Article

Impact of Metaheuristic Iteration on Artificial Neural Network Structure in Medical Data

Ihsan Salman [1,2,*] , Osman N. Ucan [1], Oguz Bayat [1] and Khalid Shaker [3]

[1] Altinbas University Graduate School of Science and Engineering, Istanbul 34217, Turkey;
    osman.ucan@altinbas.edu.tr (O.N.U.); oguz.bayat@altinbas.edu.tr (O.B.)
[2] Computer Department, College of Basic Education, University of Diyala, Diyala 32001, Iraq
[3] College of Computer Science and Information Technology, University of Anbar, Ramadi 31001, Iraq;
    khalidalhity@gmail.com
[*] Correspondence: ihsan1975.83@gmail.com; Tel.: +90-553-521-4448

check for
updates

Abstract: Medical data classification is an important factor in improving diagnosis and treatment and can assist physicians in making decisions about serious diseases by collecting symptoms and medical analyses. In this work, hybrid classification optimization methods such as Genetic Algorithm (GA), Particle Swam Optimization (PSO), and Fireworks Algorithm (FWA), are proposed for enhancing the classification accuracy of the Artificial Neural Network (ANN). The enhancement process is tested through two experiments. First, the proposed algorithms are applied on five benchmark medical data sets from the repository of the University of California in Irvine (UCI). The model with the best results is then used in the second experiment, which focuses on tuning the parameters of the selected algorithm by choosing a different number of iterations in ANNs with different numbers of hidden layers. Enhanced ANN with the three optimization algorithms are tested on biological gene sequence big dataset obtained from The Cancer Genome Atlas (TCGA) repository. GA and FWA are statistically significant but PSO was statistically not, and GA overcame PSO and FWA in performance. The methodology is successful and registers improvements in every step, as significant results are obtained.

Keywords: classification; metaheuristic algorithms; ANN; PSO; FWA; GA; data mining

1. Introduction

Medical data classification is an important factor in enhancing diagnosis and treatment. Moreover, this field continues to grow for computer researchers because of the major role played by medical data in human life. Classifying medical data can assist physicians in making decisions about serious diseases by collecting symptoms and medical analyses. Symptoms of patients are used as attributes for a disease data set, which also considers the number of instances.

The large amount of available medical data might be useful in healthcare. Data mining can be used in analyses of medical centers for providing sufficient sources, timely detection, and prevention of diseases, and avoiding high expenses caused by undesired and costly medical tests [1]. Numerous data mining approaches are implemented by scientists for diagnosing and treating various diseases, such as diabetes [2], liver disorder [3], Parkinson's [4], and cancer [5]. The artificial neural network (ANN) is used widely in disease mining classification and prediction; when performed with conventional backpropagation training, ANN improves accuracy and efficiency Mandal and Banerjee [6].

Heider et al. proposed a neural network cluster, which consists of four subfamily networks, to assign a small GTPase to one of the subfamilies and a filter network to identify small GTPases [7]. Desell et al. presented an ant colony optimization algorithm to evolve the structure of deep recurrent

neural networks. By contrast, ant colony optimization algorithm versions for continuous parameter optimization have been used to train the weights of neural networks [8]. Mizuta et al. [9] proposed genetic algorithms to design and train neural networks, in order to obtain the best network structure and the optimized parameter set concurrently. Meanwhile, a fitness function depending on output errors and simplicity in the structure of the network is introduced. Blum and Socha initially presented an ant colony optimization approach for discrete optimization problems [10].

Örkcü et al. developed a hybrid intelligent model (hybrid genetic algorithm and simulated annealing) for training ANNs that aim to exploit the advantages of the genetic and simulated annealing algorithms and alleviate their limitations. The model was applied to three benchmark data sets, namely, Breast Cancer Wisconsin (WDBC), Pima Indians Diabetes (PID), and Liver Disorders (LD) [5].

Seera and Lim proposed a hybrid intelligent system that integrates fuzzy min–max neural network with classification and regression tree and random forest. This hybrid system aims to exploit the advantages and reduce the limitations of the constituent models, learn incrementally using fuzzy min–max neural network, explain its predicted outputs with classification and regression tree, and achieve high classification performance by random forest [11].

Dutta et al. proposed an improved firework with an ANN model for classifying five benchmark data sets from the University of California, Irvine (UCI) repository [2,12].

In Zainuddin et al. [13], a wavelet neural network was used as a classifier on two epileptic seizure benchmark data sets from the UCI repository; an algorithm based on enhanced harmony search was used for feature selection.

Varma et al. developed an approach to tackling boundaries of decision trees and identifying split points using the Gini index in diagnosis (PID) [14].

Maddouri and Elloumi [15] presented four separate machine learning approaches for biological sequences classification. For recent progress of big data applications in the health-care domains the readers, refer to Luo et al. [16]. Guarracino et al. proposed a feature selection technique to train a generalized classifier with a significantly smaller subset of points and features of the original data [17]. To help biomarker discovery in cancer Celli et al. [18] studied DNA methylation data.

Nearly all of these proposed methods and algorithms focus on hybridizing ANN with one or more optimization algorithms [5], or attempt to improve or change the kernel of ANN [19], but does not bind between the structural change of ANN and the number of iterations of the metaheuristic algorithm. This limitation led to the idea of studying the impact of selecting different numbers of iterations for more than one structure of ANN.

This paper is structured as follows. The next section discusses the materials, methodology (including the algorithms), proposed work, and specifications of the medical datasets. Section 3 presents the results of the experiments. Section 4 discusses the simulation results in detail. Section 5 concludes the paper and presents recommendations for future work.

2. Materials and Methods

2.1. Artificial Neural Network

Essentially, ANN can be used as a classification model by mapping input data to the approximate desired output and consists of an input layer (the layer that receives inputs), output layer (the layer that provides outputs), and hidden layer(s) between them according to the model.

In this study, the inputs to the ANN are the attributes from the disease data set. This input goes through the input layer and is multiplied by weights, which are initialized randomly in the beginning, with respect to neurons in the hidden layers where the summation is provided to the activation function, as shown in Equations (1) and (2).

$$s(x) = \sum_{i=1}^{n} x_i w i \qquad (1)$$

The output of the neuron is specified after examining the summation results obtained with the activation function. In the proposed model, the sigmoid function is used, as shown in Equation (2).

$$f = \frac{1}{1 + e^{-s(x)}} \tag{2}$$

In this study, the data set is divided into training (40%), validation (30%), and test (30%) data sets. The trained neural network structure and the training set weights ($w1, \ldots, wn$) are used to evaluate the populations and as input to the metaheuristic algorithms, respectively.

2.2. Optimization Algorithms

This study aims to enhance ANN by hybridizing it with three optimization algorithms, namely, genetic algorithm (GA), particle swarm optimization (PSO), and fireworks algorithm (FWA), which are considered population-base metaheuristic algorithms.

