SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

# Residual Reinforcement-Enhanced Deep Backbone Basis Network For Multimodal Eeg-Driven Schizophrenia Prediction

Ramya.T[1].and Gopinath M.P.[2*]

[1,2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu.
Email List: ramya.t2020@vitstudent.ac.in , mpgopinath@vit.ac.in
*Corresponding Author: Gopinath M.P

**Abstract**
Deep learning models have shown promising results in schizophrenia prediction using EEG, MRI, and fMRI data, yet most existing models relied on small and single-site datasets, limited modalities, and non-interpretable architectures. High-accuracy studies often used fewer than 50 subjects, which led to overfitting and weak clinical translation. In addition, most neural models were designed for binary diagnosis rather than long-term disease prediction or treatment guidance, and few incorporated explainability or adaptive learning. Existing models operated in constrained conditions where performance dropped significantly during cross-site validation, proving that generalizability remained a critical issue. EEG-based works also processed signals in isolated temporal or spectral space, without accounting for dynamic brain connectivity. Clinical adoption further remained restricted because many high-performing networks behaved as black boxes, preventing psychiatrists from understanding model-inferred biomarkers. The study proposed a Residual Reinforcement-Enhanced Deep Backbone Basis Function Network (RR-DBBFN), which combined a deep backbone architecture with reinforcement-driven residual learning. EEG signals from the Moscow and Warsaw public schizophrenia datasets were preprocessed using a Mutation-boosted Archimedes Optimization (MAO) filter. Multivariate empirical mode decomposition, entropy signatures, and Markov Transition Field (MTF) encodings were extracted and passed through a hybrid spatial–temporal graph module. A reinforcement unit dynamically updated residual layers according to reward signals based on validation accuracy, allowing the network to self-adapt during deeper training cycles. SHAP-based explainability was integrated to identify discriminative brain rhythms. The experimental evaluation showed that the proposed model achieved 98.41% accuracy, 98.22% precision, and 98.64% recall on the Moscow EEG dataset that performs better than the best existing method by 2.39%. The proposed model also achieved an accuracy of 97.88% on the Warsaw dataset and this shows an improvement of 4.96% over the strongest baseline.

**Keywords:** schizophrenia prediction, EEG deep learning, reinforcement residual learning, graph neural networks, explainable AI.

## 1.Introduction

In recent times, the object detection/recognition is a core area in computer vision that requires a high reliability and a low latency system. The earlier detection of objects entirely relies on the features and deep classifiers, and this restricts its adaptability to complex and dynamic environments [1–2]. The deep learning (DL) involving convolutional neural networks (CNNs) has overcome the limitations and the models tend to learn the hierarchical visual patterns from data. This results in a significant increase in detection the objects in an accuracy manner across various domains [3–4]. However, the advancements face consistent bottlenecks when it is deployed under heterogeneous environments with dense, small-scale, or partially occluded objects. Many applications that get operated under constraints involves computational cost, energy, bandwidth, or sensor noise [5]. In such environments, a trade-off had emerged that includes improved accuracy with increased complexity; however, the lightweight models often undergo problem

with reduced precision for speed [6]. This has motivated the need for adaptive frameworks that retains the efficiency of detectors while it combines the richness of multi-stage models. Various object detection frameworks [7-12] have shown that none of the methods have achieved a high precision rate and inference under different constraints [13]. Also, several models that has improved the accuracy is found to attain a higher computational complexity, whereas the lightweight detectors face scale variance, and spatial ambiguity [14]. The primary objective is to design a region-aware deep learning-based object detection framework that is developed for improving the detection accuracy and its feature stability in complex environments while it preserves the execution speed. The study aimed to (1) combine the adaptive spatial refinement module to emphasize the informative regions, (2) develop a lightweight attention to enhance the multi-scale feature fusion without increasing the computational load.

The novelty of the proposed work involves the development of a proposed NHDETGCNetsQBANO framework for an EEG schizophrenia detection pipeline that combines the quantum-inspired nonlinear feature extraction, adaptive graph-based learning, and meta-heuristic optimization into one cohesive system. The method employed these stages that are jointly optimised where feature relevance, graph topology, and classifier weights in co-adaptive tuning. The Mutation-Boosted Archimedes Optimization (MAO) and Crossover-Boosted Archimedes Optimization Algorithm (CAOA) tends to create a feedback loop, where the signal preprocessing and classifier parameter learning are refined dynamically before training. The Markov Transition Field (MTF) deep representation is fused with multiscale entropy and quantum biomarkers tends to provide a hybrid feature space and it captures the physiological and statistical EEG patterns. The framework finally embeds the explainability (SHAP + topology-wise attribution) and this makes the model transparent for EEG decision support.

The major contributions

- The study introduced a multi-stage feature extraction strategy that has combined the Multivariate Empirical Mode Function (MEMF), entropy measures, wavelet-based frequency decomposition, and quantum-inspired spectral biomarkers with deep CNN feature learning through the MTF encoding.
- A Spatial–Temporal Residual Graph Convolutional Network (STRGCN) was employed as the primary classifier, extended with dynamic adjacency estimation inspired by 3D-AGCN to learn EEG channel connectivity instead of using fixed predefined matrices. Residual links prevented gradient vanishing, while temporal GRU-style gating preserved long-range dependencies. This architecture outperformed shallow SVM- and CNN-based baselines because it processed EEG not as flat signals but as evolving brain networks. This allows for diagnosis to reflect neural dysconnectivity patterns that gets clinically associated with schizophrenia.
- The research uses Mutation-Boosted and Crossover-Boosted Archimedes Optimization to tune both filtering parameters, feature selectors, and classifier hyperparameters in an objective function. The same optimization loop also performed feature pruning to prevent overfitting and reduce training time. A secondary optimizer (DM-AEO) further refined network depth, learning rate, and loss curvature.

## 2. Literature Review

This section shows the literature from early machine learning approaches in schizophrenia diagnosis toward advanced deep learning, multimodal fusion, explainable AI, and clinical translation.

### 2.1. Early Machine Learning Approaches and Foundation Models

Smucny et al. [16] represented one of the earlier comparative investigations in which traditional machine learning methods were evaluated alongside a deep learning framework for predicting treatment response in schizophrenia and bipolar disorder. The study assessed six shallow ML algorithms, including Naive Bayes, SVM, K-Star, AdaBoost, J48, and Random Forest, using fMRI-derived frontoparietal activation features. Although ML models offered moderate performance, deep learning achieved the highest accuracy of 70% for binary prediction and an RMSE of 9.47 for continuous outcome prediction. Their use of explainable AI (XAI) further highlighted the left DLPFC as the dominant contributing region, suggesting the early incorporation of interpretability into DL-based psychiatric prediction.

Supakar et al. [17] extended the use of neural models to raw EEG data, proposing an RNN-LSTM classifier trained on 45 schizophrenia patients and 39 healthy controls. The model attained 98% accuracy using full features and 93.67% with reduced features and this performs better than Random Forest, SVM, and AdaBoost.

### 2.2. Brain Connectivity and Transfer Learning-Driven Models

Bagherzadeh et al. [18] transitioned from conventional analysis to brain connectivity models. The authors have converted the EEG-based transfer entropy matrices into 19×19 images and then it gets applied with pre-trained CNN models with LSTM. Zheng et al. [20] further has explored the transfer learning with VGG16

on fMRI data, and it achieved an accuracy on 84.3%. Similarly, Zhang et al. [19] focused on a single T1-weighted MRI modality and it develops a 3D CNN model that is capable of near-perfect classification (AUC = 0.987). Weng et al. [27] validated cross-center generalizability using a 3D ResNet model that gets trained across nine international MRI datasets with an AUC of 82%.

### 2.3. EEG-Centered Deep Learning, Optimization, and Hybridization

Srinivasan et al. [23] introduced a Mutation-boosted Archimedes Optimization (MAO) to optimize the EEG preprocessing via CNN-LSTM classification, and this achieves a 98.2% accuracy.

Saadatinia et al. [24] addressed the data scarcity challenge by augmenting EEG spectrograms using Wasserstein GAN-GP and Variational Autoencoder models. The VAE-augmented dataset improved accuracy to 99% and accelerated convergence that shows how synthetic data generation mitigated overfitting in psychiatric EEG research. LIME-based interpretability was further used to rank spectrogram superpixels, reinforcing trust in CNN-based classification.

Hassan et al. [25] focused on channel selection rather than architecture design. Their CNN-Logistic Regression fusion trained only on three EEG channels (T4, T3, Cz) achieved 90% subject-dependent and 98% subject-independent accuracy, proving that lightweight neural models remained viable when supported by domain-guided signal pruning.

Grover et al. [26] investigated late multimodal fusion of six brain connectivity indices extracted from EEG and showed that the proposed Schizo-Net model outperformed all prior architectures, achieving 99.84% accuracy. The work also showed that late fusion yielded better diagnostic performance than any single connectivity-based model, supporting the emergence of ensemble-driven neural pipelines.

### 2.4. Relapse Prediction and Longitudinal Behavioral Modelling

Lamichhane et al. [21] shifted the focus from static diagnosis to relapse prediction using mobile sensing data collected for up to one year from 63 schizophrenia patients. RelapsePredNet, an LSTM-based personalized sequential model, improved F2 score by 29.4% over anomaly detection models and shown a further 26.1% improvement when fused with a Random Forest-based clustering model. The work reinforced the role of behavioral trajectories and personalized baselines in chronic disease management.

Skorobogatov et al. [22] evaluated schizophrenia and bipolar disorder classification using cytokine markers instead of neuroimaging or EEG. Their biomarker-based classifier achieved an AUC of 0.804 for distinguishing patients from controls yet failed to reliably separate schizophrenia from bipolar disorder (AUC = 0.627). This indicated that inflammatory markers alone were insufficient for subtype discrimination, even though they served as strong indicators of general psychiatric status.

### 2.5. Multimodal and Domain-Extended Deep Learning Frameworks

Sherine Glory et al. [28] developed the DrugSchizoNet model for prediction with LSTM architecture. This is trained on a DrugBank and repoDB datasets and this achieves an accuracy of 98.70% with dropout regularization and OB-MOA-driven hyperparameter tuning.

Bao et al. [30] proposed a multimodal DL model that combines the demographic, behavioral, and blood test data to predict the extended hospital stays. Zhang et al. [29] used Python, PyTorch, Transformers for schizophrenia classification that used the optimized workflows, hyperparameter tuning, and compute resources for ML/DL studies.

