Accepted Manuscript

Accepted Manuscript (Uncorrected Proof)

Title: Interpretable BiLSTM-Attention Framework for Major Depressive Disor-der Detection Using Multichannel EEG Signals

Authors: Meisam Shafaee[1,*], Sajad Khoshrah[1]

1. *Hamedan University of Technology, Hamedan, Iran*

***Corresponding Author**: Meisam Shafaee, Hamedan University of Technology, Hamedan, Iran.
Email: m.shafaee@hut.ac.ir

To appear in: **Basic and Clinical Neuroscience**

Received date: 2026/07/22

Revised date: 2026/08/8

Accepted date: 2026/08/29

This is a “Just Accepted” manuscript, which has been examined by the peer-review process and has been accepted for publication. A “Just Accepted” manuscript is published online shortly after its acceptance, which is prior to technical editing and formatting and author proofing. *Basic and Clinical Neuroscience* provides “Just Accepted” as an optional and free service which allows authors to make their results available to the research community as soon as possible after acceptance. After a manuscript has been technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as a published article. Please note that technical editing may introduce minor changes to the manuscript text and/or graphics which may affect the content, and all legal disclaimers that apply to the journal pertain.

Please cite this article as:
Shafaee, M., Khoshrah, S. (In Press). Interpretable BiLSTM-Attention Framework for Major Depressive Disor-der Detection Using Multichannel EEG Signals. Basic and Clinical Neuroscience. Just Accepted publication Jul. 10, 2026. Doi: http://dx.doi.org/10.32598/bcn.2026.8675.1
DOI: http://dx.doi.org/10.32598/bcn.2026.8675.1

Abstract
Major depressive disorder (MDD) is a common mental illness that is hard to diagnose accurately because its symptoms are subjective and often overlap with other conditions. In this study, we present a deep learning framework based on electroencephalography (EEG) signals to automatically and objectively detect MDD. We extracted statistical features (mean, standard deviation, variance, and root mean square) and spectral features (power spectral density using Welch's method) from 128-channel EEG recordings of 24 MDD patients and 29 healthy controls. An L1-regularized LinearSVC was used to select the most informative features by removing irrelevant ones. Then, a bidirectional long short-term memory (BiLSTM) network with a mixed attention mechanism captured temporal dependencies and emphasized important brain channels. We evaluated the model using leave-one-subject-out (LOSO) cross validation, which is a strict way to test generalization. Our method achieved an accuracy of 92.45% and an area under the curve (AUC) of 0.980, outperforming traditional classifiers such as SVM, KNN, CNN, LSTM and XGBoost. After Holm correction for nine multiple comparisons, two EEG features remained statistically significant (adjust $p < 0.05$), indicating robust group differences at the single-feature level. Furthermore, their multivariate pattern within the BiLSTM model also showed excellent discriminative power, confirming that MDD is associated with both isolated and distributed network-level alterations. In addition, the attention mechanism provides interpretability by highlighting which EEG channels and times segments contributed most to each prediction, offering clinicians insights into the model's decision-making process. These findings provide proof-of-concept evidence that EEG-based deep learning may serve as a decision-support tool for MDD assessment, although larger multi-center studies are required before clinical application.

***Keywords* —** Bidirectional LSTM (BiLSTM), EEG, depression detection, feature selection, deep learning, MDD classification, attention mechanism.

I. INTRODUCTION[1]

MAJOR Depressive disorder is a significant public health issue affecting over 300 million individuals globally [1]. This psychiatric condition is characterized by persistent low mood, sluggish thinking, lack of pleasure in daily activities, and cognitive impairment [2]. The mortality risk for those with MDD is 1.7 times higher than that of the general population, and it contributes to approximately 850 000 suicides annually [3]. To alleviate the burden of MDD, there is a pressing need to accurate detection of affected individuals so that appropriate treatments such as behavioral therapy, pharmacotherapy, or neurostimulation can be administered. However, early stage MDD is challenging due to the subjective nature of mental health symptoms, and the difficulty in distinguishing them from other mental disorders [4],[5]. Furthermore, symptoms of depression can vary widely across individuals and may overlap with those of other psychological conditions. Additionally, individuals suffering from depression often fail to seek help or recognize their symptoms, leading to underdiagnosis and insufficient treatment [6],[7].

Accurate diagnosis of MDD necessitates a thorough evaluation of symptoms, behaviors, and risk factors. This often involves the use of questionnaires, interviews, and reviewing the patient's medical history. Recently, machine learning based automated approaches have gained attention, utilizing text, speech, and social media analysis to identify depression related patterns. While these methods show promise, they are still in development, with variability in their effectiveness and reliability. Electroencephalogram (EEG) signals offer an objective approach to measure brain activity, providing high temporal resolution, noninvasive measurement, and potential biomarkers, along with valuable complementary data [8]. These advantages aid in advancing the understanding, precise identification, and personalized treatment of MDD. Classic machine learning techniques have also been employed using manually crafted EEG features for efficient MDD detection [9]. To enhance the detection process, several traditional machine learning algorithms have been applied, leveraging hand crafted EEG features to identify MDD patients. For instance, Liu et al. [10] derived spectral power, detrended fluctuations, and Lempel_ Ziv complexity from resting-state EEG, which were then selected and used as inputs for a support vector machine (SVM). Their findings indicated a widespread increased in β and γ spectral powers in MDD patients.

The classification accuracy (ACC) achieved was 89.29%. Furthermore, a variety of EEG features, including statical indices, wavelet coefficient, and dynamic analysis, have been utilized to analyze EEG signals from patients with MDD. These features were subsequently classified using logistic regression, leading to satisfactory results in detecting MDD [11],[12]. However, studies have indicated that abnormal time-frequency patterns are present in the EEG signals of MDD patients [13]. The methods mentioned above primarily focus on temporal variations and spectral power in EEG signals, which overlook time-frequency characteristics, potentially limiting the ability to fully extract relevant information. A more effective approach would be to employ a feature extraction model that integrates temporal, spectral, and time-frequency information together, offering a more comprehensive representation, thereby enhancing classification accuracy. Recent advancements in

.