2.2.1. Genetic Algorithm GA

GA applies evolutionary algorithm on a set of populations to search for optimal solutions according to fitness functions as in Holland, Mitchell and Goldberg [20–22], and the functionality of the algorithm is represented in the selection, crossover, and mutation, as shown in Figure 1.
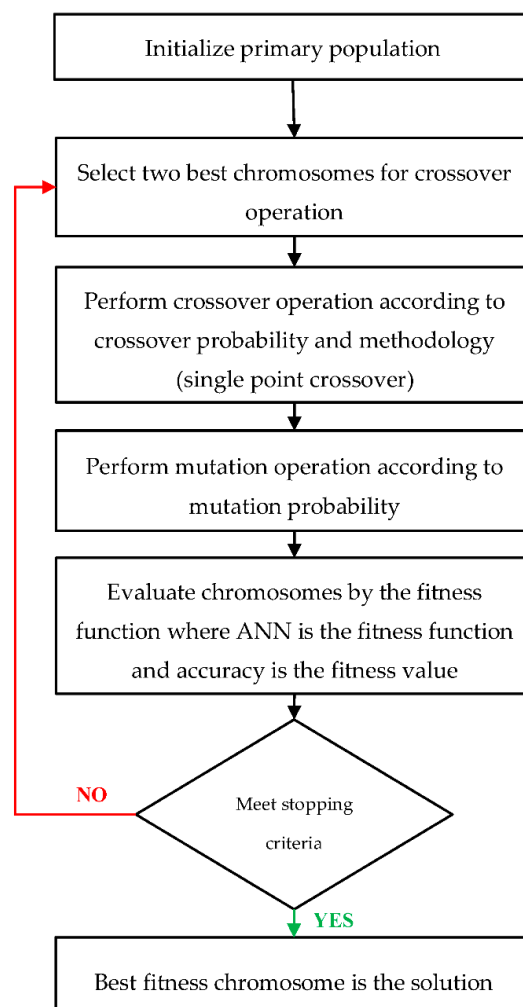

Figure 1. Flowchart of Genetic Algorithm (GA).

GA considers population evolutionary algorithms used to search for solutions similar to genetic evolution in nature.

2.2.2. Particle Swarm Optimization

PSO is a stochastic, organized, and autonomous algorithm Adrian et al. [23] and a decentralized population-based evolutionary technique Ardjani et al. [24].

The method behind the algorithm is moving particles (population of solution symbols) toward the solution in the search space.

The set of particle swarms (n) changes its position in each iteration process (i) according to the velocity vector, which governs the movement direction and distance of each particle, as shown in Equations (3) and (4).

The PSO process is shown in Figure 2.

$$
\begin{aligned}
\text{velocity}(i+1) = {}& \omega * \text{velocity}(i) + \varphi 1 * (\text{bestLocalPostion} - \text{postion}(i)) \\
& + \varphi 2 * (\text{bestGlobalPosition} - \text{position}(i))
\end{aligned}
\tag{3}
$$

$$
\text{position}(i+1) = \text{position}(i) + \text{velocity}(i+1)
\tag{4}
$$

where:

i: number of iterations;

ω: velocity coefficient or scale is normally 1 and decreases during iterations;

$\varphi 1$: fixed scale of difference between current position and local position; and

$\varphi 2$: fixed scale of difference between current position and global position.

2.2.3. Fireworks Algorithm

FWA is an intelligent swarm optimization algorithm created by Tan Ying in 2010 [25]. The algorithm mimics fireworks, which send spark showers that explode around them, Liu et al., Tan, and Dutta et al. [2,25,26].

In the algorithm, (N) fireworks represent the initial population, and the generated sparks are potential solutions around the population. Each iteration (firework) generates a number of sparks and then performs a random mutation for random fireworks to keep variants (diversification).

Sparks of various magnitudes and directions are then generated by the fireworks.

The maximum number of sparks (M) is generated from the firework as follows:

$$
S_i = M * \frac{f(X_i)}{\sum_{i=1}^{N} f(X_i)}
\tag{5}
$$

where:

S_i: number of sparks for firework (X_i);

M: maximum number of sparks; and

$f(X_i)$: activation function of firework (X_i).

The explosion spark amplitude can be determined as follows:

$$
A_i = A * \frac{F(X_{(N-i+1)})}{\sum_{i=1}^{N} F(X_i)}
\tag{6}
$$

where:

A_i: explosion spark amplitude for firework (X_i);

A: maximum amplitude; and

$F(X_i)$: activation function of firework (X_i).
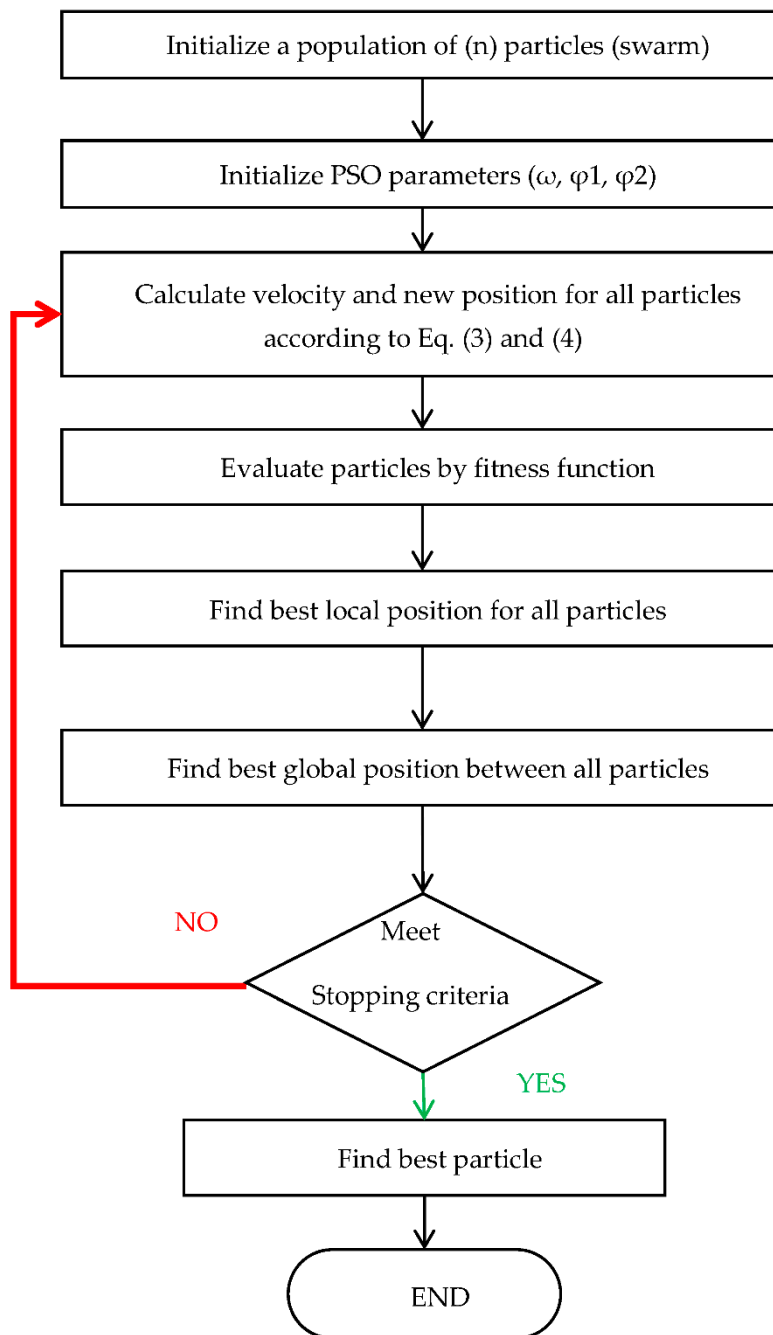
Figure 3 show the steps of FWA.

