Whale Optimization-based Prediction for Medical Diagnostic

Ali Asghar Rahmani Hosseinabadi[1], Mehdi Sadeghilalimi[1], Morteza Babazadeh Shareh[2], Malek Mouhoub[1][a] and Samira Sadaoui[1]

[1]University of Regina, 3737 Wascana Parkway, Regina, Canada

[2]Islamic Azad University, Babol Branch, Babol, Iran

Keywords: Feature Importance, Meta-heuristics Optimization, Nature-inspired Techniques, Combinatorial Optimization, Supervised Learning.

Abstract: This study aims to improve disease detection accuracy by incorporating a discrete version of the Whale Optimization Algorithm (WOA) into a supervised classification framework (KNN). We devise the discrete WOA by redefining the related components to operate on discrete spaces. More precisely, we redefine the notion of distance (between individuals in WOA), and propose a random exploration function to include more diversity. The latter includes the random move defined in the WOA algorithm, as well as two other random techniques based on the crossover and mutation operators. To assess the performance of our proposed method, we conducted experiments on two benchmark medical datasets. The results demonstrate the efficacy of the hybrid approach, WOA+KNN.

1 INTRODUCTION

Nowadays, automated diagnostic systems are an integral part of numerous medical applications. Initially, researchers adopted machine learning algorithms to assist physicians in their decision-making tasks. More recently, researchers were motivated to incorporate meta-heuristic optimization methods to improve the prediction outcome and reduce false alarms. For instance, the authors in (Shankar and Manikandan, 2019) proposed a new approach combining the Grey Wolf Optimization Algorithm (GWOA) with Fuzzy Logic. First, the diagnostic model is formed based on a set of Fuzzy rules. Then this set of rules is optimized using GWOA, which is found to be more efficient than the Ant Colony Optimization (ACO) algorithm. In (Alirezaeia et al., 2019), the authors utilized K-means for clustering patient data and removing outliers. Then, they developed four multi-objective meta-heuristic optimization methods that are integrated into a Support Vector Machine (SVM) classifier to select the features more accurately. The experimental results demonstrated that the accuracy increased significantly. The study in (Bhuvaneswari and Manikandan, 2018) proposed a new diagnostic system by devising a classifier called "Temporal Feature Selection and Temporal Fuzzy Ant Miner Tree". A modified version of the Genetic Algorithm was also utilized in the proposed method, to increase diagnosis detection. This new method has a high detection capability compared to other techniques. Lastly, the work in (Giveki and Rastegar, 2019) adopted the RBF-based Neural Network together with the Harmony Search optimization algorithm to improve efficiency in medical diagnosing. Combining these two methods showed a higher performance.

This present study is preliminary work on enhancing machine learning algorithms to improve medical diagnostic performance. Following on previous research works, we develop a hybrid approach based on the Whale Optimization Algorithm (WOA) (Mirjalili and Lewis, 2016) and K-Nearest Neighbors (KNN) algorithm to tackle disease diagnostic more efficiently. Generally speaking, we determine the optimal weights of the predictive features using the discretization of WOA, train KNN on the weighted data, and test the learned model on unseen data. For this purpose, we incorporate a discrete version of WOA intending to optimize the weights of the feature space of the training medical datasets. We define the discrete version of WOA by redefining the related components to operate on discrete spaces. In addition, we devise a random function to better conduct the exploration phase. This latter is based on the Ge-

[a] https://orcid.org/0000-0001-7381-1064

Hosseinabadi, A., Sadeghilalimi, M., Shareh, M., Mouhoub, M. and Sadaoui, S.
Whale Optimization-based Prediction for Medical Diagnostic.
DOI: 10.5220/0010802200003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 3*, pages 211-217
ISBN: 978-989-758-547-0; ISSN: 2184-433X

netic Algorithm's operators to add more diversity. We note that various discrete variants of WOA have been reported in the literature. These variants have been mainly defined to tackle specific learning problems including: image segmentation (Aziz et al., 2017), parameter tuning of neural networks (Aljarah et al., 2018), parameter estimation of solar cells (Oliva et al., 2017), and classical combinatorial problems such as the Knapsack problem (Li et al., 2020) and the Traveling Salesman Problem (TSP) (Zhang et al., 2021). The main difference between our proposed method and the above variants is regarding the discretization of the equations guiding the WOA, as well as the operator we suggest to conduct better exploration. In this regard, we propose a function that performs one of the following three random movements depending on the value of a given random parameter: (1) shrinking towards a random whale, (2) a crossover with a random whale, and (3) a random mutation. Our main objective is to improve medical diagnosis by optimizing the WOA fitness function during the training phase.

