An Enhanced Approach in Predicting and Classifying Major Depressive Disorder Using Bi-GRU with Attention Mechanism and Transfer Learning

Udutala Mahender
Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
mahenderudutala@gmail.com (corresponding author)

S. Arivalagan
Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
arivucseau@gmail.com

V. Sathiyasuntharam
Department of Cyber Security, CMR Engineering College, Hyderabad, India
sathiya4196@gmail.com

P. Sudhakar
Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
kar.sudha@gmail.com

Received: 16 May 2025 | Revised: 10 July 2025 | Accepted: 19 July 2025

 DOI: https://doi.org/10.48084/etasr.12176

ABSTRACT

Major Depressive Disorder (MDD) is a serious issue in medical research due to its high prevalence and significant impact on the quality of life of individuals, leading to disability, comorbidity, and an increased risk of suicide. Accurate and early diagnosis is crucial for effective treatment, yet predicting MDD remains challenging due to its complex etiology, overlapping symptoms with other psychiatric disorders, and the subjective nature of traditional diagnostic methods. This study proposes RAG-EEGNet, a novel approach to detect MDD using EEG data, integrating advanced feature extraction and selection techniques with deep learning models. Initially, Boruta was used for comprehensive feature extraction, followed by Cuckoo Search Optimization (CSO) to select impactful features. A hybrid model is employed for classification, combining ResNet-50 and a Bi-GRU enhanced by an attention mechanism. The results show a significant improvement in the detection of MDD with an accuracy of 99.01%, precision of 100%, recall of 99.24%, F1-score of 99.12%, and ROC-AUC of 99.0%, demonstrating the efficacy of the proposed approach, highlighting the critical role of EEG data in diagnosing and predicting mental diseases.

Keywords-Major Depressive Disorder(MDD); Boruta feature extraction; Cuckoo Search Optimization (CSO); ResNet-50; Bi-GRU; attention mechanism

I. INTRODUCTION

Depression is a prevalent mental health issue that affects individuals and society, projected by the WHO to become the most prevalent mental illness by 2030, and often leads to severe consequences, including suicide, highlighting the need for effective diagnostics and treatments [1]. Despite its seriousness, there are currently no accurate and clinically relevant diagnostic tests available [2]. Diagnostic challenges are further compounded by the subjective nature of symptom reporting, patient education, cognitive capabilities, and honesty [3]. Effective diagnosis requires extensive clinical training and is highly dependent on the clinician's skill and commitment [3].

The medical diagnosis of Major Depressive Disorder (MDD) is complicated by its variability. MDD affects millions of people around the world and has a lifetime prevalence of 14% [4]. It causes pessimism, an inability to improve mood, loss of interest in activities, lowers life quality, and raises suicide risk [5]. Early and correct diagnosis is essential for effective treatment, but no biomarkers exist for MDD diagnosis, progression, or treatment response [6]. Clinical interviews use subjective, non-specific patient-reported symptoms to diagnose, especially in early stages [2]. Anxiety, substance use, and other comorbidities hinder treatment [7]. MDD affects work, family, and life for over 300 million people [8, 9]. Depression increased from 9.7% in 2019 to 19.8% in 2020 [10]. Depression onset is influenced by genetics, environment, experience, physiology, and gender, with females having higher rates [11].

The diagnostic accuracy for depression disorder has improved with intelligent technologies [12]. EEG [13], MEG [14], and fMRI [15] are utilized to examine changes in brain function and automate the diagnosis of mental disorders [16]. Their simple, inexpensive, and non-invasive features make EEGs popular for diagnosing depression [17]. MDD patients' EEGs show changes in numerous frequency bands and brain locations, especially the frontal lobe [13, 18]. Frontal EEG power dynamics affect antidepressant response [19]. MDD can be diagnosed using Deep Learning (DL) techniques on six-channel frontal EEGs [19].

Modern MRIs detect MDD-related brain changes with high spatial resolution. MDD patients have thinner insula and orbitofrontal regions and smaller hippocampi than Healthy Controls (HCs) [20]. Small effects and group inference limit clinical use, but using ML methods on multivariate brain feature combinations and individual inference can increase it. Resting-state functional MRI (rs-fMRI) shows brain functional connectivities through cerebral blood flow [15]. MDD patients have impaired default mode, subcortical, and sensorimotor resting Functional Connectivity (FC), which correlates with depression severity [20]. MDD patients can be distinguished from HCs using FC-derived neural network markers [21].

Due to these challenges, DL methods are needed to predict the severity of depression, improving diagnosis and treatment decisions. DL methods can reduce the public health burden of depression and MDD by improving early identification and treatment. This study presents the RAG-EEGNet model, a DL architecture for MDD diagnosis utilizing EEG data. Figure 1 illustrates the proposed method. Preprocessing and normalization were applied to an EEG disorder dataset to improve data quality and recording consistency, thus improving training stability and model convergence. High-level features were extracted with ResNet-50 [22] and Bi-GRU with attention [23], handling both spatial and temporal EEG correlations. Boruta [24] and Cuckoo Search Optimization (CSO) [25] helped determine the most significant features, making the model more applicable across patient demographics and reducing the demand for large annotated datasets. After extraction and selection, the fitness functions assessed feature quality to ensure feature selection precision. The performance and robustness of the RAG-EEGNet model were optimized, including hyperparameter tuning. Evaluation metrics included accuracy, precision, recall, F1-score, and AUC-ROC to quantify the performance of the model. The results show that the RAG-EEGNet model is capable of handling the complexities of MDD diagnosis using EEG data and its flexibility across different groups of patients without requiring separate training data for each group.

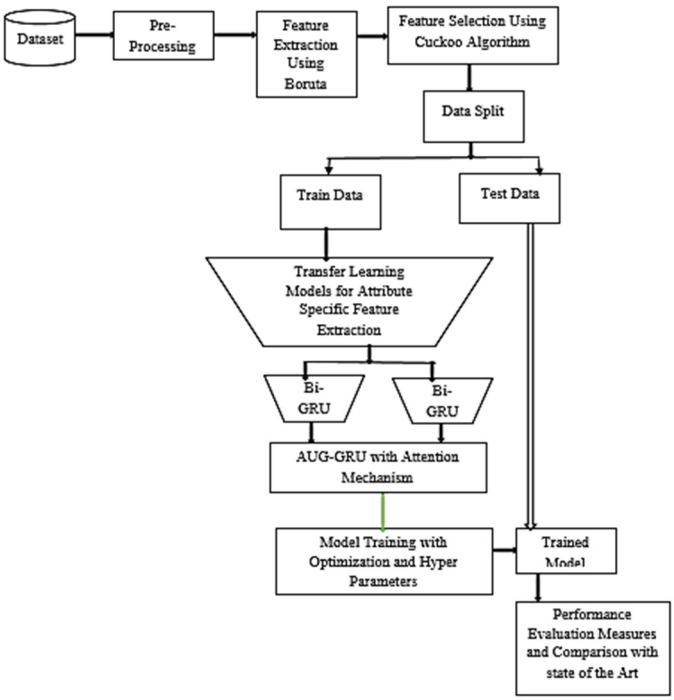

Fig. 1. Proposed method.

