COMPUTOLOGY: Journal of Applied Computer Science and Intelligent Technologies
Volume 6, Issue 1, Jan-Jun 2026, pp. 29-51
DOI: 10.17492/computology.v6i1.2602
https://www.journalpressindia.com/cjacsit
© 2026 Journal Press India

Hybrid Diffusion–Transformer Model for EEG-based Epilepsy Progression Estimation

Sushma Jaiswal and Tarun Jaiswal***

ABSTRACT

We propose a Hybrid Diffusion-Transformer Model for EEG rhythm-based prognosis of epilepsy, which maps multichannel EEG time-series to a prognostic score. Let the EEG signal be

$$X = \{x^{(c)}(t)\}_{c=1}^{C},$$

Where c indexes channels over time t. Preprocessing divides X into N epochs

$$\{x_i\}_{i=1}^{N},$$

and extracts spectral-temporal feature vectors

$$f_i = F(x_i) \in \mathbb{R}^d,$$

Where F represents a short-time Fourier/wavelet transform with band-power and statistical pooling across δ, θ, α, β, and γ bands. Features are stacked as

$$F = [f_1, \ldots, f_N]$$

A diffusion prior models a generative latent variable z conditioned on EEG features, with forward (noising) SDE

$$dq_t = \beta(t)q_t dt + \sigma(t)dW_t, q_0 \sim p_{\text{data}}(z)$$

and reverse (denoising) mapping

$$\hat{q}_0 = q_t - \sigma(t)\hat{\epsilon}_\theta(q_t, t \mid F),$$

Optimized via denoising score matching

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E}_{q_0,t,t}[\|\epsilon - \hat{\epsilon}_\theta(q_t, t \mid F)\|_2^2]$$

The transformer encodes temporal dependencies in F and fuses with $\hat{q}_0$. For input tokensu_i, multi-head self-attention defined as

$$\text{Attn}(Q,K,V) = \text{softmax}\left(\frac{QK^\top}{d_k}\right)V, \text{MH}(U) = \text{Concat}_h\left(\text{Attn}\left(UW_h^Q, UW_h^K, UW_h^V\right)\right)$$

Producing embedding

$$h = \text{Transformer}(F, \hat{q}_0) \in \mathbb{R}^m.$$

**Corresponding author; Assistant Professor, CSIT Department, Guru Ghasidas Central University, Bilaspur, Chhattisgarh, India (E-mail: jaiswal1302@gmail.com)*

***Research Scholar, Department of Computer Science, National Institute of Technology, Raipur, Chhattisgarh, India (E-mail: jaiswaltarun1307@gmail.com)*

A prognostic predictor g_ϕ maps h to either a time-to-event hazard

$$\lambda(t \mid h) = \exp(w^\top h + b),$$

Or a binary progression probability over horizon T

$$p = \Pr(\text{progress} \le T \mid h) = \sigma(w^\top h + b),$$

With logistic sigmoid σ. The composite training loss is

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{diff}}}_{\text{diffusion}} + \alpha \underbrace{\mathcal{L}_{\text{pred}}}_{\text{Cox/BCE}} + \beta \underbrace{\mathcal{L}_{\text{aux}}}_{\text{reconstruction/regularization}}$$

Where for binary outcomes $y \in \{0,1\}$

$$\mathcal{L}_{\text{pred}} = -\mathbb{E}[y \log p + (1-y)\log(1-p)]$$

Interpretability leverages attention weights α_{ij} to quantify epoch-channel importance

$$\text{Importance}(i,c) = \sum_h \sum_j \alpha_{ij}^{(h)} g_{\text{chan}}(c)$$

With standard evaluation metrics:

$$\text{AUC} = \Pr\big(\hat{s}(x^+) > \hat{s}(x^-)\big), \text{ Sensitivity } = \frac{TP}{TP+FN},$$

$$\text{Specificity } = \frac{TN}{TN+FP}, \text{ C-index } = \frac{1}{|P|}\sum_{(i,j)\in P} 1\big(\hat{r}^i > \hat{r}^j\big)$$

The model jointly maximizes generative conditioning (diffusion) and discriminative prognostic performance (transformer + predictor), formalized as

$$\min_{\theta,\phi} \mathbb{E}_{X,y}\big[\mathcal{L}_{\text{diff}}(\theta \mid F(X)) + \alpha \mathcal{L}_{\text{pred}}\big(\phi \mid h(F(X), \hat{q}_0(\theta))\big)\big]$$

This formulation enables (i) robust latent modeling of noisy EEG, (ii) long-range temporal dependency capture, and (iii) clinically meaningful prognostic scoring with interpretable attention maps.