Figure 2. Flowchart of Particle Swarm Optimization (PSO).

2.3. Proposed Work

In this study, ANN is hybridized with three metaheuristic methods, namely, GA, PSO, and FWA, which are constructed by obtaining the best set of weights for ANN to produce improved accuracy performance.

The set of weights (w1, . . . , wn) is obtained after training and testing ANN in the first phase. In the second phase, an improved set of weights (wˆ1, . . . , wˆn) is obtained for the trained ANN using the optimization algorithms. This new set of weights (wˆ1, . . . , wˆn) should perform most accurately with the proposed ANN structure.
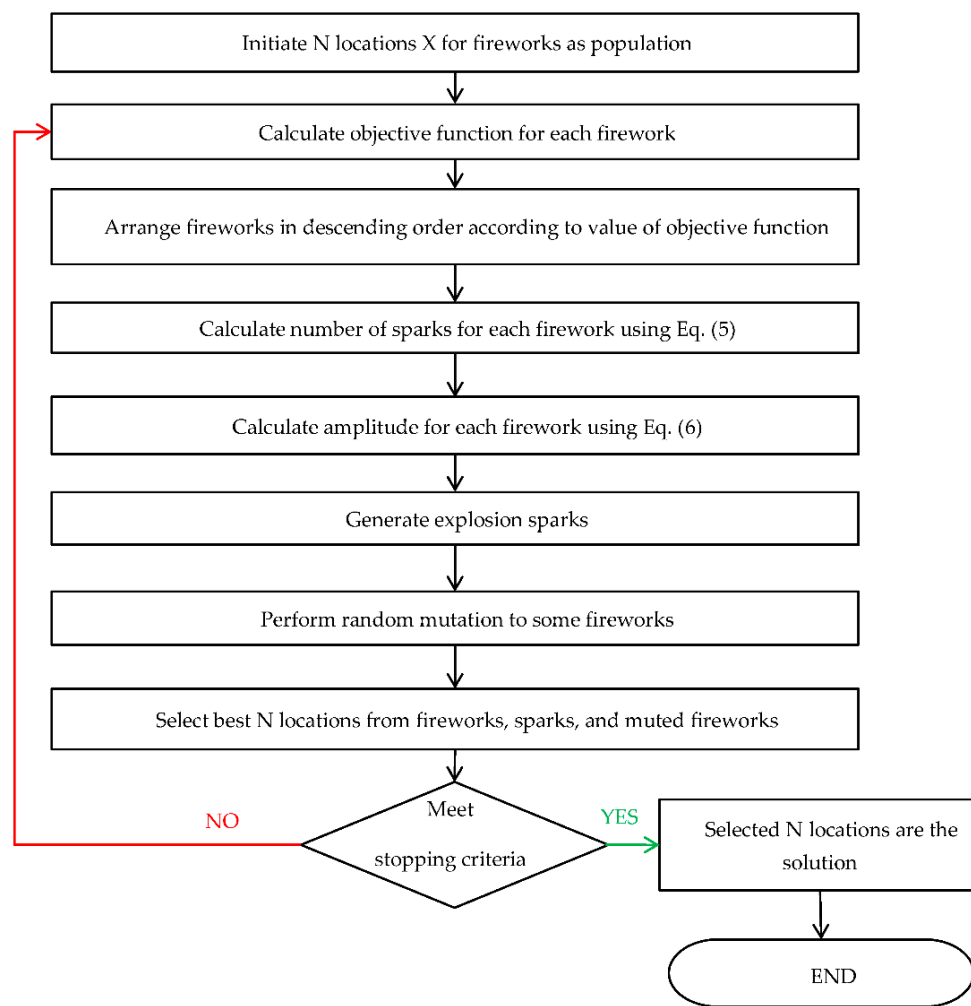
Figure 3. Flowchart of Fireworks Algorithm (FWA).

Figure 4 depicts the general structure of this study. The first training medical data set with ANN obtains the set of weights.

2.4. Diseases Data sets

Hybridization between classification methods and optimization algorithms are applied to the following disease data sets [12].

(1)   Pima Indian Diabetes (PID)

According to the National Institute of Diabetes and Digestive and Kidney Diseases, patients in this data set are pregnant females who are at least 21 years old and of Pima Indian heritage.
Number of instances: 768
Number of attributes: 9 (including class attribute)

(2)   Wisconsin Breast Cancer (WDBC)

This data set describes characteristics of the cell nuclei present in the image. Features are computed from a digitized image of a fine needle aspirate of a breast mass by the University of Wisconsin.
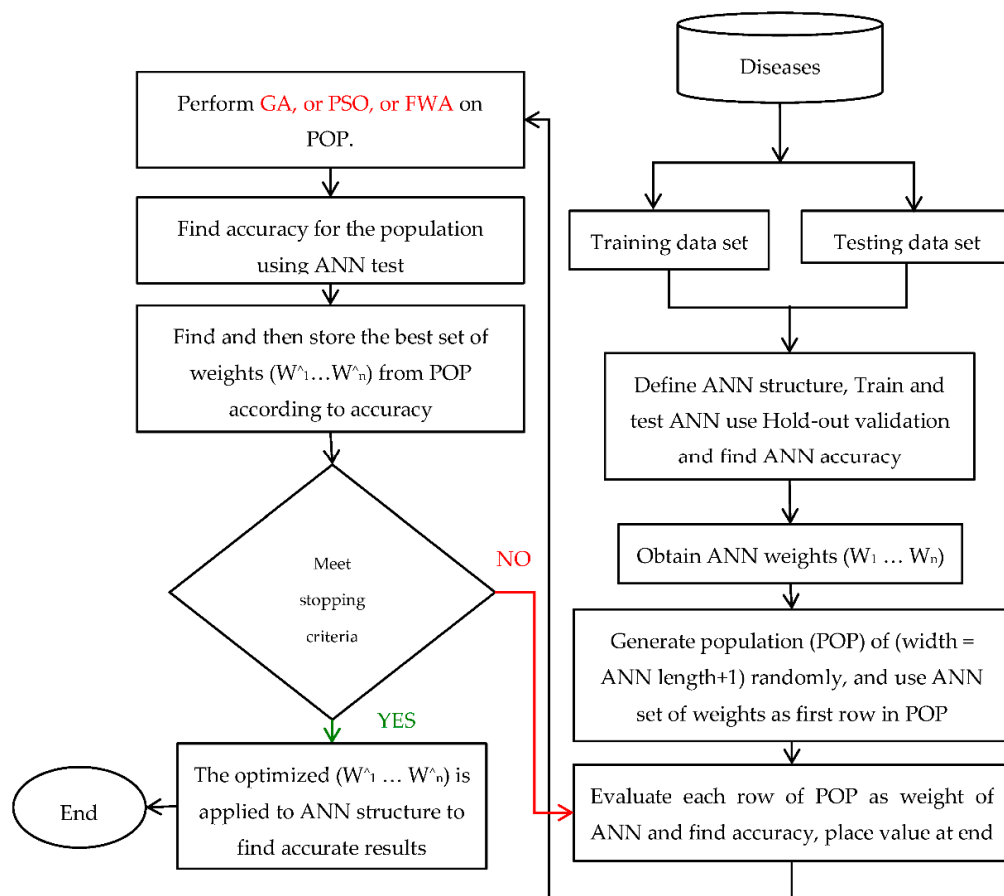Number of instances: 569
Number of attributes: 32

Figure 4. Flowchart of optimized classification by an Artificial Neural Network (ANN) with metaheuristic algorithms.

