World competitive contest-based artificial neural network: A new class-specific method for classification of clinical and biological datasets

Zohre Arabi Bulaghi[a], Ahmad Habibizad Navin[b], Mehdi Hosseinzadeh[c,d,*], Ali Rezaee[a]

[a] Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[c] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam
[d] Mental Health Research Center, Psychosocial Health Research Institue, Iran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

ABSTRACT

Many data mining methods have been proposed to generate computer-aided diagnostic systems, which may determine diseases in their early stages by categorizing the data into some proper classes. Considering the importance of the existence of a suitable classifier, the present study aims to introduce an efficient approach based on the World Competitive Contests (WCC) algorithm as well as a multi-layer perceptron artificial neural network (ANN). Unlike the previously introduced methods, which each has developed a universal model for all different kinds of data classes, our proposed approach generates a single specific model for each individual class of data. The experimental results show that the proposed method (ANNWCC), which can be applied to both the balanced and unbalanced datasets, yields more than 76% (without applying feature selection methods) and 90% (with applying feature selection methods) of the average five-fold cross-validation accuracy on the 13 clinical and biological datasets. The findings also indicate that under different conditions, our proposed method can produce better results in comparison to some state-of-art meta-heuristic algorithms and methods in terms of various statistical and classification measurements. To classify the clinical and biological data, a multi-layer ANN and the WCC algorithm were combined. It was shown that developing a specific model for each individual class of data may yield better results compared with creating a universal model for all of the existing data classes. Besides, some efficient algorithms proved to be essential to generate acceptable biological results, and the methods' performance was found to be enhanced by fuzzifying or normalizing the biological data.

1. Background

Classification is a process of placing samples in the related groups. The clinical data classification which is an important step in the clinical research, can be used to diagnose the seriously life-threatening diseases such as heart diseases or diabetes [1].

Based on the total number of samples, the clinical datasets are divided into two types [2]: balanced and unbalanced. Each class has an almost identical number of samples in the balanced datasets in contrast to the unbalanced datasets, in which the number of samples of every class is not alike. To deliver better results, a classifier inserts samples into the class with the largest number of samples. Since the produced model is unsuitable for classes with a small number of samples, learners will have problems with classes that have small numbers of samples. To deal with such limitations, our proposed algorithm, ANNWCC, can not only generate a specific model for every one of the existing classes but

also apply to both the balanced and unbalanced data with several different class sizes.

There are many methods and algorithms which can be applied to solve the problem of classification. A learner such as an ANN, for example, can be applied to find the connection among samples in a specific class. Also, the researchers have proposed many new approaches to training artificial neural networks and have shown the usefulness of their approach in data classification. Leema et al. used the particle swarm optimization (PSO) algorithm to train the neural networks. For this purpose, first they applied their method to three datasets obtained from the University of California Irvine (UCI), then compared their acquired results with those delivered by the previously introduced methods. Although their proposed method seemingly had better performance compared to the other existing methods, it appears that the state-of-art meta-heuristic optimization algorithms had the ability to enhance its performance [3] Masoudi-Sobhanzade et al. introduced a

new optimization algorithm and showed that their optimization algorithms can outperform other algorithms [4].

In the present study, the works related to the clinical data classification fall into several main categories, including:

i) **The Supervised machine learning approaches**: These methods aim to develop a model based on the machine learning methods such as ANNs [5], support vector machines [6], Bayesian networks [7] etc. To generate the model, first the data are divided into two training and testing sets, then the machine is trained and validated using the training and testing sets, respectively. Although the mentioned methods produce remarkable results, the state-of-art techniques can enhance their performance. Our proposed method, which is a combination of an ANN and the WCC algorithm, belongs in this class.

ii) **The Unsupervised machine learning approaches**: In this class of related works, unlike the first group, the existing data do not include any labels; and based on the relationship among data, they are classified into different categories. In this category, The K-nearest neighbour (KNN) is one of the most popular techniques used for clustering [8], Our proposed method is also fall into this class.

iii) **The Semi-supervised machine learning approaches:** In the third group of machine learning techniques, some of the existing data include label whereas others not. A combination of supervised and unsupervised methods is usually employed for generating a model [9]. The ensemble techniques may also function properly in this category [10].

iv) **The Heuristic-based approaches**: This class of related works searches for some connections among the existing data and arranges data in groups using a heuristic algorithm. An efficient algorithm, proposed based on Gauss-Newton [11], a [12] combination of Expectation maximization (EM) and principal component analysis (PCA) [13], and an improved ReliefF algorithm [14], are several instances of works that belong to the fourth group. Although the mentioned methods have big advantages over the other ones and produce acceptable results, they have low performance compared with the meta-heuristic methods [15].

In order to present a suitable computer-aided diagnostic model, a 3-layer architecture based on artificial intelligence has been presented [16]. In this architecture the data received from the sensors, are classified in two categories of healthy and suspicious and if suspicious, finally the doctor will be informed for further examinations [16]. The suggested architecture, requires efficient methods for data classification, whose efficiency in suspicious case diagnosis could be enhanced with ANNWCC.

To ease the Identification of disease-causing genes a tool named GPS has been introduced, whose main job is to prioritize and detect the patients genes [17]. To detect the most important disease-causing genes, this tool implements functional linkage networks. The procedure first uses HIPPIE datasets and various algorithms to examine different gens, then the candidate genes are prioritized. Finally, the classification of prioritized genes is done based on scoring schemes [17].

TRIZ article has also presented a method for choosing promising genes, which is able to detect the genes that have the most biological relevance with the disease [18]. For formulation, TRIZ makes use of optimization search operators along with modified BAT algorithm. In this method, rMRMR is used as a filter approach, while SVM and MBA are used for classification approach. Hence, not only does it give a wide coverage in gene search space, it also finds the interaction between them [18].

To cure cancer patients, classified studies of cancer subgroups is required. In article [19], to classify the cancer molecular subtypes, in addition to SVM algorithms, optimization algorithms have also been used. The findings show efficient subsets obtained from miRNA and mRNA which are considered as diagnostic biomarkers and can classify

molecular cancer groups [19].

In order to carefully study the protein structure, which is of most important biological macromolecules, a step-by-step classification algorithm has been presented [20]. Instead of classes related to the primary structure of proteins, this algorithm uses spatial configuration of proteins by predicting the structural classes of the secondary structure of proteins. Thus, extracting features from secondary structure series leads to a more accurate classification than previous methods [20].

Since the collected data are of the utmost importance in the field of biology, and so are the studies which should be conducted based on these data [21,22], some related works have made efforts to generate the clinical and biological datasets [23–26]. Despite the existence of many datasets in the field of biology, the data pre-processing and their missing values [27–29] remain a challenge [30]. In addition to the cited literature works, some other works have investigated different methods from various perspectives [31,32] and discussed their advantages and disadvantages as well as limitations [33,34]. Such studies give a few new directions for future reference.

The method suggested in this article is a method to classify the clinical and biological datasets. Employing the presented classification method in above-mentioned articles, could lead to detection of probable biomarkers, which is more reliable than other methods.

The biggest weakness of the methods introduced in the above-mentioned studies is their low performance based on the classification criteria. To address such a limitation, the current study proposes a method, named ANNWCC, which is a combination of a multi-layer perceptron artificial neural network (ANN) and the world competitive contests (WCC) algorithm ANNWCC incorporates several steps, including: (i) the data pre-processing, in which the mentioned biological data are normalized or fuzzified and then passed to the learner, (ii) the data separating, in which the data are categorized and put in the corresponding groups according to the total number of distinct labels, (iii) the model generating, in which the proposed approach (ANNWCC), unlike the other methods that have developed a universal model for all of the data classes, generates a specific model for every individual group of data, and (iv) the data classifying, in which a sample is placed in the group with a minimum value of error between the sample and the group's model. If the values of error between a sample and three models are 0.01, 0.002, and 0.03, respectively, the sample will be placed in the second class.

The proposed method uses a multi-layer artificial neural network to generate a classifier model. ANNs, inspired by the biological neural network systems, are computational models that consist of several neurons connected to other neurons via synapses. The role of the synapses is to transmit signals from a neuron to another. An ANN is a framework or an algorithm, which can be applied to the machine learning practices such as the biological data classification [35]. Fig. 1 shows an example of ANN which includes two hidden layers, which are those between the input and output layers.

As can be seen from Fig. 1 (where B represents bias) there are two neurons in each hidden layer. The output of a neuron can be calculated using Eq. (1):

$$O_{i,j} = \sum_{k=1}^{n} W_{k,j} + B_j \tag{1}$$

where $O_{i,j}$ is the output of the jth neuron in the ith hidden layer, B and W are the bias and edge weights respectively, and $W_{k,j}$ is the weight of the edge (synapse) between the kth neuron in the $(i-1)$th hidden layer and the jth neuron in the ith hidden layer.