We fully implement our hybrid approach WOA+KNN using the MATLAB toolkit. We utilize two benchmark medical diagnostic datasets to evaluate the proposed method performance: the diabetic Type 2 dataset (called PID) and the electrocardiogram dataset (called ECG200). We select these two datasets given that they are challenging; both are of small size, which may under-fit the learned classification models, and the second dataset's size is relatively small compared to its dimensionality.

We organize the paper as follows. The following section describes the original WOA (Mirjalili and Lewis, 2016). Section 3 explains the discretization of WOA and its integration into a supervised classification framework. Section 4 evaluates the proposed hybrid approach using the two medical datasets. Finally, Section 5 lists concluding remarks and ideas for future works.

2 WHALE OPTIMIZATION ALGORITHM (WOA)

WOA is a population-based algorithm drawn from the collective hunting of whales (humpback whales) representing potential solutions (Mirjalili and Lewis, 2016; Sangaiah et al., 2020). In this regard, the foraging behavior of whales is performed via crating bubbles in a spiral manner. Whales movement towards the prey follows an exploitation/exploration strategy (Mirjalili and Lewis, 2016). Exploitation is achieved through shrinking encircling and spiral motions. In the former, each whale approaches the prey by rotating around it. In the second method, each whale approaches the prey by following a spiral curve. Following the assumption that the prey's position is the same or close to the one of the Best Whale (BW), each of the other whales will update its position during a shrinking encircling motion, through the following equations defined in (Mirjalili and Lewis, 2016).

$$\begin{array}{rcl} D & = & |CX^*(t) - X(t)| \\ X(t+1) & = & X^*(t) - AD \end{array} \tag{1}$$

In the above, t and $t+1$ are the current and the next iterations, respectively. X^* and X are respectively the position of BW and a given whale. A and C are computed as follows.

$$\begin{array}{rcl} A & = & 2ar - a \\ C & = & 2r \end{array} \tag{2}$$

a and r are random parameters in [0,2] and [0,1], respectively. Note that a decreases at each iteration from 2 to 0 to achieve shrinking. The Spiral motion is obtained through the following equation (Mirjalili and Lewis, 2016).

$$X(t+1) = D' e^{bl} cos(2\pi l) + X^*(t) \tag{3}$$

In the above, l is a random number in [-1,1], b is a constant used to set the spiral curve. The distance between each whale and the prey, D', is computed as follows (Mirjalili and Lewis, 2016).

$$D' = |X^*(t) - X(t)| \tag{4}$$

In exploration, each whale searches for the prey randomly by updating its position according to a randomly chosen whale (instead of moving towards BW, as done above in shrinking encircling and spiral motion). The related equations for the random move are as follows (Mirjalili and Lewis, 2016).

$$\begin{array}{rcl} D & = & |CX_{rand}(t) - X(t)| \\ X(t+1) & = & X_{rand}(t) - AD \end{array} \tag{5}$$

In the above, X_{rand} is a randomly chosen whale.

Given the above equations guiding the exploitation and exploration strategies, the WOA algorithm works as follows. A random number named p is taken from [0,1]. If p is greater or equal to 0.5, then the whale moves in spiral motion. Otherwise, we need to check the value of A. If $|A|$ is less or equal to 1, then the shrinking encircling operation will be executed. Otherwise, the whale will move randomly as described above. Note that at the beginning of the algorithm, the value of A is likely greater than 1, which

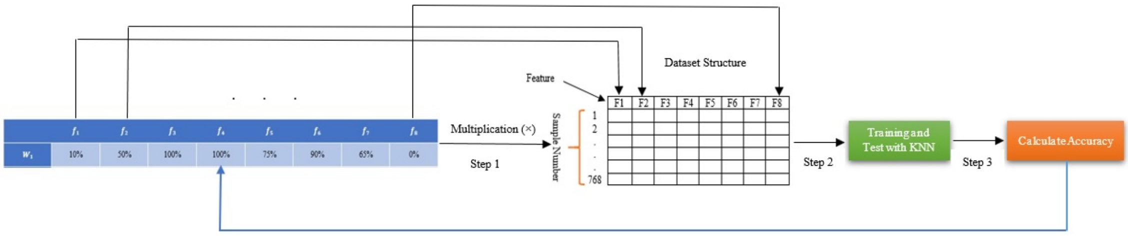


Figure 1: Process of calculating the accuracy (fitness) of an individual.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
W_1	%10	%50	%100	%100	%75	%90	%65	%0
W_{best}	%50	%100	%0	%0	%100	%25	%70	%100
$D = distance(w_1, w_{best})$	-40%	-50%	100%	100%	-25%	65%	-5%	-100%

Figure 2: Distance between two individuals.

will allow more exploration. Then, as the value of a decreases, the value of A will often become less than 1, which will favor more shrinking. This reflects the behaviour of the algorithm with more exploration at the beginning, followed by more exploitation at the end.

