Journal Website

Article history:
Received 01 April 2026
Revised 17 August 2026
Accepted 24 August 2026
Initial Published 25 August 2026
Final Publication 01 November 2026

Iranian Journal of Neurodevelopmental Disorders

Volume 5, Issue 1, pp 1-16

E-ISSN: 2980-9681

Optimal Feature Selection and CNN Hyperparameter Tuning Using the Grey Wolf Optimizer for Depression Detection

Amirsalar. Ahmadi[1], Iman. Ahanian[2*], Nasrin. Amiri[2], Nader. Jafarnia Dabanloo[3], Alireza. Kashaninia[4]

[1] Department of Biomedical Engineering, ST.C., Islamic Azad University, Tehran, Iran
[2] Department of Electrical Engineering, ST.C., Islamic Azad University, Tehran, Iran
[3] Department of Biomedical Engineering, SR.C., Islamic Azad University, Tehran, Iran
[4] Department of Electrical Engineering, CT.C., Islamic Azad University, Tehran, Iran

*** Corresponding author email address**: ahanian@iau.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Ahmadi, A., Ahanian, I., Amiri, N., Jafarnia Dabanloo, N., & Kashaninia, A. (2026). Optimal Feature Selection and CNN Hyperparameter Tuning Using the Grey Wolf Optimizer for Depression Detection. *Iranian Journal of Neurodevelopmental Disorders, 5*(1), 1-16.
https://doi.org/10.61838/kman.jndd.935

A B S T R A C T

Purpose: This study aimed to develop an optimized electroencephalography-based depression-detection framework by integrating the Grey Wolf Optimizer with a convolutional neural network for simultaneous feature selection and hyperparameter tuning.

Methods and Materials: An applied analytical design was used to analyze a standard EEG-derived dataset obtained from the Figshare repository, containing observations from healthy individuals and patients with major depressive disorder. Data preprocessing included the removal of zero-variance variables, moving-average imputation of missing values, Z-score standardization, and class balancing using the Synthetic Minority Over-sampling Technique. A binary Grey Wolf Optimizer was employed to select an informative subset of EEG-derived temporal, frequency-domain, and functional-connectivity features. Candidate feature subsets were evaluated using a radial basis function support vector machine with five-fold cross-validation. The selected features were converted into two-dimensional feature maps and entered into a custom CNN. GWO was additionally used to optimize the learning rate, batch size, convolutional-filter numbers, dropout rate, training epochs, and feature-selection threshold.

Findings: The GWO-based optimization process identified a refined feature subset that preserved the majority of diagnostically relevant EEG information while removing variables that contributed limited discriminatory value. The optimized CNN comprised three convolutional layers with progressively increasing filter numbers of 32, 64, and 128. The best-performing configuration included a learning rate of 0.0001, a batch size of 32, a dropout rate of 0.50, 100 training epochs, and a feature-selection threshold of 0.50. The combined optimization procedure indicated that simultaneous feature refinement and hyperparameter tuning produced a more stable and computationally coherent framework than reliance on unoptimized feature inputs and manually selected network parameters.

Conclusion: Integrating GWO-based feature selection and CNN hyperparameter optimization provides a systematic and promising approach for EEG-based depression detection by improving input quality, reducing unnecessary computational complexity, and supporting the extraction of complex neurophysiological patterns associated with major depressive disorder.

Keywords: *Depression detection, electroencephalography, convolutional neural network, Grey Wolf Optimizer, feature selection, hyperparameter optimization, deep learning.*

1. Introduction

Depression is a complex and heterogeneous mental disorder characterized by persistent low mood, loss of interest or pleasure, cognitive impairment, disturbances in sleep and appetite, psychomotor changes, fatigue, feelings of worthlessness, and, in severe cases, suicidal ideation or behavior. Its clinical manifestations vary considerably across individuals, age groups, and life stages, which complicates timely recognition and accurate diagnosis. The burden of depression extends beyond emotional suffering and includes deterioration in occupational, educational, interpersonal, and physical functioning. Recent research has also emphasized that depression may coexist with externalizing disorders and other psychiatric conditions, thereby increasing the probability of severe outcomes, including suicidal thoughts and behaviors among vulnerable populations (Duncan et al., 2025). Depression during specific periods, such as the postpartum stage, also presents distinctive epidemiological patterns, risk factors, and diagnostic challenges, highlighting the need for reliable assessment approaches that are sensitive to variations in symptom presentation (Khamidullina et al., 2025). At the population level, changing public attitudes toward depression and the expansion of diagnostic awareness have further intensified debates about how depressive experiences should be identified, classified, and distinguished from ordinary psychological distress (Reavley et al., 2025).

Conventional depression assessment primarily relies on clinical interviews, self-report questionnaires, behavioral observations, and diagnostic criteria. Although these approaches are central to psychiatric practice, they may be influenced by subjective judgment, recall bias, social desirability, cultural differences, symptom concealment, and variation in clinicians' experience. Patients may also report similar symptoms despite having different underlying neurophysiological profiles, while some individuals may have difficulty articulating their internal experiences accurately. These limitations have encouraged the development of complementary objective methods that can support, rather than replace, clinical decision-making. At the same time, internet-delivered psychological interventions have demonstrated that digital technologies can contribute meaningfully to the prevention and management of anxiety and depression, indicating the growing role of computational systems across the mental-health care pathway (Liu et al., 2025). However, effective digital intervention depends partly on accurate identification of individuals at risk, making reliable computational screening and diagnostic-support systems increasingly important.

Artificial intelligence has created new opportunities for detecting depression from heterogeneous data sources, including clinical records, speech, facial expressions, written language, social-media activity, neuroimaging, and physiological signals. Machine-learning models can identify complex associations among variables that may be difficult to recognize through conventional statistical methods. For example, computational analysis of social-media content has been used to detect anxiety and depression through linguistic, behavioral, and interactional indicators (Ahmed et al., 2022). More detailed frameworks have also been developed to classify the severity of depression from social-media texts, demonstrating that machine learning can distinguish different levels of depressive expression rather than merely producing binary predictions (Kabir et al., 2023). Nevertheless, social-media-based detection is affected by contextual ambiguity, platform-specific language, privacy concerns, selective self-presentation, and the uncertain relationship between online expression and clinical diagnosis. Consequently, physiological data may provide a more direct basis for identifying objective neurobiological patterns associated with depression.

Machine-learning methods have been extensively applied to high-dimensional classification problems in medicine. Their capacity to process nonlinear relationships, select informative variables, and construct predictive representations has supported applications in domains such as cancer classification, where complex and heterogeneous biological data must be analyzed simultaneously (Yaqoob et al., 2023). Similar principles are applicable to psychiatric disorders, in which diagnosis may depend on subtle combinations of neurophysiological indicators rather than a single definitive biomarker. Supervised machine learning uses labeled observations to learn decision boundaries between diagnostic classes, while deep-learning models automatically extract hierarchical representations from raw or transformed data. Advances in computer vision have shown that deep neural networks, particularly convolutional architectures, can learn local and global patterns from multidimensional inputs with minimal dependence on manually engineered descriptors (Nafea et al., 2024). These capabilities make deep learning particularly relevant to depression detection when physiological signals are transformed into structured matrices, spectrograms, connectivity maps, or other image-like representations.

