

# EEG Channel Selection Using A Modified Grey Wolf Optimizer

Hussien Rezk Hussien, El-Sayed M. El-Kenawy, Ali I. El-Desouky

**Abstract** — Consider an increasingly growing field of research, Brain-Computer Interface (BCI) is to form a direct channel of communication between a computer and the brain. However, extracting features of random time-varying EEG signals and their classification is a major challenge that faces current BCI. This paper proposes a modified grey wolf optimizer (MGWO) that can select optimal EEG channels to be used in (BCIs), the way that identifies main features and the immaterial ones from that dataset and the complexity to be removed. This allows (MGWO) to opt for optimal EEG channels as well as helping machine learning classification in its tasks when doing training to the classifier with the dataset. (MGWO), which imitates the grey wolves leadership and hunting manner nature and which consider metaheuristics swarm intelligence algorithms, is an integration with two modification to achieve the balance between exploration and exploitation the first modification applies exponential change for the number of iterations to increase search space accordingly exploitation, the second modification is the crossover operation that is used to increase the diversity of the population and enhance exploitation capability. Experimental results use four different EEG datasets BCI Competition IV-dataset 2a, BCI Competition IV- data set III, BCI Competition II data set III, and EEG Eye State from UCI Machine Learning Repository to evaluate the quality and effectiveness of the (MGWO). A cross-validation method is used to measure the stability of the (MGWO).

**Index Terms** — Brain-Computer interface, EEG signals, Channel selection, Feature selection, Grey wolf optimizer, Metaheuristics.

## I. INTRODUCTION

The human brain which consists of the brainstem, the cerebrum, and the cerebellum controls most of the activities of the body, processing, integrating, and coordinating the information it receives from the sense organs and making decisions as to the instructions sent to the rest of the body[1]. The brain is one of the most intricate and marvelous elements in our body, it is the command center for the human nervous system. To understand how the brain works and how the transmission of electrical signals to the brain through neurons and study how it works, one of the most important pieces of research in this area is BCI [2].

Nowadays Brain-Computer Interface (BCI) represents a highly growing field of research, it aims to make a direct communication link between the brain and a computer [3], BCI interprets brain signals of human interaction into the

desired output. The required output is used to control external devices like a speller, a wheelchair, games, a robotic arm, and entertainment, and so on [4]. This application of signal processing and neuroscience, Brain-computer interface (BCI) is greatly challenging and also called brain-machine interface (BMI).

EEG method is likewise utilized in extracting capabilities of the mind sign even though the problem isn't always in a country to take care of the stimuli. The human neurons communicate with every different through electric impulses. The electrodes are positioned at the scalp to degree the amplitude of electrical impulses. While recording a brain signal, numerous artifacts include the signal. Different sorts of artifacts affecting the signal are open and last of eyes all through the signal acquisition process, muscular sports, and sports occurring withinside the background [6]. Therefore, EEG alerts ought to be recorded in de-noised labs the use of machines which might be loose from interferences, artifacts, or every other style of noise [7].

The nature of EEG alerts could be very complex because they're now no longer correlated however random. EEG dimension is based on loads of things especially called the individual's age, gender, Psychological country, and intellectual country of the problem [8]. Hence, the know-how of the conduct and motion of mind cells entails many linear and nonlinear signal-processing strategies which bring about the final results associated with the physiological country and occasions of the problem. Several strategies advised extracting the dynamic capabilities and abrupt adjustments that could occur. The first element is preprocessing which incorporates the recording of alerts, elimination of artifacts, signal averaging, thresholding of the output, and improving the ensuing signal. The 2nd step withinside the operation is the function extraction scheme to decide a feature vector from an everyday vector [9].

One of the maximum demanding situations in modern BCI studies is extracting capabilities of random time-various EEG alerts and classifying the alerts as correctly as possible [10]. So we ought to communicate approximately Feature selection. Feature choice offers a manner to become aware of the vital capabilities and dispose of immaterial ones from the dataset [11]. The function choice goals are decreasing information dimensions, enhancing prediction performance, and know-how information nicely for the exceptional device getting to know applications [12]. In actual applications, information illustration frequently makes use of many capabilities with redundancy, this means that sure capabilities are unimportant and ought to be removed. Moreover, the interdependence capabilities affect the output and incorporate vital statistics with a purpose to be unknown if any of them is removed [13]. A function represents a unique property, EEG alerts lets in the direct evaluation of the country of a human, which is taken into consideration as a critical parameter in Brain-Computer-Interface. Several strategies for function extraction had been studied and the

---

Published on January 12, 2021.

Hussien Rezk Hussien, Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Egypt. (corresponding e-mail: hussienrezk@students.mans.edu.eg)

El-Sayed M.El-Kenawy Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology (DHIET), Egypt. (e-mail: skenawy@ieeee.org)

Ali I. El-Desouky, Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Egypt. (e-mail: prof.desoky@gmail.com).

choice of each suitable capabilities and vicinity of the electrode. For the EEG alerts Feature extraction scheme is supposed to pick out the capabilities or statistics this is the maximum vital for Machine getting to know category exercising while doing schooling to the classifier with dataset [14].