Given that WOA was proposed to solve continuous optimization problems (Mirjalili and Lewis, 2016), we need to adapt it to discrete spaces to be able to tune the weights of the features of the patient dataset. The discretization of the different components of WOA is described in the next section.

3 DISCRETIZATION OF WOA

3.1 Individual Representation and Fitness Function

Each individual (whale) is represented with a vector of weights, each corresponding to the degree of importance/participation of the related KNN feature. For instance, if the weight (degree of importance) of a given feature is 100%, then the latter should be fully present in the classification. A weight of 0% indicates that the feature should not be considered in the classification. If the weight is equal to 50%, then the feature should participate in the classification with half of its power. The length of the vector is equal to the dimensionality of the dataset. The left table in Figure 1 shows an example of a vector corresponding to a dataset with eight features. The fitness function of a given whale (individual) corresponds to the accuracy of the KNN model using the weights (features degree of importance) listed in the corresponding vector. Figure 1 shows how the fitness (accuracy) is computed, through the following three steps.

1. The value of each feature weight is multiplied by the corresponding feature column of the dataset. For example, the weight of the first feature (10%) is multiplied by the 768 samples in the first column.
2. The classification will be performed by KNN algorithm according to the vector of weights, following 60% for training and 40% for testing, for instance.
3. The obtained accuracy will be considered as the corresponding fitness value.

3.2 Distance, Spiral and Shrinking Functions

We define the distance between two individuals as the pairwise difference between the entries in the related vectors. Note that the distance is not symmetric. Figure 2 shows an example of a distance between two individuals.

Shrinking and spiral functions are implemented based on the equations we listed in the previous section, and the new notion of distance we defined above. More precisely, at each step of the algorithm, the distance is calculated according to equations 1 or 4 (for each pair of features in both $X^*(t)$ and $X(t+1)$) and a percentage, A, of the result will be subtracted from $X^*(t)$ to get the value for $X(t+1)$. This will allow $X(t)$ to get closer to $X^*(t)$ via shrinking or spiral motion.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
W_1	10%	50%	100%	100%	75%	90%	65%	0%
W_2	50%	100%	0%	0%	100%	25%	70%	100%
$join(w_1, w_2)$	10%	50%	100%	0%	100%	25%	70%	100%

Figure 3: Example of crossover function with crossover point equals to 3.

3.3 Exploration: Random Function

Rather than having one random function (as described in the previous section) where whales move toward a random one via shrinking, we define a random function offering more diversity. The pseudo-code, listed in Figure 4, performs one of the following three operations, depending on the value of a given random parameter *rand*: shrinking (toward a random whale, as in the traditional WOA described in the previous section), crossover, or mutation. The crossover function performs a 1-point crossover (according to a random crossover point) between W_i and another randomly selected whale (W_j). The mutation function selects some values from W_i and then changes them randomly.

```
Random Function (W_i)
rand = a random number in [0,1];
randomly select W_j
if (rand < 0.25)
        W_i = shrinking(W_i, W_j);
elseif (0.25 ≤ rand < 0.5)
        W_i = crossover(W_i, W_j);
else (rand ≥ 0.5)
        W_i = mutation(W_i);
return W_i;
```

Figure 4: Pseudo-code of the random function.

4 VALIDATION

4.1 PID Dataset

Through a well-known benchmark dataset, the Pima Indian Diabetes (PID) (Giveki and Rastegar, 2019), we assess our hybrid method's performance. The PID dataset consists of 768 patients, where all are females aged at least 21 years old. The balanced dataset has eight numerical features listed in Table 1. The target class is either diabetic or healthy. Since the features possess different scales, we normalize all of them to the range of [0, 1]. We tune the WOA's parameters as presented in Table 2. We train the KNN classifier with the training dataset (70%) and then assess its predictive performance on the testing dataset (30%).

Table 1: Features of PID Dataset (Giveki and Rastegar, 2019).