Training of an ANN, in which the values of synapses or edges are determined, is a non-deterministic polynomial (NP) problem. In the case of these problems, it is not possible to find an optimal solution in a polynomial-time order and for this reason, the heuristic and meta-heuristic methods such as the optimization algorithms are applied to the train of ANN training [4]. WCC, which has been inspired by the
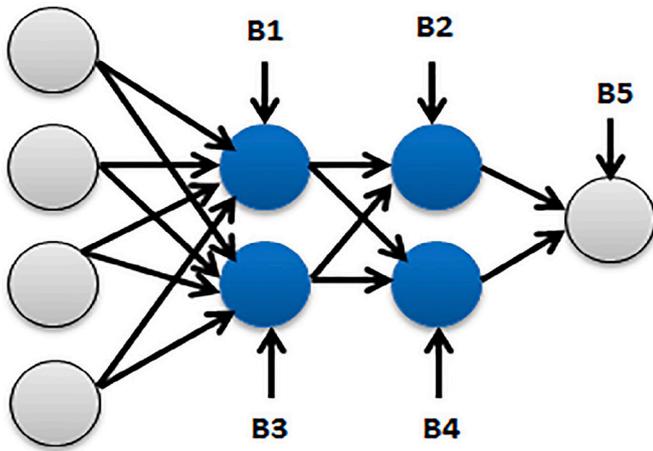
Fig. 1. An example of an ANN.

rules of human sports, is an optimization algorithm [36] with a meta-heuristic search strategy and shows a high level of performance compared to other optimization algorithms. The current study formulates the WCC algorithm to the training of the ANN problem and investigates and its contribution to classify of clinical and biological data.

The introduced method of ANNWCC can be applied in the third layer of the above-mentioned architecture.

2. Methods

ANNWCC, whose framework has been depicted in Fig. 2, is being applied at several stages. In the first step, the data pre-processing is carried out on a biological dataset. Normalization, in which the data of the groups are changed in the same range using Eq. (2), and fuzzification, which determines the degree to which each datum belongs to a group (Eq. (2)), are used at the data pre-processing stage.

$$nrvi = a + \frac{(vi - vmin) \times (b - a)}{vmax - vmin} \tag{2}$$

where a and b are the upper and lower bounds of $[a, b]$, in which the groups' data should be converted, and v_i, v_{min}, v_{max}, and nrv_i are the current values that should be normalized, the minimum value of the related group, the maximum value of the related group, and the normalized value, respectively.

In Eq. (3), which is a fuzzification equation, AVG, v_{min}, v_i, and fv_i are the average value of a group, the minimum value of a group, the current value which should be converted into a fuzzy value, and the fuzzified value, respectively. A group's data change in the range of [0,1] using Eq. (3).

$$fvi = \frac{|(vi - AVG)|}{(AVG - vmin)} \tag{3}$$

In the next step, ANNWCC divides a clinical or biological dataset into several groups based on the total number of class labels and puts every sample in its relevant group. After dividing the datasets of each class into training and testing sets, ANNWCC generates a model for each class of data using an ANN.

The number of the hidden layers in an ANN and their neurons pose a design challenge because the computation time will increase as the number of the hidden layers goes up. Thus, there should be a trade-off between these factors. Eqs. (4) and (5) show the way the number of the hidden layers and their neurons are determined in ANNWCC:

$$nh = \log(N) \tag{4}$$

$$nn_i = \log(nn_{i-1}), \ 1 < i \leq nh, \ nn_1 = N \tag{5}$$

Where N, nh, and nn_i are the number of neurons in the input layer,

the number of the hidden layers, and the number of neurons in the ith hidden layer, respectively.

2.1. The WCC-based ANN method

The ANNWCC training which is the final step toward generating the desired model, is an NP problem that cannot be solved within a reasonable period of time. Studies have shown that using optimization algorithms to train ANNs yields comparatively more acceptable results. We use WCC for ANN training due to its remarkable performance in the NP problems. By the approach employed in the present study, each team is considered a potential model for the clinical data classification. The number of players are computed using Eq. (6):

$$np = \sum_{i=1}^{nh} (nni \times nni - 1) + nni \tag{6}$$

Where np is the number of players on a team.

The score function which determines the suitability of a potential answer, is defined using Eq. (7). As shown in this equation, if a particular model in a relative class assigns the input samples to the closer output values, it will perform better than others. Eq. (7) is known as the root mean squared error (RMSE).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y - pi)^\wedge 2}{n}} \tag{7}$$

where n, y, and p_i are the number of samples in a class, the value assigned to a group's samples, and the predicted value of a sample, respectively.

ANNWCC generates an individual model for each existing class. After samples of a dataset are evaluated by every one of the created models, a test sample is labelled with a class number that reflects the closeness of its model's output to the mean of its class. Eq. (8) shows the way a test sample is assigned to a class.

$$L_i = \{k \ | \ (o_{i,k} - avg_k) \text{ is a minimum for all } 1 \leq k \leq m\} \tag{8}$$

where $o_{i,k}$, avg_k, L_i, and m are the kth model's output for the ith test sample, the average value of the training samples obtained from the kth model in the kth class, the label determined for the ith test sample, and the number of classes, respectively.

In the proposed method, parameters, including the total number of teams, groups, and attempts that improve and fulfil the scoring function, should be defined at the beginning of the algorithm. Like other optimization algorithms, WCC begins with the first population of the potential answers named *Teams*. Every team, which is generated randomly, is considered a candidate solution that determines the weights of the ANN's edges. Then, the generated teams are randomly placed in some groups which have equal numbers of teams. Teams compete with each other and improve their scores (RMSE) through four operations. The WCC algorithm consists of four operations, including shooting, attacking, passing, and crossing which change the candidate solution or the weights of an ANN and introduce some new potential answers. In this study, shooting, attacking, passing, and crossing operations have been formulated using Eqs. (9), (10), (11) and (12), respectively.

$$\sum_{i=1}^{k} (Team_i(v1) = Team_j(v2)) \tag{9}$$

Where k and *Team* are a random integer value between 1 and n (the total number of the ANN's edges) and a candidate solution that determines the weights of an ANN. $v1$ and $v2$ are two integer values between 1 and n.

$$\sum_{i=1}^{k} (Team_i(rand(n)) = rand(RNG)) \tag{10}$$

Where *RNG* determines the range of values assigned to variables or
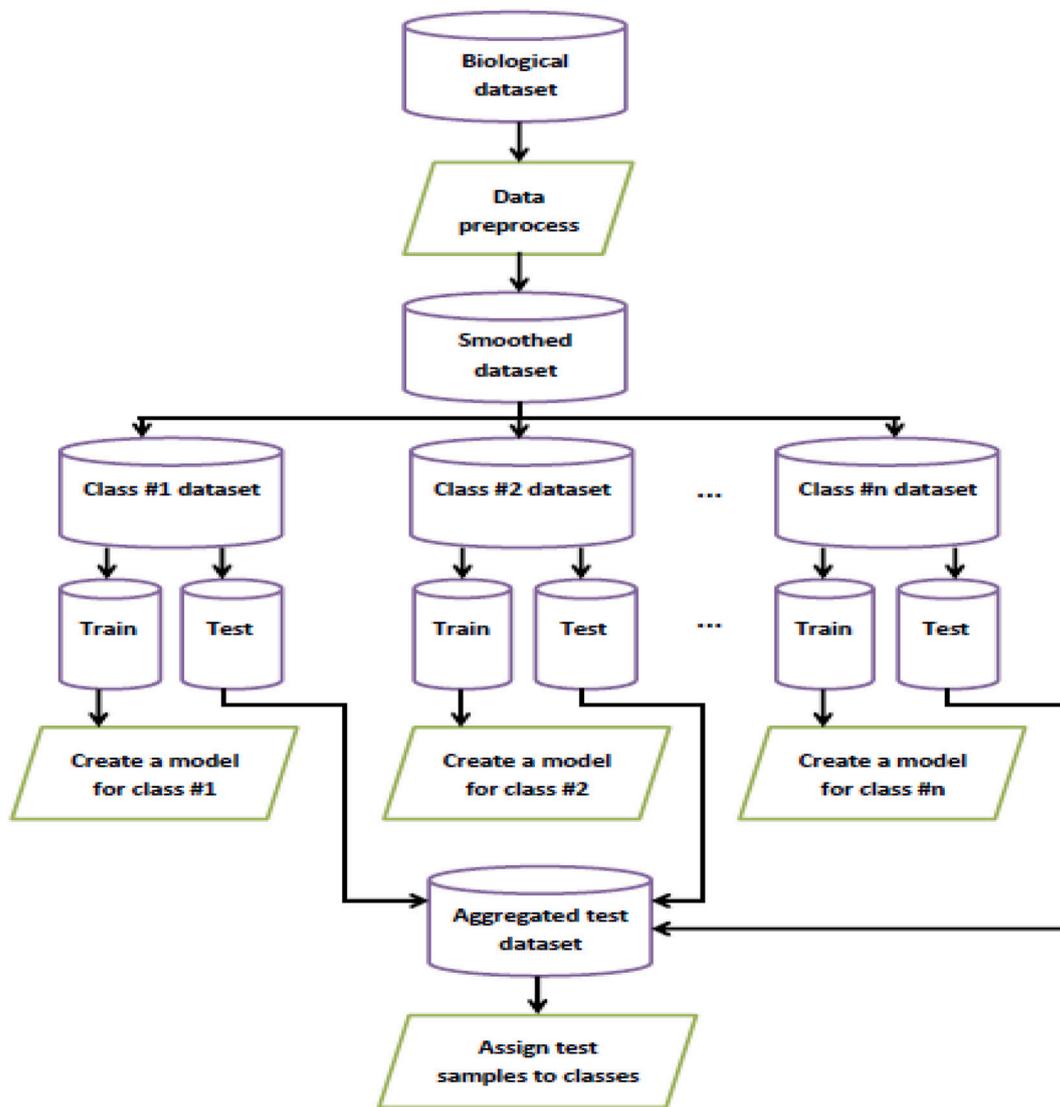
Fig. 2. The framework of the proposed method.

edges of ANN.

$$\sum_{i=1}^{k} \left(Team_i(v1) = Team_i(v2) \&\& Team_i(v2) = Team_i(v1) \right) \tag{11}$$

$$\sum_{i=m}^{k} \left(Team_i(m) = Team_i(m + 1) \right) \tag{12}$$

Where k and m ($k > m$) are two integer values between 1 and n.