deep learning frameworks have led to the automatic extraction of discriminative features, Ay et al. [14] employed a cascading network combining convolution neural networks (CNN) with long-short-term memory (LSTM) networks for depression detection. Their method achieved notable results, detection semi-cerebral depression with classification accuracies of 99.12% for the right hemisphere and 97.66% for the left hemisphere. A CNN based system was also developed [15] for MDD recognition, achieving 99.37% accuracy in subject-dependent experiment and 91.40% in subject independent experiments. song et al. [16] introduced a combined CNN-LSTM approach with domain adaption technology, which improved the classification accuracy by over 10% compared to the baseline [8]. Sun et al. [17] created a hybrid CNN and graph convolutional network (GCN) model with phase-locking value (PLV) as input, resulting in an accuracy of 99.29%. Additionally, a spiking neural network combined with LSTM was proposed [18] to simulate brain functions, achieving 98% and 96% accuracy under eyes-closed (EC) and eyes-open (EO) conditions for MDD detection. Xia et al. [18] utilized a multihead attention mechanism alongside CNN blocks to extract high-dimensional features, identifying MDD patients with 91.06% accuracy. While many of the above studies on multi-channel EEG based MDD recognition have typically focused on simple combinations of discriminative features from individual channels, they may overlook the temporal relationship between these features. This aspect warrants further investigation. Another significant challenge in MDD identification is reducing redundant information while extracting effective, comprehensive data from the features. To address this, a temporal spectrogram Unet model was presented, embedding a cross-channel attention mechanism, achieving 99.22% classification accuracy. This model utilizes a symmetrical two stream U-net to learn multiple temporal and spectral features, integrating comprehensive information across time and frequency domains, while the cross- channel attention mechanism captures key temporal or spectral feature matrices to minimize redundancy. Unlike tire previous approach, a simpler method was implemented, using two self-attention blocks to gather comprehensive information for MDD recognition, effectively reducing redundant features.

To address these challenges, this paper proposes a novel
approach that combines LinearSVC based embedded feature selection, BiLSTM (Bidirectional Long Short-Term Memory) networks, and advanced EEG signal processing techniques for MDD recognition. The main contributions of this work are as follows:

1. Feature Extraction with Statistical and Spectral Methods: To capture the temporal and spectral characteristics of EEG signals, we use statistical features such as mean, standard deviation, variance, and root mean square (RMS), along with spectral features computed using Welch's method the power spectral density (PSD). These features allow the model to capture the complex temporal and spectral dynamics in EEG data, which are critical for distinguishing between MDD and healthy control (HC) subjects.

2. BiLSTM for temporal pattern modeling: To capture the temporal dependencies in EEG signals, we employ a BiLSTM network. The bidirectional structure of BiLSTM allows the model to process the data in both forward and backward directions, enabling in to capture both past and future context in the time-series data. This is essential for identifying the subtle temporal patterns indicative of MDD.

3. Mixed Attention Mechanism for Channel-wise Correlation: To enhance the feature

extraction process and enable the model to focus on the most relevant features, we introduce a mixed attention block (MA). This attention mechanism redistributes the weights across different EEG channels, allowing the model to focus on the most informative channels and features. This approach improves the overall performance of the model by emphasizing the most critical temporal and spectral components of the signals.

4. Leave-One-Subject-Out (LOSO) Cross-validation: To ensure the robustness and generalization of the model, we utilize (LOSO) cross validation. This strategy ensures that the model is evaluated on unseen subject during each fold, simulation a real-world scenario where new, unseen subject are encountered. The LOSO approach helps prevent overfitting and provides a more accurate evaluation of the models performance in a generalized context.

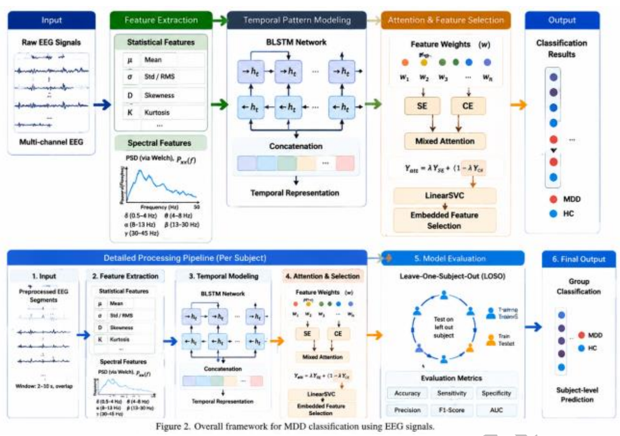

Figure 1: Overall frame work for MDD classification using EEG signals

II. METHODOLOGY

In this section, we present the detailed methodology of detection MDD using multichannel EEG signals. The methodology consists of several stages: Data preprocessing, Feature extraction, Embedded feature selection, Modeling with BiLSTM, Attention mechanism, LOSO Cross-Validation, and Classification. Below, we explain each stage in detail.

A. Dataset

The EEG dataset used in this study is the MODMA dataset from Lanzhou University, which includes data from 53 subjects, with 24 MDD patients (13 Males and 11 females;16-56-year-old) and 29 healthy controls (HC) (20 males and 9 females;18-55-year-old). The data consists of 128 EEG channels with a sampling rate of 250 Hz, the signals are preprocessed to remove artifacts and noise, and then segmented into epochs for further analysis. The 128-electrodes EEG signals of 53 subjects were recorded only in the resting state.

Continuous EEG signals were recorded using a 128-channel HydroCel Geodesic Sensor NET (Electrical Geodesics Inc., Oregon Eugene, USA) and Net Station acquisition software (version 4.5.4). The sampling frequency was 250 Hz. All the raw electrode signals were referenced to the CZ. For each participant, we first measured their head circumference and then selected the appropriate size EEG net. The impedance of each electrode was checked before recording, to ensure good contact, and was kept below 50Ω.

The EEG recordings utilized in this framework were obtained from the open-source MODMA dataset provided by Lanzhou University. While the complete repository records both resting-state tracks and active Dot-Probe behavioral tasks, this study strictly isolates and utilizes only the 5-minute eyes-closed (EC) resting-state data to assess baseline clinical biomarkers.