Feature selection strategies may be classified as following filter, wrapper, and hybrid-primarily based totally [15]. The filter-primarily based totally characteristic choice strategies or conventional characteristic choice strategies have a bonus that it's miles pace and cap potential to scale to huge datasets. The system of characteristic choice is regularly maximum beneficial in conditions wherein wrappers may also overfit [16]. which includes Information Gain (IG). IG measuring how a good deal "information" a characteristic offers us approximately the magnificence and it's miles beneficial in lowering the variety of capabilities that deliver us greater accuracy withinside the class model [17]. The Wrapper strategies of characteristic choice lessen seek area for choosing capabilities. The wrappers consist of the mastering set of rules as part of their choose function. Wrappers offer greater accuracy however it takes a good deal greater time (extra slowly). Such as Genetic algorithms (GA), (GAs) are randomly primarily based algorithms at the system of herbal choice underlying organic evolution. They may be implemented to many challenges, optimization, system mastering issues, and characteristic choice [18]. To do wrapper Feature selection we want to make use of an optimization set of rules, the classical optimization strategies are by some means constrained in fixing the issues so that evolutionary computation (EC) algorithms are the opportunity for fixing those obstacles and trying to find the greatest answer of the issues. Evolutionary computation (EC) algorithms are stimulated via way of means of nature, social behavior, and organic behavior of (animals, birds, fish, a bat, Firefly, wolves, etc.) [19].

The rest of the paper is organized as follows: Section II presents related work for optimization algorithms, EEG, and the background of grey wolf optimization (GWO). The proposed new version of grey wolf optimization (MGWO) describes in Section III. Section IV Modified version of grey wolf optimization for feature selection. The experimental results are argued in Section V. Finally, conclusions are stated in Section VI.

## II. RELATED WORK

The optimization method is current in numerous studies regions including engineering, medicine, agriculture, computer science, feature selection. In optimization, the principle goal is to choose the finest solution to given trouble from the to-be had answered in regards to the trouble description. Moreover, in optimization algorithms, there's a goal that has to be minimized or maximized in keeping with the trouble to be solved [20].

Recently, many sorts of studies used optimization to resolve given trouble like Whale Optimization Algorithm (WOA) [21], WOA turned into used to locate the ultimate weights for schooling the neural community and advanced a multi-goal model of WOA is advanced and implemented it to the hassle of forecasting the wind speed. Moreover, WOA turned into extensively utilized to decide the ultimate

placement and length of capacitors used withinside the radial system [22]. Furthermore, they applied WOA withinside the hassle of locating the ultimate length utilized by a disbursed generator [23], they take the gain of the usage of WOA for MRI picture segmentation [24]. Furthermore, some other set of rules is Grey Wolf Optimizer (GWO), GWO is a brand new optimization set of rules which simulates the gray wolves' management and looking way in nature [25]. GWO is characterized through simplicity, flexibility, deprivation-unfastened mechanism, and neighborhood optima avoidance Because of that, it's been used in lots of studies regions withinside the closing years which include function subset selection [26], DC automobiles control [27], financial emission dispatch problems [28], picture registration [29], Radial Basis Function (RBF) networks schooling and fixing ultimate reactive electricity dispatch hassle, additionally use the GWO set of rules to educate the MLP community.

### A. EEG Signal

BCI has various applications from controlling video games to artificial limbs and robotic arms. These applications can be categorized due to the field such as environmental control, locomotion, entertainment, and multimedia. There are many searching algorithms used to determine the optimal feature from channels of EEG signals, like the branch and bound search (BB), sequential forward selection (SFS), Tabu search (TB), Simulated annealing algorithm (SA), Genetic Algorithm (GA), Binary Particle Swarm Optimization (BPSO). One of these algorithms is the Improved Binary Gravitation Search Algorithm (IBGSA) used to find the best optimal channels and decrease the number of EEG signals, IBGSA [30] is a heuristic optimization algorithm that was used. GSA is an optimization algorithm that was designed depending on Newtonian laws of gravity and motion [31]. This algorithm was utilized in solving several optimization problems such as feature selection [32]. To deal with binary problems, IBGSA which is the binary version of GSA could be used.

### B. Traditional Gray Wolf

The Grey Wolf Optimizer (GWO) is a swarm-primarily based meta-heuristic this is stimulated through the social intelligence, leadership, and looking conduct of gray wolf packs. In general, the gray wolves stay in agencies of 5-12 individuals. As proven in Fig. 1, the gray wolf percent has a not unusual place social hierarchy that includes 4 degrees. The maximum effectiveness and the chief of the percent are referred to as Alpha this is answerable for handling the percent, taking selections in looking and migration. The 2nd stage is beta this is answerable for assisting alphas selections and handling the percent if alpha is unwell or dead. The different degrees are delta and omega that comply with the coaching of better degrees and feature much less dominance as proven withinside the figure.