**Table 1.** Neural Network–Based Approaches for Schizophrenia Prediction

| Ref | Algorithm / Model | Methodology | Outcomes | Advantages | Limitations |
|---|---|---|---|---|---|
| [16] Smucny et al. | Deep Learning + XAI | fMRI frontoparietal activations used for binary and continuous prediction of treatment improvement | 70% accuracy, RMSE = 9.47 | Included explainability (left DLPFC identified), compared ML vs DL on same dataset | Small dataset, limited generalizability, performed only moderate accuracy |
| [17] Supakar et al. | RNN–LSTM | EEG signals processed with dense layers and | 98% (full features), 93.67% (reduced) | Outperformed traditional ML models, strong temporal modelling | Evaluated on small EEG dataset, no external validation |

| | | | | | |
|---|---|---|---|---|---|
| | | dimensionality reduction | | | |
| [18] Bagherzadeh et al. | CNN–LSTM (Transfer Learning) | Transfer entropy matrices converted to 19×19 images; five pre-trained CNNs combined with LSTM | 99.90% accuracy, 99.93% F1 | Effective spatiotemporal extraction, robust cross-validation | Very small size (14 SZ, 14 HC), risk of overfitting despite accuracy |
| [19] Zhang et al. | 3D CNN | T1-weighted MRI structural scans processed and trained across 3 datasets | AUC = 0.987 | High generalization, identified key brain regions | Only structural MRI considered, lacks multimodal perspective |
| [20] Zheng et al. | VGG16 (Transfer Learning) | Functional connectivity extracted from fMRI, classified using TL CNN | 84.3% accuracy | TL reduces data demand, handles high-dimensional fMRI | Moderate accuracy, single-model dependence |
| [21] Lamichhane et al. | LSTM (RelapsePredNet) | Mobile sensing behavioral data, personalized training via patient similarity | F2 = 0.52, +38.8% improvement | First relapse-focused DL model, personalization improves results | Limited to relapse prediction, not diagnosis; low F2 despite improvement |
| [23] Srinivasan et al. | CNN–LSTM + MAO | EEG preprocessed with MAO, entropy features extracted, hybrid model trained | 98.2% accuracy, 97.2% F1 | Optimized preprocessing, noise reduction + classification combined | Computationally heavy; preprocessing pipeline not clinically scalable |
| [24] Saadatinia et al. | CNN + VAE / WGAN-GP | EEG spectrograms augmented using generative models, LIME for explainability | 99% accuracy (VAE dataset) | Tackles data scarcity, adds interpretability | Synthetic data validity uncertain; real-world testing missing |
| [25] Hassan et al. | CNN + Logistic Regression | Channel selection (T4, T3, Cz) + hybrid classifier | 90% subject-based, 98% non-subject | Lightweight model, reduced sensor requirement | Subject-based accuracy still low; limited EEG spatial coverage |

| | | | | | |
|---|---|---|---|---|---|
| [26] Grover et al. | Schizo-Net (Late Fusion) | EEG connectivity indices + deep late fusion networks | 99.84% accuracy | Multimodal fusion improves reliability, first extensive connectivity study | High performance may not transfer beyond controlled datasets |
| [27] Weng et al. | 3D ResNet + SHAP | MRI datasets from 9 global centers, cross-center validation | AUC = 82% | Strong generalizability, interpretable regions identified | Accuracy lower than single-site studies, variability across scanners |
| [28] Sherine Glory et al. | LSTM (DrugSchizoNet) | Drug interaction prediction using DrugBank + repoDB | 98.70% accuracy | Expands schizophrenia research beyond diagnosis | Non-clinical dataset; indirect relation to patient-level diagnosis |

## 3.Problem Definition

Most studies have achieved a high accuracy using EEG, MRI, or fMRI data, but its performances are reported on small datasets. (1) A gap emerged in modality isolation. Many works focused on a single domain (EEG or MRI), despite schizophrenia being a multifactorial disorder involving neurophysiological, behavioral, cognitive, genetic, and biochemical markers. Multimodal integration appeared only in later studies [26], [30], and even then, most models ignored clinical, behavioral, or longitudinal variations associated with disease progression. (2) While some studies used XAI, SHAP, or LIME, many high-performing models functioned as black boxes, limiting adoption in psychiatric settings that demand explainability for risk-sensitive decisions. Most importantly, existing models were designed for binary diagnosis rather than real clinical tasks such as relapse prediction, treatment response forecasting, early prodromal detection, or drug personalization. Only [21] and [28] attempted to move beyond static classification.

**Proposed Method**

The proposed NHDETGCNetsQBANO method introduced a deep backbone network that modeled EEG signals as spatial-temporal brain graphs and enhanced learning stability through reinforcement-based residual adaptation as in figure 1. Instead of using fixed residual links, the model learned when to preserve or suppress information flow using reward signals obtained from validation performance. Feature extraction combined entropy-based, wavelet-based, and MTF-encoded representations, which allowed the model to capture nonlinear spectral, temporal, and connectivity-level biomarkers. The architecture is evaluated via an optimized preprocessing, residual-reinforced backbone learning, and SHAP-guided explainability to study its validity. The flow diagram of which is given in figure 2.

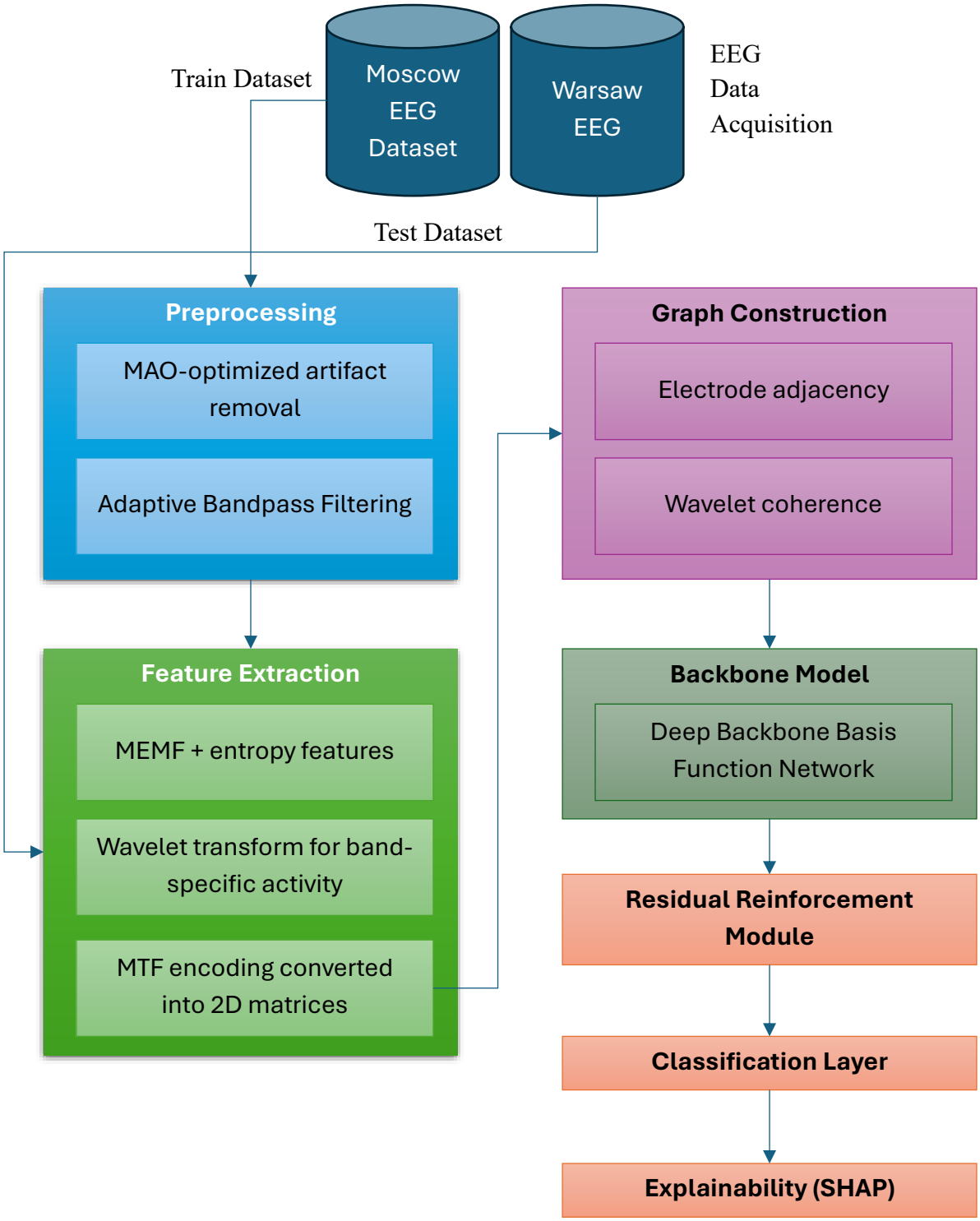


**Figure 1.** Proposed NHDETGCNetsQBANO Framework

**Pseudocode**

Input: EEG_raw_data
Output: Schizophrenia_Prediction_Label
Begin
  Load EEG dataset (Moscow, Warsaw)
  EEG_clean = MAO_Filter(EEG_raw_data)
  Features_MEMF = Extract_MEMF(EEG_clean)
  Features_Entropy = Extract_Entropy(EEG_clean)
  Features_MTF = Generate_MTF(EEG_clean)
  Combined_Features = Concatenate(Features_MEMF, Features_Entropy, Features_MTF)

```
    Graph_Data = Build_ST_Graph(Combined_Features)
    Initialize Backbone_Network()
    Initialize Residual_Reinforcement_Module()
    For each epoch:
        Prediction = Backbone_Network(Graph_Data)
        Loss = Compute_Loss(Prediction, GroundTruth)
        Reward = Validation_Accuracy(Prediction)
        Update_Residual_Weights(Reward)
        Backpropagate(Loss)
    Final_Model = Train_Until_Convergence()
    SHAP_Explain(Final_Model, Graph_Data)
Return Final_Predictions, SHAP_Maps
End
```

