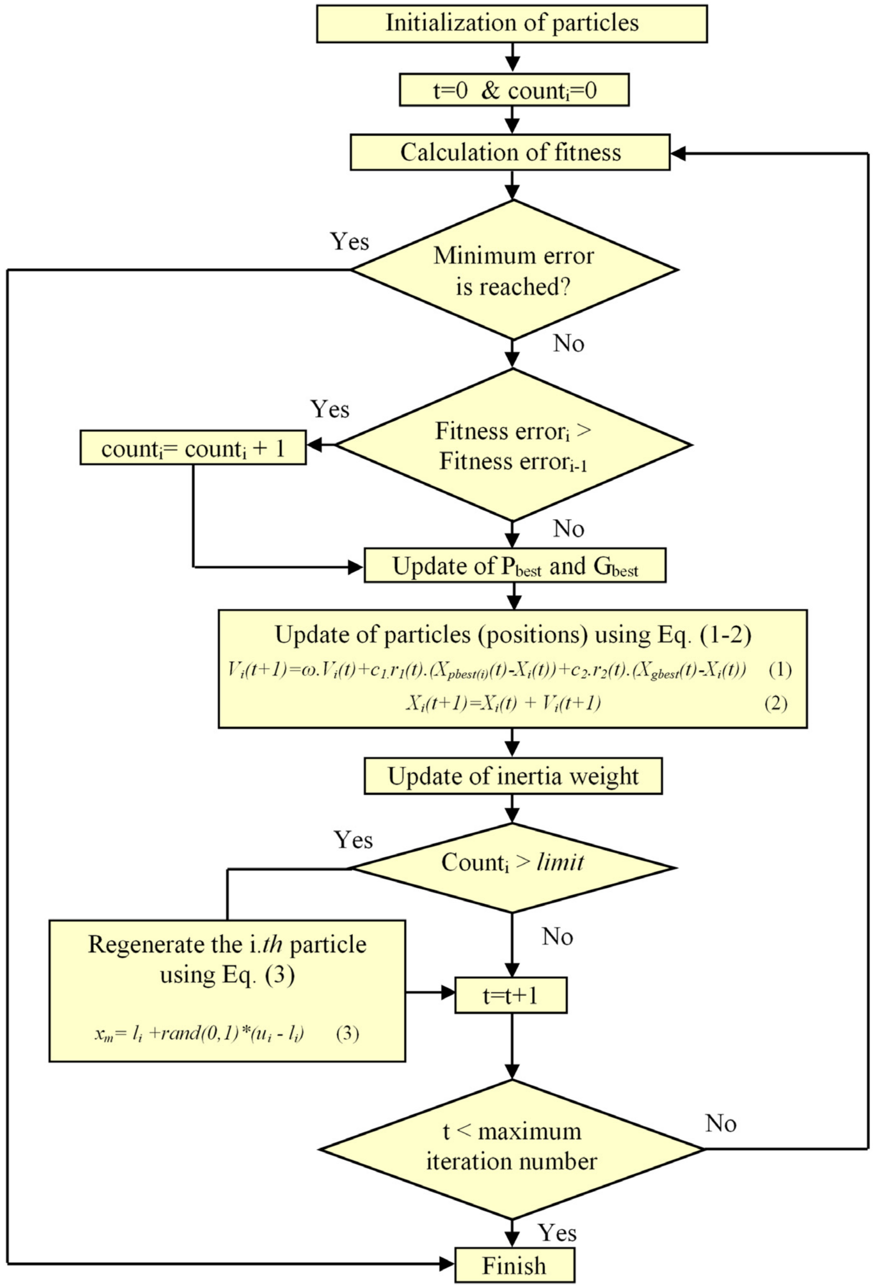

Fig. 1. Flowchart of ScPSO algorithm.

- At the third part, weight and bias vectors are updated according to the error rates. Herein, update process is realized by using the phenomena in optimization algorithms. In PSO-NN, Eqs. (1) and (2) are utilized to regenerate the vector values. In ABC-NN, Eqs. (3)–(6) are used for regeneration process. In ScPSO-NN, Eqs. (1)–(3) are used to assign the new values of vectors. At the fourth part, maximum iteration number and minimum error rate are determined to decide the termination of training process.
- At the last part, test is performed by using the weight and bias vector achieving to the minimum error rate.

Fig. 3 shows the operation of hybrid NNs based RF ensemble classifiers.

In Fig. 3;

- The first three processes (change of feature vectors' places, separation of the data into subsets and bootstrap) ensures the diversity that is needed for better accuracy. Herein, the change of feature vectors' places and bootstrap are applied to the original dataset with different combinations for obtaining different-diversified data to be used as the input of classifiers. The separation of data into subsets (subset number) is kept fixed for all diversified data.
- In the obtainment of subsets, the data matrix is divided to subsets including different features.

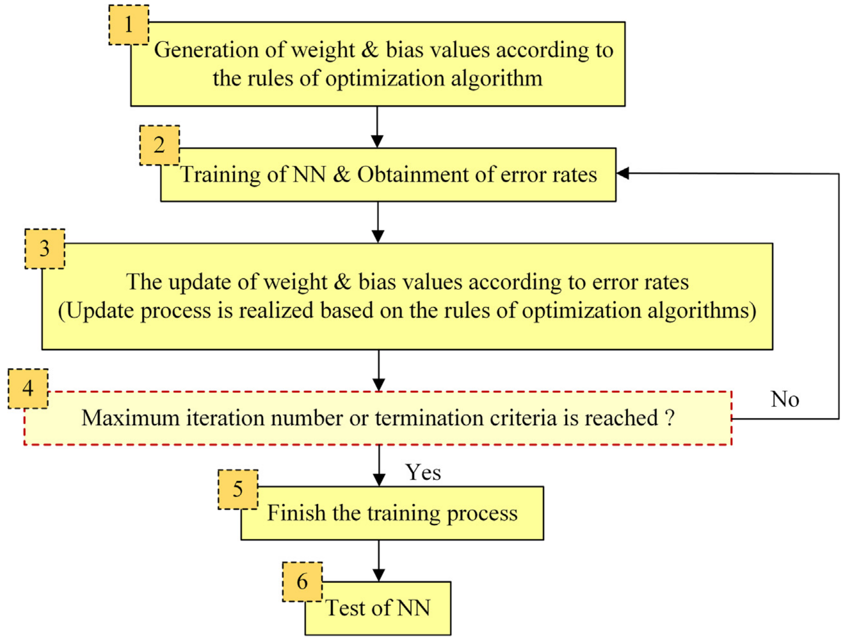

Fig. 2. The operation of hybrid NNs.

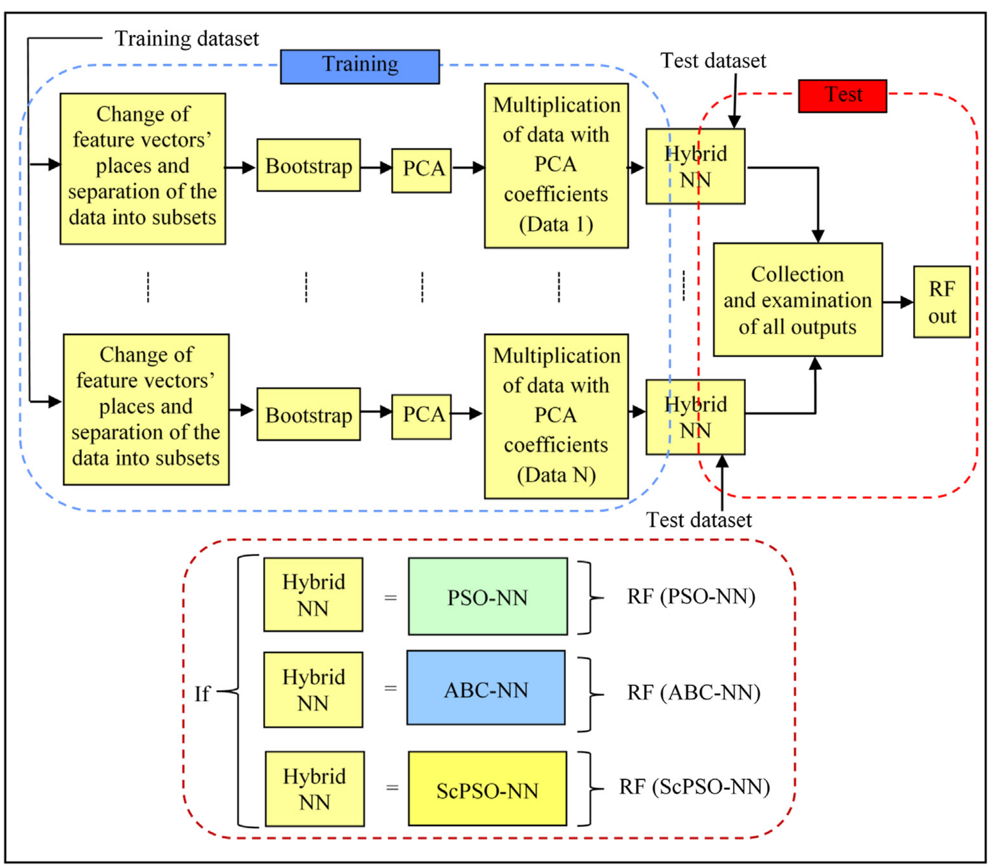

Fig. 3. The operation of RF (hybrid NN) ensemble classifiers.

- Bootstrap is performed to every subset which contains randomly chosen patterns, and the number of patterns is kept as %75 of the whole subset as in the original pseudocode of RF (Rodriguez et al., 2006).
- PCA method is utilized to obtain the eigenvector coefficients used in the obtainment of diversified data. At this point, PCA is performed to all bootstrap-applied subsets, and square matrices are obtained at the size of feature numbers for all subsets.