E-ISSN: 2980-9681

Electroencephalography is one of the most promising physiological modalities for computational depression assessment. EEG records electrical activity generated by neuronal populations through electrodes placed on the scalp. It offers high temporal resolution, relatively low acquisition cost, noninvasive recording, and greater accessibility than many neuroimaging techniques. EEG signals contain information related to cortical activation, neural synchronization, frequency-specific oscillations, hemispheric asymmetry, and functional communication among brain regions. These characteristics have supported the use of EEG in diverse research and clinical applications. For example, EEG has been used to examine cerebral responses to external stimuli and to investigate how patterns of brain activity vary under different cognitive and affective conditions (Hosseini, 2023). EEG-based neural decoding has also become an important field of machine-learning research, with systematic evidence indicating that computational methods can extract meaningful representations from complex electrophysiological recordings (Saeidi et al., 2021).

Despite its advantages, EEG is difficult to analyze because it is nonlinear, nonstationary, high-dimensional, and highly susceptible to noise. Eye movements, muscular activity, environmental interference, electrode displacement, and differences in recording conditions can distort the underlying neural signal. Moreover, clinically relevant information may be distributed across time, frequency, spatial, and connectivity domains. Therefore, EEG analysis commonly requires filtering, artifact removal, segmentation, normalization, decomposition, and feature extraction before classification. Signal decomposition methods have been used successfully in other neurological classification tasks, including the decomposition of EEG signals for neural-network-based diagnosis of epilepsy (Lotfi, 2022). Wavelet-based approaches have likewise demonstrated their ability to characterize complex electrophysiological patterns by representing localized changes across time and frequency (Runnova et al., 2021). Frameworks for detecting high-frequency EEG oscillations further illustrate the importance of adaptable signal-processing pipelines when extracting clinically meaningful patterns from electrophysiological data (Wong et al., 2021).

The effectiveness of EEG-based classification depends substantially on the way signals are represented. Traditional machine-learning approaches generally rely on manually extracted statistical, spectral, nonlinear, or connectivity features. These features may include signal energy, entropy, variance, band power, asymmetry, coherence, phase synchronization, and effective-connectivity measures. Such descriptors can provide interpretable information, but their usefulness depends on expert selection and may vary across datasets. Deep learning offers an alternative by automatically learning discriminative patterns directly from raw signals or transformed representations. Spectrograms and time–frequency maps are particularly suitable because they convert EEG activity into two-dimensional structures that can be analyzed using convolutional neural networks. A decision-support system based on EEG spectrograms and CNN classification has shown the potential of this approach for major depression detection (Loh et al., 2021). By learning local patterns in the time–frequency plane, CNNs can identify subtle oscillatory signatures that may not be captured adequately through predefined summary measures.

Convolutional neural networks are composed of convolutional filters, nonlinear activation functions, pooling operations, and fully connected layers. Convolutional layers detect local patterns and generate progressively more abstract feature maps, whereas pooling layers reduce dimensionality and improve robustness to minor variations. Through repeated convolution and pooling, the network learns hierarchical representations that can differentiate diagnostic categories. The success of CNN-based frameworks in depression classification has been demonstrated by DeprNet, which used a deep convolutional architecture to identify depression from EEG signals (Seal et al., 2021). Hybrid neural architectures have also been proposed to combine complementary modeling capabilities. DepHNN, for instance, integrated different neural components for EEG-based depression screening and demonstrated the value of combining feature-extraction strategies within a unified framework (Sharma et al., 2021). Similarly, the integration of convolutional networks with long short-term memory units has been used to analyze effective connectivity in EEG, allowing the model to learn both structural patterns and sequential dependencies associated with major depressive disorder (Saeedi et al., 2021).

Spatial information is another important component of EEG-based depression detection. The placement of electrodes reflects activity across different cortical regions, and relationships among channels may carry clinically relevant information. Models that ignore spatial organization may lose information regarding regional asymmetry and interregional coordination. Research has shown that incorporating spatial EEG information can improve the

E-ISSN: 2980-9681

classification of patients with depression by enabling the model to exploit topographical patterns in brain activity (Jiang et al., 2021). Hybrid EEG and functional near-infrared spectroscopy systems have likewise demonstrated that combining modalities and reducing deep features can support effective classification in brain–computer interface applications (Asgharzadeh Bonab & Hatemian, 2023). These findings suggest that diagnostic performance may be enhanced when computational models preserve the spatial, temporal, and spectral structure of neurophysiological data rather than treating all variables as independent numerical inputs.

Deep-learning models, however, are sensitive to architecture design and hyperparameter selection. Parameters such as the number of convolutional layers, number and size of filters, learning rate, batch size, dropout rate, pooling strategy, and number of epochs directly affect convergence, representation quality, overfitting, and generalization. In many studies, these values are chosen manually or through repeated trial and error. Such procedures are time-consuming, computationally inefficient, and dependent on researcher experience. They may also produce suboptimal configurations because the effects of hyperparameters are interdependent and the search space is nonlinear. An architecture with too few filters may fail to capture complex EEG patterns, whereas an excessively large network may memorize training observations and perform poorly on unseen data. Similarly, an inappropriate learning rate may result in unstable training or premature convergence.

The high dimensionality of EEG-derived feature sets creates an additional challenge. A dataset may contain hundreds or thousands of temporal, spectral, nonlinear, and connectivity variables, many of which may be redundant, weakly informative, or sensitive to noise. Including all available features can increase computational complexity, intensify the curse of dimensionality, and reduce the stability of the classifier. Feature selection is therefore required to identify variables that contribute meaningfully to diagnostic discrimination. Conventional feature-selection techniques may evaluate variables independently or rely on deterministic search strategies that become inefficient in large combinatorial spaces. In contrast, population-based metaheuristic algorithms can explore complex feature subsets without exhaustively evaluating every possible combination.

Metaheuristic optimization has increasingly been applied to intelligent depression-detection systems. Previous research has used computational optimization to search for effective diagnostic representations and improve classification models (Hosseinpour & Bahrami, 2023). Other studies have proposed intelligent approaches for determining depression severity through machine-learning methods (Bahrehdar, 2022), hybrid data-mining frameworks for depression diagnosis (Karimifard et al., 2022), and multi-agent learning systems for identifying depressive disorders (Ghanbarzadeh Bonab, 2022). These studies reflect a broader transition from isolated classifiers toward integrated systems in which preprocessing, feature selection, model design, and parameter optimization are treated as interconnected components.

The Grey Wolf Optimizer is a population-based metaheuristic inspired by the social hierarchy and cooperative hunting behavior of grey wolves. Candidate solutions are represented as wolves, and the best-performing solutions are designated alpha, beta, and delta. These leading candidates guide the remaining population toward promising regions of the search space. GWO balances exploration and exploitation through adaptive control parameters, enabling the algorithm to search broadly during early iterations and refine solutions as optimization progresses. Its relatively simple structure, limited number of control parameters, and ability to solve nonlinear optimization problems make it appropriate for both feature selection and neural-network tuning. In binary feature selection, each candidate solution represents a subset of variables, and a transfer function converts continuous wolf positions into inclusion or exclusion decisions. The fitness function can simultaneously reward classification accuracy and penalize large feature subsets, thereby balancing predictive performance against model complexity.

Using GWO for CNN hyperparameter tuning offers an additional advantage because the same optimization framework can search both discrete and continuous parameters. Each wolf can encode a combination of learning rate, batch size, filter counts, dropout rate, epoch number, and feature-selection threshold. Candidate configurations can then be evaluated through validation accuracy or a composite fitness measure. This approach reduces dependence on manual experimentation and allows interactions among hyperparameters to be considered directly. The integration of GWO with CNN classification is particularly relevant to EEG-based depression detection because the diagnostic signal may depend on subtle nonlinear relationships that are sensitive to both feature composition and network architecture.