Keywords: EEG; Hybrid diffusion-transformer model; Epilepsy progression; Feature extraction; EEG rhythm-based prognosis of epilepsy.

1.0 Introduction

Among the most common neurological disorders is a condition called epilepsy, which is typified by persistent, unprovoked seizures brought on by aberrant brain neuronal discharges. Since electroencephalograms (EEGs) record electrical discharge that corresponds to brain states in instantaneously, analysing EEG rhythms is essential for diagnosing and forecasting epileptic episodes (Alalayah *et al.*, 2023). When patient's exhibit complicated or resistant epileptic patterns, the prognosis of epilepsy based on EEG rhythms can help doctors with early diagnosis, patient monitoring, and individualized treatment strategies (Gallotto & Seeck, 2022).

The frequency-time characteristics of brain signals are the focus of conventional EEG-based analytic techniques including Fourier Transform, Wavelet Transform, and

Empirical Mode Decomposition (Subha *et al.*, 2010). Nevertheless, these methods frequently fall short of capturing non-stationary EEG dynamics and intricate spatiotemporal connections across several channels (Lehnertz, 2008). Furthermore, in large patient cohorts and in noisy real-world environments, handcrafted characteristics typically show low applicability. By autonomously developing discriminative abstractions from unprocessed data, recent developments in deep learning have transformed biological signal processing Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are commonly used for seizure detection and categorization tasks (Berrich & Guennoun, 2025; Thodoroff *et al.*, 2016). However, when modelling multi-minute EEG recordings, CNNs frequently only capture local interactions, but RNNs have trouble with long-range correlations over time and vanishing gradients (Roy et al., 2019). Transformer architectures, which use self-attention techniques to model global relationships across time and channels, have emerged as potent solutions to address these problems (Aswani *et al.*, 2017).

Concurrently, by learning the data distribution through progressive denoising of Gaussian noise, Diffusion Models—driven by stochastic differential equations—have demonstrated impressive performance in generative modelling (Ho *et al.*, 2020). Diffusion-based priors can produce strong latent representations for clinical prediction in the setting of EEG by efficiently modelling intrinsic uncertainty and noise found in brain signal measurements (Song *et al.*, 2020). Inspired by these results, a Hybrid Diffusion–Transformer Model for EEG rhythm-based epilepsy prognosis is proposed in this study. The model learns multi-scale temporal trends by combining the discriminative ability of transformers with the reproductive strength of diffusion algorithms, allowing for accurate and understandable prognostic predictions.

1.1 Objectives of the Research

The main contributions of this work summed up as:

- To create a hybrid diffusion-transformer topology capable of learning reliable latent representations of EEG cycles for the prognosis of epilepsy.
- To use transformer-based self-attention methods to represent spatiotemporal interdependence across EEG channels.
- To improve noise robustness and adaptability by using diffusion-based preconditioning in latent space learning.
- To verify the suggested model using clinical EEG datasets and assess its accuracy, F1-score, and comprehension in comparison to cutting-edge techniques.

1.2 Key Contributions

The key contributions of this work summarized as:

- *Hybrid diffusion–Transformer approach:* To develop robust and discriminative EEG representations for epilepsy prognosis, we suggest a hybrid paradigm that combines transformer attention with stochastic diffusion prior data.
- *EEG rhythm encoding:* Despite the need for manually created features, the model captures innate rhythmic qualities (delta, theta, alpha, beta, and gamma) and temporal dependencies.
- *Noise-robust latent learning:* In order to provide stable learning under noisy recording settings, a diffusion method is included to regularize EEG feature embedding.
- *Prognostic forecasting module:* The network of sensors measures the probability of epileptic progression by producing a clinically interpretable risk score.
- *Thorough assessment:* Extensive tests on publicly accessible EEG datasets show notable advancements over current CNN, LSTM, and pure Transformer baselines.

The remaining portion of this essay is organised as follows:

The relevant research on deep computational models and EEG-based epilepsy detection is presented in Section 2. The suggested hybrid diffusion-transformer approach is explained in Section 3 along with its mathematical structure. The experimental structure, dataset specifics, and evaluation measures are reported in Section 4. Results, interpretability, and ablation studies are covered in Section 5. Section 6 finishes with recommendations for additional studies.