- Before the multiplication of coefficients, every square matrix is settled in a zero valued-matrix as being diagonal, and the diagonal matrix is formed.
- After the obtainment of diagonal matrix, the features in matrix should be returned to their initial order as in the original data (to the order that is same with the order in original data).
- The multiplication process is performed by multiplying the coefficients of diagonal matrix and original data to attain the diversified data to be used in Hybrid NN.
- Herein, the four processes (the change of feature vectors' places, separation of the data into subsets, bootstrap, PCA, multiplication of coefficients) diversify the used data, and these parts are operated in training phase.
- Hybrid NNs are used in both training and test phases. In training phase, the diversified RF data are presented as the input of hybrid NNs. In test phase, best operation condition of hybrid NNs are kept stable and test process is performed by using the hybrid NNs trained with different-diversified data.
- At the last part, the outputs of hybrid NNs are weighted as being equal, and the collaborative output (the output of RF ensemble) is achieved by using the class with the largest confidence (Eq. (8)). Herein, it should be specified that every section (change of features' vector places, separation of the data into subsets, bootstrap, PCA, multiplication of the data with PCA coefficients) is applied individually for every classifier as being diverse from each other. Thus, the diversity can be effective on classification process. At this point, this process positively effects the classification results, if the necessary tradeoff between diversity vs. accuracy (change of parameters) are met.

In RF (hybrid NNs) techniques, the whole structure is changed and named according to the type of used hybrid NNs. In other words, the type of RF ensemble is connected with the used hybrid NN. Consequently, three kinds of RF ensemble can be obtained: 1- RF (PSO-NN), 2- RF (ABC-NN), 3- RF (ScPSO-NN).

3. Experimental results

In experiments, not only the accordance of hybrid NNs with RF is examined, but also the optimization algorithms are evaluated in terms of running compatible with the designed RF structure and with hybrid NNs. Thus, a detailed analysis of hybrid methods and more complex algorithms is realized for the task of pattern classification.

In this study, the classifier structures (hybrid NNs) are examined experimentally whether these methods run in accordance with the diversity operators (the change of feature vectors' places and separation of the data into subsets & bootstrap) and with others or not as in general RF approach in which DT are utilized. The results are evaluated by utilizing from seven metrics that constitute the most preferable measurement approaches to test the performance of a novel classification algorithm. In other words, an objective comparison of methods is wanted to be realized for a detail analysis of the designed techniques.

3.1. The used medical datasets

RF (hybrid NN) structures are tested via 5 well-known medical datasets taken from UCI machine-learning repository (Lichman, 2013). Bupa Liver Disorders (BLD), Heart Statlog (HS), Parkinson (PD) (Little et al., 2009), Pima Indian Diabetes (PID) and Wisconsin Breast Cancer (Original) (Wolberg and Mangasarian, 1990) datasets are used as input information to NN, hybrid NNs, RF (NN) and to RF (Hybrid NN) classifiers. Features of these datasets are shown in Table 1.

The BLD dataset is a combination of mass volume and chemical measurements in blood. The Parkinson dataset (PD) is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson disease (Little et al., 2009). Each column in PD is a particular voice measurement, and each row corresponds to one of 195 voice recordings from these individuals. As separator features, PD contains various frequency measurements and some statistical metrics. PID dataset is primarily based on the blood results. HS dataset includes blood tests, electrocardiogram (ECG) signal parts, chest examinations etc. In WBC, samples were prepared by Dr. Wolberg from his clinical cases. WBC uses cell-based features to categorize the patterns (Wolberg and Mangasarian, 1990).

3.2. Three staggered performance comparison via 7 diverse statistical metrics

In literature, most of the studies assess their methods with Classification Accuracy (CA), some with Area under ROC Curve (AUC), a few with Specificity and Sensitivity, and some with other methods. Ordinarily, CA is chosen as a descriptive metric on performance, but in some circumstances, CA is not seen as a sufficient statistical metric to determine good performance (Huang and Ling, 2005; Tang et al., 2009). Consequently, the significance of other statistical metrics becomes important in order to measure real performance. In other studies based on pattern classification, AUC is the second most chosen statistical parameter after CA. ROC graphs are two-dimensional graphs in which the *true positive* (*positive and, correctly classified as positive*) rate is settled on the *Y* axis and the *false positive* (*actually negative, and classified as positive*) rate is placed on the *X* axis. This graph reveals the relative trade-offs between the benefits (true positives) and costs (false positives). At this point, AUC expresses the area under the ROC curve (Fawcett, 2006). Alongside CA and AUC metrics, other statistical measurements can be used together for a much more sensitive analysis.

In this study, we used 7 different statistical metrics (Eqs. (9)–(14)) for the assessment of classifier architectures (Sensitivity, Specificity, *G*-mean, *F*-measure, Precision, AUC and CA) (Bradley, 1997; Fawcett, 2006).

$$Sensitivity = TP/(TP + FN) \tag{9}$$

$$Specificity = TN/(TN + FP) \tag{10}$$

$$Precision = TP/(TP + FP) \tag{11}$$

$$F - measure = 2 * [(precision * recall)/(precision + recall)] \tag{12}$$

$$G - mean = \sqrt{Sensitivity * Specificity} \tag{13}$$

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

In order to enhance the diversity and to provide an objective view, a three-step comparison was applied to classifier systems via 7 statistical metrics. Additionally, Total Statistical Success (TSS) values were used to measure the success of classifiers. Herein, TSS is defined as the number of best performances (or the number of obtained first rows). TSS (7 metrics) is the basis of comparisons, since it symbolizes the total success and ensures an equilibrium sight.

- TSS comparison based on 7 metrics (sensitivity, specificity, *F*-measure, *G*-mean, Precision, AUC and CA) → TSS (7 metrics)
- TSS comparison based on CA → TSS (CA)
- TSS comparison based on AUC → TSS (AUC)

Table 1
The used medical datasets (continued).

Datasets	Attribute information	Pattern Number/Number of Attributes (including class)
Bupa Liver Disorders	• Mcv mean corpuscular volume • Alkphos alkaline phosphotase • Sgpt alamine aminotransferase • Sgot aspartate aminotransferase • Gammagt gamma-glutamyl transpeptidase • Drinks number of half-pint equivalents of alcoholic beverages drunk per day	345/7
Heart Statlog	• Age • Sex • Chest pain type (4 values) • Resting blood pressure • Serum cholesterol in mg/dl • Fasting blood sugar > 120 mg/dl • Resting electrocardiographic results (values 0,1,2) • Maximum heart rate achieved • Exercise induced angina • Oldpeak = ST Depression induced by exercise relative to rest • The slope of the peak exercise ST segment • Number of major vessels (0–3) colored by fluoroscopy • Thal: 3 = Normal; 6 = Fixed defect; 7 = Reversible defect	270/13
Parkinson	• MDVP:Fo(Hz) – Average vocal fundamental frequency • MDVP:Fhi(Hz) – Maximum vocal fundamental frequency • MDVP:Flo(Hz) – Minimum vocal fundamental frequency • MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP – Several measures of variation in fundamental frequency • MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3,Shimmer:APQ5, MDVP:APQ, Shimmer:DDA – Several measures of variation in amplitude • NHR,HNR – Two measures of ratio of noise to tonal components in the voice • Status – Health status of the subject (one) – Parkinson's, (zero) – healthy • RPDE, D2 – Two nonlinear dynamical complexity measures • DFA – Signal fractal scaling exponent • spread1, spread2, PPE – Three nonlinear measures of fundamental frequency variation	197/23
Pima Indian Diabetes	• Number of times pregnant • Plasma glucose concentration a 2 h in an oral glucose tolerance test • Diastolic blood pressure (mm Hg) • Triceps skin fold thickness (mm) • 2-Hour serum insulin (mu U/ml) • Body mass index (weight in kg/(height in m)^2) • Diabetes pedigree function • Age	768 / 8
Wisconsin Breast Cancer	• Clump thickness • Uniformity of cell size • Uniformity of cell shape • Marginal adhesion • Single epithelial cell size • Bare nuclei • Bland chromatin • Normal nucleoli • Mitoses	683 / 10

With the first step, optimum operation conditions were found based on TSS (7 metrics), and the results obtained in Table 4–8 were rearranged. In this regard, Table 9 is obtained by comparing the classifiers according to TSS (7 metrics) at their optimum operation conditions. After this process, CA and AUC based comparisons are obtained by means of 7 metrics. According to the three-step comparison stated above, the results that reveal the best classification structure are presented in Table 10. 2-fold cross validation method is used in all trials for performance evaluation of proposed techniques.

3.2.1. Detection of optimum operation conditions

Parameter values of the NN, PSO-NN and ScPSO-NN classifiers were adjusted as advised in study of Koyuncu and Ceylan (2015). Moreover, the ABC-NN parameters were stable at their optimal conditions determined by trial and error. In the hybrid NNs (PSO-NN, ABC-NN and ScPSO-NN) architectures, the unique variable was the hidden node number that changed per the number of input nodes (feature number of dataset). In trials, hidden node number at which the highest TSS (7 metrics) value was achieved was chosen as the optimum hidden node number for the classifier.

RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) structures contain two variable parameters: the base classifier number and the subset number. These parameters were changed according to the output of the system. Firstly, the subset number was obtained based on RF (ABC-NN) architecture. Then, the best subset number with the highest TSS (7 metrics) was chosen as the optimal choice for other RFs. Here, the usage of RF (ABC-NN) was fulfilled as a result of providing an objective operation condition. In other words, RF (ScPSO-NN) was not chosen because of ensuring the impartiality. Additionally, the ABC algorithm exhibits a different algorithm from the PSO and ScPSO methods. Secondly, the base classifier number was chosen according to the success based on TSS (7 metrics). In all RF systems, the hidden node number of hybrid NNs remained same according to the optimum node numbers found in trials of hybrid NNs. Table 2 shows the optimum operation conditions and changeable parameter(s).

Table 2
Optimum operation conditions for classifier systems.

Classifier systems	Stable parameters & values	Variable parameters
NN	H = 10 lr = 0.1	–
PSO-NN	Maxiter = 500 Pop-size = 5 c_1 = 2.08 & c_2 = 1.92 inertia_w = 0.9 → 0.4	• H
ABC-NN	Maxiter = 500 Pop-size = 5 u_i = 10 & l_i = −10 Limit = 100	
ScPSO-NN	Maxiter = 500Pop-size = 5c_1 = 2 & c_2 = 2inertia_w = 0.9 → 0.4 Limit = pop-size * dim/12	
RF (NN) RF (PSO-NN) RF (ABC-NN) RF (ScPSO-NN)	H = 10 lr = 0.1 PSO-NN parameters are used ABC-NN parameters are used ScPSO-NN parameters are used	• Base classifier number • Subset Number

(maxiter = maximum iteration number, pop-size = population size, H = hidden node number, inertia_w = inertia weight, lr = learning rate, dim = dimension).