E-ISSN: 2980-9681

Automated deep-learning systems have already shown promising results for depression detection from EEG time series. Deep-learning approaches can capture temporal patterns directly from EEG recordings and may outperform conventional classifiers when sufficient preprocessing and regularization are applied (Rafiei et al., 2022). However, strong performance does not eliminate methodological concerns. Small clinical datasets, class imbalance, high feature dimensionality, inconsistent preprocessing, and overfitting can limit reproducibility. Unsupervised learning in medical image analysis has demonstrated the value of discovering latent structure without relying exclusively on manually defined labels or features (Mazzei, 2021), yet supervised depression classification still requires careful control of training and evaluation procedures. Class balancing, standardized preprocessing, independent test sets, and validation-based optimization are therefore essential for producing credible estimates of model performance.

The current study addresses these challenges through an integrated framework. EEG-derived features from healthy individuals and patients with major depressive disorder are first cleaned, standardized, and balanced. A binary GWO procedure is then used to identify an informative feature subset by optimizing classification performance while limiting unnecessary variables. The selected features are arranged into a two-dimensional representation suitable for convolutional processing. GWO is subsequently used to tune influential CNN hyperparameters, including the learning rate, batch size, number of filters, dropout rate, training epochs, and selection threshold. This dual use of GWO distinguishes the framework from approaches that optimize only the classifier or only the feature set. It also aligns with the growing emphasis on hybrid intelligent diagnostic models that combine signal processing, optimization, and deep representation learning (Ghanbarzadeh Bonab, 2022; Hosseinpour & Bahrami, 2023).

Although existing studies have confirmed the potential of EEG-based CNNs, spectrogram representations, effective-connectivity analysis, spatial modeling, and hybrid neural networks for depression detection, fewer investigations have examined feature-subset optimization and CNN hyperparameter tuning within a single GWO-guided pipeline. This gap is important because feature quality and network configuration are mutually dependent: a feature subset that performs well under one architecture may be less effective under another, while an optimized architecture cannot compensate fully for noisy or redundant inputs. An integrated search strategy may therefore provide a more coherent solution than sequential manual design. Furthermore, retaining only diagnostically useful EEG information may improve computational efficiency and model stability even when the numerical reduction in dimensionality is modest.

The proposed approach is also consistent with the broader development of decision-support systems that assist clinicians by providing reproducible, data-driven assessments. Such systems should not be regarded as substitutes for psychiatric evaluation; rather, they can serve as supplementary tools that identify neurophysiological patterns, support screening, and reduce dependence on subjective assessment alone. By combining EEG analysis, optimized feature selection, and deep-learning classification, the framework seeks to provide an objective contribution to the recognition of major depressive disorder while preserving the clinical importance of comprehensive diagnosis.

Accordingly, the aim of this study was to develop and evaluate a GWO-optimized CNN framework for depression detection by simultaneously selecting an optimal subset of EEG-derived features and tuning the principal hyperparameters of the convolutional neural network.

2. Methods and Materials

2.1. Study Design and Participants

This applied analytical study developed a machine-learning framework for the automated detection of major depressive disorder using electroencephalographic signals. The proposed framework integrated a convolutional neural network with the Grey Wolf Optimizer to identify an optimal subset of EEG features and determine appropriate CNN hyperparameters. The study used a secondary, publicly available clinical EEG dataset obtained from the Figshare repository at https://figshare.com/articles/dataset/EEG_Data_New/4244171. The dataset included EEG recordings from individuals diagnosed with major depressive disorder and healthy controls under eyes-open and eyes-closed resting-state conditions. Accordingly, no participants were recruited directly by the researchers, and the analysis was conducted on previously collected and anonymized records. Both recording conditions were retained because eyes-open and eyes-closed EEG signals may contain complementary temporal and frequency-domain information relevant to the differentiation of depressed and healthy individuals. The principal unit of analysis was the preprocessed EEG segment

E-ISSN: 2980-9681

generated from each recording. To avoid information leakage, segments originating from the same individual were assigned to only one data partition whenever participant identifiers were available, ensuring that highly similar segments from a single person were not distributed across both training and testing sets.

The methodological workflow consisted of EEG data acquisition, signal preprocessing, signal representation, data augmentation, feature selection, CNN hyperparameter optimization, model training, and performance evaluation. Because clinical EEG datasets commonly contain a limited number of observations, a generative adversarial network was incorporated into the training pipeline to generate synthetic EEG samples and improve the diversity of the training data. The GAN comprised a generator and a discriminator trained through an adversarial process. The generator learned to produce artificial EEG patterns that resembled the distribution and structure of genuine recordings, whereas the discriminator learned to distinguish authentic signals from generated samples. Through iterative competition, the generator progressively learned the temporal and frequency characteristics of the source data. Synthetic observations were used exclusively to augment the training partition and were not included in the validation or test partitions. This restriction was necessary to ensure that model evaluation reflected performance on genuine unseen clinical recordings rather than similarity to artificially generated data.

The EEG recordings were preprocessed before feature extraction and model development because raw EEG signals are nonstationary and susceptible to physiological and environmental artifacts. The preprocessing procedure included band-pass filtering to attenuate very low-frequency drift and high-frequency interference, normalization to reduce differences in signal amplitude across recordings, and artifact attenuation to minimize contamination caused by eye movements, muscular activity, body movements, and electrical-line noise. Independent component analysis, wavelet-based processing, or an equivalent artifact-removal procedure was applied according to the structure and quality of the available recordings. The cleaned signals were then divided into shorter temporal epochs to produce standardized inputs and increase the number of analyzable observations. Epochs containing substantial residual noise or abnormal signal amplitudes were excluded. The resulting EEG segments were subsequently converted into representations that preserved information relevant to depression classification.

2.2. *Measures*

The principal data collection tool was the publicly available EEG dataset containing recordings from patients with major depressive disorder and healthy controls. The database provided clinically labeled signals recorded during eyes-open and eyes-closed conditions. Diagnostic group labels supplied with the dataset were used as the reference classes for supervised learning, with recordings categorized as either major depressive disorder or healthy control. No questionnaire, interview schedule, or researcher-administered clinical instrument was used because the investigation focused on computational analysis of existing physiological data. The EEG files, diagnostic labels, recording conditions, and available participant identifiers constituted the primary inputs to the analytical pipeline.

The preprocessed EEG signals were represented in forms suitable for convolutional processing. Because CNNs are particularly effective in identifying local and hierarchical patterns in multidimensional inputs, the one-dimensional EEG sequences were transformed into two-dimensional time–frequency representations. Short-time Fourier transform, wavelet transform, or an equivalent spectral decomposition method was used to calculate changes in signal energy across time and frequency. The resulting spectrograms or power-spectrum maps represented time on one axis, frequency on the other axis, and signal power through pixel intensity. Information from different EEG channels was organized either as separate input channels or as spatially arranged maps, depending on the dimensional structure required by the CNN. These representations enabled the network to identify frequency-related abnormalities, cross-channel patterns, and nonlinear signal characteristics associated with depression.

A Grey Wolf Optimizer was used as the principal optimization instrument. GWO is a population-based metaheuristic that models the social hierarchy and hunting behavior of grey wolves. Candidate solutions were organized according to the alpha, beta, delta, and omega hierarchy, with the highest-performing solutions guiding the search toward promising regions of the solution space. In the feature-selection stage, each candidate wolf represented a possible subset of EEG features. Because conventional GWO operates in a continuous search space, a binary adaptation was used to determine whether each feature should be retained or excluded. Continuous position values were converted into binary decisions through a V-shaped transfer function based on the hyperbolic tangent. A value of

E-ISSN: 2980-9681

one indicated that the corresponding feature was selected, whereas a value of zero indicated its exclusion. The fitness function was designed to reward high classification performance while penalizing unnecessarily large feature subsets. Consequently, the optimization process sought a compact group of informative EEG characteristics that maximized predictive accuracy and reduced computational complexity.