(3)   Liver Disorder (LD)

This data set contains seven attributes, five of which are blood test results related to liver disorders caused by alcohol consumption. The sixth attribute is the number of drinks per day. The seventh attribute represents class and reveals whether a patient has the disorder or not.
Number of instances: 345
Number of attributes: 7 (including class attribute)

(4)   Haberman Surgery Survival (HSS)

This data set is from a study conducted by Billings Hospital of the University of Chicago between 1958 and 1970 on the survival of patients who had undergone surgery for breast cancer.
Number of instances: 306
Number of attributes: 4 (including class attribute)

(5)   Parkinson's (PD)

The data set was created by Max Little of the University of Oxford in collaboration with the National Centre for Voice and Speech in Denver, Colorado and contains 23 attributes, which are sound measures, and 197 instances that represent sound records for 31 individuals, 23 of whom have Parkinson's disease.

(6)   Gene expression cancer (RNA-Seq)

RNA-Seq is a random extraction of gene expression of patients with the following tumor types: BRCA (breast), KIRC (kidney), COAD (colon), LUAD (lung) and PRAD (prostate).

The dataset is part of The Cancer Genome Atlas Pan-cancer Analysis Project Yuan et al. [27]. RNA-Seq dataset mentioned in literature as in Cestarelli et al. [28].

RNA-Seq gene expression levels measured by the Illumina HiSeq platform (DNA-to-Data solutions), to cope with gene expressions as numerical sequences inside the dataset.

Number of instances: 801

Number of attributes: 20531

Number of classes: 5

The specifications of the data sets are shown in Table 1.

Table 1. Specifications of disease datasets and their attributes' and classes' numbers.

DATA SET	NUMBER OF INSTANCES	NUMBER OF ATTRIBUTES	CLASS 1 (0) Value	CLASS 2 (1) Value
Pima Indian Diabetes (PID)	768	8	(500) not infected	(268) infected
Wisconsin Breast Cancer (WDBC)	569	31	(357) benign	(212) malignant
Liver Disorder (LD)	345	7	(145) not	(200) disorder
Haberman Surgery Survival (HSS)	306	3	(225) lived 5 years	(81) deceased
Parkinson's (PD)	195	23	(48) normal	(147) abnormal
Gene expression cancer (RNA-Seq)	801	20531	5 cancer types: BRCA, KIRC, PRAD, LUAD, COAD	

2.5. Problem Solving Strategy

The strategy of the study is tuning the ANN architecture with specific optimization algorithm parameters and number of iterations.

This study attempts to prove that changing the neural network structure by tuning metaheuristic algorithm parameters and increasing the number of iterations improves the performance of ANN and provides improved accuracy in disease diagnosis problems.

The proposed study is divided into two experiments. The first experiment determines the best of the three algorithms using the abovementioned strategy. The selected algorithm is used in the second experiment, in which the specified algorithm parameters are tuned to achieve improved results. A statistical *t*-test was used to find the mean differences between ANN and (ANN + PSO) in the five bench marks. The conditional probability p value obtained from the *t*-test was used to determine the statistical significance of the method. Finally, the hybridized algorithms were tested using the big dataset with multiclass to determine the algorithms' performance with this dataset.

After several tests, the algorithm parameters were set as follows:

(1) Genetic algorithm GA

Population = 10

Probability of crossover = 0.7

Probability of mutation = 0.1

(2) Particle swarm optimization PSO

Swarm size = 100

Velocity scalar coefficient (w) = 1.0

Velocity change in each iteration = 0.99

First velocity equation coefficient (c1) = 2.0

Second velocity equation coefficient (c2) = 2.0

(3)　Fireworks Algorithm FWA

　　　Number of fireworks = 50
　　　Number of sparks = 5
　　　Maximum function evaluation = 50,000
　　　Gaussian number for mutation = 5
　　　Probability of mutation = 0.3
　　　Spark upper bound = 10
　　　Spark lower bound = −10

3. Results

3.1. Experiment 1

This experiment tests the change in the ANN structure (number of hidden layers). Two and three of the hidden layers are tested with 500 and 1000 iterations, respectively, for the three proposed hybridization algorithms on the five disease benchmark data sets.

The experiment design is as follows:

- Hybrid ANN with (GA, PSO, and FWA) for five data sets (two hidden layers for ANN with 500 iterations) (Table 2);
- Hybrid ANN with (GA, PSO, and FWA) for five data sets (two hidden layers for ANN with 1000 iterations) (Table 3);
- Register the improvements between using 500 and 1000 iterations of (GA, PSO, and FWA) in two hidden layers for the five datasets;
- Hybrid ANN with (GA, PSO, and FWA) for five data sets (three hidden layers for ANN with 500 iterations) (Table 4);
- Hybrid ANN with (GA, PSO, and FWA) for five data sets (three hidden layers for ANN with 1000 iterations) (Table 5);
- Register the improvements between using 500 and 1000 iterations of (GA, PSO, and FWA) in three hidden layers for the five datasets.

3.2. Results of Experiment 1

The results of implementing the three proposed algorithms on the benchmarks are shown in Tables 2–5.

Table 2. Two-hidden-layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 500 iterations on UCI datasets: Pima Indian Diabetes PID, Wisconsin Breast Cancer WDBC, Liver Disorder LD, Haberman Surgery Survival HSS, and Parkinson's PD.

2 Hidden_ 500 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		74.35	97.17	57.10	72.88	75.38
Hybridized	ANN + GA	78.78	97.72	72.75	78.10	90.26
	ANN + PSO	**79.82**	97.89	**74.52**	**78.76**	**90.77**
	ANN + FWA	79.56	**98.07**	73.07	78.76	89.23

Table 3. 2-hidden–layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 1000 iterations.

2 Hidden_1000 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		74.35	97.17	57.10	72.88	75.38
Hybridized	ANN + GA	79.40	98.07	74.52	78.10	90.77
	ANN + PSO	79.82	**98.42**	**74.52**	**79.08**	**91.28**
	ANN + FWA	**79.95**	98.07	73.33	78.43	91.28

Tables 2 and 3 show the following improvements:

- ANN + GA: 4 out of 5 benchmarks are improved.
- ANN + PSO: 4 out of 5 benchmarks are improved.
- ANN + FWA: 3 out of 5 benchmarks are improved.

Table 4. 3–hidden-layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 500 iterations.

3 Hidden_ 500 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		76.48	95.39	56.81	73.57	80.51
Hybridized	ANN + GA	78.39	97.89	71.59	77.78	90.77
	ANN + PSO	**79.82**	98.24	**73.91**	**78.43**	**89.74**
	ANN + FWA	78.39	**98.42**	70.14	77.78	89.28

Table 5. Three-hidden-layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 1000 iterations.

3 Hidden_ 1000 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		76.48	95.39	56.81	73.57	80.51
Hybridized	ANN + GA	78.91	98.07	71.59	78.10	90.77
	ANN + PSO	**80.21**	98.07	**75.65**	**78.76**	**91.28**
	ANN + FWA	79.82	**98.42**	73.91	78.10	90.77

Tables 4 and 5 show the following improvements:

- ANN + GA: Three out of five benchmarks are improved.
- ANN + PSO: Four out of five benchmarks are improved.
- ANN + FWA: Four out of five benchmarks are improved.