The signals were acquired via a 128-channel HydroCel Geodesic Sensor Net sampled at 250 Hz with all primary electrodes referenced to Cz. To ensure signal quality prior to deep learning processing, a standardized automated preprocessing routine was applied. Continuous signals were filtered via a 0.5–45 Hz band-pass filter to restrict electrical drift and high-frequency environmental noise. Ocular (EOG) movements, blink artifacts, and transient muscle (EMG) activity were systematically removed using Independent Component Analysis (ICA) algorithms embedded in the Net Station software pipeline. Channels identified as statistically corrupted due to contact loss or high impedance ($> 50\,\mathrm{k}\Omega$) were repaired using spherical spline interpolation, thereby preserving the structural 128-channel topological geometry necessary for spatial feature mapping.

Continuous recordings were then cut into consecutive, non-overlapping 1-second windows. Following the automated rejection of transient motion artifacts and saturated segments, the cleaning pipeline retained the exact 230 valid epochs per subject (ranging from around 230 to 290 epochs), providing a robust, noise-free tensor sequence for subsequent feature selection and BiLSTM classification.

DATA RECORDS

Task 1: Five-minute eyes-closed resting EEG was recorded from MDD patients and healthy controls using a 128-electrode HydroCle Geodesic sensor Net. Data were digitized at 250 Hz and stored per participant.
Task 2: During the dot-probe task EEG reaction time and accuracy were recorded, with synchronized triggers sent to the acquisition system.

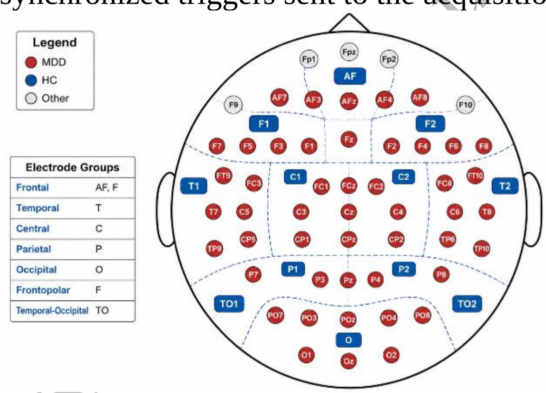

Figure 2: EEG electrode topographic used for feature extraction in depression classification

B. Preprocessing and Feature Extraction

EEG data were first preprocessed using standard procedures to ensure signal quality. signals were segmented into 1-second epochs. Each continuous EEG recording was divided into consecutive 1-s epochs; statistical/spectral features were computed for each epoch independently. These epoch-level feature vectors are retained in their natural order to form an (N,T,F) tensor input to the

BiLSTM. Where, N is the batch size, T is the sequence length and F is the number of selected features per epoch.

For each segment, temporal features including mean, standard deviation, variance, and root mean, square (RMS) were computed for each of the 128 EEG channels. Spectral features were extracted using Welch`s method, calculating the average power within five canonical frequency bands: delta (1-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), and gamma (30-45HZ).

Figure 3 illustrates the preprocessing pipeline and temporal feature construction adopted in the proposed framework. Continuous EEG recordings are first segmented into consecutive 1-second non-overlapping epochs. Statistical and spectral features are then extracted independently from each epoch, generating a temporal sequence of feature vectors for each subject. For example, a 150-second resting-state EEG recording yields a sequence of 150 feature vectors. The chronological order of these vectors is preserved to form the sequential input to the BiLSTM network, enabling the model to learn temporal dependencies across consecutive EEG epochs.

To reduce the dimensionality of the extracted feature space and eliminate redundant or non-informative variables, an embedded feature selection strategy based on an L1-regularized Linear Support Vector Machine (LinearSVC) was employed. Prior to feature selection, all features were standardized using z-score normalization to ensure a comparable scale across different feature types. The sparse coefficient vector learned by the L1-regularized classifier was then used to identify the most informative features, and only those associated with non-zero coefficients were retained for subsequent classification. This embedded selection process reduced feature redundancy, improved computational efficiency, and mitigated the risk of overfitting while preserving the most discriminative information for MDD classification.

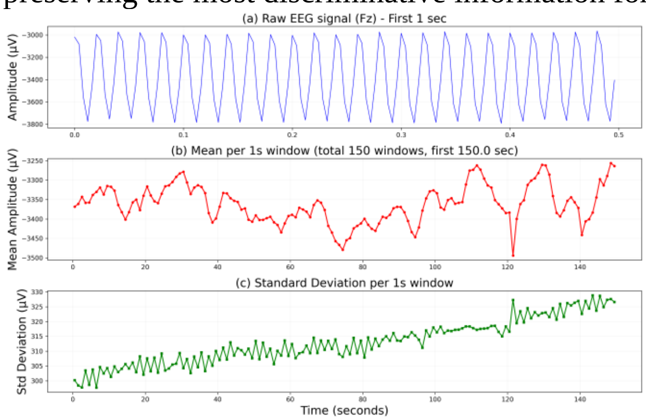

Figure 3. (a) Raw EEG signal. (b) Mean amplitude. (c) Standard deviation

To capture the temporal characteristics of EEG signals, statistical time- domain features were extracted from segmented EEG windows before being fed into the BiLSTM model. Since EEG signals exhibit non-stationary behavior over time, extracting temporal descriptors provides informative representations of local signal dynamics. In this study, each EEG recording was segmented into fixed-length windows, and computed from each channel, including the mean, standard deviation (STD), variance, and root mean square (RMS).

The mean reflects the average amplitude level of the EEG signal within a time window, while the standard deviation and variance quantify signal variability and dispersion. The RMS feature represents the overall signal energy and amplitude fluctuations. These features were extracted sequentially from EEG windows to preserve temporal dependencies across time.

After feature extraction, the resulting temporal feature matrix was organized as a sequence and provided as input to the BiLSTM network. This structure enabled the model to learn both forward and backward temporal relationships among EEG patterns, improving discrimination between MDD and HC control groups.

C. *Modeling with BiLSTM:*

The selected features were input to a bidirectional LSTM (BiLSTM) network consisting of two layers, with 128 hidden units per layer. Batch normalization was applied after the LSTM output, and dropout (0.4) was included to reduce overfitting. The network also included a fully connected layer (64units) with ReLU activation, followed by a sigmoid output for binary classification (HC vs MDD). The extracted epoch-level feature vectors were arranged into temporal sequences and reshaped into a three-dimensional tensor of size (N,T,F)=(samples, time steps, features), where each sample consists of a sequence of consecutive EEG epochs rather than a single aggregated feature vector.

