scientific reports

OPEN

EEG-based optimization of eye state classification using modified-BER metaheuristic algorithm

Ahmed M. Elshewey[1✉], Amel Ali Alhussan[2], Doaa Sami Khafaga[2], El-Sayed M. Elkenawy[3,4,6,7] & Zahraa Tarek[5]

This article introduces the Modified Al-Biruni Earth Radius (MBER) algorithm, which seeks to improve the precision of categorizing eye states as either open (0) or closed (1). The evaluation of the proposed algorithm was assessed using an available EEG dataset that applied preprocessing techniques, including scaling, normalization, and elimination of null values. The MBER algorithm's binary format is specifically designed to select features that can significantly enhance the accuracy of classification. The proposed algorithm and competing ones, namely, Al-Biruni Earth Radius (BER), Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WAO), Grey Wolf Optimizer (GWO) and Genetic Algorithm (GA) were evaluated using predefined sets of assessment criteria. The statistical analysis employed the ANOVA and Wilcoxon signed-rank tests and assessed the effectiveness and significance of the proposed algorithm compared to the other five algorithms. Furthermore, A series of visual depictions were presented to validate the effectiveness and robustness of the proposed algorithm. Thus, the MBER algorithm outperformed the other optimizers on the majority of the unimodal benchmark functions due to these considerations. Different ML models were used for classification, e.g., DT, RF, KNN, SGD, GNB, SVC, and LR. The KNN model achieved the highest values of Precision (PPV) (0.959425), Negative Predictive Value (NPV) (0.964969), FScore (0.963431), accuracy (0.9612), Sensitivity (0.970578) and Specificity (0.949711). Thus, KNN serves as a fitness function and is optimized by the utilization of Modified Al-Biruni earth radius (MBER). Finally, the accuracy of eye state classification achieved 96.12% using the proposed algorithm.

Keywords Al-Biruni Earth radius optimization, Feature selection, Eye state, Classification, Meta-heuristic optimization

Brain–Computer Interface (BCI) stands at the forefront of advanced technologies, ushering in a new era where the computational prowess of the human brain can be harnessed for various applications[1]. In the not-so-distant past, the idea of developing BCIs seemed confined to the realms of science fiction. However, a transformative breakthrough emerged with the discovery of electroencephalography (EEG), altering the landscape of possibilities[2]. EEG, a revolutionary technique, enabled researchers to capture and interpret electrical signals directly from the brain. This pivotal development shattered preconceived notions and ignited a fervor among scientists and innovators to delve into the intricate realm of decoding these EEG signals[3]. The realization that the human brain could communicate its intricate patterns through measurable electrical activity opened doors to a multitude of research avenues. Fueling the pursuit of understanding and harnessing the potential of EEG signals, researchers embarked on a journey to unlock the secrets of the brain's computational power[4]. This newfound capability not only transcended the boundaries of imagination but also fueled a wave of collaborative efforts to bridge the gap between neuroscience and technology. The once fantastical concept of BCI began to materialize,

[1]Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.Box: 43221, Suez, Egypt. [2]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia. [3]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt. [4]MEU Research Unit, Middle East University, Amman 11831, Jordan. [5]Department of Computer Sciences, Faculty of Computers and Information, Mansoura University, Mansoura 35561, Egypt. [6]School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain. [7]Jadara University Research Center, Jadara University, Jordan, Jordan. ✉email: ahmed.elshewey@fci.suezuni.edu.eg

offering a glimpse into a future where the synergy between the human brain and computational systems could revolutionize how we interact with and manipulate technology[5].

The BCI domain has a clear and intertwined division into four major stages, all vital for practically interpreting the brain's language. These phases, namely, signal acquirement, signal processing, feature extraction and classification from an empirical perspective, form the framework that defines how the human brain interacts with computational systems efficiently[6]. In the first step of Signal Acquisition, the process is carried out by recording EEG currents directly on the scalp without necessarily having to penetrate the head. This is done via the positioning of electrodes; these are instruments that receive signals of an electrical nature produced by the brain. Following this, the acquired signals are amplified through signal amplifiers, which aim to increase the signal's power and improve the process's quality. After amplification, the signals are converted into digital form, which involves converting them into a more computational format. In signal processing, attention is turned to improving the obtained data and its preparation for more stages. It is a process that entails using strategies that seek to eliminate noise and other artifacts[7]. Any external acoustic interferences are the natural source of noise, while the movements of muscles or blinks of eyes are the artifacts forming the additional interference. Signal processing has a central role in handling these problems to achieve the signal's cleanliness before further analysis[8]. These then become defined in forms likely to be quickly processed in the following stage. Thus, by carefully passing through these preliminary phases, the BCI system prepares for extracting valuable signals from the complex neural topologies of the brain. The Signal Processing stage, which in particular plays a filter role—arranges the raw data and transforms it into suitable input for the subsequent stages of the BCI procedure, making way for the successful integration of the human brain with the state-of-the-art computational systems[9]. Augmenting the ever-famous optimizer and ensemble regression models based on machine learning can produce promising results for classification and regression issues[10].

In the realm of EEG-based classification, researchers often integrate machine learning (ML) techniques to establish a mapping between EEG input signals from the brain and desired outputs, such as classes, commands, or targets, within BCI systems[11]. However, a notable challenge in the application of deep learning to EEG lies in the tendency of researchers to devise task-specific data processing and ML classification pipelines. These pipelines are designed to address challenges like noisy and non-stationary EEG signals, the need for feature engineering, and the inherent variability in EEG data. Moreover, the current landscape reveals a common practice where researchers create bespoke data processing pipelines, intertwined with ML hyperparameter tuning, to navigate the intricacies of EEG data[12]. This includes addressing challenges like non-stationarity, conducting feature engineering, and managing variability in EEG signals. Additionally, this process often involves a manual element, with researchers inspecting and fine-tuning the system based on their insights and domain knowledge[13]. In essence, the convergence of EEG-based neural classification, machine learning techniques, and deep learning methodologies holds immense promise for transforming the lives of individuals with disabilities, offering novel means of communication, device control, and capability enhancement. However, the challenges posed by noisy and dynamic EEG data necessitate ongoing research and innovation in developing adaptable and robust data processing and classification pipelines[14]. Consequently, researchers typically opt to collect and select only a fraction of the available information, engaging in the visual inspection of data to identify and discard artifacts. Manual parameter tuning is then applied, adding complexity to the processing pipeline. This approach introduces a multitude of possible combinations and demands a substantial depth of task-specific domain expertise[15].

Today, the employed models that utilize deep learning and machine learning approaches reveal excellent performance in numerous domains, especially in computer vision and medicine. This has opened the stage for much progress once these advanced models are mingled with modern technology. Furthermore, numerous articles and research papers have been systematically reviewed in the literature, which aim to improve the outcomes of both deep learning and machine learning methods[16]. This upswing of investigation shows an interest in fine-tuning these models to fit the current difficulties and complexities in the actual applications. These models also reveal effectiveness in tasks of image recognition, object detection, scene classification, etc., and tasks in computer vision as a step towards creating novel intelligent systems that can read visual references with immense accuracy[17]. Likewise, in the competency of medicine, the use of deep learning and machine learning models has shaped the new generation of diagnostic and prognostic measurements and how treatments can be planned. From identifying tumors in an image to assessing a patient's prognosis from copious and detailed data generated, these models have begun to illustrate their ability to revolutionize healthcare[18]. The list of current and future studies indicates that this line of work is indeed active and progressive, and, thus, it requires consistent development and furthering. These combined efforts help develop and sustain the continuous advancement of deep learning and machine learning models, determining their relevance to addressing the complex problems of modern computer vision and medicine[19].

In the domain of EEG classification, metaheuristics play a crucial role in optimizing the performance of classification algorithms for EEG data. EEG signals are complex and dynamic, often containing noise and artifacts that can pose challenges for accurate classification. Metaheuristic approaches are employed to enhance the efficiency and effectiveness of traditional EEG classification algorithms by optimizing parameters, configurations, and decision boundaries[20]. The significance of metaheuristics becomes particularly pronounced when grappling with intricate or high-dimensional datasets, where conventional algorithms may encounter limitations. In such scenarios, metaheuristics offer a dynamic and adaptive approach to refining and fine-tuning classification processes[21]. Illustrative examples of metaheuristics tailored for classification purposes encompass genetic algorithms, simulated annealing, particle swarm optimization, ant colony optimization, and evolutionary algorithms. These metaheuristic methods introduce a level of sophistication by navigating through the solution space more efficiently. Consequently, this often leads to superior outcomes in terms of classification accuracy, processing speed, and overall robustness[22]. The utilization of metaheuristics in classification tasks reflects a strategic and versatile problem-solving approach. By embracing a diversity of solutions and surmounting the constraints encountered by traditional algorithms, metaheuristics empower machine learning systems to navigate the intricacies of complex data patterns with increased adaptability and efficacy.

Literature review

Electroencephalography (EEG) has become a cornerstone technology in the study and diagnosis of neurological conditions, as well as in the development of Brain-Computer Interfaces (BCIs). The ability to classify and interpret EEG signals accurately is essential for various applications, ranging from epilepsy detection to lie detection and mental state recognition[23]. Recent advancements in machine learning and signal processing have significantly enhanced the precision and reliability of EEG-based classification systems[24,25].

Polat et al.[26] identified epileptic seizures in EEG signals by employing a hybrid system that incorporates a decision tree classifier and fast Fourier transform, achieving a notable classification accuracy of 98.72%. Chandaka et al.[27] presented a classifier based on support vector machines (SVM) assisted by cross-correlation. Their approach resulted in a classification accuracy of 95.96% when distinguishing between normal and epileptic EEG data. Farwell et al.[28] conducted a "Guilty Knowledge Test" using Event-Related Brain Potentials (ERPs) to determine the culpability or innocence of individuals. This examination involved three stimulus types: probe stimuli, target stimuli, and irrelevant stimuli. The occurrence of a P300 response is notable when an individual encounters a familiar object, such as the target stimuli. Even when an individual possesses "guilty information" but provides false information, a P300 response is still triggered, indicating awareness of specific details about the displayed object. The system's performance was assessed through Bootstrapped analysis, revealing no instances of false positives or negatives in the experiment. However, in 12.5% of cases, intermediate results were observed.