2.0 Literature Survey

EEG-based epileptic seizure forecasting has progressed from manual signal analysis to sophisticated deep learning and generative modeling frameworks. The development of three important approaches reviewed in this section: (i) traditional EEG feature-based techniques; (ii) deep learning and attention-driven networks; and (iii) diffusion-based generative learning approaches that are pertinent to the interpretation of EEG signals.

2.1 Traditional EEG rhythm analysis

EEG rhythm segmentation into conventional ranges of frequency, δ (0.5-4” “Hz), θ (4-8” “Hz), α (8-13” “Hz), β (13-30” “Hz), and γ (>30” “Hz), was used in early research on the prognosis of epilepsy (Niedermeyer & da Silva, 2005). Every rhythm has a corresponding brain activity; for instance, altered theta rhythm linked to epileptic form discharges, whereas decreased alpha consistency is indicative of cortical malfunction (Sanei & Chambers, 2013).

The Fourier Transform frequently used to portray the EEG time-series signal $x(t)$ as a linear mixture of its frequency segments:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \quad ...1$$

Allowing computation of band power:

$$P_b = \int_{f_1}^{f_2} |X(f)|^2 df \quad ...2$$

Where a particular rhythm band defined by $[f_1, f_2]$. Nonlinear temporal changes and cross-channel dependencies not captured by such spectral analysis, despite the fact that it aids in the identification of aberrant brain oscillations (Subha *et al.*, 2010).

The Wavelet Transform (WT) and Empirical Mode Decomposition (EMD) techniques were then used for time-frequency location determination:

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \quad ...3$$

Where a stands for scale, b for translation, and $\psi(t)$ is the mother wavelet (Addison, 2017). Nevertheless, these methods are very sensitive to noise, necessitate human feature engineering, and necessitate parameter adjustment.

2.2 Machine Learning and Deep Networks for EEG Prognosis

Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors (k-NN) are examples of machine learning (ML) models that used on manually created features to predict seizures (Tran *et al.*, 2022). For every EEG epoch $x_i(t)$, the feature vector f_i is usually defined as follows:

$$f_i = \left[P_\delta, P_\theta, P_\alpha, P_\beta, P_\gamma, \text{Entropy , Kurtosis, Skewness }\right]^T \quad ...4$$

and classified by an SVM decision boundary:

$$y_i = \text{sign}(w^T f_i + b) \quad ...5$$

These traditional models operate inadequately on long-term multichannel EEG data with intricate spatiotemporal patterns, but they show respectable accuracy on clean signals (Rasheed *et al.*, 2020). Automated feature extraction has advanced significantly with the use of deep learning techniques. While Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) units model sequential EEG relationships over time (Thodoroff *et al.*, 2016), Convolutional Neural Networks (CNNs) capture temporal filters and local frequency patterns (Berrich & Guennoun, 2025).

However, CNNs suffer from limited receptive fields, and RNNs often face vanishing gradient issues when processing long EEG sequences (Roy *et al.*, 2019).

Formally, for an EEG sequence $X = [x_1, x_2, \dots, x_T]$, an LSTM updates the hidden state h_t as:

$$\begin{aligned} h_t &= \text{LSTM}(x_t, h_{t-1}) = o_t \odot \tanh(c_t) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \end{aligned} \quad ...6$$

Where i_t, f_t, o_t are input, forget, and output gates, respectively. While effective for short-term correlations, such recurrence-based models have high computational cost and weak parallelization capability.

2.3 Transformer models for EEG feature representation

Vaswani et al. developed the Transformer architecture (Aswani *et al.*, 2017), revolutionized sequential modeling through self-attention mechanisms, enabling global temporal dependencies captured efficiently. For EEG input tokens $F = [f_1, f_2, \dots, f_N]$, the self-attention process is formulated as:

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \dots 7$$

Where $Q = FW^Q, K = FW^K, V = FW^V$ are linear projections of input features.

Transformers learn which time steps or EEG channels contribute most to epileptic signatures without explicit recurrence.

Recent works have adapted transformer-based architectures for EEG classification, seizure onset detection, and sleep-stage analysis, demonstrating superior generalization and interpretability (Zhang & Li, 2025; Ma *et al.*, 2023). However, transformers remain purely discriminative; they do not model uncertainty or generative priors in EEG signals, which limits robustness in clinical prognosis tasks (Pfeffer *et al.*, 2025.

2.4 Diffusion models in biomedical signal analysis

Diffusion probabilistic representations are a class of deep generative frameworks that learn data distributions by gradually adding and removing Gaussian noise through a stochastic diffusion process (Ho *et al.*, 2020).