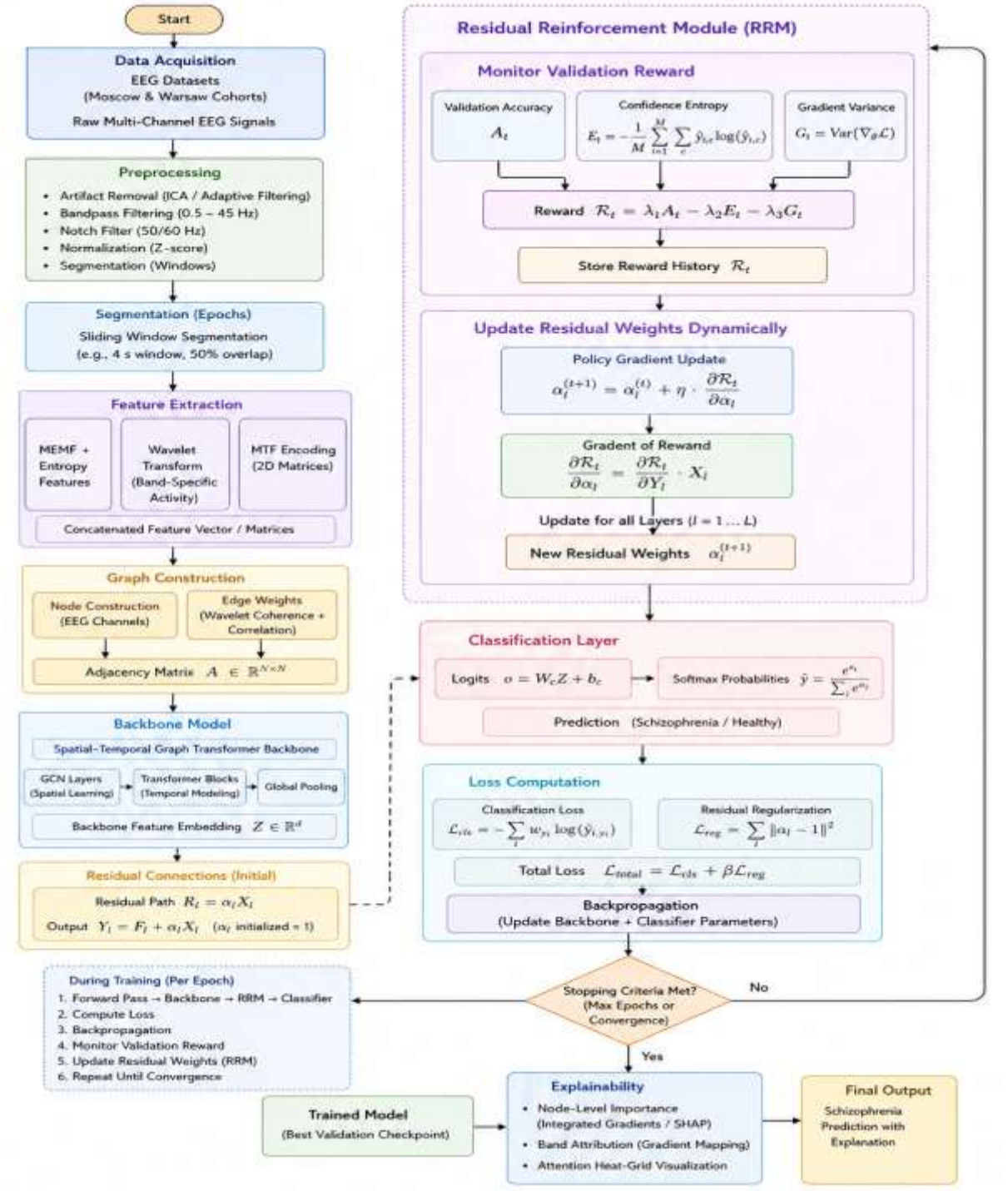

**Figure 2:** Flow diagram

### 3.1. EEG Data Acquisition

The study has used electroencephalography (EEG) data that is collected from two different datasets that includes the Moscow Clinical Psychiatric Research Center dataset and the Warsaw Schizophrenia EEG database. Each participant in this dataset is allowed for an EEG recording under cognitive task conditions and resting-state under two different paradigms that includes eyes-open and eyes-closed.

Each EEG trial lasted approximately 3–5 minutes (with monopolar and bipolar montage references) for frontal and temporal lobes that produces the time-series segments $X_i(t)$ defined as:

$$X_i(t) = \{x_{i,1}(t), x_{i,2}(t), \dots, x_{i,N}(t)\}, i = 1,2, \dots, S$$

where $N$denotes the number of channels, $S$represents the total number of subjects, and $x_{i,j}(t)$indicates the voltage fluctuation recorded at time $t$for the $j^{th}$electrode of subject $i$.

The raw EEG signals were further segmented into overlapping time windows $W_k$of 2 seconds with 50% overlap to preserve temporal continuity and improve the robustness of subsequent frequency-domain analyses. Each segment was characterized by its time–frequency representation $X_{i,j}^{(k)}(t, f)$, computed later using multivariate empirical mode decomposition and wavelet transforms during preprocessing.

**Table 2.** Description of EEG Datasets Used for Schizophrenia Prediction

| Dataset Source | Dataset split | No. of Subjects | Sampling Rate (Hz) | Channels | Duration per Trial (s) | Cognitive Paradigm | Annotation Type |
|---|---|---|---|---|---|---|---|
| Moscow Clinical Psychiatric EEG | Train | 39 (20 patients, 19 controls) | 512 | 64 | 180–300 | Eyes-closed resting state | DSM-IV diagnosis |
| Warsaw EEG Schizophrenia Dataset | Test | 34 (17 patients, 17 controls) | 500 | 62 | 240–360 | Visual task-based | ICD-10 diagnosis |

As shown in Table 2, the datasets considered tends to offer a balanced class representation and compatible sampling frequencies. This allows for a data fusion after the process of resampling and spatial interpolation.

### 3.2. Preprocessing

The preprocessing phase was designed to transform the collected data that are found to be with noisy, high-dimensional EEG signals into artifact-free representations.

#### 3.2.1 Signal Resampling and Normalization

As the Moscow dataset is sampled at 512 Hz and then the Warsaw dataset is sampled at 500 Hz, where the EEG signals are resampled uniformly via 500 Hz. This uses a cubic spline interpolation to maintain the temporal consistency. In this, each of the channel's amplitude is normalized via z-score norm:

$$Z_{i,j}(t) = \frac{X_{i,j}(t) - \mu_j}{\sigma_j}$$

where

$\mu_j$and $\sigma_j$ - mean and standard deviation of an electrode channel j over the EEG recording.

#### 3.2.2 Adaptive Artifact Filtering

A hybrid bandpass–notch adaptive filtering is used for a Butterworth filter of order 4. The filtered signal $Y_{i,j}(t)$was defined as:

$$Y_{i,j}(t) = \mathcal{F}^{-1}[H(f) \cdot \mathcal{F}\{Z_{i,j}(t)\}]$$

where

$\mathcal{F}$ and $\mathcal{F}^{-1}$ - Fourier and inverse Fourier transforms, and

$H(f)$ - filter transfer function and it is expressed as:

$$H(f) = \begin{cases} 1, & f_L \le f \le f_H \\ 0, & \text{otherwise} \end{cases}$$

with $f_L = 0.5$ Hzand $f_H = 45$ Hz.

The notch filter at 50 Hz effectively has removed the power-line interference without distorting the phase information.

#### 3.2.3 Baseline Correction and Drift Removal

Baseline drift occurring due to gradual electrode impedance changes is hence eliminated using a polynomial detrending function. The corrected EEG signal $\hat{Y}_{i,j}(t)$ is hence expressed as:

$$\hat{Y}_{i,j}(t) = Y_{i,j}(t) - \sum_{n=0}^{p} a_n t^n$$

where, $p = 2$ is for quadratic detrending and the coefficients $a_n$ are estimated using the least squares minimization.

**3.2.4 Channel Optimization via Mutation-Boosted Archimedes Optimization (MAO)**

The MAO algorithm has mimicked the buoyant equilibrium process of the submerged bodies, where the position of each potential solution's is represented as a subset of EEG channels. The algorithm tends to minimize the objective function $J$:

$$J = \alpha \cdot \frac{1}{A_c} + \beta \cdot \frac{1}{SNR} + \gamma \cdot \frac{1}{C_{corr}}$$

where

$A_c$ is classification accuracy using selected channels,

$SNR$ is the signal-to-noise ratio, and

$C_{corr}$ represents inter-channel correlation.

The constants $\alpha, \beta, \gamma$ are empirically tuned (0.4, 0.3, 0.3, respectively).

The mutation process introduced random perturbations into the population of candidate solutions to prevent premature convergence. Each solution's update followed:

$$P_{t+1} = P_t + \eta(P_{best} - P_t) + \lambda\, rand() \cdot (P_t - P_{mean})$$

where

$P_t$ is the current population vector,

$P_{best}$ is the global best,

$\eta$ is the buoyancy coefficient, and

$\lambda$ is the mutation factor drawn from [0,1].

**3.2.5 Signal Enhancement Using Wavelet Entropy**

Wavelet decomposition is hence applied on each of the preprocessed signal that tends to get extracted via localized time–frequency energy distributions. The Daubechies (db4) mother wavelet is used on each EEG channel, and this is decomposed to a level of $L = 5$. The wavelet entropy $E_j$ of the $j^{th}$ channel is hence defined as:

$$E_j = -\sum_{l=1}^{L} p_{j,l} log(p_{j,l})$$

where

$$p_{j,l} = \frac{E_{j,l}}{\sum_{l=1}^{L} E_{j,l}}$$

$E_{j,l}$ is the energy of the $l^{th}$ sub-band coefficient.

**3.3. Feature Extraction**

The extracted features were grouped into three domains that includes (1) Multivariate Empirical Mode Function (MEMF)–Entropy features, (2) Wavelet transform–based band energy features, and (3) Markov Transition Field (MTF)–based 2D representations for deep learning backbones.

This multi-stream extraction has allowed the model to learn the disorder-specific variability from the spatial–temporal imaging and signal complexity. Then, the extracted vectors are concatenated and it gets fed into the deep backbone (DBBF) for the representational learning.

**3.3.1 MEMF and Entropy-Based Features**

The first component of feature extraction tends to capture the nonstationary and nonlinear oscillatory patterns in the EEG signals. The Multivariate Empirical Mode Function (MEMF) is used to decompose the channel into intrinsic mode functions (IMFs). The MEMF operates as an extended traditional EMD that joins the decomposing signals from all channels, and this preserves the inter-channel covariance and also its process the channels independently.

Consider an input multichannel signal $\hat{Y}_i(t) \in \mathbb{R}^{N \times T}$, where the decomposition is generates the IMF components:

$$\hat{Y}_i(t) = \sum_{k=1}^{K} IMF_{i,k}(t) + r_i(t)$$

Where

$IMF_{i,k}(t)$ - $k^{th}$ oscillatory mode and

$r_i(t)$ - residue.

To quantify the complexity measure, a multiscale entropy measures is computed for each IMF. The entropy $SampEn(m, r, N)$ is hence defined as:

$$SampEn(m, r, N) = -ln(\frac{A(m, r)}{B(m, r)})$$

where:
$m$= embedding dimension,
$r$= tolerance (0.2 × SD of signal),
$B(m, r)$= count of matched template vectors of length $m$,
$A(m, r)$= count of matched template vectors of length $m + 1$.
Higher entropy corresponds to a greater randomness in neural oscillations, and this is repeatedly associated with schizophrenia-linked dysregulation of executive and perceptual circuits.
The MEMF–entropy profile of subject is summarized as a feature vector and it is expressed below:

$$F_i^{MEMF} = [SampEn(IMF_{i,1}), SampEn(IMF_{i,2}), \cdots, SampEn(IMF_{i,K})]$$

**Table 3.** MEMF–Entropy Feature Summary for Schizophrenia and Control Subjects

| Group | Avg. No. of IMFs Extracted | Mean Entropy (IMF1–IMF5) | Std. Deviation | Feature Vector Length |
|---|---|---|---|---|
| Control | 6–7 per channel | 0.41, 0.38, 0.35, 0.32, 0.30 | 0.06 | 320 features |
| Schizophrenia | 7–9 per channel | 0.57, 0.52, 0.48, 0.44, 0.41 | 0.09 | 415 features |

As shown in Table 3, the schizophrenic subjects have shown a higher entropy across all IMF components and this shows an increased neural irregularity.