Feature	Diagnosis	Unit
#1	Number of pregnancies	Integer
#2	Plasma glucose concentration	Mg/dl
#3	Diastolic blood pressure	$mmHg$
#4	Triceps skin fold thickness	Mm
#5	2-h serum insulin	MuU/mL
#6	Body mass index	Kg/m^2
#7	Diabetes pedigree function	Integer
#8	Age	Year

Table 2: Parameter Tuning for PID Dataset.

Parameter	Definition	Value
n	Number of whales	200
k	Number of whale movements	500
prob	Possibility of changing whale cell	rand(100)
a	Change the discovery phase to optimization	2
train	Percentage of learning data	70%

The best solution for the diabetes diagnostic returned by the WOA+KNN method is exposed in Table 3. We observe that the two features "Plasma glucose concentration" (F2) and "Blood pressure" (F3) are the most important to identify diabetes. The least relevant feature is the "Amount of insulin" (F5). The values of the columns in the normalized PID dataset must be multiplied by the optimal values of Table 3 to increase the detection accuracy. As observed in Table 4, WOA+KNN outperforms KNN across all the quality metrics, with an increase of 9.84% in Accuracy and 12.22% in F1-score. This increase is important in medical diagnostic to increase the disease diagnostic.

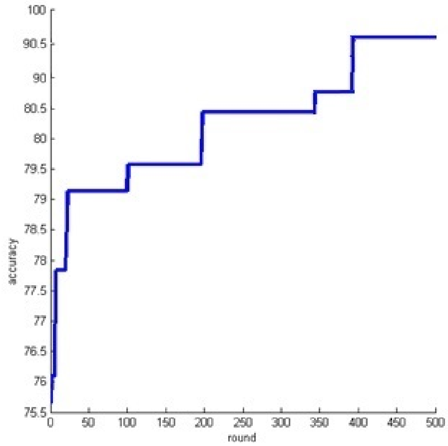
Table 3: Best Solution for PID Dataset.

$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	$F8$
39	52	52	11	2	47	22	35

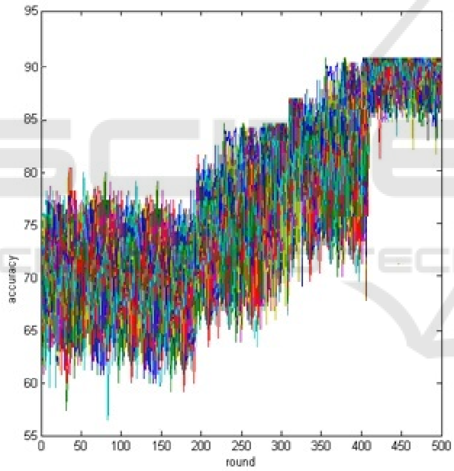
One of the essential points to consider in the proposed algorithm is the trend of changes in the best answer's value. To this end, the best whale's fit value,

Table 4: Predictive Performance with/without optimization for PID Dataset.

Classifier	Accuracy	Recall	Precision	F1-Score
WOA+KNN	91.29	92.79	91.18	91.97
KNN	81.45	81.29	78.27	79.75



(a)



(b)

Figure 5: (a) Change in amount of fit to the best answer, (b) Process of changes in fit of all whales.

in each round of the algorithm, is kept in an array, and a diagram is then drawn from it. Figure 5 depicts the best fit changes during the 500 rounds of the algorithm run. As we can observe, the fit value is optimized gradually during run-time. After each iteration, the next optimization occurs at a greater distance. In fact, the higher the performance of the algorithm, the more difficult the optimization response becomes. The whale's hunting inspires the WOA behavior. Consequently, its function should be such that, eventually, all the whales get closer to the prey. In other words, at the end of the algorithm, all whales must be close to the optimal whale. This approach should happen progressively. Figure 5 illustrates the process of changes in the fit of all the whales. As seen in the figure, the whales are initially far from the optimal answer but eventually move around the best whales.

4.2 ECG200 Dataset

Table 5: Parameter Tuning for ECG200 Dataset.

Parameter	Definition	Value
n	Number of whales	100
k	Number of whales' movements	100
prob	Possibility of changing whale cell	rand(100)
a	Change the discovery phase to optimization	2
train	Percentage of learning data	50%

We experiment with the proposed approach this time using a higher dimensional dataset, called ECG200, described in (Ding et al., 2008). ECG200 is the most frequently used benchmark dataset for evaluating time-series. It consists of 200 ECG signals and 96 quantitative features. For each timestamp, a record reflects one heartbeat signal. Out of the 200 records, 133 were annotated as Normal, while 67 as Abnormal (cardiac disease). The imbalanced class distribution ratio is low, so no need to re-balance the dataset.

