DOI: 10.15849/IJASCA.230320.08

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 1, March 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Intelligent Diagnostic Prediction and Classification System for Parkinson's Disease by Incorporating Sperm Swarm Optimization (SSO) and Density-Based Feature Selection Methods

Hisham A. Shehadeh[1, *], Iqbal H. Jebril[2], Ghaith M. Jaradat[1], Dyala Ibrahim[1], Rami Sihwail[1], Husam Al Hamad[1], Shu-Chuan Chu[3,4], Mohammad A. Alia[2]

[1]Faculty of Computer Science and Informatics, Amman Arab University, Amman 11953, Jordan.
[h.shehadeh, g.jaradat, d.ibrahim, r.sihwail, hhamad] @aau.edu.jo
[2]Faculty of Science and Information Technology, Al-. Zaytoonah University of Jordan, Amman, 11733 , Jordan.
[3]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China.
[4]College of Science and Engineering, Flinders University, 1284 South Road, Tonsley SA 5042, Australia.
*(corresponding author: Hisham A. Shehadeh)

Abstract

The systems of healthcare are being updated with modern capabilities, such as “Machine Learning (ML)”, “Data Mining (DM)”, and “Artificial Intelligence (AI)” in order to provide humans with more expert and intelligent services of healthcare. This paper provides a medical system of an intelligent prediction and classification for Parkinson’s disease incorporating the “Density based Feature Selection (DFS)” with our optimization algorithm namely, “Sperm Swarm Optimization (SSO) algorithm”. Prior to the SSO-based classifier construction, the proposed intelligent system (D-SSO) eradicates redundant or irrelevant features using DFS. Preprocessing, “Feature Selection (FS)”, and classification are the three phases of the proposed D-SSO framework. Moreover, the D-SSO algorithm is tested using a benchmark of Parkinson’s dataset, which the performance of D-SSO is examined using various evaluation factors. Mainly, the D-SSO algorithm is compared to existing approaches, which the proposed intelligent system outperforms the others, and gets an ideal recognition rate.

***Keywords**: Medical System, Data Mining (DM), Feature Selection (FS), Sperm Swarm Optimization (SSO), Parkinson’s Disease, Wrapper Based Approach (WBA).*

Received 6 October 2022; Accepted 22 January 2023

1 Introduction

Recently, the creation and developments branches of information technology include Internet of Things (IoT) [1, 2], mobile communication system, big data, and wearable computing are utilized in the area of healthcare. Generally speaking, systems of healthcare are constituted with help of mobile computing and big data to offer expert and intellectual services [3]. In addition, the huge amount of medical data leads various issues for managing, processing, and storing data. Persistent, disease of Parkinson's is a type of disorder of brain that connects to difficulty with coordination walking, stiffness, and shaking. Over time, this symptoms mainly starts gradually and being worse. As the disease progresses, humans may have difficulty talking and walking [4]. This disease affects one to two per one thousand of population, which its pervasiveness is dramatically increasing. Parkinson’s disease happened when cells of nerve in the area in the human brain that manages and controls movement become die or impaired, which this area is called the “Basal Ganglia” [5]. The “Basal Ganglia” is depicted in Fig.1 [6]. Normally, there is a brain chemical called as dopamine, which are generated by these neurons or nerve cells. When these neurons or nerve cells become impaired or die, they trigger less dopamine, which leads to Parkinson’s disease [5].

This disease has four major symptoms, which are listed as follows [7, 8]:

- Tremor (trembling) in arms, hands, jaw, head, or legs;
- Sluggishness of motion;
- Limbs and trunk of stiffness;
- Impaired in both coordination and balance, sometimes causes to falls.

There are other symptoms, which are not major. These symptoms can be listed as follows [7, 8]:

- Depression and other changes of emotion;
- Skin problems;
- Difficulty speaking, chewing, and swallowing;
- Constipation or urinary problems;
- Disruptions of sleep.

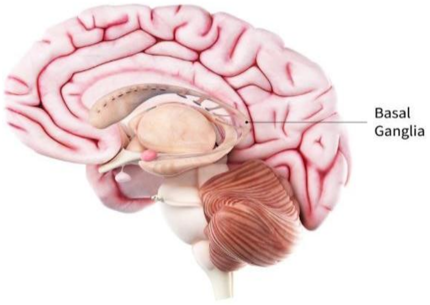

Fig 1: Basal Ganglia location in human brain [6].

Different studies reported that the earlier diagnoses and detection of Parkinson's disease could diminish the growth of disease even doctors or nurses of a primary care or the specialization of mental health care. Generally, approaches of imaging this disease are utilized to identify the presence of it. However, duo to of huge number of patients, it is difficult to make a check on each person. On the other hand, persons with a higher possibility of getting Parkinson's disease will be advised to undergo comprehensive and extensive testing. Recently, the storage of clinical and medical database be very complicated procedure in the industry of healthcare. These databases hold various diagnosis and features connected to disease, which are required to be equipped to gain a good quality of services. Since the archived data in the hospitals and health care centers may have missing as well as unimportant data. These data may become burdensome to mine the data of patients. Based on that, the data mining will take a step in the procedure to normalize these data, which data reduction will be applied. Hence, the process of identification of this disease becomes faster and simpler when the available data is reliable and accurate. Identification of Parkinson's disease from the data of patients can be utilized as an issue of data classification. In the classification process, the supervised learning task can be applied, which deduces a connection among class labels and features. A predication and classification technique uses the training data to generate model, which is utilized to test the performance of the prediction process [9].

Recently, “Artificial Intelligence (AI)” methods can be utilized to enhance the available models of classification. Simultaneously, the available of various features in the medical data of high dimension can be deduced in different problems, such as low interoperability, overfitting, and complexity of high computation of the finishing model. There is an important technique, namely “Feature Selection (FS)” in which is used to reduce the number of features by selection an important feature and eliminating the irrelevant ones. This will reduce the execution time of the procedure, which will increase the overall efficiency of the model [10, 11]. FS approaches are utilized in various applications, such as pattern recognition, “Machine Learning (ML)”, and “Data Mining (DM)”.

There are different methods of FS that are utilized as validation parameters. These methods can be classified into three approaches, such as “Filter Based Approach (FBA)”, “Wrapper Based Approach (WBA)”, and “Embedded Based Approach (EBA)” [12]. In FBA, the validation of features can be used by fixed measures instead of selected and learners features. On the other hand, the WBA is based on learning technique as a sub-procedure of estimating of betterment of the selecting of feature set. This approach is commonly applied, however, it faces many problems, such as identifying user-defined learner parameter, inbuilt learner constraints, and complexity of high computation. In contrast, EBA has less complexity compared to the aforementioned techniques, which integrates the wrapper and filter approaches and element their limitations. These approached have improved the discrimination of features or classification. Moreover, the procedure of FS has not improved the classifier but enhanced the features. In addition, as it mentioned before, hybrid and wrapper approaches have high complexity of computation [13].