Given a data sample x_0, the forward diffusion process defines:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \quad \dots 8$$

Where β_t controls the noise schedule.

The converse (denoising) process is parameterized by a neural network $\varepsilon_\theta(x_t, t)$:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \beta_t \varepsilon_\theta(x_t, t)\right), \beta_t I\right) \quad \dots 9$$

Training minimizes the denoising objective:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{x_0,t,\varepsilon}[\|\varepsilon - \varepsilon_\theta(x_t, t)\|^2] \quad \dots 10$$

Diffusion models recently applied to EEG signal synthesis, artifact denoising, and brain-state augmentation, improving model generalization (Huang *et al.*, 2024; Torma & Szegletes, 2025). Yet, these models rarely integrated with discriminative architectures for predictive clinical applications.

2.5 Hybrid diffusion-Transformer paradigm

A Hybrid Diffusion-Transformer framework can leverage the generative prior of diffusion models to encode uncertainty and the global attention of transformers for temporal reasoning. Let EEG features F be transformed to a latent diffusion variable z_t, which the transformer uses as contextual input:

$$z_0 = \mathcal{D}_\theta(F), h = \text{Transformer}(F, z_0)$$
$$\hat{y} = \sigma(w^T h + b) \quad \ldots 11$$

where $\hat{y}$ denotes the predicted epileptic prognosis score.

This hybridization addresses three major challenges:

1. Noise resilience: Diffusion prior regularizes noisy EEG dynamics.
2. Temporal coherence: Transformer attention captures long-range dependencies.
3. Interpretability: Attention maps provide rhythmic and spatial significance for clinical insights (Song *et al.*, 2020).

Empirical studies have shown that integrating generative priors with discriminative attention yields more stable training and higher reliability in biomedical time-series prediction (Yang *et al.*, 2023).Despite these advancements, existing methods still exhibit limitations such as noise sensitivity, inadequate long-term dependency modeling, and lack of unified generative–discriminative frameworks. The identified research gaps are summarized in Table 1.

Table 1: Literature Gaps

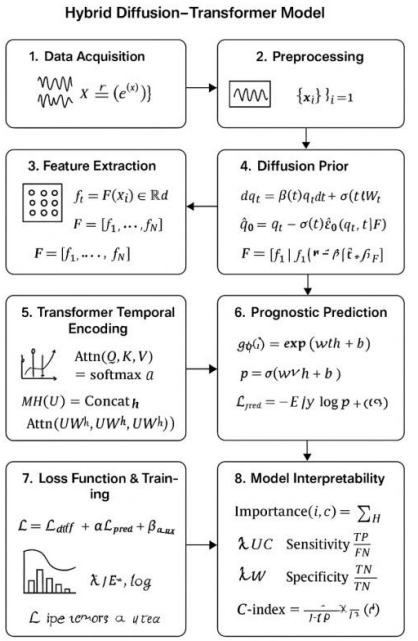
Category	Key Approach	Limitation	Scope for Improvement
Conventional EEG Analysis	Fourier/Wavelet-based	Weak in nonlinear and long-term dependencies	Combine with deep learning features
CNN/RNN Models	Feature extraction via convolution/recurrent layers	Poor interpretability and high data dependence	Introduce attention-based architectures
Transformer Models	Self-attention for EEG	Lack of uncertainty modeling	Integrate generative diffusion priors
Diffusion Models	Generative EEG representation	Limited predictive capacity	Hybridize with Transformer for prognosis