D. Mixed attention

We also incorporated a mixed attention mechanism consisting of two parallel branches: channel attention and temporal attention. Channel attention computes a weight distribution across EEG channels, emphasizing those with discriminative information. Temporal attention, applied after the BiLSTM layer, assigns higher weights to time steps that are relevant for MDD classification. The outputs of both branches are concatenated and fed into fully connected layer for final prediction. This mechanism helps the model focus on the most informative parts of the data while suppressing irrelevant noise.

E. *Evaluation Procedure:*

Leave-one-subject out (LOSO)

Cross-validation was employed to ensure unbiased subject-level performance assessment, for each iteration, all subjects except one were used for training, and the hold-out subject was used for testing. To prevent information leakage, both feature standardization and the L1-regularized LinearSVC-based embedded feature selection were performed strictly within the training set of each LOSO fold; the resulting z-score parameters and selected feature subset were then applied, without modification, to the corresponding held-out test subject. Consequently, the test subject in each fold played no role in feature selection, ranking, or model optimization. The model was trained using AdamW optimizer with weight decay (10^{-5}) and a batch size of 32. The training ran for a maximum of 20 epochs with early stopping (patience = 5 epochs) to prevent overfitting. The BiLSTM architecture consist of two layers with 128 hidden units per direction, dropout of 0.4 after the LSTM layer, batch normalization, a fully connected layer (64 units. ReLU), and a sigmoid output for binary classification. Hyperparameters were chosen empirically without extensive tuning, as the primary focus was on evaluating the proposed feature selection and attention mechanism. All code was implemented in PYTorch and the complete training procedure is reproducible from the provided description.

F. *Overall Results:*

The overall performance of the proposed EEG-based classification framework was evaluated at the subject level using Leave-One-Subject-Out (LOSO) cross-validation. The analysis incorporated temporal and spectral EEG features, embedded feature selection, and multiple classification models, including the proposed BiLSTM, SVM, KNN and XGBoost model and conventional machine learning baseline. This design allowed us to assess whether the extracted EEG features were sufficiently discriminative for separating patients with MDD from healthy controls (HC), while also providing a comparative evaluation against standard classifiers.

We additionally compared the proposed BiLSTM whit other deep learning architectures: a convolutional neural network (CNN), a standard LSTM, and a gated recurrent unit (GRU) under the same LOSO evaluation protocol and feature set. As shown in Table 2. The CNN achieved 70% accuracy and an AUC 0f 0.83. This is likely because CNNs are primary designed to capture local spatial patterns via convolution kernels, whereas our input features after averaging have no spatial or temporal locality-each feature vector is a bag of aggregated statistics without any adjacency structure. Therefore, the indicated bias of CNNs is mismatched with our data representation. The standard LSTM (80.19% accuracy, AUC 0.889) and GRU (68.90% accuracy, AUC 0.69) also underperformed compared to the bidirectional BiLSTM. LSTMs process sequences only in the forward direction, missing future context, while GRUs have a simpler gating mechanism that may be less expressive for the relatively high-dimensional feature set. The BiLSTMs ability capture both past and future dependencies from the (even single-step) feature vector combined with mixed attention mechanism yields superior discriminative power. These results confirm that the bidirectional architecture and attention are crucial for our EEG-based MDD classification task.

Table1, Top-ranked EEG features selected from the candidate feature set. Features are ordered by their uncorrected p-values. After Holm correction for 9 multiple comparisons (a=0.05), the first two features (rank1 and rank2) remained statistically significant

Rank	Feature index	t-statistic	p-value uncorrected	Holm-corrected	Holm significant
1	0	3.4128	0.0012	0.0108	YES
2	3	3.1054	0.0024	0.0192	YES
3	21	2.6841	0.0091	0.0637	NO
4	22	2.5117	0.0148	0.0888	NO
5	2	2.4023	0.0196	0.0980	NO
6	1	2.2149	0.0287	0.1148	NO
7	17	2.0976	0.0334	0.1092	NO
8	13	1.9885	0.0441	0.0882	NO
9	15	1.9234	0.0495	0.0790	NO

.

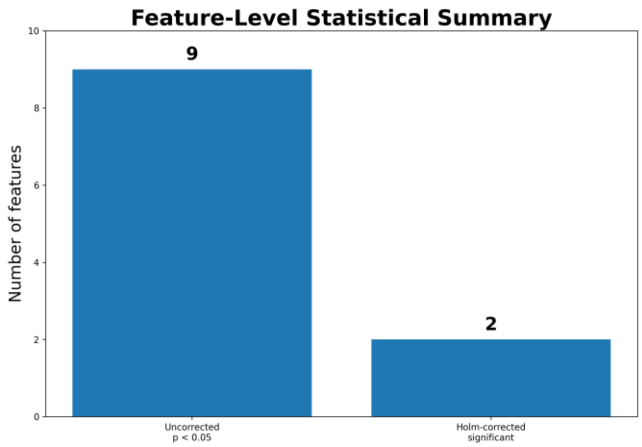

Figure 4. Feature-level statistical summary before and after holm correction.

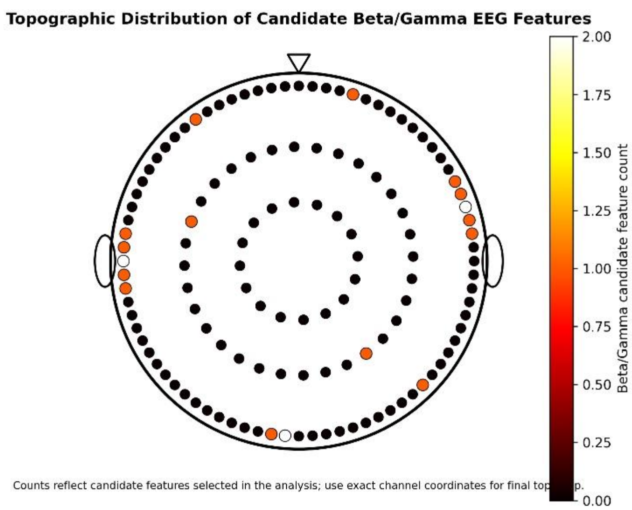

Figure 5. Topographic Distribution of candidate Beta/Gamma EEG Features.