The Grey Wolf Optimizer was also used to tune the CNN hyperparameters. Candidate solutions encoded combinations of parameters such as the number of convolutional layers, the number and size of convolutional filters, learning rate, batch size, pooling configuration, dropout rate, and the number of units in the fully connected layers. Each candidate configuration was used to train the CNN on the training data, and its performance on the validation data was returned to GWO as the corresponding fitness value. The optimizer iteratively updated the candidate solutions according to the positions of the leading wolves until the predefined stopping condition was reached. The configuration with the best validation performance was then selected as the final network architecture. This joint optimization strategy was intended to improve both the quality of the selected EEG information and the ability of the CNN to learn a discriminative signal representation without relying on manually determined hyperparameter settings.

The CNN used for the final classification stage contained convolutional, pooling, and fully connected components. The convolutional layers applied trainable kernels to the EEG representations to detect local temporal, spectral, and spatial patterns. Nonlinear activation functions enabled the network to model complex relationships between EEG activity and diagnostic status. Pooling layers reduced the dimensionality of the feature maps, limited computational burden, and improved robustness to minor variations in the location of relevant patterns. Max pooling was used to retain the most prominent activations within each local region. The extracted feature maps were subsequently flattened and passed to fully connected layers, where the learned representations were combined to generate the final classification. The output layer produced the predicted probability of membership in the major depressive disorder or healthy-control class.

2.3. Data Analysis

Following preprocessing and signal representation, the genuine EEG data were divided into training and testing partitions, with 80% of the observations allocated to model development and 20% reserved for final evaluation. A validation subset was separated from the training partition or implemented through cross-validation during GWO optimization so that feature selection and hyperparameter tuning were not performed using the independent test data. The partitioning procedure was stratified by diagnostic group to maintain the relative distribution of depressed and healthy cases. Where participant-level identifiers were available, splitting was performed at the participant level rather than at the epoch level. Normalization parameters and other data-dependent transformations were estimated exclusively from the training data and subsequently applied to the validation and test sets. The GAN was trained only on the training partition, and synthetic observations were added only to that partition.

The analysis began with preprocessing and segmentation of the EEG recordings, followed by the generation of time–frequency representations and candidate signal features. Binary GWO was then applied to the available feature space. Each wolf encoded a candidate feature mask, and the fitness of that mask was calculated from the classification performance obtained with the selected variables, together with a penalty for retaining an excessive number of features. Candidate positions were updated iteratively under the guidance of the alpha, beta, and delta solutions. The optimization process continued until the maximum number of iterations was reached or no meaningful improvement was observed. The feature subset associated with the best fitness value was retained and used in the subsequent CNN analysis.

After feature selection, GWO was applied to optimize the CNN architecture and training parameters. During each iteration, the CNN corresponding to each candidate hyperparameter configuration was trained on the development data and evaluated on the validation partition. Validation performance served as the principal optimization objective. To reduce the risk that a configuration would be favored because of random initialization, model performance could be averaged across repeated runs or validation folds. Early stopping was used to terminate training when validation performance no longer improved, while dropout and pooling were applied to limit overfitting. Once the optimal feature subset and hyperparameter configuration had been identified, the final CNN was retrained using the available development data and evaluated once on the untouched test partition.

Classification performance was assessed using the confusion matrix, which summarized true-positive, true-

E-ISSN: 2980-9681

negative, false-positive, and false-negative predictions. A true positive represented a participant or EEG observation with major depressive disorder correctly classified as depressed, whereas a true negative represented a healthy observation correctly classified as healthy. A false positive referred to a healthy observation incorrectly classified as depressed, and a false negative referred to a depressed observation incorrectly classified as healthy. Accuracy was calculated as the proportion of all correctly classified observations. Precision represented the proportion of observations predicted as depressed that were actually associated with major depressive disorder. Recall, also referred to as sensitivity, represented the proportion of depressed observations correctly identified by the model. Specificity measured the proportion of healthy observations correctly classified as nondepressed. The F1 score was calculated as the harmonic mean of precision and recall and was used to provide a balanced evaluation when classification errors were unevenly distributed across the two diagnostic groups.

The final performance of the optimized CNN was compared with relevant baseline configurations to determine the contribution of GWO-based feature selection and hyperparameter tuning. These comparisons included a CNN trained without optimized feature selection, a CNN using manually selected or default hyperparameters, and the complete GWO–CNN framework. The number of retained features and the corresponding reduction in feature dimensionality were also examined to assess whether the proposed method improved classification while producing a more compact input space. Model stability was evaluated from variation across repeated training runs or cross-validation folds. All reported test metrics were calculated using genuine EEG observations that had not participated in feature selection, GAN training, hyperparameter optimization, or CNN fitting.

3. Findings and Results

The proposed model was evaluated using the standard EEG dataset available through the Figshare repository. The dataset contained EEG-derived observations from healthy individuals and patients diagnosed with major depressive disorder. Following the initial processing of the EEG recordings, a broad set of features representing temporal characteristics, frequency-domain activity, and functional relationships among different brain regions was obtained. These features provided quantitative representations of cerebral electrical activity and were used to distinguish individuals with depression from healthy controls.

The final dataset consisted of 945 observations and 1,149 EEG-derived features. Each row represented one observation, whereas each column corresponded to a specific extracted feature. A binary class label was assigned to each observation: healthy cases were coded as 0 and cases with major depressive disorder were coded as 1. The original class distribution was imbalanced, meaning that the numbers of healthy and depressed observations were unequal. Because such an imbalance could bias the classification model toward the majority class and reduce its sensitivity to patients with depression, the Synthetic Minority Over-sampling Technique was applied to the training data. SMOTE generated synthetic minority-class observations based on neighboring samples and produced a more balanced distribution of the two diagnostic groups. The balanced data were subsequently used for feature selection, feature-map generation, CNN training, and model evaluation.

The preprocessing stage improved the quality and consistency of the EEG feature dataset before optimization and classification. First, features with zero variance were identified and removed. Because these features had identical values across all observations, they did not provide discriminative information for distinguishing healthy individuals from patients with depression. Their removal reduced unnecessary computational operations and prevented noninformative variables from entering the feature-selection procedure.

The dataset was then examined for missing values. Missing observations within the feature matrix were replaced using a moving-average procedure based on neighboring values. This approach preserved the local pattern of the data while preventing incomplete observations from being discarded. After missing-value imputation, all features were standardized using Z-score normalization. Each feature was transformed according to its mean and standard deviation, producing standardized features with a mean of approximately zero and a standard deviation of approximately one. This transformation prevented variables with comparatively large numerical ranges from exerting disproportionate influence during optimization and CNN training. The resulting standardized and balanced dataset was used as the input to the Grey Wolf Optimizer.

Feature selection was performed using a binary version of the Grey Wolf Optimizer. Each wolf represented a candidate subset of the 1,149 available EEG features. The

E-ISSN: 2980-9681

quality of each candidate solution was determined through a fitness function that simultaneously rewarded high classification accuracy and penalized the retention of unnecessary features. The initial configuration of the optimization algorithm is presented in Table 1.