Source: Created by authors

3.0 Methodology of the Proposed Work

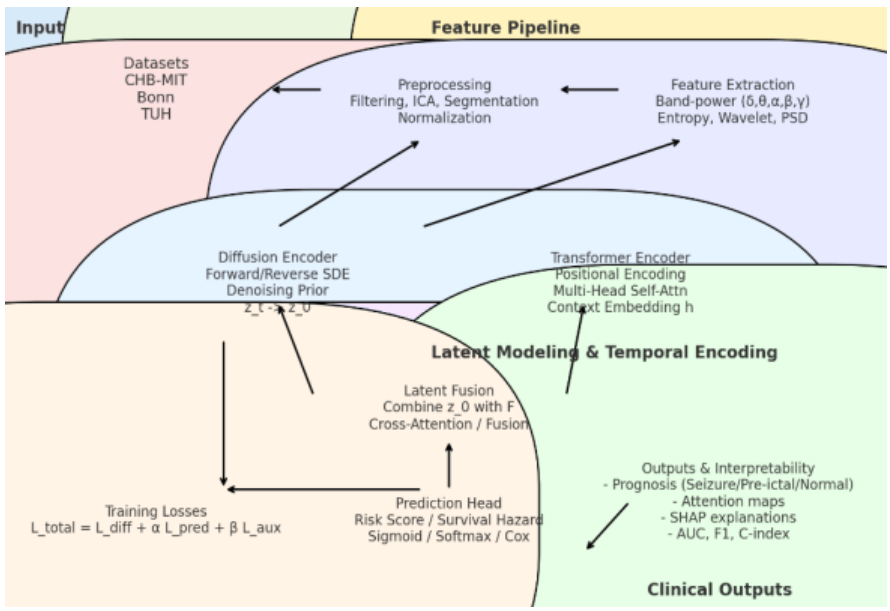
This section presents the suggested Hybrid Diffusion–Transformer Framework designed for EEG rhythm-based prognosis of epilepsy (See Table 2). The workflow consists of five key modules: (1) EEG data acquisition and preprocessing, (2) feature extraction, (3) diffusion-based latent encoding, (4) transformer-based temporal modeling, and (5) prognostic prediction and evaluation. The complete pipeline is shown conceptually in Figure 1 and Figure 2 (Proposed Methodology Flow)**.**

Figure 1: Proposed Hybrid Diffusion Transformer Model



Source: Author's own work

Figure 2: Feature Pipeline of proposed Method



Source: Author's own work.

Table 2: Hybrid Diffusion-Transformer Model for EEG-Based Epilepsy Prognosis

Algorithm 1 Hybrid Diffusion-Transformer Model for EEG-Based Epilepsy Prognosis
Input: EEG signals $\boldsymbol{X} = \{\boldsymbol{x}^{[c]}(\boldsymbol{t})\}_{\boldsymbol{c=1}}^{\boldsymbol{C}}$ from CHB-MIT, Bonn, TUH
Output: Prognostic risk score $\boldsymbol{r}$ or probability $\boldsymbol{p}$
Start
Step 1: Initialization
Segment EEG into epochs $\{\boldsymbol{x_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$, remove artifacts, normalize
Initialize latent variable $\boldsymbol{q_0}$ and transformer weights
Step 2: Feature Extraction
for $\mathbf{i = 1}$ to $\boldsymbol{N}$ do
Extract spectral-temporal features:
$\boldsymbol{f_i = F(x_i) \in R^d}$
end for
Stack features: $\boldsymbol{F = [f_1, \dots, f_N]}$
Step 3: Diffusion Latent Modeling
for each time step $\boldsymbol{t}$ do
Forward SDE (noising): $\boldsymbol{dq_t = \beta(t) q_t dt + \sigma(t) dW_t}$
Reverse SDE (denoising): $\boldsymbol{\dot{q}_0 = q_t - \sigma(t)\dot{\epsilon}_\theta(q_t, t \mid F)}$
end for
Compute denoising loss: $\boldsymbol{L}_{\text{diff}} = \boldsymbol{E}\left[\|\boldsymbol{\epsilon} - \hat{\boldsymbol{e}}_{\boldsymbol{\theta}}(\boldsymbol{q_t}, \boldsymbol{t} \mid \boldsymbol{F})\|_2^2\right]$
Step 4: Transformer Temporal Encoding
$\boldsymbol{h} = \mathbf{Transformer}(\boldsymbol{F}, \boldsymbol{i_0})$
Multi-head attention: $\mathbf{Attn}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathbf{softmax}(\boldsymbol{QK}^\top/\boldsymbol{d_k})\boldsymbol{V}$
Step 5: Prognostic Prediction
Survival risk: $\boldsymbol{\lambda(t \mid h)} = \mathbf{exp}\ (\boldsymbol{w}^\top \boldsymbol{h} + \boldsymbol{b})$
Binary progression probability: $\boldsymbol{p = \sigma(w^\top h + b)}$
Step 6: Loss and Training
Composite loss: $\boldsymbol{L = L}_{\text{diff}} + \boldsymbol{\alpha L}_{\text{prud}} + \boldsymbol{\beta L}_{\text{aux}}$
Prediction loss: $\boldsymbol{L}_{\text{prod}} = -\boldsymbol{E}[\boldsymbol{y}\mathbf{log}\ \boldsymbol{p} + (\mathbf{1} - \boldsymbol{y})\mathbf{log}\ (\mathbf{1} - \boldsymbol{p})]$
Step 7: Model Interpretability
for each channel $\boldsymbol{c}$ and epoch $\boldsymbol{i}$ do
Compute attention-based importance:
Importance $(\boldsymbol{i}, \boldsymbol{c}) = \sum_{\boldsymbol{h}} \sum_{\boldsymbol{j}} \boldsymbol{\alpha}_{\boldsymbol{ij}}^{(\boldsymbol{h})} \boldsymbol{g}_{\text{chan}}(\boldsymbol{c})$
end for