We apply WOA+KNN to the EGC dataset and set the parameters presented in Table 5. The method optimizes the weights of the entire feature space, and in Table 6, among the 96 features, we expose the top 40 features and their weights. Features F#8 and F#28 are the most relevant to the target class.

In Table 7, we compute the average for each metric for ten runs. The hybrid method WOA+KNN attained high Accuracy of 97% and F1-Score of 97.67%, even though we are dealing with high dimensionality. The outcome of the KNN classifier without any feature optimization is only 89% for Accuracy and 88% for F1-score. As seen, using the metaheuristic optimization technique improved the KNN classification performance. However, WOA+KNN took much more time, which is expected.

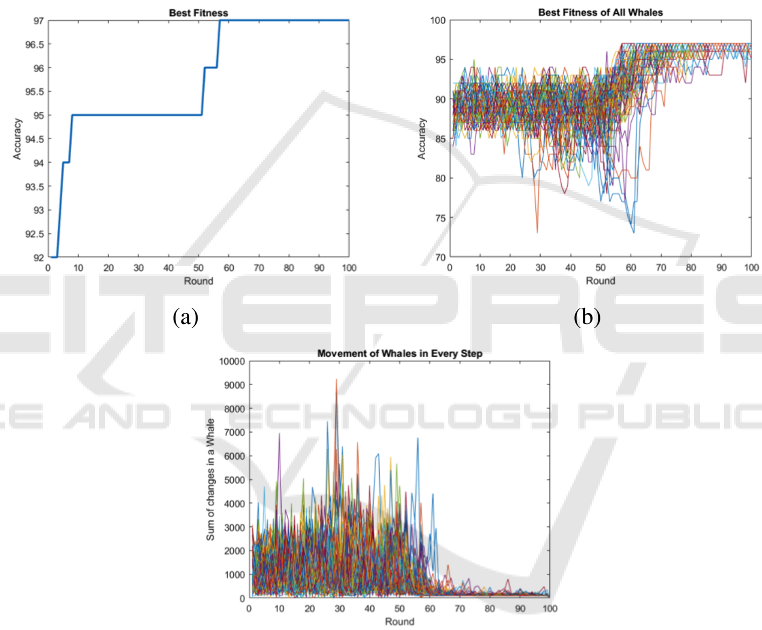
The research (Anowar et al., 2021) adopted different categories of feature extraction methods, such as unsupervised vs. supervised, linear vs. non-linear, and manifold vs. random projection, which is combined with the supervised classification framework

Table 6: Best Solution For ECG200 Dataset (top 40 features).

F1 55	F2 32	F3 51	F4 38	F6 57	F8 87	F11 38	F14 45	F15 75	F16 32
F17 78	F21 44	F24 45	F25 50	F28 96	F29 30	F31 52	F32 30	F33 40	F36 38
F37 71	F38 31	F41 80	F42 67	F44 41	F46 55	F51 58	F56 57	F62 70	F65 32
F66 35	F67 59	F72 62	F75 34	F77 58	F78 38	F79 44	F81 46	F85 69	F93 38

Table 7: Predictive Performance with/without optimization for ECG200 Dataset.

Classifier	**Accuracy**	**Recall**	**Precision**	**F1-Score**	**Time**
WOA+KNN	97	98.43	96.93	97.67	180s
KNN	89	88	89	88.49	4s



(a) (b)

(c)

Figure 6: (a): Best whale's fitness in each round, (b): Whales' fitness in each round, (c): Whales' movement in each round.

Kernel SVM. More precisely, the authors assessed and compared the performances of KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE using the same electrocardiography dataset. The highest F1-score of 90.03% was obtained with the Kernel PCA+Kernel SVM. Therefore, the WOA+KNN approach may outperform dimensionality reduction methods.