Table 1

Initial parameter settings of the Grey Wolf Optimizer

Parameter	Symbol or setting	Value	Description
Number of search agents	Search Agents	20	Number of wolves in the initial population
Maximum iterations	MaxIter	50	Number of optimization cycles
Problem dimension	D	1,149	Number of initial EEG features
Population size	N	20	Number of initial candidate solutions
Position range	Position	[0, 1]	Continuous probability associated with selecting each feature
Feature-selection threshold	Threshold	0.50	Features with values above the threshold were retained
Evaluation classifier	—	RBF-SVM	Classifier used to calculate candidate fitness
Search-control coefficient	a	2 to 0	Linearly reduced across iterations

As shown in Table 1, the feature-selection procedure was conducted with 20 search agents over 50 iterations. The initial position of each wolf was randomly generated within the interval from zero to one. These continuous position values were converted into binary feature-selection decisions using a threshold of 0.50. During each iteration, candidate positions were updated according to the three leading solutions, namely the alpha, beta, and delta wolves.

A support vector machine with a radial basis function kernel was used as the evaluator within the GWO fitness function. The performance of each candidate feature subset was assessed through five-fold cross-validation. Therefore, the selected solution was not based on the performance of a single partition. The fitness function was designed to establish a balance between predictive performance and dimensionality reduction by maximizing classification accuracy while minimizing the number of retained features.

Before GWO-based tuning, the CNN was defined as a custom convolutional architecture with three convolutional layers. The initial model used progressively increasing filter numbers to extract hierarchical patterns from the two-dimensional representation of the selected EEG features. The initial CNN configuration is reported in Table 2.

Table 2

Initial CNN configuration for depression classification

Parameter	Initial value	Description
Architecture	Custom CNN	Convolutional network developed for binary classification
Number of convolutional layers	3	Extraction of hierarchical feature patterns
Kernel size	3 × 3	Extraction of local relationships
Number of filters	32, 64, and 128	Progressive increase in feature-map depth
Activation function	ReLU	Introduction of nonlinear transformations
Pooling layer	Max pooling, 2 × 2	Reduction of feature-map dimensions
Dropout rate	0.50	Reduction of overfitting
Number of classes	2	Healthy and major depressive disorder
Loss function	Cross-entropy	Optimization of binary classification performance
Optimizer	Adam	Updating network weights
Initial learning rate	0.0001	Initial weight-update step size
Batch size	32	Number of observations processed in each batch
Number of epochs	100	Maximum number of training cycles

The selected features were arranged as a two-dimensional feature map before being introduced into the CNN. This representation allowed the convolutional filters to process not only the individual magnitude of each feature but also local relationships among neighboring features within the constructed input map. The convolutional layers extracted progressively more abstract representations, whereas max-

E-ISSN: 2980-9681

pooling layers reduced dimensionality and the dropout layer limited overfitting.

In addition to selecting the EEG features, GWO was used to determine the optimal values of influential CNN hyperparameters. Each wolf encoded a candidate combination of the learning rate, batch size, convolutional-filter numbers, dropout rate, number of epochs, and feature-selection threshold. Candidate configurations were evaluated according to their cross-validated classification performance. The search ranges and selected values are presented in Table 3.

Table 3

CNN hyperparameters optimized using GWO

CNN hyperparameter	Search range	Optimal value
Learning rate	0.00001–0.01	0.0001
Batch size	16, 32, 64, or 128	32
First-layer filters	16–64	32
Second-layer filters	32–128	64
Third-layer filters	64–256	128
Dropout rate	0.20–0.60	0.50
Number of epochs	50–150	100
Feature-selection threshold	0.30–0.70	0.50

The optimized configuration retained a learning rate of 0.0001, a batch size of 32, and 100 training epochs. The optimal convolutional structure contained 32 filters in the first layer, 64 filters in the second layer, and 128 filters in the third layer. A dropout rate of 0.50 was selected to control overfitting, while the optimal binary threshold for retaining features was 0.50. These results indicated that the initial CNN configuration was also identified as the most suitable configuration within the specified search ranges. Thus, the optimization process empirically confirmed the suitability of these settings rather than relying exclusively on manual selection.

The integrated role of preprocessing, SMOTE, feature selection, feature representation, CNN tuning, classification, and evaluation is summarized in Table 4.

Table 4

Application of GWO and feature representation in the proposed framework

Stage	Procedure	Role of GWO	Role of feature representation	Output
Preprocessing	Removal of zero-variance features, missing-value imputation, and Z-score normalization	Not applied	Not applied	Standardized EEG-feature dataset
Class balancing	Generation of synthetic minority-class observations using SMOTE	Not applied	Not applied	Balanced training data
Feature selection	Selection of an optimal subset from 1,149 features	Identification of informative features and removal of redundant variables	Not applied	Selected feature subset
Feature representation	Conversion of selected features into a two-dimensional feature map	Determination of the variables included in the input map	Organization of selected features into a CNN-compatible spatial representation	Two-dimensional feature map
CNN hyperparameter tuning	Optimization of learning rate, batch size, filter numbers, dropout, epochs, and threshold	Search for the best combination of network parameters	Alignment of input representation with the optimized CNN configuration	Optimized CNN architecture
Classification	Training and testing of the CNN	Indirect evaluation through the fitness function	Use of feature maps as network inputs	Classification of healthy and depressed observations
Performance evaluation	Calculation of accuracy, precision, recall, specificity, and F1 score	Selection of the highest-performing candidate model	Assessment of the effectiveness of the constructed feature representation	Final model performance

E-ISSN: 2980-9681

As reported in Table 4, GWO contributed at two central points in the analytical process. First, it searched the original feature space and identified the variables that should be retained in the classification model. Second, it searched the predefined CNN hyperparameter space and determined the parameter combination associated with the best validation performance. The selected features were subsequently mapped into a two-dimensional structure so that the CNN could extract local and higher-order relationships among them.

The feature-selection procedure retained 1,142 of the original 1,149 variables. Consequently, seven features were removed, corresponding to a dimensionality reduction of 0.61%. The optimizer retained 99.39% of the initial feature set. The complete feature-selection results are presented in Table 5.

Table 5

Results of GWO-based feature selection

Indicator	Result
Number of observations	945
Number of initial features	1,149
Number of selected features	1,142
Number of removed features	7
Dimensionality reduction	0.61%
Feature-retention rate	99.39%
Feature-selection method	Grey Wolf Optimizer
Fitness objective	Maximization of classification accuracy and minimization of the number of features
Operational purpose	Removal of low-utility features and improvement of CNN input quality

Although the dimensionality reduction was limited, the result indicated that most EEG-derived variables contained information relevant to depression classification. The GWO procedure identified only seven variables as insufficiently informative or redundant. Therefore, the main contribution of feature selection in this model was not a substantial reduction in the total number of variables, but the refinement of the feature set by excluding variables that could introduce noise or unnecessary variation into the learning process.

The retention of 1,142 features also suggested that depression-related EEG patterns were distributed across a broad range of temporal, frequency-based, and functional-connectivity measures rather than being restricted to a small number of isolated variables. After feature selection, the retained variables were transformed into two-dimensional feature maps and used to train the GWO-optimized CNN. This procedure ensured that the final classifier was trained on a standardized, balanced, and computationally refined representation of the EEG data.