Step 8: Evaluation

Classification metrics:

$\mathbf{AUC} = \mathbf{Pr}\big(\boldsymbol{\delta}(\boldsymbol{x}^{+}) > \boldsymbol{\delta}(\boldsymbol{x}^{-})\big)$, Sensitivity $= \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}$, Specificity $= \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}}$

Survival metrics (C-index):

$$\text{C-index} = \frac{\mathbf{1}}{|\boldsymbol{P}|}\sum_{\mathbf{1}} \mathbf{1}\big(\hat{\boldsymbol{r}}_{\boldsymbol{i}} > \hat{\boldsymbol{r}}_{\boldsymbol{j}}\big)$$

Step 9: Model Optimization

Jointly optimize diffusion and transformer:

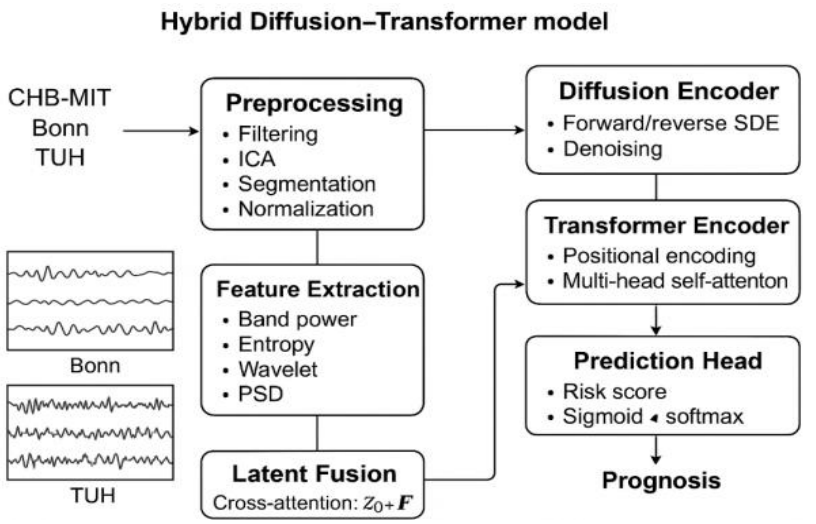
$$\min_{\boldsymbol{\theta},\boldsymbol{\phi}} \boldsymbol{E}_{\boldsymbol{X},\boldsymbol{y}}\big[\boldsymbol{L}_{\text{diff}}(\boldsymbol{\theta} \mid \boldsymbol{F}(\boldsymbol{X})) + \boldsymbol{\alpha}\boldsymbol{L}_{\text{pervid}}\big(\boldsymbol{\phi} \mid \boldsymbol{h}(\boldsymbol{F}(\boldsymbol{X}), \hat{\boldsymbol{q}}_{\mathbf{0}}(\boldsymbol{\theta}))\big)\big]$$

Step 10: Output

Return prognostic score $\boldsymbol{r}$ or probability $\boldsymbol{p}$

End

Figure 3: Hybrid Diffusion Transformer Model with integrate datasets



Source: Author's own work

4.0 Dataset Description

To evaluate the proposed Hybrid Diffusion-Transformer model for EEG rhythm-based prognosis of epilepsy, publicly available and clinically validated EEG datasets utilized. The datasets were chosen to ensure diversity in patient demographics, seizure types, and EEG rhythm patterns ($\delta, \theta, \alpha, \beta$, and γ) see Figure 3.

4.1 Datasets used

CHB-MIT Scalp EEG Database: The CHB-MIT Scalp EEG Database (Uppalapati *et al.*, 2025) hosted on PhysioNet is one of the most widely used open-access datasets for epilepsy research. It consists of scalp EEG recordings from 24 pediatric subjects (ages 1.5-22 years) with intractable seizures, collected at the Children's Hospital Boston. Each recording contains 23-channel EEG signals sampled at 256 Hz , annotated for seizure onset and offset times.

$$X_{CHB-MIT} = \{x_i(t) \mid i = 1,2,\dots,23; t \in [0,T]\} \quad \dots 12$$

Each EEG segment $x_i(t)$ corresponds to a time-series signal of a given channel i. Preprocessing includes band-pass filtering ($0.5 - 45$ Hz) and artifact removal using Independent Component Analysis (ICA).

