ISSN: 2664-9640

Detection Methods for Depressive Disorders based on Resting-State EEG Signals

Haoshu Zhang, Ping Yang*

School of Electronic and Information Engineering, Beijing Union University, Beijing, China

*Corresponding Author: Ping Yang (E-mail: xxtyangping@buu.edu.cn)

Abstract

Major depressive disorder (MDD) is primarily diagnosed through structured interviews, symptom-rating scales, and clinical judgment, whereas objective neurophysiological biomarkers have not yet been standardized for routine practice. Resting-state electroencephalography (rs-EEG) is noninvasive, relatively inexpensive, has high temporal resolution, and is suitable for repeated recordings. It has therefore become an important platform at the intersection of electronic information engineering and psychiatry. This review synthesizes representative studies of multicenter validation, functional brain networks, adaptive signal decomposition, convolutional neural networks, recurrent and spiking neural networks, Transformers, graph neural networks, and graph-convolutional Transformer architectures to analyze the complete technical chain of rs-EEG-based depression detection. The review first examines eyes-open and eyes-closed paradigms, sampling and referencing, filtering and artifact rejection, empirical mode decomposition (EMD) and variational mode decomposition (VMD), spectral and nonlinear complexity measures, functional connectivity, and graph-theoretic indices. It then compares support vector machines, random forests, convolutional neural networks, long short-term memory networks, spiking neural networks, self-attention, graph convolutional networks, and hybrid architectures. Representative findings are interpreted according to subject-independent partitioning, independent test sets, cross-site heterogeneity, data-leakage risk, and the comparability of evaluation metrics. Available evidence indicates that interchannel connectivity may reflect the network-level abnormalities of MDD more robustly than isolated channel power. The integration of graph-based spatial modeling with global temporal modeling is methodologically well motivated, but high accuracy in a small cohort should not be interpreted as clinical generalizability. Current work remains constrained by limited sample sizes, inconsistent diagnostic labels, distribution shifts across devices and sites, sensitivity to graph construction, insufficient interpretability, privacy barriers to data sharing, and the difficulty of lightweight deployment. Future studies should prioritize self-supervised pretraining, dynamic brain graphs, cross-domain and federated learning, uncertainty calibration, low-channel wearable systems, prospective clinical validation, and interpretable neurophysiological biomarkers. These directions are essential for moving rs-EEG-assisted diagnosis from algorithmic demonstrations toward reproducible, explainable, and deployable clinical decision-support tools.

Keywords

Resting-state EEG; Major Depressive Disorder; Functional Connectivity; Deep Learning; Transformer; Graph Neural Network; Computer-aided Diagnosis.

ISSN: 2664-9640

1. Introduction

1.1. Background and Engineering Requirements

Major depressive disorder is characterized by persistent depressed mood, diminished interest, and impairments in cognition and social functioning. Current clinical procedures can establish a diagnosis, but uncertainty may arise from self-reported scales, incomplete interview information, comorbid symptoms, medication effects, and inter-clinician variability. For electronic information engineering, the objective is not to replace psychiatrists with an algorithm. A more realistic goal is to build a physiological-signal system that can be recorded repeatedly, analyzed quantitatively, and used to provide risk alerts or supplementary evidence for clinical decision making. EEG directly records scalp-potential fluctuations with millisecond temporal resolution and has relatively low cost, portability, and suitability for longitudinal follow-up; these characteristics provide clear engineering advantages for objective depression assessment [1-3].

To address this objective, rs-EEG-based detection has developed into a complete chain of acquisition, preprocessing, representation, modeling, validation, and interpretation. Wu et al. compared band power, coherence, and fractal dimension in 400 participants from four medical centers and evaluated the frozen model on an independent test set; their results suggested that coherence-based connectivity was comparatively stable in heterogeneous multicenter data [4]. Zhang et al. constructed a 64-channel brain functional network using the phase-lag index (PLI) and reported associations between selected network measures and PHQ-9 scores [5]. Shao et al. further built connectivity networks at different intrinsic-mode-function scales using an improved EMD procedure [6]. In end-to-end learning, Li et al. applied a CNN to both depression identification and severity classification [7], whereas Sam et al. integrated LSTM and spiking neural networks for scale-defined multilevel depression classification [8]. GCTNet used residual graph convolution and a Transformer in parallel, illustrating the current trend toward fusion of spatial topology and global temporal context [9].

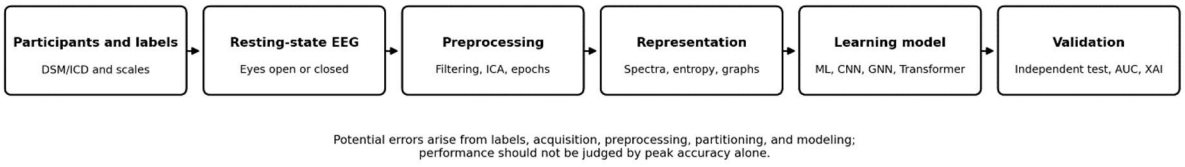

Figure 1. Complete technical pipeline for rs-EEG-based depression detection

1.2. Characteristics of rs-EEG compared with Other Neuroimaging Techniques

Functional magnetic resonance imaging (fMRI), positron-emission tomography (PET), magnetoencephalography (MEG), and EEG can all be used to study brain function, but they address different methodological goals. fMRI provides comparatively high spatial resolution and is suitable for localizing large-scale networks. PET can characterize metabolic or receptor-related processes, but ionizing radiation limits repeated measurement. MEG combines high temporal resolution with better source-localization potential than scalp EEG, yet it is expensive and requires a highly controlled environment. EEG has more limited spatial localization, but it can record rapid neural oscillations at low cost and can be used in repeated follow-up, community screening, and wearable scenarios. Therefore, EEG is particularly suitable for building a high-temporal-resolution, repeatable, and deployable depression-assessment system rather than for pursuing precise anatomical localization alone [10-11].

ISSN: 2664-9640

Table 1. Comparison of common methods for measuring brain function

Technique	Temporal resolution	Spatial resolution	Cost and portability	Relevance to MDD-assisted detection
EEG	Milliseconds	Relatively low	Low cost; suitable for repeated recording	Useful for oscillations, connectivity, and longitudinal follow-up
MEG	Milliseconds	Moderate to high	Very high cost; poor portability	Good source localization, but difficult to scale
fMRI	Seconds	High	High cost and strict environmental requirements	Suitable for resting-state network localization and multimodal studies
PET	Minutes	Moderate	High cost; ionizing radiation	Useful for metabolism and receptor mechanisms; unsuitable for frequent measurement
fNIRS	Subsecond to seconds	Low to moderate	Moderate cost; portable	Useful for frontal hemodynamics and fusion with EEG

Note: The general temporal, spatial, and deployment characteristics of the listed technologies are summarized at a methodological level; the brain-network perspective is supported by [10-11].

1.3. Scope, Label Boundaries, and Hierarchy of Evidence

This review focuses on eyes-open or eyes-closed resting-state EEG and mainly discusses the classification of clinically diagnosed MDD or broadly defined depressive states versus healthy controls. Task-evoked EEG, including emotion-elicitation paradigms [12], sleep EEG, and prediction of antidepressant treatment response are mentioned only when they help explain methodological transfer and are not included in the core performance comparison. Three label categories must be distinguished. The first is clinically confirmed MDD based on DSM or ICD criteria and a structured interview. The second is depression severity defined by thresholds on scales such as the Beck Depression Inventory (BDI), Patient Health Questionnaire-9 (PHQ-9), or Self-Rating Depression Scale (SDS). The third is a broadly defined depression state determined by symptom scales together with preliminary clinical judgment. These labels have different clinical meanings and should not be ranked solely by classification accuracy.

