Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

HYBRIDIZATION OF ARITHMETIC OPTIMIZATION WITH GREAT DELUGE ALGORITHMS FOR FEATURE SELECTION PROBLEMS IN MEDICAL DIAGNOSIS

Mohammed Alweshah

(Received: 13-Dec.-2021, Revised: 24-Jan.-2022, Accepted: 13-Feb.-2022)

ABSTRACT

In the field of medicine, there is a need to filter data to find information that is relevant for specific research problems. However, in the realm of scientific study, the process of selecting the appropriate data or features is a substantial and challenging problem. Therefore, in this paper, two wrapper feature selection (FS) methods based on novel metaheuristic algorithms named the arithmetic optimization algorithm (AOA) and the great deluge algorithm (GDA) were used to attempt to tackle the medical diagnostics challenge. Two methods, AOA and AOA-GD were tested on 23 medical benchmark datasets. According to all of the experimental data, the hybridization of the GDA with the AOA considerably increased the AOA's search capability. The AOA-GD method was then compared with two previous wrapper FS approaches; namely, the coronavirus herd immunity optimizer with greedy crossover operator (CHIO-GC) and the binary moth flame optimization with Lévy flight (LBMFO_V3). When applied to the 23 medical benchmark datasets, the AOA-GD achieved an accuracy rate of 0.80, thereby surpassing both the CHIO-GC and LBMFO V3.

KEYWORDS

Medical diagnosis, Feature selection, Arithmetic optimization algorithm, Great deluge algorithm.

1. INTRODUCTION

In the last decade, artificial intelligence (AI) has undergone significant development and there are signs that it has already reached the level of being able to give genuine solutions to healthcare problems, heralding the dawn of a revolution in the field of medicine [1]. However, the use of AI raises some challenges, which mostly concern the extent of the ability of AI to simulate human skills, such as logical thinking. Moreover, it excels at analyzing huge data and reaching correct scientific findings throughout record durations. In recent years, a broad variety of AI initiatives have been proposed for gathering and analyzing massive amounts of health data, the most significant of which is machine learning (ML) [2]-[3].

Machine learning is now widely utilized for evaluating medical data and much work has been done in the area of medical diagnosis to address specific diagnostic issues [4]-[5]. In specialist hospitals or departments, data on proper diagnoses is frequently available in the form of medical records [3], [6]. All that is required is to enter the patient data with known proper diagnoses into a computer software program that then runs a learning process [7]-[9]. However, there is still a need to improve classifier performance, which has led to the usage of the feature selection (FS) approach as a means of simplifying the already available classifiers [10]-[12].

In data preprocessing, FS is an essential technique that is used to identify a sub-set of associated attributes. Feature selection is particularly important in supervised learning, because it optimizes a specific function to improve prediction accuracy by picking the relevant features in a given class label. For the goal of optimizing the prediction model, several FS approaches are utilized and have been developed [13]-[14].

To put it simply, the FS process finds and retains only those features that are the most relevant to the problem at hand [15]. By eliminating irrelevant as well as common characteristics, the FS method decreases the number of features that a classifier has to learn, which then reduces the training time and the number of features that the classifier has to evaluate, thereby resulting in improved classification performance [16]-[17]. The FS method is used to choose the optimal sub-set of features from the whole feature space in order to provide the necessary elucidations about the learning operations [18].

M. Alweshah is with Prince Abdullah Bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt, Jordan. Email: weshah@bau.edu.jo

Feature selection consists of four major steps: starting up, searching, evaluating sub-sets of features and reaching the stopping condition [19]. To date, several FS approaches have been developed and utilized to attempt to optimize prediction models [20].

Researchers are also constantly attempting to find ways to enhance the accuracy of ML by utilizing another algorithm with a classifier algorithm in the learning model [21]-[23]. To express a single metaheuristic mechanism as an optimizer, one algorithm is used in conjunction with the learning model [24]. Many research studies have recently proven the efficacy of similar algorithms in achieving improved outcomes and in improving the method for picking a variety of characteristics [1]. Nature has inspired some of the most successful metaheuristics [25]. When tracing the search method, metaheuristics takes the information obtained into account [10]. Furthermore, it generates new ways of connecting a single, efficient approach or more [26].

In this paper, the arithmetic optimization algorithm (AOA), a novel metaheuristic algorithm created by Abualigah et al. in 2020 [27] for the field of medical diagnostics is utilized to address FS concerns. Multiplication, division, subtraction and addition are the arithmetic operators employed by the AOA; these operators reflect the standard calculation techniques used to investigate numbers. These basic operators are used as a mathematical optimization to choose the assessment that can help from a set of candidate replacements consisting of a set of criteria (solutions).

The AOA is employed in two techniques in this study to choose the most valuable and first used qualities in medical datasets in their basic form. Then, the AOA is hybridized with the great deluge algorithm (GDA) [28] in an attempt to increase its exploration capability. The hybrid approach is named the AOA-GD. The two suggested methods, AOA and AOA-GD are implemented in a wrapper model using a K-nearest neighbor (KNN) classifier and their performance is compared with those of other methods in the literature.

The rest of this paper is organized as follows: First, the AOA, the GDA and the proposed approaches for FS are discussed in Sections 2, 3 and 4, respectively. Then, the experiments and results are presented and discussed in Section 5. Finally, a conclusion is presented in Section 6.

2. THE ARITHMETIC OPTIMIZATION ALGORITHM

Arithmetics, along with geometry, algebra and analysis, are among the fundamental components of number theory and among the most significant aspects of modern mathematics. The conventional calculation methods that are used to analyze numbers are multiplication, division, subtraction and addition and are called arithmetic operators [29]. These basic operators have been utilized in mathematical optimization techniques to select the best element from a group of potential alternatives based on certain criteria (solutions) [30]. In problems with optimization, there are numerous quantitative domains, such as engineering, economics and computer science, as well as operations research and industry and the creation of improved solution methodologies has therefore long been of interest to mathematicians [31].

Regardless of the variations in the metaheuristic algorithms established for population-based optimization techniques, the optimization process is divided into two stages: exploration and exploitation. To prevent local solutions, the former includes the deployment of search agents that cover a large search field. The latter is the enhancement of the correctness of the obtained solutions during the exploration phase [32].The mathematical model is utilized to make a recommendation for the AOA [33].

2.1 Initialization Phase

The AOA optimization procedure begins with a set of randomly generated candidate solutions (X), as seen in the matrix below (Equation (1)) and the best candidate solution found in each iteration is deemed to be the best-obtained or nearly optimum solution so far.

Before starting its work, the AOA must decide on the search phase (i.e., exploration or exploitation). This is done by using the math optimizer accelerated (MOA) function, which is a coefficient that is calculated using Equation (2) and employed in the succeeding search phases:

$$X=\begin{cases} X1,1 & \dots & \dots & X1,J & X1,N-1 & X1,N \\ X2,1 & \dots & \dots & X2,J & X2,N-1 & X2,N \\ X3,1 & \dots & \dots & X3,J & X3,N-1 & X3,N \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ XN,1 & \dots & \dots & XN,J & XN,N-1 & XN,N \end{cases} \quad (1)$$

$$\text{MOA (C\_Iter)} = \text{Min} + \text{C\_Iter} \times \text{Max} - \left\{\frac{\text{Min}}{\text{M\_Iter}}\right\} \quad (2)$$

MOA (C Iter) is the function value at the 't'th iteration as calculated by Eq. (2). The current iteration is the number of iterations between 1 and the maximum number of iterations (M Iter) (C Iter). The accelerated function's minimum and maximum values are presented in Min and Max, respectively.