Table 3
Optimum subset number for RF ensemble classifiers.

Dataset	Subset Number	2	3	4	5	6	7
BLD (Sub_num = 4)	Sensitivity	51,72	47,59	48,28			
	Specificity	56,00	62,50	**80,00**			
	Gmean	53,82	54,54	**62,15**			
	Precision	46,01	47,92	**63,64**			
	F-measure	48,70	47,75	**54,90**			
	AUC	53,86	55,04	**64,14**			
	CA	54,20	56,23	**66,67**			
HS (Sub_num = 6)	**Subset Number**	**2**	**3**	**4**	**5**	**6**	**7**
	Sensitivity	71,67	79,17	80,00	**80,83**	76,67	73,33
	Specificity	82,67	80,67	84,67	83,33	**90,67**	88,67
	Gmean	76,97	79,91	82,30	82,07	**83,37**	80,64
	Precision	76,79	76,61	80,67	79,51	**86,79**	83,81
	F-measure	74,14	77,87	80,33	80,17	**81,42**	78,22
	AUC	77,17	79,92	82,33	82,08	**83,67**	81,00
	CA	77,78	80,00	82,59	82,22	**84,44**	81,85
PD (Sub_num = 2)	**Subset Number**	**2**	**3**	**4**	**5**	**6**	**7**
	Sensitivity	97,96	97,28	92,52	**98,64**	95,24	94,56
	Specificity	**60,42**	54,17	50,00	39,58	**60,42**	56,25
	Gmean	**76,93**	72,59	68,01	62,49	75,85	72,93
	Precision	**88,34**	86,67	85,00	83,33	88,05	86,88
	F-measure	**92,90**	91,67	88,60	90,34	91,50	90,55
	AUC	**79,19**	75,72	71,26	69,11	77,83	75,40
	CA	**88,72**	86,67	82,05	84,10	86,67	85,13
PID (Sub_num = 4)	**Subset Number**	**2**	**3**	**4**			
	Sensitivity	54,10	41,04	**58,21**			
	Specificity	87,20	**91,60**	85,80			
	Gmean	68,69	61,32	**70,67**			
	Precision	69,38	72,37	68,72			
	F-measure	60,80	52,38	**63,03**			
	AUC	70,65	66,32	**72,00**			
	CA	75,65	73,96	**76,17**			
WBC (Sub_num = 4)	**Subset Number**	**2**	**3**	**4**	**5**	**6**	
	Sensitivity	95,82	96,65	**97,91**	94,98	95,82	
	Specificity	96,62	96,62	95,95	96,62	**96,85**	
	Gmean	96,22	96,64	**96,92**	95,80	96,33	
	Precision	93,85	93,90	92,86	93,80	**94,24**	
	F-measure	94,82	95,26	**95,32**	94,39	95,02	
	AUC	96,22	96,64	**96,93**	95,80	96,33	
	CA	96,34	96,63	**96,63**	96,05	96,49	

(The best results are presented as bold numbers, sub_num = subset number).

Table 4
Optimum values for variable parameters (on BLD).

Classifier	H	3	4	5	6	8	10	12	14	16	18	20
PSO-NN (H = 18)	Sensitivity	28,97	30,34	37,24	**52,41**	42,07	50,34	**52,41**	37,93	42,76	46,21	45,52
	Specificity	**90,00**	77,50	85,00	71,50	77,50	75,00	70,50	73,50	71,50	83,50	77,50
	Gmean	51,06	48,49	56,26	61,22	57,10	61,45	60,79	52,80	55,29	**62,12**	59,39
	Precision	**67,74**	49,44	64,29	57,14	57,55	59,35	56,30	50,93	52,10	67,00	59,46
	F-measure	40,58	37,61	47,16	54,68	48,61	54,48	54,29	43,48	46,97	**54,69**	51,56
	AUC	59,48	53,92	61,12	61,96	59,78	62,67	61,46	55,72	57,13	**64,85**	61,51
	CA	64,35	57,68	64,93	63,48	62,61	64,64	62,90	58,55	59,42	**67,83**	64,06
ABC-NN (H = 18)	Sensitivity	24,83	56,55	44,14	42,07	46,90	48,28	52,41	57,24	33,10	**64,83**	43,45
	Specificity	**85,50**	70,00	70,00	53,00	80,50	80,00	82,50	71,50	81,00	72,00	83,00
	Gmean	46,07	62,92	55,58	47,22	61,44	62,15	65,76	63,97	51,78	**68,32**	60,05
	Precision	55,38	57,75	51,61	39,35	63,55	63,64	**68,47**	59,29	55,81	62,67	64,95
	F-measure	34,29	57,14	47,58	40,67	53,97	54,90	59,38	58,25	41,56	**63,73**	52,07
	AUC	55,16	63,28	57,07	47,53	63,70	64,14	67,46	64,37	57,05	**68,41**	63,22
	CA	60	64,35	59,13	48,41	66,38	66,67	**69,86**	65,51	60,87	68,99	66,38
ScPSO-NN (H = 20)	Sensitivity	**54,48**	41,38	29,66	53,10	35,86	33,10	48,28	35,86	48,28	49,66	51,03
	Specificity	70,00	78,50	81,00	71,00	83,00	80,50	68,50	89,00	80,50	73,00	82,00
	Gmean	61,76	56,99	49,01	61,40	54,56	51,62	57,51	56,50	62,34	60,21	**64,69**
	Precision	56,83	58,25	53,09	57,04	60,47	55,17	52,63	70,27	64,22	57,14	67,27
	F-measure	55,63	48,39	38,05	55,00	45,02	41,38	50,36	47,49	55,12	53,14	**58,04**
	AUC	62,24	59,94	55,33	62,05	59,43	56,80	58,39	62,43	64,39	61,33	**66,52**
	CA	63,48	62,90	59,42	63,48	63,19	60,58	60,00	66,67	66,96	63,19	**68,99**

Classifier	BCN	5	10	15	20	25	30	35	40
RF (NN) (BCN = 20)	Sensitivity	20,69	48,97	48,97	44,14	28,28	51,72	41,38	35,86
	Specificity	**95,50**	63,00	68,00	84,00	90,00	54,00	77,50	77,50
	Gmean	44,45	55,54	57,70	**60,89**	50,45	52,85	56,63	52,72
	Precision	**76,92**	48,97	52,59	66,67	67,21	44,91	57,14	53,61
	F-measure	32,61	48,97	50,71	**53,11**	39,81	48,08	48,00	42,98
	AUC	58,09	55,98	58,48	**64,07**	59,14	52,86	59,44	56,68
	CA	64,06	57,10	60,00	**67,25**	64,06	53,04	62,32	60,00
RF (PSO-NN) (BCN = 10)	Sensitivity	**44,83**	32,41	31,03	26,90	20,00	31,72	34,48	31,03
	Specificity	65,50	92,50	88,50	91,50	**96,50**	91,50	89,00	92,00
	Gmean	54,19	54,76	52,41	49,61	43,93	53,88	**55,40**	53,43
	Precision	48,51	75,81	66,18	69,64	**80,56**	73,02	69,44	73,77
	F-measure	46,59	45,41	42,25	38,81	32,04	44,23	46,08	43,69
	AUC	55,16	**62,46**	59,77	59,20	58,25	61,61	61,74	61,52
	CA	56,81	**67,25**	64,35	64,35	64,35	66,38	66,09	66,38
RF (ABC-NN) (BCN = 5)	Sensitivity	48,28	43,45	48,97	43,45	32,41	43,45	43,45	44,83
	Specificity	80,00	67,50	63,00	67,50	**85,00**	81,50	78,00	79,50
	Gmean	**62,15**	54,15	55,54	54,15	52,49	59,51	58,21	59,70
	Precision	**63,64**	49,22	48,97	49,22	61,04	63,00	58,88	61,32
	F-measure	**54,90**	46,15	48,97	46,15	42,34	51,43	50,00	51,79
	AUC	**64,14**	55,47	55,98	55,47	58,71	62,47	60,72	62,16
	CA	**66,67**	57,39	57,10	57,39	62,90	65,51	63,48	64,93
RF (ScPSO-NN) (BCN = 25)	Sensitivity	47,59	30,34	64,83	48,97	55,17	51,72	43,45	30,34
	Specificity	55,50	82,50	59,00	63,00	83,00	79,50	77,50	**87,50**
	Gmean	51,39	50,03	61,85	55,54	**67,67**	64,13	58,03	51,53
	Precision	43,67	55,70	53,41	48,97	**70,18**	64,66	58,33	63,77
	F-measure	45,54	39,29	58,57	48,97	**61,78**	57,47	49,80	41,12
	AUC	51,54	56,42	61,91	55,98	**69,09**	65,61	60,47	58,92
	CA	52,17	60,58	61,45	57,10	**71,30**	67,83	63,19	63,48

(The best results are presented as bold numbers, H = hidden node number, BCN = base classifier number).

Table 5
Optimum values for variable parameters (on HS).