Figure 5, shows the performance of each proposed algorithm in enhancing the ANN with three hidden layers using 1000 iterations. PSO with 3-hidden-layer ANN and 1000 iterations performed the best in most cases. Therefore, hybridized PSO with ANN is approximately the most promising algorithm in terms of enhancement.
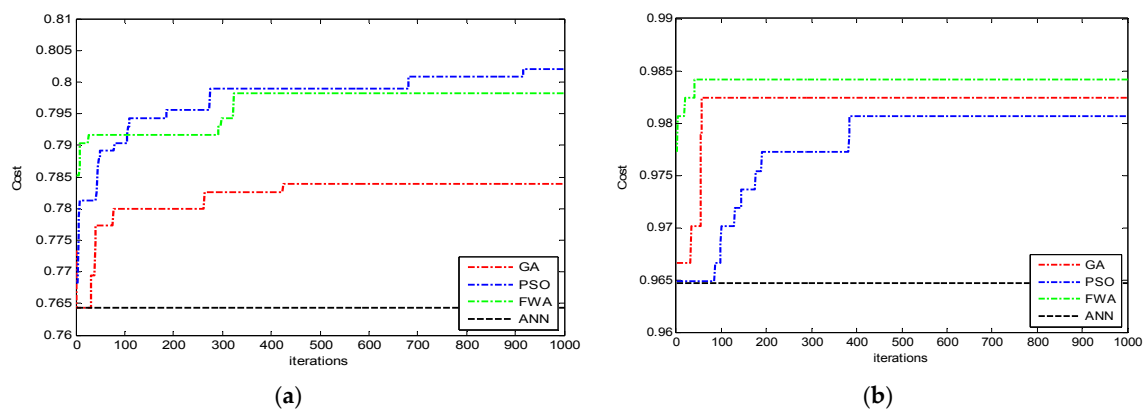

(**a**)

(**b**)

Figure 5. *Cont.*
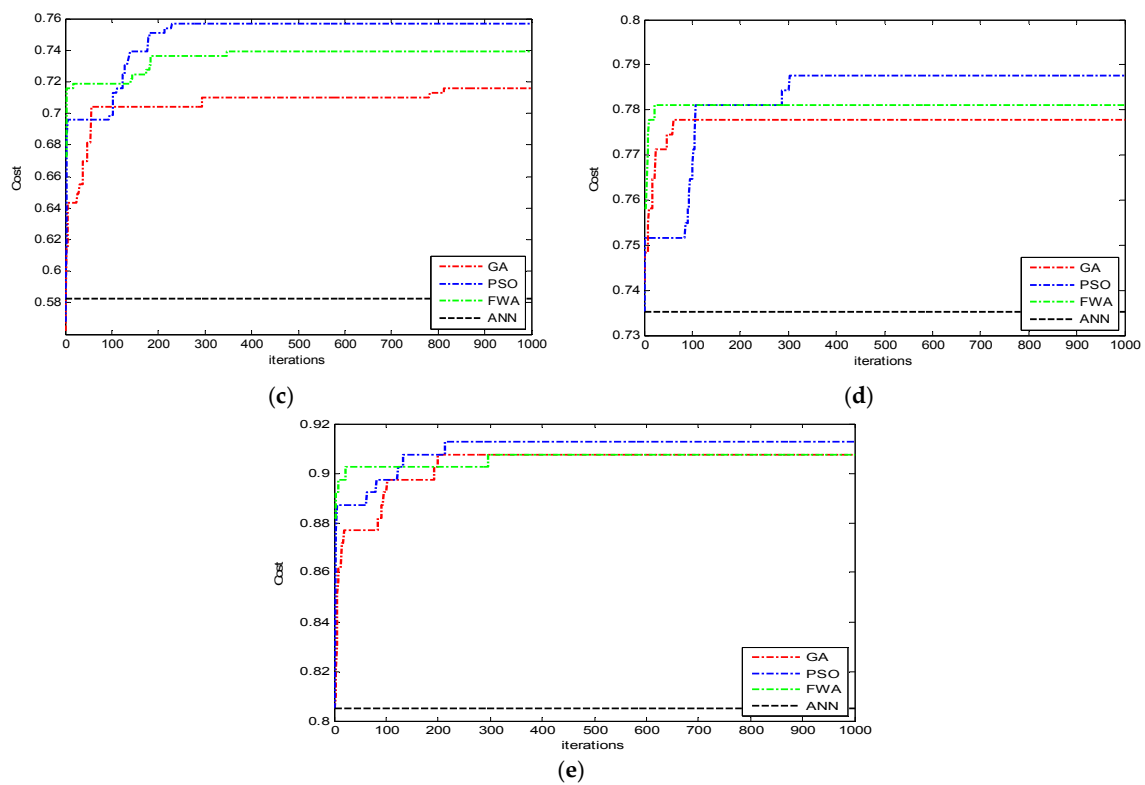
(c)

(d)

(e)

Figure 5. Shows experiment 1, the classification performance of the three proposed hybrid algorithms GA, PSO, and FWA applied on the five diseases data sets: (**a**) PID; (**b**) WDBC; (**c**) LD; (**d**) HSS; (**e**) PD.

3.3. Experiment 2

In Experiment 1, PSO obtained the best results on the five benchmark disease data sets except WDBC, which had the most attributes among the data set's 30 features. In Experiment 2, ANN was enhanced with only PSO by tuning parameters with different ANN architectures.

Experiment 2 involved changing the initialize condition of PSO and the swarms fully initialized around the set of weights obtained from the first ANN training to enforce PSO to start searching locally in order to increase intensification. Hybridization was performed according to the following strategy in order to explore search space and guarantee diversification.

- Three hidden layers for ANN with 1000 iterations and 100 swarms.
- Four hidden layers for ANN with 1000 iterations and 100 swarms.
- Four hidden layers for ANN with 1000 iterations and 200 swarms.

3.4. Results of Experiment 2

The results of Experiment 2 are shown in Table 6. Figure 6 show the performances of PSO with 4 hidden layers using 1000 iterations and 200 swarms (PSO population) applied on the five disease data sets.

Table 6. Experiment 2 results, 3- and 4-hidden-layer ANN hybridized with PSO algorithm for 1000 iterations.