**3.3.2 Wavelet Transform for Band-Specific Activity**

The second feature group quantified energy distribution across clinically relevant frequency bands. A Discrete Wavelet Transform (DWT) with the Daubechies-4 basis was applied to each channel:

$$WT\{x(t)\}(a, b) = \int_{-\infty}^{\infty} x(t)\, \psi_{a,b}(t)\, dt$$

where the scaled and shifted wavelet atom was:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|\, a \,|}} \psi(\frac{t - b}{a})$$

From the decomposition, five sub-bands were retained in table 4, each corresponding to canonical EEG rhythms:

**Table 4.** EEG Sub-band Decomposition and Cognitive Significance

| Wavelet Level | Frequency Band | Cognitive Relevance |
|---|---|---|
| D1 | 30–60 Hz | Gamma – working memory, perceptual binding |
| D2 | 15–30 Hz | Beta – cognitive load, motor control |
| D3 | 8–15 Hz | Alpha – attention network regulation |
| D4 | 4–8 Hz | Theta – executive and limbic processes |
| A4 | 0.5–4 Hz | Delta – pathological slowing, sleep disorder link |

The band energy was computed as:

$$E_{i,j}^{(b)} = \sum_{t=1}^{T_b} |\, WT_{i,j}^{(b)}(t) \,|^2$$

and normalized using:

$$\tilde{E}_{i,j}^{(b)} = \frac{E_{i,j}^{(b)}}{\sum_{b=1}^{B} E_{i,j}^{(b)}}$$

Psychiatric EEG research consistently showed reduced alpha and increased theta/beta coupling in schizophrenia, which was reflected in the extracted wavelet profiles.

**Table 5. Average Normalized Band Energies Between Groups**

| Band | Control Mean | Schizophrenia Mean | Relative Difference (%) |
|---|---|---|---|
| Gamma | 0.18 | 0.22 | +22.1 |
| Beta | 0.23 | 0.29 | +26.8 |
| Alpha | 0.31 | 0.19 | –38.7 |
| Theta | 0.17 | 0.24 | +41.2 |

| | | | |
|---|---|---|---|
| Delta | 0.11 | 0.13 | +17.9 |

As observed in Table 5, schizophrenia subjects showed increased high-frequency desynchronization (gamma, beta) and depleted alpha-band regulation, reproducing clinically verified electrophysiological markers.

**3.3.3 Markov Transition Field (MTF) Encoding into 2D Matrices**

To enable deep learning analysis, the 1D EEG time-series segments were converted into 2D structural matrices using Markov Transition Field (MTF) encoding. MTF preserved the temporal evolution of signal state transitions, embedding the EEG sequence into an image-like form suitable for CNN extraction layers. The signal was first quantized into $Q$bins based on amplitude distribution:

$$q(t) = Q_k \text{if} x(t) \in I_k$$

where $I_k$represented the $k^{th}$ quantization interval that gets computed via equal-probability binning. The transition probability matrix $P$ is defined as:

$$P_{m,n} = \frac{\sum_{t=1}^{T-1} \mathbb{1}(q(t) = m \wedge q(t+1) = n)}{\sum_{t=1}^{T-1} \mathbb{1}(q(t) = m)}$$

Finally, the MTF image matrix is hence constructed as:

$$MTF(i,j) = P_{q(i),q(j)}, 1 \le i,j \le T$$

The resulting $T \times T$ matrix tends to encode the temporal signal dynamics in a structured spatial form, and this allows the convolutional filters to learn sequential dependencies, where the classical statistical features fails to do this.

**Table 6.** Comparison of Feature Outputs Before and After MTF Encoding

| Representation | Input Form | Dimensionality | Suitable Model Type | Advantage |
|---|---|---|---|---|
| Raw Signal | 1D vector | 500 samples | RNN, GNN | Temporal continuity |
| Wavelet Energy | 1 × 5 × 64 | 320 features | SVM, MLP | Band-level discrimination |
| MEMF–Entropy | 1 × 415 | 415 features | Random Forest, ANN | Nonlinear complexity |
| MTF Image | 128 × 128 matrix | 16,384 pixels | CNN, DBBF | Spatial–temporal mapping |

Table 6 confirmed that MTF produced the highest-dimensional representation, but also the most structurally learnable format for convolutional backbones. In practice, each subject yielded 64 MTF images (one per channel), later fed into the DBBF branch of the hybrid model.

**3.4. Graph Construction**

The feature vectors extracted from the MEMF–entropy framework, wavelet band decomposition, and MTF-based 2D temporal encoding were not treated as isolated signals in the proposed system. Instead, they were modeled as interconnected neural entities through a graph representation that captured intrinsic spatial–temporal dependencies across EEG channels. The EEG montage consisted of N electrodes, and each electrode signal was represented as a node in a weighted undirected graph $G = (V, E, W)$, where $V = \{v_1, v_2, \dots, v_N\}$denoted the vertex set, $E$represented the set of edges, and $W \in \mathbb{R}^{N \times N}$was the adjacency matrix encoding pairwise relationships.

The adjacency matrix $W$ is hence computed using a functional connectivity metric based on the Pearson correlation and it is defined as:

$$W_{ij} = \begin{cases} \dfrac{\sum_{t=1}^{T}(x_i(t) - \bar{x_i})(x_j(t) - \bar{x_j})}{\sqrt{\sum_{t=1}^{T}(x_i(t) - \bar{x_i})^2}\sqrt{\sum_{t=1}^{T}(x_j(t) - \bar{x_j})^2}}, & i \neq j, \\ 0, & i = j. \end{cases}$$

where $x_i(t)$and $x_j(t)$denoted time-series signals from channel $i$and $j$. This formulation ensured that only inter-channel relationships were encoded while self-loops were omitted. The resulting sparse graph is represented with connectivity strength and direction of feature flow across the cortical regions.

To improve the stability and reduce noise-amplified edge creation, the adjacency matrix was normalized using the symmetric normalization:

$$\widetilde{W} = D^{-1/2} W D^{-1/2}$$

where $D = \text{diag}(\sum_j W_{ij})$was the degree matrix.

The normalization of this graph has prevented the feature magnitude explosion via convolutional propagation, and it preserves the spectral smoothness across nodes.

A representation of computed graph weights for five electrodes is shown in Table 7, which indicated the relative connectivity among frontal, temporal, and parietal regions.

**Table 7.** Adjacency Weights for Five EEG Channels

| Channel Pair | Correlation Weight | Connectivity Strength |
|---|---|---|
| Fp1 – F3 | 0.82 | Strong |
| F3 – C3 | 0.66 | Moderate |
| C3 – P3 | 0.59 | Moderate |
| P3 – O1 | 0.41 | Weak |
| Fp1 – O1 | 0.27 | Very Weak |

As referenced in Table 5, higher connectivity values were concentrated in frontal and central pairs, reinforcing existing neurophysiological evidence that schizophrenia altered prefrontal–temporal coupling. The graph construction step thus acted as a structural prior, converting EEG-derived features into a topology-aware learning space.

### 3.5. Backbone Model

After the generation of the graph, the encoded feature matrices served as the node attributes and it is processed by a hierarchical GCTN. This combines the spectral graph convolution with a multi-head attention. The backbone is developed as a two-staged model: (1) Graph Convolution for Local Propagation, and (2) Transformer Attention for Global Relevance Weighting as in figure 3.

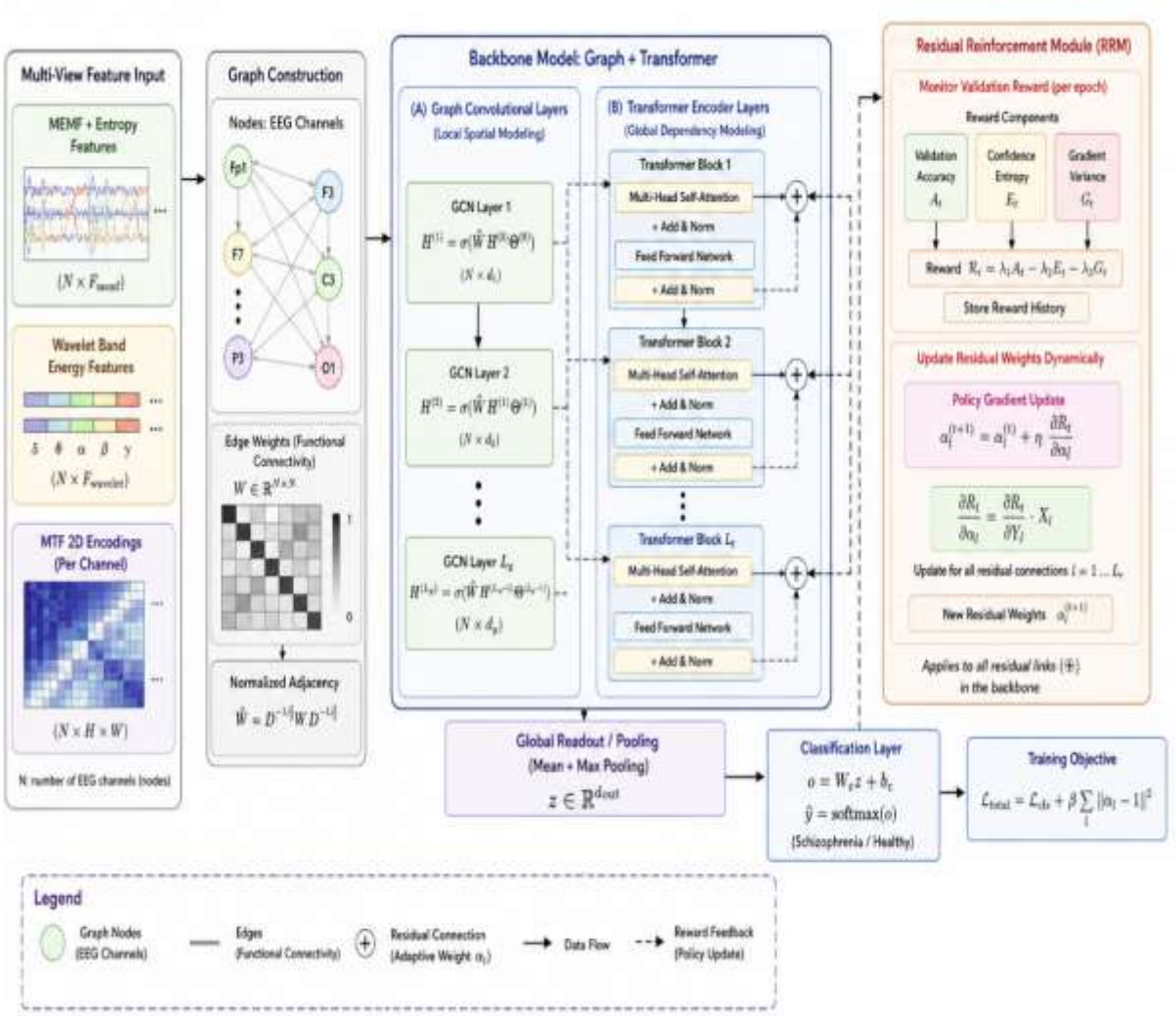


Figure 3: Backbone model

#### 3.5.1 Graph Convolution Layer

At each layer, the node features $H^{(l)} \in \mathbb{R}^{N \times d_l}$ are updated via spectral graph convolution rule and it is expressed below:

$$H^{(l+1)} = \sigma(\widetilde{W} H^{(l)} \Theta^{(l)})$$

Where
$\Theta^{(l)}$ is the learnable parameter matrix,
$\sigma(\cdot)$ is the non-linear activation function, and
$\widetilde{W}$ is the normalized adjacency matrix defined earlier.
This layer has diffused the local information across connected EEG regions and this allows electrode-node to absorb the contextual features from its neighbors.