Haider et al.[29] presented a lie detection technique using Linear Discriminant Analysis (LDA) to differentiate between positive and negative samples. Sixteen channels were employed, and signal extraction utilized diverse methods. The methodology was implemented using MATLAB and the Xilinx tool, with the complete system deployed on an FPGA for efficiency evaluation. The proposed approach achieved an accuracy of 85% and was highlighted as a straightforward and more user-friendly alternative compared to previously proposed methods. Simbolon et al.[30] presented an approach that utilizes Support Vector Machines (SVM) to discern between guilt and innocence, relying on Event-Related Potential (ERP). The entire process was executed using MATLAB. Eleven male participants, aged between 20 and 27, were chosen for the experiment. The dataset was divided into training and testing sets, and various models were developed. The method demonstrated an accuracy reaching up to 70.83%. Feature extraction involved the use of Hjorth parameters, and the classifier employed in this study was the k-nearest neighbor.

Bablani et al.[31] presented an approach to detect deception by analyzing EEG signals from the brain. Detecting deceit is particularly challenging, as it is essential to prevent wrongful convictions of innocent individuals. Therefore, the precision and accuracy of the system's outcomes are paramount. The proposed method integrates Hjorth parameters, encompassing activity, mobility, and complexity. Following subject-specific analysis, we obtained optimal results, with the mobility parameter demonstrating a performance of up to 96.7% for subject 6. Khare et al.[32] undertook the categorization of EEG recordings associated with four distinct mental states, employing a set of five classifiers. Particularly noteworthy is the finding that the most favorable result in this study was attained by employing the resilient backpropagation method, achieving an accuracy rate of 95%. Bayram et al.[33] used planning relax dataset, obtained from the UCI Machine Learning Repository, was subjected to classification using Support Vector Machines (SVM). To optimize accuracy and identify key features, four distinct feature selection algorithms were employed on the dataset. Various SVMs were trained with the original data to ascertain the most appropriate kernel. The highest accuracy achieved with the original data was through the Radial Basis Function (RBF) kernel, reaching 71.43%. Following Sequential Forward Selection (SFS) on the data, accuracy improved to 72.53%, utilizing only one feature. A particularly noteworthy outcome was observed with Sequential Backward Elimination (SBE), yielding an accuracy of 74.73% with eleven features.

Rajaguru et al.[34] utilized Logistic Regression (LR) for its ability to facilitate result analysis in both explanatory and predictive contexts. These studies focus on signal dimensionality reduction before applying logistic regression. The results showcase a remarkable classification accuracy of 97.91% achieved with the Gaussian logistic regression model. Guerrero et al.[35], the study employed traditional classification methods on frequency data extracted from distinct channels containing informative details from EEG examinations. Fourier analysis was utilized for feature extraction within specific frequency bands. The implementation of classification techniques was carried out using Python. Following a comprehensive comparison of metrics and performance, it was concluded that artificial neural networks proved to be the most effective approach for characterizing epileptic patients, attaining an accuracy rate of 86%.

Table 1 provides a comprehensive summary of various methodologies and their corresponding accuracies in EEG signal classification. The table encapsulates a range of studies that employ diverse techniques to achieve high classification accuracy in different contexts, such as epilepsy detection, lie detection, and mental state recognition.

Each study leverages specific algorithms and methodologies to enhance the precision and reliability of EEG-based systems. From hybrid systems combining decision trees with fast Fourier transform to support vector machines and logistic regression, the table highlights the effectiveness and innovation of these approaches. This summary underscores the critical advancements in EEG signal analysis and their potential applications in medical and forensic fields.

The proposed methodology

An EEG dataset is used in this context to build an eye state classification model. During the EEG measurement, an eye camera identified the eye state and added it to the file manually after analyzing the video frames; a total of two classes were created: 1 for eye closure and 0 for eye-opening. Data is then preprocessed, which involves

Study	Objective	Methodology/technique	Accuracy	Notable details
Polat et al.[26]	Identify epileptic seizures	Decision tree classifier, FFT	98.72%	Hybrid system combining decision tree and fast Fourier transform
Chandaka et al.[27]	Classify normal vs. epileptic EEG data	SVM, cross-correlation	95.96%	Utilized SVM with cross-correlation to distinguish between normal and epileptic EEG data
Farwell et al.[28]	Guilty Knowledge Test	ERPs, P300 response, Bootstrapped analysis	–	Used ERPs to determine culpability with no false positives or negatives; 12.5% intermediate results
Haider et al.[29]	Lie detection	LDA, signal extraction, FPGA deployment	85%	Technique employed 16 channels and was implemented using MATLAB and Xilinx tool, deployed on FPGA for efficiency
Simbolon et al.[30]	Discern guilt vs. innocence	SVM, ERP, Hjorth parameters, KNN	70.83%	Experiment involved 11 male participants, aged 20–27, and feature extraction using Hjorth parameters
Bablani et al.[31]	Detect deception	Hjorth parameters (activity, mobility, complexity)	96.7%	Precision in preventing wrongful convictions; mobility parameter showed optimal performance for subject 6
Khare et al.[32]	Categorize four mental states	Five classifiers, resilient backpropagation	95%	Most favorable result obtained using resilient backpropagation method
Bayram et al.[33]	Classify planning relax dataset	SVM, feature selection (SFS, SBE)	74.73%	Highest accuracy with SBE using 11 features; original data accuracy 71.43% with RBF kernel, improved to 72.53% with SFS using one feature
Rajaguru et al.[34]	EEG signal classification	Logistic Regression (Gaussian model)	97.91%	Focused on signal dimensionality reduction before applying logistic regression
Guerrero et al.[35]	Characterize epileptic patients	Fourier analysis, ANN	86%	Traditional classification methods on frequency data from EEG; artificial neural networks were most effective

Table 1. Summary of EEG signal classification studies.

removing empty values, Min-Max scaling, and normalization. After the initialization process, the training data is collected in batches and then inputted into a model for iterative analysis. Feature selection is applied using different optimization algorithms, e.g., Al-Biruni earth radius (BER), Whale Optimization Algorithm (WAO), Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA). These algorithms are converted to binary by applying the sigmoid method to continuous data. Six evaluation parameters, average error, average select size and (average, worst, standard deviation, and best) fitness, are used to evaluate the binary optimization algorithms. The proposed Al-Biruni earth radius (bBER) algorithm for feature selection is shown in Algorithm 1.

```
Initialize the individuals of population, maximum iterations, fitness function, and BER parameters
Convert individuals to binary [0 or 1]
Compute the fitness function F_n for the individuals
Locate the optimal solution
while not reached maximum iterations, do
    For each solution in exploration group, do
        Progressing towards the optimal solution
    End for
    For each solution in exploitation group, do
        Best Solution Elitism
        Examine the region surrounding the optimal solution.
        Compare exploration group solutions and choose the optimal solution
        If there was no change in optimal fitness after 2 iterations, then
            Mutate solution
        End if
    End for
    Modify the fitness function
End while
Return the optimal solution
```

Algorithm 1. Pseudo Code of the binary BER

The dataset is partitioned into two distinct groups: 80% is allocated for the training set, while the remaining 20% is designated for the testing set. Some traditional machine learning algorithms as reference models are used to classify eye states using the embedded dataset. These algorithms are Stochastic Gradient Descent (SGD), Support Vector Classifier (SVC), Gaussian Naive Bayes (GNB), Random Forest (RF), K Nearest Neighbors (KNN), Decision Tree (DT), and Logistic Regression (LR). The evaluation metrics, Precision (PPV), Negative Predictive Value (NPV), F-score, accuracy, sensitivity, and specificity, are used in comparison between the reference models. KNN, one of the reference models, is used as a fitness function and optimized using Modified BER. The performance of the proposed optimized KNN by MBER is compared by state-of-the-art continuous algorithms, BER, PSO, GWO, WAO, and GA. The stages depicted in Fig. 1 are used in the proposed methodology for classifying eye states.

k-Nearest neighbor algorithm)KNN(

An effective and straightforward supervised ML technique utilized in regression and classification issues is the KNN algorithm. This method carries out the classification process using similarity criteria while taking the distance into account. A class is selected by tallying the majority of votes from neighboring positions that correspond to the nearest class[36]. First, the points are grouped according to their shared characteristics in an n-dimensional area, as n is the total no. of input parameters. After that, for each new position, the k closest positions are chosen and examined to see which class has the most points nearby. Since k is usually a small number, it may be expected that a hypercube would form, with its center at the new point and expanding until k points fall within of it. The points of the hypercube are then tallied, and it is established from which cluster more points are present. As a result, a new point is added to that cluster, and the cluster prediction technique is used to forecast the output of that cluster. The performance of the KNN is greatly influenced by the value of k, particularly when there are noisy individuals present. Using heuristic techniques to choose the ideal k is a successful strategy for creating a powerful machine-learning model[37,38]. Figure 2 presents the KNN algorithm and illustrates how the choice of K value has a significant impact on the outcome.

After classifying the data points, an output estimate is carried out. The weighted average approach is one way to do it. In this method, the output of the new point is more strongly influenced by the points that are nearby. The weight in this situation is the inverse of the distance. The neighbors are assigned weights, with the closest neighbor having a greater weight contribution to the average compared to the neighbor who resides farther away. Neighbors are assigned weights according to their Euclidean distance, or Manhattan, or Minkowski as shown in the following equations:

$$d = \sqrt{\sum (z_i - y_i)^2} \tag{1}$$

$$d = \sum (z_i - y_i) \tag{2}$$

$$d = \left(\sum (z_i - y_i{}^p)\right)^{\frac{1}{p}} \tag{3}$$

where z_i and y_i are the parameters of the two locations, and p is a real number within 1 and 2.

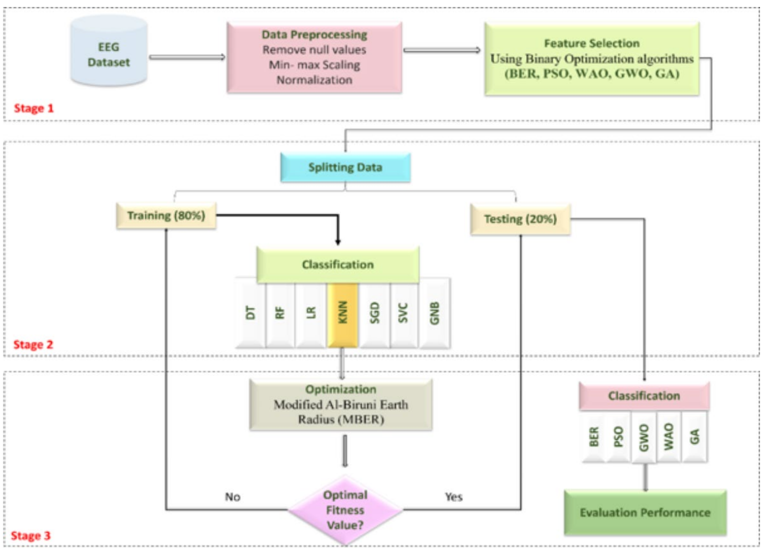

Fig. 1. The proposed framework for eye state classification.