Classifier	H	4	6	8	10	12	14	16	18	20	22	24	26	28	30
PSO-NN (H = 16)	Sensitivity	68,33	69,17	69,17	69,17	74,17	75,83	78,33	**81,67**	76,67	73,33	73,33	65,83	72,50	70,83
	Specificity	**85,33**	81,33	82,00	74,67	77,33	82,00	81,33	75,33	74,67	80,67	71,33	80,67	84,00	74,00
	Gmean	76,36	75,00	75,31	71,86	75,73	78,86	**79,82**	78,44	75,66	76,91	72,33	72,87	78,04	72,40
	Precision	**78,85**	74,77	75,45	68,60	72,36	77,12	77,05	72,59	70,77	75,21	67,18	73,15	78,38	68,55
	F-measure	73,21	71,86	72,17	68,88	73,25	76,47	**77,69**	76,86	73,60	74,26	70,12	69,30	75,32	69,67
	AUC	76,83	75,25	75,58	71,92	75,75	78,92	**79,83**	78,50	75,67	77,00	72,33	73,25	78,25	72,42
	CA	77,78	75,93	76,30	72,22	75,93	79,26	**80,00**	78,15	75,56	77,41	72,22	74,07	78,89	72,59
ABC-NN (H = 26)	Sensitivity	70,00	79,17	76,67	68,33	75,00	74,17	76,67	71,67	67,50	71,67	69,17	**80,00**	74,17	77,50
	Specificity	81,33	76,00	84,67	82,67	84,67	79,33	80,67	80,67	**85,33**	77,33	78,67	82,67	81,33	74,67
	Gmean	75,45	77,57	80,57	75,16	79,69	76,71	78,64	76,03	75,89	74,45	73,76	**81,32**	77,67	76,07
	Precision	75,00	72,52	**80,00**	75,93	79,65	74,17	76,03	74,78	78,64	71,67	72,17	78,69	76,07	70,99
	F-measure	72,41	75,70	78,30	71,93	77,25	74,17	76,35	73,19	72,65	71,67	70,64	**79,34**	75,11	74,10
	AUC	75,67	77,58	80,67	75,50	79,83	76,75	78,67	76,17	76,42	74,50	73,92	**81,33**	77,75	76,08
	CA	76,30	77,41	81,11	76,30	80,37	77,04	78,89	76,67	77,41	74,81	74,44	**81,48**	78,15	75,93
ScPSO-NN (H = 28)	Sensitivity	76,67	70,83	80,00	69,17	77,50	67,50	75,00	69,17	78,33	72,50	71,67	75,00	**80,83**	69,17
	Specificity	82,00	80,67	80,67	71,33	76,00	82,67	80,00	**88,67**	78,00	79,33	79,33	86,00	83,33	82,00
	Gmean	79,29	75,59	80,33	70,24	76,75	74,70	77,46	78,31	78,17	75,84	75,40	80,31	**82,07**	75,31
	Precision	77,31	74,56	76,80	65,87	72,09	75,70	75,00	**83,00**	74,02	73,73	73,50	81,08	79,51	75,45
	F-measure	76,99	72,65	78,37	67,48	74,70	71,37	75,00	75,45	76,11	73,11	72,57	77,92	**80,17**	72,17
	AUC	79,33	75,75	80,33	70,25	76,75	75,08	77,50	78,92	78,17	75,92	75,50	80,50	**82,08**	75,58
	CA	79,63	76,30	80,37	70,37	76,67	75,93	77,78	80,00	78,15	76,30	75,93	81,11	**82,22**	76,30

Classifier	BCN	5	10	15	20	25	30	35	40
RF (NN) (BCN = 35)	Sensitivity	78,33	75,00	74,17	75,00	74,17	77,50	**80,00**	78,33
	Specificity	83,33	84,00	84,67	82,67	**88,67**	85,33	85,33	84,67
	Gmean	80,79	79,37	79,24	78,74	81,09	81,32	**82,62**	81,44
	Precision	78,99	78,95	79,46	77,59	**83,96**	80,87	81,36	80,34
	F-measure	78,66	76,92	76,72	76,27	78,76	79,15	**80,67**	79,32
	AUC	80,83	79,50	79,42	78,83	81,42	81,42	**82,67**	81,50
	CA	81,11	80,00	80,00	79,26	82,22	81,85	**82,96**	81,85
RF (PSO-NN) (BCN = 35)	Sensitivity	72,50	**79,17**	75,83	76,67	75,00	76,67	78,33	78,33
	Specificity	85,33	79,33	86,67	84,67	**88,67**	**88,67**	**88,67**	**88,67**
	Gmean	78,66	79,25	81,07	80,57	81,55	82,45	**83,34**	83,34
	Precision	79,82	75,40	81,98	80,00	84,11	84,40	**84,68**	84,68
	F-measure	75,98	77,24	78,79	78,30	79,30	80,35	**81,39**	81,39
	AUC	78,92	79,25	81,25	80,67	81,83	82,67	**83,50**	83,50
	CA	79,63	79,26	81,85	81,11	82,59	83,33	**84,07**	84,07
RF (ABC-NN) (BCN = 5)	Sensitivity	**76,67**	76,67	75,00	75,00	75,00	47,59	75,00	72,50
	Specificity	90,67	86,00	**91,33**	90,67	90,00	55,50	86,67	88,00
	Gmean	**83,37**	81,20	82,76	82,46	82,16	51,39	80,62	79,87
	Precision	86,79	81,42	**87,38**	86,54	85,71	43,67	81,82	82,86
	F-measure	**81,42**	78,97	80,72	80,36	80,00	45,54	78,26	77,33
	AUC	**83,67**	81,33	83,17	82,83	82,50	51,54	80,83	80,25
	CA	**84,44**	81,85	84,07	83,70	83,33	52,17	81,48	81,11
RF (ScPSO-NN) (BCN = 15)	Sensitivity	75,00	**82,50**	78,33	76,67	73,33	78,33	76,67	75,83
	Specificity	86,00	84,67	**89,33**	80,67	88,00	86,67	87,33	85,33
	Gmean	80,31	83,58	**83,65**	78,64	80,33	82,39	81,83	80,44
	Precision	81,08	81,15	**85,45**	76,03	83,02	82,46	82,88	80,53
	F-measure	77,92	**81,82**	81,74	76,35	77,88	80,34	79,65	78,11
	AUC	80,50	83,58	**83,83**	78,67	80,67	82,50	82,00	80,58
	CA	81,11	83,70	**84,44**	78,89	81,48	82,96	82,59	81,11

(The best results are presented as bold numbers, H = hidden node number, BCN = base classifier number).

Table 6
Optimum values for variable parameters (on PD).

Classifier	H	5	10	15	20	25	30	35	40
PSO-NN (H = 25)	Sensitivity	91,84	89,12	**99,32**	89,12	93,20	87,76	95,92	97,28
	Specificity	52,08	52,08	43,75	**60,42**	**60,42**	45,83	33,33	39,58
	Gmean	69,16	68,13	65,92	73,38	**75,04**	63,42	56,54	62,05
	Precision	85,44	85,06	84,39	87,33	**87,82**	83,23	81,50	83,14
	F-measure	88,52	87,04	**91,25**	88,22	90,43	85,43	88,13	89,66
	AUC	71,96	70,60	71,53	74,77	**76,81**	66,79	64,63	68,43
	CA	82,05	80,00	**85,64**	82,05	85,13	77,44	80,51	83,08
ABC-NN (H = 25)	Sensitivity	94,56	91,16	92,52	92,52	88,44	94,56	91,84	**95,92**
	Specificity	60,42	62,50	60,42	64,58	**79,17**	60,42	70,83	54,17
	Gmean	75,58	75,48	74,76	77,30	**83,67**	75,58	80,65	72,08
	Precision	87,97	88,16	87,74	88,89	**92,86**	87,97	90,60	86,50
	F-measure	91,15	89,63	90,07	90,67	90,59	91,15	**91,22**	90,97
	AUC	77,49	76,83	76,47	78,55	**83,80**	77,49	81,34	75,04
	CA	86,15	84,10	84,62	85,64	86,15	86,15	**86,67**	85,64
ScPSO-NN (H = 30)	Sensitivity	97,28	93,88	95,92	91,84	89,80	89,12	93,20	91,16
	Specificity	20,83	62,50	45,83	58,33	64,58	**68,75**	41,67	41,67
	Gmean	45,02	76,60	66,30	73,19	76,15	**78,27**	62,32	61,63
	Precision	79,01	88,46	84,43	87,10	88,59	**89,73**	83,03	82,72
	F-measure	87,20	**91,09**	89,81	89,40	89,19	89,42	87,82	86,73
	AUC	59,06	78,19	70,88	75,09	77,19	**78,93**	67,43	66,41
	CA	78,46	**86,15**	83,59	83,59	83,59	84,10	80,51	78,97
Classifier	**BCN**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
RF (NN) (BCN = 30)	Sensitivity	**92,52**	91,16	87,76	90,48	90,48	91,84	89,80	87,07
	Specificity	64,58	64,58	62,50	64,58	62,50	68,75	68,75	**70,83**
	Gmean	77,30	76,73	74,06	76,44	75,20	**79,46**	78,57	78,54
	Precision	88,89	88,74	87,76	88,67	88,08	90,00	89,80	**90,14**
	F-measure	90,67	89,93	87,76	89,56	89,26	**90,91**	89,80	88,58
	AUC	78,55	77,87	75,13	77,53	76,49	**80,29**	79,27	78,95
	CA	85,64	84,62	81,54	84,10	83,59	**86,15**	84,62	83,08
RF (PSO-NN) (BCN = 35)	Sensitivity	97,28	95,92	95,92	97,28	97,28	96,60	94,56	**97,96**
	Specificity	27,08	52,08	52,08	52,08	54,17	56,25	**58,33**	50,00
	Gmean	51,33	70,68	70,68	71,18	72,59	73,71	**74,27**	69,99
	Precision	80,34	85,98	85,98	86,14	86,67	87,12	**87,42**	85,71
	F-measure	88,00	90,68	90,68	91,37	**91,67**	91,61	90,85	91,43
	AUC	62,18	74,00	74,00	74,68	75,72	76,42	**76,45**	73,98
	CA	80,00	85,13	85,13	86,15	**86,67**	**86,67**	85,64	86,15
RF (ABC-NN) (BCN = 25)	Sensitivity	97,96	99,32	99,32	96,60	95,92	**100,00**	97,96	93,88
	Specificity	60,42	52,08	50,00	47,92	**66,67**	50,00	54,17	56,25
	Gmean	76,93	71,92	70,47	68,03	**79,97**	70,71	72,84	72,67
	Precision	88,34	86,39	85,88	85,03	**89,81**	85,96	86,75	86,79
	F-measure	**92,90**	92,41	92,11	90,45	92,76	92,45	92,01	90,20
	AUC	79,19	75,70	74,66	72,26	**81,29**	75,00	76,06	75,06
	CA	**88,72**	87,69	87,18	84,62	**88,72**	87,69	87,18	84,62
RF (ScPSO-NN) (BCN = 5)	Sensitivity	**100,00**	96,60	99,32	99,32	93,88	97,28	99,32	97,96
	Specificity	54,17	43,75	47,92	47,92	**56,25**	52,08	50,00	43,75
	Gmean	**73,60**	65,01	68,99	68,99	72,67	71,18	70,47	65,47
	Precision	**86,98**	84,02	85,38	85,38	86,79	86,14	85,88	84,21
	F-measure	**93,04**	89,87	91,82	91,82	90,20	91,37	92,11	90,57
	AUC	**77,08**	70,17	73,62	73,62	75,06	74,68	74,66	70,85
	CA	**88,72**	83,59	86,67	86,67	84,62	86,15	87,18	84,62

(The best results are presented as bold numbers, H = hidden node number, BCN = base classifier number).