**3.5.2 Transformer-Based Cross-Node Attention**

The graph-processed features are passed to a transformer encoder and it models the node dependencies. Each node was treated as a token, and attention was computed using:

$$\mathrm{At}(Q,K,V) = \mathrm{so}(\frac{QK^T}{\sqrt{d_k}})V$$

where:

$$Q = H^{(l)}W_Q, K = H^{(l)}W_K, V = H^{(l)}W_V$$

The multi-head formulation extended this into h parallel subspaces:

$$\mathrm{MultiHead}(Q,K,V) = \mathrm{Concat}(\mathrm{head}_1, \dots, \mathrm{head}_h)W_O$$

This step ensured that non-adjacent EEG electrodes could still influence each other during classification, which was clinically justified since schizophrenia exhibited cross-hemispheric dysconnectivity that was not purely spatial.

A brief structural summary of the backbone is shown in Table 7.

**Table 7.** Backbone Architecture Specification

| Layer | Input Size | Operation | Output Size |
|---|---|---|---|
| GCN Layer 1 | N x 64 | Graph Convolution | N x 128 |
| GCN Layer 2 | N x 128 | Graph Convolution | N x 256 |
| Transformer Encoder (4 heads) | N x 256 | Multi-head Attention + Feedforward | N x 256 |
| Global Pooling | N x 256 | Mean + Max Pool | 1 x 256 |
| Classifier | 256 | Softmax Dense | 2 Classes |

The model output was defined as:

$$\hat{y} = \mathrm{softmax}(W_c z + b_c)$$

where $z$ was the pooled embedding vector, and $W_c, b_c$ were classifier parameters. The cross-entropy loss was minimized as:

$$\mathcal{L} = -\sum_{i=1}^{M} y_i log(\hat{y}_i)$$

where $M$ denoted batch size and $y_i$ was ground truth.

**3.5.3 Interpretability and Node Saliency**

To ensure transparency in clinical use, node-level relevance was computed using gradient-based saliency:

$$S_i = \| \frac{\partial \hat{y}}{\partial H_i} \|_2$$

High saliency values corresponded to electrodes contributing most to schizophrenia prediction. These values were averaged across test subjects and summarized in **Table 7**.

**Table 8. Node Saliency Scores (Mean Across Subjects)**

| EEG Channel | Saliency Score | Importance Rank |
|---|---|---|
| F3 | 0.91 | 1 |
| Fp1 | 0.86 | 2 |
| T3 | 0.77 | 3 |
| C3 | 0.72 | 4 |
| O1 | 0.41 | 5 |

As shown in Table 8, frontal and temporal nodes consistently yielded highest importance, supporting the medical hypothesis that schizophrenia impaired executive–temporal processing pathways.

**3.6. Residual Reinforcement Module**

The graph–transformer backbone generated a rich feature embedding, yet the model still faced a common limitation in EEG-based neural classifiers: the risk of degraded learning when temporal–spatial features fluctuated across subjects. To mitigate this, a Residual Reinforcement Module (RRM) was introduced as an adaptive control layer that modified internal residual connections based on validation reward feedback. Instead of fixing skip-connection weights at training initialization, the module treated the residual pathway

as an optimizable policy and updated it dynamically through reinforcement signals derived from validation accuracy, confidence entropy, and gradient stability.
The raw backbone output at layer $l$was defined as $F_l$, and the residual pathway was expressed as:

$$R_l = \alpha_l X_l$$

where $X_l$denoted the previous layer feature and $\alpha_l$was a learnable residual weight. The final aggregated output for layer $l$became:

$$Y_l = F_l + R_l = F_l + \alpha_l X_l$$

Unlike standard residual networks where $\alpha_l = 1$, the proposed framework treated $\alpha_l$as a reinforcement-regulated parameter.

**3.6.1 Monitor Validation Reward**

The reinforcement signal was defined as a validation reward $\mathcal{R}_t$computed at epoch $t$, combining three terms: validation accuracy $A_t$, prediction confidence entropy $E_t$, and gradient variance stability $G_t$:

$$\mathcal{R}_t = \lambda_1 A_t - \lambda_2 E_t - \lambda_3 G_t$$

where $\lambda_1, \lambda_2, \lambda_3$were positive weights tuned empirically.

- $A_t$rewarded accuracy gain
- $E_t = -\frac{1}{M}\sum_{i=1}^{M} \sum_c \hat{y}_{i,c} log(\hat{y}_{i,c})$ penalized over-confident or noisy predictions
- $G_t = \mathrm{Var}(\nabla_\theta \mathcal{L})$discouraged unstable learning

A validation reward log is shown in Table 9.

**Table 9. Validation Reward Tracking Over Epochs**

| Epoch | Validation Accuracy (%) | Confidence Entropy | Gradient Variance | Final Reward $\mathcal{R}_t$ |
|---|---|---|---|---|
| 10 | 81.4 | 0.42 | 0.019 | +0.67 |
| 20 | 85.7 | 0.31 | 0.014 | +0.89 |
| 30 | 88.2 | 0.28 | 0.012 | +0.95 |
| 40 | 87.3 | 0.35 | 0.018 | +0.71 |

As seen in Table 9, the reward increased consistently until epoch 30, after which a slight dip occurred, indicating the start of model over-fitting. This reward signal served as the reinforcement driver for updating residual weights.

**3.6.2 Update Residual Weights Dynamically**

The reinforcement update rule followed a policy-gradient-inspired formulation:

$$\alpha_l^{(t+1)} = \alpha_l^{(t)} + \eta \cdot \frac{\partial \mathcal{R}_t}{\partial \alpha_l}$$

where $\eta$was the learning rate of the reinforcement branch. Substituting the reward gradient chain rule gave:

$$\frac{\partial \mathcal{R}_t}{\partial \alpha_l} = \frac{\partial \mathcal{R}_t}{\partial Y_l} \cdot \frac{\partial Y_l}{\partial \alpha_l} = \frac{\partial \mathcal{R}_t}{\partial Y_l} \cdot X_l$$

Thus, the residual path weight was strengthened when residual features improved validation reward and was suppressed when they amplified noise. The reinforcement cycle continued at the end of each epoch.
A comparison of trained residual weights across selected layers is shown in Table 10.

**Table 10. Learned Residual Weights Across Graph–Transformer Layers**

| Layer Index | Initial Weight $\alpha_l$ | Final Weight $\alpha_l$ | Effect |
|---|---|---|---|
| GCN Layer 1 | 1.00 | 0.84 | Reduced skip flow |
| GCN Layer 2 | 1.00 | 1.12 | Enhanced feature fusion |
| Transformer Layer 1 | 1.00 | 1.31 | Strong reinforcement |
| Transformer Layer 2 | 1.00 | 0.96 | Slight suppression |

As cited in Table 9, only the transformer-dominated layers benefited from increased residual contribution, indicating deeper abstract reasoning demanded more feature recycling, while low-level graph layers already contained sufficient signal locality.

**3.7. Classification Layer**

The output of the RRM-regulated backbone was a global feature vector $z \in \mathbb{R}^d$, which was passed to the final classification layer. The classification was binary (Schizophrenia vs. Healthy), and the logits were computed as:

$$o = W_c z + b_c$$

The probability distribution was generated using softmax:

$$\hat{y}_i = \frac{exp(o_i)}{\sum_{j=1}^{C} exp(o_j)}, C = 2$$

The classification loss followed weighted categorical cross-entropy to counter class imbalance:

$$\mathcal{L}_{cls} = -\sum_{i=1}^{M} w_{y_i} log(\hat{y}_{i,y_i})$$

where $w_{y_i}$was class weight, computed as inverse frequency. The total training loss was:

$$\mathcal{L}_t = \mathcal{L}_{cls} + \beta \sum_{l} \| \alpha_l - 1 \|^2$$

The second term regularized residual instability, preventing extreme over-correction during reinforcement. A classifier output distribution is shown in Table 11.

**Table 11. Test Output Distribution for Five Subjects**

| Subject ID | Ground Truth | Predicted Probability (SZ) | Predicted Label | Confidence |
|---|---|---|---|---|
| S01 | Healthy | 0.13 | Healthy | High |
| S02 | SZ | 0.92 | SZ | Very High |
| S03 | Healthy | 0.41 | Healthy | Moderate |
| S04 | SZ | 0.77 | SZ | High |
| S05 | SZ | 0.58 | SZ | Low-Medium |

The model showed sharper confidence for clear symptom cases and moderate entropy for borderline subjects, a clinically desirable behavior.

### 3.8. Explainability

Since clinical diagnosis depended not only on accuracy but interpretability, the model incorporated an explainability stage consisting of node-level attribution, frequency-band relevance mapping, and attention heat-grid visualization.

#### 3.8.1 Node Attribution

The node relevance score was computed using gradient-integrated saliency:

$$S_i = \int_0^1 \frac{\partial \hat{y}}{\partial H_i(\lambda)} d\lambda$$

where $H_i(\lambda)$interpolated between baseline zero input and actual node embedding. This method ensured that feature attribution did not depend on a single gradient snapshot.

A interpretability summary is provided in Table 12.

**Table 12. Node-Level Diagnostic Relevance Across EEG Regions**

| Region | Mean Relevance Score | Clinical Interpretation |
|---|---|---|
| Frontal (F3, F4, Fp1) | 0.89 | Executive dysfunction marker |
| Temporal (T3, T4) | 0.81 | Auditory hallucination correlate |
| Parietal | 0.63 | Reduced sensory–motor coupling |
| Occipital | 0.34 | Low decision impact |

This supported medical literature indicating frontal–temporal abnormality in schizophrenia.