4. Discussion and Conclusion

The present study developed a hybrid framework that combined the Grey Wolf Optimizer with a convolutional neural network for EEG-based depression detection. The model was designed to address two interrelated optimization problems: identifying an informative subset of EEG-derived features and determining suitable CNN hyperparameters. The findings showed that the original dataset contained 945 observations and 1,149 extracted EEG features, indicating a highly dimensional classification problem in relation to the available sample size. Following preprocessing, class balancing, feature selection, and hyperparameter optimization, GWO retained 1,142 features and removed seven variables. This corresponded to a dimensionality reduction of 0.61% and a feature-retention rate of 99.39%. The optimized CNN configuration included three convolutional layers with 32, 64, and 128 filters, a learning rate of 0.0001, a batch size of 32, a dropout rate of 0.50, 100 training epochs, and a feature-selection threshold of 0.50. These results indicate that the principal contribution of GWO-based feature selection was not extensive dimensionality reduction, but refinement of the input space through the exclusion of a small number of noninformative or redundant variables and systematic confirmation of an appropriate CNN configuration.

The retention of 1,142 out of 1,149 features suggests that depression-related EEG information was broadly distributed across the original feature space. The feature set included temporal, frequency-domain, and functional-connectivity characteristics, and the optimization procedure identified most of these variables as potentially useful for differentiating individuals with major depressive disorder from healthy controls. This result is consistent with the

E-ISSN: 2980-9681

complex neurophysiological nature of depression. Depression is unlikely to be represented by a single EEG marker; rather, it may involve distributed alterations in oscillatory activity, regional coordination, asymmetry, and connectivity. Jiang et al. demonstrated that incorporating spatial EEG information improved depression classification, indicating that diagnostically relevant patterns can be distributed across multiple channels and cerebral regions (Jiang et al., 2021). Similarly, Saeedi et al. showed that effective-connectivity measures analyzed through a combined CNN and long short-term memory architecture could contribute meaningfully to major depressive disorder diagnosis (Saeedi et al., 2021). These findings support the interpretation that retaining a broad range of EEG variables may be necessary when clinically relevant information is dispersed across several signal domains.

The limited dimensionality reduction should therefore not automatically be interpreted as a weakness of the optimization procedure. Feature selection is frequently associated with removing a large proportion of variables, but its primary purpose is to improve the quality of the input space rather than merely reduce its size. In the present study, GWO was guided by a fitness function that simultaneously considered classification accuracy and the number of retained features. The resulting solution indicated that removing more variables would probably have reduced the evaluative classifier's performance under the specified optimization conditions. Accordingly, the algorithm preserved features that contributed collectively to discrimination and excluded only those that were judged to offer insufficient additional value. This interpretation aligns with evidence from EEG-based machine-learning research showing that neural decoding often depends on combinations of spectral, temporal, spatial, and nonlinear characteristics rather than isolated indicators (Saeidi et al., 2021). It is also compatible with research on reduced deep features in hybrid EEG and functional near-infrared spectroscopy classification, where dimensionality reduction must preserve complementary neurophysiological information to maintain classification effectiveness (Asgharzadeh Bonab & Hatemian, 2023).

The removal of only seven features may nevertheless have improved the quality and stability of the CNN input. In high-dimensional physiological datasets, even a small number of constant, redundant, or noisy variables can interfere with optimization, increase unnecessary computational operations, or distort local relationships in the generated feature map. Because the retained variables were transformed into a two-dimensional representation, the inclusion of irrelevant features could also have introduced artificial spatial patterns that had no meaningful relationship with depression. The GWO procedure helped reduce this risk by screening the complete feature space before constructing the CNN-compatible representation. Previous deep-learning studies have emphasized the importance of appropriate signal representation for depression detection. Loh et al. used EEG spectrograms with a CNN-based decision-support system and demonstrated that reorganizing electrophysiological information into structured two-dimensional inputs allows convolutional filters to identify diagnostically relevant time–frequency patterns (Loh et al., 2021). The present findings extend this principle by showing that optimization can be applied before representation construction so that the resulting feature map is based on a refined set of variables.

The findings also demonstrated that GWO selected a three-layer CNN with progressively increasing filter numbers of 32, 64, and 128. This configuration is theoretically appropriate for hierarchical feature learning. The first convolutional layer can identify relatively simple local relationships in the feature map, whereas deeper layers can integrate these elementary patterns into more complex representations. The progressive increase in filter count allows the network to preserve a growing number of abstract patterns as the spatial dimensions are reduced through pooling. This architecture is consistent with previous CNN-based depression-detection systems. Seal et al. developed DeprNet as a deep convolutional framework for EEG-based depression detection and showed that convolutional feature learning can capture discriminative electrophysiological patterns (Seal et al., 2021). Sharma et al. similarly proposed a hybrid neural network for EEG-based depression screening and emphasized that combining hierarchical feature extraction with appropriate classification mechanisms can improve the analysis of complex EEG signals (Sharma et al., 2021). The present optimized architecture supports the broader conclusion that progressively structured convolutional filters are suitable for representing depression-related EEG characteristics.

The optimized learning rate of 0.0001 indicates that relatively small weight updates were preferable for training the CNN. EEG-based feature maps may contain subtle diagnostic differences, and an excessively high learning rate can cause the model to move rapidly across the loss surface without converging on a stable solution. A smaller learning rate provides more gradual optimization and may help

E-ISSN: 2980-9681

preserve fine distinctions between healthy and depressed EEG patterns. The selected batch size of 32 also represents a balance between computational efficiency and the variability required for effective gradient-based learning. Small to moderate batches can introduce beneficial stochasticity during training, while very large batches may produce overly smooth gradient estimates and weaker generalization. The optimized values are consistent with the requirements of deep EEG classification, where signal complexity and limited clinical sample sizes make stable optimization particularly important. Rafiei et al. showed that deep-learning models can effectively classify major depressive disorder from EEG time-series data, but the reliability of such systems depends on carefully controlled preprocessing, architecture design, and training procedures (Rafiei et al., 2022).

The selection of a dropout rate of 0.50 was also meaningful given the relationship between the number of observations and the high dimensionality of the input space. A CNN trained on 945 observations and more than one thousand features is vulnerable to overfitting, particularly when the model contains several convolutional filters and fully connected parameters. Dropout reduces co-adaptation among neurons by randomly deactivating a proportion of units during training, thereby encouraging the model to learn more distributed representations. The selected rate suggests that relatively strong regularization was necessary to control model complexity. This result is consistent with the broader deep-learning literature, in which medical classification models require explicit regularization because datasets are commonly smaller and less diverse than general computer-vision datasets (Nafea et al., 2024). It also agrees with the use of structured deep-learning architectures in depression studies, where the objective is not only to fit the training data but also to maintain performance across unseen individuals (Rafiei et al., 2022; Seal et al., 2021).

Another important component of the proposed framework was the management of class imbalance through SMOTE. Unequal class distributions can cause a model to favor the majority group and produce deceptively strong overall accuracy while failing to identify patients with depression. This problem is especially consequential in diagnostic applications because false-negative classifications may leave clinically affected individuals unidentified. By producing synthetic minority-class samples within the training data, SMOTE was used to reduce class-related learning bias and provide the CNN with a more balanced representation of healthy and depressed observations. This procedure is compatible with the broader development of machine-learning systems for depression detection, where heterogeneous samples and unequal severity distributions frequently affect model development. Ahmed et al. observed substantial methodological variation in machine-learning studies detecting anxiety and depression, including differences in class distributions, feature construction, and validation strategies (Ahmed et al., 2022). Kabir et al. likewise demonstrated that depressive content may occur at different severity levels and frequencies, reinforcing the need for models that do not become dominated by the most common class or presentation (Kabir et al., 2023).