In the proposed algorithm, there is a referee who evaluates the effect of an action on a team and keeps scores. If an operation enhances the team's score, the new changes will be accepted by that team; otherwise, they will be ignored. Matches are held at two stages: 1) group matches; and 2) elimination matches. At the end of the elimination stage, a champion is selected as an answer to the problem. This scenario, in which individual iterations are called seasons, will be repeated until an acceptable response is produced. The above-mentioned operations which change a team's values have been illustrated in Fig. 3. After holding the group competitions, teams with a higher score ascend to the elimination stage and continue their competition. Like the grouping stage, the elimination stage is held based on the mentioned operations. At the end of the elimination phase, the remaining team is considered to be the answer to the problem [36].

3. Results

We implemented a multi-layer perceptron ANN in the MATLAB programing language on a system with 12 GB of RAM, corei7 CPU, and Windows 10.1 operating system. To train an ANN, we used the WCC algorithm, a discrete symbiotic optimization search (DSOS) [37], and the PSO algorithms [38]. To carry out this experiment, we used 13 clinical and biological datasets obtained from the UCI Machine Learning Repository [39]. Table 1 presents the following information about these datasets: the number of references to each dataset (REF), the name of the datasets, the number of instances (NOI), the number of features (NOF) without considering the class labels, the number of classes (NOC), the data types (DT), and the missing values (MV). The Cleveland Heart Disease (CHD) dataset was prepared in two ways: as the CHD with two classes (CHD2) and the CHD with five categories (CHD5). The CHD2 was generated by merging the risk level of heart disease and heart disease class labels. The Statlog Heart Disease (SHD) dataset consists of two classes, indicating the presence and absence of heart disease. The Pima Indian Diabetes (PID) dataset has been made up of two groups of people those with diabetes and those without diabetes. Hepatitis (HEP), Liver Disorders (LIV), Lung Cancer (LUNG), Parkinson's Disease (PAR), and the Wisconsin Diagnostic Breast Cancer (WDBC) datasets are the other ones whose features have been also
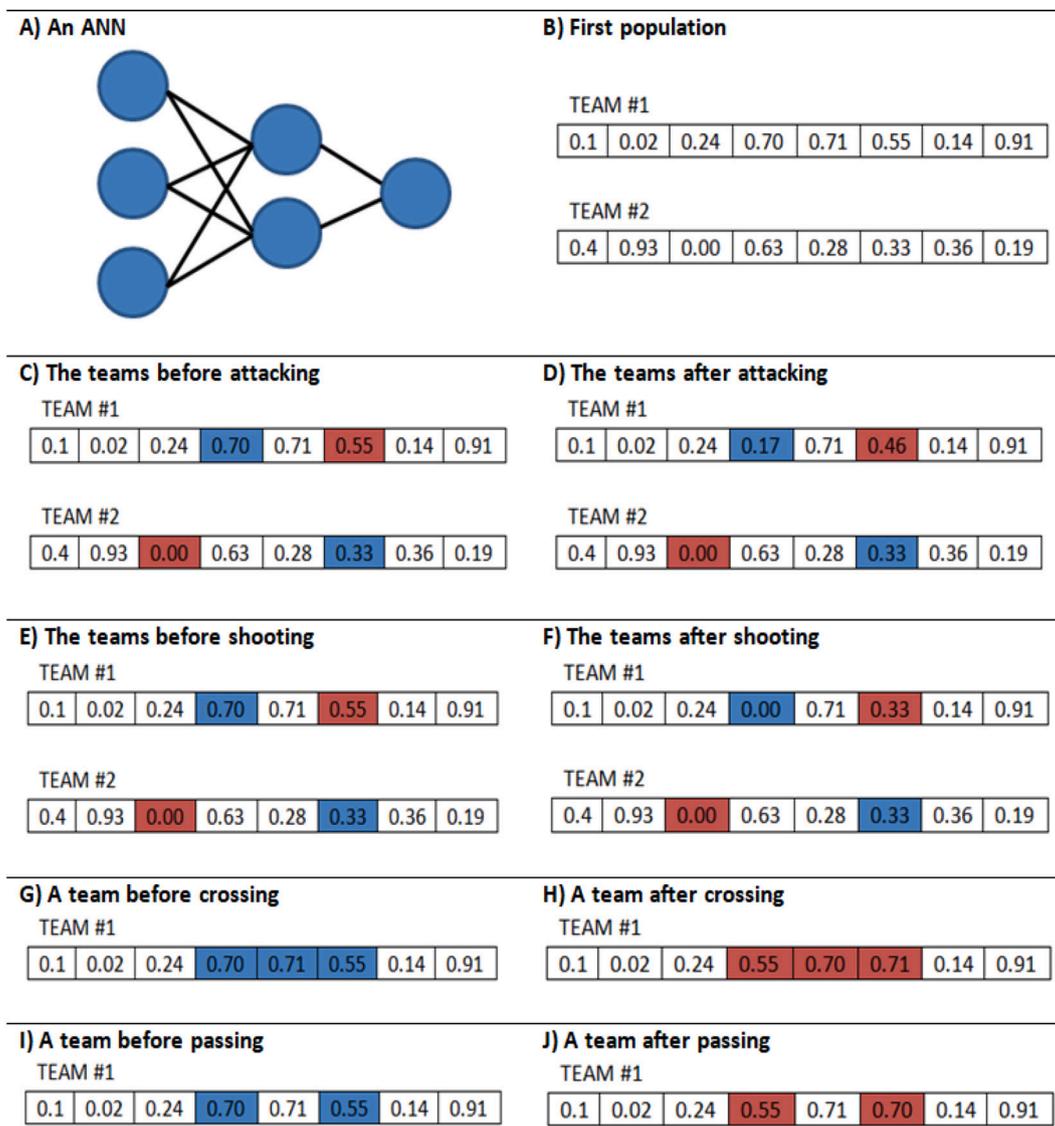
Fig. 3. An example of ANN training. The changes are accepted if they reduce RMSE.
(a) An ANN with one hidden layer and 8 edges is considered.
(b) The first population consists of two teams whose players are equal to the number of ANN's edges. Every player determines a weight for the specified edge.
(c) The status of the teams before the attacking operators are shown. Team #2 is the attacker team and selects two players randomly.
(d) The attacker team adds random values to its selected values and sends them toward to randomly selected players in Team #1.
(e) Team #2 is the shooter and selects two players randomly.
(f) The shooter team sends its selected values toward Team #1.
(g) The status of a team before crossing. In this operator, the permutation of values in a specified range is changed.
(h) The status of Team #1 is presented after crossing.
(i) the status of a team before passing. In this operator, two players who are chosen randomly change their values.
(j) The status of Team #1 after passing.

presented in Table 1.

In the first experiment, we investigated the effect of the input data type on the ANN training stage, and then compared the three algorithms' performance on the CHD2, the normalized CHD2 (NCHD2), and the fuzzified CHD2 (FCHD2) datasets. It was found that an optimization algorithm with a faster convergence speed and similar results obtained from its different executions is the most suitable algorithm. Convergence means that an optimization algorithm can yield an optimal answer as the dedicated time or the number of iterations increases. Since optimization algorithms usually have operators that change the candidate solutions randomly, the produced answers may vary to some extent. However, different executions may yield similar answers. To address such a problem, it is recommended that

optimization algorithms should be executed at least 30 individual times [52]. The variable behaviour of an optimization algorithm in its different executions should be measured from the stability perspective. An optimization algorithm with similar output values in its different executions functions better compared to others. For the purpose of classification, the convergence and stability were examined in terms of accuracy (ACC), a correlation (COR) between actual and predicted labels, and RMSE [53]. Based on the mentioned criteria, an evaluation of the WCC, DSOS, and PSO algorithms was made in Figs. 4 & 5 and Table 2, respectively.

Every algorithm was executed 30 times, with 50 seasons per execution. The best-obtained results have been illustrated in Fig. 4, which shows the convergence of the algorithms on the PID dataset and

Table 1
Datasets used.