2.2 Exploration Phase

The exploration operators of the AOA use two basic search methods (the D search strategy and the M search strategy), to randomly examine different regions of the search space in order to discover a better solution (3). This phase of searching is conditioned for the condition of r1 > MOA by the MOA function, where r1 is a random number.

The first operator (D) is conditioned by r2 0.5 in this phase, while the other operator (M) is disregarded until the former operator completes its task. If D is unable to complete its task, instead of D, the duty is handed to the second operator (M). (It should be noted that r2 is a random integer.). To replicate the behavior of arithmetic operators, the simplest rule is employed [20].

For the exploration portions, two position updating equations are provided in this paper. The following is the initial position update equation:

$$\text{xi, j } (C_{\text{Iter}} + 1) = \begin{cases} \text{best(xj)} \div (\text{MOP} + \varepsilon) \times ((\text{UBj} - \text{LBj}) \times \text{M} + \text{LBj}), & \text{r2} < 0.5. \\ \text{best(xj)} \times \text{MOP} \times ((\text{UBj} - \text{LBj}) \times \text{M} + \text{LBj}), & \text{otherwise} \end{cases} \quad (3)$$

where xi (C_ Iter + 1) signifies the ith solution for next iteration, xi and j(C_Iter) signifies the jth position of the current iteration's ith solution, best(xj) and is the best solution achieved thus far. UBj and LBj denote the upper and lower limits of the jth position, respectively. In the experiments conducted for this study, M is a control parameter for altering the search process and is set at 0.5.The second position updating equation is as follows:

$$\text{MOP (C\_Iter)} = 1 - \left\{\frac{\text{C\_Iter}^{1/\alpha}}{\text{M\_Iter}^{1/\alpha}}\right\} \quad (4)$$

where MOP (C_Iter) represents the function value at the't' iteration, C_Iter indicates the current iteration, (M_Iter) denotes the maximum number of iterations and MOP(C_Iter) indicates the cost function at the 't'th iteration and MOP (C_Iter) signifies the function value at the 't'th iteration, which is set to 5 in the experiments conducted for the current work. Note that this is a crucial parameter that influences the accuracy of exploitation across repetitions.

2.3 Exploitation Phase

Subtraction (S) and addition (A) mathematical processes provide high-density outputs which represent the exploitation search process. However, unlike the other operators (D and M), these operators (S and A) may easily reach the objective owing to their low dispersion. As a result, the exploitation search might find a near-optimal solution that can be established after several attempts (iterations).

The MOA cost function for the constraint that r1 is not greater than the current MOA(C_ Iter) value (see Equation (4)) is conditioned for this phase of searching (exploitation search by executing S or A). As shown in Equation (5), the AOA exploitation operators (S and A) delve deeply into specified dense parts of the search space in order to find a better solution.

$$\text{xi, j (C\_Iter} + 1) = \begin{cases} \text{best(xj)} - \text{MOP} \times \left((\text{UBj} - \text{LBj}) \times \text{M} + \text{LBj}\right), & \text{r3} < 0. \\ \text{best(xj)} + \text{MOP} \times ((\text{UBj} - \text{LBj}) \times \text{M} + \text{LBj}), & \text{otherwise} \end{cases} \quad (5)$$

In this phase, which involves performing a deep search of the search space, the first operator (S) is conditioned by r30.5 (first rule in Equation (5)), while the other operator (A) is disregarded until the former operator (S) completes its task. If S is unable to complete its task, the second operator (A) is

utilized to accomplish the current task instead of S. The processes in this phase are identical to the partitioning in the previous phase. This method aids exploratory search strategies in finding the ideal answer while keeping a diversity of candidate solutions.

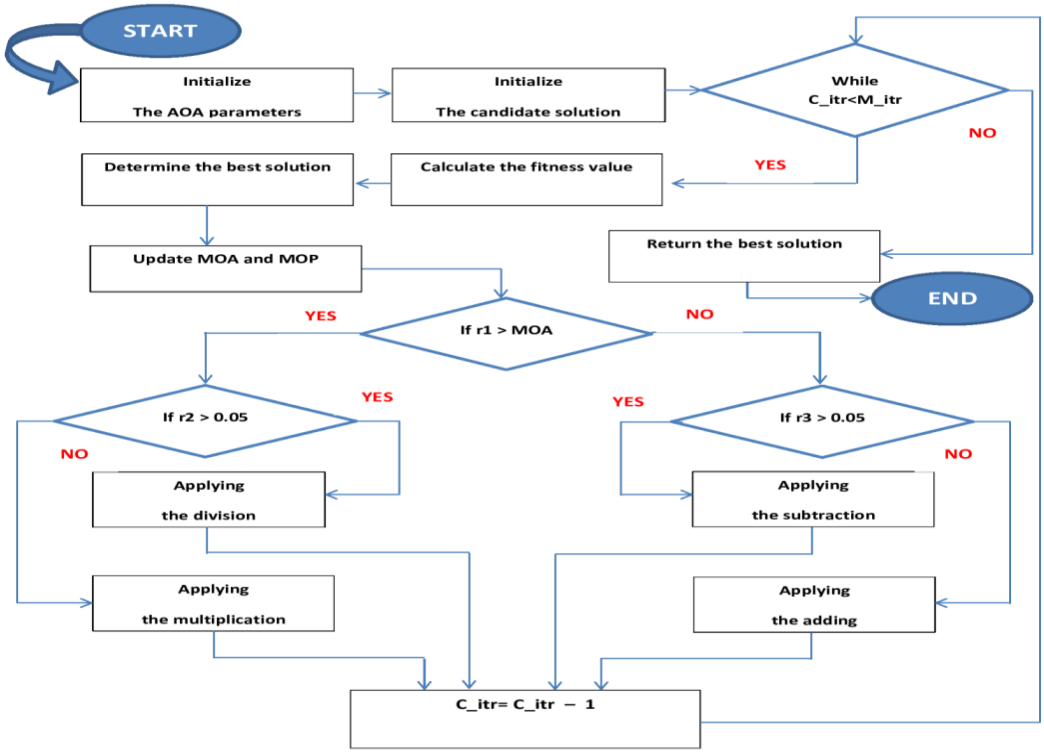

Figure 1. Flowchart of proposed AOA [20].

The M parameter is carefully chosen to create a random value at each iteration, which allows exploration to continue not only during the first but also during the last iteration. This element of the search process is highly effective when local optima stagnation occurs, particularly in the last iterations. The final location found can fall within a stochastic range specified by the search scope's locations of D, M, S and A. Various solutions update their locations stochastically about the near-optimal solution's region, whereas D, M, S and A estimate the position of the near-optimal solution in other ideas. All of these steps are clarified in the AOA flowchart illustrated in Figure 1.

All of these steps are clarified in Algorithm 1 below, which contains the pseudocode of the AOA:

Algorithm 1: The pseudo-code of the AOA.