Before the comparison process, the optimum values of variable parameters were detected via TSS (7 metrics). The BLD, HS, PD, PID and WBC dataset trials are respectively shown in Table 3, Table 4, Table 5 Table 6, Table 7 and Table 8.

Literature suggests an idea about the subset number not being available. According to Table 3, in datasets with 6-to-9 feature numbers (with BLD, PID and WBC), RF structures operate sufficiently while the subset number is 4. But the optimum subset number cannot be generalized beyond the feature number of 9. In HS (12 features) and PD (22 features) datasets, the subset number changes unpredictably. *Thus, in the RF structure, it can be suggested that the assignment of the subset number as 4 is a conceivable choice for the datasets with a feature number less than 10.*

In trials related to the BLD dataset (Table 4), optimum node numbers of PSO-NN, ABC-NN and ScPSO-NN were found as 18, 18 and 20, respectively. Optimum base classifier numbers were 20, 10, 5 and 25 respectively for RF (NN), RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) architectures.

According to trials on the HS dataset (Table 5); optimum node numbers of PSO-NN, ABC-NN and ScPSO-NN were 16, 26 and 28, respectively. Optimum base classifier numbers were 35, 35, 5 and 15 respectively for RF (NN), RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) architectures.

On the PD dataset (Table 6); optimum node numbers were 25, 25 and 30 respectively for PSO-NN, ABC-NN and ScPSO-NN methods. Optimum base classifier numbers were 30, 35, 25 and 5

Table 7
Optimum values for variable parameters (on PID).

Classifier	H	6	8	10	12	14	16	18	20
PSO-NN (H = 20)	Sensitivity	49,25	50,00	52,99	56,72	54,48	56,34	**57,46**	55,97
	Specificity	85,40	83,60	83,80	83,20	84,00	81,80	80,20	**86,40**
	Gmean	64,86	64,65	66,63	68,69	67,65	67,89	67,89	**69,54**
	Precision	64,39	62,04	63,68	64,41	64,60	62,40	60,87	**68,81**
	F-measure	55,81	55,37	57,84	60,32	59,11	59,22	59,12	**61,73**
	AUC	67,33	66,80	68,39	69,96	69,24	69,07	68,83	**71,19**
	CA	72,79	71,88	73,05	73,96	73,70	72,92	72,27	**75,78**
ABC-NN (H = 20)	Sensitivity	45,90	56,34	55,22	58,96	61,94	56,34	59,33	**62,69**
	Specificity	**87,60**	86,00	81,60	82,00	83,20	83,00	78,00	83,40
	Gmean	63,41	69,61	67,13	69,53	71,79	68,38	68,03	**72,31**
	Precision	66,49	**68,33**	61,67	63,71	66,40	63,98	59,11	66,93
	F-measure	54,30	61,76	58,27	61,24	64,09	59,92	59,22	**64,74**
	AUC	66,75	71,17	68,41	70,48	72,57	69,67	68,66	**73,04**
	CA	73,05	75,65	72,40	73,96	75,78	73,70	71,48	**76,17**
ScPSO-NN (H = 8)	Sensitivity	54,85	**63,81**	48,88	55,22	47,01	59,33	55,60	45,52
	Specificity	83,40	83,40	79,80	81,40	**87,20**	79,40	83,20	**87,20**
	Gmean	67,64	**72,95**	62,46	67,05	64,03	68,63	68,01	63,00
	Precision	63,91	**67,32**	56,47	61,41	66,32	60,69	63,95	65,59
	F-measure	59,04	**65,52**	52,40	58,15	55,02	60,00	59,48	53,74
	AUC	69,13	**73,60**	64,34	68,31	67,11	69,36	69,40	66,36
	CA	73,44	**76,56**	69,01	72,27	73,18	72,40	73,57	72,66
Classifier	**BCN**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
RF (NN) (BCN = 10)	Sensitivity	56,72	66,42	64,18	**66,79**	64,93	64,18	53,73	**66,79**
	Specificity	84,40	80,60	78,80	78,20	81,80	82,40	**87,60**	78,20
	Gmean	69,19	**73,17**	71,11	72,27	72,88	72,72	68,61	72,27
	Precision	66,09	64,73	61,87	62,15	65,66	66,15	**69,90**	62,15
	F-measure	61,04	**65,56**	63,00	64,39	65,29	65,15	60,76	64,39
	AUC	70,56	**73,51**	71,49	72,50	73,36	73,29	70,67	72,50
	CA	74,74	75,65	73,70	74,22	75,91	**76,04**	75,78	74,22
RF (PSO-NN) (BCN = 20)	Sensitivity	52,61	**64,18**	55,97	56,72	57,46	59,33	60,07	57,84
	Specificity	81,40	73,20	81,60	**86,60**	82,40	79,60	82,60	81,60
	Gmean	65,44	68,54	67,58	70,08	68,81	68,72	**70,44**	68,70
	Precision	60,26	56,21	61,98	**69,41**	63,64	60,92	64,92	62,75
	F-measure	56,18	59,93	58,82	**62,42**	60,39	60,11	62,40	60,19
	AUC	67,01	68,69	68,79	**71,66**	69,93	69,46	71,34	69,72
	CA	71,35	70,05	72,66	**76,17**	73,70	72,53	74,74	73,31
RF (ABC-NN) (BCN = 10)	Sensitivity	**58,21**	57,84	53,73	54,85	54,85	54,10	53,73	52,99
	Specificity	85,80	86,60	87,00	88,00	87,80	87,60	**88,40**	88,20
	Gmean	70,67	**70,77**	68,37	69,48	69,40	68,84	68,92	68,36
	Precision	68,72	69,82	68,90	71,01	70,67	70,05	**71,29**	70,65
	F-measure	63,03	**63,27**	60,38	61,89	61,76	61,05	61,28	60,55
	AUC	72,00	**72,22**	70,37	71,43	71,33	70,85	71,07	70,59
	CA	76,17	**76,56**	75,39	76,43	76,30	75,91	76,30	75,91
RF (ScPSO-NN) (BCN = 35)	Sensitivity	58,21	**61,94**	49,25	58,21	54,48	55,97	56,34	55,60
	Specificity	85,20	82,60	**90,80**	85,60	86,60	84,60	90,20	86,80
	Gmean	70,42	**71,53**	66,87	70,59	68,69	68,81	71,29	69,47
	Precision	67,83	65,61	74,16	68,42	68,54	66,08	**75,50**	69,30
	F-measure	62,65	63,72	59,19	62,90	60,71	60,61	**64,53**	61,70
	AUC	71,70	72,27	70,03	71,90	70,54	70,29	**73,27**	71,20
	CA	75,78	75,39	76,30	76,04	75,39	74,61	**78,39**	75,91

(The best results are presented as bold numbers, H = hidden node number, BCN = base classifier number).

respectively for RF (NN), RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) structures.

With regard to the results on the PID dataset (Table 7); optimum node numbers were 20, 20 and 8 respectively for PSO-NN, ABC-NN and ScPSO-NN techniques. Optimum base classifier numbers of RF (NN), RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) were 10, 20, 10 and 35, respectively.

As seen in trials on the WBC dataset (Table 8); optimum node numbers of PSO-NN, ABC-NN and ScPSO-NN were found as 6, 18 and 4, respectively. Optimum base classifier numbers were 15, 40, 15 and 10 respectively for RF (NN), RF (PSO-NN), RF (ABC-NN) and RF (ScPSO-NN) structures.

3.2.2. Performance comparison based on 7 metrics

We acquired optimum operation conditions at Section 3.3.1. In this part, 7 metrics based TSS values (TSS (7 metrics)) were brought together in every dataset for all classifiers. Table 9 shows the performance comparison based on TSS (7 metrics).

According to Table 9;

With the BLD dataset, the best TSS value was 5 belonging to the RF (ScPSO-NN) structure. With the HS dataset, the RF (ScPSO-NN) obtained the best scores on 4/7 of the statistical metrics meaning that the TSS of the RF (ScPSO-NN) is 4. With PD, the best TSS was 3 obtained by RF (ScPSO-NN) and RF (ABC-NN) architectures. With PID, RF (ScPSO-NN) and RF (NN) achieved the 3 best scores in 7 metrics. With WBC data, RF (ScPSO-NN) was the best again with a TSS value of 3. In brief, RF (ScPSO-NN) obtained the best scores on 18/25 of the trials with 5 datasets, while the second best algorithm (RF (ABC-NN)) achieved the best scores on 7 of 25 different conditions. As a result, it is obvious that if a comparison based on TSS (7 metrics) is realized, RF (ScPSO-NN) is the best structure among other options in binary-medical pattern classification.

Table 8
Optimum values for variable parameters (on WBC).