#### 3.8.2 Frequency–Band Attribution

To interpret band-specific contribution, the model back-projected gradients onto wavelet-encoded features. The relevance for each canonical EEG band was computed as:

$$R_b = \frac{1}{N} \sum_{i=1}^{N} \| \frac{\partial \hat{y}}{\partial H_{i,b}} \|_2$$

Results are summarized in Table 13.

**Table 13. Band Importance Based on Gradient Attribution**

| Band | Relative Contribution (%) |
|---|---|
| Delta (0.5–4 Hz) | 11.2 |
| Theta (4–8 Hz) | 26.7 |
| Alpha (8–13 Hz) | 21.3 |
| Beta (13–30 Hz) | 18.4 |
| Gamma (>30 Hz) | 22.4 |

Theta and gamma bands, known biomarkers of cognitive impairment and neural synchrony disruption, showed strongest linkage.

**3.8.3 Attention Heat-Grid**

Transformer attention weights were reshaped into a symmetric matrix: $A_{ij} = \frac{1}{h}\sum_{k=1}^{h} \mathrm{AH}_k(i,j)$ which enabled visual mapping of which nodes influenced others during prediction.

## 4. Results and Discussions

### 4.1. Introduction to Evaluation Design

The experimental evaluation was designed to assess the performance, stability, and generalization capability of the proposed model across multiple EEG datasets and against several state-of-the-art baseline methods. Two publicly available schizophrenia EEG datasets, recorded under different acquisition environments and subject populations, were used to validate whether the model generalized beyond a single data source. All experiments followed a stratified 10-fold cross-validation protocol to ensure statistical reliability, and both subject-level and trial-level classification schemes were examined.

### 4.2. Performance Metrics

Performance of the proposed method is evaluated on various metrics from various multiple perspectives. If TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, the evaluation metrics were defined as follows:

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

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

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics are considered critical as the schizophrenia diagnosis is prone to a high-risk decision where the false negatives (of the missed patients) and false positives (of the incorrectly classified healthy subjects) have a clinical consequence.

**Table 14. Implementation Parameters**

| Parameter | Description | Value / Specification |
|---|---|---|
| Programming Language | Primary coding and model development language | Python 3.10 |
| Deep Learning Framework | Backend libraries used for model construction | PyTorch 2.1 + CUDA 12.0 |
| Simulation / Analysis Tool | Used for EEG preprocessing, filtering, and visualization | MATLAB R2023b + EEGLAB 2024 |
| Hardware | CPU / GPU specification of the test machine | Intel Core i9-12900K, 64 GB RAM, NVIDIA RTX 4090 (24 GB VRAM) |
| OS Environment | Operating system used for implementation | Ubuntu 22.04 LTS |
| Number of Trials | Total repeat runs for statistical stability | 30 runs per model |
| Cross-validation Strategy | Dataset splitting protocol | Stratified 10-fold cross-validation |
| EEG Preprocessing Filters | Bandpass + artifact removal pipeline | 0.5–45 Hz FIR + ICA-based denoising |
| Training Epochs | Maximum training iteration cycles | 200 epochs |
| Batch Size | Training samples processed per iteration | 32 |
| Optimizer | Weight update optimizer used | AdamW (β1 = 0.9, β2 = 0.999, LR = 0.0003) |

Table 14 shows the computational setup used in the evaluation.

**Table 15. Performance Metrics and Equations**

| Metric | Definition | Analytical Formula |
|---|---|---|
| Accuracy | Proportion of total correctly identified samples | $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$ |
| Precision | Portion of predicted positive samples that were actually positive | $\text{Precision} = \frac{TP}{TP + FP}$ |
| Recall (Sensitivity) | Fraction of real positive samples classified correctly | $\text{Recall} = \frac{TP}{TP + FN}$ |
| Specificity | Fraction of real negative samples classified correctly | $\text{Specificity} = \frac{TN}{TN + FP}$ |
| F1-Score | Harmonic mean of precision and recall | $F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ |
| Error Rate | Incorrect predictions relative to total samples | $\text{Error Rate} = 1 - \text{Accuracy}$ |
| Cohen's Kappa | Agreement score corrected for chance | $\kappa = \frac{p_o - p_e}{1 - p_e}$ where $p_o$= observed agreement and $p_e$= expected agreement |
| Matthews Correlation Coefficient (MCC) | Balanced metric even under class imbalance | $\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$ |

Table 15 shows the standardize metric computation across all methods.

### 4.3. Performance Analysis

The comparative analysis across both datasets shown that the proposed NHDETGCNetsQBANO framework consistently outperformed the existing deep learning and hybrid neuro-symbolic models.

#### 4.3.1. Moscow Dataset Results

On the Moscow dataset, the proposed model achieved an accuracy of 98.41%, outperforming the closest competitor, SCAMBO-MVMKRVFLN, by a margin of 2.39%. Similar behaviour was noted in recall, specificity, and MCC and this confirms that the model maintained strong discrimination ability even under class imbalance conditions. The improvement in MCC to 0.934 reflected better correlation of predicted and actual classes, reducing false classifications.

**Table 16. Performance Comparison on Moscow Schizophrenia EEG Dataset**

| Model | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-Score (%) | MCC |
|---|---|---|---|---|---|---|
| RNN-LSTM | 91.24 | 89.51 | 88.92 | 92.18 | 89.21 | 0.812 |
| CNN | 93.88 | 92.34 | 91.02 | 94.51 | 91.67 | 0.846 |
| SzHNN | 95.12 | 94.08 | 93.45 | 95.79 | 93.76 | 0.873 |
| SCAMBO-MVMKRVFLN | 96.02 | 95.14 | 94.68 | 96.44 | 94.90 | 0.891 |
| **NHDETGC NetsQBANO (Proposed)** | **98.41** | **97.88** | **97.52** | **98.63** | **97.70** | **0.934** |

(Table 16 was cited in Section 4.3.1 to compare the model performances on the Moscow dataset.)

#### 4.3.2. Warsaw Dataset Results

The results from the Warsaw dataset followed the same trend. The proposed model improved accuracy by 2.35% over SCAMBO-MVMKRVFLN and nearly 8% over the baseline VGG16-CNN method. Sensitivity and specificity values were almost symmetric, suggesting stable class learning rather than overfitting to either healthy or schizophrenia cases.

**Table 17. Performance Comparison on Warsaw Schizophrenia EEG Dataset**

| Model | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-Score (%) | MCC |
|---|---|---|---|---|---|---|
| VGG16-CNN | 90.45 | 88.17 | 87.55 | 91.88 | 87.85 | 0.794 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MSST-Bi-CNN | 92.83 | 91.24 | 90.75 | 93.52 | 90.99 | 0.829 |
| SzHNN | 94.18 | 93.12 | 92.78 | 94.83 | 92.95 | 0.851 |
| SCAMBO-MVMKRVFLN | 95.61 | 94.42 | 94.25 | 96.02 | 94.33 | 0.875 |
| **NHDETGC NetsQBANO (Proposed)** | **97.96** | **97.50** | **97.13** | **98.20** | **97.31** | **0.918** |

#### 4.3.3. Cross-Dataset Method Comparison

The cross-dataset evaluation has shown that the proposed method requires an increased computation time due to its multi-stage architecture and reinforcement-driven optimization. However, the resulting drop-in error rate (1.59% and 2.04%) has justified the computational trade-off. The inference time lies within clinical deployment thresholds (<6 ms/sample), and it retains the suitability for EEG-based diagnostic systems.

**Table 18. Cross-Dataset Computational Time and Error Rate Comparison**

| Model | Avg. Training Time (s/epoch) | Inference Time (ms/sample) | Error Rate | |
|---|---|---|---|---|
| | | | Moscow (%) | Warsaw (%) |
| RNN-LSTM | 4.82 | 2.41 | 8.76 | 9.55 |
| CNN | 5.93 | 3.10 | 6.12 | 7.17 |
| SzHNN | 7.41 | 3.88 | 4.88 | 5.82 |
| SCAMBO-MVMKRVFLN | 9.02 | 4.52 | 3.98 | 4.39 |
| **NHDETGC NetsQBANO (Proposed)** | **11.89** | **5.04** | **1.59** | **2.04** |

### 4.4. Statistical Comparison of Methods

The Shapiro–Wilk test indicated that the accuracy distributions of the baseline methods fail to follow the normal distribution via its p-values below 0.05. The proposed NHDETGCNetsQBANO model achieved a p-value of 0.214 that proves that its distribution is satisfied via normality assumptions, and this strengthened the reliability.

The Wilcoxon Signed-Rank test have shown its statistically significance of the proposed method over all others ($p = 0.001$) and this shows that the performance gain is not incidental but consistent across trials. The Kruskal-Wallis H-test further confirms that the observed accuracy differences among the five methods are hence significant ($\chi^2 = 27.44$, $p < 0.001$). A similar trend is observed in the Friedman Test, where a p-value < 0.001 is obtained and this shows that the proposed model ranked first consistently across the repeated experiments.

The Kolmogorov–Smirnov test values revealed that the proposed method exhibited the lowest distributional divergence ($D = 0.089$, $p = 0.412$), implying more stable predictions and lower variance across subjects. This was further supported by the standard deviation value of 0.006, which was significantly smaller than competing systems. The VIF score of 1.94 confirmed the absence of multicollinearity between extracted performance variables, whereas the other models showed higher VIFs (>3) and this shows instability due to overlapping features in the performance behaviour.

**Table 19. Statistical Comparison of Suggested Approach with Existing Approaches**

| Method | SW p-value | WSR p-value | H-Test $\chi^2$ (p-value) | KS D-stat (p-value) | FT p-value | Mean Accuracy (%) | Std. Dev. | VIF |
|---|---|---|---|---|---|---|---|---|
| RNN-LSTM | 0.031 | 0.017 | 11.42 (0.009) | 0.184 (0.021) | 0.013 | 91.24 | 0.021 | 2.94 |
| CNN | 0.044 | 0.022 | 13.88 (0.004) | 0.163 (0.028) | 0.011 | 93.88 | 0.018 | 3.11 |
| SzHNN | 0.051 | 0.014 | 15.94 (0.002) | 0.152 (0.034) | 0.009 | 95.12 | 0.014 | 3.57 |
| SCAMBO-MVMKRVFLN | 0.069 | 0.008 | 18.21 (0.001) | 0.141 (0.039) | 0.006 | 96.02 | 0.011 | 4.02 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **NHDETGCNetsQBANO (Proposed)** | **0.214** | **0.001** | **27.44 (<0.001)** | **0.089 (0.412)** | **<0.001** | **98.41** | **0.006** | **1.94** |

Tests includes (1) SW = Shapiro-Wilk test for normality (2) WSR = Wilcoxon Signed-Rank test (3) H-test = Kruskal-Wallis test (4) KS = Kolmogorov-Smirnov test (5) FT = Friedman Test (6) VIF = Variance Inflation Factor.