1. Set the parameters for the AOA.
2. Randomize the positions of the solutions.
3. Determine the fitness value.
4. While C_itr<M_itr
5. Find the best solution.
6. Update the MOA value.
7. Update the MOP value.
8. Calculate the Fitness Function for the new solution.
9. for (i = 1 to Solution) do
10. if rand < 0.5 then
11. Create a set of random numbers between [0, 1].
12. if r1 > MOA then
13. if r2 > 0.5 then
14. Using the first rule in Equation (3), modify the positions of the i^{th} solutions.
15. Else
16. Using the second rule in Equation (3), modify the positions of the i^{th} solutions.
17. End if
18. Else

19. if r3 > 0.5 then
20. Using the first rule in Equation (5), modify the positions of the i^{th} solutions.
21. Else
22. Using the second rule in Equation (5), modify the positions of the i^{th} solutions.
23. Enf if
24. End if
25. End if
26. End for
27. C_Iter = C_Iter + 1
28. End while
29. Return the best solution.

3. GREAT DELUGE ALGORITHM

The great deluge algorithm (GDA) was invented by Dueck in 1993 [34]. The GDA operates similarly to simulated annealing (SA), with the exception that the GDA uses an upper limit (commonly referred to as the water level) as the acceptability barrier rather than a temperature. The GD technique starts with a boundary equal to the original solution's quality. If the cost (objective value) is much less than the boundary, which is reduced at a predetermined rate in each iteration, it accepts lesser options (known as the decay rate). The GDA has just one parameter (the decay rate), which gives it an advantage over the SA, since the effectiveness of a metaheuristic method is dependent on parameter tweaking [35].

Additionally, in comparison to SA, the GD method is less dependent on parameters. In actuality, the GD has only two parameters: the quality of the solution and the amount of time it takes to compute the solution [36]. The best solution is always accepted by the GDA. The worst solution, on the other hand, might be retained if its quality is less than or equal to a defined upper limit (water level), which is implemented to cope with minimization concerns. The water level is used as the initial value of the solution objective function and it is iteratively increased by a constant upper pound (UP) while the algorithm runs [37]. Algorithm 2 shows the pseudo-code for the typical GD algorithm.

Algorithm 2: The pseudo-code for the typical GD algorithm.

1: Pick a good initial setting.
2: Pick “rain speed” UP > 0.
3: Get an initial WATER-LEVEL > 0.
4: Use a new setting, which is a stochastic tiny setting.
5: Updates to the old setup
6: Measure E:= new quality setting
7: If E WATER-LEVEL is true, then
8: Then the previous setting is equal to the new setting.
9: WATER-LEVEL:= WATER-LEVEL+UP
10: If there has been no improvement in quality for a long period or if there have been too many iterations
11: Stop
12: End if
13: Output

4. PROPOSED METHOD

Two models are suggested in the current work. Both models use a wrapper FS method to select the most significant features inside a dataset. The first is based on the basic AOA, while the second involves combining the AOA with the GDA to achieve a balance between exploration and exploitation in the AOA.

In more detail, the current study proposes a progressive hybridization of the AOA and GD. The GD is incorporated into the AOA improvement process during the hybridization process. The hybridization process begins with a certain number of repetitions of the AOA. After the specified number of iterations, the GDA receives the best solution and highest fitness identified so far by the AOA and starts its improvement process. The obvious solution and fitness that the GDA discovers are then submitted back to the AOA to continue the process of development. This reciprocal procedure continues until all of the AOA iterations have been completed and the stopping condition has been fulfilled.

The proposed AOA-GD approach uses AOA to produce the initial (solution) population of possible solutions in the AOA stage. The GDA calculates the fitness value of all candidates in the second stage to identify better solutions, ensuring efficient convergence, high-quality solutions and finally obtaining the ideal parameter values and therefore improving classification accuracy. Figure 2 depicts the suggested AOA-GD with KNN solution to the FS issue.

The current study investigates the accuracy of two techniques, AOA and AOA-GD, for the FS process in the area of medical diagnostics utilizing a wrapper FS methodology based on KNN. The KNN classifier was selected, since previous research has shown that it has a good classification efficiency when used to FS issues. Throughout the investigation, the number of closest neighbors (K) was fixed at five. The 5-NN approach was used to assess fitness during the training phase using internal N-fold cross-validation, with a total of five folds and the average error rate in the classification methodology was computed for each fold. The number of folds (N) and the number of nearest neighbors were determined using previous work (K).

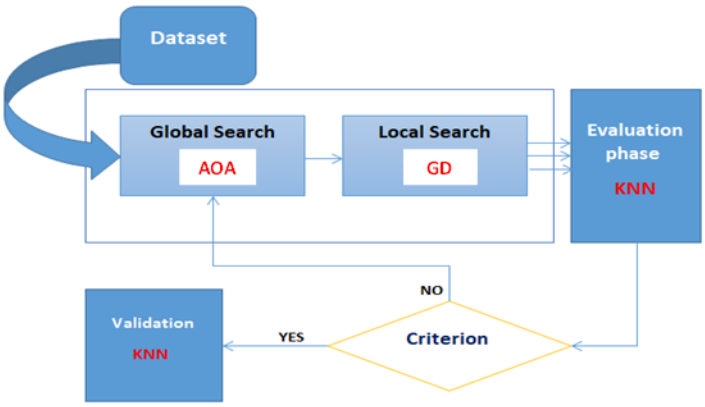

Figure 2. Proposed wrapper FS model based on AOA-GD method.

5. Experimental Results and Discussion

This part explains the experimental design and the data processing procedures used to assess the suggested methodology's performance. Additionally, it compares the findings to those obtained using previous methodologies. The resulting instability is impacted by a variety of parameters, including accuracy, convergence rate and particular measures of central tilt. To guarantee a fair scientific investigation, same work settings and circumstances were employed across the trials. The program and its execution are powered by an Intel® CoreTM i7-6006U processor running at 2.00 GHz (four CPUs) and 2.0 GHz, with 8 GB of RAM. Matlab R2016a was used to create the model. Each dataset was partitioned into 70% for training and 30% for testing. The tests were run 30 times for each dataset, with each run consisting of 100 iterations.

5.1 Parameter Settings

The findings of some preliminary tests were used to specify the input parameters in the experiments, allowing the recommended technique to provide improved results. In order for the results to be consistent, the algorithm parameters were kept constant throughout the trial. Table 1 shows the parameter values utilized in each experiment.

Table 1. Parameter setups.

Parameter	Value
σ	0.1
ɯ	0.5
P	30
Max-itr	100
LB (lower bound)	0
UB (upper bound)	1

5.2 Datasets' Description

Medical data is defined as any information about an individual's health that is used to make normal patient care choices or to conduct diagnostic trials. Examples include administrative data, claims data, patient illness data and clinical trial data. The trial results were assessed using a collection of 23 medical benchmark datasets. The datasets were obtained from a number of sources, including UCI, KEEL and Kaggle, in addition to other well-known websites that provide FS medical datasets. Table 2 summarizes the properties of these datasets.

Table 2. Description of the datasets.