Figure 6(a) shows the fitness value of the best whale in each round. At the beginning of the algorithm, the best fitness is 92% but in less than 10 rounds, we reach the promising fitness of 95%. The nest improvement is somewhere between 50 and 60. The last optimization is after round 60 and beyond that there is no any optimization that occurred. Finally, the last and best fitness after 100 rounds of execution is 97%. Figure 6(b) depicts the value of fitness of all the whales. Due to the highly random behavior of algorithm in the early rounds, the fitness values of solutions are highly scattered, but gradually the algorithm moves from exploration phase to exploitation phase and as a result all the fitness values tend to be optimized. In the last round, all the solutions are near the best answer. Figure 6(c) shows the difference in whale motion at each stage of motion. The vertical axis is the sum of the current and previous differences of the whale vector. Here, the whales move quickly in the middle of the algorithm and get close to the prey.

5 CONCLUSION AND FUTURE WORK

Our study is ongoing research that aims to improve medical diagnostic performance by fusing a discrete version of the meta-heuristic optimization method WOA with supervised classification. We designed the discrete WOA by redefining the related components for discrete spaces and a new exploration function to include more diversity. The whale fitness is calculated based on the classification accuracy using the KNN classifier. The feature space is re-scaled based on the whales' values before the learning task. The experimental results demonstrated that the WOA+KNN approach increased the performance of machine learning algorithms.

One exciting but challenging research direction is to incorporate WOA and other nature-inspired techniques (Mouhoub and Wang, 2008; Bidar et al., 2018a; Abbasian et al., 2011; Bidar et al., 2018b; Hmer and Mouhoub, 2016) to the incremental learning setting where the classifier is updated gradually with new observations but without re-training from scratch.

REFERENCES

Abbasian, R., Mouhoub, M., and Jula, A. (2011). Solving graph coloring problems using cultural algorithms. In Murray, R. C. and McCarthy, P. M., editors, *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference, May 18-20, 2011, Palm Beach, Florida, USA*. AAAI Press.

Alirezaeia, M., Niaki, S. T. A., and Niaki, S. A. A. (2019). A bi-objective hybrid optimization algorithm to reduce noise and data dimension in diabetes diagnosis using support vector machines. *Expert Systems with Applications*, 127:47–57.

Aljarah, I., Faris, H., and Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Comput.*, 22(1):1–15.

Anowar, F., Sadaoui, S., and Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, 40:100378.

Aziz, M. A. E., Ewees, A. A., and Hassanien, A. E. (2017). Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. *Expert Systems with Applications*, 83:242–256.

Bhuvaneswari, G. and Manikandan, G. (2018). A novel machine learning framework for diagnosing the type 2 diabetics using temporal fuzzy ant miner decision tree classifier with temporal weighted genetic algorithm. *Computing*, 100:759—772.

Bidar, M., Kanan, H. R., Mouhoub, M., and Sadaoui, S. (2018a). Mushroom reproduction optimization (mro): a novel nature-inspired evolutionary algorithm. In *2018 IEEE congress on evolutionary computation (CEC)*, pages 1–10. IEEE.

Bidar, M., Sadaoui, S., Mouhoub, M., and Bidar, M. (2018b). Enhanced firefly algorithm using fuzzy parameter tuner. *Comput. Inf. Sci.*, 11(1):26–51.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008). Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1(2):1542–1552.

Giveki, D. and Rastegar, H. (2019). Designing a new radial basis function neural network by harmony search for diabetes diagnosis. *Optical Memory and Neural Networks*, 28(4):321—-331.

Hmer, A. and Mouhoub, M. (2016). A multi-phase hybrid metaheuristics approach for the exam timetabling. *International Journal of Computational Intelligence and Applications*, 15(04):1–22.

Li, Y., He, Y.-C., Liu, X., Guo, X., and Li, Z. (2020). A novel discrete whale optimization algorithm for solving knapsack problems. *Applied Intelligence*, 50:3350–3366.

Mirjalili, S. and Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95:51 – 67.

Mouhoub, M. and Wang, Z. (2008). Improving the ant colony optimization algorithm for the quadratic assignment problem. In *2008 IEEE Congress on Evolutionary Computation*, pages 250–257. IEEE.

Oliva, D., Abd El Aziz, M., and Ella Hassanien, A. (2017). Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm. *Applied Energy*, 200:141–154.

Sangaiah, A. K., Hosseinabadi, A. A. R., Shareh, M. B., Bozorgi Rad, S. Y., Zolfagharian, A., and Chilamkurti, N. (2020). Iot resource allocation and optimization based on heuristic algorithm. *Sensors*, 20(2).

Shankar, G. S. and Manikandan, K. (2019). Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization. *Pattern Recognition Letters*, 125:432–438.

Zhang, J., Hong, L., and Liu, Q. (2021). An improved whale optimization algorithm for the traveling salesman problem. *Symmetry*, 13(1).