To overcome these issues, in this paper, a newly “Wrapper Based Approach (WBA)” is proposed for Parkinson's disease identification by incorporating “Sperm Swarm Optimization (SSO)” algorithm and “Density-Based FS (DFS)”. The DFS method is a heuristic based procedure, which is applied to estimate the worthiness of a feature. In addition, it is utilized to eliminate the unnecessary features and help to increase the accuracy of SSO. The SSO method is applied to a benchmark Parkinson's disease from UCI repository. The rest sections of the study are organized as follows: Sec. 2 presents literature review. Section 3 discusses the proposed approach. Sec. 4 investigates the results obtained by the proposed approach. The conclusion of this work is made in section 5.

2 Related Work

There are different approaches have been proposed in the literature to predict of various diseases in patient's medical data. In this section, we can summarize few of them as follows:

Cai *et al.* [14] have proposed an enhancement on "fuzzy k-nearest neighbor (FKNN)" method to predict a Parkinson's disease for a set of patients. This method is coupled with an approach of instance-based learning, which is "Chaotic Bacterial Foraging Optimization with Gauss Mutation (CBFO)". This approach is compared with different approaches, such as "Support Vector Machine (SVM)", "Genetic Algorithm (GA)" based learning approach, etc. the outcomes showed that the proposed method outperformed the other approaches in the term of accuracy of prediction. In different view, Mathur *et al.* [15] has the advantages of prediction the Parkinson's disease based on different methods. They have merged the "K-Nearest Neighbor (KNN)" algorithm with "Artificial neural network (ANN)" in this prediction. The results illustrated that the proposed algorithm outperformed the "KNN-AdaBoosta", and KNN algorithms. On the other hand, Zue *et al.* [16] looked to further details by predicting the Parkinson's disease for a set of patient using "Fuzzy k-nearest neighbor (FKNN)" methods. This method is combined with an approach of instance-based learning, such as binary "Particle Swarm Optimization (PSO)". The outcomes presented that the proposed method outperformed the other existing methods in the literature.

On the other hand, Zhao *et al.* [17] have proposed a novel method for Parkinson's disease prediction. This method is "An Ensemble K-Nearest Neighbor (EnKNN)". The results showed that the proposed approach got an accuracy of 95.02% in the process of predication the disease. In a different discussion, Pahuj *et al.* [18] looked to further details by comparing a set of classifiers while they are utilized in the procedure of predicting Parkinson's disease. These classifiers are "Support Vector Machine, Multilayer Perceptron, and K-Nearest Neighbor". The results showed that "Artificial Neural Network (ANN)" has the highest classification accuracy, which is 95.89%.

Later on, Asmae *et al.* [19] have proposed a comparative study that compare between "K Nearest Neighbors (KNN)" and "Artificial Neural Networks (ANN)" algorithms in the process of predicting the Parkinson's disease. They used a newly dataset that are taken from UCI repository. This dataset consists of thirty-one subjects of which twenty-three were diagnosed with Parkinson's disease. The results showed that the ANN has the highest classification accuracy, which is 96.7%.

On the other hand, Gupta *et al.* [20] have proposed an optimized version of "Crow Search Algorithm (OCSA)" to predict the Parkinson's disease. The authors compared the proposed method with the standard one, which is "Chaotic Crow Search Algorithm (CCSA)". The results showed that the optimized version of CSA has the highest prediction accuracy of the disease.

The aforementioned studies have the advantages of studying the process of prediction and classification of Parkinson’s disease using different techniques. However, they did not use the wrapper method to select the appropriate features to get more accurate percentage of prediction accuracy. Based on that, in this work, we are motivated to study the prediction and classification of Parkinson’s disease based on wrapper method by incorporating “Sperm Swarm Optimization (SSO)” algorithm and “Density-Based FS (DFS)” method. The next subsection summarizes the concept of “Sperm Swarm Optimization (SSO)”.

2.1 Standard “Sperm Swarm Optimization (SSO)”

“Sperm Swarm Optimization (SSO)” is a modern created swarm-based approach inspired by the attitude of flock of sperms while fertilizing process, which is proposed by Shehadeh et al. [21–28]. Fertilization is an epic and complex story of a single sperm that unions with an Ovum (egg). Through the process of fertilization, the whole swarm floating in a path between two important points, which are Cervix and Fallopian Tube. Generally speaking, in the insemination process, the number of sperms that are swimming in the aforementioned path can be counted up to one hundred thirty million cells. Between all of these cells there is just one sperm that will fertilize the egg. The procedure of fertilization can be summarized in the following three velocities based on Shehadeh *et al.* [21–28]:

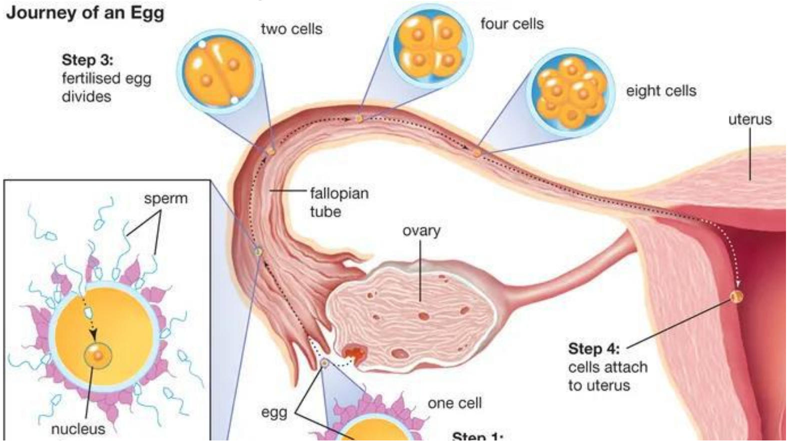

Fig 2: The procedure of fertilization [21].

First of all, the swarm of sperm is triggered by the male reproductive system in side the beginning of the path, which is called Cervix zone. This point is the starting zone of the fertilization journey. The procedure of fertilization is depicted in Fig (2). Based on that, each sperm will gain a random location in that point to get ready of that journey, where each cell has a value of velocity on the “*Cartesian plane*”. Mathematically speaking, in SSO, the initial velocity of swarm can be calculated according to the following equation [21–28]:

$$Initial\_Velocity = D \cdot V_i(t) \cdot Log_{10}(pH\_Rand_1) \qquad (1)$$

where,

- v_i – is the velocity of cell i at iteration t;
- D – id the factor of velocity damping, which is a random parameter in the range of 0–1;
- pH_Rand_1 – is the reached location pH value, which is random parameter in the range of 7–14;

Currently, every cell in the swarm becomes stand by to swim from the past point until getting closer the Ovum outer surface. The scientist in this field noticed that these cells float in the surface as “flock or swarm”, which swims from the zone of low temperature to the zone of higher temperature. Moreover, they noticed that the Ovum triggers a chemical to pull the whole swarm in which this task is called “*Chemotactic*”. The scientist also noticed that swarm beat in the same synchronicity as their frequency of tail movements through the “flocking and grouping”. The Ovum and its site in the Fallopian Tubes is depicted in Fig (2). Depends on Shehadeh *et al.* [21–28], this velocity is denoted by the velocity of personal best of the sperm, which is adjusted in the memory based on the prior site until getting closer to the optimal value (location of egg). Mathematically speaking, in SSO, this velocity can be formulized as follows:

$$Current\_Best\_Solution = Log_{10}(pH\_Rand_2)\cdot Log_{10}(Temp\_Rand_1)\cdot(x_{sbest_i}[]-x_i[]) \qquad (2)$$

As aforementioned above, in normal case, there is only one cell can fertilize the egg. Based on that, Shehadeh *et al.* [21–28] gave a name for this cell as the winner. The flock of sperm and winner are depicted in Fig (3).