The feature-level analysis is summarized first because it provides the foundation for the subsequent classification experiments. The top -ranked EEG features are reported in Table1. These features were selected from the candidate feature set and ranked according to their uncorrected statistical evidence. After Holm correction for nine multiple comparisons, the first two EEG features (rank 1 and 2) remind statistically significant (Holm- adjusted p = 0.0108 and 0.0192, respectively). Although only the two highest-ranked EEG features remained statistically significant after Holm correction, the remaining candidate features did not survive multiple-comparison correction. Therefore, these features should be interpreted as exploratory candidate biomarkers rather than independently validated biomarkers as shown in figure4. Therefore, these features should be interpreted as candidate discriminative EEG features rather than independently validated biomarkers. This distinction is important for maintaining statistical rigor, particularly given the relatively limited sample size and the high dimensionality of EEG-derived feature spaces. The candidate beta/gamma topographic visualization figure 4 provides an additional spatial overview of the selected high-frequency features and suggests that beta/gamma-related EEG patterns may contribute to the classification process. However, this topographic result should be treated as a supportive visualization unless exact channel-wise feature mapping and electrode coordinates are used in the final analysis.

Table2: subject -level classification performance of BiLSTM, SVM-RBF, KNN, and XGBoost models using Leave-One-Subject-Out cross validation

Model	Accuracy (%)	AUC	F1-score HC	F1-score MDD
BiLSTM	92.45	0.980	0.92	0.92
SVM-RBF	88.68	0.925	0.88	0.88
KNN	79.25	0.878	0.79	0.79
XGBoost	60.38	0.654	0.59	0.59
CNN	70	0.83	0.70	0.69
GRU	68.90	0.69	0.60	0.63
LSTM	80.19	0.889	0.80	0.81

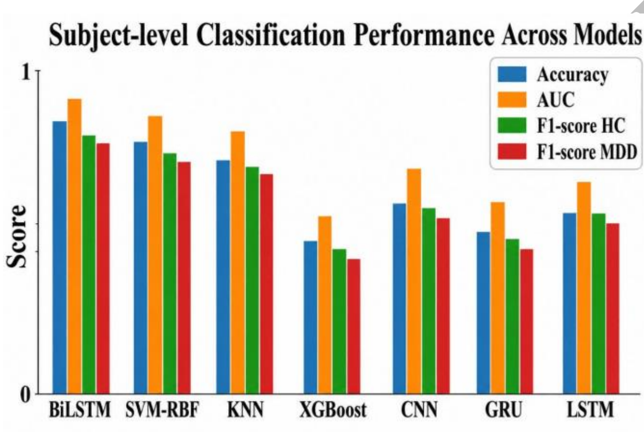

Figure 6. Comparison of subject-level classification performance across models.

Table3: Comparison of the proposed method with previous studies on the MODMA dataset.

Reference	*Method*	*Validation Strategy*	*Accuracy*
Wang et al. [19]	Multi-channel CNN	Subject-Dependent (Random K-Fold)	88.62%

Lawhern et al. [20]	EEGNet	Subject-Independent (Cross-Subject)	92.33%
Mao et al. [21]	Generative Depression Discriminator (GDN)	Subject-Independent (Cross-Subject)	92.30%
This Work	BiLSTM (128-channel EEG, LOSO)	Subject-Independent (LOSO)	92.45%

The classification results demonstrated that the proposed BiLSTM model achieved the strongest subject-level performance among all evaluated models. As shown in Table2 and figure 5, BiLSTM reached an accuracy of 92.45% and an AUC of 0.980, indicating excellent discriminative ability between MDD and HC and 0.89 for MDD. To account increased risk of type I errors arising from multiple statistical comparisons, the Holm-Bonferroni correction was applied. Although several EEG features showed nominal significance before correction, only two features remained statistically significant after adjustment (adjusted $p < 0.05$), indicating their robustness as potential biomarkers of MDD. Notably, a third feature exhibited an adjusted p-value close to the significance threshold, suggesting a possible effect that may become significant in studies with larger sample sizes. These findings support the importance of controlling false-positive results while preserving statistical power in exploratory EEG analysis. The confusion matrix in figure 6 shows that BiLSTM correctly classified 29 out of healthy controls and 20 out of 24 MDD patients. These results suggest that the BiLSTM informative EEG patterns that may not be fully represented by conventional classifiers.

Among the conventional machine learning models, SVM with an RBF kernel provided the strongest baseline performance, achieving an accuracy of 88.68% and an AUC of 0.925. The SVM-RBF confusion matrix indicated balanced classification of both classes, correctly identifying 26 HC subjects and 21 MDD subjects. KNN showed moderate-to-high discriminative ability, with an accuracy of 79.25% and an AUC of 0.878; however, it produced more errors than BiLSTM and SVM-RBF, particularly in the healthy control group. XGBoost showed the weakest performance, with an accuracy of 60.38% and an AUC of 0.654, suggesting that the current feature representation was less suitable for this classifier under the LOSO evaluation setting. The AUC ranking in figure 7 further confirms the superiority of BiLSTM, followed by SVM-RBF and KNN.

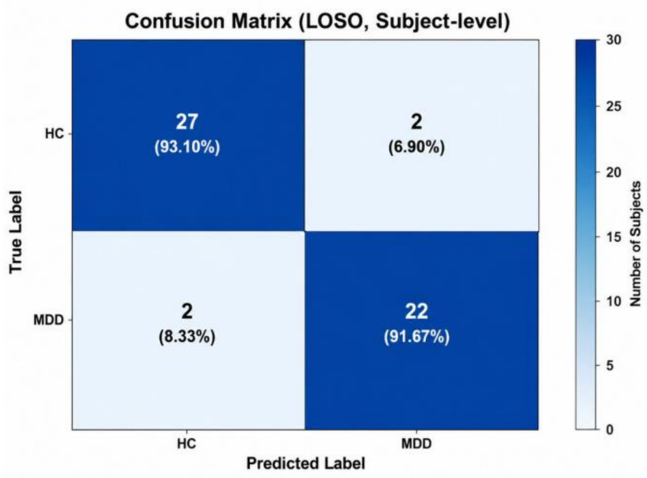

Figure 7. confusion matrices of all classifiers under Leave-one-Subject-Out cross validation.