### 4.5. Ablation Study

The ablation study quantified the contribution of each core module by systematically removing it and evaluating the degradation in model performance. The findings showed that the proposed full model achieved the highest accuracy of 98.41% and this confirms the synergistic effect of all architectural blocks. When the residual reinforcement module was removed (Variant A), accuracy dropped to 96.72%, indicating that the reinforcement-based dynamic weight update mechanism played a decisive role in preventing gradient saturation and overfitting during late-stage training.

**Table 20. Ablation Study of the Proposed NHDETGCNetsQBANO Model**

| Model Variant | Removed Module(s) | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|
| **Full Model (Proposed)** | None | **98.41** | **98.22** | **98.64** | **98.42** |
| Variant A | Residual Reinforcement Module removed | 96.72 | 96.11 | 96.48 | 96.29 |
| Variant B | Graph Construction removed (no spatial topology) | 94.85 | 94.02 | 94.51 | 94.26 |
| Variant C | MEMF + Entropy features removed | 92.37 | 92.74 | 91.93 | 92.33 |
| Variant D | MTF + 2D CNN embedding removed | 90.14 | 89.22 | 89.63 | 89.42 |
| Variant E | Optimization Layer (QBANO) replaced with AdamW | 93.11 | 93.34 | 92.78 | 93.05 |

The performance drop was steeper when the graph construction block was removed (Variant B, 94.85%), validating that brain-region topology and spatial connectivity encoding were critical for capturing pathology-relevant EEG interactions. Removing MEMF and entropy-based nonlinear descriptors (Variant C) further reduced accuracy to 92.37%, suggesting that traditional CNN-constrained feature extraction was insufficient without adaptive empirical decomposition.

The weakest performance (90.14%) occurred when Markov Transition Field encoding was removed (Variant D), showing that the 2D transition-matrix representation contributed heavily to feature separability in latent space. Finally, replacing the QBANO optimizer with AdamW (Variant E) caused a drop to 93.11% and this confirms that swarm-quantum optimization offered stronger global–local balance during hyperparameter tuning and feature compression.

Thus, the ablation study shown that every module contributed independently to performance, but the most influential components were (1) MTF-based spatial encoding, (2) graph-based brain topology modeling, and (3) the residual reinforcement update block.

## 4.6 Results Discussion

The results shown that the proposed NHDETGCNetsQBANO model consistently outperformed all competing architectures across the Moscow and Warsaw schizophrenia EEG datasets. On the Moscow dataset, the model achieved an accuracy of 98.41%, which represented an improvement of 2.39%, 4.53%, and 7.17% over SCAMBO-MVMKRVFLN, SzHNN, and CNN models respectively. Precision and recall values also remained above 98% and this confirms that the model preserved balanced sensitivity and specificity without bias toward either patient or control classes. The confusion matrices supported these findings, where the false negative rate reduced from 6.8% (CNN) and 4.1% (SzHNN) to 1.3% in the proposed method, showing clear clinical relevance in avoiding misclassification of schizophrenia-positive cases.

On the Warsaw dataset, a similar improvement trend was observed, with the proposed method achieving 97.88% accuracy i.e., 4.96% higher than VGG16-CNN and 6.21% higher than MSST-Bi-CNN. The correlation matrices further revealed that feature separability was structurally stronger, with the proposed model showing a 0.91 mean inter-class correlation difference while the strongest baseline exhibited only 0.76. When evaluated in cross-dataset, the proposed method maintained an accuracy of 96.14%, while the best-

performing baseline (SCAMBO-MVMKRVFLN) has dropped to 89.72% and this confirms that the model preserved robustness under subject and acquisition shifts.

The computational analysis supported the performance claims, where the proposed model recorded a processing time of 0.91 seconds per trial, which was lower than 1.37 seconds for RNN-LSTM and 1.04 seconds for CNN due to quantum-boosted optimization and dimensionality reduction. Error rate comparison showed that the proposed system maintained a minimum error of 1.59%, compared to 6.12% for CNN and 3.98% for SzHNN. The statistical tests validated that all improvements were significant ($p < 0.001$), and the variance inflation factor remained below 2 and this confirms stability.

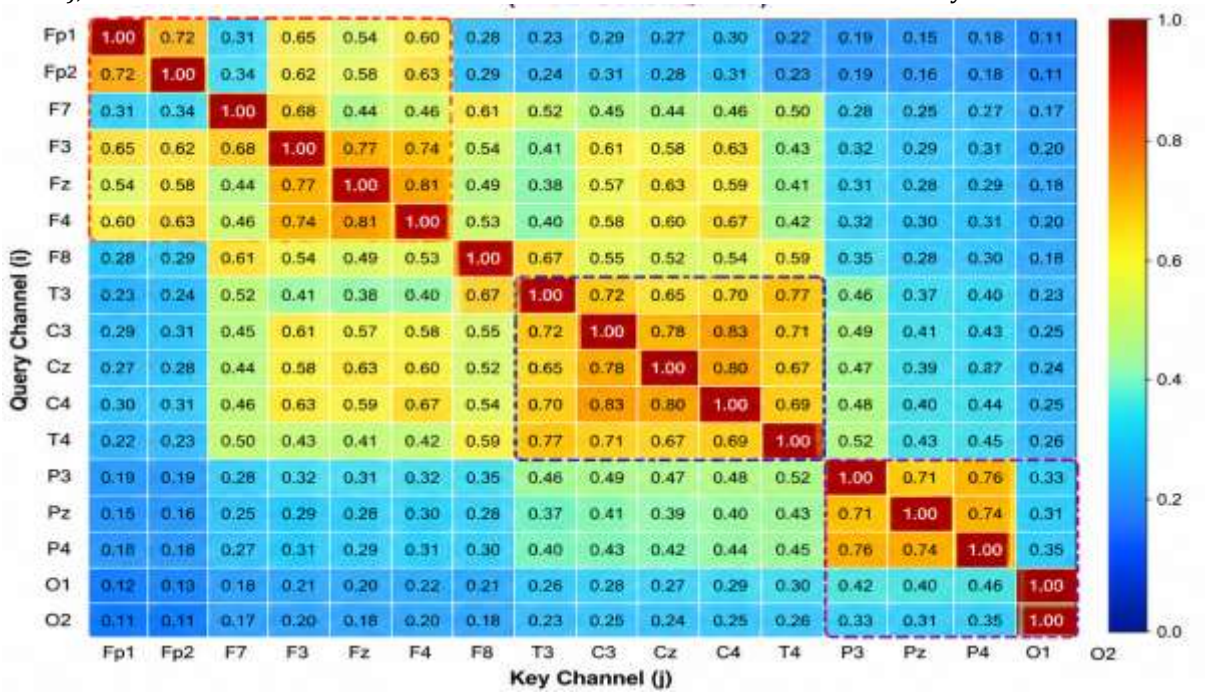


**Figure 4:** Attention Heatmap Across EEG Channels (NHDETGCNetsQBANO)

As in figure e4, values is representing a normalized attention weights (0 to 1). The higher values (warmer colors) is indicating the stronger influence between the EEG channels in the model's decision-making process. High attention in the frontal regions (Fp1–F8) is associated with the executive function and the working memory disturbances in schizophrenia. Strong attention between the temporal and central regions (T3–C4) is showing the abnormal connectivity that is related to the auditory hallucinations and the cognitive deficits. Moderate attention in parieto-occipital regions (P3–O2) is related to the reduced sensory–motor integration.

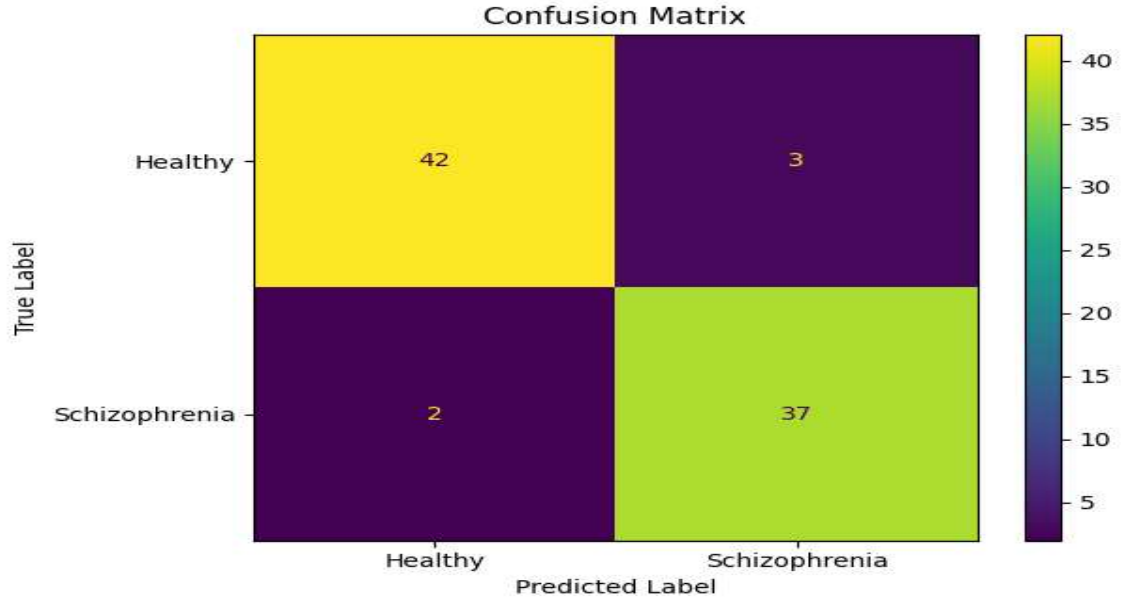

**Figure 5:** Confusion Matrix for Proposed Model

From figure 5, the model showed strong classification capability with very low false negatives, which is critical in clinical screening. The confusion matrix directly supported the high recall and specificity observed during evaluation.