As per above, this method utilizes a set of sperms (potential solutions), which floating in the entire domain of search space to discover and obtain the optimal solution. Concurrently, the potential solutions will consider the best sperm in their path, which is called the winner (the nearest sperm to the egg). In the other meaning, the flock will be considered the position of the winner as well as the position of the its prior best solution. In this method, every sperm enhances its prior location toward the optimum by considering its current velocity, current location, the location of both sperm best solution and global best solutions (the winner) as well. Mathematically speaking, in SSO, the winner velocity can be summarized as follows:

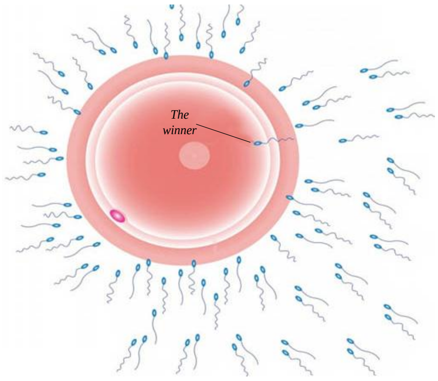

Fig 3: flock of sperm and the winner value [1, 12, 23, 24].

$$Global\_Best\_Solution(the\_winner) = Log_{10}(pH\_Rand_3)\cdot Log_{10}(Temp\_Rand_2)\cdot(x_{sgbest_i}[]-x_i[]) \qquad (3)$$

Depends on the prior equations, the total velocity rule $V_i(t)$ can be modeling as follows [21–28]. The pseudocode of SSO is stated as in Algorithm 1.

$$\begin{aligned} V_i(t)=&\overbrace{D\cdot Log_{10}(pH\_Rand_1)}^{\text{Sperm initial velocity}}\cdot V_i+\overbrace{Log_{10}(pH\_Rand_2)\cdot Log_{10}(Temp\_Rand_1)}^{\text{Personal best solution}}\\ &\cdot(x_{sbest_i}-x_i(t))+\underbrace{Log_{10}(pH\_Rand_3)\cdot Log_{10}(Temp\_Rand_2)\cdot(x_{sgbest}-x_i(t))}_{\text{Global best solution}} \end{aligned} \tag{4}$$

Mathematically speaking, in SSO, the swarm updated their locations according to the following model:

$$x_i(t)=x_i(t)+v_i(t) \tag{5}$$

The symbols of the prior equations, Eq.(4), and Eq.(5) are as follows:

- *pH_Rand_2, and pH_Rand_3* – are the reached location pH values, which are random
- parameter in the range of 7–14;
- *$Temp\_Rand_1$, $Temp\_Rand_2$* – are the reached location temperature values, which are
- random parameter in the range of 35.1–38.5;
- x_i – current position of potential solution *i* at iteration *t*;
- x_{sbest} – personal best location of potential solution *i* at iteration *t*;
- x_{sgbest} – global best location of the flock.
- where,
- v_i – is the velocity of cell *i* at iteration *t*;
- x_i – current position of cell *i* at iteration *t*.

The pseudocode of this algorithm can be structured as follows [21–28]:

```
Algorithm 1 “Sperm Swarm Optimization (SSO)”
Begin
Step 1: Initialize potential solutions.
Step 2: for i=1: size of flock do
Step 3: apply the fitness for potential solution.
            if obtained fitness > best solution of the potential solution then
            give the current value as the best solution of the potential solution.
            end if
        end for
Step 4: depends on the winner, give the value of winner.
Step 5: for i=1: size of flock do
            Perform Eq. (4)
            Perform Eq. (5).
        end for
Step 6: while final iterations is not reached go to Step 2.
End.
```

Based on the theory and rules of SSO, it can be noticed that the low of velocity is affected by two tuner parameters in which are the temperature and pH values. The former one can be changed depends on circulation of blood pressure of reproduction system, which can gain a value between 35.1 to 38.5 C_o randomly. On the other hand, the later one can be varied depends on a set of things include kinds of food consumed, and mood status of female, including, sadness or happiness. The temperature tuner parameter can gain a value between 7 to 14. To mimic the speed of real sperm, Shehadeh *et al.* [21–28] applied the logarithm to the aforementioned tuner parameters in their theory. In the other meaning, the logarithm rule is applied to velocity model to normalize the potential solutions.

As we mentioned previously, SSO is a swarm-based variant, which mimics the metaphor of natural fertilization. Aside from all of the aforementioned pros and strengths of this variant, the SSO has few cons in terms of its major performance. On the other hand, this algorithm needs enhancement to work on classification problems. Based on that, in this paper, we are motivated to incorporate it with the "Density based Feature Selection (DFS)" for classification purpose.

3 Proposed Approach

The D-SSO algorithm is depicted in Fig 4. Preprocessing, FS, and classification are three stages of the proposed work. Because the database may include noisy and redundant data, the preprocessing stage is the most critical step. Different processes are carried out as a result of examining the data, including filling in missing values and removing excess values, all of which degrade performance. There are a total of 24 features in this work, and DFS is used to select a few of them. The wrapper method aims to find the best subset of features. It continuously generates a set of features until the best subset is found by DFS. To classify the data as the presence of Parkinson's disease or the absence of Parkinson's disease, an SSO-based classification approach is used to indicate the acquired vector of features. The D-SSO algorithm combines DFS and SSO, which allows users to predict and diagnose Parkinson's disease or any type of disease. The proposed D-SSO will achieve optimal measurements and high performance of classification with few features. In addition, the D-SSO algorithm's process is illustrated in Fig. 4, and the pseudo-code is provided in Algorithm 2 with parameter settings.

Preprocessing. For data mining processes to function effectively and affordably, the quality of the data must be trustworthy. The Parkinson's dataset as a whole needs to have the database's missing values filled in. In some cases, the methods can be synchronized to produce discrete traits when continuous features are present. There are some noisy and missing values in each scenario. Preprocessing is done on the initial data to enhance how medical data behaves [13].