REF	Name	NOI	NOF	NOC	DT	MV
[40]	CHD2	303	13	2	Numerical and binary	NO
[40]	CHD5	303	13	5	Numerical and binary	NO
[41]	SHD	270	13	2	Numerical and binary	NO
[42]	PID	768	8	2	numerical	NO
[43]	HEP	150	19	2	Numerical and binary	YES
[44]	LIV	345	6	2	Numerical and binary	NO
[45]	LUNG	32	56	3	Numerical and binary	YES
[46]	PAR	197	22	2	Real	YES
[47]	WDBC	569	31	2	Real	NO
[48]	ARCENE	900	10,000	2	Numerical	NO
[49]	ARRYTH	452	279	16	Double	YES
[50]	PARKINSON	756	754	2	Numerical and binary	NO
[51]	GENEEXPR	801	20,531	5	Double	NO

demonstrates that all the three algorithms converge successfully on the non-preprocessed, fuzzified, and normalized PID datasets. The lower is the value of RMSE, the more effective is the algorithm. Although the model generated by each algorithm is different from the ones produced by the other two algorithms, every algorithm's performance is remarkable on the PID datasets their RMSE value is almost zero. As indicated in Fig. 4, when data pre-processing is not carried out, the performance of the algorithms will be low in terms of accuracy and correlation. The WCC algorithm achieves 100% of accuracy and correlation on both the fuzzified and normalized datasets, whereas the DSOS and PSO algorithms reach 100% of classification accuracy and correlation only on the normalized datasets. Fig. 5 shows the stability of the algorithms in 30 individual executions.

Fig. 5 illustrates the performance of the algorithms from the stability perspective in terms of error, accuracy, and correlation on the non-preprocessed, fuzzified, and normalized PID datasets. Although WCC, PSO, and DSOS all show acceptable as well as similar results, WCC outperforms others. Figs. 4 and 5 show that the pre-processing stage is essential in the ANNWCC's training and improves the performance of the algorithms. The normalized datasets yield better results on the PID datasets compared with the other dataset types.

To investigate the performance of the algorithms on all the three types of PID datasets in detail, an evaluation of the algorithms' performance on the training and testing datasets with different sizes was made. To this end, the data were divided into various parts, and the methods were applied to them. The results have been shown in Table 2. As observed earlier, once again algorithms produced acceptable results even with a minimum size of the training datasets. In comparison to PSO and DSOS, the WCC algorithm yielded better results. It was revealed that DSOS and PSO function approximately in a similar way. The results reported by the current investigation, show that the proposed method can be a suitable option for a dataset with a small size of samples. However, the performance of the obtained models can be enhanced by applying various feature selection approaches.

In the third experiment, we evaluated the algorithms' performance on all the datasets. For this purpose, the datasets were divided into five folds. In the 5 iterations, one of the folds was used as the testing set and the other ones as the training sets. The algorithms were ran in 30 individual executions on the three types of data: non-pre-processed, fuzzified, and normalized. The results have been shown in Table 3. This table presents the name of the algorithms (AN), the type of datasets on which the algorithms have maximum performance, the elapsed time (ET) of the algorithms which is presented in seconds, accuracy (ACC), the confidence interval (CI) which determines the accuracy range with a specific probability, and the p-value. Since the order of the input has been changed to calculate the *p*-value and CI, the cross-validation results were reported in different runs. Accuracy, specificity (SPC), sensitivity (SEN), and the false-positive rate (FPR) were calculated using Eqs. (13) through (16), where n is the number of classes, m is the

number of samples, and TP_i, TN_i, FP_i and FN_i are the true positive, true negative, false positive, and false-negative rates of the ith class, respectively.

$$ACC = \frac{\sum_{i=1}^{n} ((TP_i + TN_i)/m)}{n} \tag{13}$$

$$SPC = \frac{\sum_{i=1}^{n} ((TN_i)/(TN_i + FP_i))}{n} \tag{14}$$

$$SEN = \frac{\sum_{i=1}^{n} ((TP_i)/(TP_i + FN_i))}{n} \tag{15}$$

$$FPR = 1 - SPC \tag{16}$$

As indicated in Table 3, the performance of WCC, which is approximately the same for both the normalized and fuzzified datasets, is better than the other two algorithms. Besides, PSO and DSOS deliver better results only on the normalized data sets. Also, as evidenced by Table 3, algorithms demonstrate distinct performance on different data types; while DSOS produces the best results on the CDH5 and HEP datasets, PSO yields the most desirable effects on the PAR dataset. Also, WCC delivers the most appropriate results on the CHD2, LUND, and WDBC datasets. In the case of the SHD, PID, and LIV datasets, each algorithm performs in much the same way as the other two algorithms do. In general, although all of the algorithms function properly, compared to DSOS and PSO, WCC seems to produce the most desired results in terms of CI, *P*-value, and the elapsed time.

Table 4 summarizes and compares the three algorithms based on the above-mentioned criteria which have been computed using the "ANOVA 1" function (MATLAB) with the default parameters. As shown in Table 4, although the results produced by the other two algorithms are suitable, once again WCC delivers the most favourable results on all the datasets. In the case of DSOS and PSO, although they produce similar results, DSOS' performance is slightly better than PSO's. Since feature selection (FS) is one of the primary data pre-processing steps which may help produce better results, the effects of FS on the performance of the proposed method was also investigated. For this purpose, after the desired features were selected using the FeatureSelect software application [15], they were applied to the methods and the final models were generated using the selected features. The obtained datasets, called the FS-based datasets, together with their average results, have been shown in Table 4.

The receiving operation characteristic (ROC) curve is a graphical method for determining the separation ability of the models in the samples of classes [54]. An algorithm with a curve close to the top left-hand side of the graph has a high ability to separate the data. Based on the average of the best 30 individual executions of algorithms on all the datasets, Fig. 6 is represented. As it is shown in Fig. 6, although the ROC curves of all the three algorithms show their remarkable separation ability, WCC outperforms DSOS and PSO in its ability to separate the data. Moreover, the separation ability of DSOS is better than that of PSO.

The vertical axis of the ROC curves marks the sensitivity of the algorithms, which is computed using Eq. (16), and the horizontal axis presents the false positive rate (FPR) which is equal to $(1 - SPC)$ and is calculated by Eq. (16). The Purple, blue, red, and yellow curves show the ROC curves of the random guess, WCC, DSOS, and PSO algorithms, respectively. Although all the three algorithms' ROC curves are approximately in harmony/agreement, the area under the WCC's curve (AUC) is larger than the ones under the other two algorithms' curves.

The PID, CHD, and SHD datasets, whose attributes and properties such as the total number of samples were presented in Table 1, have been frequently used in various studies. In the final experiment, we made a comparison between the results of some previously done studies in the literature works and those of the three algorithms described in this paper. The research references, publication dates, authors' names, methods' names, and the best value for the accuracy obtained by each