II. RELATED WORKS

Deep neural networks excel in pattern recognition [26], but overfitting and training times are concerns. RBM-based Deep Belief Networks [27] accelerate training. CNNs extract discriminative features using convolution, pooling, rectifiers, and normalization [28]. CNNs excel in image analysis but not in temporal decoding. LSTM RNNs tackle long-term dependency concerns such as vanishing gradients in sequential data processing [29]. LSTMs are suitable for time-sequential applications such as the representation of textual emotional states [28]. DL has helped mental health researchers diagnose and predict MDD, as they can find hidden patterns in large datasets and make accurate predictions [12]. CNNs are good at extracting EEG information, improving classifier generalization [17]. RNNs, especially LSTM networks, are useful for modeling MDD and MCI time series [18]. In [28], a CNN-LSTM-based hybrid model classified EEG data from the left and right hemispheres with 99.12% and 97.66% accuracy, showing DL's promise in depressive disorder identification.

Graph Neural Networks (GNNs) and other DL approaches are gaining popularity for graph-structured data. GNNs have been used to diagnose MDD and MCI [30]. In [31], a GCN was 30% more accurate in early AD diagnosis in cognitively unimpaired persons. In [32], GNN and LSTM networks on

functional brain network dynamics predicted MDD with 80% accuracy. In [33], a multi-task learning model with tensors and a regularizer predicted MDD with 85% accuracy using rs-fMRI data. In [34], an ensemble learning strategy diagnosed MDD with 88% accuracy using a transformer and CNN to extract features from DTI, fMRI, and sMRI data. In [35], graph models classified 16 locations, and in [36], graph encoders were used to classify multi-site MDD.

Recent works have improved DL models for MDD and related disease diagnosis. In [37], using brain connection modeling, a two-branch CNN model with multi-head self-attention on resting-state EEG data achieved 91.06% accuracy. In [38], DepressionGraph surpassed traditional GNNs in 533 runs, utilizing a transformer-encoder GNN architecture for MDD diagnosis from rs-fMRI data, despite gender distribution bias. In [39], the FBANet method using HTP drawings to automatically diagnose depression, achieving 99.07% validation accuracy and 97.71% average accuracy with minimal preprocessing. In [40], MRCNN-LSTM and MRCNN-RSE models accurately detected depression in EEG recordings (98.48 ± 0.22%), making them helpful for therapy and efficacy assessment. In [41], an LSTM-based RNN accurately detected depressive symptoms in text data, demonstrating significant potential.

III. PROPOSED METHOD

The proposed EEG-based MDD detection method emphasizes rigorous data preprocessing and division into training and test sets to improve model resilience. For feature extraction and temporal dependency capture, the hybrid RAG-EEGNet architecture uses transfer learning with ResNet-50, along with a Bi-GRU with an attention mechanism. To address the complexity of MDD diagnosis, RAG-EEGNet used Boruta for feature extraction and CSO for feature selection.

A. Preprocessing

EEG signal preparation used artifact removal and re-referencing to improve data quality. To separate the effects of internal and external noise in the EEG data, Independent Component Analysis (ICA) was used to rectify muscle contractions and eye blink disturbances, increasing signal clarity and minimizing artifacts. EEG data were prepared as NumPy arrays for analysis and feature extraction using a 50-Hz bandpass filter to reduce spike noise from the power spectral distribution.

EEG data were preprocessed using NeuroGuide Deluxe 3.0.5 in many phases to improve quality. To control computational load and focus on key frequency ranges, the sample rate was first decreased to 128 Hz. A 40 Hz bandpass filter isolated baseline EEG signals, followed by a second filter to refine them. NeuroGuide's splicing method reduced artifacts by integrating 600-ms-long edited EEG segments. To increase the signal-to-noise ratio and overall quality, the baseline EEG was filtered using a Butterworth high-pass filter set at 1 Hz and a low-pass filter with a frequency set at 55 Hz. In the global 10-20 system, 19 ear channels were evaluated, including FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T5, P3, Pz, P4, T6, O1, and O2, ensuring thorough EEG coverage. Finally, EEG signals were normalized using z-scores.

Inter-subject variability in EEG signals was addressed through subject-wise normalization following artifact removal using ICA. Z-score normalization was applied independently for each subject, with mean and standard deviation computed from individual recordings. This method preserved intra-subject dynamics while standardizing feature distributions across subjects, enhancing model stability and generalization during classification.

Linear and nonlinear features were retrieved from preprocessed EEG data to distinguish between healthy and mentally ill power spectrums. The power of the alpha, beta, delta, and theta was linear, as were the mean, median, and lowest values. EEG data were investigated for nonlinear features, saving all extracted linear and nonlinear characteristics for analysis.

B. Feature Extraction Using the Boruta Algorithm

The proposed method for MDD prediction and classification utilizing EEG data relies on the Boruta algorithm for robust feature extraction. The high-dimensional and complicated nature of EEG signals requires excellent feature extraction to identify relevant information for successful prediction models. Boruta finds features with significant target variable information, keeping the most informative and removing unnecessary features by iteratively comparing original and randomized shadow features. This stringent feature selection procedure helps the model focus on crucial EEG data, boosting MDD classification and prediction accuracy. The process is as follows.

Let $X \in R^{n \times p}$ be the database that has n number of samples and p attributes. Let $y \in R^n$, be the target vector indicating the presence or absence of MDD.

Generate shadow features X_{shadow} by randomly permuting the values of each feature x_i in X.

Concatenate the original and shadow features to form the extended dataset $X_{ext} = [X_{shadow}]$.

Train a random forest classifier on X_{ext} with y. Compute the importance score I_i for each original feature x_i and the shadow features I_{shadow}. Then, calculate the maximum importance of shadow features:

$$I_{shadow}^{max} = \max(I_{shadow}) \quad (1)$$

Compare each original feature's importance I_i to I_{shadow}^{max}.