Table 2. Evolution of methods for rs-EEG-based MDD detection

Stage	Representative representation/model	Primary contribution	Typical limitation
Handcrafted-feature stage	PSD, entropy, fractal dimension + SVM	Established interpretable spectral and complexity markers	Expert-dependent features and limited cross-dataset stability
Functional-connectivity stage	Coherence, PLI, graph measures	Shifted analysis from isolated channels to network relations	Sensitive to referencing, thresholds, and volume conduction
End-to-end deep-learning stage	CNN, LSTM, SNN	Learned local or temporal representations automatically	Overfitting in small cohorts and weaker interpretability
Graph-learning and attention stage	GCN, GRU, Transformer	Fused electrode topology with long-range dependencies	Sensitive graph construction and higher training cost
Clinical-translation stage	Multicenter, federated, calibrated, lightweight models	Emphasized cross-domain generalization and deployment reliability	Insufficient standard databases and prospective trials

ISSN: 2664-9640

The evidence in this review is evaluated according to label reliability, sample size, independence of the data split, external validation, and methodological transparency. A structured clinical diagnosis generally provides stronger evidence than a single scale threshold. Subject-level partitioning is stronger than randomly assigning segments from the same participant to both training and test sets. Evaluation on a frozen, multicenter independent test set provides stronger evidence than internal cross-validation on a single dataset. Full reporting of sensitivity, specificity, and AUC is more informative than reporting only the highest accuracy. Clear preprocessing descriptions, available code, and ablation studies further improve reproducibility. This hierarchy does not invalidate exploratory small-sample studies; instead, it defines the appropriate boundary of their conclusions. Small cohorts are useful for proposing mechanistic or architectural hypotheses, whereas clinical utility requires stronger validation designs [2-4].

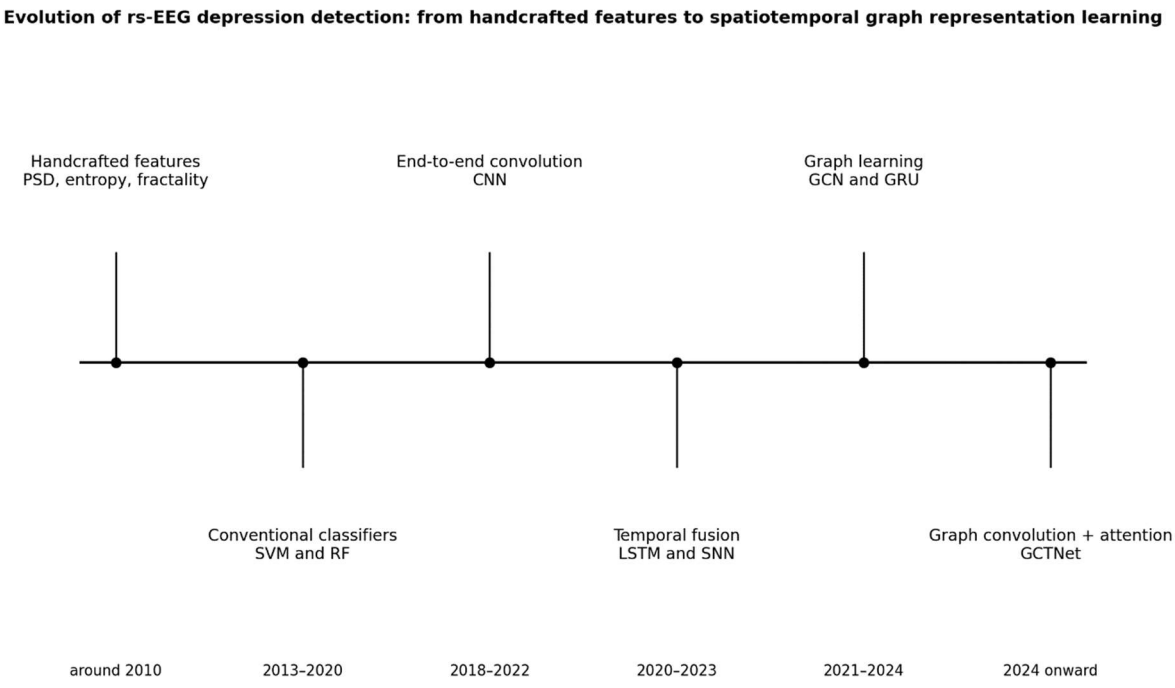

Figure 2. Evolution of methods for rs-EEG-based depression detection

1.4. Organization and Central Arguments

Section 2 discusses recording, preprocessing, signal decomposition, feature extraction, and brain-network representation. Section 3 analyzes conventional machine-learning approaches and classification with functional brain-network features. Section 4 reviews CNNs, LSTMs, SNNs, Transformers, GCNs, and GCTNet. Section 5 provides a unified comparison based on datasets, validation protocols, and evaluation metrics. Section 6 summarizes major challenges and future directions, and Section 7 presents the conclusions. The central arguments are as follows. First, connectivity relations are more suitable than isolated-channel features for representing the network-level abnormalities of MDD, although connectivity estimation itself can be unstable. Second, combining graph convolution with a Transformer is methodologically reasonable, but its benefit must be demonstrated through subject-independent partitioning, ablation studies, and cross-site testing. Third, the evaluation objective for clinical translation should shift from maximum accuracy toward a multiobjective balance of reproducibility, calibration, interpretability, and deployability.

ISSN: 2664-9640

2. The rs-EEG Data Chain: Acquisition, Preprocessing, and Representation

2.1. Recording Paradigms and Data Heterogeneity

A resting-state experiment generally asks participants to remain awake and relaxed, minimize movement, and undergo continuous recording under eyes-open fixation or eyes-closed conditions. Eyes-closed recording reduces visual input and usually enhances occipital alpha activity, but it increases the likelihood of drowsiness. Eyes-open recording helps maintain vigilance but introduces more blinks and eye movements. Studies differ substantially in channel count, sampling rate, recording duration, reference montage, medication status, age distribution, and diagnostic scale. This heterogeneity directly changes spectral and connectivity estimates.

Wu et al. used 32-channel eyes-open rs-EEG from 400 participants across four centers and separated the data into a development set and a completely independent test set [4]. Zhang et al. used 64-channel recordings to construct PLI networks for 24 patients with MDD and 24 healthy controls [5]. Shao et al. analyzed high-density EEG from 30 participants using improved EMD and network measures [6]. Li et al. enrolled 82 participants; the CNN achieved a diagnostic accuracy of 66.40%, and the comparatively conservative result reflected the realistic difficulty introduced by simplified preprocessing and clinical heterogeneity [7]. GCTNet used subject-independent ten-fold partitioning on an in-house cohort of 85 participants and on a public dataset, providing a stricter framework for assessing stability across datasets [9].

Table 3. Data acquisition and task settings in representative studies

Study	Sample and labels	Channels/paradigm	Validation design	Primary purpose
Wu et al. [4]	400 participants; clinically diagnosed MDD/HC	32 channels; eyes-open rest	Five-fold development + independent test set of 120 participants	Multicenter generalization and feature stability
Zhang et al. [5]	24 MDD + 24 HC	64 channels; resting state	Internal classification and statistical analysis	PLI networks and graph measures
Shao et al. [6]	15 MDD + 15 HC	High-density resting-state EEG	Small-sample internal comparison	Improved-EMD multiscale networks
Li et al. [7]	41 depression + 41 HC	Routine clinical resting-state EEG	Internal CNN validation	Diagnosis and severity classification
Sam et al. [8]	Scale-defined severity groups	Eyes open/eyes closed	Multiclass classification	LSTM-SNN severity modeling
GCTNet [9]	85 participants + public data	56/19 channels; eyes closed	Subject-independent ten-fold validation	Spatial graph + global temporal fusion
W-GCN-GRU [13]	In-house data + MODMA	Resting state	Within-dataset evaluation	Weighted features, GCN, and GRU
Song and Zhang [14]	Public dataset	Multichannel resting-state EEG	Comparative experiments	APSO-VMD and a hybrid deep model

Note: Sample, paradigm, and validation information was summarized from the core studies [4-9, 13-14]. Labels across tasks are not clinically equivalent.

2.2. Preprocessing and Reproducibility

The objective of preprocessing is to suppress device noise, baseline drift, power-line interference, ocular and muscular artifacts, and transient outliers while retaining disease-

ISSN: 2664-9640

relevant neural information. A common workflow includes rereferencing, band-pass and notch filtering, bad-channel detection, independent component analysis (ICA), automatic component labeling, segmentation, threshold-based rejection, and normalization. EEGLAB and FieldTrip provide mature implementations for many of these operations [15-16]. Nevertheless, aggressive manual cleaning may retain only the segments that are easiest to classify, producing excellent performance on sanitized data while reducing robustness to routine clinical recordings.