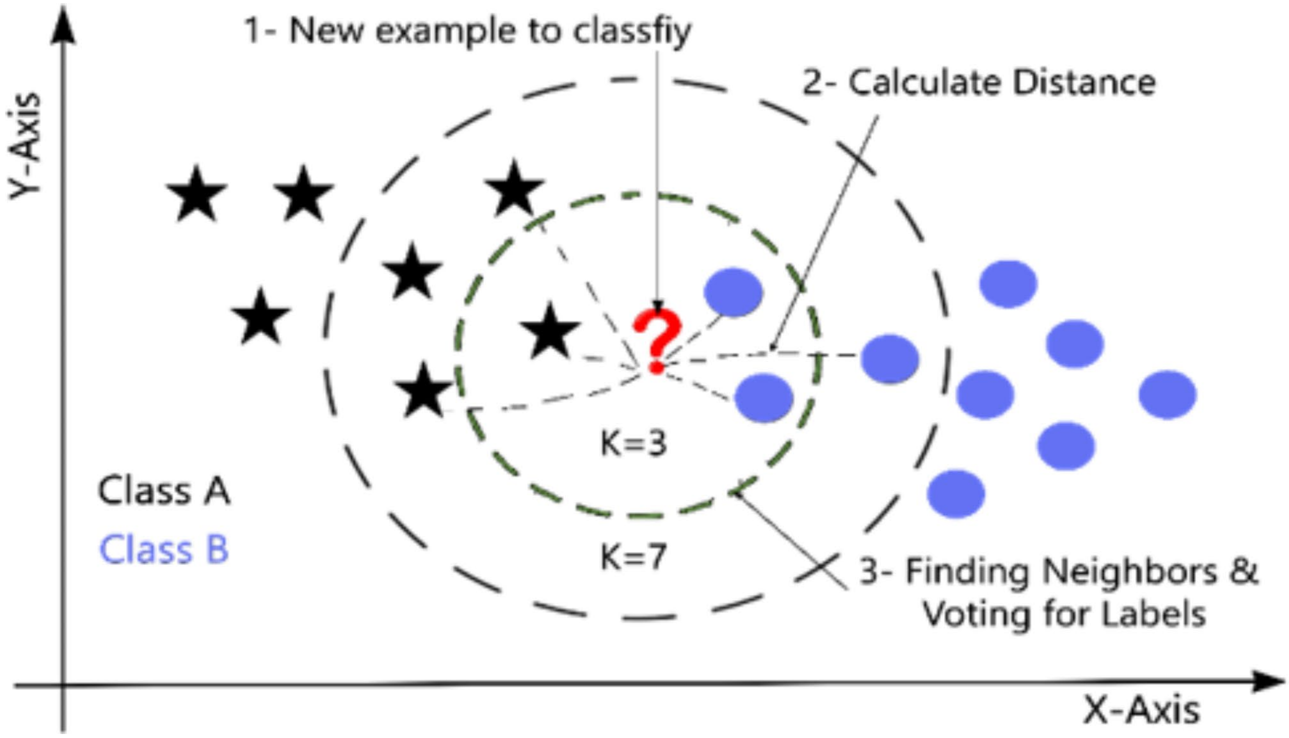

Fig. 2. An example of the KNN model.

Al-Biruni earth radius optimization algorithm

An optimization technique called the Al-Biruni Earth Radius (BER) technique first chooses a population at random and then divides it into exploration and exploitation groups. To improve the ability of balancing between the exploitation and exploration processes, agents are divided into Subcategories and dynamically altered within each subgroup. In comparison to the exploitation group, which makes up 30% of the population, the exploration group constitutes 70%[39,40]. One individual from the total population may be represented as the S vector which is presented as $S^{\rightarrow} = \{\mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3, \ldots \ldots, \mathrm{S}_d\} \in \mathrm{R}_d$, where $\mathrm{S_i}$ indicates the space size of search, and d is the feature in the optimization issue. Prior to BER initiating the optimization procedure, it needs the population size, the dimension, the fitness function, and the upper and lower bounds for every solution. The exploitation and exploration groups' members are modified to accommodate a more significant rise in the global average fitness of individuals in order to boost the fitness levels of the individuals in each group. Mathematically, the agents of the exploration group examine the search space for possible locations surrounding its present location. Searching individuals for an ideal vector that optimizes fitness function is done using these phases of the optimization approach. This is achieved by continuously looking over the surrounding feasible possibilities for a higher alternative in terms of fitness value[41]. Figure 3 describes the steps and equations of the BER algorithm.

Exploration operation

Exploration is in charge of not only the identification of interesting regions in the search space, but also the escaping from local optima stagnancy which is discussed below through moving towards the best solution found so far.

Heading towards the best solution The strategy is employed by the individual in the exploration group to find out the potential areas in the search space neighbouring its current position. This is done by continually searching for a better solution among the neigboring set of feasible solutions according to the fitness value. The following equations are utilized in the BER investigation for this purpose:

$$r = h\,\frac{cos\,(x)}{1-\cos\,(x)} \tag{4}$$

$$\overrightarrow{D} = \overrightarrow{r_1}.\left(\overrightarrow{S}\,(t)-1\right) \tag{5}$$

$$\overrightarrow{S}\,(t+1) = \overrightarrow{S}\,(t) + \overrightarrow{D}.\,(2\overrightarrow{r_2}-1) \tag{6}$$

where $0 < x \leq 180$, h is a number that is randomly selected from the range [0, 2], $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are coefficient vectors whose values are measured by these equations, $\overrightarrow{S}\,(t)$ is the solution vector at iteration t, and $\overrightarrow{D}$ is the diameter of the circle in which the search agent will look for promising areas.

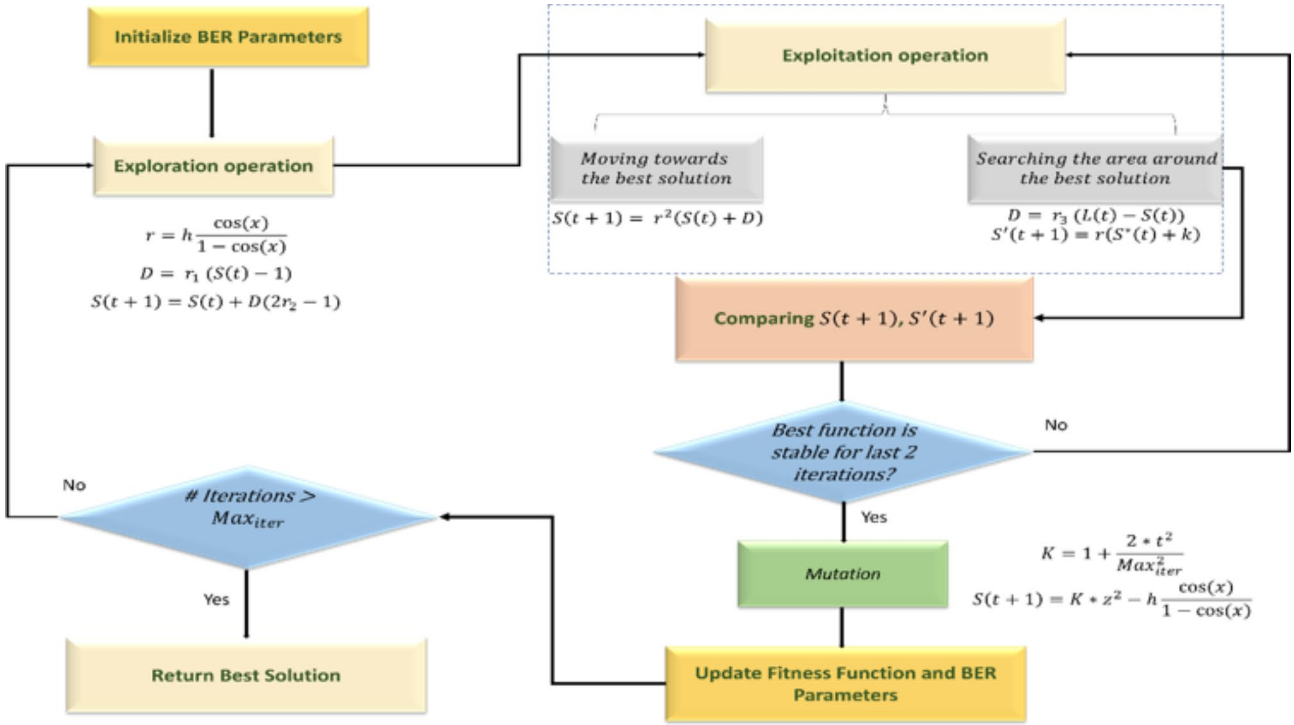

Fig. 3. The framework of BER algorithm steps and equations.

Exploitation operation
The exploitation team is in charge of improving existing solutions. The BER calculates all individuals' fitness values at each cycle and distinguishes the best individual. The BER employs two different approaches to achieve exploitation, as detailed in the sections below.

Heading towards the best solution The following equations are used to move the search agent towards the best solution:

$$\vec{S}\,(t+1) = r^2\left(\vec{S}\,(t) + \vec{D}\right) \tag{7}$$

$$\vec{D} = \vec{r_3}\left(\vec{L}\,(t) - \vec{S}\,(t)\right) \tag{8}$$

where $\vec{r_3}$ is a random vector calculated that controls the movement steps towards the best solution, $\vec{S}\,(t)$ is the solution vector at iteration t, $\vec{L}$ is the best solution vector, and $\vec{D}$ refers to the distance vector.

Investigating area around best solution Most promising is the region surrounding the best solution (leader). As a result, some individuals hunt in the vicinity of the best solution with the potential of finding a better solution. To realize this operation, the BER utilizes the following Eq.

$$\vec{S}\,(t+1) = r\left(\vec{S*}\,(t) + \vec{k}\right) \tag{9}$$

$$\vec{k} = z + \frac{2\,\times\,t^2}{N^2} \tag{10}$$

where $\vec{S*}$ refers to the best solution, z is a random number in the range [0, 1], t is the iteration number, and N is the total number of iterations.

Modified-BER optimization algorithm

In the Modified-BER optimization algorithm (MBER), a significant enhancement is introduced by selecting the top two leaders that yield the best results. This involves the calculation of the average fitness of these leaders using specific equations within the algorithm. Notably, the modification focuses on identifying and utilizing the most promising solutions to guide the algorithm's evolution. By emphasizing the performance of the leaders, MBER aims to enhance the convergence speed and overall effectiveness of the optimization process. This strategic selection process contributes to the algorithm's ability to adapt and refine its solutions, ultimately leading to improved optimization outcomes. The Pseudo Code of the MBER algorithm is demonstrated in Algorithm 2.