Optimal FS. To choose the best features, the proposed algorithm uses the following procedures. In this case, DFS is used, and each iteration chooses a set of features. A subset of the ideal features from the raw dataset is the feature that matters most for the classification process. A heuristic method for assessing feature merits is the DFS method.

In general, a feature should be viewed favorably if it overlaps with the other classes less frequently. When exploring and determining ranks, the DFS algorithm takes into account the distribution of features among classes as well as their correlation. The first step in DFS involves computing the "Probability Density Function (PDF)" of each feature in each class individually. The features are then ranked according to the extent of their overlap. The two main methods for calculating various types of PDF are non-parametric and parametric. The first approach makes the assumption that the data has a Gaussian distribution, so the task of density estimation is restricted to figuring out the distribution's proper mean and variance values.

Contrarily, nonparametric approaches compute the density directly from the instances rather than making any assumptions about the shape of the density function. For estimating the density of primary data, many pattern recognition applications lack a set format. However, non-parametric techniques can be applied to distributions of random regardless of the shape of the underlying densities. Therefore, the suggested method uses a parametric approach and is known as follows:

$$p(x) \cong \frac{k}{NV} \qquad (6)$$

where k denotes the number of instances in V, N denotes the overall number of instances, and p(x) denotes the obtained PDF value for instance x. The precise PDF can be found by using increased N and decreased V. The next step is to assess the feature's value using the calculated PDFs across classes after each class's PDF has been estimated. A feature is deemed effective when there is less overlap between each class and the other classes, as was previously mentioned. Estimates of the overlap between a particular feature instances of classes are made using PDFs for each feature and class label. As the overlapping zone for a feature increases, its significance for class label prediction declines, and its consideration results in worsened performance of classification. The value of overlapping for a feature f in class cl is determined by Eq. (7).

$$Overlapping(f, cl) = \int Min\ (PDF(cl), Max\ (PDF(cl_j))) \qquad (7)$$

where $1 < j \leq num_{classes}\ and\ j \neq cl.$

Classification of Parkinson. For the classification task, our algorithm, called “Sperm Swarm Optimization (SSO)” algorithm [21–28] is used to extract classification rules from sperm behavior and “Data Mining (DM)” methods. This method aims to assign each instance to a class from a set of predefined classes using the values of some features [13]. Eq. (8) generally defines the information discovered during the classification process.

IF < conditions > THEN < class > (8)

In the rule's predecessor ((IF part), (AND)), a logical conjunction operator connects a group of conditions. The rule subsequent lists the predicted classes for cases whose predictor features satisfy each term represented in the rule antecedent.

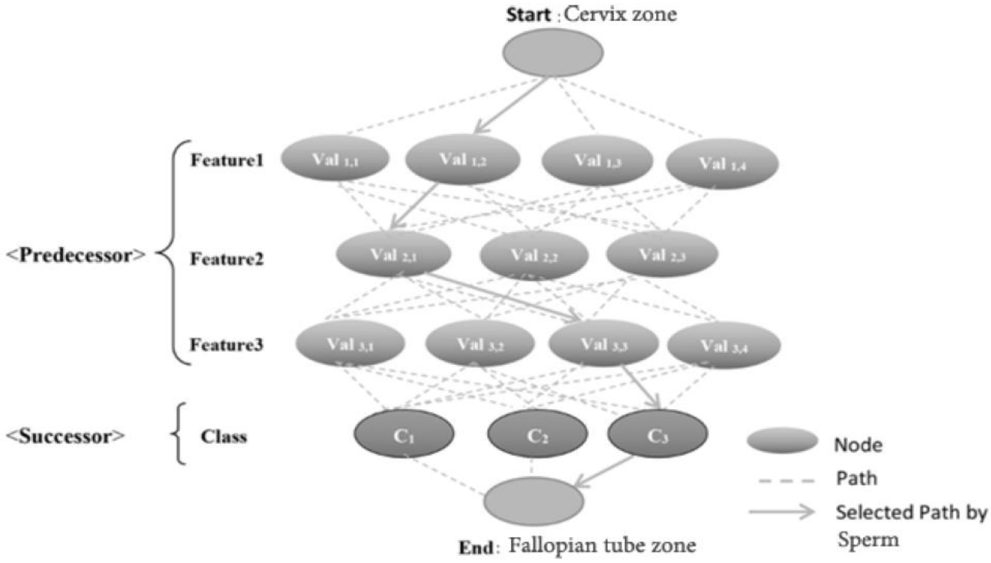

Fig 4: Structural schema of SSO.

The following procedures are involved in the SSO algorithm's application to the Parkinson's classification task:

- Structural schema;
- Rules generation;
- Velocity function;
- Rules pruning;
- Sperm location update;
- Using discovered rules.

Structural schema. The presented classification model's structural schema is shown in Fig. 4. The topmost start node, which acts as a virtual Cervix zone, is where the sperms start their journey. Each feature represented by the lower-level nodes has a unique set of values. A feature is defined as f m and V mn, where I represents the features' series number and j represents the value's series number. The class is the last feature, and the values for the class are written as C k, where k is the sequence value for the class.

The sperm begins its journey from the source, selecting a class value until reaching the Fallopian tube where the Ovum is located there, as shown in the Fig 4. After the traversal process is finished, each class will be assigned a value. A sufficient number of sperms must follow the same path to discover the rules, as explained below. The discovered path is illustrated by a solid line in this case (see Fig. 4): Start-Val1,2-Val2,1-Val3,3-C3-End. The pseudocode of the D-SSO
can be structured as follows:

```
Algorithm 2. "Density based feature selection with Sperm Swarm Optimization (D-SSO)
for Data Classification"

Input: X= {xi, x2, x3„. $x_n$} where n=Total number of instances
Input: F= {f1, f2, f3„, $f_m$} where m=Total number of features
Input: L= {11, 12, 13.„ $l_k$} where k=Total number of class
Intermediate output: $\delta$ ranked features
Final output: Classification accuracy
Begin Algorithm
Step 1: For f = 1 to n do
            Calculate Probability Density Function (PDF) of feature fin each class $L_i(1\leq i\leq k)$
            For L = 1 to k do
                        Add each feature, which are all selected
            End For
End For
Step 2: Initialize Selected Feature in Dataset $D_T$
           Store the discovered rules in rule list ← [ ]
```

```
        While ($D_T$ > Max_UC)  //Training set ($D_T$)
         t ←1   //Sperm index
         calculate the sperm velocity
         update sperm location
         update the winner
**Repeat** the prior steps **Until** (t ≥ No_of_sperms) or (j ≥ No_rules _converge)
                rule $_{\text{best\_Sperm(the winner)}}$ ←ConstructRule()
```

4 Performance Analysis

For the validation of the D-SSO algorithm, An Intel Core i7 running on Windows 10 is used to run MATLAB R2018a on a general-purpose PC with 2TB of storage and 6GB of RAM.