Li et al. deliberately reduced preprocessing and obtained an accuracy of approximately 66%, arguing that intensive cleaning may cause earlier studies to overestimate clinical performance [7]. GCTNet removed ocular, muscular, and cardiac components using ICA and ICLabel, downsampled the data to 100 Hz, segmented it with a 5-s window and a 1-s step, and partitioned data at the participant level [9]. These choices show that preprocessing and validation cannot be considered separately. Even when each segment is clean, a model may learn participant-specific signatures rather than disease-related patterns if adjacent windows from one participant appear in both the training and test sets.

$$z_{c,t} = \frac{x_{c,t} - \mu_c}{\sigma_c + \varepsilon} \quad (1)$$

Equation (1) denotes channel-wise standardization, where x(c,t) is the voltage of channel c at time t, and μc and σc are the mean and standard deviation estimated from the training data. These statistics must be computed only from the training set; otherwise, information from the test distribution is introduced into model development.

Table 4. Common preprocessing steps and associated risks in rs-EEG

Step	Typical method	Primary purpose	Potential risk and recommendation
Rereferencing	Average reference; linked mastoids	Reduce reference-electrode bias	The reference scheme affects power and connectivity; report it in full
Filtering	0.5/1-45/50 Hz; notch filtering	Remove drift and power-line noise	High-order filters may create edge effects and phase distortion
Artifact handling	ICA, ICLabel, regression	Remove ocular, muscular, and cardiac artifacts	Automatically labeled components should be visually audited
Segmentation	Sliding windows of 2-10 s	Increase samples and approximate stationarity	Segments from one participant must not cross train/test sets
Bad-epoch rejection	Amplitude, peak-to-peak, or gradient thresholds	Remove transient abnormalities	Overly strict thresholds produce selection bias
Normalization	Z-score or robust scaling	Stabilize optimization	Statistics must be estimated on the training set only
Data augmentation	Cropping, noise, channel masking	Reduce overfitting	Augmentation must remain physiologically plausible

2.3. Adaptive Decomposition: EMD, Improved EMD, and VMD

EEG is nonstationary, and fixed frequency bands do not necessarily match the individual boundaries of neural oscillations. Empirical mode decomposition separates a signal into intrinsic mode functions (IMFs) and a residual through iterative operations based on local extrema and envelopes, making it suitable for nonlinear and nonstationary signals [17]. Its basic representation is:

ISSN: 2664-9640

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) \tag{2}$$

Classical EMD is susceptible to mode mixing, endpoint effects, and inconsistent numbers of IMFs across channels. Shao et al. proposed an improved EMD procedure for high-density EEG so that the number of modes was more stable across channels. Correlation, coherence, and PLI networks were then calculated at different IMF scales [6]. Network differences at selected IMFs were more evident than those obtained from fixed bands, but the cohort contained only 30 participants. The findings are therefore more appropriate as a multiscale network hypothesis than as evidence of clinical accuracy.

Variational mode decomposition formulates decomposition as a constrained variational optimization problem in which each mode has a compact bandwidth around an adaptively updated center frequency; it is generally more mathematically stable than EMD [18]. A simplified objective is:

$$\min_{\{u_k\},\{\omega_k\}} \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2, \quad \text{s.t.} \sum_k u_k = x \tag{3}$$

The key VMD parameters are the number of modes K and the penalty factor alpha. Song and Zhang used adaptive particle swarm optimization to jointly select K and alpha, reconstructed the EEG based on the similarity between each component and the original signal, extracted frequency-spatial features, and fed them to a 2D-CNN-BiLSTM-SA network. The reported best accuracy was 94.47% [14]. The strength of this pipeline lies in the coordinated design of denoising, frequency-spatial representation, and deep learning. Its weakness is the complexity of both the optimizer and the network. If parameter search is performed before data partitioning or with information from all participants, hidden leakage may occur.

Table 5. Comparison of EMD-family and VMD-family methods

Method	Optimization mechanism	Advantage	Limitation	Role in MDD research
EMD	Local-extrema and envelope sifting	Data-driven; no predefined basis	Mode mixing and endpoint effects	Extraction of multiscale nonstationary components
Improved EMD	Stabilized IMF count and sifting procedure	Suitable for synchronized high-density multichannel analysis	Strong dependence on implementation details	Construction of IMF-scale functional networks [6]
EEMD/CEEMDAN	Noise-assisted ensemble averaging	Reduces mode mixing	High computation and residual noise	Improves decomposition stability
VMD	Constrained variation and center-frequency updates	Compact bands and mathematical stability	Requires selection of K and alpha	Denoising and frequency-spatial representation
APSO-VMD	Swarm-based joint search of K and alpha	Reduces manual parameter selection	High search cost; requires nested validation	Combined with CNN-BiLSTM-SA for recognition [14]

ISSN: 2664-9640

2.4. Spectral, Nonlinear, and Statistical Features

Conventional approaches commonly calculate absolute power, relative power, peak frequency, hemispheric asymmetry, and band ratios in the delta, theta, alpha, beta, and gamma ranges. Power spectral density can be estimated using a periodogram or Welch method; its idealized definition is:

$$S_x(f) = \lim_{T\to\infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} x(t) e^{-j2\pi ft}\, dt \right|^2 \tag{4}$$

In addition to spectral power, sample entropy, approximate entropy, permutation entropy, Higuchi fractal dimension (HFD), and Katz fractal dimension (KFD) describe complexity and nonlinear dynamics. Hosseinifard et al. combined linear and nonlinear features with machine learning to classify MDD [19]. Ahmadlou et al. analyzed frontal fractal characteristics in MDD [20], and Mahato and Paul compared linear and nonlinear representations and emphasized the value of their combination [21]. In the large multicenter study by Wu et al., approximately 98% of the selected features were coherence-based connectivity measures. This result suggests that, as heterogeneity increases, the stability of isolated-channel power and fractal measures may be lower than that of interchannel relationships [4].

$$\mathrm{SampEn}(m, r) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{5}$$

Table 6. Common EEG feature categories and their interpretation

Feature category	Representative measures	Interpretation	Main strength	Main limitation
Spectral	Absolute/relative PSD; band ratios	Energy of neural oscillations	Simple computation and clinical readability	Affected by reference, eye state, and individual alpha peak
Asymmetry	Frontal alpha asymmetry; power differences/ratios	Interhemispheric activity differences	Linked to theories of emotion regulation	Direction and stability vary across studies
Complexity	Entropy, HFD, KFD	Irregularity and nonlinear dynamics	May reveal dynamical changes	Sensitive to length, noise, and parameter selection
Statistical	Mean, variance, skewness, kurtosis	Shape of the time-domain distribution	Fast and suitable for lightweight systems	Limited disease specificity
Time-frequency	Wavelets, IMFs, VMD-mode energy	Nonstationary multiscale information	Represents transient changes	High dimensionality and complex parameterization
Connectivity	Coherence, PLV, PLI, wPLI	Synchronization and network coordination	Consistent with the network nature of MDD	High computational cost and graph-construction dependence

2.5. Functional connectivity and volume conduction

Functional connectivity describes statistical or phase-based associations between channels. Magnitude-squared coherence combines auto-spectral and cross-spectral information and can be written as:

ISSN: 2664-9640

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \tag{6}$$

Coherence includes both amplitude and phase relations, but common zero-lag sources can produce spuriously high values. Phase-locking value (PLV) considers only the stability of phase differences [22]:

$$\mathrm{PLV}_{xy} = \left| \frac{1}{N} \sum_{n=1}^{N} e^{j(\phi_x(n) - \phi_y(n))} \right| \tag{7}$$

The phase-lag index measures the directional bias of the sign of the imaginary component of phase differences and reduces the influence of zero-lag common sources [23]. The weighted phase-lag index further weights this sign by the magnitude of the imaginary component, reducing flips caused by weak noise [24]. However, reduced sensitivity to volume conduction does not mean complete elimination of bias. Phase estimation, window length, filter edges, reference choice, and sample size can all alter results. Studies should report the precise connectivity definition, band, window, threshold, and sparsification strategy.

$$\mathrm{PLI}_{xy} = |E[\mathrm{sign}(\sin(\Delta\phi_{xy}))]| \tag{8}$$

Table 7. Comparison of common functional-connectivity measures

Measure	Information used	Sensitivity to volume conduction	Advantage	Limitation
Pearson correlation	Linear time-domain amplitude relation	High	Simple and fast	Ignores frequency and nonlinear dependence
Coherence	Frequency-domain amplitude and phase	Relatively high	Stable in a large multicenter study [4]	Zero-lag common sources may create false connections
PLV	Stability of phase differences	Moderate	Suitable for weighted phase networks	Biased in small samples and affected by reference
PLI	Direction of nonzero phase lag	Relatively low	Suppresses zero-lag common sources [23]	Discards phase-lag magnitude; weak edges are unstable
wPLI	Magnitude-weighted imaginary phase lag	Relatively low	More robust to noise and sample-size bias [24]	More complex computation and interpretation
Mutual information	Linear and nonlinear dependence	Moderate	Captures complex coupling	Estimation depends on binning or kernel parameters

2.6. From Connectivity Matrices to Graph-theoretic Measures

Treating electrodes as nodes and connectivity strengths as edges yields a weighted undirected graph G=(V,E,A). Graph construction may retain the complete weighted matrix or apply fixed density, statistical significance, a minimum spanning tree, or a learnable adjacency matrix. A fixed threshold is simple, but it can produce different network densities across participants. Fixed density facilitates comparison but may preserve noisy edges. A fully learnable adjacency is flexible, yet it may be less interpretable and less stable across datasets.