The names of datasets	The no. of instances	The no. of classes	The no. of features	The source
1. Diagnostic	569	2	30	Source:UCI
2. Original	699	2	9	Source:UCI
3. Prognostic	194	2	33	Source:UCI
4. Coimbra	115	2	9	Source:UCI
5. BreastEW	596	2	30	Source:UCI
6. Retinopathy	1151	2	19	Source:UCI
7. Dermatology	366	6	34	Source:UCI
8. ILPD-Liver	583	2	10	Source:UCI
9. Lymphography	148	4	18	Source:UCI
10. Parkinsons	194	2	22	Source:UCI
11. ParkinsonC	755	2	753	Source:UCI
12. SPECT	267	2	22	Source:KEEL
13. Cleveland	297	5	13	Source:KEEL
14. HeartEW	270	2	13	Source:KEEL
15. Hepatitis	79	2	18	Source:KEEL
16. SAHear	461	2	9	Source:KEEL
17. Spectfheart	266	2	43	Source:KEEL
18. Thyroid0387	7200	3	21	Source:KEEL
19. Heart	302	5	13	Source:Kaggle
20. Pima-diabetes	768	2	9	Source:Kaggle
21. Leukemia dataset	72	2	7129	The source: https: //jundongl.github.io/scikit-feature/ datasets.html
22. Colon dataset	62	2	2000	The source: https: //jundongl.github.io/scikit-feature/ datasets.html
23. ProstateGE dataset	102	2	5966	The s ource :https: //jundongl.github.io/scikit-feature/ datasets.html

As shown in Table 2, the 23 datasets cover a range of case studies on medical diagnosis with varying architectures. Hence, the efficacy of the AOA and AOA-GD was determined by testing them on various problems with varying features. 70% of the datasets were utilized for training purposes and 30% for testing. Each dataset was tested 30 times, with each run consisting of 100 iterations.

5.3 Experimental Result

The recall, accuracy, precision, F-measure, error rate, number of features selected and convergence speed of the two proposed approaches, AOA and AOA-GD, were evaluated. Table 3 compares the accuracy rate and selection size of the two approaches after 30 runs on each dataset.

Table 3. AOA and AOA-GD accuracy and selection size results.

The dataset name		Accuracy		Selection Size	
		(AOA)	(AOA-GD)	(AOA)	(AOA-GD)
1	Diagnostic	.8470	.9097	15.381	12.9140
2	Original	.9300	.9762	7.0005	5.6201
3	Prognostic	.5888	.6501	17.9302	13.9271
4	Coimbra	.8105	.9100	4.0081	3.3810
5	BreastEW	.9017	.9531	14.9917	14.0027
6	Retinopathy	.5047	.6594	8.0590	6.7015
7	Dermatology	.7101	.8206	19.5103	17.0099
8	ILPD-Liver	.6273	.7619	4.9140	4.0000

9	Lymphography	.7593	.8511	11.4291	9.6910
10	Parkinsons	.6482	.7833	11.0610	10.7103
11	ParkinsonC	.6995	.8391	370.1920	360.0081
12	SPECT	.6109	.7108	10.0003	9.7201
13	Cleveland	.4792	.6163	8.1935	6.0003
14	HeartEW	.8741	.9207	6.0080	6.0619
15	Hepatitis	.6592	.7888	9.9104	9.0996
16	GDHear	.6208	.7259	5.0017	3.5039
17	Spectfheart	.6891	.7419	21.2710	19.6914
18	Thyroid0387	.9000	.9607	10.0041	7.0071
19	Heart	.7194	.8192	8.9914	7.1900
20	Pima-diabetes	.7005	.8201	5.8099	6.2814
21	Leukemia	.9819	.9914	3599.0092	3559.6091
22	Colon	.6430	.7228	1010.8192	989.2715
23	Prostate_GE	.4917	.6307	3038.0197	2961.0041

As seen in Table 3, the AOA-GD method bested the AOA in the whole datasets in terms of accuracy. This implies that if the search process of the AOA is changed, it is capable of producing more trustworthy findings.

As regards the number of features selected (FS size), the AOA-GD outperformed the AOA in 22 out of 24 datasets. Only in the Pima-diabetes and HeartEW datasets, the AOA able to outperform the AOA-GD. These findings illustrate the effectiveness of the AOA-GD modification in improving the exploratory search capacity of the AOA that enables it to discover the best basic solutions.

To further evaluate the findings and the classifier's capacity to provide trustworthy, correlated and similar solutions in all sequences for each dataset, the precision, recall and F-measure values were determined. Precision is the ratio of properly identified true positive IDs, recall is the ratio of correctly detected true positive IDs and the F-measure is the balance of the recall and precision ratios. Precision, recall and F-measure were computed as in [38]. Table 4 demonstrates how the AOA and AOA-GD methodologies' efficiency has been changed and focused over wholly datasets utilized in the experiment.

To compare the effectiveness of the AOA and AOA-GD approaches, a T-test was used. The presented

Table 4. Classification results for precision, recall and F-measure using AOA and AOA-GD.

Datasets		Precision Result		Recall Result		F- Measure Result	
		AOA	AOA-GD	AOA	AOA-GD	AOA	AOA-GD
1	-Diagnostic	.950	.970	.960	.970	.960	.960
2	-Original	.930	.950	.950	.980	.930	.940
3	-Prognostic	.930	.960	.950	.980	.930	.940
4	-Coimbra	.880	.900	.940	.000	.660	.670
5	-BreastEW	.750	.790	.780	.750	.750	.760
6	-Retinopathy	.800	.860	.700	.850	.830	.820
7	-Dermatology	.870	.960	.880	.960	.870	.830
8	-ILPD-Liver	.920	.960	.930	.980	.960	.920
9	-Lymphography	.810	.900	.790	.890	.790	.710
10	-Parkinsons	.890	1.000	.880	1.000	.860	.930
11	-ParkinsonC	.820	.950	.830	.940	.630	.700
12	-SPECT	.780	.890	.870	.970	.740	.730
13	-Cleveland	.810	.900	.790	.880	.800	.780
14	-HeartEW	.740	.790	.750	.790	.730	.710
15	-Hepatitis	.930	.980	.930	.970	.920	.890
16	-GDHear	.780	.770	.720	.760	.770	.770
17	-Spectfheart	.950	.970	.960	.970	.960	.960
18	-Thyroid0387	.930	.950	.950	.980	.930	.940
19	-Heart	.930	.960	.950	.980	.930	.940
20	-Pima-diabetes	.880	.900	.940	1.000	.660	.670
21	-Leukemia	.750	.790	.780	.750	.750	.760
22	-Colon	.800	.860	.700	.850	.830	.820
23	-Prostate_GE	.870	.960	.880	.960	.870	.830

methodologies are used to calculate the findings statistics depending on the accuracy of the findings point to each dataset. Table 5 demonstrates the results of T-test along with a 95% confidence interval (the alpha value = 0.05) as well as the p-values and classification accuracy produced by the AOA and AOA-GD.

Table 5. T-test results for AOA and AOA-GD.