Classifier	H	4	6	8	10	12	14	16	18	20
PSO-NN (H = 6)	Sensitivity	97,49	**99,58**	81,17	94,14	95,82	96,23	94,14	97,49	94,56
	Specificity	96,62	95,95	94,14	96,62	95,72	95,95	96,62	96,17	**96,85**
	Gmean	97,05	**97,75**	87,42	95,37	95,77	96,09	95,37	96,83	95,70
	Precision	93,95	92,97	88,18	93,75	92,34	92,74	93,75	93,20	**94,17**
	F-measure	95,69	**96,16**	84,53	93,95	94,05	94,46	93,95	95,30	94,36
	AUC	97,06	**97,76**	87,66	95,38	95,77	96,09	95,38	96,83	95,70
	CA	96,93	**97,22**	89,60	95,75	95,75	96,05	95,75	96,63	96,05
ABC-NN (H = 18)	Sensitivity	93,72	97,07	93,72	95,82	95,40	94,56	93,31	**97,49**	95,82
	Specificity	96,17	95,27	95,95	96,40	96,40	96,17	**96,62**	96,40	95,95
	Gmean	94,94	96,17	94,83	96,11	95,90	95,36	94,95	**96,94**	95,88
	Precision	92,95	91,70	92,56	93,47	93,44	93,00	**93,70**	93,57	92,71
	F-measure	93,33	94,31	93,14	94,63	94,41	93,78	93,50	**95,49**	94,24
	AUC	94,95	96,17	94,83	96,11	95,90	95,37	94,96	**96,94**	95,88
	CA	95,31	95,90	95,17	96,19	96,05	95,61	95,46	**96,78**	95,90
ScPSO-NN (H = 4)	Sensitivity	**98,33**	95,40	50,21	95,82	91,63	95,40	96,65	95,40	95,40
	Specificity	96,40	**96,85**	96,17	**96,85**	96,62	96,62	96,62	96,17	95,72
	Gmean	**97,36**	96,12	69,49	96,33	94,09	96,01	96,64	95,78	95,56
	Precision	93,63	94,21	87,59	**94,24**	93,59	93,83	93,90	93,06	92,31
	F-measure	**95,92**	94,80	63,83	95,02	92,60	94,61	95,26	94,21	93,83
	AUC	**97,36**	96,12	73,19	96,33	94,13	96,01	96,64	95,78	95,56
	CA	**97,07**	96,34	80,09	96,49	94,88	96,19	96,63	95,90	95,61
Classifier	**BCN**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	
RF (NN) (BCN = 15)	Sensitivity	93,72	96,23	**96,65**	94,98	95,82	**96,65**	94,14	**96,65**	
	Specificity	97,30	96,17	97,30	97,30	**97,52**	96,62	97,30	96,85	
	Gmean	95,49	96,20	**96,97**	96,13	96,67	96,64	95,71	96,75	
	Precision	94,92	93,12	95,06	94,98	**95,42**	93,90	94,94	94,29	
	F-measure	94,32	94,65	**95,85**	94,98	95,62	95,26	94,54	95,45	
	AUC	95,51	96,20	**96,98**	96,14	96,67	96,64	95,72	96,75	
	CA	96,05	96,19	**97,07**	96,49	96,93	96,63	96,19	96,78	
RF (PSO-NN) (BCN = 40)	Sensitivity	95,40	97,07	97,49	97,49	97,49	97,07	97,49	**98,33**	
	Specificity	96,62	**97,30**	97,07	96,62	96,85	97,07	97,07	96,85	
	Gmean	96,01	97,18	97,28	97,05	97,17	97,07	97,28	**97,58**	
	Precision	93,83	**95,08**	94,72	93,95	94,33	94,69	94,72	94,38	
	F-measure	94,61	96,07	96,08	95,69	95,88	95,87	96,08	**96,31**	
	AUC	96,01	97,18	97,28	97,06	97,17	97,07	97,28	**97,59**	
	CA	96,19	97,22	97,22	96,93	97,07	97,07	97,22	**97,36**	
RF (ABC-NN) (BCN = 15)	Sensitivity	**97,91**	96,23	**97,91**	96,65	96,23	95,82	94,98	97,07	
	Specificity	95,95	**97,52**	97,07	97,30	96,40	97,07	97,30	96,62	
	Gmean	96,92	96,88	**97,49**	96,97	96,32	96,44	96,13	96,85	
	Precision	92,86	**95,44**	94,74	95,06	93,50	94,63	94,98	93,93	
	F-measure	95,32	95,83	**96,30**	95,85	94,85	95,22	94,98	95,47	
	AUC	96,93	96,88	**97,49**	96,98	96,32	96,44	96,14	96,85	
	CA	96,63	97,07	**97,36**	97,07	96,34	96,63	96,49	96,78	
RF (ScPSO-NN) (BCN = 10)	Sensitivity	97,91	**98,74**	98,33	97,07	97,91	96,23	96,65	97,91	
	Specificity	96,62	96,62	96,40	96,40	96,62	**97,52**	96,17	96,62	
	Gmean	97,26	**97,68**	97,36	96,73	97,26	96,88	96,41	97,26	
	Precision	93,98	94,02	93,63	93,55	93,98	**95,44**	93,15	93,98	
	F-measure	95,90	**96,33**	95,92	95,28	95,90	95,83	94,87	95,90	
	AUC	97,26	**97,68**	97,36	96,73	97,26	96,88	96,41	97,26	
	CA	97,07	**97,36**	97,07	96,63	97,07	97,07	96,34	97,07	

(The best results are presented as bold numbers, H = hidden node number, BCN = base classifier number).

After the 7 metrics based TSS comparison was completed, we wanted to enforce the possibilities and realize a comparison based on the AUC metric. The area under the ROC Curve (AUC) was chosen among other metrics according for the fact that AUC was the second most preferred metric in the pattern classification. In this part, only the AUC metric of classifiers was compared. For this reason, the AUC values (stated in Table 9) were individually described in graphs for every dataset. Fig. 4 shows the AUC based comparison.

According to Fig. 4;

On BLD, HS and WBC datasets, the RF (ScPSO-NN) structure achieved the best AUC percentages (69.09%, 83.83% and 97.68%, respectively). With PID, RF (NN) obtained the highest AUC (73.51%). With PD, RF (ABC-NN) took the first place with an AUC value of 81.29%. Finally, the TSS of RF (ScPSO-NN) is 3. Moreover, ScPSO outperforms other base classifiers (NN, PSO-NN, ABC-NN) on 4 medical datasets (except the WBC dataset). As seen in AUC based comparison, an optimal choice, obtaining the best AUC values on all trials, is not available. However, RF (ScPSO-NN) obtains the TSS value of 3 among five trials that brings this method into the forefront more than others.

After the AUC based TSS comparison (TSS (AUC)), the CA metric was used for comparing the classification methods, that owed to its being the most preferable one in the pattern classification area. Fig. 5 shows the classification performances via the CA results taken from Table 9.

According to Fig. 5;

Table 9
Performance comparison of classifier systems (based on TSS).

Datasets	Technique	NN	PSO-NN	ABC-NN	ScPSO-NN	RF (NN)	RF (PSO-NN)	RF (ABC-NN)	RF (ScPSO-NN)
BLD	Sensitivity	42,07	46,21	45,52	51,03	44,14	32,41	48,28	**55,17**
	Specificity	84,50	83,50	83,00	82,00	84,00	**92,50**	80,00	83,00
	Gmean	59,62	62,12	61,46	64,69	60,89	54,76	62,15	**67,67**
	Precision	66,30	67,00	66,00	67,27	66,67	**75,81**	63,64	70,18
	F-measure	51,48	54,69	53,88	58,04	53,11	45,41	54,90	**61,78**
	AUC	63,28	64,85	64,26	66,52	64,07	62,46	64,14	**69,09**
	CA	66,67	67,83	67,25	68,99	67,25	67,25	66,67	**71,30**
TSS		0	0	0	0	0	2	0	**5**
HS	Sensitivity	72,50	78,33	80,00	**80,83**	80,00	78,33	76,67	78,33
	Specificity	78,67	81,33	82,67	83,33	85,33	88,67	**90,67**	89,33
	Gmean	75,52	79,82	81,32	82,07	82,62	83,34	83,37	**83,65**
	Precision	73,11	77,05	78,69	79,51	81,36	84,68	**86,79**	85,45
	F-measure	72,80	77,69	79,34	80,17	80,67	81,39	81,42	**81,74**
	AUC	75,58	79,83	81,33	82,08	82,67	83,50	83,67	**83,83**
	CA	75,93	80,00	81,48	82,22	82,96	84,07	**84,44**	**84,44**
TSS		0	0	0	1	0	0	3	**4**
PD	Sensitivity	85,03	93,20	91,16	89,12	91,84	**100,00**	95,92	**100,00**
	Specificity	**68,75**	60,42	66,67	**68,75**	**68,75**	45,83	66,67	54,17
	Gmean	76,46	75,04	77,96	78,27	79,46	67,70	**79,97**	73,60
	Precision	89,29	87,82	89,33	89,73	90,00	84,97	89,81	86,98
	F-measure	87,11	90,43	90,24	89,42	90,91	91,88	92,76	**93,04**
	AUC	76,89	76,81	78,91	78,93	80,29	72,92	**81,29**	77,08
	CA	81,03	85,13	85,13	84,10	86,15	86,67	**88,72**	**88,72**
TSS		1	0	0	1	1	1	**3**	**3**
PID	Sensitivity	61,19	55,97	62,69	63,81	**66,42**	56,72	57,84	56,34
	Specificity	82,60	86,40	83,40	83,40	80,60	86,60	86,60	**90,20**
	Gmean	71,10	69,54	72,31	72,95	**73,17**	70,08	70,77	71,29
	Precision	65,34	68,81	66,93	67,32	64,73	69,41	69,82	**75,50**
	F-measure	63,20	61,73	64,74	65,52	**65,56**	62,42	63,27	64,53
	AUC	71,90	71,19	73,04	**73,60**	73,51	71,66	72,22	73,27
	CA	75,13	75,78	76,17	76,56	75,65	76,17	76,56	**78,39**
TSS		0	0	0	1	**3**	0	0	**3**
WBC	Sensitivity	94,56	**99,58**	97,49	98,33	96,65	98,33	97,91	98,74
	Specificity	96,85	95,95	96,40	96,40	**97,30**	96,85	97,07	96,62
	Gmean	95,70	**97,75**	96,94	97,36	96,97	97,58	97,49	97,68
	Precision	94,17	92,97	93,57	93,63	**95,06**	94,38	94,74	94,02
	F-measure	94,36	96,16	95,49	95,92	95,85	96,31	96,30	**96,33**
	AUC	95,70	97,76	96,94	97,36	96,98	97,59	97,49	**97,68**
	CA	96,05	97,22	96,78	97,07	97,07	**97,36**	**97,36**	**97,36**
TSS		0	2	0	0	2	1	1	**3**

(The best results are presented as bold numbers).