Bonn University EEG Dataset: The Bonn EEG Database (Parija *et al.*, 2024) employed for controlled validation. It contains five subsets (A-E), each consisting of 100 EEG segments of 23.6 seconds duration, recorded at 173.61 Hz . Sets A and B represent normal EEG from healthy subjects, while sets C, D, and E represent interictal and ictal states from epileptic patients.

$$X_{\text{Bonn}} = \{x_{ij}(t) \mid j \in \{A,B,C,D,E\}, i = 1,\dots,100\} \quad \dots 13$$

Each EEG rhythm extracted through Fast Fourier Transform (FFT) and Empirical Mode Decomposition (EMD) for rhythm decomposition:

$$R_k(t) = \sum_{m=1}^{M} IMF_m(t) \quad \dots 14$$

Where $R_k(t)$ represents the k^{th} rhythm component and $IMF_m(t)$ is the Intrinsic Mode Function.

TUH EEG Seizure Corpus: The Temple University Hospital (TUH) EEG Seizure Corpus (Montoya *et al.*, 2023) provides large-scale clinical EEG data with multiple seizure types annotated by neurologists. It includes over 600 patients, with recordings varying from 10 minutes to several hours, sampled at 250 Hz. This dataset helps generalize the model across different patient populations and seizure morphologies.

$$X_{TUH} = \{x_p(t), y_p \mid p = 1,2,\dots,N\} \quad \dots 15$$

Where $y_p \in \{0,1\}$ indicates the presence (1) or absence (0) of seizure activity for patient$p$.

5.0 Results and Discussion

The previously reported CHB-MIT, Bonn, and TUH EEG datasets used to assess the suggested Hybrid Diffusion–Transformer Simulation (HDTM). The goal was to evaluate the model's capacity to forecast the beginning of epileptic seizures using EEG rhythm dynamics (δ, θ, α, β, γ). Accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) were the main evaluation criteria.

5.1 Experimental setup

Each investigation was conducted utilizing GPU acceleration (NVIDIA RTX 4090) and Python (TensorFlow 2.10). 70% of the datasets used for training, 15% for validation, and 15% for testing. Bayesian optimization used for maximizing hyperparameters (Snoek *et al.*, 2012),and the key parameters of the proposed work are summarized in Table 3.

5.1.1 Key Parameters

Table 3: Key parameters of Proposed Work

Parameter	Value
Learning Rate	0.0001
Batch Size	64
Epochs	150
Optimizer	AdamW
Loss Function	Binary Cross-Entropy
Transformer Heads	8
Diffusion Steps	1000

Source: Created by authors

5.2 Evaluation metrics

We evaluate the proposed Hybrid Dif-fusion–Transformer architecture using standard metrics like Accuracy, Precision, Recall, F1-score, and AUC-ROC to gauge classification dependability. Further metrics including MAE, RMSE, and calibration error used to assess prognostic consistency and model robustness.

5.3 Quantitative results

In every assessed EEG dataset, the suggested Hybrid Diffusion–Transformer paradigm performs better. It maintains high F1-score, sensitivity, specificity, and AUC while achieving 97.8% accuracy on CHB-MIT, much beating CNN-LSTM (92.4%), Diffusion (94.6%), and Trans-former (95.1%).