The local clustering coefficient describes the compactness of a node's neighborhood, the average shortest-path length describes global integration, and small-worldness characterizes the balance between local segregation and global integration [10-11, 25]. Zhang et al. reported a trend toward network randomization in MDD, associations between selected theta/alpha2

ISSN: 2664-9640

measures and PHQ-9, and high internal classification performance using graph features [5]. Because graph measures change markedly with thresholds and density, a multithreshold sensitivity analysis is preferable to reporting only the optimal threshold.

$$C_i = \frac{2e_i}{k_i(k_i-1)}, \qquad L = \frac{1}{N(N-1)}\sum_{i\neq j} d_{ij} \tag{9}$$

Table 8. Interpretation and cautions for graph-theoretic brain-network measures

Measure	Primary meaning	Interpretation in MDD studies	Methodological caution
Node degree/strength	Number or total weight of connections	Identification of potential abnormal hubs	Strong dependence on network density
Clustering coefficient	Local neighborhood clustering	Local segregation and modular organization	Binary or weighted definition must be specified
Characteristic path length	Distance of global information transfer	Efficiency of global integration	Treatment of disconnected graphs changes the value
Global/local efficiency	Parallel information-transfer capacity	Related to cognition and emotion regulation	Requires matched random-network normalization
Betweenness centrality	Frequency on shortest paths	Identification of critical relay nodes	Sensitive to a small number of edge changes
Small-world coefficient	Balance of clustering and path length	Network randomization or weakened organization	Random graph and density must be matched

2.7. Summary

Effective rs-EEG modeling is not a single-algorithm problem but a complete data-chain problem. The recording paradigm and reference determine the original distribution; preprocessing controls signal-to-noise ratio and information fidelity; decomposition and feature extraction define the representation visible to the model; and connectivity and graph construction introduce a brain-network prior. Current evidence favors comparing representations under strict subject-independent validation rather than repeatedly optimizing preprocessing and features on one dataset and reporting only the best result.

3. Conventional Machine Learning and Functional Brain-Network Classification

3.1. Handcrafted Features and Feature Selection

The conventional machine-learning pipeline first constructs spectral, complexity, statistical, and connectivity features; applies feature selection or dimensionality reduction; and finally trains a classifier such as an SVM, linear discriminant analysis (LDA), K-nearest neighbors (KNN), or random forest. Its strengths include a relatively small parameter count, suitability for small samples, and direct interpretation of input features. Its principal weakness is that performance depends heavily on expert feature design. In addition, feature selection performed before cross-validation exposes information from the test distribution and leads to optimistically biased results.

Wu et al. generated 1,859 candidate features and used sequential backward selection (SBS) to identify an optimal subset before training a conformal-kernel SVM [4]. The work contributed not only a five-fold cross-validation accuracy of 91.07% but also an accuracy of 84.16% on a completely independent test set, thereby quantifying the gap between internal cross-validation and realistic generalization. Coherence accounted for approximately 98% of the final selected features, again suggesting that interchannel relationships may be more stable than isolated-channel descriptors in large multicenter data.

ISSN: 2664-9640

Table 9. Comparison of common feature-selection methods

Method	Category	Advantage	Limitation	Recommended use
t-test/effect size	Filter	Simple and interpretable	Ignores feature interactions and multiple testing	Perform within each training fold and correct significance
Mutual information/mRMR	Filter	Balances relevance and redundancy	Estimation depends on sample size	Useful for high-dimensional screening
SFS/SBS	Wrapper	Directly optimizes classifier performance	Computationally expensive and prone to selection bias	Requires nested cross-validation
LASSO	Embedded	Sparse and interpretable	Unstable when features are strongly collinear	Useful with linear or generalized linear models
Tree-based importance	Embedded	Models nonlinear effects	May favor high-cardinality variables	Prefer permutation importance for interpretation
PCA	Unsupervised reduction	Reduces redundancy and stabilizes computation	Principal components have weak clinical meaning	Use for compression rather than biomarker claims

3.2. Support Vector Machines and Kernel Methods

An SVM obtains a decision hyperplane by maximizing the margin between classes and is therefore well suited to high-dimensional problems with limited samples. The primal soft-margin optimization problem is:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i, \qquad \text{s.t. } y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i \tag{10}$$

A kernel function K(x_i,x_j)=phi(x_i)^Tphi(x_j) represents nonlinear decision boundaries without explicitly computing the high-dimensional mapping. Wu et al. used a conformal kernel to rescale the local geometry of the original kernel space and obtain a more suitable boundary near difficult samples [4]. In multicenter data, this approach outperformed standard SVM, KNN, and LDA. However, feature selection, scaling, and hyperparameters must all be frozen before evaluation on an external cohort; otherwise, repeated optimization can still overfit the development data.

Hosseinifard et al. provided early evidence that nonlinear EEG features and machine-learning classifiers can distinguish patients with depression from healthy participants [19]. Mumtaz et al. proposed an EEG-based computer-aided diagnostic framework [26], and Wan et al. showed that a single prefrontal channel may have value for prescreening MDD [27]. These studies demonstrate the potential of conventional models for lightweight devices, but their reported results vary widely because of differences in cohort size, labels, preprocessing, and partitioning.

$$K_{CK}(x_i, x_j) = D(x_i)K(x_i, x_j)D(x_j) \tag{11}$$

3.3. Random Forests, KNN, and Linear Discriminant Analysis

A random forest constructs an ensemble of decision trees using random sampling of participants and features. It can model nonlinear relations and provide estimates of feature importance. KNN has almost no explicit training process, but distance becomes less informative in high-dimensional spaces and predictions can be slow. LDA assumes class-specific means with a shared covariance matrix; it is parsimonious and transparent but cannot represent complex nonlinear boundaries. For small EEG datasets, a more complex model is not necessarily superior. The critical issues are feature stability, independent partitioning, and nested hyperparameter selection.

ISSN: 2664-9640

For the handcrafted-feature baseline in GCTNet, the authors extracted temporal, spectral, nonlinear, and combined features and compared SVM, random forest, naive Bayes, and KNN. Combined features generally performed better, but the optimal classifier and the optimal number of features differed across the two datasets [9]. This result reinforces that a best feature subset may be dataset-specific and should not automatically be promoted as a universal biomarker.

3.4. Brain-network Features with Conventional Classifiers

A functional brain-network approach transforms a connectivity matrix into clustering coefficients, path length, efficiency, centrality, and small-world measures and then applies a conventional classifier. Its major strength is that model inputs can be linked to network organization and therefore yield more interpretable neurophysiological hypotheses. Zhang et al. reported abnormal synchronization and topology in frontal, temporal, parietal, and occipital regions in a 64-channel PLI network and achieved a best internal classification accuracy of 93.31% [5]. Shao et al. compared correlation, coherence, and PLI networks across multiple IMF scales and found that graph measures in selected modes had strong discriminative potential [6].

Interpretability, however, is not guaranteed merely because graph measures are used. If a threshold is chosen retrospectively according to test accuracy, or if only significant measures at one optimal density are reported, multiple comparisons and selection bias may dominate the conclusion. A more reliable protocol determines graph-construction parameters on the training data, evaluates consistency across a range of densities, reports effect sizes and confidence intervals, and validates the direction of network abnormalities in an independent cohort.