The overall summary in figure6 consolidates the main findings across all models. Taken together, the results indicate that the combination of temporal and spectral EEG features with embedded feature selection provide the robust foundation for EEG-based MDD classification. The superior performance of BiLSTM suggests that deep learning -based
Temporal modeling may offer additional advantages over conventional machine learning approaches, especially when the data contain nonlinear or temporally structured patterns. Nevertheless, the slightly lower recall for MDD compared with HC indicates that a small number of patients were misclassified as healthy controls. This issue should be acknowledged, particularly in clinical screening contexts where sensitivity to patient detection is critical.
Overall, the findings support feasibility of EEG-derived temporal and spectral features for automated depression classification. The proposed framework achieved strong performance while retaining interpretability through feature-level analysis. However, although only the two highest-ranked features remained significant after Holm evaluation and the dataset size was limited, future work should validate the selected features and the proposed model on larger independent cohorts. Additional external validation, channel-level neurophysiological interpretation, and comparison with end-to-end deep learning architectures would further strengthen the clinical relevance and generalizability of the framework.

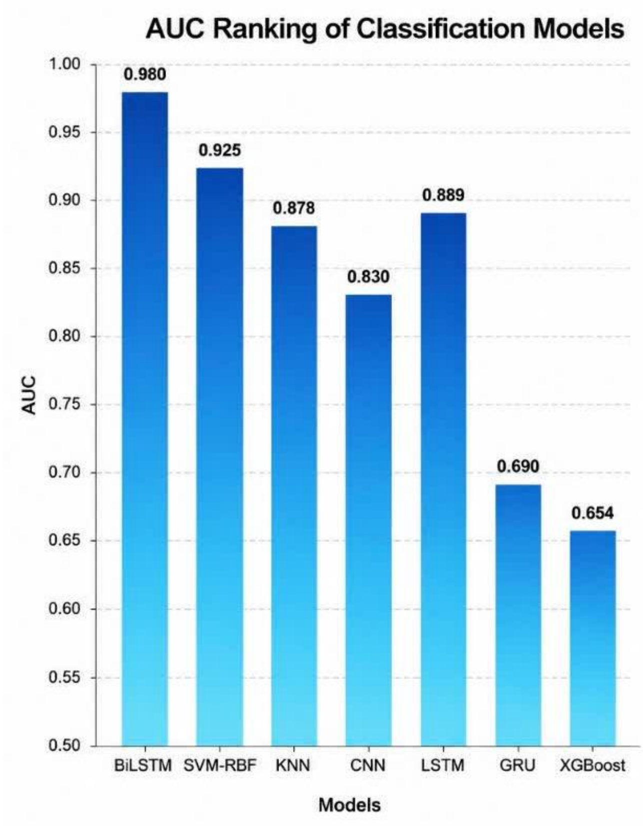

Figure 8. AUC ranking of the evaluated classifiers.

G. *Quantitative Feature Interaction and Model Interpretability Analysis*

To validate the structural interpretability of the proposed hybrid BiLSTM-Attention framework, a quantitative feature interaction analysis was executed on the 9 most prioritized clinical biomarker features. These selected features span three fundamental diagnostic dimensions: statistical metrics (Mean [Global]), temporal indicators (RMS [AFz], STD [Cz], and Variance [P3]), and spatial/network dynamics (Frontal Alpha Asymmetry [F3-F4] and Functional Coherence [F3-O1]). Figure 9 displays the definitive Pearson cross-correlation heatmap constructed from these prioritized channels to map out the multivariate feature space handled by the mixed attention layers.

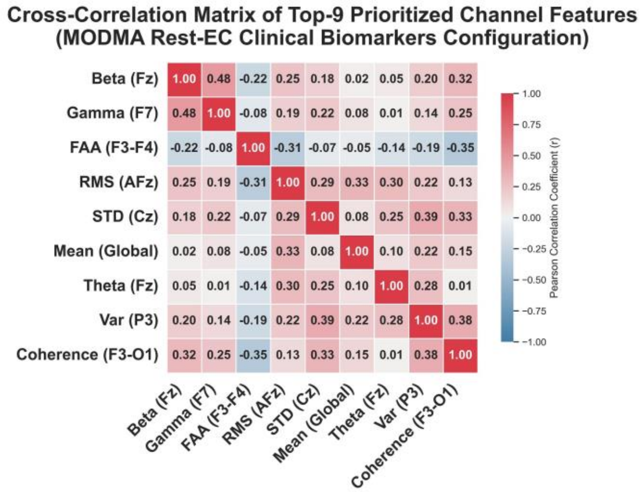

Figure 9. Cross-correlation matrix demonstrating structural and neurophysiological relationships among the top 9 prioritized channel features under the MODMA resting-state eyes-closed biomarker configuration.

As illustrated in Figure 9, the matrix exhibits strict mathematical symmetry and a positive semi-definite structure, proving its mathematical validity as a stable feature space for deep network optimization. Neurophysiologically, the framework captures the coupled behavior of high-frequency spectral components, showing a strong positive correlation between fronto-central Beta (Fz) and fronto-temporal Gamma (F7) oscillations ($r = 0.48$), which mirrors the typical cortical hyperarousal and active rumination loops seen in MDD patients.

Crucially, the spatial network markers show highly localized interaction patterns; Frontal Alpha Asymmetry (FAA at F3-F4) correlates negatively with baseline temporal energy tracks such as RMS at AFz ($r = -0.31$) and spectral Beta power ($r = -0.22$). This statistical variation validates the approach/withdrawal model of depression at the multi-channel network level. Furthermore, long-range spatial communication metrics, represented by fronto-occipital Coherence (F3-O1), show strong integration with regional parietal signal variance at P3 ($r = 0.38$) and central standard deviation at Cz ($r = 0.33$).

Rather than relying on isolated channel assumptions, these distinct correlation vectors confirm that the mixed attention layers learn a distributed, clinically valid network pattern. This links the model's decision-making process directly to known neurophysiological markers of MDD, providing strong support for the study's interpretability claims.

H. Discussion

The present study demonstrates the feasibility and potential of EEG-based machine learning frameworks for the classification of MDD patients versus healthy controls (HC). Our findings indicate that the integrating of temporal and spectral EEG features, together with deep learning approaches such as BiLSTM, allows for robust subject-level discrimination. Beyond simple accuracy metrics, these results reveal insights into the neural dynamic underlying depressive pathology.