Initialize individuals $\vec{d}_i (i = 1,2,\dots,s)$ with size s, maximum iterations Max_{iter}, fitness function F_n
Initialize BER variables
Set t = 1
Compute fitness F_n, for every individual in $\vec{d}_i$
Locate optimal solution d^*
while $t \leq \mathrm{Max}_{iter}$ do
    For every solution in exploration group do
        Progressing towards the optimal solution
        $\mathrm{x} = h \frac{\cos(\mathrm{r})}{1-\cos(\mathrm{r})}$
        $\vec{S} = \overrightarrow{x_1}(d(t) - 1)$
        $\vec{D}(t+1) = \vec{D}(t) + \vec{S}(2\overrightarrow{x_2} - 1)$
    end for
    For every solution in exploitation group do
        Choose of the optimal solution
        $\vec{S} = \overrightarrow{x_2}\left(\vec{L}(t) - \vec{D}(t)\right)$
        $\overrightarrow{D_1}(t+1) = \left(\frac{x_1}{x_1+x_2}\right)\left(\vec{D}(t) + \vec{S}\right)$
        Investigating the space around optimal solution
        $\vec{k} = 1 + \frac{2 \times t^2}{\mathrm{Max}_{\mathrm{iter}}}$
        $\overrightarrow{D_2}(t+1) = \left(\frac{x_2}{x_1+x_2}\right)\left(\overrightarrow{D^*}(t) + \vec{k}\right)$
        Compare $\overrightarrow{D_1}(t+1)$ and $\overrightarrow{D_2}(t+1)$ and choose the optimal solution $\overrightarrow{D^*}$
        If the best fitness remained unchanged in the last two rounds, then
            mutate solution
            $\vec{D}(t+1) = \vec{k} * z^2 - h \frac{\cos(\mathrm{r})}{1-\cos(\mathrm{r})}$
        end if
    end for
    Modify fitness F_n for each $\vec{D}$
end while
Return optimal solution $\overrightarrow{D^*}$

Algorithm 2. Pseudo Code of the MBER

Fitness function

In order to achieve peak performance from our classification model, it's crucial that we optimize it to the fullest extent. Only then can we ensure that we're making the most of our resources and achieving accurate and reliable results, the focal point of our research lies in the introduction of an fitness mathematically expressed as

$$F_n = \alpha \; Error\,(P) + \beta \,\frac{|S|}{|A|} \tag{11}$$

This meticulously crafted equation encapsulates the essence of our optimization strategy, where the parameters α and β hold pivotal roles in shaping the fine balance between two fundamental aspects of model performance. The significance of the chosen features in the population is reflected by the values of $\alpha \in [0,1]$, $\beta = 1 - \alpha$.

Our goal is to strategically optimize these parameters, guiding the optimization process to identify an optimal set of inputs (P) and selected features ($|S|$) that strike a balance between high accuracy and minimal feature complexity. This fitness function serves as the linchpin of our optimization strategy, driving the pursuit of an efficient and effective classification model.

The combination of a well-curated dataset, a powerful machine learning algorithm, and an innovative optimization approach contributes to the creation of a robust predictive analytics framework. The demonstrated success of the Random Forest model, optimized using the BER algorithm, reinforces the effectiveness of this methodology in understanding and predicting students' academic performance.

Evaluation metrics

When evaluating the efficiency of our models, we employed a range of performance criteria tailored for classification tasks. These included Accuracy, Precision (PPV), Negative Predictive Value (NPV), FScore, Sensitivity and Specificity, specifically for the process of EEG classification[42]. By leveraging these metrics, we could thoroughly assess and quantify the model's effectiveness across various aspects, encompassing its capability to accurately discern True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN)[43]. These classification performance metrics are expressed through the below Equations:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{12}$$

$$\mathrm{Precision}\,(\mathrm{PPV}) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$\text{NegativePredictiveValue (NPV)} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{14}$$

$$\text{FScore} = \frac{2^*\text{TP}}{2^*\text{TP} + \text{FP} + \text{FN}} \tag{15}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

$$\text{Specitivity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{17}$$

Experimental results

This section delineates the assessment of the proposed algorithm across several scenarios in experimental settings. The tests utilized conventional mathematical functions as benchmarks to determine their minimal values within a specific search space. These functions are commonly employed in literature to assess the performance of optimization techniques, and multiple optimization approaches are accessible in the literature. This work conducted a comparative analysis between the proposed algorithm, bMBER, and five established optimization algorithms in order to showcase its superiority and efficacy. The algorithms MBER, BER, PSO, WAO, GWO, and GA were selected because of their widespread recognition and practical value.

The algorithm was implemented and tested in a Jupyter Notebook environment, which provides an interactive platform for running and analyzing the code. The implementation was done entirely in Python, leveraging its extensive libraries for optimization, machine learning, and data analysis. Python was chosen due to its flexibility, ease of use, and the availability of robust libraries such as NumPy, SciPy, and scikit-learn, which were integral to the development and evaluation of the proposed algorithm.

Dataset description

The data provided is derived from a continuous EEG reading using the Emotiv EEG Neuroheadset. The reading continued for a length of 117 seconds. The dataset consisting of 14 columns and 14,981 rows. During the EEG measurement, a camera was used to identify the eye state. After examining the video frames, the eye state was manually inserted to the file. A value of '1' indicates that the eyes are closed, whereas a value of '0' indicates that the eyes are opened. The values are organized chronologically, starting at the top of the data with the oldest measured value. The dataset can be downloaded from the Kaggle platform (https://www.kaggle.com/datasets/robikscube/eye-state-classification-eeg-dataset), where comprehensive instructions on dataset utilization and interpretation are also provided. Figure 4 illustrates the scatter plot of the 14 columns and eye state variables.

The correlation matrix, a useful tool for statistical analysis to examine the relationship between factors in the dataset, is shown in Fig. 5. Often, it produces a matrix that shows all the variables' pairwise correlations. The relative intensity and direction of the correlations are shown by the correlation coefficients, which have a range of − 1 to + 1. The correlation matrix seeks to identify elements that show positive or negative correlations in order to analyze the relationships, patterns, and potential predictors in the data. Predictive modeling benefits

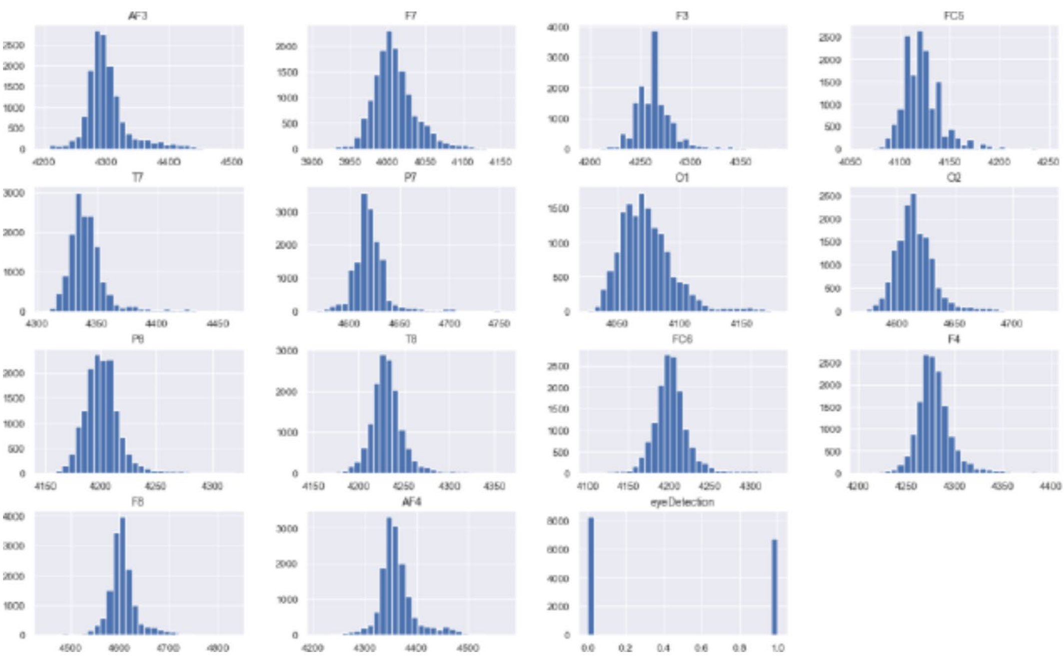

Fig. 4. Scatter plot of the input features and output one.

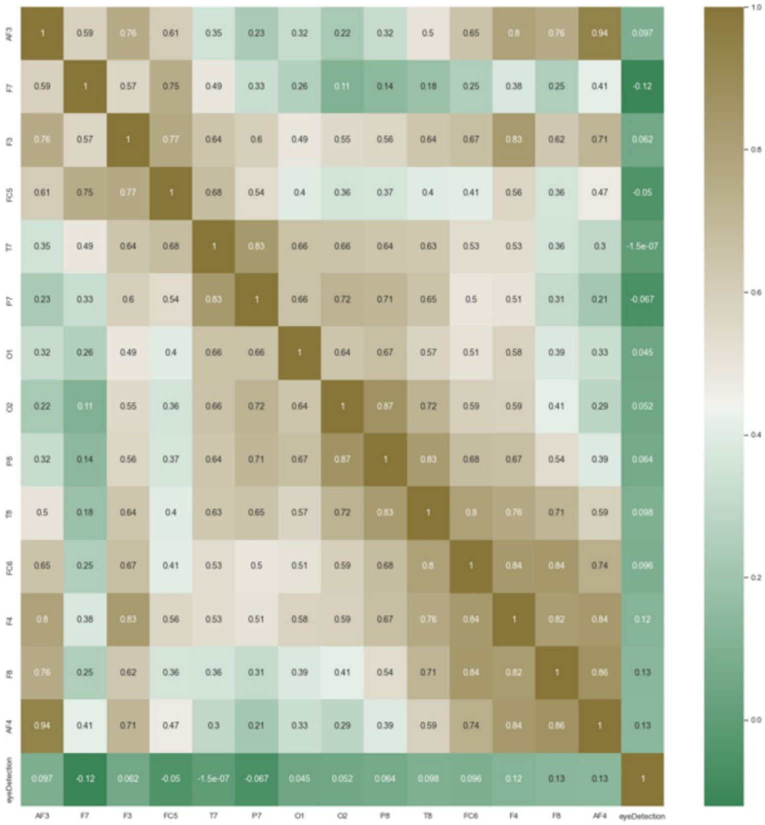

Fig. 5. A correlation matrix between dataset features.