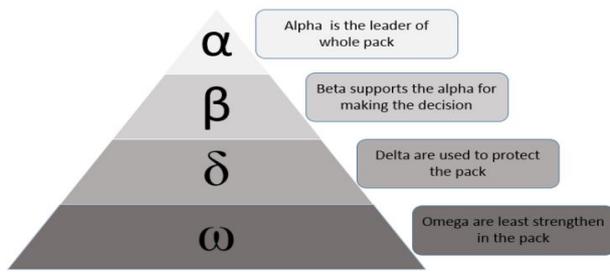


Fig. 1. Hierarchy of Gray Wolves.

As a population-based meta-heuristic, GWO starts the optimization process with an initial random population. Each individual in the population represents a candidate solution to the problem being solved. Throughout iterations, GWO ensures a proper balance between exploration and exploitation. In each iteration, the fitness is calculated for each individual, then the Alpha, Beta, and Delta are determined based on their fitness. The position of each individual is updated based on the position of the three main leaders (Alpha, Beta, and Delta). The position is updated based on the following equations:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}$$
(1)

where  $t$  refers to the current iteration,  $\vec{A}$  and  $\vec{C}$  are coefficient vectors,  $\vec{X}_p$  is the preposition, and  $\vec{X}$  is the position of the gray wolf. The vectors are calculated using the following equation:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$$

$$\vec{C} = 2 \cdot \vec{r}_2$$
(2)

where  $a$  is linearly decreased from 2 to 0 throughout iterations and  $r_1$  and  $r_2$  are random numbers in the range [0, 1]. The whole pack reaches the prey and attack by updating the position based on the best locations of the alpha, beta, and delta using the following equations:

$$\vec{D}_\alpha = |\vec{C}_1 * \vec{X}_\alpha - \vec{X}|$$

$$\vec{D}_\beta = |\vec{C}_2 * \vec{X}_\beta - \vec{X}|$$
(3)

$$\vec{D}_\delta = |\vec{C}_3 * \vec{X}_\delta - \vec{X}|$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 * \vec{D}_\alpha$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 * \vec{D}_\beta$$
(4)

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 * \vec{D}_\delta$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}$$
(5)

The pseudo-code of the classical GWO algorithm is described as following in Fig 2:

```

Input Xi (i = 1, 2, ..., n)
Initialize the grey wolf population Xi (i = 1, 2, ..., n)
while (t < Max number of iterations)
    Calculate the fitness of each search agent
    Calculate alpha, beta, and delta
    Calculate A as per Equation  $\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$ 
    for each search agent
        Update A, and C
        Calculate X $\alpha$ , X $\beta$ , and X $\delta$ 
    Solution (t+1) = (X $\alpha$ +X $\beta$ + X $\delta$ )/3
end for
t=t+1
end while
return alpha
    
```

Fig. 2. Pseudo-code of the GWO algorithm.

### III. MGWO OPTIMIZER

The procedure of locating the worldwide minimal is a hard task. GWO, as a population-primarily based totally set of rules, makes use of strategies to gain its work: exploration and exploitation. Exploration is the procedure of locating promising factors withinside the seek area, while exploitation is the procedure of locating higher factors close to correct answers observed so far. Exploration allows GWO to keep away from stagnation in neighborhood optimums. GWO encourages individuals to discover the quest area withinside the early iterations. Then, withinside the closing iterations, people carry out exploitation the use of the amassed statistics to converge on the worldwide minimal. GWO achieves stability among exploration and exploitation through the use of the 2 parameters  $A$  and  $a$ . The fee of  $a$  decreases linearly from 2 to zero at some point of the iterations in keeping with equation (3.1). Half of the iterations are dedicated to exploration ( $|A| > 1$ ) and the closing 1/2 is dedicated to exploitation ( $|A| < 1$ )

To decorate the stability among exploration and exploitation, we modified equation (4.1) so that the fee decreases exponentially over the route of the new release as defined in equation (3.2). By observe exponential extrude, the range of iterations used for exploration is improved and therefore the changed GWO achieves better exploration of the quest area for greater iterations. Fig (3) illustrates the distinction between a linear and exponential extrude of the fee of  $a$ . As proven withinside the figure, withinside the exponential approach, exploration is completed for an extra range of iterations.