Dataset:

To evaluate the performance of the D-SSO model, a benchmark Parkinson's dataset from the UCI repository is used [29]. The Parkinson's dataset holds a sum of 756 instances with 754 features. All the data descriptions are shown in Table 1.

Table 1: Database Description,

#	Specification	Values
1	Task	Classification
2	Features Characteristics & Datatype	Multivariate & Numerical
3	# of Instances	756
4	# of Features	754
5	# of Classes	2
6	Class datatype	Nominal
7	% of Positive instances	564
8	% of Negative instances	192
9	# of Patients with Parkinson disease	188 (107 Men / 81 Women)
10	Ages	33-87
11	# of Healthy individuals	64 (23 men / 41 women)
12	Ages	41-82
13	Data Source	UCI ML Repository

For features, the wavelet transform based features, vocal fold features, and TWQT features are depended. All the details of features are shown in Table 2.

Table 2: Features Description.

#	Features Information	
1	**Baseline Features**	21
2	Parameters of Intensity	3
3	Formant Frequencies	4
4	Parameters of Bandwidth	4
5	Vocal Fold	22
6	MFCC	84
7	Wavelet Features	182
8	TQWT Features	432

As shown in Table 2. The baseline features are the most important ones for the classification task. A research work reported that they have considered only those 21 baseline features ignoring the rest completely. The authors as long as what can recall, made around 6 experiments with different classifiers from different families, and got accuracies range from 71.xx% to 87.xx% to 90.xx% to 93.5%. One work considered both baseline and vocal features and ended up in the very same 21 baseline features. One other work considered all 754 features and came up with all (21) baseline features, (5) vocal fold, (14) MFCC, and (30) from both Wavelet and TQWT features, being selected for the classification task. Selected features (ALL of which are Baseline Features) by US, as shown in Table 3, which the 21 features that are used in this work.

Table 3: Selected features.

#	**Feature Name**
1	PPE
2	DFA
3	RPDE
4	numPulses
5	numPeriodsPulses
6	meanPeriodPulses
7	stdDevPeriodPulses
8	locPctJitter
9	locAbsJitter
10	rapJitter
11	ppq5Jitter
12	ddpJitter
13	locShimmer
14	locDbShimmer
15	apq3Shimmer
16	apq5Shimmer
17	apq11Shimmer
18	ddaShimmer
19	meanAutoCorrHarmonicity
20	meanNoiseToHarmHarmonicity
21	meanHarmToNoiseHarmonicity

In Fig. 5(a), eF chart represents number of selected features under 1000 iterations and Fig. 5(b), comparing accuracy of selected features over 1000 iterations.

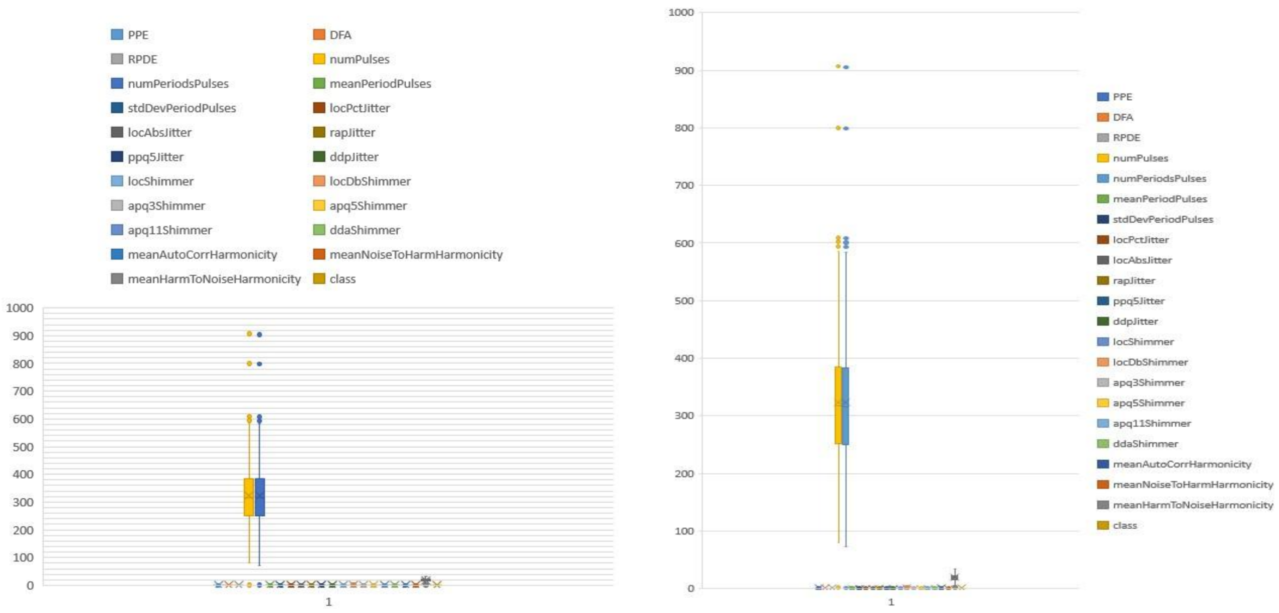

Fig 5: (a) Number of selected features under 1000 iterations and (b) comparing accuracy of selected features over 1000 iterations.

The rest of features are almost ineffective in getting higher accuracy rate. Baseline, Vocal, MFCC, Wavelet, TQWT are not accurate, which got (49% ~ 56% ~ 69%) accuracy respectively. Baseline, Vocal, Wavelet, TQWT are not accurate, which got (51% ~ 54% ~ 71%) accuracy respectively. Baseline, MFCC, Wavelet, TQWT are not accurate, which got (44% ~ 53% ~ 65%) accuracy respectively. Baseline, Vocal, MFCC are not accurate, which got (34% ~ 59%) accuracy respectively. Baseline, Wavelet, TQWT are not accurate, which got (71% ~ 87% ~ 90%) accuracy respectively. On the other hand, with a clear presence of overfitting, Baseline, Wavelet are not accurate, which got 90% accuracy with overfitting. Baseline, and TQWT are not accurate, which got (88%) accuracy with overfitting. Wavelet, and TQWT are not accurate, which got (44%) accuracy. Vocal, and MFCC are not accurate, which got 41% accuracy. Baseline are not accurate, which got (87% ~ 90% ~ 93.5%) accuracy respectively. The sample frequency distribution and class distribution of the 21 features are shown in Fig. 6.