Datasets		The Method Used	The Mean Value	The Std. Deviation Value	The Std. Error Mean Value	P-value Result
1-	-Diagnostic	AOA	0.8540	0.02673	0.00.488	00.00
		AOA-GD	0.9033	0.04046	0.00739	
2-	-Original	AOA	0.9233	0.02523	0.00461	00.00
		AOA-GD	0.9710	0.01062	0.00194	
3-	-Prognostic	AOA	0.5293	0.04638	0.00847	00.00
		AOA-GD	0.6716	0.04442	0.00811	
4-	-Coimbra	AOA	0.8006	0.04548	0.00830	00.00
		AOA-GD	0.8896	0.00999	0.00182	
5-	-BreastEW	AOA	0.8993	0.02100	0.00383	00.00
		AOA-GD	0.9400	0.01912	0.00349	
6-	-Retinopathy	AOA	0.4660	0.06106	0.01115	00.00
		AOA-GD	0.6436	0.02553	0.00466	
7-	-Dermatology	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.8006	0.04548	0.00830	
8-	-ILPD-Liver	AOA	0.6423	0.02609	0.00476	00.00
		AOA-GD	0.7716	0.01744	0.00318	
9-	-Lymphography	AOA	0.7606	0.03483	0.00636	00.00
		AOA-GD	0.8343	0.02661	0.00486	
10-	-Parkinsons	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.7903	0.01903	0.00347	
11-	-ParkinsonC	AOA	0.6856	0.06611	0.01207	00.00
		AOA-GD	0.8400	0.02197	0.00401	
12-	-SPECT	AOA	0.6073	0.02840	0.00518	00.00
		AOA-GD	0.6960	0.06667	0.01217	
13-	-Cleveland	AOA	0.4896	0.04173	0.00762	00.00
		AOA-GD	0.5966	0.02496	0.00456	
14-	-HeartEW	AOA	0.8540	0.02673	0.00488	00.00
		AOA-GD	0.9116	0.01783	0.00325	
15-	-Hepatitis	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.7903	0.01903	0.00347	00.00
16-	-SAHear	AOA	0.6420	0.03089	0.00564	00.00
		AOA-GD	0.7036	0.01066	0.00195	
17-	-Spectfheart	AOA	0.6716	0.04442	0.00811	00.00
		AOA-GD	0.7303	0.03178	0.00580	
18-	-Thyroid0387	AOA	0.8986	0.04455	0.00813	00.00
		AOA-GD	0.9603	0.01474	0.00269	
19-	-Heart	AOA	0.7316	0.04009	0.00732	00.00
		AOA-GD	0.8126	0.02612	0.00477	
20-	-Pima-diabetes	AOA	0.7153	0.05557	0.01015	00.00
		AOA-GD	0.7956	0.03757	0.00686	
21-	-Leukemia	AOA	0.9876	0.01736	0.00317	00.00
		AOA-GD	0.9900	0.01017	0.00186	
22-	-Colon	AOA	0.6203	0.03634	0.00663	00.00
		AOA-GD	0.7176	0.05380	0.00982	
23-	-Prostate_GE	AOA	0.4750	63290.0	11550.0	00.00
		AOA-GD	0.6010	25370.0	04630.0	
24-	-Covid-19 dataset	AOA	0.9135	0.02523	0.00461	00.00
		AOA-GD	0.9370	0.01912	0.00349	

Table 5 demonstrates that the AOA-GD is more efficient than the original AOA, where the p-values in all datasets are below 0.0001. This data indicates that the AOA-GD is useful for resolving FS issues. It is generally understood that a consistent and fast ratio of convergence leads to superior solutions. Therefore, to further evaluate the efficiency of the proposed AOA and AOA-GD methods, their

convergence speed behavior curves were studied in detail. Figure 3 shows the convergence speeds of the two proposed methods when employed to each of the 24 datasets over 30 runs.

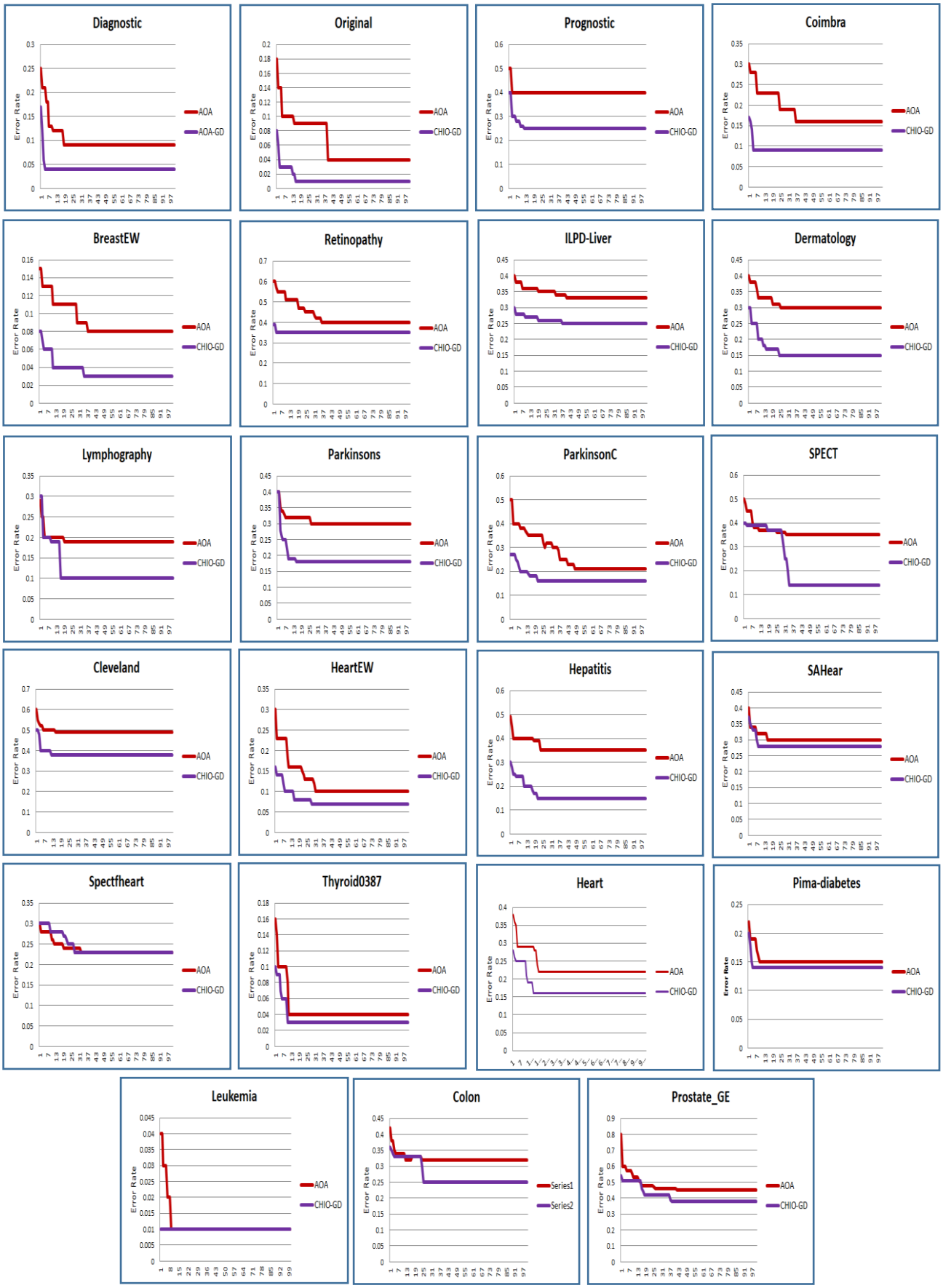

Figure 3. Convergence speeds of AOA and AOA-GD.

The findings in Figure 3 show that the AOA-GD was able to increase classification accuracy at a faster convergence time than the AOA. This was achieved by the GDA boosting the global search capacity of the original AOA.