$$a = 2 - t \frac{2}{MaxIter},$$
(6)

$$a = 2 - t^2 \frac{2}{MaxIter^2}$$
(7)

where  $t$  is the iteration's wide variety and  $MaxIter$  is the widest variety of iterations. The crossover is the operation that mixes facts of 3 leaders (alpha, beta, and gamma) to generate new offspring, it's far the manner to stochastically generate new answers from a present population. The crossover operation will increase the variety of the population and complements exploitation capability. Mutation operator random modifications one or extra additives of the offspring. It is used to save you untimely

convergence. The mutation operation is hired to decorate the location of a selected answer round randomly selected leaders. The resultant solution from the move over operation is hired as equation (8):

$$offspring = CrossOver(X_{\alpha}, X_{\beta}, X_{\gamma}) \quad (8)$$

$$\vec{X}(t+1) = Mutate(offspring)$$

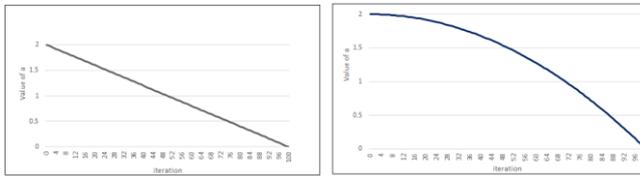


Fig. 3 difference between a linear and exponential change of the value of a. The pseudo-code of the MGWO algorithm is presented in Fig. 4.

```

Input Xi (i = 1, 2, ..., n)
Initialize the grey wolf population Xi (i = 1, 2, ..., n)
while (t < Max number of iterations)
  Calculate the fitness of each search agent
  Calculate α, β, and δ
  Calculate A as per Equation a = 2 - t^2 / MaxIter^2
  for each search agent
    Update A, and C
    Calculate Xα, Xβ, and Xδ
    offspring = crossover (Xα, Xβ, Xδ)
    solution (t+1) = Mutate (offspring)
  end for
  t=t+1
end while
return α
    
```

Fig. 4. Pseudo-code of the MGWO algorithm.

The Flow-chart for the proposed MGWO algorithm is presented in Fig. 5.

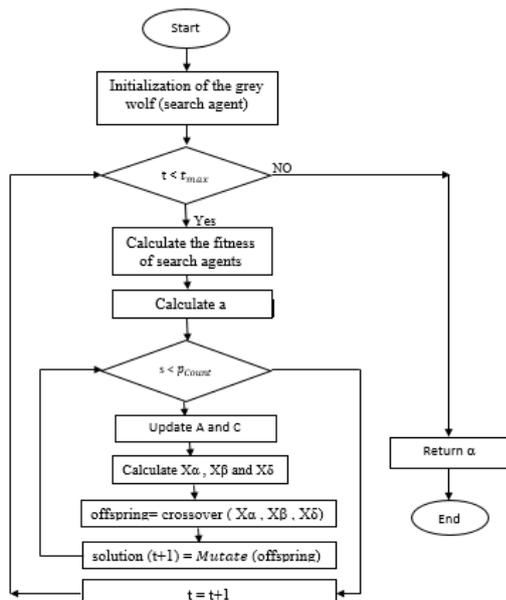


Fig. 5. Flow-chart for proposed MGWO algorithm.

To summarize, we introduce two different modifications

to the original GWO. The first modification enforces the parameter “a” to change exponentially and hence increases the number of iterations for exploration. On the other hand, the second modification is applying crossover and mutation on  $X_1$ ,  $X_2$ , and  $X_3$ . The crossover operator enhances the exploitation process while the mutation operator enhances the exploration process. By merging all these modifications, the proposed MGWO has a higher exploration capability than the original GWO. Moreover, MGWO has enhanced exploration capabilities than the original GWO.

#### IV. MGWO FOR FEATURE SELECTION

The problem with the feature is so unique because the search space is limited to two binary values 0 and 1. Therefore the traditional continuous version of the optimizer for the grey wolf should be updated to function properly for this problem. bMGWO was introduced to change GWO's continuous values to binary values which can be used for feature selection.

##### A. Binary Modified Grey Wolf Optimizer

To convert the values of standard GWO from continuous to binary values, we modified Eq. (5) to the following equations:

$$x_d^{t+1} = \begin{cases} 1 & \text{if } \text{sigmoid}\left(\frac{\alpha x_1 + \beta x_2 + \delta x_3}{\alpha + \beta + \delta}\right) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $x_d^{t+1}$  are the updated binary position of the dimension d at iteration t and  $\alpha, \beta$  and  $\delta$  are the fitness values for alpha, beta, and gamma leaders.

$$\text{sigmoid} = \frac{1}{1 + e^{-10(x-0.5)}} \quad (10)$$

The role of the sigmoid is to scale the continuous values of the GWO between 0 and 1. Then the usage of the situations we determine whether or not the cost of the size maybe 0 or one.

##### B. Solution Representation

For the hassle of feature selection, the answer may be represented as a vector of features with length h in which h is the wide variety of capabilities. Each object in that vector is a binary value (zero or 1) in which zero methods that the feature isn't always covered and 1 method the feature is covered.

##### C. Fitness Function

Fitness functions are used to measure the quality of each solution of the GWO. The fitness function depends on two factors: the classification error rate and the number of selected features., the following equation is used:

$$Fitness = h_1 E(D) + h_2 \frac{|s|}{|f|} \quad (11)$$

where  $h_1 \in [0, 1]$ ,  $h_2 = 1 - h_1$  E (D) is the classification error, s is the selected feature's number, and, d f is the feature's number.

#### D. K-Nearest Neighbor

In this paper, K-Nearest Neighbor (KNN) has been used, Wrapped feature selection technique Includes an algorithm for learning to evaluate the quality and power. [33].