Table 10. Applicability of conventional machine-learning approaches

Model	Strength	Limitation	Suitable EEG representation	Clinical-translation perspective
LDA/logistic regression	Few parameters and high transparency	Strong linear assumptions	A small set of stable features	Appropriate as an interpretable baseline
SVM/CK-SVM	Effective for high-dimensional small samples; flexible kernels	Hyperparameter selection and probability calibration are nontrivial	Spectral, entropy, connectivity, and graph features	Supported by multicenter independent testing [4]
Random forest	Nonlinear and provides variable importance	Importance estimates are unstable in small samples	Mixed handcrafted features	Useful as a robust comparator
KNN	Simple implementation and no parametric training	Distance failure in high dimensions; slow inference	Low-dimensional standardized features	Weak choice as a standalone clinical model
Brain network + SVM	Mechanistically interpretable	Depends on graph construction and threshold	PLI/coherence and graph measures	Requires cross-site validation of network stability
PCA + classifier	Reduces dimensionality and collinearity	Weakens interpretability	High-dimensional connectivity matrices	Useful for compression, not direct biomarker claims

3.5. Value and Limitations of Conventional Methods

Conventional methods remain indispensable in small-sample, low-channel, and highly interpretable settings. They provide strong baselines, expose problems in the data pipeline, and help determine whether the improvement of a deep model truly arises from representation learning. Their main limitations are that handcrafted features incompletely describe cross-channel, cross-frequency, and long-term dependencies; high-dimensional connectivity matrices require more samples; feature selection and hyperparameter tuning can create optimistic bias; and decision scores are often poorly calibrated. Conventional models should therefore be combined with strict validation rather than dismissed as merely simple methods.

ISSN: 2664-9640

4. Deep Learning, Transformers, and Graph Neural Networks

4.1. Convolutional Neural Networks: Learning Local Spatiotemporal Patterns

A convolutional neural network learns representations from raw EEG, spectra, time-frequency maps, or scalp topographies through local receptive fields and weight sharing. For a one-dimensional sequence, the l-th convolutional layer may be expressed as:

$$h_k^{(l)}(t) = \sigma\left(\sum_c \sum_\tau w_{k,c}^{(l)}(\tau) h_c^{(l-1)}(t-\tau) + b_k^{(l)}\right) \tag{12}$$

Acharya et al. applied a deep CNN to automated depression screening and helped establish end-to-end EEG classification [28]. Uyulan et al. subsequently compared multiple CNN architectures for MDD classification [29]. Li et al. enrolled 82 participants and reported a diagnostic AUC of 0.74 with an accuracy of 66.40%; the severity-classification AUC was 0.70 with an accuracy of 66.93% [7]. These numbers are lower than those in several small-cohort reports, but they have substantial methodological value. The study retained more realistic noise by simplifying preprocessing, demonstrating that clinically usable classification is much more difficult than classification of highly curated segments.

The principal limitation of CNNs is that a local kernel does not directly represent long-range dependencies. Larger kernels, deeper stacks, and dilated convolution increase the receptive field but also increase model size and optimization difficulty. When channels are arranged in a two-dimensional matrix, artificial neighborhood relations may be introduced. Multichannel EEG studies should therefore describe the spatial mapping explicitly and verify that performance is not driven by an arbitrary channel order.

4.2. LSTM and Spiking Neural Networks

Long short-term memory networks use input, forget, and output gates to control a persistent state and alleviate the vanishing-gradient problem of ordinary recurrent neural networks [30]. The core updates are:

$$f_t = \sigma(W_f[x_t, h_{t-1}] + b_f), \quad i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \tag{13}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c[x_t, h_{t-1}] + b_c), \quad h_t = o_t \odot \tanh(c_t) \tag{14}$$

LSTM is suitable for modeling consecutive windows, frequency-band sequences, or sequences of local features. Its disadvantages are serial computation and high cost for long signals. A spiking neural network uses discrete spikes and synaptic-plasticity rules to approximate neural information transmission; it is biologically inspired and may support low-power neuromorphic deployment.

Sam et al. combined LSTM with a three-dimensional brain-template SNN for multilevel depression classification defined by BDI. Average accuracies were approximately 96% in the eyes-open condition and 98% in the eyes-closed condition [8]. The work illustrates the potential of SNNs to visualize brain-region dynamics. Nevertheless, the labels were scale-based severity levels rather than a clinically diagnosed binary MDD task, and the participant population and task differed from those in the studies by Wu et al. and Li et al. and in the GCTNet study. The reported performance should therefore not be ranked directly against MDD-versus-control results.

ISSN: 2664-9640

4.3. Transformers and EEG Conformer

A Transformer uses self-attention to calculate relations between arbitrary positions in a sequence and can learn long-range dependencies in parallel [31]. Scaled dot-product attention is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

A pure Transformer has weak inductive bias for local waveforms, and the computational cost of attention increases quadratically with sequence length. EEG Conformer first uses temporal and spatial convolution to extract local features, treats temporal patches as tokens, and applies a lightweight Transformer before classification [32]. The original paper evaluated motor-imagery and emotion-recognition datasets and provided class-activation topographies. Removing the Transformer substantially reduced mean performance, indicating that global dependencies are valuable for general EEG decoding [32].

EEG Conformer is not an MDD-detection study. It provides a methodological backbone and potential pretraining architecture, but its motor-imagery accuracy is not clinical evidence for depression diagnosis. When a generic EEG architecture is transferred to MDD, subject-independent performance, the stability of attention, and the neuroanatomical interpretation must all be re-evaluated.

4.4. Graph Convolutional Networks: Incorporating Electrode Relations

Multichannel EEG can be represented as a non-Euclidean graph through physical or functional relations between electrodes; graph neural networks therefore provide a natural framework for multichannel EEG. This rationale is supported by spatial-information-based depression classification [33] and by general graph-neural-network and geometric-deep-learning theory [34-39]. The propagation rule of a classical graph convolutional network is [34]:

$$H^{(l+1)} = \sigma(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H^{(l)}W^{(l)}) \tag{16}$$

Here, the augmented adjacency matrix A + I introduces self-loops, the corresponding degree matrix is D, and H contains node features. Graph attention networks learn adaptive weights for neighboring nodes [35], whereas GraphSAGE aggregates neighborhoods and supports inductive inference on unseen graphs [36]. In EEG applications, node features may be raw signals, band powers, entropy measures, or hybrid features; adjacency matrices may be based on physical distance, correlation, coherence, PLV, or learnable parameters.

W-GCN-GRU first used Spearman correlation to select six sensitive features from 23 handcrafted descriptors and then weighted them according to AUC. A correlation matrix served as the GCN adjacency, the weighted features served as node attributes, and a GRU modeled temporal information. The reported best accuracy was 94.72% [13]. The model combines feature selection, a spatial graph, and temporal learning, but the handcrafted weights and correlation graph may change substantially across datasets; cross-site transfer remains to be demonstrated.

4.5. GCTNet: Parallel Fusion of Spatial Graphs and Global Temporal Context

GCTNet contains a residual graph-convolution module, a Transformer module, and a readout module [9]. Its residual design follows the broader principle of residual learning [40]. Preprocessed EEG enters two parallel branches. The ResGCN extracts spatial relations between

ISSN: 2664-9640

electrodes from a functional graph reconstructed using PLV. The Transformer combines convolution, positional embedding, and a learnable classification token to capture global temporal dependencies. The readout module fuses the two branches and maps the representation into a classification space and a projection space. A contrastive cross-entropy loss combines class supervision with an explicit clustering constraint on the learned features.

On the in-house dataset, GCTNet achieved a mean accuracy of 74.84% and an AUC of 0.7693. On the public dataset, the corresponding values were 95.63% and 0.9755 [9]. The large performance difference shows that architecture alone does not determine results; channel count, signal quality, labels, and population characteristics also matter. Ablation experiments showed performance degradation after removing the Transformer, removing the ResGCN, or replacing the contrastive loss with ordinary cross-entropy. The benefit therefore arose from the joint contribution of the spatial graph, global temporal modeling, and the loss function rather than from one component in isolation.