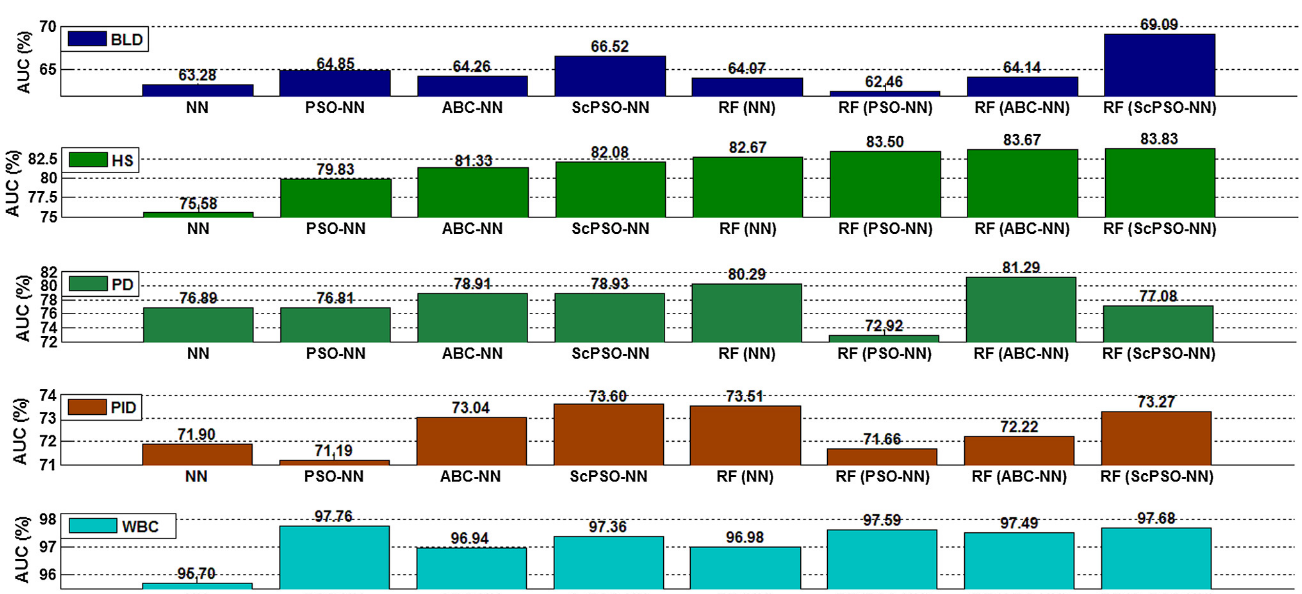

Fig. 4. Comparison based on AUC.

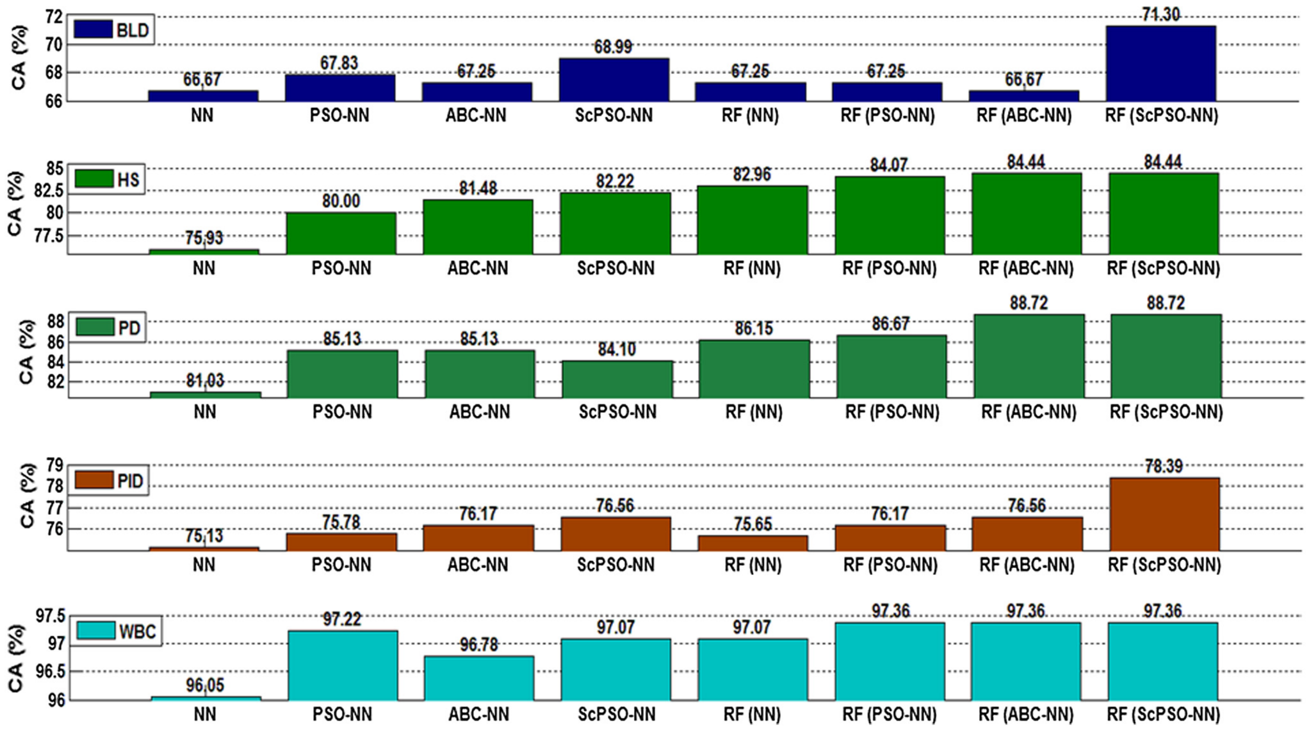

Fig. 5. Comparison based on CA.

On BLD and PID datasets, the best performance was obtained with RF (ScPSO-NN). Similarly, RF (ABC-NN) and RF (ScPSO-NN) methods shared the highest CA rates in HS, PD and WBC datasets. In addition, the performance of RF (PSO-NN) was equal to RF (ABC-NN) and RF (ScPSO-NN) architectures in WBC. Consequently, the TSS value of RF (ScPSO-NN) was 5, while the TSS of the second best structure (RF (ABC-NN)) was 3. Among base classifiers, ScPSO-NN achieves better results with BLD, HS and PID datasets. At this point, it should be specified that CA rates of RF (ScPSO-NN) never falls behind of other techniques (NN, hybrid NNs and other RF ensembles) on all trials. In brief, RF (ScPSO-NN) is the most robust one among others on account of TSS (CA) based comparison.

4. Discussions

RF uses Decision Tree classifiers and achieves approximately the same performance with hybrid classifiers. In this study, the hybrid NN based RFs were preferably generated, since these hybrid algorithms could easily outperform with a base classifier like DT. For this purpose, NN, PSO-NN, ABC-NN and ScPSO-NN algorithms were settled in the base classifier unit of RF. Therefore, RF (NN) method was formed, and RF (PSO-NN), RF (ABC-NN), RF (ScPSO-NN) architectures were designed.

With PSO-NN, ABC-NN and ScPSO-NN architectures, optimization algorithms were used for updating the weight and bias parameters in NN. In other words, optimization algorithms were used instead of backpropagation algorithm.

The test process was realized with a three-step statistical comparison for an objective view. With this aim, TSSs (7 metrics), TSSs (AUC) and TSSs (CA) were calculated in five medical datasets for all classification structures.

According to experiments;

Performance of RF (ScPSO-NN) is generally prior to other structures on 7 metrics based comparison. Also its performance is better than others by means of obtained AUC values on average. More importantly, its performance never falls behind of other structures on CA based comparisons meaning that it is more robust than others to changeable conditions (to different datasets).

As stated in Section 1, our second aim was to compare the ScPSO with PSO and ABC algorithms in hybrid NNs and in hybrid NN based RF. In other words, second aim was to reveal that ScPSO prior to PSO and ABC algorithms is owed to providing coherence in a complex system. This experimentation was performed, and ScPSO demonstrated prior performance not only in NN, but also in RF (Hybrid NNs). Herein, if a detailed scan is achieved between hybrid classifiers via TSS (7 metrics), TSS (AUC) and TSS (CA), ScPSO-NN demonstrates a surpassing performance on average in comparison to other base classifiers because of its ability to hold a better optimization algorithm. Besides, it is more coherent with hybrid NN based RF than others, since RF (PSO-NN) and RF (ABC-NN) obtains worse results than hybrid NNs in some trials with CA based comparisons. Thus, we can infer that ScPSO technique is more convenient to use in hybrid NNs and in hybrid NN based RF. Furthermore, usage of RF (hybrid NNs) has a meaning if optimization method in NN is chosen as ScPSO.

However, our first aim was to test RF (Hybrid NNs) in order to obtain prior performance with a base classifier, hybrid classifiers, RF (NN) and literature studies. As seen in results, hybrid NN method plays a key role on performance of RF (hybrid NN) architecture. But, it's seen that all hybrid NNs are not reliable to use in RF (hybrid NN) method meaning that the useful optimization algorithm should be chosen in NN for the obtainment of a robust classifier. Otherwise, complex structure cannot outperform to hybrid NN or RF (NN) techniques. As a result, the first three parts of our first aim were realized and RF (ScPSO-NN) was found the optimum choice among others. At the last part of our first objective was accomplished as in indicated in Table 11.

As demonstrated in Table 10, RF (ScPSO-NN) is a promising architecture that exhibits a challenging performance not only in new hybrid methods, but also in other ensemble algorithms (Random Forest, Rotation Forest with DT, Derivatives of RF, MmGA-based ensemble optimizer).

Table 10
The comparison of RF (ScPSO-NN) with literature studies.