If $I_i > I_{shadow}^{max}$, x_i is important; otherwise, unimportant.

C. Feature Selection Using CSO

Feature selection was used in the proposed EEG-based MDD detection technique to improve efficiency and performance. From the high dimensional EEG dataset, the most relevant features were extracted using CSO. CSO uses Lévy flights to search space, inspired by cuckoo brood parasitism, which is a metaheuristic optimization. CSO optimizes feature selection to maximize forecast accuracy while minimizing redundancy and unnecessary information. The process is as follows.

Initialize a population of nests (feature subsets) $N \in \{0,1\}^{m \times p}$, where m is the number of nests and p is the total number of features.

Randomly initialize each nest, using a binary vector where 1 indicates that the feature is selected and 0 indicates it is not.

Define the fitness function $f(N_i)$ for each nest N_i to evaluate the classification accuracy of the selected features.

$$f(N_i) = Accuracy(X_N, y) \quad (2)$$

where X_N, represents the dataset with features selected by N_i. Then use Levy flights and new solutions. Generate a new solution by:

$$N_i^{new} = clip(N_i + \alpha . Levy(\lambda), 0,1) \quad (3)$$

where $Levy(\lambda)$ denotes a Levy distribution, and α is the step size. To ensure binary values, the new solutions are clipped to the range [0, 1].

Then, evaluate and select features, using the fitness function on the new solution $f(N_i^{new})$, and evaluate its fitness compared to the previous one. If $f(N_i^{new}) > f(N_i)$, update the nest:

$$Ni = N_i^{new} \quad (4)$$

With a probability P_a, replace some nests with new random solutions to maintain diversity:

$$N_i \sim \{0,1\}^p, \text{ if } rand() < P_a \quad (5)$$

The optimization process continues iteratively, with each nest being evaluated and updated based on the fitness function. The incorporation of Lévy flights ensures a global search capability, avoiding local optima. The randomization step helps maintain diversity within the population, preventing premature convergence.

D. The Proposed RAG-EEGNet Model

1) ResNet-50

The RAG-EEGNet architecture for MDD prediction and classification uses ResNet-50 to extract complex features from preprocessed input data. Batch normalization stabilizes and speeds training, while global average pooling creates a one-dimensional feature vector from feature maps. After capturing complex feature interactions with a dense layer and ReLU activation, a 50% dropout layer prevents overfitting. A final 2-unit Dense layer with softmax activation classifies items. Figure 2 shows the architectural flow of RAG-EEGNet.

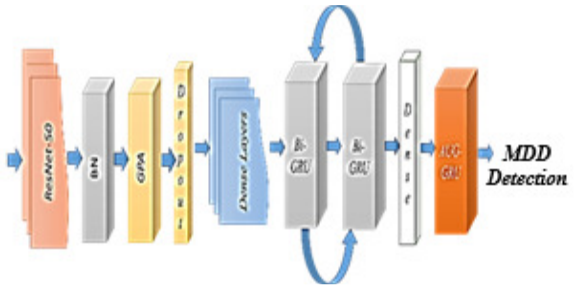

Fig. 2. Architecture of RAG-EEGNet.

RAG-EEGNet uses ResNet-50, a powerful DL architecture that addresses accuracy deterioration and gradient vanishing, to improve MDD detection using EEG data. ResNet-50, a 50-layer residual network, learns residual mappings by skipping connections. These connections directly propagate input to deeper layers, improving training gradient flow. ResNet-50 simplifies residual information learning, allowing it to extract and use complicated EEG features. ResNet-50 was chosen for its improved feature extraction, based on its performance in image segmentation, object detection, and classification. ResNet-50's 16 residual blocks are staged. Max-pooling and convolutional layers are used in the first stage, followed by stages with different numbers of residual blocks and convolutional layers: three, four, and six. This hierarchical structure helps ResNet-50 capture and interpret complicated EEG signals, improving RAG-EEGNet's MDD detection.

A convolution operation is performed by:

$$y = (x * W) + b \quad (6)$$

where x is the input tensor, W is a convolution kernel (filter), $*$ is a convolution operation, b is a bias term, and y is the output tensor after convolution. Residual connections work by:

$$y = F(x, \{Wi\}) + x \quad (7)$$

where x is an input tensor (identity mapping), $F(x, \{W_i\})$ is a function representing the residual block, typically a series of convolutions, batch normalization, and ReLU activations, and y is the output tensor after adding the input x and the function $F(x, \{W_i\})$. An expanded form for a typical residual block is:

$$y = ReLU\big(BN\big(W2 * ReLU\big(BN(W1 * x + b1)\big) + b2\big)x\big) \quad (8)$$

$$x' = \frac{x - \mu}{\sqrt{\sigma 2 + \varepsilon}} \quad (9)$$

$$y' = \gamma x + \beta \quad (10)$$

where x is the input to the batch normalization layer, μ is the mean of the input, σ^2 is the variance of the input, ε is a constant to avert division by zero, x' is a normalized input, and γ and β are learnable parameters. The ReLU (Rectified Linear Unit) function is defined as:

$$ReLU(x) = max(0, x) \quad (11)$$

After ResNet50, the model was optimized by adding batch normalization, a dense layer of 1024, and a dropout of 50%.

ResNet50 requires a labeled dataset. To improve accuracy, the model was trained using mini-batch stochastic gradient descent and adjusted hyperparameters, such as the learning rate. The ResNet-50 transfer learning architecture was used on an EEG dataset to reduce computational complexity and improve model performance. The output was obtained as feature maps based on transfer learning to obtain high-level features.

Although EEG data is inherently temporal, the choice of ResNet-50 is justified by its ability to extract high-level spatial representations from multi-channel EEG matrices, which are often structured as 2D topographic feature maps (e.g., time-

frequency spectrograms or spatial electrode arrangements). Before input into ResNet-50, EEG signals were transformed into spatially structured 2D arrays that encode both electrode location and frequency content. This enables ResNet-50 to learn inter-channel spatial dependencies and cross-regional brain patterns that are clinically significant. Moreover, ResNet's residual connections mitigate vanishing gradients and enable deeper feature learning, which 1D-CNN or ConvLSTM alone may not fully capture. To handle temporal dependencies, ResNet-50 was combined with a Bi-GRU with an attention mechanism, allowing the model to learn both spatial and temporal EEG features in a hybrid manner.