The methodological value of GCTNet lies in its parallel design, which avoids the information bottleneck of strictly graph-first or sequence-first pipelines, and in its use of ablation and two datasets to support architectural reasoning. Limitations include the manually constructed PLV graph, a sensitivity of only 66.73% on Dataset I, and greater parameter and training cost than conventional models. Clinical use requires further testing of probability calibration, failure cases, and robustness across sites.

$$\mathcal{L}_{\mathrm{CCE}} = \mathcal{L}_{\mathrm{CE}} + \lambda \mathcal{L}_{\mathrm{contrast}} \tag{17}$$

4.6. Integrated Comparison of Deep Architectures

Table 11. Representation capabilities of deep-learning architectures

Model	Local temporal patterns	Long-range dependence	Electrode topology	Primary strength	Critical limitation
CNN	Strong	Limited	Depends on input mapping	Efficient end-to-end learning	Limited receptive field; arbitrary 2-D layouts
LSTM/GRU	Moderate	Strong	Usually implicit	Suitable for continuous dynamics	Serial computation and long-sequence cost
SNN	Spike dynamics	Architecture dependent	Can incorporate a brain template	Biologically inspired and potentially low power	Complex training and reproducibility
Transformer	Usually requires convolution	Strong	Usually implicit	Parallel global-dependency modeling	Data hungry and quadratic attention cost
GCN/GAT	Depends on node features	Limited after several layers	Strong	Uses connectivity and electrode relations	Adjacency sensitivity and oversmoothing
W-GCN-GRU	Handcrafted local features	Modeled by GRU	Strong	Joint feature, spatial, and temporal modeling	Depends on handcrafted features and weights
GCTNet	Supported by convolution	Strong Transformer branch	Strong ResGCN branch	Parallel fusion and contrastive loss	More parameters and incomplete cross-site validation

ISSN: 2664-9640

Table 12. Tasks and results of representative deep-learning studies

Study	Task/label	Model	Validation and result	Appropriate interpretation
Acharya et al. [28]	MDD/HC	Deep CNN	High performance in a small cohort	CNNs can learn features automatically; segment leakage must be excluded
Li et al. [7]	Depression/HC and severity	CNN	Accuracy 66.40%/66.93%	More realistic noise; lower performance should not be dismissed
Sam et al. [8]	Four BDI levels	LSTM + SNN	Eyes open 96%; eyes closed 98%	Scale-defined severity is not equivalent to clinical MDD diagnosis
EEG Conformer [32]	Motor imagery/emotion	CNN + Transformer	State-of-the-art general EEG decoding	Methodological foundation, not MDD clinical evidence
W-GCN-GRU [13]	Depression state	Weighted features + GCN + GRU	Best accuracy 94.72%	Performance depends on handcrafted features and partitioning
GCTNet [9]	MDD/HC	ResGCN + Transformer	Accuracy 74.84% and 95.63%	The dataset gap exposes the problem of cross-study comparability
Song and Zhang [14]	Depression recognition	APSO-VMD + 2D-CNN-BiLSTM-SA	Best accuracy 94.47%	Joint optimization is promising but requires nested validation

4.7. Summary

The evolution of deep learning is not a simple replacement of SVMs with increasingly complex networks. Rather, each architecture expands the information that can be modeled. CNNs capture local spatiotemporal patterns; LSTM and GRU capture sequential dependence; SNNs introduce neural dynamics; Transformers model global relations; and graph neural networks explicitly use electrode topology and functional connectivity. The most promising direction is joint modeling of local waveforms, temporal dynamics, and spatial graphs. As complexity increases, however, the need for strict partitioning, ablation, calibration, and external validation becomes more-not less-important.

5. Datasets, Validation Protocols, and Performance Comparability

5.1. Why Reported Accuracies Cannot Be Ranked Directly

Reported accuracies in EEG-based depression studies range from approximately 60% to almost 100%, but these numbers may describe completely different tasks: clinically diagnosed MDD versus healthy controls, scale-defined severity, prescreening of a depression state, or segment-level classification. They may also arise from different sample sizes, channel configurations, cleaning intensity, and validation procedures. A model evaluated on 5-s windows randomly split from 30 participants does not provide evidence equivalent to a model evaluated on an independent set of participants from four centers.

Random segment-level partitioning is a major source of leakage. Adjacent windows from the same participant share head geometry, electrode impedance, baseline spectra, and noise. A model may therefore perform pseudo-classification by recognizing participant identity. The correct procedure is to partition participants into training, validation, and test sets before windowing within each subset. Normalization, feature selection, graph thresholds, and hyperparameter search must likewise be conducted only within the training folds.

ISSN: 2664-9640

Table 13. Relative strength of validation in representative studies

Study	Sample size	Partitioning	External/independent test	Evidence-strength interpretation
Wu et al. [4]	400 participants; four centers	Five-fold development	Independent test of 120 participants	Relatively strong: multicenter and frozen test set
Zhang et al. [5]	48 participants	Internal classification	None	Moderate to low: valuable mechanistic analysis; generalization unresolved
Shao et al. [6]	30 participants	Internal comparison	None	Exploratory: suitable for multiscale-network hypotheses
Li et al. [7]	82 participants	Internal CNN validation	None	Moderate: realistic noise but limited sample
Sam et al. [8]	Scale-defined groups	Multiclass classification	No clear cross-site test	Different task; not equivalent to MDD diagnosis
GCTNet [9]	85 participants + public data	Subject-independent ten-fold	Two datasets but not a matched external cohort	Moderate to high: strict partitioning; true external testing remains needed
W-GCN-GRU [13]	In-house + MODMA	Within-dataset evaluation	Multiple datasets	Moderate to high if all data-driven steps remain independent
Song and Zhang [14]	Public data	Comparative experiments	No prospective external test	High methodological performance; clinical generalization unresolved

5.2. Evaluation Metrics and Class Imbalance

Accuracy is suitable for a reasonably balanced binary dataset. Sensitivity measures the proportion of patients detected, whereas specificity measures the proportion of healthy participants correctly excluded. Precision depends strongly on disease prevalence. The F1-score balances precision and recall, and AUC measures ranking ability across thresholds but does not specify the consequences of a particular clinical operating point. In screening, the costs of missed cases and false alarms differ, so sensitivity, specificity, positive predictive value, negative predictive value, and decision-curve analysis should be reported together.

Let TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad \text{Specificity} = \frac{TN}{TN + FP} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad F_1 = \frac{2\,\text{Precision}\,\text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

Probability calibration is as important as discrimination. If a model outputs a risk of 0.9, approximately 90% of similar cases should actually be positive. Brier score, calibration curves,

ISSN: 2664-9640

and expected calibration error evaluate probability reliability. Most EEG-MDD studies report only accuracy and AUC, which is insufficient for communicating individual risk.

Table 14. Classification metrics and their clinical interpretation

Metric	Question answered	Strength	Limitation	Recommendation
Accuracy	How many predictions are correct overall?	Intuitive	Misleading with class imbalance	Report only with sensitivity and specificity
Sensitivity	How many patients are detected?	Important for screening	Higher sensitivity may increase false alarms	Report confidence intervals and threshold
Specificity	How many healthy participants are excluded?	Describes false-alarm control	Depends on the evaluation population	Choose operating point according to resources
Precision/PPV	How many positive predictions are true cases?	Related to clinical workload	Strongly prevalence dependent	Re-estimate in the target setting
F1-score	How well are precision and recall balanced?	Useful for imbalanced data	Ignores true negatives	Cannot replace specificity
AUC	How well are cases ranked across thresholds?	Threshold independent	Does not describe calibration or clinical decisions	Also report PR-AUC and a working point
Brier/calibration error	Are predicted probabilities reliable?	Directly relevant to risk output	Needs sufficient sample size	Essential before clinical deployment

5.3. Performance and Boundaries of Representative Studies

Table 15. Unified interpretation of performance in representative studies

Study	Main result	Validation characteristic	Supported conclusion	Unsupported conclusion
Wu et al. [4]	CV 91.07%; independent test 84.16%	Four centers and independent test	Coherence features and CK-SVM generalize comparatively well	Applicability to all devices and populations is not established
Zhang et al. [5]	Best accuracy 93.31%	Internal validation in 48 participants	PLI graph measures have discriminative potential	The result is not a multicenter clinical accuracy estimate
Li et al. [7]	Accuracy 66.40%; AUC 0.74	82 participants and simplified preprocessing	CNN retains moderate discrimination under realistic conditions	CNN should not be declared ineffective
Sam et al. [8]	Eyes closed 98%; eyes open 96%	Multilevel BDI task	LSTM-SNN is suitable for scale-defined severity modeling	Direct comparison with clinical MDD/HC is invalid
W-GCN-GRU [13]	Best accuracy 94.72%	In-house and MODMA data	Combining spatial graphs with temporal modeling is useful	Dataset and feature-selection dependence cannot be excluded
GCTNet [9]	Dataset I 74.84%; Dataset II 95.63%	Subject-independent ten-fold and two datasets	Parallel graph-Transformer architecture is effective	Dataset differences and limited sensitivity cannot be ignored
Song and Zhang [14]	Best accuracy 94.47%	Comparative experiments on public data	APSO-VMD and the hybrid model are promising	External prospective validation is still required

ISSN: 2664-9640

5.4. Evidence Quality and Reproducibility Checklist

High-quality research should satisfy requirements in data, methods, and reporting. Data reporting should specify inclusion and exclusion criteria, medication, comorbidities, device, reference, and recording paradigm. Methodological reporting should describe subject-level partitioning, ensure that all data-driven operations are performed within training folds, and provide preprocessing, random seeds, and hyperparameters. Result reporting should include a confusion matrix, confidence intervals, calibration, computational complexity, failure cases, and code availability.