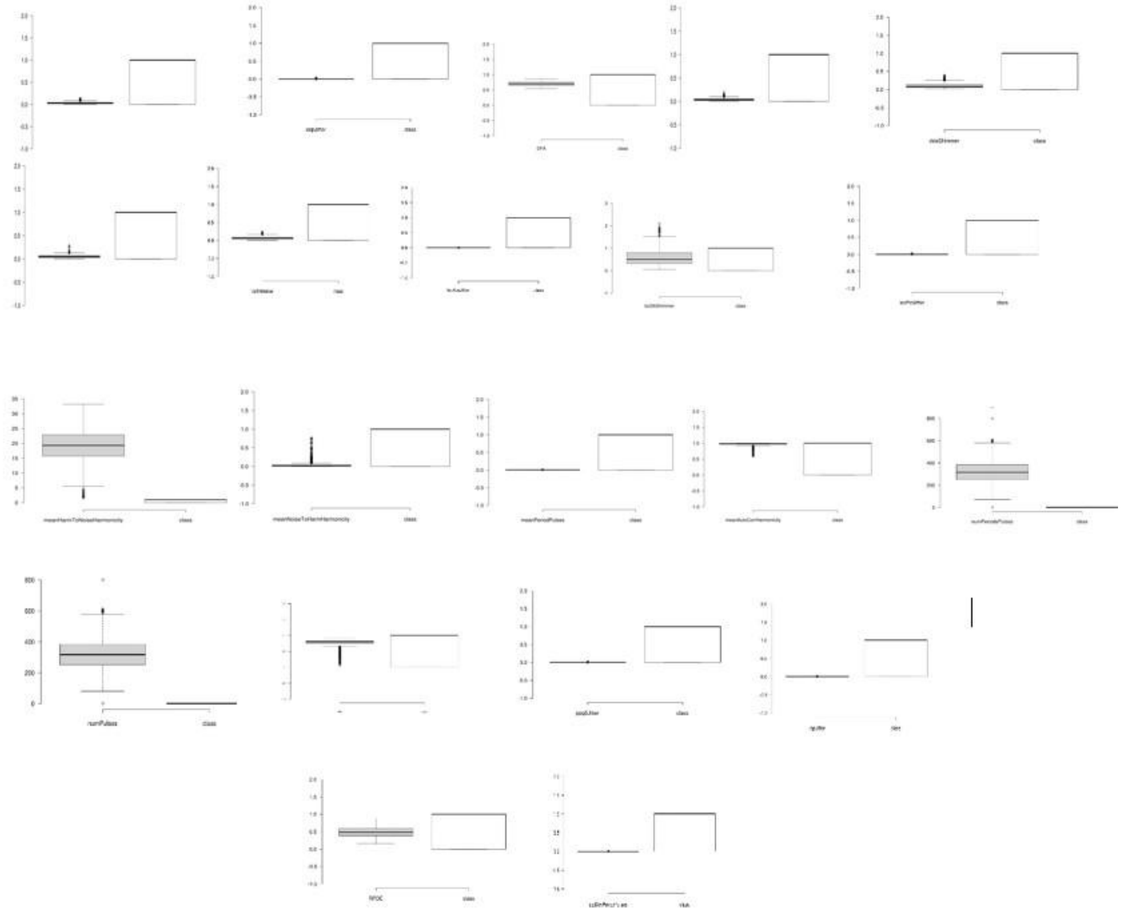

Fig 6: Sample class distribution of 21 features.

5 Results Analysis and Performance

To highlight the accuracy of the D-SSO method on the Parkinson's dataset, a set of performance measures include accuracy, Kappa, "True Positive rate (TPR)", "False Positive rate (FPR) rate", precision, recall, F-score, and ROC area. Before going into detail on the evaluation criteria, the concept of a confusion matrix is covered. In order to evaluate the classification performance of any classifiers, confusion matrix is crucial. The facts on the existing and anticipated classifications are presented in a 2×2 matrix. The confusion matrix consists of four parts as follows, FN, FP, TN, and TP. Using these four elements, the classification measures can be defined as given in Fig. 7 and Fig. 8.

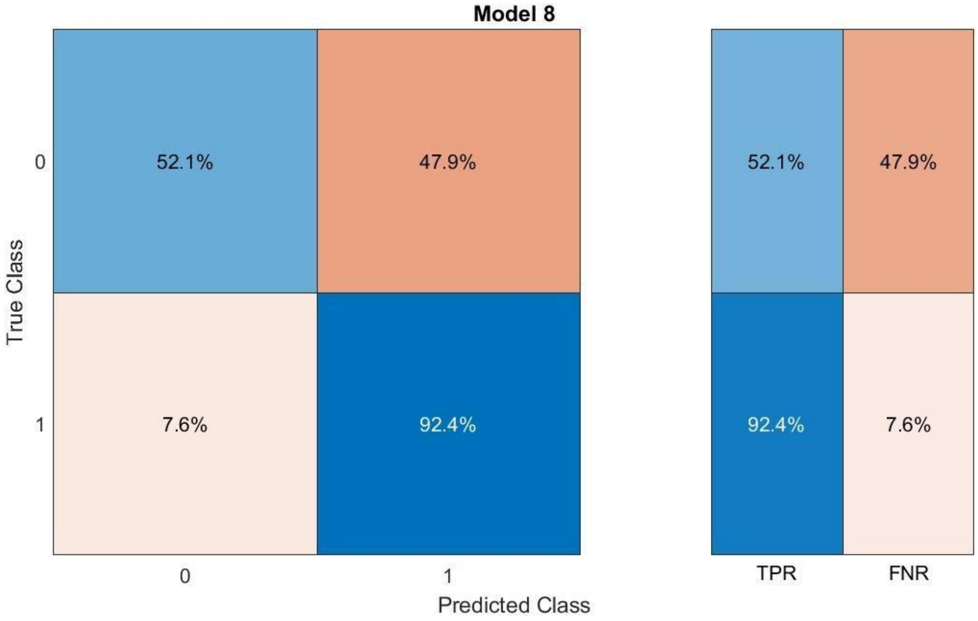

Fig 7: The comparison results for the Classification measures for Predicted class.

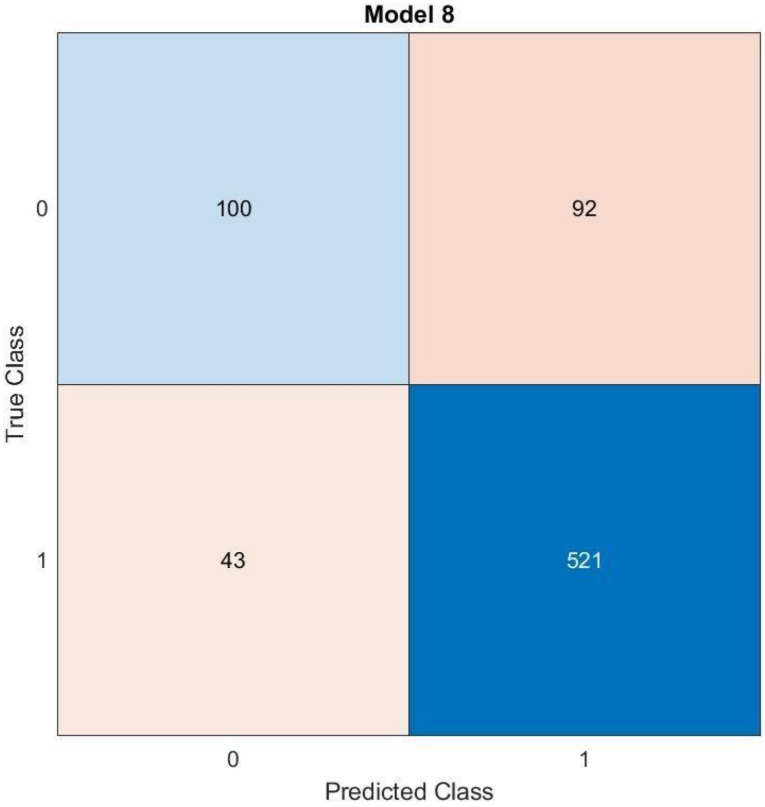

Fig 8: The results for the classification for True and Predicted classes.