#### V. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate our proposed algorithm, four publicly available datasets for EEG signal analysis were used. The first dataset is the BCI Competition IV. BCI competition IV was held in 2008 by the Graz University of Technology in Austria. In our study, dataset 2a of the mentioned competition is used for further analysis. This dataset is publicly available in [49]. The dataset consists of EEG data of nine healthy subjects. The subjects were healthy and normal individuals. They were asked to sit in a comfortable armchair in front of an LCD monitor to perform the motor imagery tasks. Subjects performed four different types of motor imageries to complete all the tasks. These tasks included the imagination of left or right hand, either feet or tongue movements. To perform the experimental paradigm, first, a short acoustic beep was presented. Then, a fixation cross appeared on the LCD monitor for two seconds and was replaced with an arrow pointing to one of the directions of up, down, right, or left. According to the direction of the arrow, the subjects performed one of the imagery tasks of the tongue, feet, and left hand or right-hand movements. The performer subject kept the imagination of the selected item for about three seconds until the fixation cross was disappeared and the LCD became completely black. Then, a short break of approximately two seconds was followed and the next task was started again. This process was done 72 times for each one of the four tasks, yielding a total of 288 cases of motor imageries per subject [34].

The second and third dataset is BCI Competition Data set III, which supplied via way of means of the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. (Gert Pfurtscheller). This dataset became recorded from an ordinary issue (female, 25y) all through a comments session. The issue sat in a chilled chair with armrests. The assignment became to manipulate a comments bar making use of imagery left or proper-hand movements. The order of left and proper cues became random. The test includes 7 runs with forty trials each. All runs had been performed at the equal day with numerous mins wreck in among. Given are 280 trials of 9s length. The first 2s became quiet, at t=2s an acoustic stimulus shows the start of the trial, The EEG became sampled with 128Hz, it became filtered among 0.5 and 30Hz. The facts aren't always posted yet, comparable experiments are defined in [35]. The trials for schooling and checking out had been randomly selected. This needs to save you any systematic impact because of the comments. The fourth dataset is EEG Eye State Data Set, the facts set supplied via way of means of the UC Irvine Machine Learning Repository. The UCI Machine Learning Repository is a set of databases, area theories, and facts turbines that are utilized by the device getting to know the network for the empirical evaluation of device getting to know algorithms. EEG Eye State Data Set is from one non-

stop EEG dimension with the Emotive EEG Neuroheadset. The length of the dimension became 117 seconds. The eye country became detected thru a digital digicam all through the EEG dimension and brought later manually to the document after studying the video frames. '1' shows the eye-closed and '0' the eye-open country. All values are in chronological order with the primary measured cost on the pinnacle of the facts.

TABLE I: DATASETS DESCRIPTION

Data-set Name	Number of rows	Number of columns (Feature)
BCI Competition IV- dataset 2a	45,000	22
BCI Competition IV- data set III	64,000	19
BCI Competition II data set III	120,000	32
EEG Eye State	14,000	14

#### A. Configurations

Each dataset is divided into three randomly equal-size parts: training, validation, and test. Training is used to train the KNN classifier during the learning phase. Validation is used to test. When calculating the fitness function for a specific solution.

TABLE II: EXPERIMENTS CONFIGURATION

Parameter	value
No of the search agents	10
No of iterations	80
Problem dimension	Number of features in the data
Search domain	[0 1]
No. repetitions of runs	20
Inertia factor of PSO	0.1
$\alpha$ Parameter in the fitness function	0.99
$\beta$ Parameter in the fitness function	0.01

#### B. Evaluation Metric

- **Classification average Error** calculated in equation (12).

$$AvgPref = \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N mse(C_i, L_i) \quad (12)$$

- **Best Fitness** can be formulated in equation (13).

$$best = \min_{i=1}^M g_*^i \quad (13)$$

- **Worst Fitness.** can be formulated in equation (14).

$$worst = \max_{i=1}^M g_*^i \quad (14)$$

- **Average Fitness size.** This measure can be formulated as in equation (15).

$$AVG Selection Size = \frac{1}{M} \sum_{i=1}^M \frac{size(g_*^i)}{D} \quad (15)$$

- **Mean.** can be formulated in equation (16).

$$Mean = \frac{1}{M} \sum_{i=1}^M g_*^i \quad (16)$$

- **Std (Standard Deviation).** Std is formulated as in equation (17).

$$Std = \sqrt{\frac{1}{M-1} \sum (g_*^i - Mean)^2} \quad (17)$$

C. Experimental Result and Analysis

Seven experimental are done to evaluate the performance of the MGWO optimizer these are shown below. In experimental 1, the average error of different algorithms is proven in Table three. The lower error suggests that the optimizer has decided on the right set of features that can train the classifier and convey a decrease in blunders at the hidden test data. We can commentary from desk three that the bottom blunders are finished through the proposed MGWO which proved to fairly discover the hunt space. After that, the proposed MGOW makes use of the crossover operator to transport in the direction of the optimal solution which incorporates the choicest subset of features that can reduce the blunders.