Table 16. Proposed grading of evidence quality for EEG-MDD studies

Grade	Data and labels	Validation design	Result reporting	Appropriate conclusion
A	Adequate multicenter sample with clinical diagnosis	Frozen independent external or prospective test	Discrimination, calibration, confidence intervals, failure analysis	Preliminary clinical utility may be discussed
B	Moderate single-center sample with clinical diagnosis	Subject-independent nested cross-validation	Complete metrics and ablation studies	Methodological effectiveness can be supported
C	Small clinical sample or scale-defined labels	Subject-independent internal validation	Primary discrimination metrics	Supports an exploratory hypothesis
D	Random segment split or unclear labels	Potential information leakage	Only the highest accuracy	At most a preliminary engineering demonstration

5.5. Summary

The central task in model comparison is not to order accuracy values from high to low but to determine under which labels, samples, partitions, and preprocessing procedures each value was produced. Independent testing, participant-level partitioning, and multicenter validation often reduce apparent performance while increasing the credibility of the evidence. In a thesis or journal paper, validation design should receive at least the same attention as model architecture.

6. Key Challenges and Future Research Directions

6.1. Small Samples, Heterogeneity, and Cross-site Generalization

MDD is clinically heterogeneous. Patients differ in symptoms, illness duration, medication, sleep, anxiety comorbidity, and cognitive state. EEG is additionally affected by hardware, montage, reference, impedance, and recording environment, so differences between sites may exceed differences between diagnostic groups. In the multicenter study by Wu et al., independent-test accuracy was approximately seven percentage points lower than internal cross-validation accuracy [4]. GCTNet differed by more than 20 percentage points across its two datasets [9]. These observations show that generalization cannot be solved simply by deepening the network.

Future work should establish a minimum acquisition standard and metadata specification, use stratified sampling to balance age, sex, and site, and evaluate leave-one-site-out validation, domain generalization, and test-time adaptation. Models should separate disease-related features from site-related features. Data augmentation should simulate plausible hardware or physiological variation rather than arbitrarily altering the waveform.

ISSN: 2664-9640

6.2. Self-supervised Pretraining and Foundation Models

Large quantities of unlabeled EEG are easier to obtain than clinically confirmed MDD data. Self-supervised learning can learn general representations through temporal-order prediction, masked reconstruction, contrastive learning, cross-channel prediction, or frequency-domain consistency and can then be fine-tuned using a small labeled MDD cohort. Transformers and graph networks can act as pretraining backbones, but their parameter counts should be controlled to avoid training from scratch on very small datasets.

An EEG foundation model should be assessed not only by average accuracy across tasks but also by transfer across devices, channel montages, and sampling rates. A reasonable design uses local convolution as a low-level signal encoder, electrode positions and connectivity as structural priors, and masked temporal or graph-contrastive objectives for pretraining. Diagnostic, medication, and demographic biases in the pretraining data must be audited; otherwise, the model may encode and amplify site bias.

6.3. Dynamic Brain Graphs and Learnable Connectivity

Most studies use a static connectivity matrix computed over an entire recording or a fixed window, although depression-related brain states may fluctuate over time. A dynamic brain graph estimates time-varying connectivity through sliding windows, state-space models, or time-dependent attention and then uses a spatiotemporal GNN to represent transitions. Compared with static PLV, a dynamic method can capture short-lived synchronization, state dwell time, and transition probability, but it also introduces more parameters and higher estimation noise.

Learnable adjacency enables data-driven optimization of edge weights but may weaken neurophysiological interpretability. A more defensible strategy is a prior graph with learnable corrections. Physical distance, PLI/wPLI, or a group-average network can form the base graph, while attention learns a small residual adjustment. Sparsity, symmetry, cross-band consistency, and anatomical constraints can further regularize the learned graph.

6.4. Multimodal Fusion and Federated Learning

EEG alone has limited specificity for MDD because anxiety, insomnia, medication, and fatigue can produce overlapping changes. EEG can be fused with functional near-infrared spectroscopy, eye tracking, heart-rate variability, speech, symptom scales, or clinical text to improve robustness across electrophysiological, hemodynamic, autonomic, and behavioral levels. Decision-level fusion or architectures robust to missing modalities are preferable because not every modality is available in routine clinical practice.

Privacy constraints limit centralized multicenter datasets. Federated learning allows each center to train locally while exchanging parameters or gradients. In EEG, non-independent and non-identically distributed data, device differences, and imbalanced site sizes can make ordinary federated averaging ineffective. Personalized federation, domain alignment, secure aggregation, and differential privacy should be explored together with the effect of privacy noise on sensitivity and calibration.

6.5. Interpretability, Uncertainty, and Clinical Workflow

Interpretability should not be reduced to displaying one attention heat map. A credible explanation should satisfy stability, counterfactual consistency, and neurophysiological plausibility. Explanations should be similar across random seeds and adjacent windows; removing highly important channels should reduce prediction confidence; and highlighted regions should agree with independent statistical or connectivity analyses. The class-activation topographies of EEG Conformer and the t-SNE visualization in GCTNet provide preliminary approaches [9, 32], but they do not yet constitute participant-level clinical explanations.

ISSN: 2664-9640

A deployed model should also quantify uncertainty. For an out-of-distribution participant, severe artifact, or an unusual comorbidity, the system should abstain rather than force a confident label and should recommend rerecording or human review. Deep ensembles, temperature scaling, isotonic regression, and conformal prediction can provide calibrated intervals or controlled error rates. In a clinical workflow, the algorithm is better positioned for prescreening, repeat-test prompts, and risk stratification than for independent diagnosis.

6.6. Lightweight, Low-channel, and Wearable Deployment

High-density EEG provides richer spatial information but requires long preparation time and higher cost, limiting use in community clinics and homes. Single-channel or few-channel studies suggest that prefrontal EEG has some prescreening potential [27], although low channel count inevitably removes network information. Knowledge distillation, channel selection, low-rank graph convolution, quantization, and edge inference may reduce cost while retaining critical frontotemporal or frontoparietal relationships.

Deployment studies should report parameter count, floating-point operations, inference latency, memory, battery consumption, and recalibration frequency. A lightweight model must also tolerate electrode displacement and household noise, and should include a signal-quality index and an automatic rerecording mechanism.

Table 17. Current problems, recommended technical directions, and validation requirements

Current problem	Recommended direction	Required additional validation	Expected value
Small samples and expensive labels	Self-/semi-supervised learning; active learning	Learning curves and cross-cohort fine-tuning	Reduce dependence on large labeled datasets
Cross-site distribution shift	Domain generalization; leave-one-site-out; test-time adaptation	External-site testing and subgroup analysis	Improve transfer across devices and hospitals
Insufficient static graphs	Dynamic connectivity and spatiotemporal GNNs	Window sensitivity and state stability	Characterize fluctuations in brain state
Privacy restrictions	Federated learning and secure aggregation	Non-IID simulation and privacy-performance curves	Enable multicenter collaboration
Weak interpretability	Prior constraints, counterfactuals, stability analysis	Deletion/insertion tests and expert agreement	Produce credible neurophysiological evidence
Unreliable probabilities	Calibration and conformal prediction	Brier score, ECE, and coverage	Support risk communication and abstention
High deployment cost	Channel selection, distillation, and quantization	Latency, energy, and home-noise testing	Support wearable and primary-care use
Insufficient clinical-utility evidence	Prospective and multistage decision studies	Decision curves, workload, and patient outcomes	Move from model metrics to clinical value

6.7. Summary

The major future direction is not simply increasing network depth. It is expanding the amount and diversity of effective training data, controlling distribution shift, developing dynamic and interpretable representations, and incorporating probability reliability and deployment cost into the optimization objective. Only when a model remains stable at an independent site, under realistic noise, and within an actual clinical workflow can rs-EEG-based systems progress from exploratory biomarkers to decision-support tools.