The comparison of classification performance is reported in Table 4. The main proves for training and testing procedures are done as follows: The percentage splitting 70% x 30% training set vs. testing set including validation set. Percentage Splitting 80% x 20% training set vs. testing set including validation set overfitting was an issue in many cases, so that the proposed algorithm addressed this issue and solve it by implementing. Then, the best-obtained accuracy and avoiding the model overfitting are achieved by cross validation K= 10 folds. The proposed algorithm got 90% with K= 5 folds. Then the proposed algorithm gained a better accuracy 93.5% with k=10 folds. The enhanced results of D-SSO algorithm are due to the addition of SSO, which eliminates the unrequired features to improve results of the classifier. Comparative results of diverse models of classification involving different metrics under different measures are given previously to prove the efficiency of the proposed algorithm. Finally, the outcomes of various algorithms of classification on Parkinson's dataset in terms of different performance measures revealed that the proposed D-SSO method is found to be accurate on the classification of Parkinson's dataset. This is because of the pros of DFS as well as the features of wrapper method, which continuously operate the DFS and SSO algorithm consecutively.

Table 4: Performance Evaluation of Parkinson using D-SSO method with different classifiers.

	NB	KNN	PSO	ACO	SVM	LR	D-SSO
Accuracy	34.78%	68.91%	75.26%	75.52%	76.19%	78.17%	93.57%
Kappa	0.007	0.186	0.154	0.166	0.201	0.308	0.743
TP rate	0.348	0.689	0.753	0.755	0.762	0.782	0.930
FP rate	0.335	0.501	0.634	0.626	0.603	0.524	0.103
Precision	0.632	0.692	0.713	0.718	0.731	0.761	0.930
Recall	0.348	0.689	0.753	0.755	0.762	0.782	0.930
F-Measure	0.321	0.690	0.700	0.705	0.717	0.754	0.929
ROC Area	0.580	0.603	0.559	0.565	0.579	0.624	0.957

6 Conclusion

Our algorithm, namely "Sperm Swarm Optimization (SSO)" is a promising bioinspired based method that is inspired by the motility of swarm of sperm. This method proved its efficiency in solving various optimization problems, but requires an enhancement to work on classification problems. This study proposes the DFS with SSO method, also known as the D-SSO algorithm, for the classification of the Parkinson's dataset. It is an intelligent system for classification and prediction in the healthcare industry. Hence, the proposed D-SSO framework performs SSO-based learning while simultaneously eliminating unnecessary features. By utilizing a benchmark of Parkinson's dataset, the efficacy of the DSSO algorithm is estimated, and existing approaches are also compared. The proposed D-SSO method outperformed the other classification approaches in a variety of ways, with improved classification performance when compared to the current methods. Overall, the D-SSO method is determined to be a suitable classifier for the detection of Parkinson's disease.

Conflicts of Interest:
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Author contributions:
The concept, design and evaluation of medical system is done by Hisham A. Shehadeh, Ghaith M. Jaradat. Dyala Ibrahim structures the paper and devices experiments with selected benchmarking dataset; Rami Sihwail, Iqbal H. Jebril, Husam Al Hamad and Mohammad Alia contribute in the results evaluation.

Funding:
There is no any fund for this research.

Declarations:
Competing interests The authors declare that they have no competing interests.

Human and Animal Ethics:
“The authors of this paper undersign and certificate that the research conducted complied with the ethical standards in accordance with Amman Arab University, as well as national regulations in the field.”

Ethical Approval & Consent for publication:
We give our consent for the publication of identifiable details, which can include photograph(s) and/or videos and/or case history and/or details within the text (“Material”) to be published in the above Journal and Article. We confirm that we have seen and been given the opportunity to read both the Material and the Article (as attached) to be published by your journal. In Addition, a sample of data of this paper will be available upon request.
The **open source code** of our algorithm, namely, **SSO** is available via the following links:
[*] https://www.mathworks.com/matlabcentral/fileexchange/92150-sperm-swarm-optimization-sso
[**] https://www.mathworks.com/matlabcentral/fileexchange/92130-hssogsa

Data Availability Statements
The data that supports the findings of this study are openly available in [UCI Machine Learning Repository] at
https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification, reference number [29].

References

[1] Jaradat, Y., Masoud, M., Jannoud, I., Manasrah, A., Zerek, A. (2021). Comparison of Genetic Algorithm Crossover Operators on WSN Lifetime. 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering, MI-STA 2022 - Proceeding, 2022, pp. 356–360.

[2] Hamdan, M., Bani-Yaseen, M., & Shehadeh, H. A. (2018) Multi-objective Optimization Modeling for the Impacts of 2.4-GHz ISM Band Interference on IEEE 802.15. 4 Health Sensors. In Information Innovation Technology in Smart Cities, Springer, Singapore, (pp 317–330). ACM. https://doi.org/10.1007/978-981-10-1741-4_21

[3] Yassein, M. B., Hamdan, M., Shehadeh, H. A., and Mrayan, L. A. (2017) Novel Approach for Health Monitoring System Using Wireless Sensor Network. International Journal on Communications Antenna and Propagation (IRECAP) 7(4), 271. https://doi.org/10.15866/irecap.v7i4.11311

[4] Tysnes OB, Storstein A (2017) Epidemiology of Parkinson's disease. J Neural Transm (Vienna) 124(8), 901–905. https://doi.org/10.1007/s00702-017-1686-y

[5] Kalia, L. V., Lang, A. E. (2015) Parkinson's disease. the lancet 386(9996), 896–912, https://doi.org/10.1016/S0140-6736(14)61393-3

[6] Miall, R. C. (2013). Basal Ganglia: Basic Principles. In: Pfaff, D.W. (eds) Neuroscience in the 21st Century. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-1997-6_37

[7] Erro, R., & Reich, S. G. (2022). Rare Tremors and Tremors Occurring in Other Neurological Disorders. Journal of the Neurological Sciences, 435(1),120200. https://doi.org/10.1016/j.jns.2022.120200