a)
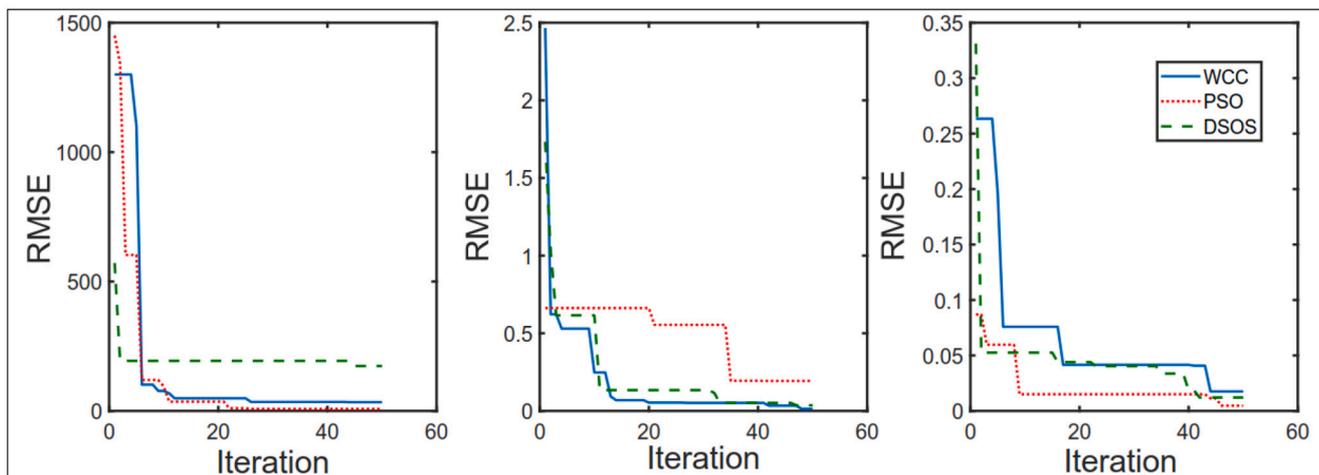
b)
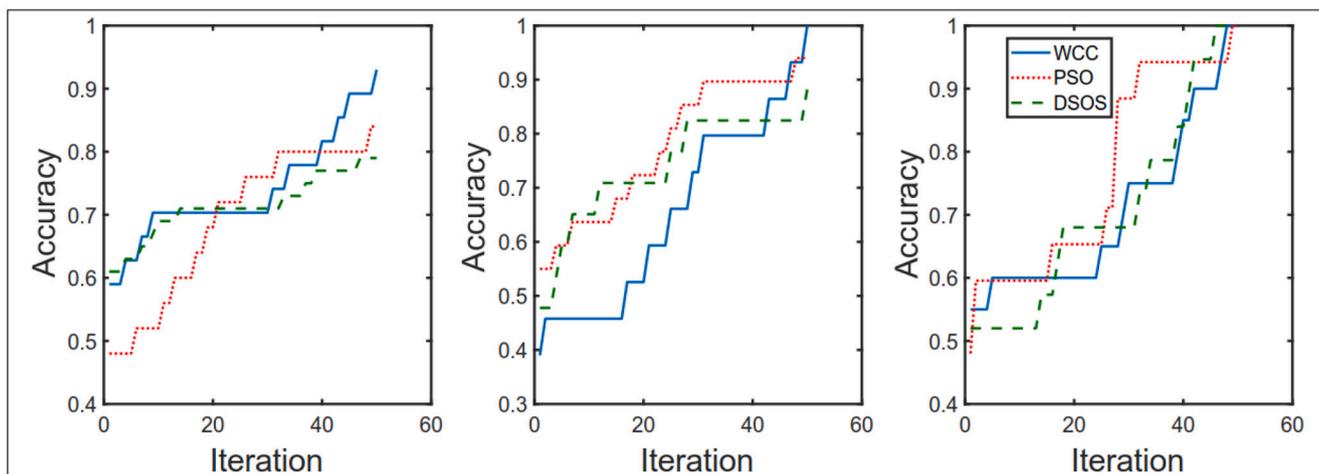
c)
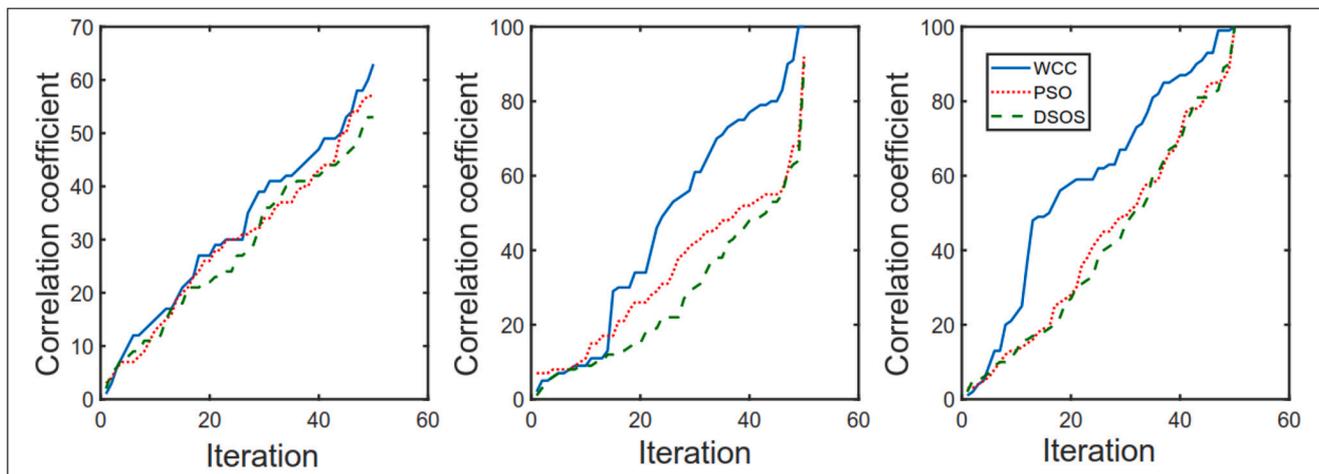
Fig. 4. An evaluation of the algorithms' performance based on the RMSE, accuracy, and correlation. The WCC, PSO and DSOS algorithms are shown by the blue, red and green curves, respectively. The first, second and third columns show the results obtained for the non- preprocessed PID, fuzzified PID and normalized PID datasets, respectively. (a) Comparison of convergence based on the RMSE criterion. (b) Comparison of convergence based on the accuracy criterion. (c) Comparison of convergence based on the correlation criterion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a)

b)

c)
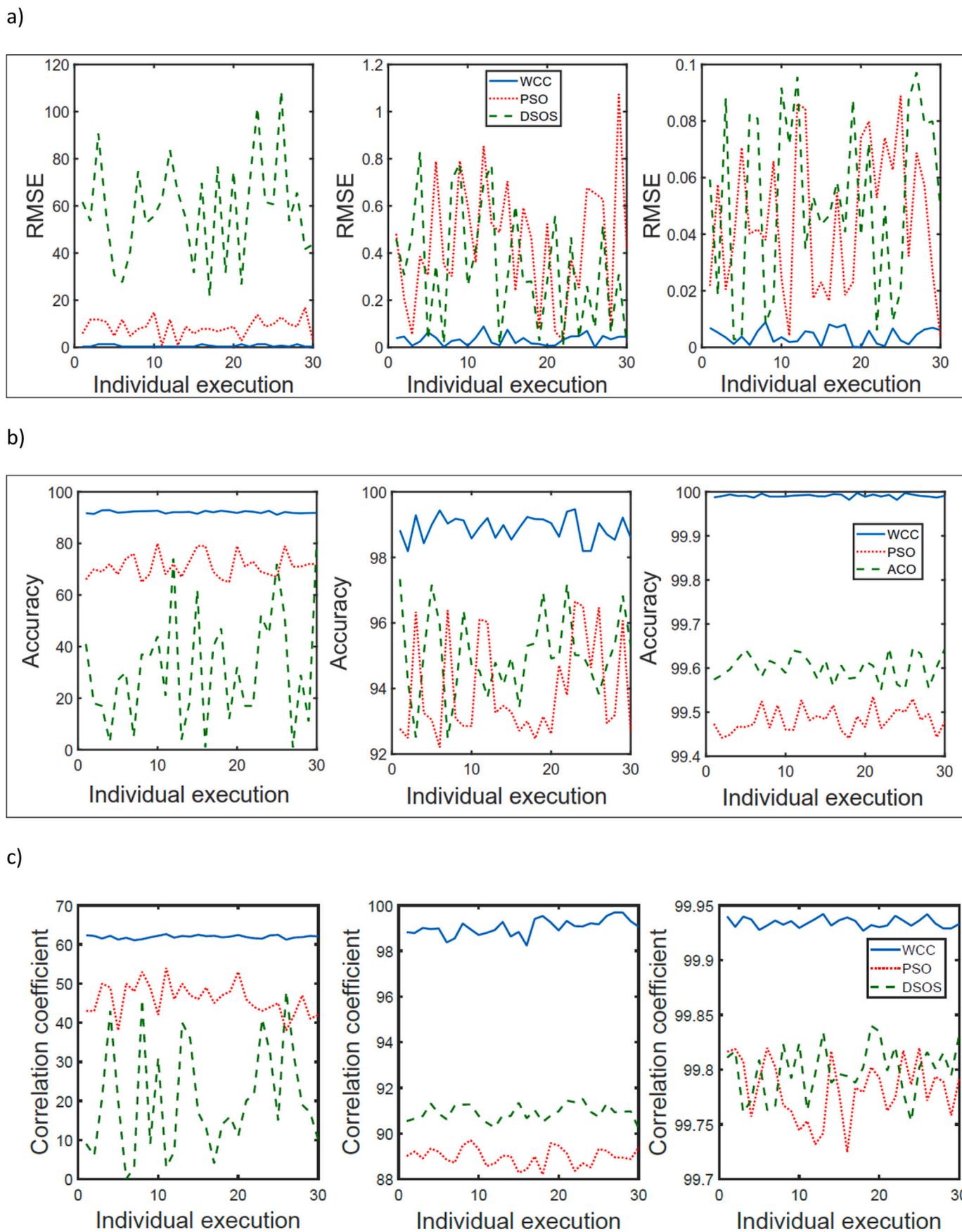
Fig. 5. Evaluation of the algorithms based on the RMSE, accuracy and correlation criteria. WCC, PSO, and DSOS are shown by the blue, red and green curves, respectively. The first, second and third columns show the results obtained for the non-preprocessed, fuzzified and normalized PID datasets, respectively. (a) Comparison of stability based on the RMSE criterion, (b) Comparison of stability based on the accuracy criterion. (c) Comparison of stability based on the correlation criterion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
An evaluation of the algorithms' performance on the training and testing datasets with different sizes.

Dataset	AL_name	ACC	ACC	ACC	ACC
		Train = 20% Test = 80%	Train = 40% Test = 60%	Train = 60% Test = 40%	Train = 80% Test = 20%
No preprocessed PID	WCC	62	94	94	94
	PSO	60.19	60.19	87.96	83
	DSOS	52.31	74.04	79	79
Fuzzified PID	WCC	62.96	100	100	100
	PSO	62.96	80.56	93	93
	DSOS	52.78	76.39	89	89
Normalized PID	WCC	80.56	100	100	100
	PSO	64.81	93.06	100	100
	DSOS	71.76	100	100	100

Table 3
Comparison of algorithms' performance on all of the datasets.