2) BiGRU with Attention

This study developed an improved GRU model that fixes its structural overfitting issues and employs an attention mechanism module to extract features from the input. The objective is to predict cases of MDD in a timely and reliable manner. Using the input r_t and the hidden layer output g_{t-1}, the GRU network arrives at the output result g_t. The GRU internal unit uses an update gate u_t and a reset gate v_t. Despite this, the GRU network's core architecture is rather simple.

$$v_t = \sigma(e_v.[g_{t-1}, r_t]) \quad (12)$$

$$u_i = \sigma(e_u.[g_{t-1}, r_t]) \quad (13)$$

$$\bar{g}_t = \tanh\left(w_{\bar{g}}.[v_t * g_{t-1}, r_t]\right) \quad (14)$$

$$g_t = (1 - u_t) * g_{t-1} + u_t * \bar{g}_t \quad (15)$$

One common use of the sigmoid function is in binary classification problems; in this case, w is the weight matrix and σ is the activation function used to calculate the gate's output. The quantity of state information added to the current iteration may be regulated by adjusting the value of u_t, which depends on the previous one. A higher number suggests a greater amount of information is supplied. The reset gate v_t controls the amount of state-related data from the previous time step that affect the potential candidate set of the current state. The quantity of data from the previous iterations can be decreased by adjusting the reset gate to a lower value.

The attention technique is formulated in (16), where s_q is the total of the prospective embedding vectors. The matrix c represents the model factors, an element in $R^{n_G \times n_Q}$, where n_G is the size of the hidden layer, and n_Q is the size of the candidate embedding vectors. An effective method for illustrating the connection across the input r_t and the potential products is by using the attention q_t value:

$$q_t = \frac{exp(r_t c s_q)}{\sum_{i=1}^{n} exp(r_i c s_q)} \quad (16)$$

Currently, many studies disregard dimensional differences and handle the hidden layer's update state directly using the attention score q_t, excluding the update gate u_t. This work develops the Attention Update Gate GRU (AUG-GRU) by combining the attention system with the update gate.

To minimize the impact of irrelevant input on the model and scale attention scores while keeping the original input dimensions, the proposed AUG-GRU network adjusts the update gate across all of its dimensions. This adaptation allows AUG-GRU to enhance interest extraction more smoothly and effectively mitigate interference from sudden changes in focus. Equations (17) and (18) formulate the transformation from AUG-GRU to GRU, where u_t characterizes the original GRU, $u_t{}'$ denotes the redesigned AUG-GRU's update gate, and g_t and $g_{t-1}{}'$ represent the internal layer states of AUG-GRU.

$$u'_t = q_t * u_t \quad (17)$$

$$g'_t = (1 - u'_t).g'_{t-1} + u'_t * g'_t \quad (18)$$

The integration of the AUG-GRU mechanism allows the model to adaptively weight temporal segments of EEG sequences based on their relevance to MDD-related patterns. Unlike standard GRUs, which treat all time steps equally, AUG-GRU dynamically modulates the update gate using attention scores, allowing the network to prioritize time windows that exhibit symptom-related abnormalities such as emotional dysregulation or neural slowing. This selective focus enhances interpretability by identifying temporally significant EEG segments, which may correspond to clinically meaningful changes in symptom progression or cognitive state. As a result, AUG-GRU contributes not only to improved accuracy but also to greater transparency in how the model associates temporal dynamics with depressive symptomatology.

The extraction of patient depressed sequence patterns using a GRU model represents the most significant enhancement of the proposed system compared to conventional depression detection systems. This approach ensures the accurate extraction of user sequence characteristics by employing a GRU structure that incorporates an attention function (AUG-GRU), which can adjust more effectively to the evolving trends in patients' interests and symptoms related to MDD. The choice of Bi-GRU with an attention mechanism was guided by the need for a lightweight yet effective temporal modeling approach. Compared to Bi-LSTM, Bi-GRU has a simpler gating structure and fewer parameters, allowing training and reducing the risk of overfitting on limited EEG datasets while retaining the ability to model long-range dependencies. Although transformer encoders and Temporal Convolutional Networks (TCNs) offer powerful temporal modeling capabilities, they typically require larger datasets and higher computational resources. The Bi-GRU with attention balances complexity and performance, allowing the model to capture bidirectional temporal patterns in EEG sequences and selectively focus on relevant time steps, improving both accuracy and interpretability without imposing significant computational overhead.

3) Algorithm of the Proposed Model

The RAG-EEGNet algorithm described below is a sophisticated approach to MDD prediction and classification. It leverages the strengths of the ResNet-50 transfer learning architecture and then synergistically combines with a Bi-GRU with an attention updated gate.

```
Algorithm: RAG-EEGNet Model
Step 1: Load pre-trained ResNet-50 model
        from Deep Learning Toolbox
```

```
Step 2: Prepare a Bi-GRU-based sequence
        model - Initialize with default
        parameters or custom settings
Step 3: Feature extraction using Boruta
Step 4: Feature selection using CSO
Step 5: Transfer learning with ResNet-50
Step 6: Sequence processing with Bi-GRU
        with Attention Mechanism
```

IV. RESULTS AND ANALYSIS

TensorFlow was used to develop the RAG-EEGNet model. This framework streamlines development, incorporating models such as GRU [42]. Table I details the experimental setup used to develop the model.

TABLE I. COMPUTING ENVIRONMENT

CPU	Intel i5
GPU	P-100
RAM	16GB
Language	Python
Platform	Tensorflow

A. Dataset Description

The dataset in [43] was collected using electrodes for brain wave monitoring during sleep or mental activities, recording 120 MDD patients' EEGs. It has 1149 columns for age, gender, medication history, symptom intensity, and comorbidities. Timestamps, EEG channel data, clinical annotations, and preprocessing details are important columns. The dataset includes seven main disorder classes and 12 specific ones, such as depression and healthy. EEG annotations mark MDD-related features and preprocessing filters to remove artifacts and improve data quality. Metadata includes sample rate, recording time, and electrode montage. This CSV-format dataset helps psychiatric researchers identify MDD EEG biomarkers and improve diagnostic tools, revealing neural mechanisms.

B. Performance Evaluation

The evaluation metrics used provide a complete picture of the model's performance. Accuracy measures a model's data fit. Recall measures the model's capacity to discover actual positives. Precision focuses on reducing false positives. F1-score balances recall and accuracy to measure performance. These measurements reveal the model's strengths and limitations across several scenarios [44]:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (19)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (20)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TN} + \text{FP})} \quad (21)$$

$$\text{F1} - \text{score} = \frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision} + \text{Recall})} \quad (23)$$

where the acronyms TP, TN, FP, and FN stand for True Positives, True Negatives, False Positives, and False Negatives, respectively.