[8] Zahed, H., Zuzuarregui, J. R. P., Gilron, R. E., Denison, T., Starr, P. A., & Little, S. (2021) The Neurophysiology of Sleep in Parkinson's Disease. Movement Disorders:36(7):1526–1542. https://doi.org/10.1016/j.clinph.2012.02.043

[9] Khater, B. S., Abdul Wahab, A. W., Idris, M. Y. I., Hussain, M. A., Ibrahim, A. A., Amin, M. A., & Shehadeh, H. A. (2021) Classifier Performance Evaluation for Lightweight IDS Using Fog Computing in IoT Security. Electronics, 10(14),1633. https://doi.org/10.3390/electronics10141633

[10] Abu Khurmaa, R., Aljarah, I., Sharieh, A. (2021) An Intelligent Feature Selection Approach Based on Moth Flame Optimization for Medical Diagnosis. Neural Computing and Applications 33(12):7165–204. https://doi.org/10.1007/s00521-020-05483-5

[11] Goel, N. (2021). Performance Analysis of Classification Techniques with Feature Selection Method for Prediction of Chronic Kidney Sisease. In Innovations in Digital Branding and Content Marketing, IGI Global, pp. 220–244.

[12] Xue, B., Zhang, M., Browne, W. N., Yao, X. (2015) A Survey on Evolutionary Computation Approaches to Feature Selection. IEEE Transactions on Evolutionary Computation 20(4):606–26. https://doi.org/10.26686/wgtn.14214497.v1

[13] Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease. Scientific reports, 9(1):1–14. https://doi.org/10.1038/s41598-019-46074-2

[14] Cai, Z., Gu, J., Wen, C., Zhao, D., Huang, C., Huang, H., & Chen, H. (2018). An Intelligent Parkinson’s Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach. Computational and mathematical methods in medicine 2018(1):1–24. https://doi.org/10.1155/2018/2396952

[15] Mathur, R., Pathak, V., Bandil, D. (2019) Parkinson Disease Prediction Using Machine Learning Qlgorithm. In Emerging Trends in Expert Applications and Security, Springer, Singapore (pp 357–363), Springer. https://doi.org/10.1007/978-981-13-2285-3_42

[16] Zuo, W. L., Wang, Z. Y., Liu, T., Chen, H. L. (2013) Effective Detection of Parkinson's Disease Using an Adaptive Fuzzy K-Nearest Neighbor Approach.

Biomedical Signal Processing and Control 8(4), 364–373. https://doi.org/10.1016/j.bspc.2013.02.006

[17] Zhao, H., Wang, R., Lei, Y., Liao, W. H., Cao, H., Cao, J. (2022) Severity Level Diagnosis of Parkinson’s Disease by Ensemble K-Nearest Neighbor Under Imbalanced Data. Expert Systems with Applications, 189(1),116113. https://doi.org/10.1016/j.eswa.2021.116113

[18] Pahuja, G., Nagabhushan,T. N. (2021) A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection. IETE Journal of Research 67(1),4–14. https://doi.org/10.1080/03772063.2018.1531730

[19] Asmae, O., Abdelhadi, R., Bouchaib, C., Sara, S., Tajeddine, K. (2020, April) Parkinson’s Disease Identification Using KNN and ANN Algorithms Based on Voice Disorder. In IEEE 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, pp 1–6, IEEE. https://doi.org/10.1109/iraset48871.2020.9092228

[20] Gupta, D., Sundaram, S., Khanna, A., Hassanien, A. E., De Albuquerque, V. H. C. (2018) Improved diagnosis of Parkinson's Disease Using Optimized Crow Search Algorithm. Computers & Electrical Engineering, 68(1),412–424. https://doi.org/10.1016/j.compeleceng.2018.04.014

[21] Shehadeh, H. A. (2021) A Hybrid Sperm Swarm Optimization and Gravitational Search Algorithm (HSSOGSA) for Global Optimization. Neural Computing and Applications, 33(18),11739–11752. https://doi.org/10.1007/s00521-021-05880-4

[22] Shehadeh, H. A., Idris, M. Y. I., Ahmedy, I. (2017) Multi-objective Optimization Algorithm Based on Sperm Fertilization Procedure (MOSFP). Symmetry, 9(10), 241. https://doi.org/10.3390/sym9100241

[23] Shehadeh, H. A., Ahmedy, I., Idris, M. Y. I. (2018) Sperm Swarm Optimization Algorithm for Optimizing Wireless Sensor Network Challenges. In Proceedings of the ACM International Conference on Communications and Broadband Networking (ICCBN), Singapore, ACM, pp. 53 59. https://doi.org/10.1145/3193092.3193100

[24] Shehadeh, H. A., Ahmedy, I., Idris, M. Y. I. (2018) Empirical Study of Sperm Swarm Optimization Algorithm. In book: Volume 869 of the Advances in Intelligent Systems and Computing series, Arai K, Kapoor S, Bhatia R (eds.): In Proceedings of SAI Intelligent Systems Conference, Springer, Cham , (pp. 1082–1104), Springer. https://doi.org/10.1007/978-3-030-01057-7_80

[25] Shehadeh, H. A., Mustafa, H. M., Tubishat, M. (2022) A Hybrid Genetic Algorithm and Sperm Swarm Optimization (HGASSO) for Multimodal Functions. International Journal of Applied Metaheuristic Computing (IJAMC), 13(1), 1-33. https://doi.org/10.4018/ijamc.292507

[26] Shehadeh, H. A., Shagari, N. M. (2022) A Hybrid Grey Wolf Optimizer and Sperm Swarm Optimization for Global Optimization. Eds: Manshahia M S, Kharchenko V, Munapo E, Thomas J J, Vasant P. Handbook of Intelligent Computing and Optimization for Sustainable Development, 1:487–507. https://doi.org/10.1002/9781119792642.ch24

[27] Shehadeh, H. A., Idna Idris, M. Y., Ahmedy, I., Ramli R., Mohamed Noor, N. (2018) The Multi-objective Optimization Algorithm Based on Sperm Fertilization Procedure (MOSFP) Method for Solving Wireless Sensor Networks Optimization Problems in Smart Grid Applications. Energies 11(1), 97. https://doi.org/10.3390/en11010097

[28] Shehadeh, H. A., Jebril, I. H., Wang, X., Chu, S. C., & Idris, M. Y. I. (2022). Optimal Topology Planning of Electromagnetic Waves Communication Network for

Underwater Sensors Using Multi-objective Optimization Algorithms (MOOAs). *Automatika*, 63(4), 1–12. https://doi.org/10.1080/00051144.2022.2123761

[29] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Apaydin, H. (2019) A Comparative Analysis of Speech Signal Processing Algorithms for Parkinson’s Disease Classification and the Use of the Tunable Q-Factor Wavelet Transform. Applied Soft Computing, 74(1):255–263. https://doi.org/10.1016/j.asoc.2018.10.022