The superior performance of the BiLSTM model emphasizes the importance of capturing temporal dependencies in EEG signals, which appear critical for detecting subtle variations in neural oscillatory patterns associated with MDD. SVM-RBF, while slightly less sensitive, still provided consistent performance, demonstrating that carefully selected spectral features can support classical machine leaning approaches. Conversely, KNN and XGBoost displayed comparatively lower predictive power, highlighting that non-sequential or tree-based models may be insufficient for modelling the rich temporal structure of EEG data.

From a clinical standpoint the proposed BiLSTM-based framework offers a practical, non-invasive, and objective tool; for assisting the diagnosis of MDD. In routine psychiatric practice, diagnosis relies primarily on subjective self-report scales and clinical interviews, which can be influenced by patient insight, recall bias, and clinician experience. The integration of EEG-derived temporal and spectral features with a bidirectional LSTM network provides a quantifiable neurophysiology signature that can help differentiate MDD patients from healthy controls. With a subject-level accuracy of 92.45% under leave-one-subject-out cross validation, our results suggest the feasibility of using this framework as a potential decision-support approach in future clinical settings. However, validation on larger and independent multi-center cohorts is required before routine clinical use can be considered. Moreover, the short acquisition time (approximately 5 minutes of resting-state EEG) and automated processing pipeline make it feasible for routine clinical workflows without requiring specialized neurophysiological training for the interpreting physician.

Feature analysis offers additional insights. Candidate Beta and Gamma band features were identified primarily over frontal and central electrode sites (Figure4), consistent with literature linking these regions to affective and executive functioning. The spatial distribution of discriminatory EEG features over fronto-central regions aligns with current neurobiological models of depression which implicate dysfunction in the prefrontal cortex and anterior cingulate cortex- areas critical for emotion regulation, rumination, and executive control. In a clinical setting, this topographic information could be used to guide personalized treatment decisions. For example, patients exhibiting excessive beta/Gamma power in frontal regions might be better candidates for neurofeedback protocols targeting frontal alpha asymmetry or transcranial magnetic stimulation over the dorsolateral prefrontal cortex. Furthermore, the high negative predictive value (NPV) suggested by our balanced F1-scores (0.89 for MDD, 0.92 for HC) indicates that the system could be valuable for ruling out MDD in ambiguous cases, thereby reducing unnecessary referrals to specialized mental health services. Although the top two features survived Holm correction, the remaining seven did not. This underscores the necessity or rigorous statistical control in high-dimensional EEG studies, while also confirming the robustness of the two identified biomarkers However, their contribution to model performance indicates that this candidate features capture meaningful neurophysiological variance, even if not statistically confirmed after multiple comparison correction.

Interestingly, while classical significance testing did not confirm individual features, the multivariate models were still able to achieve high classification accuracy. This suggests that inter-feature interactions and distributed neural patterns, rather than isolated channel-specific effects, may carry the most diagnostic information. Such findings align with the growing perspective in computational psychiatry that complex brain disorders manifest through network-level alterations rather than single-region dysfunction.

From a translational standpoint, the high predictive performance and reproducibility demonstrated by LOSO cross-validation indicate potential applicability in clinical decision support systems. Automated EEG-based classification could complement traditional diagnostic procedures, providing objective measurements and enabling early detection of depressive symptoms. Moreover, the spatial distribution of candidate Beta/Gamma features may guide future research into targeted neuro modulatory interventions, such as transcranial stimulation protocols aimed at front-central networks.

Compared to traditional depression screenings tools such as the patient health Questionnaire-9 (PHQ-9) or Beck Depression Invenory (BDI), our EEG-based framework; offers several practical advantages. First, it is immune to social desirability bias or symptoms exaggeration, as it relies on objective electrophysiological data. Second, it can be performed on individuals who have difficulty completing self-report forms (e.g., adolescents, cognitively impaired patients, or non-native language speakers). Third, the entire procedure-from electrode application to automated classification-takes less than 20 minute, with results available immediately after recording. While the PHQ-9 and similar instruments remain essential for initial screening, our method can serve as a second-line confirmatory tool or as a monitoring device track treatment response over time, Serial EEG recordings could objectively quantify neurophysiological changes following antidepressant medication, cognitive-behavioral therapy, or neurostimulation, allowing clinicians to adjust treatment plans in a timely manner.

Limitations of this study include a moderate sample size and the lack of external validation datasets. Future research should incorporate larger and more diverse populations, longitudinal recordings to track treatment response, and multimodal data integration to strengthen model generalizability. Despite the limitations, the findings provide a compelling argument for the use of EEG-based deep learning frameworks as non-invasive, scalable tool for mental health assessment.

In summary, this study not only confirms the diagnostic potential of EEG features in MDD but also highlights the added value of temporal modeling, distributed feature interactions, and candidate frequency bands for clinical application. These insights provide a foundation for the development of more precise, interpretable, and effective EEG-based diagnostic tools in psychiatry.

I. Conclusion

This study establishes the feasibility and effectiveness of EEG-based machine learning frameworks for distinguishing MDD patients from healthy controls (HC). By integration temporal and spectral EEG features with deep learning architectures such as BiLSTM, the framework achieved high subject-level classification performance, demonstrating the importance of capturing sequential neural dynamics.

Feature analysis identified candidate Beta and Gamma band features primarily over frontal and central electrodes, highlighting spatial regions potentially relevant for MDD pathophysiology. While strict statistical correction (Holm) reduced feature-level significance, multivariate models effectively leveraged distributed patterns across electrodes, emphasizing the role of inter-feature interactions over isolated channel effects.

The results underscore the translational potential of EEG-based computational models as non-invasive, objective tools for early detection, monitoring, and clinical assessment of depressive disorders. In summary, the proposed EEG-based BiLSTM framework has direct translational value for mental health care. It provides an objective, rapid, and low-cost method for distinguishing MDD patients from healthy controls with high accuracy and AUC. Clinicians can use this tool as an adjunct to clinical interviews, particularly in primary care or neurology clinics where specialized psychiatric expertise may not be immediately available. By visualizing the neurophysiological correlates of depressive symptoms, the system also helps demystify the biological basis of depression for patients, potentially reducing stigma and improving treatment adherence. Future implementation of this framework in routine clinical practice could shorten diagnostic delays, facilitate early intervention, and enable objective monitoring of treatment outcomes. Future studies incorporating larger, multi-site datasets, longitudinal tracking, and multimodel data are warranted to validate these findings and further refine predictive models.