Evaluation metrics	bMBER	bBER	bPSO	bWAO	bGWO	bGA
Average error	0.61698	0.63418	0.66798	0.66778	0.65428	0.66638
Average select size	0.56978	0.76978	0.76978	0.93318	0.69258	0.80428
Average fitness	0.68018	0.69638	0.69478	0.70258	0.70248	0.74668
Best fitness	0.58198	0.61668	0.67508	0.66668	0.68028	0.66538
Worst fitness	0.68048	0.68358	0.74278	0.74278	0.75648	0.76298
Standard deviation fitness	0.50248	0.50718	0.50658	0.50878	0.50778	0.54338

Table 2. Comparison of the proposed (bMBER) with other competing feature selection algorithms.

greatly from this concepts as it assists in selecting relevant features, reducing dimensionality, and identifying multicollinearity issues[44–46].

Feature selection results

This study used feature selection techniques to implement six optimization strategies in binary format—namely, the Modified BER, Al-Biruni earth radius (BER), Particle Swarm Optimizer (PSO), Whale Optimization Algorithm (WOA), Genetic Algorithm (GA), and Gray Wolf Optimizer (GWO). Table 2 displays an assessment of the outcomes attained by various feature selection algorithms. The table clearly demonstrates that the outcomes obtained by the proposed bMBER algorithm surpass those acquired by other algorithms for feature selection.

The average error of proposed bMBER alongside five alternative feature selection algorithms is shown in Fig. 6. The proposed algorithm produced the lowest average error, demonstrating its resilience. as the figure presents.

Figure 7 displays the bMBER algorithm's convergence curve compared to other binary optimization algorithms. The figure shows the algorithm's exploitation potential and ability to prevent potential local optima from arising throughout the optimization process.

To determine if the proposed bMBER is statistically superior, p-values were calculated by contrasting the outcomes of every pair of algorithms. This study used Wilcoxon's rank-sum test to conduct the investigation. The Wilcoxon rank-sum test results are shown in Table 3. The statistical superiority of the proposed algorithm is demonstrated by its lower p-value ($p < 0.005$) when compared to the other optimization algorithms. To find out if the proposed bMBER algorithm and the other algorithms differed statistically significantly from one another,

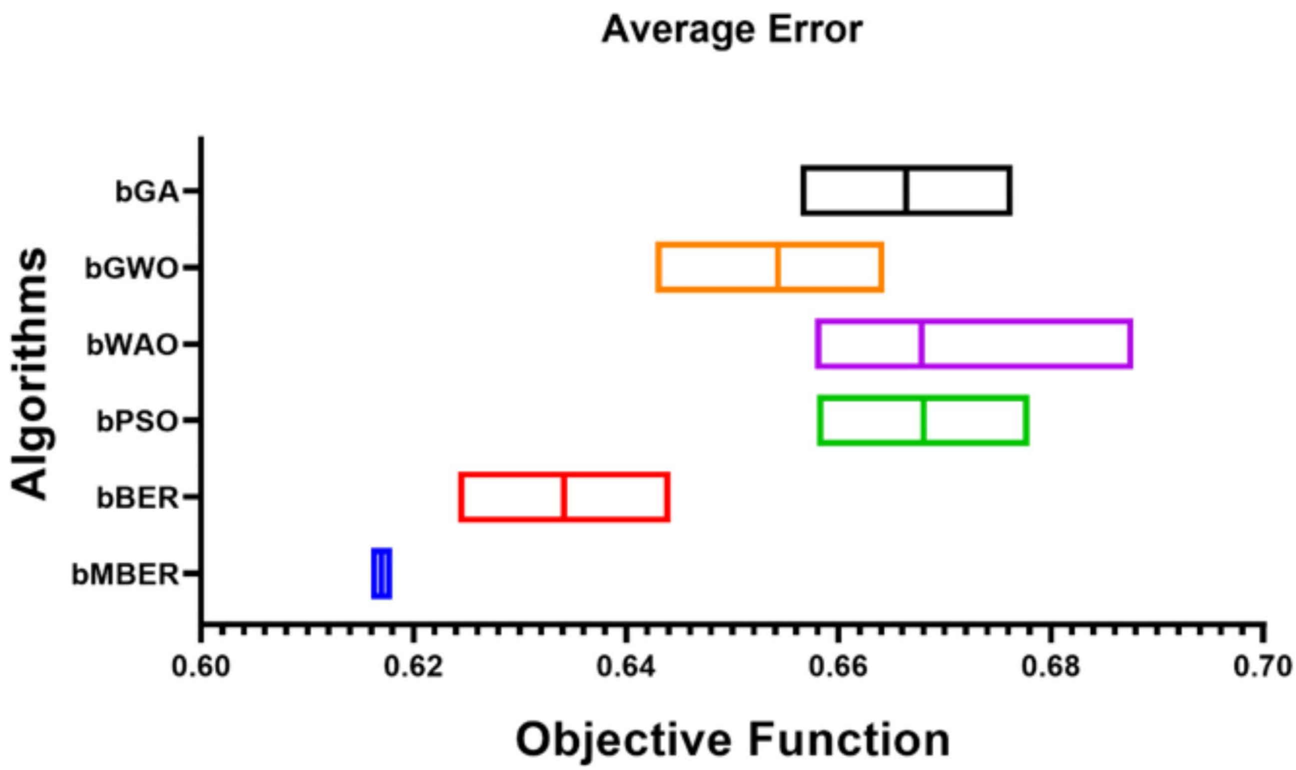

Fig. 6. The average error of the outcomes obtained using bMBER.

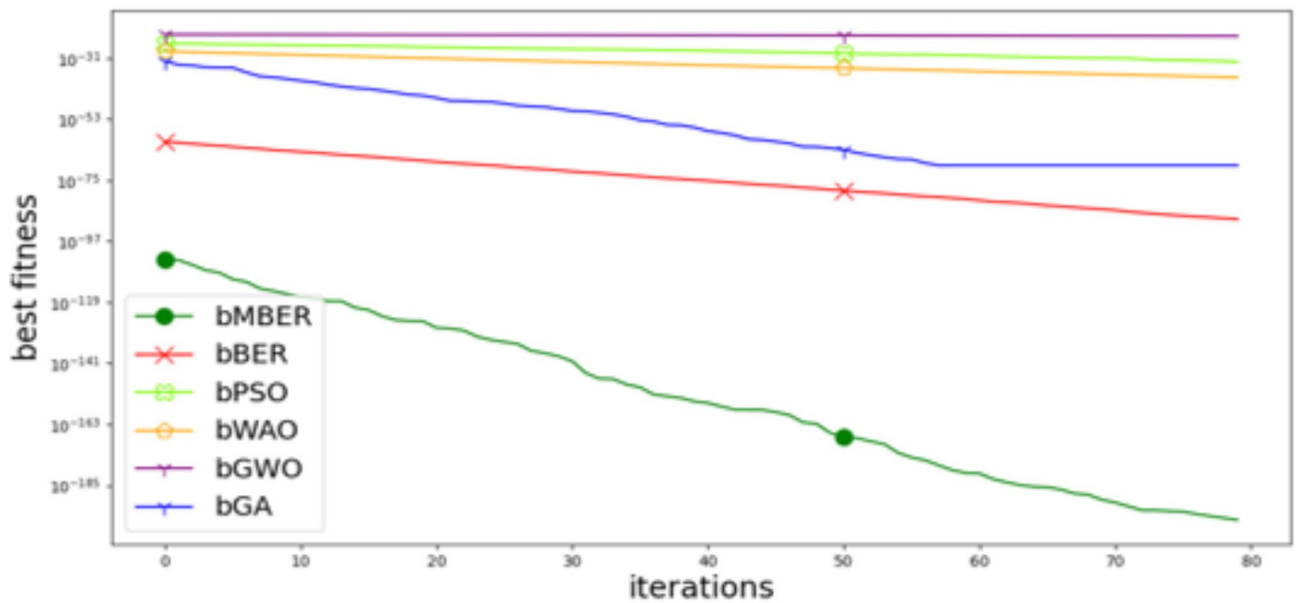

Fig. 7. The convergence curve of the bMBER algorithm in comparison to other methods.

a one-way analysis of variance (ANOVA) test was performed. Table 4 presents the results of the ANOVA test. The results shown in these tables confirm the proposed feature selection algorithm's superiority, significance, and effectiveness.

In Fig. 8, the plots demonstrate the outcomes obtained from the proposed feature selection algorithm. The figure utilizes residual plots, quartile–quartile (QQ), homoscedasticity, and heatmap to demonstrate the robustness and efficacy of the proposed algorithm. The values displayed in the QQ plot exhibit a close approximation to a linear trend, indicating the robustness of the chosen features in accurately identifying the eye state. Furthermore, the emphasis on outcomes is further highlighted by the recorded results in the residual

	bMBER	bBER	bPSO	bWAO	bGWO	bGA
Theoretical median	0	0	0	0	0	0
Actual median	0.617	0.6342	0.668	0.6678	0.6543	0.6664
Number of values	10	10	10	10	10	10
Wilcoxon signed rank test						
Sum of signed ranks (W)	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0
Sum of positive ranks	55	55	55	55	55	55
P value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002
P value summary	**	**	**	**	**	**
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?						
Discrepancy	0.617	0.6342	0.668	0.6678	0.6543	0.6664

Table 3. Wilcoxon signed-rank test for assessing the proposed (bMBER) compared to other feature selection algorithms.

ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
Treatment (between columns)	0.02297	5	0.004594	F (5, 54) = 187.5	*P* < 0.0001
Residual (within columns)	0.001323	54	2.45E-05		
Total	0.02429	59			

Table 4. The analysis-of-variance (ANOVA) test for evaluating the proposed algorithm, bMBER.