5.4 Comparison with Previous Methods

The better of the two suggested approaches, AOA-GD, was compared to the CHIO-GC [39] (M1) and LBMFO-V3 (M2) [40] on the 23 medical datasets to examine the dependability of the proposed algorithm and its capacity to create a high degree of classification accuracy while decreasing the number of chosen characteristics. The classification accuracy and FS size of the AOA-GD were associated with those of the LBMFO-V3 and the CHIO-GC utilizing 23 medical datasets. The classification accuracy and FS size are calculated as average values over 30 runs. Table 6 displays the results.

Table 6. Comparison of the AOA-GD, CHIO-GC and LBMFO-V3 on the 23 medical benchmark datasets.

Datasets		Average of Accuracy			Selection Size		
		AOA-GD	M1	M2	AOA-GD	M1	M2
1	Diagnostic	0.9097	0.9033	0.9100	12.9140	13.3700	13.9991
2	Original	0.9762	0.9710	0.9683	5.6201	5.1040	5.5000
3	Prognostic	0.6501	0.6716	0.5851	13.9271	14.6202	15.0126
4	Coimbra	0.9100	0.8896	0.9312	3.3810	3.6007	3.5103
5	BreastEW	0.9531	0.9400	0.9398	14.0027	13.7303	13.9714
6	Retinopathy	0.6594	0.6436	0.5380	6.7015	.72647	6.9002
7	Dermatology	0.8206	0.8006	0.8442	17.0099	18.4900	18.3541
8	ILPD-Liver	0.7619	0.7716	0.7143	4.0000	4.0000	4.0000
9	Lymphography	0.8511	0.8343	0.8002	9.6910	.100622	9.7520
10	Parkinsons	0.7833	0.7903	0.7689	10.7103	9.7383	10.3584
11	ParkinsonC	0.8391	0.8400	0.8190	360.0081	365.8322	369.1070
12	SPECT	0.7108	0.6960	0.6576	9.7201	9.6050	10.7832
13	Cleveland	0.6163	0.5966	0.5333	6.0003	6.8097	6.6899
14	HeartEW	0.9207	0.9116	0.9388	6.0619	7.0105	6.3100
15	Hepatitis	0.7888	0.7903	0.7500	9.0996	8.2011	8.3569
16	SAHear	0.7259	0.7036	0.6992	3.5039	.31551	3.2222
17	Spectfheart	0.7419	0.7303	0.7013	19.6914	21.0030	20.4598
18	Thyroid0387	0.9607	0.9603	0.9776	7.0071	8.0116	8.4563
19	Heart	0.8192	0.8126	0.7603	7.1900	6.1505	6.2752
20	Pima-diabetes	0.8201	0.7956	0.8065	6.2814	.68387	6.7612
21	Leukemia	0.9914	0.9900	1.0000	3559.6091	3560.5107	3570.7137
22	Colon	0.7228	0.7176	0.6667	989.2715	1000.0067	991.5551
23	Prostate_GE	0.6307	0.6010	0.5056	2961.0041	2979.4116	2984.7153
Average		0.8071	0.7983	0.7746	349.6698	351.4142	351.9462

In 18 datasets, including Diagnostic original, Coimbra, BreastEW, Retinopathy, Dermatology, Lymphography, SPECT, Cleveland, HeartEW, SAHear, Spectfheart, Thyroid0387, Heart, Pima-diabetes, Leukemia, Colon and Prostate GE, the AOA-GD method outperformed the CHIO-GC method with regards to classification accuracy. Additionally, the AOA-GD approach outperformed the LBMFO-V3 technique in 17 datasets, including the Original, Prognost, BreastEW, Retinopathy, Lymphography SPECT, Parkinsons, ILPD-Liver, Colon, Cleveland, Hepatitis, SAHear, Spectfheart,,Heart, Parkinson C, Pima-diabetes and Prostate GE. The AOA-GD strategy outperformed the CHIO-GC and LBMFO-V3 approaches across each dataset, with on average accuracy of (0.8071).

In 16 datasets, the AOA-GD technique outperformed the CHIO-GC method in terms of selection size, including Diagnostic, Prognostic, Coimbra, Retinopathy, Dermatology, ILPD-Liver, Lymphography, ParkinsonC, Cleveland, HeartEW, Spectfheart, Thyroid0387, Pima-diabetes, Leukemia, Colon and Prostate GE. Additionally, the AOA-GD technique outperformed the LBMFO-V3 method in 17 datasets, including Diagnostic, Prognostic, Coimbra, Retinopathy, Dermatology, ILPD-Liver, Lymphography, ParkinsonC, SPECT, Cleveland, HeartEW, Spectfheart, Thyroid0387, Pima-diabetes,

Leukemia, Colon and Prostate GE. Across all datasets, the AOA-GD had an average selection size of 349.6698 features. Figure 4 graphically illustrates, respectively, the average accuracy and average selection size of the three methods.

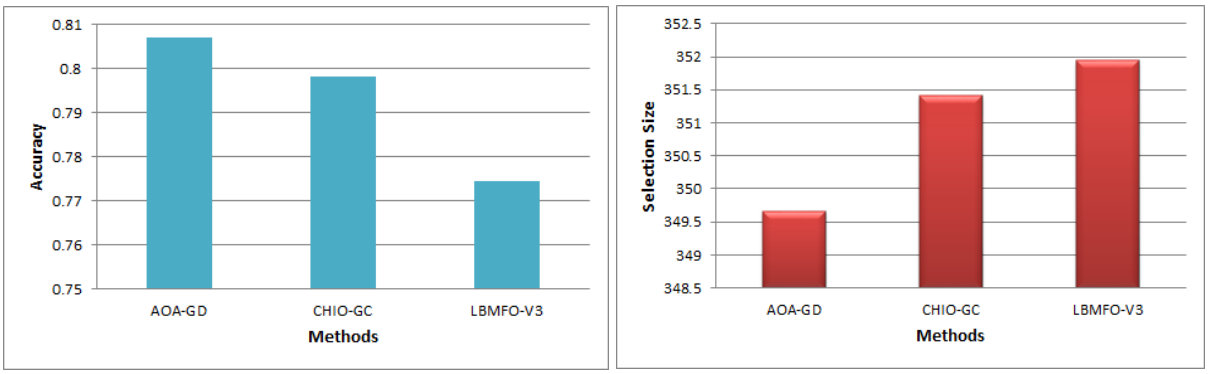

Figure 4. Average accuracy and average selection size of all three methods on the 23 medical benchmark datasets.

5.5 Discussion

The initial results for the AOA demonstrated that the search mechanism of the algorithm finds a reasonable balance between exploration and exploitation [41]. The capacity of metaheuristic algorithms to uncover optimum solutions throughout the search process is one of its most important characteristics. According to the findings, the suggested hybridization of the GDA with the AOA improves the AOA's exploration capabilities, allowing it to better choose the initial characteristics necessary to fulfill the aim of optimizing the solution. In other words, the suggested change aided in improving the balance between exploration and exploitation.

In all 23 datasets, Table 4 demonstrates that the AOA-GD approach surpassed the AOA approach in terms of classification accuracy, highlighting the utility of the proposed hybrid AOA-GD technique for striking an acceptable balance between exploration and exploitation. Additionally, as shown in Table 4, Figure 6 and Figure 7, the AOA-GD excelled in terms of the highest and lowest accuracy achievable throughout each run. Finally, the AOA-GD demonstrated to be beneficial in narrowing the accuracy gap between maximum and minimum values and accelerating convergence.