C. Preprocessing and Feature Selection Using Boruta and CSO

The Boruta and CSO algorithms were used to extract and select features, capture essential MDD EEG signal components, and optimize model performance. Table II shows the mean distributions of the depressive disorder EEG channels, and Table III shows the key attributes selected by Boruta and CSO.

TABLE II. CHANNEL MEAN DISTRIBUTION FOR EEG SIGNALS

Channel	Value	Channel	Value
FP1	3.681332	C4	3.041254
FP2	3.840776	T4	3.387337
F7	3.586139	T5	2.435776
F3	3.380585	P3	2.736861
Fz	3.279946	Pz	2.799705
F4	3.411508	P4	2.865227
F8	3.784580	T6	2.966232
T3	4.228335	O1	4.257993
C3	3.165926	O2	3.868594
Cz	3.080554		

TABLE III. MOST PROMINENT FEATURE ATTRIBUTES SELECTED USING BORUTA AND CSO

AB.D.beta.q.T6	AB.A.delta.q.T6	AB.A.delta.r.O1
AB.D.beta.e.Fz	AB.D.beta.r.O1	AB.D.beta.g.F8
AB.D.beta.d.F3	AB.A.delta.m.T5	COH.C.alpha.b.FP2.d.F3
COH.B.theta.b.FP2.h.T3	COH.B.theta.h.T3.j.Cz	AB.D.beta.h.T3
AB.D.beta.c.F7	AB.D.beta.f.F4	AB.D.beta.p.P4

D. Ablation Study on Feature Selection Techniques

An ablation study was carried out to evaluate the individual and combined contributions of Boruta and CSO to the overall performance of the model. Three experimental settings were considered: (i) Boruta-only, (ii) CSO-only, and (iii) the combined Boruta+CSO pipeline. The results in Table IV demonstrate that Boruta alone enhanced feature relevance, while CSO improved subset compactness. However, the combined approach delivered the highest classification performance in all metrics, indicating that Boruta and CSO function complementarily, with Boruta providing all-relevant feature filtering and CSO refining the selection to optimize predictive accuracy.

TABLE IV. ABLATION STUDY ON BORUTA AND CSO FOR FEATURE SELECTION

Feature selection method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Boruta only	96.72	97.10	95.86	96.47	96.12
CSO only	97.38	98.24	96.10	97.16	97.01
Boruta+CSO (Proposed)	99.01	100.00	99.98	99.12	99.00

The impact of the Boruta and CSO-based feature selection pipeline on computational efficiency was evaluated by recording the training time and memory usage of the model before and after feature reduction. The original EEG dataset contained 1149 features, which were reduced to 15 key ones through the combined selection process. As shown in Table V,

this dimensionality reduction achieved a 53% decrease in training time and a 42% reduction in memory consumption, while maintaining model performance. These findings demonstrate the computational advantages of the feature selection pipeline, enhancing scalability for real-time or resource-limited deployment.

TABLE V. EFFECT OF FEATURE SELECTION ON COMPUTATIONAL COMPLEXITY

Configuration	Feature count	Training time (s/epoch)	Memory usage (MB)
Before feature selection	1149	2.74	615
After Boruta+CSO	15	1.29	357

E. RAG-EEGNet Performance

Table VI shows epochs, parameter combinations, and model setups for RAG-EEGNet training on the EEG Disorder Dataset for MDD detection. To maximize training efficacy, the model continually modified internal parameters to identify patterns and extract key information. Cross-epoch accuracy and loss metrics show the model's performance progression, verifying its MDD identification and classification accuracy. Figure 3 shows a robust accuracy over 20 epochs.

TABLE VI. RAG-EEGNET MODEL-TRAINING PARAMETERS

	Parameter used	Value
1	Training epochs	20
2	Batch size	32
3	Kernel size	5
4	Optimizer	Adam
5	Learning rate	0.001
6	Drop out	0.50
7	Early stopping	Yes
8	Reduce LR	Yes
9	Data shuffle	True
10	Loss function	Sparse_Categorical_Cross Entropy,
11	Activation function	ReLU, Softmax

Training RAG-EEGNet for 20 epochs improved its MDD prediction and classification using EEG data, having initially 93% training and 96% validation accuracy. Through 20 epochs, these metrics improved to 98% training accuracy and 98.59% validation accuracy. The final training loss was 0.0718 and validation loss 0.0358, indicating strong convergence and low errors. RAG-EEGNet's high accuracy and low loss values throughout training demonstrate its robustness and effectiveness in detecting MDD. Figure 3 shows the loss and accuracy of RAG-EEGNet on the EEG Disorder Dataset over 20 epochs. The first graph shows decreasing training and validation loss, indicating convergence and low error. The second graph shows high and consistent accuracy throughout training, approaching near-perfect by the final epochs. The model's MDD classification accuracy is supported by the training and validation curves' close alignment, indicating robust generalization without overfitting.

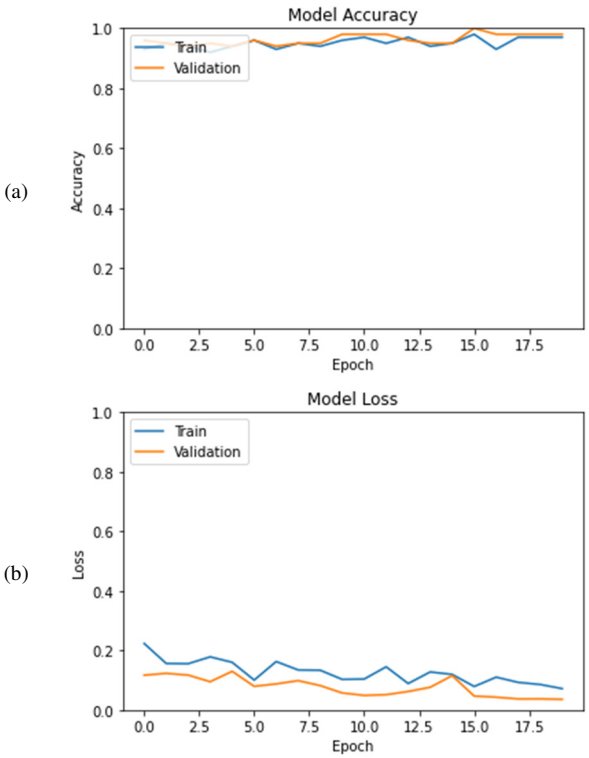

Fig. 3. Training accuracy (a) and loss (b) of RAG-EEGNet on the EEG Disorder Dataset for 20 epochs.