Experimental 2, the average selected features by different optimizers are shown in Table 4. Although, choosing a lower number of features indicates that the optimizer performed feature selection, maintaining lower error is more important. For that reason, the fitness function assigns a higher weight for the classification error and still encourages the optimizer to choose the lower number of features. As shown in the table, the proposed MGWO was able to find the least number of channels for (BCI Competition IV- dataset 2a, BCI Competition IV- data set III, and EEG Eye State) datasets and gave the lower classification for them. However, MGWO choose a higher number of features for

the BCI Competition II- Data set III dataset, it maintains the smallest error for that dataset.

The statistical results for experimental 3, experimental 4, experimental 5, and experimental 6 (Average, Best, Worst, and Standard Deviation) are reported in Tables 5, 6, 7, and 8. According to table 5, the proposed bMGWO turned into capable of locating the bottom health cost for all datasets because of this that it could pick the foremost subset of features that deliver the bottom class error. The cause for this excessive overall performance is the cooperative nature of the individuals of the GWO which applied the proposed amendment of “a” parameter and the mutation operator to fairly discover the quest area for special solutions. Moreover, the proposed crossover better the exploitation process. As according to desk 6, bMGWO turned into capable of locating the first-rate fitness in comparison to different optimizers for runs. On the alternative hand, in Table 7, bMGWO did now no longer locate the worst health in comparison to different optimizers which proves the functionality of the proposed MGWO to locate the foremost subset of features. Table 8 outlines the usual deviation for statistical results. As proven withinside the table, the proposed bMGWO has the lowest popular deviation in comparison to different optimizers. The lowest popular deviation proves the stability and the robustness of the proposed bMGWO.

TABLE III: AVERAGE ERROR RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

Dataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	0.09458	0.10820	0.13371	0.05813	0.11136	0.11619	0.13104
BCI Competition IV- dataset 2a	0.05629	0.45704	0.45795	0.45419	0.45820	0.45893	0.46269
BCI Competition IV- data set III	0.00097	0.00221	0.00193	0.00144	0.00211	0.00227	0.00320
BCI Competition II data set III	0.45503	0.05852	0.06599	0.09164	0.06058	0.06158	0.06766

TABLE IV: AVERAGE SELECT SIZE RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

Dataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	0.49773	0.83571	0.59844	0.98929	0.84286	0.79286	0.72143
BCI Competition IV- dataset 2a	0.48947	0.56818	0.60909	0.85682	0.60682	0.58864	0.92500
BCI Competition IV- data set III	0.46094	0.56579	0.61316	0.80263	0.59737	0.62368	0.67045
BCI Competition II data set III	0.60000	0.52656	0.71786	0.82188	0.62344	0.57969	0.63158

TABLE V: AVERAGE FITNESS RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

Dataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	3.4943	3.4988	3.5324	3.4979	3.5064	3.5039	11.3548
BCI Competition IV- dataset 2a	7.2952	7.3013	7.3023	7.3036	7.3003	7.2976	11.3548
BCI Competition IV- data set III	11.3528	11.3528	11.3543	11.3529	11.3534	11.3530	11.3548
BCI Competition II data set III	12.7987	12.7986	12.8026	12.7954	12.8034	12.7959	11.3548

TABLE VI: BEST FITNESS RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

Dataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	3.4935	3.4983	3.5301	3.4965	3.5060	3.4983	11.3525
BCI Competition IV- dataset 2a	7.2915	7.2984	7.2992	7.2970	7.2984	7.2959	11.3525
BCI Competition IV- data set III	11.3522	11.3523	11.3526	11.3524	11.3527	11.3523	11.3525
BCI Competition II data set III	12.7977	12.7962	12.7960	12.7868	12.8008	12.7902	11.3525

TABLE VII: WORST FITNESS RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

Dataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	3.5095	3.4994	3.5347	3.4994	3.5068	3.4950	11.3577
BCI Competition IV- dataset 2a	7.2989	7.3042	7.3053	7.2989	7.3102	7.2993	11.3577
BCI Competition IV- data set III	11.3536	11.3535	11.3566	11.3538	11.3544	11.3536	11.3577
BCI Competition II data set III	12.7996	12.8011	12.8091	12.8041	12.8060	12.8017	11.3577

TABLE VIII: STANDARD DEVIATION FITNESS RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

ataset	bMGWO	bGWO	bPSO	bWAO	bMVO	bFFA	bGA
EEG_Eye_State	0.0038	0.0006	0.0016	0.0006	0.0008	0.0007	0.0023
BCI Competition IV- dataset 2a	0.0039	0.0041	0.0043	0.0053	0.0027	0.0093	0.0023
BCI Competition IV- data set III	0.0051	0.0007	0.0032	0.0021	0.0079	0.0010	0.0023
BCI Competition II data set III	0.0005	0.0007	0.0032	0.0021	0.0079	0.0010	0.0023