6. CONCLUSION

The FS problem is one of the most pressing concerns for academics across a broad range of disciplines; metaheuristics has been widely used in recent years in feature services (FS) to reduce the number of features mandatory to obtain satisfactorily honest results, with the objective of growing to have reliability and performance. The AOA metaheuristics was used in the current work for two proposed models for addressing FS issues in medical diagnosis. The first model was based on the basic AOA, while the second entailed integrating the AOA with the GDA in order to obtain a better balance between exploration and exploitation in the AOA. 23 medical datasets were used to evaluate the suggested approaches. Many measures were used to compare the two strategies, including number of selected features, classification accuracy, recall, precision, convergence speed, F-measure and T-test. The findings of all data analyses indicated that the AOA-GD enhanced the exploratory capabilities of the original AOA. Moreover, the feature selection size and classification accuracy of the AOA-GD were compared to those of other approaches previously published. The investigation's findings indicated that the AOA-GD technique outperformed the CHIO-GC and LBMFO-V3 wrapper approaches. AOA-GD surpassed the CHIO-GC and LBMFO-V3 in the majority of medical benchmark datasets with an accuracy rate of 0.80 and a selection size rate of 349 features, according to the results of the study. These encouraging findings were obtained as a consequence of a well-balanced AOA-GD search phase during the identification of appropriate solutions, which increased the pace of convergence. This optimum balance was reached by combining the AOA with the GDA, since the GDA was able to correct the unacceptable solutions that were obtained during premature convergence and while confined in a narrow optimal search space, respectively.

ACKNOWLEDGEMENTS

The research reported in this publication was supported by the Deanship of Scientific Research and Innovation at Al-Balqa Applied University in Jordan. Grant Number: DSR-2021#369.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

[1] M. Alweshah, S. A. Khalaileh, B. B. Gupta, A. Almomani, A. I. Hammouri and M. A. Al-Betar, "The Monarch Butterfly Optimization Algorithm for Solving Feature Selection Problems," Neural Computing and Applications, vol. 2020, pp. 1-15, 2020.

[2] C. Lam, A. Siefkas, N. S. Zelin et al., "Machine Learning As a Precision-medicine Approach to Prescribing COVID-19 Pharmacotherapy with Remdesivir or Corticosteroids," Clinical Therapeutics, vol. 43, no. 5, pp. 871-885, 2021.

[3] M. Alweshah, M. Al-Sendah, O. M. Dorgham, A. Al-Momani and S. Tedmori, "Improved Water Cycle Algorithm with Probabilistic Neural Network to Solve Classification Problems," Cluster Computing, vol. 23, pp. 2703-2718, 2020.

[4] S. Uddin, A. Khan, M. E. Hossain and M. A. Moni, "Comparing Different Supervised Machine Learning Algorithms for Disease Prediction," BMC Medical Informatics and Decision Making, vol. 19, pp. 1-16, 2019.

[5] M. Alweshah, W. A. AlZoubi and A. Alarabeyyat, "Cluster Based Data Reduction Method for Transaction Datasets," Proc. of IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 78-83, Langkawi, Malaysia, 2015.

[6] K. Zukotynski, V. Gaudet, C. F. Uribe et al., "Machine Learning in Nuclear Medicine: Part 2—Neural Networks and Clinical Aspects," Journal of Nuclear Medicine, vol. 62, pp. 22-29, 2021.

[7] N. Wang, Y. Huang, H. Liu, Z. Zhang, L. Wei, X. Fei and H. Chen, "Study on the Semi-supervised Learning-based Patient Similarity from Heterogeneous Electronic Medical Records," BMC Medical Informatics and Decision Making, vol. 21, pp. 1-13, 2021.

[8] E. H. Houssein, I. E. Ibrahim, N. Neggaz, M. Hassaballah and Y. M. Wazery, "An Efficient ECG Arrhythmia Classification Method Based on Manta Ray Foraging Optimization," Expert Systems with Applications, vol. 181, p. 115131, 2021.

[9] E. H. Houssein, M. E. Hosney and D. Oliva, "A Hybrid Seagull Optimization Algorithm for Chemical Descriptors Classification," Proc. of the International Mobile, Intelligent and Ubiquitous Computing Conference (MIUCC), pp. 1-6, Cairo, Egypt, 2021.

[10] A. Almomani, M. Alweshah, S. Al Khalayleh, M. Al-Refai and R. Qashi, "Metaheuristic Algorithms-based Feature Selection Approach for Intrusion Detection," Chapter in Book: Machine Learning for Computer and Cyber Security, 1st Edn., CRC Press, pp. 184-208, 2019.

[11] H. Al Nsour, M. Alweshah, A. I. Hammouri, H. Al Ofeishat and S. Mirjalili, "A Hybrid Grey Wolf Optimiser Algorithm for Solving Time Series Classification Problems," Journal of Intelligent Systems, vol. 29, no. 1, pp. 846-857, 2018.

[12] M. Alweshah, L. Rababa, M. H. Ryalat, A. Al Momani and M. F. Ababneh, "African Buffalo Algorithm: Training the Probabilistic Neural Network to Solve Classification Problems," Journal of King Saud University-Computer and Information Sciences, DOI: 10.1016/j.jksuci.2020.07.004, 2020.

[13] O. Dorgham, M. Alweshah, M. Ryalat, J. Alshaer, M. Khader and S. Alkhalaileh, "Monarch Butterfly Optimization Algorithm for Computed Tomography Image Segmentation," Multimedia Tools and Applications, pp. 1-34, DOI: 10.1007/s11042-020-10147-6, 2021.

[14] M. Alweshah, "Solving Feature Selection Problems by Combining Mutation and Crossover Operations with the Monarch Butterfly Optimization Algorithm," Applied Intelligence, vol. 51, no. 1, pp. 1-24, 2020.

[15] M. Alweshah, H. Rashaideh, A. I. Hammouri, H. Tayyeb and M. Ababneh, "Solving Time Series Classification Problems Using Support Vector Machine and Neural Network," International Journal of Data Analysis Techniques and Strategies, vol. 9, pp. 237-247, 2017.

[16] B. Sathiyabhama et al., "A Novel Feature Selection Framework Based on Grey Wolf Optimizer for Mammogram Image Analysis," Neural Computing and Applications, vol. 33, pp. 14583–14602, 2021.

[17] M. Alweshah, E. Ramadan, M. H. Ryalat, M. Almi'ani and A. I. Hammouri, "Water Evaporation Algorithm with Probabilistic Neural Network for Solving Classification Problems," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 1, pp. 1-15, 2020.

[18] M. Alweshah, O. A. Alzubi, J. A. Alzubi and S. Alaqeel, "Solving Attribute Reduction Problem Using Wrapper Genetic Programming," International Journal of Computer Science and Network Security (IJCSNS), vol. 16, p. 77, 2016.

[19] M. Alweshah, S. Alkhalaileh, D. Albashish, M. Mafarja, Q. Bsoul and O. Dorgham, "A Hybrid Mine Blast Algorithm for Feature Selection Problems," Soft Computing, vol. 25, pp. 517-534, 2021.