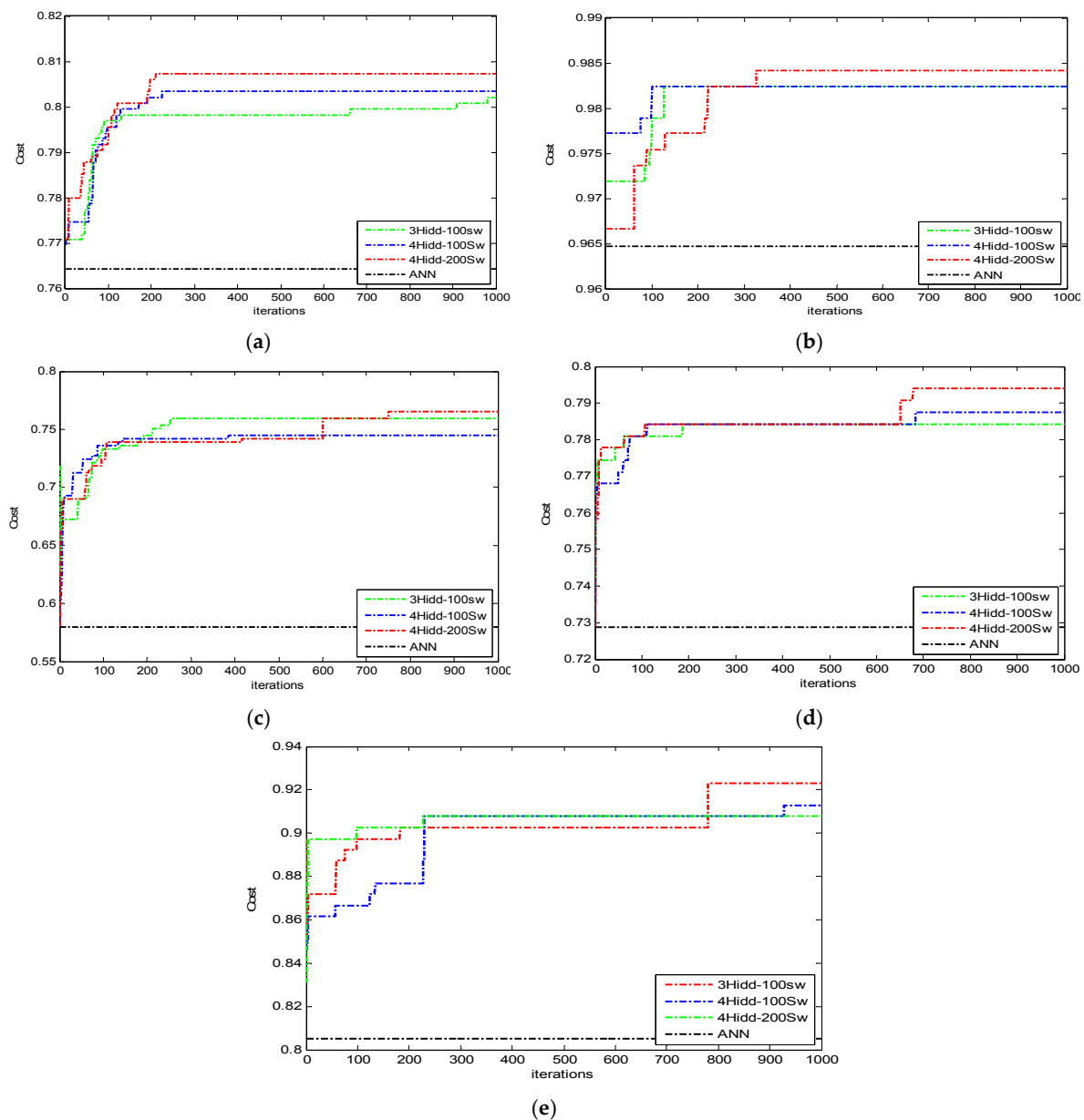|	Experiment 2	PID	WDBC	LD	HSS	PD
	Classical ANN	74.35	95.39	57.10	72.88	75.38
Hybridized	ANN + PSO(3,1000), 100 swarm	80.21	98.42	74.49	78.43	90.26
	ANN + PSO(4,1000), 100 swarm	79.56	97.01	76.52	78.43	91.28
	ANN + PSO(4,1000), 200 swarm	**80.34**	**98.42**	**76.52**	**79.41**	**92.31**

Figure 6. The performances of PSO with four hidden layers using 1000 iterations and 200 swarms (PSO population) applied on the five disease data sets: (**a**) PID; (**b**) WDBC; (**c**) LD; (**d**) HSS; (**e**) PD.

A *t*-test is a statistical hypothesis test that compares the means of two algorithms. *T*-test uses correlation and regression to determine how two algorithms vary. The *t*-test statistic is converted to a conditional probability called *p*-value. The *p*-value answers the question, 'If the null hypothesis is true, what is the probability of observing the current data or data that is more extreme?'

A small *p*-value provides evidence against the null hypothesis because the observed data are unlikely when the null hypothesis is true (Afifi and Azen 2014) [29].

We considered the default value (0.05) as an indicator to test significance, and ($p < 0.05$) was considered significant.

The formula involved in the computation of *p*-value, which is integral to the t-distribution probability density function, is detailed below in Equation (7):

$$\frac{1}{\sqrt{df}\beta\left(\frac{1}{2}, \frac{df}{2}\right)} \int_{-t}^{t} \left(-\frac{x^2}{df}\right) - \frac{df+1}{2} dx \tag{7}$$

where df is the degree of freedom and β is the beta function [30].

Table 7 shows that the statistical test on the five UCI benchmarks are significant according to finding the differences between ANN and optimized ANN + PSO.

Table 7. Statistical test and of 5 University of California in Irvine (UCI) datasets: Pima Indian Diabetes PID, Wisconsin Breast Cancer WDBC, Liver Disorder LD, Haberman Surgery Survival HSS, and Parkinson's PD, between ANN and ANN + PSO.

	PIMA	WDBC	LD	HSS	PD
Mean difference	0.02889000	0.00929500	0.17477500	0.02254000	0.03258000
T score	16.3009	9.9128	64.9535	30.3677	5.3323
standard error of difference	0.002	0.001	0.003	0.001	0.006
Difference verdict	Significant	Significant	Significant	Significant	Significant

Testing the Hybridized Algorithms with RNA-Seq

Large medical datasets were used in the proposed study and tested with the Cancer RNA-Seq Data Set as in Cestarelli et al. [28], which is a random extraction of gene expression of patients with the following tumor types: BRCA, KIRC, COAD, LUAD and PRAD.

GA, PSO and FWA registered enhancements when hybridized with ANN-4 hidden layers and applied on the RNA-Seq for classification, as shown in Table 8. GA overcame the other optimization algorithms.

Table 8. Accuracy of RNA-Seq in ANN and the three algorithms.

	ANN	ANN + PSO	ANN + GA	ANN + FWA
RNA-Seq	93.68	95.03	98.75	96.31

The statistical test shows that GA and FWA are statistically significant, whereas PSO was not significant, as shown in Table 9.

Table 9. Statistical test of RNA-Seq between ANN and the three algorithms GA, FWA, and PSO.

	ANN + GA	ANN + FWA	ANN + PSO
Mean difference	0.004328	0.033836	−0.007176
T score	2.9800	7.6776	0.4819
standard error of difference	0.001	0.004	0.015
Difference verdict	Significant	Significant	Not Significant

4. Discussion

The number of hidden layers is one of the most effective factors in ANN learning because increasing the number of these layers will increase the complexity of ANN and cause overfitting in its learning. Moreover, increasing this number allows the model to obtain precise classification results. To address the overfitting problem and improve the classification, metaheuristic algorithms are used for the ANN generalization.

Metaheuristic algorithms are search algorithms that optimize mathematical models to obtain better solutions [31,32]. These algorithms iteratively employ different strategies. For instance, tuning the parameters accurately within a single iteration of the algorithm will prevent algorithm divergence, and increasing the number of iterations will enable the mathematical model (the ANN model in this paper) to dive deeper and search for better solutions within the search space.

In experiment 1, we observe the following:

- As shown in Tables 2 and 3, increasing the number of iterations in two hidden layers from 500 to 1000 will produce the following effects:

  (1) ANN + GA: four out of five benchmarks are improved.
  (2) ANN + PSO: four out of five benchmarks are improved.
  (3) ANN + FWA: three out of five benchmarks are improved.