A subject-independent 10-fold cross-validation protocol was implemented to ensure reliable evaluation and prevent data leakage. Each fold consisted of entirely disjoint subjects across training, validation, and test sets, preventing overlap of EEG samples from the same individual. Accuracy, precision, recall, F1-score, and ROC-AUC were averaged across all folds. Exceptionally high values, including 100% precision and 99.98% recall, remained consistent with low standard deviation, reflecting the strong generalizability and robustness of the RAG-EEGNet model.

Class-specific performance and potential imbalance issues were evaluated using a confusion matrix and class-wise metrics for the two primary categories: MDD and HC. As shown in Table VII, the model demonstrates consistent performance across both classes. High precision and recall values indicate balanced classification behavior, with no bias toward the majority class. These results suggest robust handling of class distribution and effective generalization across both affected and control populations.

TABLE VII. CLASS-WISE METRICS

Class	Precision (%)	Recall (%)	F1-Score (%)
MDD	100.00	99.96	99.98
HC	99.94	100.00	99.97

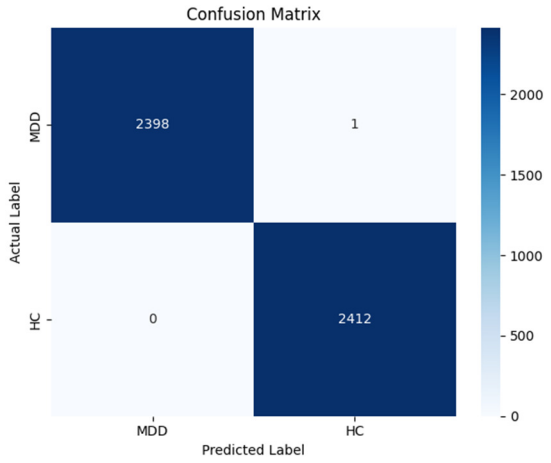

Fig. 4. Confusion matrix for RAG-EEGNet.

The feasibility of deploying RAG-EEGNet in wearable EEG systems was evaluated by analyzing the model size and inference latency. The model contains approximately 23.6 million parameters, largely from ResNet-50, with minimal overhead from the Bi-GRU and attention layers. Inference latency was recorded at ~47 ms per EEG segment on an Intel i5 CPU and ~12 ms on a mobile-class GPU (NVIDIA Jetson Nano). These findings indicate the suitability for real-time MDD screening in resource-constrained wearable devices, with further gains possible through pruning or quantization.

F. Performance Comparison with State-of-the-Art (SOTA) Methods

Table VIII compares RAG-EEGNet with SOTA models for MDD. The balanced F1-score confirms that RAG-EEGNet is the best-performing model in this evaluation, ensuring a reliable detection of MDD.

TABLE VIII. CLASSIFICATION COMPARISON FOR MDD DIAGNOSIS WITH SOTA MODELS USING EEG DATA

Model	Accuracy	Precision	Recall	F1-score
MRCNN-LSTM [40]	95.38	95.49	95.84	95.66
MRCNN-RSE [40]	98.47	98.63	98.66	98.65
EEGNet [45]	90.07	90.55	90.83	90.65
RAG-EEGNet (Proposed)	99.01	100	99.98	99.12

G. Statistical Significance Testing Between RAG-EEGNet and Baseline Models

Performance improvements of the proposed RAG-EEGNet model were validated using paired t-tests against SOTA models. Each model underwent 10 independent training and evaluation runs using identical data splits and evaluation protocols. The metrics compared included accuracy, precision, recall, and F1-score, with the null hypothesis assuming no significant difference. As shown in Table IX, all resulting p-values were below 0.01, confirming statistically significant performance gains at 99% confidence level, attributable to the model's architectural and algorithmic enhancements.

TABLE IX. PAIRED T-TEST P-VALUES COMPARING RAG-EEGNET WITH SOTA MODELS (N = 10 RUNS)

Model	Accuracy *p*-value	Precision *p*-value	Recall *p*-value	F1-score *p*-value
1DCNN-LSTM [40]	0.0035	0.0031	0.0029	0.0030
2DCNN-LSTM [40]	0.0019	0.0024	0.0020	0.0018
EEGNet [45]	0.0011	0.0013	0.0010	0.0015

The model was evaluated using subject-independent cross-validation and diverse participant subsets within the EEG dataset to assess generalizability. Although this study did not perform external validation on additional datasets or independent cohorts, these results indicate strong robustness, supported by consistent cross-fold performance and minimal variance. However, validating RAG-EEGNet across multi-site datasets or demographically diverse populations is necessary to confirm its scalability and clinical applicability in real-world environments.

H. ROC Analysis for RAG-EEGNet on MDD Using EEG Data

Figure 5 displays the ROC curve for RAG-EEGNet. With an AUC of 0.99, the model shows exceptional precision in distinguishing between MDD-positive and negative cases. The closeness of the curve to the upper left corner shows a very sensitive system with few false positives and negatives, which affirms the model's reliability for clinical MDD prediction.

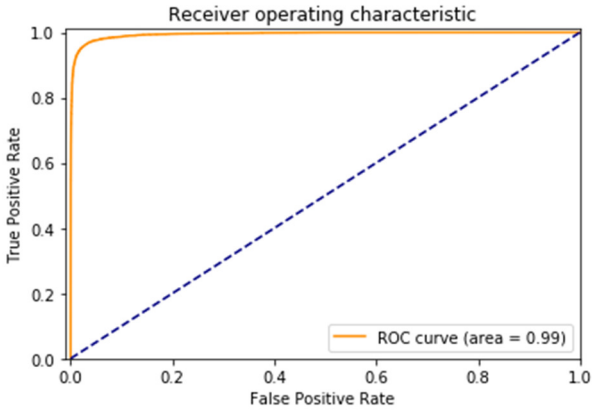

Fig. 5. AUC-ROC curve for RAG-EEGNet on an EGG dataset.