[20] H. Sun, J. Jin, R. Xu and A. Cichocki, "Feature Selection Combining Filter and Wrapper Methods for Motor-Imagery Based Brain–Computer Interfaces," International Journal of Neural Systems, vol. 31, no. 9, p. 2150040, 2021.

[21] N. Darabi, A. Rezai and S. S. F. Hamidpour, "Breast Cancer Detection Using RSFS-based Feature Selection Algorithms in Thermal Images," Biomedical Engineering: Applications, Basis and Communications, vol. 33, no. 3, p. 2150020, 2021.

[22] M. Salimian, A. Rezai, S. Hamidpour and F. Khajeh-Khalili, "Effective Features in Thermal Images for Breast Cancer Detection," Proc. of the 2nd National Conf. on New Technologies in Electrical and Computer Engineering, Isfahan, Iran, pp. 1-7, 2019.

[23] M. Zarei, A. Rezai and S. S. Falahieh Hamidpour, "Breast Cancer Segmentation Based on Modified Gaussian Mean Shift Algorithm for Infrared Thermal Images," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 9, no. 6, pp. 1-7, 2021.

[24] H. Y. Chong, H. J. Yap, S. C. Tan et al., "Advances of Metaheuristic Algorithms in Training Neural Networks for Industrial Applications," Soft Computing, vol. 25, pp. 11209–11233, 2021.

[25] Y. Xue, Y. Tang, X. Xu, J. Liang and F. Neri, "Multi-objective Feature Selection with Missing Data in Classification," IEEE Transactions on Emerging Topics in Computational Intelligence, DOI: 10.1109/TETCI.2021.3074147, pp. 1-10, 2021.

[26] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk and W. Al-Atabany, "Honey Badger Algorithm: New Metaheuristic Algorithm for Solving Optimization Problems," Mathematics and Computers in Simulation, vol. 192, pp. 84-110, 2022.

[27] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz and A. H. Gandomi, "The Arithmetic Optimization Algorithm," Computer Methods in Applied Mechanics and Engineering, vol. 376, p. 113609, 2021.

[28] D. Simon, "Biogeography-based Optimization," IEEE Transactions on Evolutionary Computation, vol. 12, pp. 702-713, 2008.

[29] S. Khatir, S. Tiachacht, C. Le Thanh, E. Ghandourah, S. Mirjalili and M. A. Wahab, "An Improved Artificial Neural Network Using Arithmetic Optimization Algorithm for Damage Assessment in FGM Composite Plates," Composite Structures, vol. 273, p. 114287, 2021.

[30] L. Abualigah, A. Diabat, P. Sumari and A. H. Gandomi, "A Novel Evolutionary Arithmetic Optimization Algorithm for Multilevel Thresholding Segmentation of Covid-19 ct Images," Processes, vol. 9, p. 1155, 2021.

[31] M. Premkumar, P. Jangir, B. S. Kumar, R. Sowmya, H. H. Alhelou, L. Abualigah, A. R. Yildiz and S. Mirjalili, "A New Arithmetic Optimization Algorithm for Solving Real-world Multiobjective CEC-2021 Constrained Optimization Problems: Diversity Analysis and Validations," IEEE Access, vol. 9, pp. 84263-84295, 2021.

[32] Y. Wu, "A Survey on Population-based Meta-heuristic Algorithms for Motion Planning of Aircraft," Swarm and Evolutionary Computation, vol. 62, p. 100844, 2021.

[33] D. Gürses, S. Bureerat, S. M. Sait and A. R. Yıldız, "Comparison of the Arithmetic Optimization Algorithm, the Slime Mold Optimization Algorithm, the Marine Predators Algorithm, the Salp Swarm Algorithm for Real-world Engineering Applications," Materials Testing, vol. 63, pp. 448-452, 2021.

[34] G. Dueck, "New Optimization Heuristics: The Great Deluge Algorithm and the Record-to-record Travel," Journal of Computational Physics, vol. 104, pp. 86-92, 1993.

[35] D. Landa-Silva and J. H. Obit, "Great Deluge with Non-linear Decay Rate for Solving Course Timetabling Problems," Proc. of the 4th International IEEE Conference on Intelligent Systems, pp. 8-11-8-18, Varna, Bulgaria, 2008.

[36] S. Kifah and S. Abdullah, "An Adaptive Non-linear Great Deluge Algorithm for the Patient-admission Problem," Information Sciences, vol. 295, pp. 573-585, 2015.

[37] M. Mohmad Kahar and G. Kendall, "A Great Deluge Algorithm for a Real-world Examination Timetabling Problem," Journal of the Operational Research Society, vol. 66, pp. 116-133, 2015.

[38] S. Kassaymeh, S. Abdullah, M. A. Al-Betar and M. Alweshah, "Salp Swarm Optimizer for Modeling the Software Fault Prediction Problem," Journal of King Saud University-Computer and Information Sciences, In Press, DOI: 10.1016/j.jksuci.2021.01.015, 2021.

[39] M. Alweshah, S. Alkhalaileh, M. A. Al-Betar and A. A. Bakar, "Coronavirus Herd Immunity Optimizer with Greedy Crossover for Feature Selection in Medical Diagnosis," Knowledge-based Systems, vol. 235, p. 107629, DOI: 10.1016/j.knosys.2021.107629, 2021.

[40] R. Abu Khurmaa, I. Aljarah and A. Sharieh, "An Intelligent Feature Selection Approach Based on Moth Flame Optimization for Medical Diagnosis," Neural Computing and Applications, vol. 33, pp. 7165-

7204, 2021.

[41] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk and W. Al-Atabany, "Archimedes Optimization Algorithm: A New Metaheuristic Algorithm for Solving Optimization Problems," Applied Intelligence, vol. 51, pp. 1531-1551, 2021.

ملخص البحث:

في حقْل الطّبّ، ثمّة حاجةٌ الى تصفية البيانات لإيجاد المعلومات ذات العلاقة بمسائل بحثية معينة. ومع ذلك، في إطار الدّراسة العلمية، فإنّ عملية انتقاء البيانات أو السِّمات الملائمة تبقى مسألةً أساسيةً تنطوي على الكثير من التّحديات.

لذا، نقدّم في هذه الورقة طريقتين من طُرق انتقاء السِّمات (FS) بناءً على خوارزميات عالية التّوجيه، وبالذّات خوارزمية الأمثَلة الحسابية (AOA) وخوارزمية الغَمْر العظيم (GDA)، في محاولةٍ لمواجهة تحدّي التّشخيص. وقد تمّ اختبار نموذجيْن: (AOA) و (AOA-GD) على 23 من مجموعات البيانات المرجعية الطبّية. وبناءً على جميع البيانات التّجريبية، فإنّ تهجين الأمثَلة الحسابية بخوارزمية الغَمْر العظيم زاد على نحوٍ ملموس من قُدرة الأمثَلة الحسابية على البحث. بعدئذٍ، تمّت مقارنة النّظام الهجين المقترح باثنتيْن من طُرق انتقاء السِّمات المستخدَمة سابقاً (CHIO-GC) و (LBMFO_V3). وعند تطبيق الطّريقة المقترحة في هذا البحث والطّريقتين السّابقتين المشار إليهما آنفاً على مجموعات البيانات الثلاث والعشرين، حقّقت الطّريقة المقترحة معدّل دقّةٍ بلغ 0.80، وتجاوزت بذلك الطّريقتين السابقتين.