- As shown in Tables 2 and 3, increasing the number of iterations in three hidden layers from 500 to 1000 will produce the following effects:

  (1) ANN + GA: four out of five benchmarks are improved.
  (2) ANN + PSO: four out of five benchmarks are improved.
  (3) ANN + FWA: three out of five benchmarks are improved.

The results of experiment 1 prove that the hybrid ANNPSO algorithm is the most suitable methodology for this paper.

In experiment 2, we use the best algorithm identified in experiment 1 and fine-tune the parameters of the PSO algorithm. We start from population initialization. In experiment 1, the population initialization for the three algorithms was two halves, first half constructed from ANN set of weights (w1, . . . ,wn) and around neighbors, and the second half initialized randomly. In experiment 2, all of the initial population are taken from the ANN set of weights and neighbors.

Population density is useful in finding better solutions in the hyperspace. Therefore, we test 100 and 200 swarms as discussed in Section 3.3, experiment 2, and Table 6. The empirical results of experiment 2 revealed the following:

- ANN + PSO (3 hidden layers, 100 swarms): 2 out of 5 medical datasets are improved.
- ANN + PSO (4 hidden layers, 100 swarms): 2 out of 5 medical datasets are improved.
- ANN + PSO (4 hidden layers, 200 swarms): all medical datasets are improved.

These results indicate that PSO is the best of all algorithms tested in this study. The results of this algorithm also compete well with those of some state-of-the-art algorithms as reported in Table 10. Figure 7, shows PSO algorithm enhancement on ANN in the five medical datasets.

Table 10. Comparative results.

Pima Indian Diabetes (PID)	77.60% (Au et al., 2001)	[33]
	75.29% (P. Luuka, 2009)	[34]
	77.60% (Örkcü, 2015)	[5]
	80.34% (proposed)	
Wisconsin Breast Cancer (WDBC)	94.00% (Örkcü, 2011)	[35]
	97.49% (P. Luuka, 2011)	[36]
	97.29% (Seara & Lim, 2014)	[11]
	98.42% (proposed)	
Liver Disorder (LD)	70.25% (Luuka, 2009)	[34]
	74.86% (Lee and Mangasarian, 2001)	[37]
	76.52% (proposed)	
Haberman Surgery Survival (HSS)	72.70% (Pham, 2011)	[38]
	51.96% (Yang et al., 2017)	[39]
	79.41% (proposed).	
Parkinson's (PD)	85.03% (p. Luuka, 2011)	[36]
	81.34% (Rustempasic & Can 2013)	[40]
	93.60% (Shrivastava 2017)	[4]
	92.31% (proposed).	

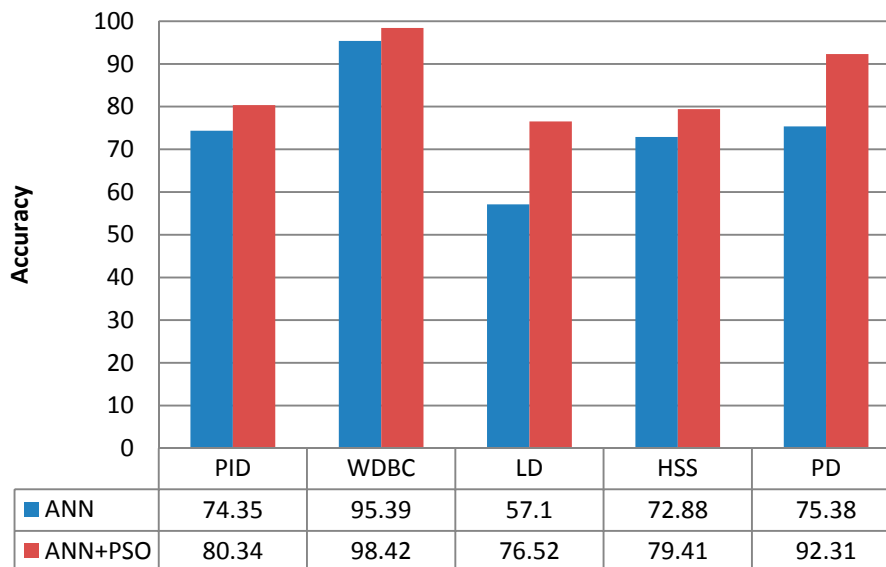|	PID	WDBC	LD	HSS	PD
■ ANN	74.35	95.39	57.1	72.88	75.38
■ ANN+PSO	80.34	98.42	76.52	79.41	92.31

Figure 7. Bar charts show accuracy enhancement of experiment 2.

The statistical test clearly shows that the differences between accuracies in ANN and ANN + PSO are significant in the five UCI datasets.

The above results did not mean that hybridized PSOANN is a general universal approach, and this notion is incorrect. The No Free Lunch theorem [27,35] declares that no superior optimization algorithm exists for all problems. This finding emerged from applying the same optimization algorithms on RNA-Seq dataset, and we can note that PSO, which was competitive in first five datasets, failed with big data in the RNA-Seq.

The aim of this study was to discuss the effect of population-metaheuristic number of iterations on different ANN structures and the influence of tune-metaheuristic parameters. We tested the performance of the population algorithms in classifying the large biomedical dataset, which allowed us to experiment with the No Free Lunch theorem. Future work can extend the findings of this study to different fields and applications. For instance, future research can study the impact of choosing different ANN kernel functions on hybridization or on selecting different types of metaheuristic algorithms, such as trajectory algorithms. The metaheuristic mathematical model for balancing exploration with explanation can be effectively improved by hybridizing metaheuristic and machine learning.

5. Conclusions

This paper investigates the hybridization of ANN with three optimization algorithms, namely, GA, PSO, and FWA. PSO is used to hybridize ANN with different layer structures, and the algorithm parameters are tuned. Comparisons of results on the five utilized medical data sets indicate that PSO is competitive in comparison with most of the other approaches in the literature (Table 10). The RNA-Seq classification experiment was a practical application of the No Free Lunch theorem, which establishes that a no universal optimization algorithm is found in all other algorithms for all optimization problems. In future studies, our approach can be applied to different medical problems.