	NA	Type	ET	ACC %	CI	P_value
CHD2	WCC	normalized	80.01	93.25	[91 92]	1.6E-15
	PSO	Normalized	139.91	85.90	[77 81]	0.003
	DSOS	normalized	238.64	90.10	[72 89]	0.025
CHD5	WCC	Normalized and fuzzified	238	71.5	[68 69]	9.2E-11
	PSO	fuzzified	412.08	72.5	[69 70]	5.3E-7
	DSOS	fuzzified	1015.11	73.5	[71 72]	1.7E-15
SHD	WCC	Normalized and fuzzified	97.77	96.5	[95 96]	6.1E-65
	PSO	normalized	63.58	96.5	[94 96]	4.1E-23
	DSOS	normalized	333.46	96.5	[94 96]	4.5E-29
PID	WCC	normalized	50.07	97.13	[95 97]	3.64E-92
	PSO	normalized	119.55	96.84	[95 96]	9.2E-105
	DSOS	normalized	242.88	96.84	[95 96]	8.8E-109
HEP	WCC	normalized	409.63	71.83	[70 71]	3.7E-65
	PSO	normalized	204.76	77.6	[62 65]	3.9E-12
	DSOS	normalized	866.36	72.55	[70 71]	4.4E-25
LIV	WCC	Normalized and fuzzified	305	73.13	[70 72]	1.5E-15
	PSO	normalized	240.45	71.63	[66 70]	0.00001
	DSOS	normalized	1337.08	72.13	[65 69]	0.0098
PAR	WCC	normalized	421.9	68.47	[65 66]	6.8E-71
	PSO	normalized	156.05	69.12	[62 67]	0.00055
	DSOS	normalized	504.26	66.62	[63 65]	0.00002
LUNG	WCC	Normalized and fuzzified	813.44	71.23	[69 70]	4.8E-93
	PSO	fuzzified	1497.14	53.66	[52 52]	9.6E-11
	DSOS	fuzzified	3603.6	63.08	[54 59]	0.0037
WDBC	WCC	normalized	76.19	86.1	[83 84]	5.9E-33
	PSO	normalized	57.74	85.6	[78 82]	0.0004
	DSOS	normalized	196.24	80.35	[77 79]	3.1E-6
ARCENE	WCC	normalized	11,367.94	61.18	[57 59]	3.3E-11
	PSO	normalized	3609.55	62.11	[57 60]	2.2E-4
	DSOS	normalized	17,512.47	60.02	[57 59]	3.6E-9
ARRYTH	WCC	fuzzified	1074.63	74.3	[71 72]	8.8E-21
	PSO	fuzzified	569.66	73.9	[69 71]	4.6E-8
	DSOS	fuzzified	2441.01	75.67	[71 74]	2.11E-4
PARKINSON	WCC	fuzzified	13,598.26	57.3	[54 57]	2.55E-6
	PSO	fuzzified	7955.52	54.7	[52 53]	9.63E-29
	DSOS	fuzzified	26,304.08	56.44	[53 55]	3.7E-5
GENEEXPR	WCC	normalized	54,244.9	89.8	[85 87]	7.02E-4
	PSO	normalized	21,673.86	84.5	[81 83]	4.88E-4
	DSOS	normalized	56,412.33	81.14	[77 80]	1.2E-3

Table 4
Overall comparison of algorithms' performance.

	AN	ET	ACC %	CI	P_value
All datasets	WCC	6367.518	77.824	[75.69 75.46]	5.42E-5
	PSO	2823.065	75.735	[71.23 71.84]	3.67E-4
	DSOS	8539.04	75.764	[73.23 71.61]	3.07E-3
FS-based datasets	WCC	4005.55	90.003	[86.29 88.44]	3.66E-11
	PSO	1899.62	86.243	[83.03 85.85]	3.05E-11
	DSOS	5883.741	86.243	[83.96 84.43]	3.71E-11

DSOS outperformed the other previously proposed approaches.

4. Discussion

In this part, first, we make a comparison between ANNWCC and the other clinical data classification approaches, then discuss its capabilities from different perspectives:

1- The Applied method aspect: The proposed method, employed for classification, is different from the other introduced methods. Unlike ANNWCC, which is a combination of the designed ANN and the WCC algorithm, a majority of classification methods are based on either an ANN, which uses classical or modified training methods [71–77] or an optimization algorithm which determines a weighting method for the classification of data [78–81]. In contrast to the other approaches which introduce a universal model for all the different kinds of data [82–84], ANNWCC generates a specific model for each class of data.
2- The Dataset types: In contrast to the previously introduced works, which have focused on just a specific dataset and based their evaluation on a single data type [85–87], ANNWCC's performance / effect has been investigated on 13 different medical and biological datasets. ANNWCC is also operable on both the balanced and unbalanced datasets, while a majority of the other approaches have focused on just the balanced data [88–90]. In addition to the above-mentioned findings, it was revealed that there is a connection between the number of hidden layers and the computational time enhancing.
3- The Scalability aspect: Unlike the other approaches which have only one specific application, ANNWCC's results indicate that it can be used for various practical purposes, [91–94]. ANNWCC can be also employed as a reference in future studies.

5. Conclusion

In this paper, we proposed a new method based on a combination of ANN and the WCC algorithm to classify the clinical datasets. For the purpose of this classification, we employed several real datasets obtained from the UCI Machine Learning Repository and different literary

method on the above-mentioned datasets have been listed in Table 5. As shown in Table 5, the five-fold cross-validation results have been acquired under similar conditions.

The best values for each group have been presented in bold print. The findings revealed that the FELM method yielded the best results on the PID and CHD5 datasets as well as the WCC, PSO, and DSOS algorithms. WCC produced the best results on the CHD2 dataset. And finally, the results delivered by WCC, PSO, and DSOS on the SHD dataset were identical. In most cases, based on the findings, WCC, PSO, and

Fig. 6. ROC curves for the three algorithms on all the datasets.

Table 5
Comparison of the proposed method with other methods.

Reference	Date	Authors	Method name	Accuracy			
				PID	CHD5	CHD2	SHD
[55]	2013	Aslam et al	GP-KNN	80.50	–	–	–
[56]	2010	Patil et al	K-means with C4.5	92.38	–	–	–
[57]	2014	Alneamy et al	TLBO-FWNN	90.29	–	–	–
[58]	2014	Varma et al	Gini-Fuzzy	75.80	–	–	–
[59]	2014	Seera et al	FMNN-CART-RF	78.39	–	–	–
[60]	2012	Kalpana et al	FIS	90.38	–	–	–
[61]	2015	Christopher et al	PSO-C4.5	83.99	–	77.89	–
[62]	2010	Lee et al	Fuzzy Expert system	91.20	–	–	–
[63]	2014	Dennis et al	AGFS	89.90	76.67	–	–
[64]	2015	Subbulakshmi and Deepa	SRLPSO-ELM	93.09	–	91.33	89.96
[65]	2008	Bhatia	SVM-GA	–	72.55	90.57	–
[66]	2012	Mattila et al	Generic Statistical approach	78.04	–	85.01	–
[67]	2012	Anooj	Weighted fuzzy rule	–	–	62.35	–
[68]	2009	Setiawan et al	Fuzzy based roughest	–	–	83.00	–
[69]	2008	Kahramanli et al	Hybrid of AIS and ANN	–	–	87.40	–
[70]	2015	Nahato et al	RS-BPNN	–	–	–	90.40
[1]	2016	Kindie et al	FELM	92.54	**73.77**	93.55	94.44
–	–	Proposed method	ANNWCC	**97.13**	71.5	**94.5**	**96.5**
–	–	Proposed method	ANNPSO	96.84	72.5	87.40	**96.5**
–	–	Proposed method	ANNDSOS	96.84	73.5	91.20	**96.5**

works. We separated each group's data and generated a model for each one. Then we evaluated the performance of the algorithms on various experiments and described their functionality. Convergence, stability, error, the correlation between predicted and actual labels, accuracy, sensitivity, and specificity were the main criteria used for the comparison of the algorithms. Based on the above-mentioned criteria, the results produced by the WCC algorithm seemed to be better than those delivered by the other ones. We also compared the current studies' reported results on the PID, CHD, and SHD datasets with the results reported by the previously done studies on these datasets, and found out that the proposed method can achieve remarkable results in terms of the average value of accuracy for the PID and SHD datasets.

Author contribution statements

Zohre Arabi designed the project, prepared the manuscript, and performed the analysis. Ali Rezaee and Ahmad Habibizad Navin reviewed the manuscript and supervised the project. Mehdi Hosseinzadeh designed the project, prepared data for the analysis, reviewed the manuscript and supervised the project. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

None.

References

[1] K.B. Nahato, K.H. Nehemiah, A. Kannan, Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets, Inform. Med. Unlocked 2 (2016) 1–11.
[2] S. Shilaskar, A. Ghatol, P. Chatur, Medical decision support system for extremely imbalanced datasets, Inf. Sci. 384 (2017) 205–219.

[3] N. Leema, H.K. Nehemiah, A. Kannan, Neural network classifier optimization using differential evolution with global information and back propagation algorithm for clinical datasets, Appl. Soft Comput. 49 (2016) 834–844.