Dataset	Study	Classification structures	Test method	Classification accuracy (%)
BLD	Hsieh et al. (2014)	HRCNN	%50–%50 training-test split	65
	Aldape-Pérez et al. (2012)	AMBC	%50–%50 training-test split	65.40
	Aldape-Pérez et al. (2012)	Rotation Forest	%50–%50 training-test split	68.98
	Tao et al. (2015)	MCSSVM	5-fold CV	69.52
	Aldape-Pérez et al. (2012)	Random Forest	%50–%50 training-test split	70.14
	Shi et al. (2015)	FR-KECA	%50–%50 training-test split	70.44
	Li et al. (2011)	PM-SVM(Gaus)	200 data → train 145data → test	70.85
	This study	**RF (ScPSO-NN)**	**2-fold CV**	**71.30**
HS	Li and Leng (2015)	SAMA	10-fold CV	66.96
	Aldape-Pérez et al. (2012)	Random Forest	%50–%50 training-test split	79.25
	Ozcift and Gulten (2011)	RF classifier ensembles	10-fold CV	80.49
	Aldape-Pérez et al. (2012)	Rotation Forest	%50–%50 training-test split	80.74
	Kim et al. (2011)	WAVE	10-fold CV	82.71
	Aldape-Pérez et al. (2012)	AMBC	%50–%50 training-test split	83.33
	Tan et al. (2014)	MmGA-based ensemble optimizer	10-fold CV	83.68
	This study	**RF (ScPSO-NN)**	**2-fold CV**	**84.44**
PD	Hsieh et al. (2014)	HRCNN	%50–%50 training-test split	75
	Li and Leng (2015)	SAMA	10-fold CV	81.84
	Hsieh et al. (2014)	PFHRCNN	%50–%50 training-test split	82.4
	Ozcift and Gulten (2011)	RF classifier ensembles	10-fold CV	87.13
	This study	**RF (ScPSO-NN)**	**2-fold CV**	**88.72**
PID	Li and Leng (2015)	SAMA	10-fold CV	68.57
	Aldape-Pérez et al. (2012)	AMBC	%50–%50 training-test split	70.57
	Aldape-Pérez et al. (2012)	Random Forest	%50–%50 training-test split	74.34
	Ozcift and Gulten (2011)	RF classifier ensembles	10-fold CV	74.47
	Xiang et al. (2015)	GA-KNN	Leave-One-Out	74.9
	Couellan and Wang (2015)	Bi-level SVM	5-fold CV	75.07
	Li et al. (2011)	PM−SVM(Gaus)	700 data → train 68data → test	75.36
	Xiang et al. (2015)	BPSO-KNN	Leave-One-Out	75.5
	Xiang et al. (2015)	QBPSO-KNN	Leave-One-Out	76.1
	Tao et al. (2015)	MCSSVM	5-fold CV	76.11
	Xiang et al. (2015)	BGSA-KNN	Leave-One-Out	76.3
	Aldape-Pérez et al. (2012)	Rotation Forest	%50–%50 training-test split	76.82
	Xiang et al. (2015)	BIGSA-KNN	Leave-One-Out	77.9
	Je and Je (2011)	GEP-B	10-fold CV	78.12
	Shi et al. (2015)	FR-KECA	5-fold CV	78.12
	This study	**RF (ScPSO-NN)**	**2-fold CV**	**78.39**
WBC	Hsieh et al. (2014)	HRCNN	%50–%50 training-test split	90
	Hsieh et al. (2014)	PFHRCNN	%50–%50 training-test split	96
	Kim et al. (2011)	WAVE	10-fold CV	96.24
	Couellan and Wang (2015)	Bi-level SVM	5-fold CV	96.50
	Aldape-Pérez et al. (2012)	Random Forest	%50–%50 training-test split	96.92
	Zhang and Suganthan (2014)	Random Forests ensemble	10-fold CV	97.16
	Je and Je (2011)	GEP-B	10-fold CV	97.21
	Aldape-Pérez et al. (2012)	Rotation Forest	%50–%50 training-test split	97.21
	Tan et al. (2014)	MmGA-based ensemble optimizer	10-fold CV	97.21
	This study	**RF (ScPSO-NN)**	**2-fold CV**	**97.36**

(The best results are presented as bold numbers).

Table 11
Complete TSS comparison.

Statistical Test	Datasets	NN	PSO-NN	ABC-NN	ScPSO-NN	RF (NN)	RF (PSO-NN)	RF (ABC-NN)	RF (ScPSO-NN)
TSSs (7 metrics)	BLD								+
	HS								+
	PD							+	+
	PID								+
	WBC								+
TSS		0	0	0	0	0	0	1	**5**
TSSs (AUC)	BLD								+
	HS								+
	PD							+	
	PID					+			
	WBC								+
TSS		0	0	0	0	1	0	1	**3**
TSSs (CA)	BLD								+
	HS							+	+
	PD							+	+
	PID								+
	WBC						+	+	+
TSS		0	0	0	0	0	1	3	**5**
Sum of TSSs		0	0	0	0	1	1	5	**13**

5. Conclusions

In literature, there are a few studies proving that the usage of RF (NN) is more appropriate than the usage of RF (DT) (Koyuncu and Ceylan, 2013). Likewise, RF (NN) outperforms the PSO-NN algorithm as demonstrated in study of Koyuncu and Ceylan (2013). In this study, it has been revealed that the usage of ScPSO-NN is more appropriate than the usage of DT and NN methods in RF ensembles. Moreover, ScPSO was deemed to be more reliable to use in NN and NN based RFs. Table 11 is formed based on the obtained results as being the summary of this work.

As evidenced in Table 11, the best complete TSS value is 13 achieved by RF (ScPSO-NN) architecture. Thus, it is obvious that RF (ScPSO-NN) is more convenient to use than a base classifier, a hybrid classifier, or other hybrid-NN based RFs. Furthermore, it is concluded that the trade-off between complexity-performance can be realized via coherence that is achieved through the hybridization of RF and ScPSO-NN.

For biomedical determination and diagnosis, a challenging ensemble structure (RF (ScPSO-NN)) has been obtained achieving higher classification performance on five diverse medical datasets.

In future work, we wish to design a new ensemble classifier technique based on different base classifiers and various hybrid architectures.

References

Aldape-Pérez, M., Yáñez-Márquez, C., Camacho-Nieto, O., Argüelles-Cruz, A.J., 2012. An associative memory approach to medical decision support systems. Comput. Meth. Prog. Biol. 106, 287–307.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. 30, 1145–1159.

Castro-Rios, A., Nevarez-Sida, A., Tiro-Sanchez, M.T., Wacher-Rodarte, N., 2014. Triggering factors of primary care costs in the years following type 2 diabetes diagnosis in Mexico. Arch. Med. Res. 45, 400–408.

Couellan, N., Wang, W., 2015. Bi-level stochastic gradient for large scale support vector machine. Neurocomputing 153, 300–308.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.

Hsieh, Y.Z., Su, M.C., Wang, P.C., 2014. A PSO-based rule extractor for medical diagnosis. J. Biomed. Inform. 49, 53–60.

Huang, J., Ling, C.X., 2005. Using AUC and accuracy in evaluating learning algorithms. IEEE T. Knowl. Data Eng. 17, 299–310.

Je, J., Je, P., 2011. Experimental evaluation of two new GEP-based ensemble classifiers. Expert Syst. Appl. 38, 10932–10939.

Karaboga, D., Akay, B., 2009. A comparative study of Artificial Bee Colony algorithm. Appl. Math. Comput. 214, 108–132.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: IEEE International Conference on Neural Networks. IEEE, pp. 1942–1948.

Kim, H., Kim, H., Moon, H., Ahn, H., 2011. A weight-adjusted voting algorithm for ensembles of classifiers. J. Korean Stat. Soc. 40, 437–449.

Koyuncu, H., Ceylan, R., 2013. Artificial neural network based on rotation forest for biomedical pattern classification. In: 36th International Conference on Telecommunications and Signal Processing (TSP 2013). IEEE, pp. 581–585.

Koyuncu, H., Ceylan, R., 2015. Scout particle swarm optimization. In: Proceedings of 6th European Conference of the International Federation for Medical and Biological Engineering, Springer, pp. 82–85.

Li, D.C., Liu, C.W., Hu, S.C., 2011. A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. Artif. Intell. Med. 52, 45–52.

Li, J., Pen, Y., Liu, Y., Li, W., Jin, Y., Tang, Z., Duan, Y., 2014. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography–mass spectrometry. Clin. Chim. Acta 436, 59–67.

Li, Y., Leng, Q., 2015. Alternating multiconlitron: A novel framework for piecewise linear classification. Pattern Recogn. 48, 968–975.

Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O., 2009. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE T. Bio-Med. Eng. 56, 1015–1022.

Ozcift, A., Gulten, A., 2011. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput. Meth. Prog. Biol. 104, 443–451.

Rehm, J., Samokhvalov, A.V., Shield, K.D., 2013. Global burden of alcoholic liver diseases. J. Hepatol. 59, 160–168.

Rodriguez, J.J., Kuncheva, L., Alonso, C.J., 2006. Rotation forest: a new classifier ensemble method. IEEE T. Pattern Anal. 28, 1619–1630.

Shi, J., Jiang, Q., Mao, R., Lu, M., Wang, T., 2015. FR-KECA: Fuzzy robust kernel entropy component analysis. Neurocomputing 149, 1415–1423.

Shieh, S.-H., Hsieh, V.C.-R., Liu, S.-H., Chien, C.-R., Lin, C.-C., Wu, T.-N., 2014. Delayed time from first medical visit to diagnosis for breast cancer patients in Taiwan. J. Formosan Med. Assoc. 113, 696–703.

Tan, C.J., Lim, C.P., Cheah, Y.N., 2014. A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models. Neurocomputing 125, 217–228.

Tang, Y., Zhang, Y.Q., Chawla, N.V., Krasser, S., 2009. SVMs modeling for highly imbalanced classification. IEEE Syst. Man Cybern. B 39, 281–288.

Tao, J., Wang, S., Hu, W., 2015. Minimum class spread constrained support vector machine. Neurocomputing 151, 481–500.

Trible, F., Riedere, P., 2008. Biomarkers for early detection of Parkinson's disease: an essential challenge. Adv. Behav. Biol. 57, 35–49.

Wolberg, W.H., Mangasarian, O.L., 1990. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In: Proceedings of the National Academy of Sciences. PNAS, pp. 9193–9196.

Xiang, J., Han, X., Duan, F., Qiang, Y., Xiong, X., Lan, Y., Chai, H., 2015. A novel hybrid system for feature selection based on an improved gravitational search algorithm and k-NN method. Appl. Soft Comput. 31, 293–307.

Zhang, L., Suganthan, P.N., 2014. Random forests with ensemble of feature spaces. Pattern Recogn. 47, 3429–3437.

[dataset] Lichman, M., 2013. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets.html.