ISSN: 2664-9640

7. Conclusion

This review systematically analyzed rs-EEG acquisition, preprocessing, adaptive decomposition, spectral and nonlinear features, functional connectivity, conventional machine learning, deep learning, Transformers, and graph neural networks for intelligent MDD detection. Conventional approaches based on interpretable handcrafted features and SVMs remain valuable for small-sample and low-channel scenarios. Multicenter evidence indicates that coherence-based connectivity may be more stable than isolated-channel power. Functional brain-network analysis extends MDD research from local oscillatory abnormalities to abnormal network organization, and measures such as PLI, clustering coefficient, and path length provide a bridge to mechanistic interpretation.

Deep learning enables automatic learning of local and temporal representations. CNNs may show high performance on intensively cleaned data, but conservative results on more realistic clinical recordings expose the true difficulty of generalization. LSTM-SNN demonstrates the potential to model multilevel depression and brain dynamics, but scale-defined tasks must be distinguished from clinical diagnosis. EEG Conformer shows that combining convolution with global attention is a useful general EEG backbone. GCN, W-GCN-GRU, and GCTNet further integrate electrode topology, functional connectivity, and temporal dependence into unified models.

Across the literature, high accuracy is credible only when labels are reliable, participants are independent across splits, preprocessing avoids leakage, hyperparameters are selected in nested validation, and the final system is tested independently. Future work should treat multicenter generalization, self-supervised pretraining, dynamic brain graphs, federated learning, calibration, and interpretability as central problems while also addressing low-channel acquisition, lightweight inference, and prospective clinical evaluation. The most realistic role of rs-EEG is not to replace diagnosis, but to provide an objective complement to structured interviews and symptom scales for screening, repeated assessment, risk stratification, and longitudinal treatment follow-up.

Acknowledgments

It is now 1:01 a.m., and the low hum of the air conditioner reverberates through the room. As I sit at my desk revising this manuscript, I suddenly recall another early morning three years ago, when I completed the final revision of my undergraduate thesis beneath the lamplight. Time has passed so swiftly that the memory feels almost dreamlike.

The verse I once included in the acknowledgements of my undergraduate thesis still captures my feelings today: “The day will come when I ride the wind and cleave the waves; then I shall set my cloudlike sail and cross the boundless sea.”

References

[1] Cai, H., Han, J., Chen, Y., et al. (2018). A pervasive approach to EEG-based depression detection. *Complexity*, 2018, Article 5238028. https://doi.org/10.1155/2018/5238028

[2] Čukić, M., López, V., & Pavón, J. (2020). Classification of depression through resting-state electroencephalogram as a novel practice in psychiatry: Review. *Journal of Medical Internet Research*, 22(11), Article e19548. https://doi.org/10.2196/19548

[3] Greco, C., Matarazzo, O., Cordasco, G., et al. (2021). Discriminative power of EEG-based biomarkers in major depressive disorder: A systematic review. *IEEE Access*, 9, 112850–112870. https://doi.org/10.1109/ACCESS.2021.3103047

[4] Wu, C.-T., Huang, H.-C., Huang, S., et al. (2021). Resting-state EEG signal for major depressive disorder detection: A systematic validation on a large and diverse dataset. *Biosensors*, 11(12), Article 499. https://doi.org/10.3390/bios11120499

[5] Zhang, B., Yan, G., Yang, Z., Su, Y., Wang, J., & Lei, T. (2021). Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 215–229. https://doi.org/10.1109/TNSRE.2020.3043426

[6] Shao, X., Sun, S., Li, J., Kong, W., Zhu, J., Li, X., & Hu, B. (2021). Analysis of functional brain network in MDD based on improved empirical mode decomposition with resting state EEG data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 1546–1556. https://doi.org/10.1109/TNSRE.2021.3092140

[7] Li, M., Liu, Y., Liu, Y., et al. (2022). Resting-state EEG-based convolutional neural network for the diagnosis of depression and its severity. *Frontiers in Physiology*, 13, Article 956254. https://doi.org/10.3389/fphys.2022.956254

[8] Sam, A., Boostani, R., Hashempour, S., Taghavi, M., & Sanei, S. (2023). Depression identification using EEG signals via a hybrid of LSTM and spiking neural networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 4725–4737. https://doi.org/10.1109/TNSRE.2023.3336467

[9] Wang, Y., Peng, Y., Han, M., et al. (2024). GCTNet: A graph convolutional transformer network for major depressive disorder detection based on EEG signals. *Journal of Neural Engineering*, 21(3), Article 036042. https://doi.org/10.1088/1741-2552/ad5048

[10] Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198. https://doi.org/10.1038/nrn2575

[11] Stam, C. J. (2014). Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15(10), 683–695. https://doi.org/10.1038/nrn3801

[12] Akar, S. A., Kara, S., Agambayev, S., & Bilgiç, V. (2015). Nonlinear analysis of EEGs of patients with major depression during different emotional states. *Computers in Biology and Medicine*, 67, 49–60. https://doi.org/10.1016/j.compbiomed.2015.09.019

[13] Wang, Z., Hu, C., Liu, W., Zhou, X., & Zhao, X. (2024). EEG-based high-performance depression state recognition. *Frontiers in Neuroscience*, 17, Article 1301214. https://doi.org/10.3389/fnins.2023.1301214

[14] Song, Y., & Zhang, B. (2026). Depression recognition based on improved variational mode decomposition of electroencephalogram signals. *Journal of Biomedical Engineering*, 43(1), 45–52, 60. [In Chinese.]

[15] Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

[16] Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, Article 156869. https://doi.org/10.1155/2011/156869

[17] Huang, N. E., Shen, Z., Long, S. R., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

[18] Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544. https://doi.org/10.1109/TSP.2013.2288675

[19] Hosseinifard, B., Moradi, M. H., & Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer Methods and Programs in Biomedicine*, 109(3), 339–345. https://doi.org/10.1016/j.cmpb.2012.10.008

[20] Ahmadlou, M., Adeli, H., & Adeli, A. (2012). Fractality analysis of frontal brain in major depressive disorder. *International Journal of Psychophysiology*, 85(2), 206–211. https://doi.org/10.1016/j.ijpsycho.2012.05.001

[21] Mahato, S., & Paul, S. (2019). Detection of major depressive disorder using linear and non-linear features from EEG signals. *Microsystem Technologies*, 25(3), 1065–1076. https://doi.org/10.1007/s00542-018-4075-z

[22] Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

[23] Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11), 1178–1193. https://doi.org/10.1002/hbm.20346

[24] Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. A. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4), 1548–1565. https://doi.org/10.1016/j.neuroimage.2011.01.055

[25] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442. https://doi.org/10.1038/30918

[26] Mumtaz, W., Xia, L., Ali, S. S. A., et al. (2017). Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical Signal Processing and Control*, 31, 108–115. https://doi.org/10.1016/j.bspc.2016.07.006

[27] Wan, Z., Zhang, H., Huang, J., Zhou, H., Yang, J., & Zhong, N. (2019). Single-channel EEG-based machine learning method for prescreening major depressive disorder. *International Journal of Information Technology & Decision Making*, 18(5), 1579–1603. https://doi.org/10.1142/S0219622019500342

[28] Acharya, U. R., Oh, S. L., Hagiwara, Y., et al. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Computer Methods and Programs in Biomedicine*, 161, 103–113. https://doi.org/10.1016/j.cmpb.2018.04.012

[29] Uyulan, C., Ergüzel, T. T., Unubol, H., et al. (2021). Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clinical EEG and Neuroscience*, 52(1), 38–51. https://doi.org/10.1177/1550059420916634

[30] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[31] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008).

[32] Song, Y., Zheng, Q., Liu, B., & Gao, X. (2023). EEG Conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 710–719. https://doi.org/10.1109/TNSRE.2022.3230250

[33] Jiang, C., Li, Y., Tang, Y., & Guan, C. (2021). Enhancing EEG-based classification of depression patients using spatial information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 566–575. https://doi.org/10.1109/TNSRE.2021.3059429

[34] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*.

[35] Veličković, P., Cucurull, G., Casanova, A., et al. (2018). Graph attention networks. In *Proceedings of the International Conference on Learning Representations*.

[36] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 1024–1034).

[37] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

ISSN: 2664-9640

[38] Zhou, J., Cui, G., Hu, S., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

[39] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42. https://doi.org/10.1109/MSP.2017.2693418

[40] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90