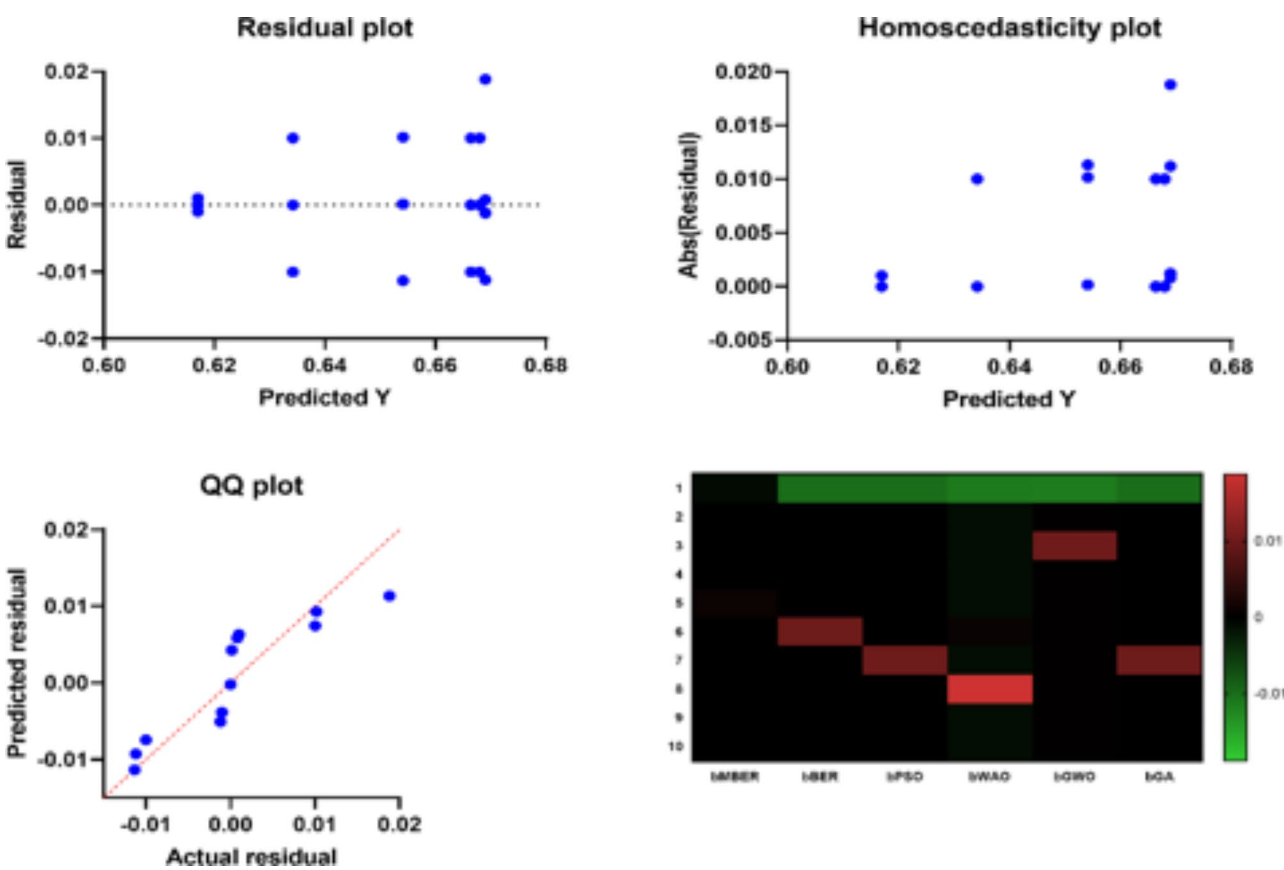

Fig. 8. Analysis plots of the obtained outcomes based on bMBER algorithm.

Models	PPV	NPV	FScore	Accuracy	Sensitivity	Specificity
SGDClassifier	0.585275	0.729289	0.8	0.6046	0.967308	0.160214
GaussianNB	0.648238	0.664424	0.583571	0.621	0.681438	0.546951
SVC	0.608502	0.739124	0.781671	0.6342	0.941155	0.258122
LogisticRegression	0.646066	0.706096	0.637552	0.6432	0.778424	0.477526
DecisionTreeClassifier	0.855364	0.851933	0.816219	0.8376	0.848529	0.82421
RandomForestClassifier	0.917393	0.936322	0.943219	0.9284	0.956048	0.894526
KNeighborsClassifier	0.959425	0.964969	0.963431	0.9612	0.970578	0.949711

Table 5. Different classifiers for eye state classification.

	MBER	BER	PSO	WAO	GWO	GA
Number of values	10	10	10	10	10	10
Minimum	0.9988	0.9798	0.9699	0.9653	0.974	0.9711
25% Percentile	0.9996	0.9877	0.9788	0.9744	0.974	0.9711
75% Percentile	0.9998	0.988	0.9788	0.9753	0.9745	0.9716
Median	0.9998	0.9877	0.9788	0.9753	0.974	0.9711
Range	0.001	0.00993	0.01	0.01	0.004	0.004
Maximum	0.9998	0.9897	0.9799	0.9753	0.978	0.9751
10% Percentile	0.9988	0.9806	0.9708	0.9659	0.974	0.9711
90% Percentile	0.9998	0.9896	0.9798	0.9753	0.9778	0.9749
95% CI of median						
Actual confidence level	97.85%	97.85%	97.85%	97.85%	97.85%	97.85%
Lower confidence limit	0.9988	0.9877	0.9788	0.9715	0.974	0.9711
Upper confidence limit	0.9998	0.9887	0.9788	0.9753	0.976	0.9731
Std. deviation	0.000422	0.002697	0.002879	0.003253	0.00135	0.00135
Mean	0.9996	0.9872	0.978	0.9739	0.9746	0.9717
Std. error of mean	0.000133	0.000853	0.00091	0.001029	0.000427	0.000427

Table 6. The descriptive statistics of the proposed optimizing MBER algorithm compared to other algorithms.

plots and homoscedasticity. The superiority of bMBER is verified by the heatmap, as it attained the most optimal outcomes in comparison to other feature selection algorithms.

Classification results

The classification results for different ML models are listed in Table 5. The proposed models show encouraging outcomes of DT, KNN, and RF than SGD, GNB, SVC, and LR. The KNN model achieved the highest values of PPV (0.959425), NPV (0.964969), FScore (0.963431), accuracy (0.9612), Sensitivity (0.970578) and Specificity (0.949711). Thus, KNN serves as a fitness function and is optimized by the utilization of Modified Al-Biruni earth radius (MBER).

The statistical results presented in Table 6 compare the efficiency of the MBER algorithm to five other optimizers (PSO, BER, WAO, GA, and GWO). The table demonstrates that the MBER algorithm had superior performance compared to the other five optimizers, owing to the utilization of two distinct exploitation strategies in every cycle. The initial strategy is moving towards the most optimal solution discovered thus far, but the subsequent strategy involves actively seeking superior solutions in proximity. By employing these strategies, the MBER algorithm can optimize its use of the search space and attain outstanding outcomes. To accomplish efficient exploitation, achieving a balance between exploring and exploiting the search space is essential. Furthermore, it's critical to start the process of exploitation early in every round and gradually expand the exploitation group's size.

Figure 9 exhibits the convergence curves by various optimization algorithms. The figure demonstrates that the suggested MBER algorithm surpasses the performance of the other methods by a substantial margin. Furthermore, the accuracy of proposed optimization algorithm MBER is evaluated compared to other models. This is demonstrated by the results displayed in Fig. 10, which presents the accuracy, and Fig. 11, which shows a histogram of the accuracy.

Wilcoxon's rank-sum and ANOVA tests are used to evaluate the statistical differences between the proposed algorithm and competing ones. Table 7 shows the results of the ANOVA. To determine if the results of the algorithms differ significantly, the Wilcoxon's rank-sum test is used, as indicated in Table 8. Significant statistical superiority is shown by a p-value that is less than 0.05. The outcomes show the statistical significance of the algorithm and show that the MBER performs better.

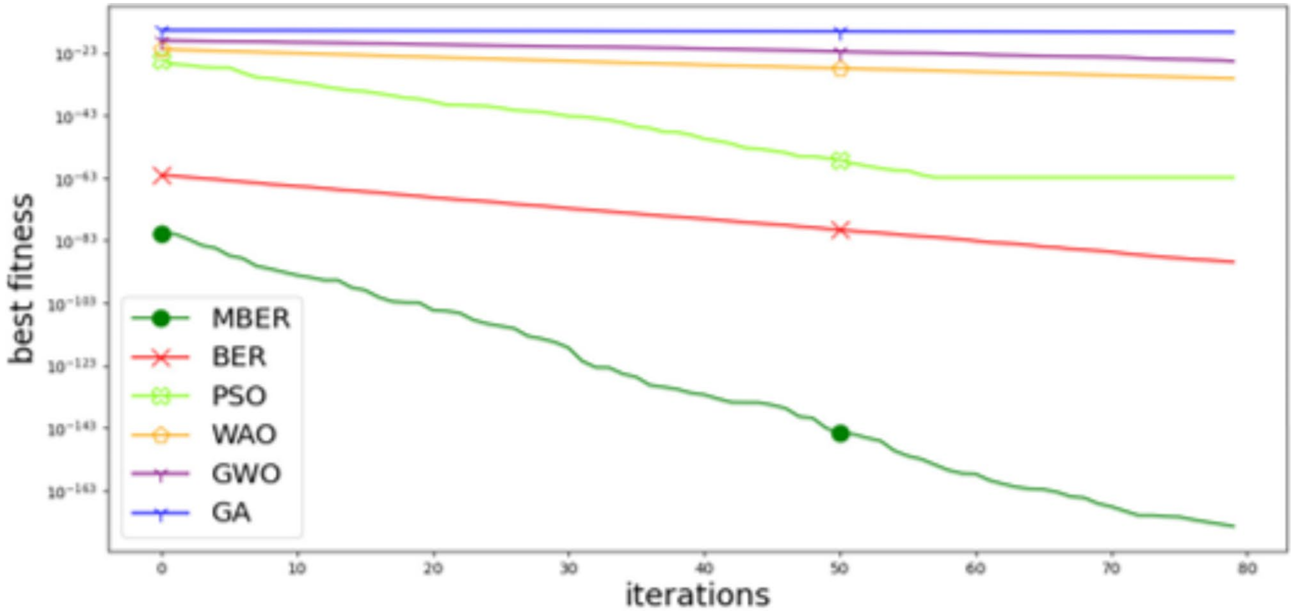

Fig. 9. The convergence curves for MBER, BER, PSO, WAO, GWO, and GA.