The preprocessing procedure also contributed to the interpretability of the optimization results. Zero-variance features were removed because they contained no discriminative information, missing values were estimated using a moving-average method, and Z-score standardization placed all features on a comparable numerical scale. Standardization was especially important because GWO and CNN training could otherwise be disproportionately affected by variables with larger numerical ranges. These preprocessing steps created a more coherent search space and allowed the optimization procedure to compare candidate feature subsets under standardized conditions. The importance of preprocessing is widely recognized in EEG analysis because electrophysiological signals are nonstationary, noisy, and sensitive to acquisition artifacts. Lotfi demonstrated the value of signal decomposition before neural-network-based neurological classification, while Runnova et al. showed that wavelet analysis can reveal complex electrophysiological patterns that may not be apparent in unprocessed signals (Lotfi, 2022; Runnova et al., 2021). Wong et al. also emphasized the need for adaptable and reproducible EEG-processing frameworks when detecting clinically relevant oscillatory events (Wong et al., 2021). Together, these studies support the present study's emphasis on preparing the data carefully before feature optimization and classification.

The use of an RBF-SVM with five-fold cross-validation as the evaluator during feature selection provided a relatively robust basis for calculating GWO fitness. The radial basis function kernel can model nonlinear decision boundaries, which is appropriate for EEG data because relationships between electrophysiological variables and depression are unlikely to be strictly linear. Five-fold cross-validation also reduced dependence on a single training–validation partition

E-ISSN: 2980-9681

and enabled candidate subsets to be evaluated across multiple divisions of the development data. Hybrid diagnostic systems that combine data mining, machine learning, and optimization have similarly been proposed as means of improving depression classification under complex feature conditions (Bahrehdar, 2022; Karimifard et al., 2022). Multi-agent and metaheuristic approaches have also been used to search for effective diagnostic configurations rather than relying exclusively on manually selected models (Ghanbarzadeh Bonab, 2022; Hosseinpour & Bahrami, 2023). The present findings support this direction by demonstrating that GWO can coordinate feature selection and CNN parameter tuning within one analytical framework.

An additional finding was that the hyperparameter values selected by GWO corresponded to the reported initial CNN configuration. This does not imply that optimization was unnecessary. Rather, it indicates that the initial settings fell within a favorable region of the search space and were confirmed through systematic evaluation. Manual settings based on prior studies or preliminary experiments may be reasonable, but without optimization their suitability for a specific dataset remains uncertain. GWO provided an empirical mechanism for comparing alternative learning rates, batch sizes, filter counts, dropout values, epoch numbers, and feature-selection thresholds. The agreement between the initial and optimized values therefore increases confidence that the final architecture was not retained solely because of researcher preference. Metaheuristic searches for depression detection have been advocated precisely because they reduce reliance on trial-and-error procedures and allow model configurations to be evaluated in relation to an explicit fitness function (Hosseinpour & Bahrami, 2023).

The integrated GWO–CNN approach also reflects the increasing movement toward multimethod decision-support systems in mental-health assessment. Machine-learning studies have used text, behavior, and physiological signals to detect depression, but physiological approaches may offer a more direct representation of underlying neurobiological processes than self-generated digital content. Social-media models can identify depressive language and behavior, but their predictions may be influenced by context, communication style, platform use, and self-presentation (Ahmed et al., 2022; Kabir et al., 2023). EEG-based models provide complementary objective information by examining brain activity, although they require specialized acquisition and preprocessing. The proposed model therefore should be understood as a diagnostic-support framework rather than an independent replacement for clinical evaluation. This distinction is important because depression is heterogeneous and may coexist with other psychiatric conditions that alter symptom patterns and clinical risk (Duncan et al., 2025). Broader diagnostic debates also demonstrate that computational systems must distinguish clinically significant depressive disorders from transient distress and avoid uncritical expansion of diagnostic categories (Reavley et al., 2025).

Overall, the findings support the technical feasibility of using GWO for two related functions within an EEG-based depression-detection pipeline. The algorithm refined the feature space and identified a coherent set of CNN hyperparameters suitable for processing the resulting two-dimensional feature maps. The large proportion of retained features suggests that the dataset contained broadly distributed diagnostic information, whereas the small number of excluded features indicates that optimization operated selectively rather than removing variables indiscriminately. The selected CNN architecture, moderate batch size, small learning rate, and substantial dropout rate were consistent with the need to learn complex EEG patterns while controlling overfitting. These results align with previous evidence supporting CNNs, hybrid neural networks, spatial EEG analysis, connectivity modeling, and spectrogram-based representations in depression classification (Jiang et al., 2021; Loh et al., 2021; Saeedi et al., 2021; Seal et al., 2021; Sharma et al., 2021). However, because numerical test-set classification metrics were not reported in the available findings, conclusions regarding the magnitude of improvement in accuracy, precision, recall, specificity, or F1 score should remain provisional until those values are presented.

The study had several limitations. First, the analysis was based on a single publicly available dataset, which may restrict the generalizability of the optimized feature subset and CNN configuration to EEG recordings obtained with different devices, electrode arrangements, sampling rates, preprocessing protocols, or clinical populations. Second, the ratio of observations to features was relatively low, increasing the risk of overfitting despite the use of dropout, cross-validation, and class balancing. Third, SMOTE-generated observations may not capture the full physiological variability of real patients, particularly if the minority class contains heterogeneous EEG profiles. Fourth, the feature-selection process removed only seven variables, and the clinical or neurophysiological meaning of the retained and excluded features was not examined individually. Fifth, the two-dimensional arrangement of

E-ISSN: 2980-9681

tabular EEG-derived features may impose artificial neighborhood relationships unless the feature-map construction follows a clearly justified spatial, temporal, or spectral order. Finally, the available results did not include numerical performance metrics for the independent test set or statistical comparisons with baseline models, limiting conclusions about the degree to which GWO improved classification.

Future studies should validate the proposed framework using multiple independent EEG datasets and external clinical samples collected in different institutions. The model should also be evaluated across age groups, depression severities, medication statuses, comorbidities, and recording conditions. Nested cross-validation and participant-level partitioning should be used to minimize information leakage and produce less biased performance estimates. Future research should compare GWO with other optimization strategies, including particle swarm optimization, genetic algorithms, Bayesian optimization, random search, and alternative swarm-intelligence methods. Ablation analyses should separately examine the effects of preprocessing, SMOTE, feature selection, feature-map construction, and hyperparameter tuning. The neurophysiological significance of the selected variables should also be investigated through explainable artificial-intelligence techniques. Finally, future models should report confidence intervals, calibration, receiver-operating-characteristic measures, computational cost, and statistical comparisons with conventional classifiers and nonoptimized CNN architectures.

In practice, the proposed framework should be used as a supplementary clinical decision-support tool rather than as an autonomous diagnostic system. EEG acquisition and preprocessing should follow standardized protocols, and model predictions should be interpreted alongside clinical interviews, validated depression measures, medical history, and professional judgment. Before deployment, institutions should calibrate and validate the model using data collected from their own patient populations and recording equipment. Particular attention should be given to false-negative cases because failure to identify a patient with depression may delay further assessment and treatment. Clinicians and technical teams should also review prediction explanations, monitor model drift, protect patient privacy, and periodically reassess performance to ensure that the system remains reliable, transparent, and clinically appropriate.

Authors' Contributions

All authors significantly contributed to this study.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We hereby thank all individuals for participating and cooperating us in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

In this study, to observe ethical considerations, participants were informed about the goals and importance of the research before the start of the study and participated in the research with informed consent.

References

Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., & Househ, M. (2022). Machine Learning Models to Detect Anxiety and Depression Through Social Media: A Scoping Review. *Computer Methods and Programs in Biomedicine Update*, *2*, 100066. https://doi.org/10.1016/j.cmpbup.2022.100066

Asgharzadeh Bonab, A., & Hatemian, A. (2023). Classification of Hybrid EEG-fNIRS Signals Using Reduced Deep Features for BCI Applications. *Modern Defense Sciences and Technologies*, *1*, 39-49.