[4] Y. Masoudi-Sobhanzadeh, Y. Omidi, M. Amanlou, A. Masoudi-Nejad, Trader as a new optimization algorithm predicts drug-target interactions efficiently, Sci. Rep. 9 (2019) 9348.

[5] P. Hobson, B.C. Lovell, G. Percannella, M. Vento, A. Wiliem, Benchmarking human epithelial type 2 interphase cells classification methods on a very large dataset, Artif. Intell. Med. 65 (2015) 239–250.

[6] A.O. Kusakci, B. Ayvaz, E. Karakaya, Towards an autonomous human chromosome classification system using competitive support vector machines teams (CSVMT), Expert Syst. Appl. 86 (2017) 224–234.

[7] M.A. Morid, M. Fiszman, K. Raja, S.R. Jonnalagadda, G. Del Fiol, Classification of clinically useful sentences in clinical evidence resources, J. Biomed. Inform. 60 (2016) 14–22.

[8] S.G. Stark, S.L. Hyland, M.F. Pradier, K. Lehmann, A. Wicki, F.P. Cruz, J.E. Vogt, G. Rätsch, Unsupervised Extraction of Phenotypes from Cancer Clinical Notes for Association Studies, arXiv Preprint arXiv:1904.12973 (2019).

[9] Q. Wang, L.-Y. Xia, H. Chai, Y. Zhou, Semi-supervised learning with ensemble self-training for cancer classification, 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, 2018, pp. 796–803.

[10] G. Tang, J. Shi, W. Wu, X. Yue, W. Zhang, Sequence-based bacterial small RNAs prediction using ensemble learning strategies, BMC Bioinforma 19 (2018) 503.

[11] L. Dora, S. Agrawal, R. Panda, A. Abraham, Optimal breast cancer classification using gauss–Newton representation based algorithm, Expert Syst. Appl. 85 (2017) 134–145.

[12] J.S. Wekesa, J. Meng, Y. Luan, Multi-feature fusion for deep learning to predict plant lncRNA-protein interaction, Genomics (2020).

[13] M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, A knowledge-based system for breast cancer classification using fuzzy logic method, Telematics Inform. 34 (2017) 133–144.

[14] Y. Huang, P.J. McCullagh, N.D. Black, An optimization of ReliefF for classification in large datasets, Data Knowl. Eng. 68 (2009) 1348–1356.

[15] Y. Masoudi-Sobhanzadeh, H. Motieghader, A. Masoudi-Nejad, FeatureSelect: a software for feature selection based on machine learning approaches, BMC Bioinforma 20 (2019) 170.

[16] Z.A. Bulaghi, A.H.Z. Navin, M. Hosseinzadeh, A. Rezaee, SENET: a novel architecture for IoT-based body sensor networks, Inform. Med. Unlocked (2020) 100365.

[17] A. Meshkin, A. Shakery, A. Masoudi-Nejad, GPS: identification of disease genes by rank aggregation of multi-genomic scoring schemes, Genomics 111 (2019) 612–618.

[18] M.A. Al-Betar, O.A. Alomari, S.M. Abu-Romman, A TRIZ-inspired bat algorithm for gene selection in cancer classification, Genomics 112 (2020) 114–126.

[19] H. MotieGhader, Y. Masoudi-Sobhanzadeh, S.H. Ashtiani, A. Masoudi-Nejad, mRNA and microRNA selection for breast cancer molecular subtype stratification using meta-heuristic based algorithms, Genomics (2020).

[20] Y. Ge, S. Zhao, X. Zhao, A step-by-step classification algorithm of protein secondary structures based on double-layer SVM model, Genomics 112 (2020) 1941–1946.

[21] Y. Masoudi-Sobhanzadeh, Y. Omidi, M. Amanlou, A. Masoudi-Nejad, DrugR + : a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy, Comput. Biol. Med. 109 (2019) 254–262.

[22] Y. Masoudi-Sobhanzadeh, Computational-based drug repurposing methods in COVID-19, BioImpacts 10 (2020) 205.

[23] C.L. Gordon, C. Weng, Bacterial Clinical Infectious Diseases Ontology (BCIDO) Dataset, Data In Brief, 8 (2016), p. 881.

[24] Y. Masoudi-Sobhanzadeh, Y. Omidi, M. Amanlou, A. Masoudi-Nejad, Drug databases and their contributions to drug repurposing, Genomics 112 (2020) 1087–1095.

[25] J. Ma, J. Wang, L.S. Ghoraie, X. Men, R. Chen, P. Dai, Comprehensive expression-based isoform biomarkers predictive of drug responses based on isoform co-expression networks and clinical data, Genomics 112 (2020) 647–658.

[26] Y. Masoudi-Sobhanzadeh, A. Masoudi-Nejad, Synthetic repurposing of drugs against hypertension: a datamining method based on association rules and a novel discrete algorithm, BMC Bioinforma. 21 (2020).

[27] J.Y. Nancy, N.H. Khanna, K. Arputharaj, Imputing missing values in unevenly spaced clinical time series data to build an effective temporal classification framework, Comput. Stat. Data Anal. 112 (2017) 63–79.

[28] T. Usman, F. Hadlich, W. Demasius, R. Weikard, C. Kühn, Unmapped reads from cattle RNAseq data: a source for missing and misassembled sequences in the reference assemblies and for detection of pathogens in the host, Genomics 109 (2017) 36–42.

[29] Q. Song, W. Xu, W. Li, S. He, J. Liu, G. Wang, L. Ma, Accurate haplotype imputation with individualized ancestry-adjusted reference panels, Genomics 110 (2018) 329–335.

[30] M.A. Dziadzko, O. Gajic, B.W. Pickering, V. Herasevich, Clinical calculators in hospital medicine: availability, classification, and needs, Comput. Methods Prog. Biomed. 133 (2016) 1–6.

[31] M. Aldape-Pérez, C. Yáñez-Márquez, O. Camacho-Nieto, I. López-Yáñez, A.-J. Argüelles-Cruz, Collaborative learning based on associative models: application to pattern classification in medical datasets, Comput. Hum. Behav. 51 (2015) 771–779.

[32] M. Maghsoudloo, S.A. Jamalkandi, A. Najafi, A. Masoudi-Nejad, An efficient hybrid feature selection method to identify potential biomarkers in common chronic lung inflammatory diseases, Genomics (2020).

[33] L. Wang, P.J. Haug, G. Del Fiol, Using classification models for the generation of disease-specific medications from biomedical literature and clinical data repository, J. Biomed. Inform. 69 (2017) 259–266.

[34] X. Xu, X. Wei, H. Xie, Advances in methods and software for RNA cytosine methylation analysis, Genomics 112 (2020) 1840–1846.

[35] E. Pournoor, N. Elmi, Y. Masoudi-Sobhanzadeh, A. Masoudi-Nejad, Disease global behavior: a systematic study of the human interactome network reveals conserved topological features among categories of diseases, Inform. Med. Unlocked 17 (2019) 100249.

[36] Y. Masoudi-Sobhanzadeh, H. Motieghader, World competitive contests (WCC) algorithm: a novel intelligent optimization algorithm for biological and non-biological problems, Inform. Med. Unlocked 3 (2016) 15–28.

[37] A.E.-S. Ezugwu, A.O. Adewumi, Discrete symbiotic organisms search algorithm for travelling salesman problem, Expert Syst. Appl. 87 (2017) 70–78.

[38] J. Kennedy, Particle swarm optimization, Encyclopedia of Machine Learning, Springer, 2011, pp. 760–766.

[39] K. Bache, M. Lichman, UCI Machine Learning Repository, (2013).

[40] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K.H. Guppy, S. Lee, V. Froelicher, International application of a new probability algorithm for the diagnosis of coronary artery disease, Am. J. Cardiol. 64 (1989) 304–310.

[41] A.K. Seewald, Towards Understanding Stacking: Studies of a General Ensemble Learning Scheme, Na, (2003).

[42] J.W. Smith, J. Everhart, W. Dickson, W. Knowler, R. Johannes, Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, Proceedings of the Annual Symposium on Computer Application in Medical Care, American Medical Informatics Association, 1988, p. 261.

[43] B. Cestnik, I. Kononenko, I. Bratko, Assistant 86: A knowledge-elicitation tool for sophisticated users, Proceedings of the 2nd European Conference on European Working Session on Learning, Sigma Press, 1987, pp. 31–45.

[44] J. McDermott, R.S. Forsyth, Diagnosing a disorder in a classification benchmark, Pattern Recogn. Lett. 73 (2016) 41–43.

[45] Z.-Q. Hong, J.-Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Pattern Recogn. 24 (1991) 317–324.

[46] M.A. Little, P.E. McSharry, S.J. Roberts, D.A. Costello, I.M. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection, Biomed. Eng. Online 6 (2007) 23.

[47] W.N. Street, W.H. Wolberg, O.L. Mangasarian, Nuclear feature extraction for breast tumor diagnosis, Biomedical Image Processing and Biomedical Visualization, International Society for Optics and Photonics, 1993, pp. 861–870.