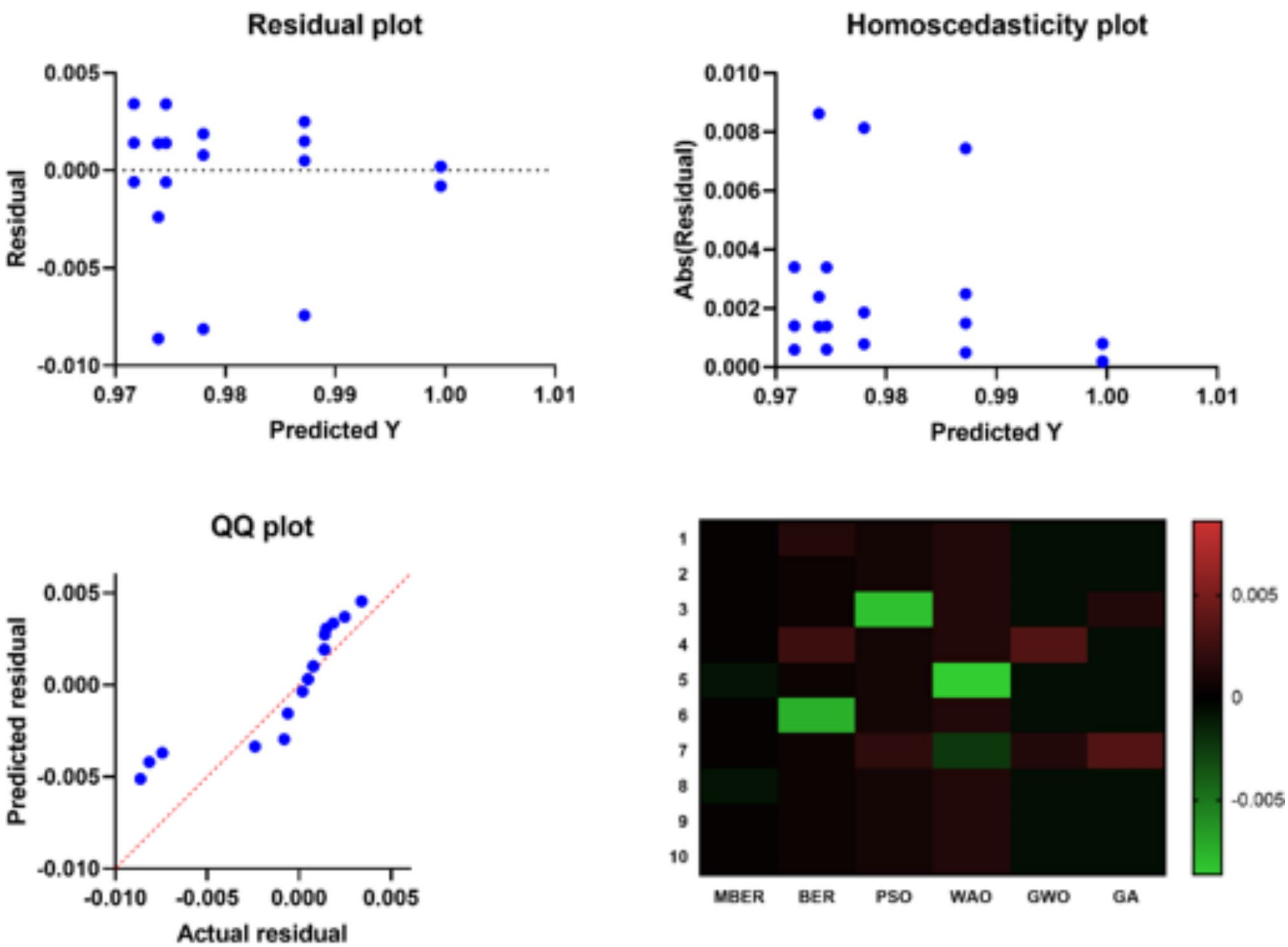

Fig. 10. Evaluating the accuracy of MBER compared to other algorithms based on the objective function.

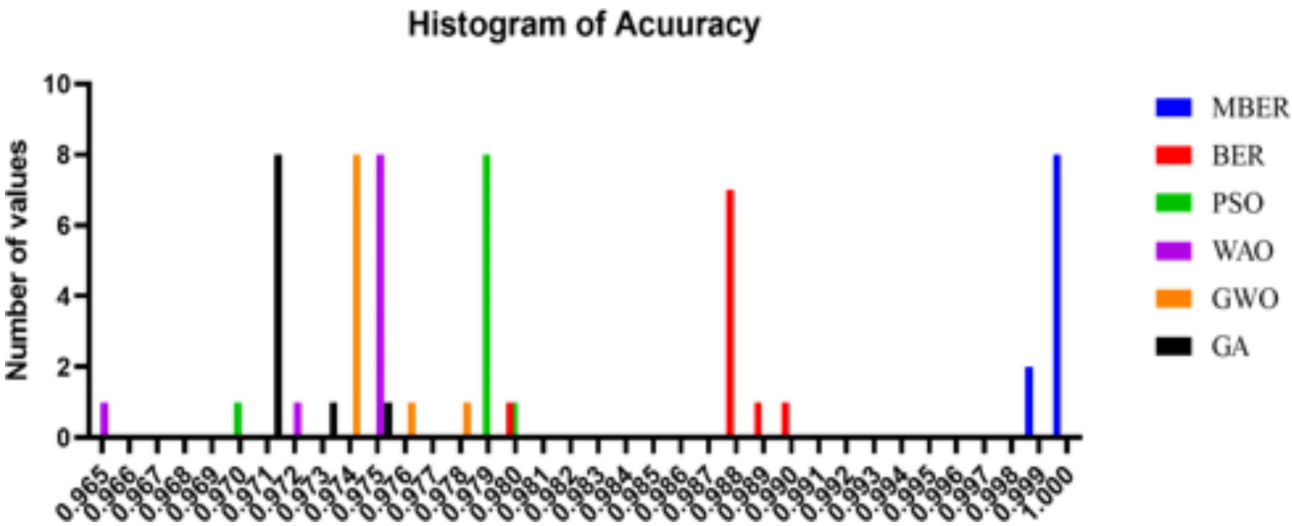

Fig. 11. Histograms of the accuracy outcomes obtained by MBER, BER, PSO, WAO, GWO, and GA algorithms.

ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
Treatment (between columns)	0.005708	5	0.001142	F (5, 54) = 228.6	*P* < 0.0001
Residual (within columns)	0.00027	54	4.99E−06		
Total	0.005977	59			

Table 7. ANOVA results of the proposed and competing models for the eye state classification.

	MBER	BER	PSO	WAO	GWO	GA
Theoretical median	0	0	0	0	0	0
Actual median	0.9998	0.9877	0.9788	0.9753	0.974	0.9711
Number of values	10	10	10	10	10	10
Wilcoxon signed rank test						
Sum of signed ranks (W)	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0
Sum of positive ranks	55	55	55	55	55	55
P value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002
P value summary	**	**	**	**	**	**
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?						
Discrepancy	0.9998	0.9877	0.9788	0.9753	0.974	0.9711

Table 8. The wilcoxon signed-rank test outcomes of MBER and competing ones for the eye state classification.

Figure 12 displays the residual plot, heteroscedasticity plot, QQ plot, and heat map for this situation. The QQ plot indicates a close alignment of the dots with the line, indicating a linear correlation between the expected and actual residuals. This verifies the effectiveness of the proposed MBER method for classifying eye states.

The analysis plots depicted in Fig. 12 illustrate the efficacy of the proposed MBER algorithm in resolving the optimization issues employed in this paper.

Conclusions

This study proposes the MBER algorithm, which aims to enhance the accuracy of classifying eye states as either open (0) or closed (1). The assessment of the proposed algorithm was conducted using an easily accessible EEG dataset that underwent preprocessing techniques such as scaling, normalization, and removal of null values. The binary format of the MBER algorithm is designed to choose the optimal features that can enhance classification accuracy. The proposed algorithms were assessed using sets of assessment criteria. The statistical investigation

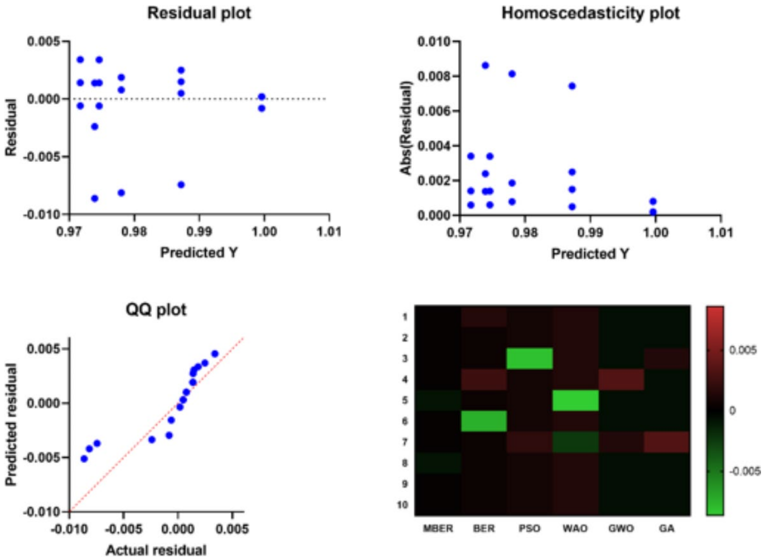

Fig. 12. Analysis plots for MBER, BER, PSO, WAO, GWO, and GA algorithms.

utilized the ANOVA and Wilcoxon signed-rank test to determine the relevance and efficacy of the proposed algorithm in comparison to five other algorithms (BER, PSO, WAO, GWO, and GA). Additionally, a series of visual depictions of the outcomes were produced to verify the strength and efficacy of the proposed algorithm. In general, the experimental and statistical findings demonstrate the superiority of the proposed algorithm in comparison to other competing optimization algorithms for the eye state classification. The next objectives of this study involve evaluating the proposed algorithm using extensive datasets and other optimization problems to ascertain its strengths and weaknesses definitively.

Future work and limitations

Future work regarding this study requires improving the stability, versatility, and efficacy of the modified Al-Biruni Earth Radius (MBER) algorithm. One of them is the application of more extensive and more diverse datasets. Incorporating other datasets derived from different EEG instruments and diverse experimental scenarios will advance the assessment of the algorithm's versatility. There are also algorithmic optimizations. Algo improvements are also significant. More improvement can be made by embracing MBER with other advanced optimization methods to come up with best-of-breed solutions." Implementing deep learning frameworks could also be integrated into MBER to enhance results. Application in real-time is advisable if the proposed algorithm is to be implemented. This paper aims to present a modified version of the weighted center MBER algorithm and the factors that should be considered to improve the algorithm's efficiency at lesser latency and faster processing time, which will be significant in real-world BCI applications. Attaining the real-time processing capability will make the algorithm useable in assistive technologies, which will immensely benefit disabled persons. Different extensions of research using the MBER algorithm in other fields can also be considered for future work. The algorithm proposed here can be applied to other domains besides the EEG-based eye state classification results: emotion recognition, cognitive workload assessment, and neurofeedback. Another aspect that remains to be researched is the customization of the MBER algorithm with certain parameters specific to the user. If models capable of learning and adjusting to other characteristic EEG patterns of different people are created, the specificity of the classification system will be significantly improved.