TABLE IX: PROCESSING TIME RESULT FROM DIFFERENT OPTIMIZER IN OUR EXPERIMENTS

	EEG_Eye_State	BCI Competition IV- dataset 2a	BCI Competition IV- data set III	BCI Competition II data set III
bMGWO	196.213	771.796	930.693	2727.028
bGWO	255.989	1072.36	761.536	2658.27
bPSO	242.088	1152.42	950.214	2987.445
bWAO	305.786	1667.598	1154.606	3620.63
bMVO	277.536	1178.843	935.171	3040.844
bFFA	269.791	1151.469	970.975	2905.012
bGA	222.164	1060.143	1072.36	2043.791

In experimental 7 the processing time by different optimizers is shown in Table 9. The decrease processing time shows that the optimizer unearths the optimum subset of features in much less time. The proposed optimizer has aggressive effects in comparison to different optimizers. The quicker convergence time proves the excessive exploitations functionality of the proposed optimizer and the capacity to keep away from neighborhood optima. This proves the robustness and reliability of bMGWO in locating the optimum subset of features in an affordable quantity of time.

As shown in the experiments result in previous tables, we can see that the bMGWO outperformed other optimizers in nine dataset, this is due to high exploration and exploitation of bMGWO which allow it to find the best subset of feature, which confirms its robustness and reliability in classification tasks in the various dataset in finding the optimal subset of features.

## VI. CONCLUSION

In this work, we proposed a modified version of the Grey wolf optimizer bMGWO that is used with the KNN classifier to select the optimal subset of features for the problem of EEG channel selection. We merge two modifications to achieve the balance between exploration and exploitation, via practice exponential change, the range of iterations used for exploration increases, and therefore the changed GWO to attain better exploration of the quest area for greater iterations. In the exponential approach, exploration is executed for an extra number of iterations and uses the crossover operation to increase the variety of the population and decorate exploitation capability. To evaluate the quality and effectiveness of the proposed solution, four different EEG datasets with a cross-validation method are used to measure the stability of the proposed optimizer. The results proved that the bMGWO has better performance than other optimizers. For future perspectives, we are planning to test the proposed solution on more complex datasets. Moreover, a parallel version of the bMGWO will be tested to reduce the processing time.

## ACKNOWLEDGMENT

The authors would like to thank the UCI Machine Learning Repository for sharing the datasets used in this research. I would also like to thank my friends and family

who supported me and offered deep insight into the study.

## REFERENCES

- [1] Sylwester, R. (1995). A celebration of neurons: An educator's guide to the human brain. Association for Supervision and Curriculum Development, 1250 N. Pitt St., Alexandria, VA 22314 (ASCD Stock No. 1-95085).
- [2] Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352.
- [3] Swain, J. E., Dayton, C. J., Kim, P., Tolman, R. M., & Völling, B. L. (2014). Progress on the paternal brain: Theory, animal models, human brain research, and mental health implications. *Infant Mental Health Journal*, 35(5), 394-408.
- [4] Fazel-Rezai, R. (Ed.). (2011). Recent advances in brain-computer interface systems. *BoD-Books on Demand*.
- [5] Vaid, S., Singh, P., & Kaur, C. (2015, February). EEG signal analysis for BCI interface: A review. In 2015 Fifth International Conference on Advanced Computing & Communication Technologies (pp. 143-147). IEEE.
- [6] Taha Al-Kasasbeh, R., Salman Shamaseen, M., & Skopin, D. E. (2008). Automated detection and selection of artifacts in encephalography signals. *Biomedical Engineering*, 42(6), 293-301.
- [7] Amjed S. Al-Fahoum and Ausilah A. Al-Fraihat, "Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains," *ISRN Neuroscience*, Hindawi Publishing Corporation, Article ID 730218, 2014.
- [8] D. Puthankattil Subha, Paul K. Joseph, Rajendra Acharya U, Choo Min Lim," EEG Signal Analysis: A Survey," *Springer Science* vol.34, pp.195-212, 2007.
- [9] Shiliang Sun and Jin Zhou," A Review of Adaptive Feature Extraction and Classification Methods for EEG-Based Brain-Computer Interfaces," in *IEEE Joint Conference (IJCNN) on Neural Networks International*, pp. 1746 - 1753, 2014.
- [10] Chizi, B., Rokach, L., & Maimon, O. (2009). A survey of feature selection techniques. In *Encyclopedia of Data Warehousing and Mining*, Second Edition (pp. 1888-1895). IGI Global.
- [11] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.
- [12] Bell, D. A., & Wang, H. (2000). A formalism for relevance and its application in feature subset selection. *Machine learning*, 41(2), 175-195.
- [13] El-kenawy, E. S. M. T. (2018). Solar Radiation Machine Learning Production Depend on Training Neural Networks with Ant Colony Optimization Algorithms. *IJARCCCE*, 7(5). doi: DOI10.17148/IJARCCCE.2018.751.
- [14] Tubishat, M., Abushariah, M. A., Idris, N., & Aljarah, I. (2019). Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. *Applied Intelligence*, 49(5), 1688-1707.
- [15] Das, S. (2001, June). Filters, wrappers and a boosting-based hybrid for feature selection. In *Icml* (Vol. 1, pp. 74-81).
- [16] Yang, C. H., Chuang, L. Y., & Yang, C. H. (2010). IG-GA: a hybrid filter/wrapper method for feature selection of microarray data. *Journal of Medical and Biological Engineering*, 30(1), 23-28.
- [17] A. M. Turing, "Computing machinery and intelligence," *Mind*, 59: 433-460, 1950.