[48] I. Guyon, S. Gunn, M. Nikravesh, L.A. Zadeh, Feature Extraction: Foundations and Applications, Springer, Verlag Berlin Heidelberg, 2008.

[49] T. Vatanen, M.-S. Paukkeri, I.T. Nieminen, T. Honkela, Analyzing authors and articles using keyword extraction, self-organizing map and graph algorithms, Proceed. AKRR (2008) 105–111.

[50] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263.

[51] J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, C.G.A.R. Network, The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (2013) 1113.

[52] M. Črepinšek, S.-H. Liu, M. Mernik, Exploration and exploitation in evolutionary algorithms: a survey, ACM Comput. Surv. (CSUR) 45 (2013) 35.

[53] H. MotieGhader, S. Gharaghani, Y. Masoudi-Sobhanzadeh, A. Masoudi-Nejad, Sequential and mixed genetic algorithm and learning automata (SGALA, MGALA) for feature selection in QSAR, Iran. J. Pharm. Res. 16 (2017) 533.

[54] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577.

[55] M.W. Aslam, Z. Zhu, A.K. Nandi, Feature generation using genetic programming with comparative partner selection for diabetes classification, Expert Syst. Appl. 40 (2013) 5402–5412.

[56] B.M. Patil, R.C. Joshi, D. Toshniwal, Hybrid prediction model for type-2 diabetic patients, Expert Syst. Appl. 37 (2010) 8102–8108.

[57] J.S.M. Alneamy, R.A.H. Alnaish, Heart disease diagnosis utilizing hybrid fuzzy wavelet neural network and teaching learning based optimization algorithm, Adv. Artif. Neural Syst. 2014 (2014) 6.

[58] K.V. Varma, A.A. Rao, T.S.M. Lakshmi, P.N. Rao, A computational intelligence approach for a better diagnosis of diabetic patients, Comput. Electr. Eng. 40 (2014) 1758–1765.

[59] M. Seera, C.P. Lim, A hybrid intelligent system for medical data classification, Expert Syst. Appl. 41 (2014) 2239–2249.

[60] M. Kalpana, A. Kumar, Design and implementation of fuzzy expert system using fuzzy assessment methodology, Int. J. Sci. Appl. Inform. Technol. 1 (2012).

[61] J.J. Christopher, H.K. Nehemiah, A. Kannan, A swarm optimization approach for clinical knowledge mining, Comput. Methods Prog. Biomed. 121 (2015) 137–148.

[62] K. Vijaya, H. Khanna Nehemiah, A. Kannan, N. Bhuvaneswari, Fuzzy neuro genetic approach for predicting the risk of cardiovascular diseases, Int. J. Data Min. Modell. Manag. 2 (2010) 388–402.

[63] B. Dennis, S. Muthukrishnan, AGFS: adaptive genetic fuzzy system for medical data classification, Appl. Soft Comput. 25 (2014) 242–252.

[64] C. Subbulakshmi, S. Deepa, Medical dataset classification: a machine learning paradigm integrating particle swarm optimization with extreme learning machine classifier, Sci. World J. 2015 (2015).

[65] S. Bhatia, P. Prakash, G. Pillai, SVM based decision support system for heart disease

classification with integer-coded genetic algorithm to select critical features, Proceedings of the World Congress on Engineering and Computer Science, 2008, pp. 34–38.

[66] J. Mattila, J. Koikkalainen, A. Virkki, M. van Gils, J. Lötjönen, A.s.D.N. Initiative, Design and application of a generic clinical decision support system for multiscale data, IEEE Trans. Biomed. Eng. 59 (2012) 234–240.

[67] P. Anooj, Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules, J. King Saud Univ. Comput. Inform. Sci. 24 (2012) 27–40.

[68] N. Setiawan, P. Venkatachalam, M. Fadzil, Rule selection for coronary artery disease diagnosis based on rough set, Int. J. Recent Trends Eng. 2 (2009) 198–202.

[69] H. Kahramanli, N. Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Expert Syst. Appl. 35 (2008) 82–89.

[70] K.B. Nahato, K.N. Harichandran, K. Arputharaj, Knowledge mining from clinical datasets using rough sets and backpropagation neural network, Comput. Math. Methods Med. 2015 (2015).

[71] S. Gazali, I. Singh Kaur, Artificial intelligence based clinical data management systems: a review, Inform. Med. Unlocked 9 (2017) 219–229.

[72] C. Raji, H. Anand, S.V. Chandra, Computer based prognosis model with dimensionality reduction and validation of attributes for prolonged survival prediction, Inform. Med. Unlocked 9 (2017) 93–106.

[73] E.D. Farahani, M.H. Moradi, Multimodal detection of concealed information using genetic-SVM classifier with strict validation structure, Inform. Med. Unlocked 9 (2017) 58–67.

[74] M. Tavana, A.-R. Abtahi, D. Di Caprio, M. Poortarigh, An artificial neural Network and Bayesian Network model for liquidity risk assessment in banking, Neurocomputing 275 (2018) 2525–2554.

[75] B.Y. Hiew, S.C. Tan, W.S. Lim, Intra-specific competitive co-evolutionary artificial neural network for data classification, Neurocomputing 185 (2016) 220–230.

[76] M. Manngård, J. Kronqvist, J.M. Böling, Structural learning in artificial neural networks using sparse optimization, Neurocomputing 272 (2018) 660–667.

[77] R. Jafari-Marandi, M. Khanzadeh, B.K. Smith, L. Bian, Self-organizing and error driven (SOED) artificial neural Network for smarter classifications, J. Comput. Des. Eng. 4 (2017) 282–304.

[78] B. Ghaddar, J. Naoum-Sawaya, High dimensional data classification and feature selection using support vector machines, Eur. J. Oper. Res. 265 (2018) 993–1004.

[79] Y. Hayashi, K. Fukunaga, Accuracy of rule extraction using a recursive-rule extraction algorithm with continuous attributes combined with a sampling selection technique for the diagnosis of liver disease, Inform. Med. Unlocked 5 (2016) 26–38.

[80] M. Dialameh, M.Z. Jahromi, A general feature-weighting function for classification problems, Expert Syst. Appl. 72 (2017) 177–188.

[81] Y. Ko, New feature weighting approaches for speech-act classification, Pattern Recogn. Lett. 51 (2015) 107–111.

[82] R. Nagarajan, M. Upreti, An ensemble predictive modeling framework for breast cancer classification, Methods 131 (2017) 128–134.

[83] A.A. Funkner, A.N. Yakovlev, S.V. Kovalchuk, Data-driven modeling of clinical pathways using electronic health records, Procedia Comput. Sci. 121 (2017) 835–842.

[84] A. Meléndez-Lazo, L. Ordeix, M. Planellas, J. Pastor, L. Solano-Gallego, Clinicopathological findings in sick dogs naturally infected with Leishmania infantum: comparison of five different clinical classification systems, Res. Vet. Sci. 117 (2018) 18–27.

[85] H. Wu, S. Yang, Z. Huang, J. He, X. Wang, Type 2 diabetes mellitus prediction model based on data mining, Inform. Med. Unlocked 10 (2018) 100–107.

[86] J.-C. Nault, V. Paradis, D. Cherqui, V. Vilgrain, J. Zucman-Rossi, Molecular classification of hepatocellular adenoma in clinical practice, J. Hepatol. 67 (2017) 1074–1083.

[87] Y. Mamatjan, S. Agnihotri, A. Goldenberg, P. Tonge, S. Mansouri, G. Zadeh, K. Aldape, Molecular signatures for tumor classification: an analysis of the Cancer genome atlas data, J. Mol. Diagn. 19 (2017) 881–891.

[88] T.P. Exarchos, G. Rigas, A. Bibas, D. Kikidis, C. Nikitas, F.L. Wuyts, B. Ihtijarevic, L. Maes, M. Cenciarini, C. Maurer, Mining balance disorders' data for the development of diagnostic decision support systems, Comput. Biol. Med. 77 (2016) 240–248.

[89] J.J. Valdés, Extreme learning machines with heterogeneous data types, Neurocomputing 277 (2018) 38–52.

[90] U.R. Acharya, P. Chowriappa, H. Fujita, S. Bhat, S. Dua, J.E. Koh, L. Eugene, P. Kongmebhol, K.-H. Ng, Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images, Knowl.-Based Syst. 107 (2016) 235–245.

[91] S. Harous, M. El Menshawy, M.A. Serhani, A. Benharref, Mobile health architecture for obesity management using sensory and social data, Inform. Med. Unlocked 10 (2018) 27–44.

[92] O. Hadjerci, A. Hafiane, D. Conte, P. Makris, P. Vieyres, A. Delbos, Computer-aided detection system for nerve identification using ultrasound images: a comparative study, Inform. Med. Unlocked 3 (2016) 29–43.

[93] Y. Chen, X. Yue, H. Fujita, S. Fu, Three-way decision support for diagnosis on focal liver lesions, Knowl.-Based Syst. 127 (2017) 85–99.

[94] J. Camps, A. Samà, M. Martín, D. Rodríguez-Martín, C. Pérez-López, J.M.M. Arostegui, J. Cabestany, A. Català, S. Alcaine, B. Mestre, Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit, Knowl.-Based Syst. 139 (2018) 119–131.