However, the study has limitations. The future potential of the results is constructive, but the weaknesses of the study can be criticized. The first and probably foremost limitation is related to the available data used as the basis for analysis and modeling. Problems associated with the study included the limitation that only one specific EEG dataset was used in the study-meaning that there could be generalizability issues if the same amount of variability were not included across different participants or indifferent experimental conditions. Therefore, the results cannot be applied to other settings without further verification. Furthermore, the analytical formulation of the proposed MBER algorithm, it incurs a higher computational complexity than the others. The steps mentioned above of feature selection and optimization are computationally expensive, and this brings forth a problem when applied to real-time applications or systems with low computational capabilities. Solving this complexity while making the required changes to improve the algorithm's efficiency will be very important for the future application of the algorithm. There are also inherent issues with noise and artifacts in the EEG data. Despite pre-processing techniques, the data of EEG is always vulnerable to noise and artifacts caused by muscle movements and other electronics devices. However, in the proposed method, the mentioned issues are reduced, but eliminating them in total causes some loss of classification. Another limitation is model interpretability. Like most machine learning algorithms, it isn't easy to understand how the MBER algorithm arrives at a particular decision. Establishing additional ways of explaining how the algorithm interacts and categorizes the EEG signals would be helpful for both the research and the customer.

Finally, there is a need to cross-validate among different populations. The current study's participant sample fails to sample the whole population of human variations in terms of the EEG results. Thus, future studies should focus on the algorithm's generalization and testing of the intervention on more extensive and different samples of participants, including different ages, genders and diseases, to consider the intervention general and efficient. In conclusion, the study demonstrated that the MBER algorithm has an excellent potential for EEG classification and BCI applications. Therefore, to utilize the full potential of the MBER algorithm, future work should address the related directions and limitations. To enhance this novel MBER algorithm, more improvements such as using all the datasets at once, varying and improving the algorithm used, implementing it in real-time, adapting the methodology across different disciplines, fine-tuning it depending on the users' specifications, and verifying it thoroughly can be done further.

Data availability

https://www.kaggle.com/datasets/robikscube/eye-state-classification-eeg-dataset.

Received: 1 May 2024; Accepted: 26 September 2024
Published online: 18 October 2024

References

1. Berger, H. *Ueber das Elektrenkephalogramm des Menschen* (Journal für Psychologie und Neurologie, 1930).
2. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.***16**(3), 031001 (2019).
3. Siuly, S., Li, Y. & Zhang, Y. EEG signal analysis and classification. *IEEE Trans. Neural Syst. Rehabili. Eng.***11**, 141–144 (2016).
4. Eid, M. M. et al. Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of Monkeypox cases. *Mathematics*. **10**, 3845 (2022).
5. Alkanhel, R. I. et al. Enhancing Wireless Sensor Network Efficiency through Al-Biruni Earth Radius optimization. *Comput. Mater. Continua*. **79**(3). (2024).
6. Smith, S. J. EEG in the diagnosis, classification, and management of patients with epilepsy. *J. Neurol. Neurosurg. Psychiatry*. **76** (suppl 2), ii2–ii7 (2005).
7. Lotte, F. et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15(3), 031005 (2018).
8. Hernandez-Pavon, J. C. et al. TMS combined with EEG: recommendations and open issues for data collection and analysis. *Brain Stimul.***16** (2), 567–593 (2023).
9. Khafaga, D. et al. Meta-heuristics for feature selection and classification in diagnostic breast Cancer. *CMC*. **73**, 749–765 (2022).
10. Rizk, F. H., Elshabrawy, M., Sameh, B., Mohamed, K. & Zaki, A. M. Optimizing Student Performance Prediction using Binary Waterwheel Plant Algorithm for Feature Selection and Machine Learning. *Full Length Article*. **7**(1), 9–19 (2024).
11. Wang, X. W., Nie, D. & Lu, B. L. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*. **129**, 94–106 (2014).
12. Alkhammash, E. H., Hadjouni, M. & Elshewey, A. M. A hybrid ensemble stacking model for gender voice recognition approach. *Electronics***11**(11), 1750 (2022).
13. Gandhi, T., Panigrahi, B. K. & Anand, S. A comparative study of wavelet families for EEG signal classification. *Neurocomputing*. **74** (17), 3051–3057 (2011).
14. Subasi, A. & Gursoy, M. I. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.***37** (12), 8659–8666 (2010).
15. Samee, N. et al. Metaheuristic optimization through Deep Learning classification of COVID-19 in chest X-Ray images. *CMC*. **73**, 4193–4210 (2022).
16. Shams, M. Y., El-kenawy, E. S. M., Ibrahim, A. & Elshewey, A. M. A hybrid dipper throated optimization algorithm and particle swarm optimization (DTPSO) model for hepatocellular carcinoma (HCC) prediction. *Biomed. Signal Process. Control* 85, 104908 (2023).
17. Gaber, K. S. et al. Optimizing marketing strategies: integration of Al-Biruni Earth Radius Algorithm for Feature Selection and Pipeline Regression Model. *J. Artif. Intell. Eng. Pract.***1** (1), 18–33 (2024).
18. Alkhammash, E. H. et al. Application of Machine Learning to Predict COVID-19 Spread via an Optimized BPSO Model. *Biomimetics***8**(6), 457 (2023).
19. Alharbi, A. H. et al. Forecasting of energy efficiency in buildings using multilayer perceptron regressor with waterwheel plant algorithm hyperparameter. *Front. Energy Res.***12**, 1393794 (2024).
20. El-Kenawy, E. S. M. et al. Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification. *IEEE Access.***9**, 36019–36037 (2021).
21. El-Kenawy, E. S. M., Eid, M. M., Saber, M. & Ibrahim, A. MbGWO-SFS: modified binary grey wolf optimizer based on stochastic fractal search for feature selection. *IEEE Access.***8**, 107635–107649 (2020).
22. Abdelhamid, A. A. et al. Robust speech emotion recognition using CNN + LSTM based on stochastic fractal search optimization algorithm. *IEEE Access.***10**, 49265–49284 (2022).
23. Peng, Y. et al. Automatic measurement of fetal anterior neck lower jaw angle in nuchal translucency scans. *Sci. Rep.***14**, 5351 (2024).
24. Peng, Y., Zeng, S., Luo, Y., Yan, L. & Yao, L. Attention mechanism optimized neural network for automatic measurement of fetal anterior-neck-lower-jaw angle in nuchal translucency tests. *Multimed. Tools Appl.***83**, 15629–15648 (2024).
25. Fu, J. et al. Low-light image enhancement base on brightness attention mechanism generative adversarial networks. *Multimed. Tools Appl.***83**, 10341–10365 (2024).
26. Polat, K. & Güneş, S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Appl. Math. Comput.***187** (2), 1017–1026 (2007).
27. Chandaka, S., Chatterjee, A. & Munshi, S. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Syst. Appl.***36** (2), 1329–1336 (2009).
28. Farwell, L. A. & Donchin, E. The truth will out: interrogative polygraphy (lie detection) with event-related brain potentials. *Psychophysiology*. **28** (5), 531–547 (1991).
29. Haider, S. K., Daud, M. I., Jiang, A. & Khan, Z. Evaluation of p300 based lie detection algorithm. *Electr. Electron. Eng.***2017** (3), 69–76 (2017).
30. Simbolon, A. I., Turnip, A., Hutahaean, J., Siagian, Y. & Irawati, N. An experiment of lie detection based EEG-P300 classified by SVM algorithm. In *2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)* 68–71 (IEEE, 2015).
31. Bablani, A., Edla, D. R. & Dodia, S. Classification of EEG data using k-nearest neighbor approach for concealed information test. *Proc. Comput. Sci.***143**, 242–249 (2018).

32. Khare, V., Santhosh, J. & Anand, S. Classification of EEG signals based on neural network to discriminate five mental states. In *Proceedings of SPIT-IEEE Colloquium and International Conference, Mumbai, India*, vol. 1, 24–26 (2007).
33. Bayram, K. S., Kızrak, M. A. & Bolat, B. Classification of EEG signals by using support vector machines. In *2013 IEEE INISTA* 1–3 (IEEE, 2013).
34. Rajaguru, H. & Prabhakar, S. K. Logistic regression Gaussian mixture model and softmax discriminant classifier for epilepsy classification from EEG signals. In *2017 International Conference on Computing Methodologies and Communication (ICCMC)* 985–988 (IEEE, 2017).
35. Guerrero, M. C., Parada, J. S. & Espitia, H. E. EEG signal analysis using classification techniques: logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon*7(6) (2021).
36. Tarek, Z. et al. Soil erosion status prediction using a novel random forest model optimized by random search method. *Sustainability*. **15** (9), 7114 (2023).
37. Alzakari, S. A., Alhussan, A. A., Qenawy, A. S. T., Elshewey, A. M. & Eed, M. *An Enhanced Long Short-Term Memory Recurrent Neural Network Deep Learning Model for Potato Price Prediction* 1–19 (Potato Research, 2024).
38. Zaki, A. M., Abdelhamid, A. A., Ibrahim, A., Eid, M. M. & El-Kenawy, E. S. M. Enhancing K-nearest neighbors algorithm in wireless sensor networks through stochastic fractal search and particle swarm optimization. *J. Cybersecur. Inf. Manag.*13(1) (2024).
39. El-kenawy, E. S. M. et al. Al-Biruni based optimization of rainfall forecasting in Ethiopia. *Comput. Syst. Sci. Eng.***46** (1), 2885–2899 (2023).
40. Ibrahim, A. et al. Al-Biruni earth radius optimization based algorithm for improving prediction of hybrid solar desalination system. *Energies***16**(3), 1185 (2023).
41. Saeed, M. A. et al. A novel voting classifier for electric vehicles population at different locations using Al-Biruni earth radius optimization algorithm. *Front. Energy Res.***11**, 1221032 (2023).
42. Elshewey, A. M., Tawfeek, S. M., Alhussan, A. A., Radwan, M. & Abed, A. H. *Optimized Deep Learning for Potato Blight Detection Using the Waterwheel Plant Algorithm and Sine Cosine Algorithm* 1–25 (Potato Research, 2024).
43. Fouad, Y., Osman, A. M., Hassan, S. A., El-Bakry, H. M. & Elshewey, A. M. Adaptive visual sentiment prediction model based on event concepts and object detection techniques in social media. *Int. J. Adv. Comput. Sci. Appl.***14**(7). (2023).
44. Abdelhamid, A. A. et al. *Potato Harvesting Prediction Using an Improved ResNet-59 Model* 1–20 (Potato Research, 2024).
45. Eed, M. et al. *Potato Consumption Forecasting Based on a Hybrid Stacked Deep Learning Model* 1–25 (Potato Research, 2024).
46. Elshewey, A. M. et al. Optimizing HCV disease prediction in Egypt: The hyOPTGB framework. *Diagnostics*. **13**(22), 3439 (2023).

Acknowledgements
Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions
All authors have contributed equally.

Funding
Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-74475-5.

Correspondence and requests for materials should be addressed to A.M.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.