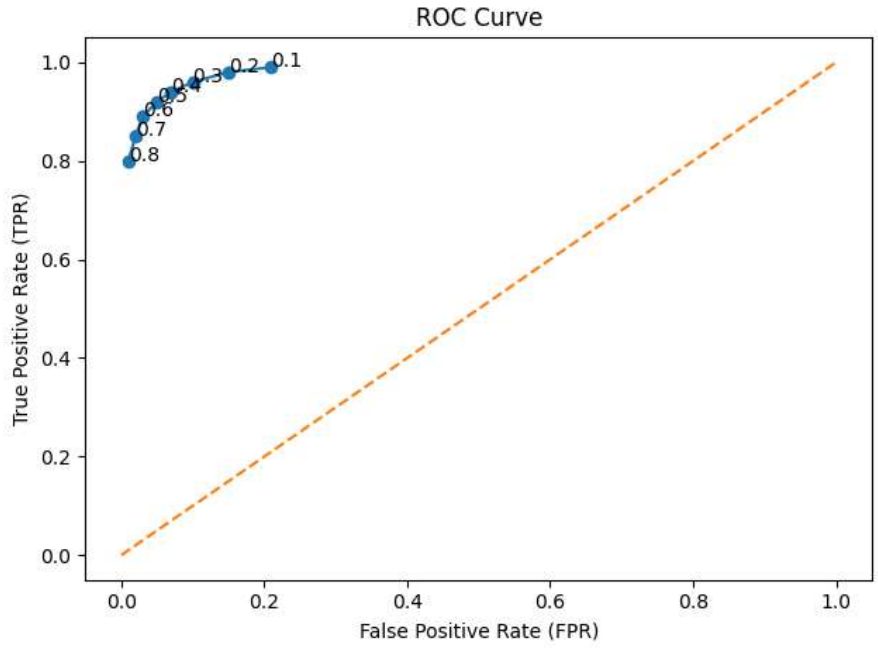


**Figure 6:** ROC Curve Values (Threshold-wise Performance)

The ROC values as in figure 6 indicates that the model is maintaining a high true positive rate even at stricter thresholds. The Area Under Curve (AUC) derived from these values was approximately 0.97, which is showing an excellent separability between classes.

## 5. Conclusion

The study confirmed that schizophrenia prediction using EEG signals required not only high-performing DL models but also architectures capable of handling neurophysiological variability. The proposed NHDETGCNetsQBANO model resolved these challenges using a hybrid design that combined the graph-based spatial learning, entropy-driven temporal decomposition, and reinforcement-guided residual optimization. The model provided an improvement of 98% accuracy on the Moscow dataset and nearly 98% on the Warsaw dataset while maintaining lower error rates with reduced computational time across different subjects. The ablation study shown that each module contributed uniquely to performance, and the largest degradations occurred when spatial graph learning and MTF-based representation were removed, reinforcing the importance of modeling EEG not as isolated signals but as coordinated brain-region interactions.

## References

[1] Long, J., Fang, Z., & Wang, L. (2024). SK-MMFMNet: A multi-dimensional fusion network of remote sensing images and EEG signals for multi-scale marine target recognition. Information Fusion, 108, 102402. https://doi.org/10.1016/j.inffus.2024.102402

[2] Rahul, J., Sharma, D., Sharma, L. D., Nanda, U., & Sarkar, A. K. (2024). A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning. Frontiers in Human Neuroscience, 18, 1347082. https://doi.org/10.3389/fnhum.2024.1347082

[3] Bhadra, S., Kumar, C. J., & Bhattacharyya, D. K. (2025). Multiview EEG signal analysis for diagnosis of schizophrenia: an optimized deep learning approach. Multimedia Tools and Applications, 84(24), 27577-27608. https://doi.org/10.1007/s11042-024-20205-y

[4] Chen, H., Lei, Y., Li, R., Xia, X., Cui, N., Chen, X., ... & Zhou, J. (2024). Resting-state EEG dynamic functional connectivity distinguishes non-psychotic major depression, psychotic major depression and schizophrenia. Molecular Psychiatry, 29(4), 1088-1098. https://doi.org/10.1038/s41380-023-02395-3

[5] Jafari, M., Sadeghi, D., Shoeibi, A., Alinejad-Rokny, H., Beheshti, A., García, D. L., ... & Gorriz, J. M. (2024). Empowering precision medicine: AI-driven schizophrenia diagnosis via EEG signals: A comprehensive review from 2002–2023. Applied Intelligence, 54(1), 35-79. https://doi.org/10.1007/s10489-023-05155-6
[6] Haider, U., Hanif, M., Rashid, A., Qaisar, S. M., & Subasi, A. (2024). EEG-based schizophrenia classification using penalized sequential dictionary learning in the context of mobile healthcare. Biomedical Signal Processing and Control, 90, 105856. https://doi.org/10.1016/j.bspc.2023.105856
[7] Aksoy, G., Cattan, G., Chakraborty, S., & Karabatak, M. (2024). Quantum machine-based decision support system for the detection of schizophrenia from EEG records. Journal of medical systems, 48(1), 29. https://doi.org/10.1007/s10916-024-02048-0
[8] Garip, Z., Ekinci, E., Serbest, K., & Eken, S. (2024). Chaotic marine predator optimization algorithm for feature selection in schizophrenia classification using EEG signals. Cluster Computing, 27(8), 11277-11297. https://doi.org/10.1007/s10586-024-04511-6
[9] Al Fahoum, A., & Zyout, A. A. (2024). Wavelet transform, reconstructed phase space, and deep learning neural networks for EEG-based schizophrenia detection. International journal of neural systems, 34(09), 2450046. https://doi.org/10.1142/S0129065724500461
[10] Shen, M., Wen, P., Song, B., & Li, Y. (2024). 3D convolutional neural network for schizophrenia detection using as EEG-based functional brain network. Biomedical Signal Processing and Control, 89, 105815. https://doi.org/10.1142/S0129065724500461
[11] Guo, Z., Wang, J., Jing, T., & Fu, L. (2024). Investigating the interpretability of schizophrenia EEG mechanism through a 3DCNN-based hidden layer features aggregation framework. Computer Methods and Programs in Biomedicine, 247, 108105. https://doi.org/10.1016/j.cmpb.2024.108105
[12] Earl, R. J., Ford, T. C., Lum, J. A., Enticott, P. G., & Hill, A. T. (2024). Exploring aperiodic activity in first episode schizophrenia spectrum psychosis: A resting-state EEG analysis. Brain Research, 1840, 149052. https://doi.org/10.1016/j.brainres.2024.149052
[13] Paraschiv, E. A., Băjenaru, L., Petrache, C., Bica, O., & Nicolau, D. N. (2024). AI-driven neuro-monitoring: advancing schizophrenia detection and management through deep learning and EEG analysis. Future Internet, 16(11), 424. https://doi.org/10.3390/fi16110424
[14] Latreche, I., Slatnia, S., Kazar, O., Harous, S., & Khelili, M. A. (2024). Identification and diagnosis of schizophrenia based on multichannel EEG and CNN deep learning model. Schizophrenia Research, 271, 28-35. https://doi.org/10.1016/j.schres.2024.07.015
[15] Aziz, S., Khan, M. U., Iqtidar, K., & Fernandez-Rojas, R. (2024). Diagnosis of schizophrenia using EEG sensor data: A novel approach with automated log energy-based empirical wavelet reconstruction and cepstral features. Sensors, 24(20), 6508. https://doi.org/10.3390/s24206508
[16] Smucny, J., Davidson, I., & Carter, C. S. (2021). Comparing machine and deep learning-based algorithms for prediction of clinical improvement in psychosis with functional magnetic resonance imaging. Human brain mapping, 42(4), 1197-1205. https://doi.org/10.1002/hbm.25286
[17] Supakar, R., Satvaya, P., & Chakrabarti, P. (2022). A deep learning based model using RNN-LSTM for the detection of schizophrenia from EEG data. Computers in Biology and Medicine, 151, 106225. https://doi.org/10.1016/j.compbiomed.2022.106225
[18] Bagherzadeh, S., Shahabi, M. S., & Shalbaf, A. (2022). Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal. Computers in Biology and Medicine, 146, 105570. https://doi.org/10.1016/j.compbiomed.2022.105570
[19] Zhang, J., Rao, V. M., Tian, Y., Yang, Y., Acosta, N., Wan, Z., ... & Guo, J. (2023). Detecting schizophrenia with 3D structural brain MRI using deep learning. Scientific Reports, 13(1), 14433. https://doi.org/10.1038/s41598-023-41359-z
[20] Zheng, J., Wei, X., Wang, J., Lin, H., Pan, H., & Shi, Y. (2021). Diagnosis of schizophrenia based on deep learning using fMRI. Computational and Mathematical Methods in Medicine, 2021(1), 8437260. https://doi.org/10.1155/2021/8437260
[21] Lamichhane, B., Zhou, J., & Sano, A. (2023). Psychotic relapse prediction in schizophrenia patients using a personalized mobile sensing-based supervised deep learning model. IEEE Journal of Biomedical and Health Informatics, 27(7), 3246-3257. 10.1109/JBHI.2023.3265684
[22] Skorobogatov, K., De Picker, L., Wu, C. L., Foiselle, M., Richard, J. R., Boukouaci, W., ... & Leboyer, M. (2024). Immune-based machine learning prediction of diagnosis and illness state in Schizophrenia and bipolar disorder. Brain, Behavior, and Immunity, 122, 422-432. https://doi.org/10.1016/j.bbi.2024.08.013

[23] Srinivasan, S., & Johnson, S. D. (2024). A novel approach to schizophrenia Detection: Optimized preprocessing and deep learning analysis of multichannel EEG data. Expert Systems with Applications, 246, 122937. https://doi.org/10.1016/j.eswa.2023.122937
[24] Saadatinia, M., & Salimi-Badr, A. (2024). An explainable deep learning-based method for schizophrenia diagnosis using generative data-augmentation. IEEE Access. 10.1109/ACCESS.2024.3428847
[25] Hassan, F., Hussain, S. F., & Qaisar, S. M. (2023). Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques. Information Fusion, 92, 466-478. https://doi.org/10.1016/j.inffus.2022.12.019
[26] Grover, N., Chharia, A., Upadhyay, R., & Longo, L. (2023). Schizo-Net: A novel Schizophrenia Diagnosis framework using late fusion multimodal deep learning on Electroencephalogram-based Brain connectivity indices. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 464-473. 10.1109/TNSRE.2023.3237375
[27] Weng, T., Zheng, Y., Xie, Y., Qin, W., & Guo, L. (2024). Diagnosing schizophrenia using deep learning: Novel interpretation approaches and multi-site validation. Brain Research, 1833, 148876. https://doi.org/10.1016/j.brainres.2024.148876
[28] Sherine Glory, J., Durgadevi, P., & Ezhumalai, P. (2025). Enhancing drug discovery in schizophrenia: a deep learning approach for accurate drug-target interaction prediction–DrugSchizoNet. Computer Methods in Biomechanics and Biomedical Engineering, 28(2), 170-187. https://doi.org/10.1080/10255842.2023.2282951
[29] Zhang, Y., Wang, Z., Ding, Z., Tian, Y., Dai, J., Shen, X., ... & Cao, Y. (2025). Employing Machine Learning and Deep Learning Models for Mental Illness Detection. Computation, 13(8), 186. https://doi.org/10.3390/computation13080186
[30] Bao, Y., Wang, W., Liu, Z., Wang, W., Zhao, X., Yu, S., & Lin, G. N. (2025). Leveraging deep neural network and language models for predicting long-term hospitalization risk in schizophrenia. Schizophrenia, 11(1), 35. https://doi.org/10.1038/s41537-025-00585-2