V. CONCLUSION AND SUMMARY

This study presented an advanced method for MDD prediction using EEG data, called RAG-EEGNet. The proposed model combines ResNet-50 for feature extraction, Bi-GRU with attention for capturing temporal dependencies, and feature selection using the Boruta and CSO algorithms. This approach significantly improved MDD detection accuracy, achieving 99.01% accuracy, 100% precision, 99.98% recall, and a 99.12% F1-score. The ROC-AUC of 0.99 further validates its robust performance, surpassing previous SOTA models. These results highlight its potential as a reliable tool for MDD diagnosis and management in clinical settings, with low false positives and high sensitivity. Future work will focus on optimizing the model for real-time processing and integration into wearable EEG devices, enhancing its practicality for continuous monitoring and instant clinical diagnosis.

REFERENCES

[1] "Depressive disorder (depression)," *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/depression.

[2] L. Cui *et al.*, "Major depressive disorder: hypothesis, mechanism, prevention and treatment," *Signal Transduction and Targeted Therapy*, vol. 9, no. 1, Feb. 2024, Art. no. 30, https://doi.org/10.1038/s41392-024-01738-y.

[3] J. D. Smith *et al.*, "Collaborative care for depression management in primary care: A randomized roll-out trial using a type 2 hybrid effectiveness-implementation design," *Contemporary Clinical Trials Communications*, vol. 23, Sep. 2021, Art. no. 100823, https://doi.org/10.1016/j.conctc.2021.100823.

[4] D. Watts, R. F. Pulice, J. Reilly, A. R. Brunoni, F. Kapczinski, and I. C. Passos, "Predicting treatment response using EEG in major depressive disorder: A machine-learning meta-analysis," *Translational Psychiatry*, vol. 12, no. 1, Aug. 2022, Art. no. 332, https://doi.org/10.1038/s41398-022-02064-z.

[5] "Major Depression - National Institute of Mental Health (NIMH)," *National Institute of Mental Health*. https://www.nimh.nih.gov/health/statistics/major-depression.

[6] E. Zhang, Z. Huang, Z. Zang, X. Qiao, J. Yan, and X. Shao, "Identifying circulating biomarkers for major depressive disorder," *Frontiers in Psychiatry*, vol. 14, Aug. 2023, https://doi.org/10.3389/fpsyt.2023.1230246.

[7] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal EEG data for depression recognition," *Information Fusion*, vol. 59, pp. 127–138, Jul. 2020, https://doi.org/10.1016/j.inffus.2020.01.008.

[8] R. Schaakxs, H. C. Comijs, R. C. van der Mast, R. A. Schoevers, A. T. F. Beekman, and B. W. J. H. Penninx, "Risk Factors for Depression: Differential Across Age?," *The American Journal of Geriatric Psychiatry*, vol. 25, no. 9, pp. 966–977, Sep. 2017, https://doi.org/10.1016/j.jagp.2017.04.004.

[9] A. U. Patil *et al.*, "Review of EEG-based neurofeedback as a therapeutic intervention to treat depression," *Psychiatry Research: Neuroimaging*, vol. 329, Mar. 2023, Art. no. 111591, https://doi.org/10.1016/j.pscychresns.2023.111591.

[10] D. F. Santomauro *et al.*, "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic," *The Lancet*, vol. 398, no. 10312, pp. 1700–1712, Nov. 2021, https://doi.org/10.1016/S0140-6736(21)02143-7.

[11] R. H. Salk, J. S. Hyde, and L. Y. Abramson, "Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms," *Psychological Bulletin*, vol. 143, no. 8, pp. 783–822, 2017, https://doi.org/10.1037/bul0000102.

[12] W. J. Yan, Q. N. Ruan, and K. Jiang, "Challenges for Artificial Intelligence in Recognizing Mental Disorders," *Diagnostics*, vol. 13, no. 1, Jan. 2023, Art. no. 2, https://doi.org/10.3390/diagnostics13010002.

[13] B. Pandurangaraju, B. Vijayalakshmi, S. A. Kumar, K. U. K. Reddy, S. Karimullah, and F. Shaik, "Optimizing Natural Language Understanding: A Comprehensive Study of Python's Top Libraries," in *2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, Hamburg, Germany, Oct. 2024, pp. 559–563, https://doi.org/10.1109/ICCCMLA63077.2024.10871883.

[14] H. Wang *et al.*, "Convergent and divergent cognitive impairment of unipolar and bipolar depression: A magnetoencephalography resting-state study," *Journal of Affective Disorders*, vol. 321, pp. 8–15, Jan. 2023, https://doi.org/10.1016/j.jad.2022.09.126.

[15] Y. Li, L. Qian, G. Li, and Z. Zhang, "Frequency specificity of aberrant triple networks in major depressive disorder: a resting-state effective connectivity study," *Frontiers in Neuroscience*, vol. 17, Jun. 2023, https://doi.org/10.3389/fnins.2023.1200029.

[16] Z. Shen *et al.*, "Aberrated Multidimensional EEG Characteristics in Patients with Generalized Anxiety Disorder: A Machine-Learning Based Analysis Framework," *Sensors*, vol. 22, no. 14, Jan. 2022, Art. no. 5420, https://doi.org/10.3390/s22145420.

[17] A. Khosla, P. Khandnor, and T. Chand, "Automated diagnosis of depression from EEG signals using traditional and deep learning approaches: A comparative analysis," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 108–142, Jan. 2022, https://doi.org/10.1016/j.bbe.2021.12.005.

[18] S. F. Begum *et al.*, "Emotion Recognition from Physiological Signals Using Ensembled Machine Learning Strategy," in *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)*, Bengaluru, India, May 2024, pp. 1–8, https://doi.org/10.1109/ICECCC61767.2024.10593862.

[19] C. T. Lin *et al.*, "Forehead EEG in Support of Future Feasible Personal Healthcare Solutions: Sleep Management, Headache Prevention, and Depression Treatment," *IEEE Access*, vol. 5, pp. 10612–10621, 2017, https://doi.org/10.1109/ACCESS.2017.2675884.

[20] L. Schmaal *et al.*, "Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group," *Molecular Psychiatry*, vol. 22, no. 6, pp. 900–909, Jun. 2017, https://doi.org/10.1038/mp.2016.60.

[21] G. Okada *et al.*, "Verification of the brain network marker of major depressive disorder: Test-retest reliability and anterograde generalization performance for newly acquired data," *Journal of Affective Disorders*, vol. 326, pp. 262–266, Apr. 2023, https://doi.org/10.1016/j.jad.2023.01.087.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[23] F. Zeng, R. Tang, and Y. Wang, "User Personalized Recommendation Algorithm Based on GRU Network Model in Social Networks," *Mobile Information Systems*, vol. 2022, no. 1, 2022, Art. no. 1487586, https://doi.org/10.1155/2022/1487586.