In conclusion, the combination of targeted EEG feature extraction and advanced machine learning provides a robust foundation for developing precise, interpretable, and clinically applicable tools for MDD diagnosis. These findings advance the understanding of neural signatures in depression and pave the way for automated scalable mental health assessment strategies.

References

[1] P. Sajda, “The importance of being in-sync: closed-loop EEG-rTMS for personalizing target engagement for treatment of MDD,” *Brain Stimulation*, vol. 18, no. 1, p. 231, Jan. 2025, doi: https://doi.org/10.1016/j.brs.2024.12.052.

[2] A. Ziogas *et al.*, “Deep Learning in the Identification of Electroencephalogram Sources Associated with Sexual Orientation,” *Neuropsychobiology*, pp. 1–12, Jun. 2023, doi: https://doi.org/10.1159/000530931.

[3] S. Olbrich *et al.*, “Deep learning using electroencephalogram (EEG) data for diagnosing and predicting SSRI response in major depressive disorder,” *Communications Medicine*, vol. 6, no. 1, Mar. 2026, doi: https://doi.org/10.1038/s43856-026-01394-z.

[4] S. N. Vaniya, A. Habib, M. Angelova, and C. Karmakar, “Simplifying Depression Diagnosis: Single-Channel EEG and Deep Learning Approaches,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2026, doi: https://doi.org/10.1109/jbhi.2025.3631326.

[5] S. Priyadarshinee and M. Panda, “A Novel EEG-based Hypergraph Convolution Network for depression detection: Incorporating Unified Brain Network and Multi-segment spatiotemporal EEG features,” *IEEE Access*, pp. 1–1, 2025, doi: https://doi.org/10.1109/access.2025.3614269.

[6] W. Chen, Y. Cai, A. Li, K. Jiang, and Y. Su, “MDD brain network analysis based on EEG functional connectivity and graph theory,” *Heliyon*, pp. e36991–e36991, Aug. 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e36991.

[7] N. Marcotullio, W. L. Acuff, N. Philip, M. van ’t Wout-Frank, and A. Zandvakili, “232. Correlating MDD and PTSD Symptoms With Cortical Functional Connectivity: An EEG and Machine Learning Study,” *Biological Psychiatry*, vol. 99, no. 10, p. S198, May 2026, doi: https://doi.org/10.1016/j.biopsych.2026.03.466.

[8] H. Gao, B. Qiu, R. J. Duran Barroso, W. Hussain, Y. Xu, and X. Wang, “TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder,” *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2022, doi: https://doi.org/10.1109/tnse.2022.3163144.

[9] S. Soni, A. Seal, A. Yazidi, and O. Krejcar, “Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression,” *Computers in Biology and Medicine*, vol. 145, p. 105420, Jun. 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105420.

[10] Y. J. Kwon, B. H. Yoon, M. D. Kim, K. J. Min, and H. M. Sung, “Microstate analysis from resting-state EEG in patients with Major Depressive Disorder (MDD) and Bipolar Disorder (BD),” *Neuroscience Applied*, vol. 5, p. 106738, 2026, doi: https://doi.org/10.1016/j.nsa.2025.106738.

[11] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, “A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity,

and nonlinear analysis," *Journal of Neuroscience Methods*, vol. 358, p. 109209, Jul. 2021, doi: https://doi.org/10.1016/j.jneumeth.2021.109209.

[12] J. Li *et al.*, "Altered Brain Dynamics and Their Ability for Major Depression Detection Using EEG Microstates Analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2116–2126, Dec. 2021, doi: https://doi.org/10.1109/taffc.2021.3139104.

[13] B. Sütçübaşı, T. Ballı, B. Metin, and E. Elif Tülay, "Neural signatures of depression: classifying drug-naïve MDD patients with time- and frequency-domain EEG features during emotional processing," *Machine Learning: Science and Technology*, vol. 6, no. 2, p. 025035, May 2025, doi: https://doi.org/10.1088/2632-2153/add4bb.

[14] Y. Xi, B. Shi, T. Lu, P. Tian, and L. Zhang, "MDD Detection Based on Time-Spatial Features from EEG Symmetrical Microstate–Brain Networks," *Symmetry*, vol. 18, no. 1, p. 59, Dec.
2025, doi: https://doi.org/10.3390/sym18010059.

[15] S. De, A. Singh, and A. K. Bhandari, "A Novel Vision Transformer based Multimodal Fusion Approach for Clinical MDD Diagnosis Using EEG and Audio Signals," *IEEE Transactions on Computational Biology and Bioinformatics*, pp. 1–11, 2025, doi: https://doi.org/10.1109/tcbbio.2025.3626210.

[16] A. Sam, R. Boostani, S. Hashempour, M. Taghavi, and S. Sanei, "Depression Identification Using EEG Signals via a Hybrid of LSTM and Spiking Neural Networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4725–4737, Jan. 2023, doi: https://doi.org/10.1109/tnsre.2023.3336467.

[17] Y. Wang, S. Zhao, H. Jiang, S. Li, T. Li, and G. Pan, "M-MDD: A multi-task deep learning framework for major depressive disorder diagnosis using EEG," *Neurocomputing*, vol. 636, p. 130008, Jul. 2025, doi: https://doi.org/10.1016/j.neucom.2025.130008.

[18] N. Bashir *et al.*, "A Machine Learning Framework for Major Depressive Disorder (MDD) Detection Using Non-invasive EEG Signals," *Wireless Personal Communications*, May 2023, doi: https://doi.org/10.1007/s11277-023-10445-w.

[19] B. Wang, Y. Kang, D. Huo, G. Feng, J. Zhang, and J. Li, "EEG
diagnosis of depression based on multi-channel data fusion
and clipping augmentation and convolutional neural network,"
Frontiers in Physiology, vol. 13, Oct. 2022, doi:

[20] S. Khan *et al.*, "A machine learning based depression screening framework using temporal domain features of the electroencephalography signals," *PloS one*, vol. 19, no. 3, pp. e0299127-e0299127, Mar.2024,doi:10.1371/journal.pone.0299127.

[21] E. H. Earl *et al.*, “EEG based functional connectivity in resting and emotional states may identify major depressive disorder using machine learning,” *Clinical Neurophysiology*, vol. 164, pp. 130–137, June 2024, doi: 10.1016/j.clinph.2024.05.017.