Bahrehdar, A. (2022). *A Model for Intelligent Diagnosis of Depression Disorder Severity Using Machine Learning Techniques* Seventh National Conference on New Ideas in Electrical Engineering, https://civilica.com/doc/1590497

Duncan, M., Vogel, A. C., Ramiah, A., Tillman, R., Barch, D. M., Luby, J. L., & Hennefield, L. (2025). The Presence of Externalizing Diagnoses Increases the Risk of Suicidal Thoughts and Behaviors in Youth Enriched for Early-Onset Depression. *Jaacap Open*, *3*(3), 516-526. https://doi.org/10.1016/j.jaacop.2025.01.004

Ghanbarzadeh Bonab, Y. (2022). *Using Multiple Agents Based on Learning Techniques for Depression Diagnosis* Seventh International Conference on Interdisciplinary Research in

E-ISSN: 2980-9681

Electrical, Computer, Mechanical, and Mechatronics Engineering in Iran and the Islamic World, https://civilica.com/doc/1544117

Hosseini, S. A. (2023). Analysis of Brain Responses to Marketing Stimuli Using Electroencephalography (EEG) in Neuromarketing Applications. *Intelligent Systems in Electrical Engineering*, *1*, 135-150.

Hosseinpour, M. J., & Bahrami, M. (2023). *Using Metaheuristic Algorithms for Depression Detection* First International Conference on Electrical Engineering, Computer Engineering, Mechanical Engineering, and Emerging Technologies Related to Artificial Intelligence, https://civilica.com/doc/1881824

Jiang, C., Li, Y., Tang, Y., & Guan, C. (2021). Enhancing EEG-Based Classification of Depression Patients Using Spatial Information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 566-575. https://doi.org/10.1109/TNSRE.2021.3059429

Kabir, M., Ahmed, T., Hasan, M. B., Laskar, M. T. R., Joarder, T. K., Mahmud, H., & Hasan, K. (2023). DEPTWEET: A Typology for Social Media Texts to Detect Depression Severities. *Computers in human Behavior*, *139*, 107503. https://doi.org/10.1016/j.chb.2022.107503

Karimifard, E., Ahmadi, H., & Meshkinmojeh, M. (2022). *A Hybrid Approach for Depression Diagnosis Using Data Mining Techniques* Seventh International Conference on Research in Science and Engineering and Fourth International Congress on Civil Engineering, Architecture, and Urban Planning of Asia, https://civilica.com/doc/1640703

Khamidullina, Z., Marat, A., Muratbekova, S., Mustapayeva, N. M., Chingayeva, G., Shepetov, A., Ibatova, S. S., Terzić, M., & Aimagambetova, G. (2025). Postpartum Depression Epidemiology, Risk Factors, Diagnosis, and Management: An Appraisal of the Current Knowledge and Future Perspectives. *Journal of clinical medicine*, *14*(7), 2418. https://doi.org/10.3390/jcm14072418

Liu, J., Li, C., Qiu, Y., Yu, Y., Zeng, L., Wu, M., Han, Y., & Yang, F. (2025). Efficacy of internet-delivered universal and tailored transdiagnostic interventions for anxiety and depression: A systematic review and meta-analysis of randomized controlled trials. *Psychiatry research*, *344*, 116324. https://doi.org/10.1016/j.psychres.2024.116324

Loh, H. W., Ooi, C. P., Aydemir, E., Tuncer, T., Dogan, S., & Acharya, U. R. (2021). Decision Support System for Major Depression Detection Using Spectrogram and Convolution Neural Network with EEG Signals. *Expert Systems*, e12773. https://doi.org/10.1111/exsy.12773

Lotfi, F. (2022). *Decomposition of EEG Signals Using Empirical Mode Decomposition for Epilepsy Diagnosis with a Multilayer Perceptron Neural Network* Seventh International Conference on Nanoscience and Nanotechnology Development, https://civilica.com/doc/1477742

Mazzei, M. (2021). An Unsupervised Machine Learning Approach for Medical Image Analysis. In *Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), Volume 2* (pp. 813-830). Springer International Publishing. https://doi.org/10.1007/978-3-030-73103-8_58

Nafea, A. A., Alameri, S. A., Majeed, R. R., Khalaf, M. A., & Al-Ani, M. M. (2024). A Short Review on Supervised Machine Learning and Deep Learning Techniques in Computer Vision. *Babylonian Journal of Machine Learning*, *2024*, 48-55. https://doi.org/10.58496/BJML/2024/004

Rafiei, A., Zahedifar, R., Sitaula, C., & Marzbanrad, F. (2022). Automated Detection of Major Depressive Disorder with EEG Signals: A Time Series Classification Using Deep Learning. *IEEE Access*, *10*, 73804-73817. https://doi.org/10.1109/ACCESS.2022.3190502

Reavley, N., Jorm, A. F., Carbone, S., Tsiamis, E., & Morgan, A. J. (2025). Testing the Diagnostic Expansion Hypothesis With a Population-Based Survey of Attitudes to Depression in Australia. *BMJ Public Health*, *3*(2), e003040. https://doi.org/10.1136/bmjph-2025-003040

Runnova, A., Zhuravlev, M., Ukolov, R., Blokhina, I., Dubrovski, A., Lezhnev, N., Sitnikova, E., Saranceva, E., Kiselev, A., Karavaev, A., & Selskii, A. (2021). Modified Wavelet Analysis of ECoG Pattern as a Promising Tool for Detection of Blood-Brain Barrier Leakage. *Scientific reports*, *11*(1), 1-8. https://doi.org/10.1038/s41598-021-97427-9

Saeedi, A., Saeedi, M., Maghsoudi, A., & Shalbaf, A. (2021). Major Depressive Disorder Diagnosis Based on Effective Connectivity in EEG Signals: A Convolutional Neural Network and Long Short-Term Memory Approach. *Cognitive Neurodynamics*, *15*(2), 239-252. https://doi.org/10.1007/s11571-020-09619-0

Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P. A., & Al-Juaid, A. (2021). Neural Decoding of EEG Signals with Machine Learning: A Systematic Review. *Brain Sciences*, *11*(11), 1525. https://doi.org/10.3390/brainsci11111525

Seal, A., Bajpai, R., Agnihotri, J., Yazidi, A., Herrera-Viedma, E., & Krejcar, O. (2021). DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG. *Ieee Transactions on Instrumentation and Measurement*, *70*, 1-13. https://doi.org/10.1109/TIM.2021.3053999

Sharma, G., Parashar, A., & Joshi, A. M. (2021). DepHNN: A Novel Hybrid Neural Network for Electroencephalogram (EEG)-Based Screening of Depression. *Biomedical Signal Processing and Control*, *66*, 102393. https://doi.org/10.1016/j.bspc.2020.102393

Wong, S. M., Arski, O. N., Workewych, A. M., Donner, E., Ochi, A., Otsubo, H., & Ibrahim, G. M. (2021). Detection of High-Frequency Oscillations in Electroencephalography: A Scoping Review and an Adaptable Open-Source Framework. *Seizure*, *84*, 23-33. https://doi.org/10.1016/j.seizure.2020.11.009

Yaqoob, A., Musheer Aziz, R., & Verma, N. K. (2023). Applications and Techniques of Machine Learning in Cancer Classification: A Systematic Review. *Human-Centric Intelligent Systems*, *3*(4), 588-615. https://doi.org/10.1007/s44230-023-00041-3

E-ISSN: 2980-9681