[24] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta – A System for Feature Selection," *Fundamenta Informaticae*, vol. 101, no. 4, pp. 271–285, Jul. 2010, https://doi.org/10.3233/FI-2010-288.

[25] A. H. Gandomi, X. S. Yang, and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Engineering with Computers*, vol. 29, no. 1, pp. 17–35, Jan. 2013, https://doi.org/10.1007/s00366-011-0241-y.

[26] R. Wang, Y. Hao, Q. Yu, M. Chen, I. Humar, and G. Fortino, "Depression Analysis and Recognition Based on Functional Near-Infrared Spectroscopy," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 12, pp. 4289–4299, Sep. 2021, https://doi.org/10.1109/JBHI.2021.3076762.

[27] Y. Dong and X. Yang, "A hierarchical depression detection model based on vocal and emotional cues," *Neurocomputing*, vol. 441, pp. 279–290, Jun. 2021, https://doi.org/10.1016/j.neucom.2021.02.019.

[28] B. Ay *et al.*, "Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals," *Journal of Medical Systems*, vol. 43, no. 7, May 2019, Art. no. 205, https://doi.org/10.1007/s10916-019-1345-y.

[29] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent Neural Network Regularization." arXiv, Feb. 19, 2015, https://doi.org/10.48550/arXiv.1409.2329.

[30] S. Zhang *et al.*, "The Combination of a Graph Neural Network Technique and Brain Imaging to Diagnose Neurological Disorders: A Review and Outlook," *Brain Sciences*, vol. 13, no. 10, Oct. 2023, Art. no. 1462, https://doi.org/10.3390/brainsci13101462.

[31] S. Y. Kim, "Personalized Explanations for Early Diagnosis of Alzheimer's Disease Using Explainable Graph Neural Networks with Population Graphs," *Bioengineering*, vol. 10, no. 6, Jun. 2023, Art. no. 701, https://doi.org/10.3390/bioengineering10060701.

[32] E. N. Pitsik *et al.*, "The topology of fMRI-based networks defines the performance of a graph neural network for the classification of patients with major depressive disorder," *Chaos, Solitons & Fractals*, vol. 167, Feb. 2023, Art. no. 113041, https://doi.org/10.1016/j.chaos.2022.113041.

[33] M. K. Myee, R. D. C. Rebekah, T. Deepa, G. D. Zion, and K. Lokesh, "Detection of Depression in Social Media Posts using Emotional Intensity Analysis," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16207–16211, Oct. 2024, https://doi.org/10.48084/etasr.7461.

[34] J. Wang *et al.*, "Automatic Diagnosis of Major Depressive Disorder Using a High- and Low-Frequency Feature Fusion Framework," *Brain Sciences*, vol. 13, no. 11, Nov. 2023, Art. no. 1590, https://doi.org/10.3390/brainsci13111590.

[35] S. Venkatapathy *et al.*, "Ensemble graph neural network model for classification of major depressive disorder using whole-brain functional connectivity," *Frontiers in Psychiatry*, vol. 14, Mar. 2023, https://doi.org/10.3389/fpsyt.2023.1125339.

[36] T. Zhao and G. Zhang, "Detecting Major Depressive Disorder by Graph Neural Network Exploiting Resting-State Functional MRI," in *Neural Information Processing*, 2023, pp. 255–266, https://doi.org/10.1007/978-981-99-1642-9_22.

[37] M. Xia, Y. Zhang, Y. Wu, and X. Wang, "An End-to-End Deep Learning Model for EEG-Based Major Depressive Disorder Classification," *IEEE Access*, vol. 11, pp. 41337–41347, 2023, https://doi.org/10.1109/ACCESS.2023.3270426.

[38] Z. Xia, Y. Fan, K. Li, Y. Wang, L. Huang, and F. Zhou, "DepressionGraph: A Two-Channel Graph Neural Network for the Diagnosis of Major Depressive Disorders Using rs-fMRI," *Electronics*, vol. 12, no. 24, Dec. 2023, Art. no. 5040, https://doi.org/10.3390/electronics12245040.

[39] B. H. Bhavani, M. Sreenatha, and N. C. Kundur, "Diagnosis and Classification of Depressive Disorders using ML and DL Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21383–21389, Apr. 2025, https://doi.org/10.48084/etasr.10017.

[40] Y. Xu *et al.*, "Depressive Disorder Recognition Based on Frontal EEG Signals and Deep Learning," *Sensors*, vol. 23, no. 20, Jan. 2023, Art. no. 8639, https://doi.org/10.3390/s23208639.

[41] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," *Neural Computing and Applications*, vol. 34, no. 1, pp. 721–744, Jan. 2022, https://doi.org/10.1007/s00521-021-06426-4.

[42] A. Gulli and S. Pal, *Deep Learning with Keras*. Packt Publishing Ltd, 2017.

[43] "EEG Psychiatric Disorders Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/shashwatwork/eeg-psychiatric-disorders-dataset.

[44] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." arXiv, Oct. 11, 2020, https://doi.org/10.48550/arXiv.2010.16061.

[45] B. Liu, H. Chang, K. Peng, and X. Wang, "An End-to-End Depression Recognition Method Based on EEGNet," *Frontiers in Psychiatry*, vol. 13, Mar. 2022, https://doi.org/10.3389/fpsyt.2022.864393.

AUTHORS PROFILE

Udutala Mahender is a PhD scholar at the Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India, Having 13 Years of Teaching experience in various engineering colleges. He earned his M.Tech. (CSE) from JNTU University, Telangana. His research interests involve Artificial Intelligence, Networks, and Data Analytics.

Dr. S. Arivalagan is an Assistant Professor in the Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India. He has 18 years of teaching and research experience. He has expertise in Computer Science, Information Systems, Image Processing, Computer Networks, Data Structures, and Database Management Systems.

Dr. V. Sathiyasuntharam is a Professor and Head of the Department of Computer Science and Engineering in the CMR Engineering College, Kandlakoya, Medchal, Hyderabad, Telangana, India. He has 17 and 10 years of teaching and research experience, respectively. He earned his B.Tech (IT) from Madras University, his M.E (CSE) from Anna University, and his Ph.D. in Computer Science and Engineering from Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.

Dr. P. Sudhakar is an Associate Professor in the Department of Computer science and Engineering, Annamalai University, Tamilnadu, India. He has 23 years of research experience. He has expertise in Computer Science, Information Systems, Image Processing, Computer Networks, Data Structures, and Data Analytics.