- [17] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371-381.
- [18] Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
- [19] Aljarah I, Faris H, Mirjalili S (2018) Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Comput* 22(1):1–15
- [20] Emary, E., Zawbaa, H. M., Grosan, C., & Hassanien, A. E. (2015). Feature subset selection approach by gray-wolf optimization. In *Afro-European Conference for Industrial Advancement* (pp. 1-13). Springer, Cham.
- [21] El-Kenawy, E. S., & Eid, M. Hybrid Gray Wolf And Particle Swarm Optimization For Feature Selection.
- [22] Hassib, E. M., El-Desouky, A. I., Labib, L. M., & El-kenawy, E. S. M. (2020). WOA+ BRNN: An imbalanced big data classification framework using Whale optimization and deep neural network. *soft computing*, 24(8), 5573-5592.
- [23] Fouad, M. M., El-Desouky, A. I., Al-Hajj, R., & El-Kenawy, E. S. M. (2020). Dynamic group-based cooperative optimization algorithm. *IEEE Access*, 8, 148378-148403.
- [24] Ibrahim, A., El-kenawy, E. S. M. (2020). Image Segmentation Methods Based on Superpixel Techniques: A Survey. *Journal of Computer Science and Information Systems*, 15 (3 October 2020).
- [25] MADADI A., MOTLAGH M.M. Optimal control of DC motor using grey wolf optimizer algorithm. *Technical Journal of Engineering and Applied Sciences*. 2014, 4(4), pp. 373–379.
- [26] El-kenawy, E. S. M. T. (2019). A Machine Learning Model for Hemoglobin Estimation and Anemia Classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 17(2).
- [27] RATHEE P., GARG R., MEENA S. Using grey wolf optimizer for image registration. *International Journal of Advance Research in Science and Engineering*. 2015, 4(4), pp. 360–364.
- [28] Ibrahim, A., El-kenawy, E. S. M. (2020). Applications and Datasets for Superpixel Techniques: A Survey. *Journal of Computer Science and Information Systems*, 15 (3 October 2020).
- [29] E. Rashedi, Nezamabadi-pour H: Feature subset selection using improved binary gravitational search algorithm, *J. Intell. Fuzzy Syst.* 26 (2014) 1211–1221.
- [30] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, BGSA: binary gravitational search algorithm, *Natural Comput.* 9 (2010) 727–745.
- [31] BCI Competition IV, <http://www.bbci.de/competition/iv/>. 2008.
- [32] El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., Eid, M. M., & Hussein, S. E. (2020). Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE Access*, 8, 179317-179335.
- [33] C. Brunner, R. Leeb, G.R. Muller-Putz, A. Schlogl, G. Pfurtscheller, BCI competition 2008–Graz data set a, in: *Institute for Knowledge Discovery & Institute for Human-Computer Interfaces*, Graz University of Technology, Austria, 2008, pp. 1–6.
- [34] Pfurtscheller G, Neuper C, Schlögl A, Lugger K. (1998) Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans Rehabil Eng.* 6(3):316-25.
- [35] El-Kenawy, E. S. M., Eid, M. M., Saber, M., & Ibrahim, A. (2020). MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection. *IEEE Access*, 8, 107635-107649.



**Hussien R. Hussien** received a B.Sc. degree in computer engineering from the Faculty of Engineering, Mansoura University, Egypt, in 2011, were currently pursuing a master's degree. involved in the software engineering area in the industry since 2012, with expertise in machine learning, Brain signal processing, and optimization algorithms.



**El-Sayed M. El-Kenawy** (Member, IEEE) is currently an Assistant Professor with the Delta Higher Institute for Engineering and Technology (DHIET), Mansoura, Egypt. He has inspired and motivated students by providing a thorough understanding of a variety of computer concepts. He has pioneered and launched independent research programs. He is interested in computer science and machine learning fields. He is adept at explaining sometimes complex concepts in an easy-to-understand manner.



**Ali Ibrahim El-Desouky** received the M.A. and Ph.D. degrees from the University of Glasgow, USA. He is currently a Full Professor with the Department of Computer Engineering and Systems, Faculty of Engineering, Mansoura University, Egypt. He is also a visiting part-time Professor with the MET Academy. He has been teaching computer engineering at American University and Mansoura University. He has taken over many positions of leadership and supervision of many scientific articles. He has published hundreds of articles in well-known international journals.