References

1. Soliman, O.S.; Aboelhamd, E. Classification of Breast Cancer using Differential Evolution and Least Squares Support Vector Machine. *Int. J. Emerg. Trends Technol. Comput. Sci.* **2014**, *3*, 155–161.
2. Dutta, R.K.; Karmakar, N.K.; Si, T. Artificial Neural Network Training using Fireworks Algorithm in Medical Data Mining. *Int. J. Comput. Appl.* **2016**, *137*, 1–5.
3. Tavakkoli, P.; Souran, D.M.; Tavakkoli, S.; Hatamian, M.; Mehrabian, A.; Balas, V.E. Classification of the liver disorders data using Multi-Layer adaptive Neuro-Fuzzy inference system. In Proceedings of the 6th International Conference on Computing, Communications and Networking Technologies (ICCCNT), Denton, TX, USA, 13–15 July 2015; pp. 13–16.
4. Shrivastava, P.; Shukla, A.; Vepakomma, P.; Bhansali, N.; Verma, K. A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease. *Comput. Methods Progr. Biomed.* **2017**, *139*, 171–179. [CrossRef] [PubMed]
5. Örkcü, H.H.; Doğan, M.Đ.; Örkcü, M. A Hybrid Applied Optimization Algorithm for Training Multi-Layer Neural Networks in Data Classification. *Math. Comput. Model.* **2015**, *28*, 115–132.
6. Mandal, S.; Banerjee, I. Cancer Classification Using Neural Network. *Int. J. Emerg. Eng. Res. Technol.* **2015**, *3*, 172–178.
7. Heider, D.; Appelmann, J.; Bayro, T.; Dreckmann, W.; Held, A.; Winkler, J.; Barnekow, A.; Borschbach, M. A computational approach for the identification of small GTPases based on preprocessed amino acid sequences. *Technol. Cancer Res. Treat.* **2009**, *8*, 333–341. [CrossRef] [PubMed]
8. Desell, T.; Clachar, S.; Higgins, J.; Wild, B. Evolving deep recurrent neural networks using ant colony optimization. In Proceedings of the European Conference on Evolutionary Computation in Combinatorial Optimization, Copenhagen, Denmark, 8–10 April 2015; Springer: Cham, Switzerland, 2015; pp. 86–98.
9. Mizuta, S.; Sato, T.; Lao, D.; Ikeda, M.; Shimizu, T. Structure design of neural networks using genetic algorithms. *Complex Syst.* **2001**, *13*, 161–176.
10. Blum, C.; Socha, K. Training feed-forward neural networks with ant colony optimization: An application to pattern classification. In Proceedings of the 2005 Fifth International Conference on Hybrid Intelligent Systems (HIS'05), Rio de Janeiro, Brazil, 6–9 November 2005; p. 6.
11. Seera, M.; Lim, C.P. A hybrid intelligent system for medical data classification. *Expert Syst. Appl.* **2014**, *41*, 2239–2249. [CrossRef]
12. Lichman, M. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA, 2013; Available online: http://archive.ics.uci.edu/ml (accessed on 13 March 2016).
13. Zainuddin, Z.; Lai, K.H.; Ong, P. An enhanced harmony search based algorithm for feature selection: Applications in epileptic seizure detection and prediction R. *Comput. Electr. Eng.* **2016**, *53*, 143–162. [CrossRef]
14. Varma, K.V.S.R.P.; Rao, A.A.; Lakshmi, T.S.M.; Rao, P.V.N. A Computational Intelligence Approach for a Better Diagnosis of Diabetic Patients. *Comput. Electr. Eng.* **2014**, *40*, 1758–1765. [CrossRef]
15. Maddouri, M.; Elloumi, M. A data mining approach based on machine learning techniques to classify biological sequences. *Knowl. Based Syst.* **2002**, *15*, 217–223. [CrossRef]
16. Luo, J.; Wu, M.; Gopukumar, D.; Zhao, Y. Big data application in biomedical research and health care: A literature review. *Biomed. Inform. Insights* **2016**, *8*, 1–10. [CrossRef] [PubMed]
17. Guarracino, M.R.; Cuciniello, S.; Pardalos, P.M. Classification and characterization of gene expression data with generalized eigenvalues. *J. Opt. Theory Appl.* **2009**, *141*, 533–545. [CrossRef]
18. Celli, F.; Cumbo, F.; Weitschek, E. Classification of large DNA methylation datasets for identifying cancer drivers. *Big Data Res.* **2018**, in press. [CrossRef]
19. Avci, D.; Dogantekin, A. An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine. *Parkinson Dis.* **2016**, *2016*. [CrossRef] [PubMed]
20. Holland, J.H. *Adaptation in Natural and Artificial Systems II*; The University of Michigan Press: Ann Arbor, MI, USA, 1975.
21. Mitchell, M. An introduction to genetic algorithms. *Comput. Math. Appl.* **1996**, *32*, 133. [CrossRef]
22. Goldberg, D.E.; Holland, J.H. Genetic Algorithms and Machine Learning. *Mach. Learn.* **1988**, *3*, 95–99. [CrossRef]

23. Serapião, A.B.S.; Corrêa, G.S.; Gonçalves, B.; Carvalho, V.O. Combining K-Means and K-Harmonic with Fish School Search Algorithm for data clustering task on graphics processing units. *Appl. Soft Comput.* **2016**, *41*, 290–304. [CrossRef]

24. Ardjani, F.; Sadouni, K.; Benyettou, M. Optimization of SVM multiclass by particle swarm (PSO-SVM). In Proceedings of the 2010 2nd International Workshop on Database Technology and Applications (DBTA), Wuhan, China, 27–28 November 2010; pp. 32–38.

25. Liu, J.; Zheng, S.; Tan, Y. The improvement on controlling exploration and exploitation of Firework Algorithm. In *ICSI 2013, Part I: LNCS 7928*; Tan, Y., Shi, Y., Mo, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 11–23.

26. Tan, Y. *Fireworks Algorithm*; Springer: Berlin, Germany, 2015; pp. 247–262. [CrossRef]

27. Yuan, Y.; Van Allen, E.M.; Omberg, L.; Wagle, N.; Amin-Mansour, A.; Sokolov, A.; Liang, H. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **2014**, *32*, 644–652. [CrossRef] [PubMed]

28. Cestarelli, V.; Fiscon, G.; Felici, G.; Bertolazzi, P.; Weitschek, E. CAMUR: Knowledge extraction from RNA-seq cancer data through equivalent classification rules. *Bioinformatics* **2016**, *32*, 697–704. [CrossRef] [PubMed]

29. Afifi, A.A.; Azen, S.P. *Statistical Analysis: A Computer Oriented Approach*; Academic Press: London, UK, 2014.

30. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*; Courier Corporation: Chelmsford, MA, USA, 1964; Volume 55.

31. Talbi, E.-G. *Metaheuristics: From Design to Implementation*; John Wiley & Sons: New York, NY, USA, 2009; Volume 74.

32. Blum, C.; Roli, A. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. *ACM Comput. Surv.* **2003**, *35*, 189–213. [CrossRef]

33. Au, W.-H.; Au, W.-H.; Chan, K.C.C. Classification with Degree of Membership: A Fuzzy Approach. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November– 2 December 2001; pp. 24–25.

34. Luukka, P. Classification Based on Fuzzy Robust PCA Algorithms and Similarity Classifier. *Expert Syst. Appl.* **2009**, *36*, 7463–7468. [CrossRef]

35. Örkcü, H.H.; Bal, H. Comparing performances of backpropagation and genetic algorithms in the data classification. *Expert Syst. Appl.* **2011**, *38*, 3703–3709. [CrossRef]

36. Pasi, L. Feature Selection Using Fuzzy Entropy Measures with Similarity Classifier. *Expert Syst. Appl.* **2011**, *38*, 4600–4607.

37. Lee, Y.-J. SSVM: A Smooth Support Vector Machine for Classification. *Comput. Opt. Appl.* **2001**, *20*, 5–22. [CrossRef]

38. Pham, H.N.A.; Triantaphyllou, E. A Meta-Heuristic Approach for Improving the Accuracy in Some Classification Algorithms. *Comput. Oper. Res.* **2011**, *38*, 174–189. [CrossRef]

39. Yang, J.; Ma, Y.; Zhang, X.; Li, S.; Zhang, Y. A Minimum Spanning Tree-Based Method for Initializing the K-Means Clustering Algorithm. *Int. Sci. Index* **2017**, *11*, 13–17.

40. Indira, R.; Can, M. Diagnosis of Parkinson's Disease Using Principal Component Analysis and Boosting Committee Machines. *Southeast Eur. J. Soft Comput.* **2013**, *2*, 102–109.