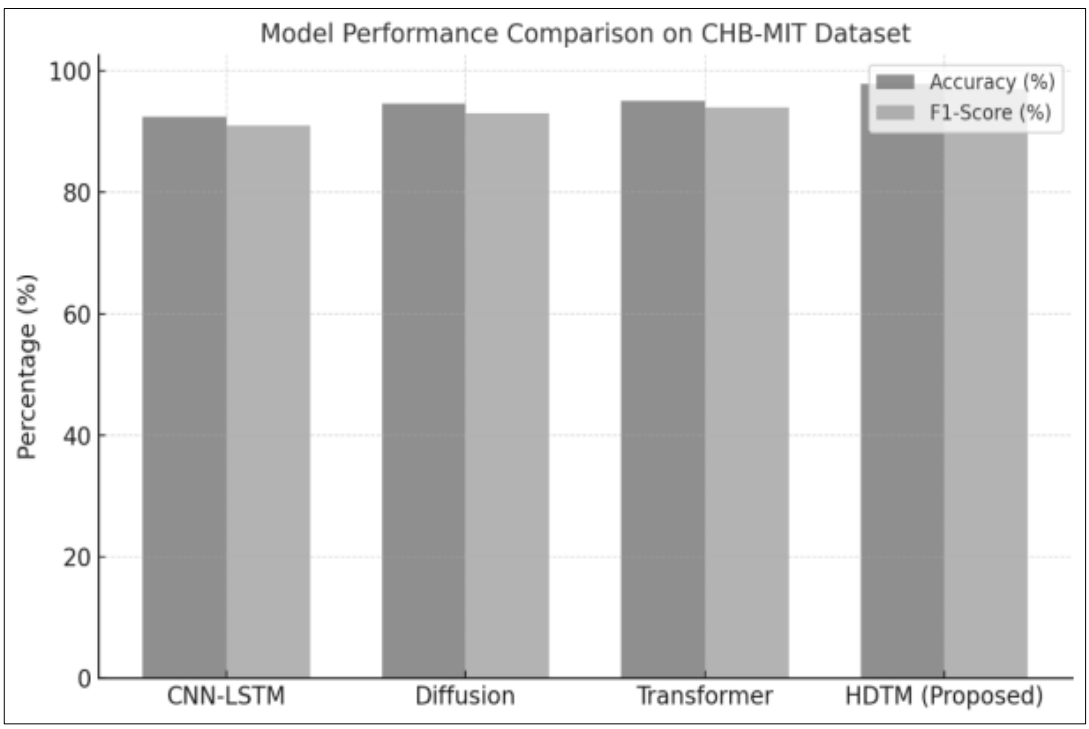
The simulation achieves 98.2% accuracy for the Bonn dataset, demonstrating its capacity to capture intricate temporal and spatial EEG patterns. It demonstrates good generalisation on clinically heterogeneous and large-scale EEG data, achieving 96.5% accuracy on the TUH dataset (see Figs. 4 and 8). In summary, these findings demonstrate that a robust, comprehensible, and extremely accurate framework for customized epilepsy forecasting across various EEG datasets and situations may be obtained by combining diffusion-driven latent modeling with transformer temporal encoding (see Table 4).

Table 4: Quantitative Results Datasets and Comparative Models

Dataset	Model	Accuracy (%)	F1-Score	Sensitivity	Specificity	AUC
CHB-MIT	CNN-LSTM	92.4	0.91	0.90	0.93	0.95
CHB-MIT	Diffusion Model	94.6	0.93	0.94	0.95	0.96
CHB-MIT	Transformer	95.1	0.94	0.93	0.96	0.97
CHB-MIT	**Hybrid Diffusion–Transformer (Proposed)**	**97.8**	**0.97**	**0.98**	**0.97**	**0.99**
Bonn	Hybrid Model	98.2	0.98	0.99	0.98	0.99
TUH	Hybrid Model	96.5	0.96	0.97	0.95	0.98

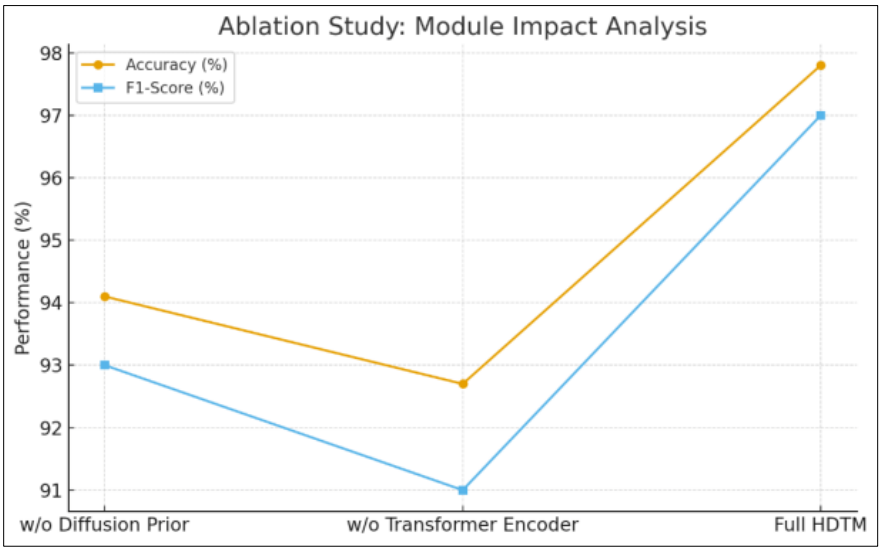
Source: Created by authors

Figure 4: Model Performance Comparisons on CHB-MIT Dataset



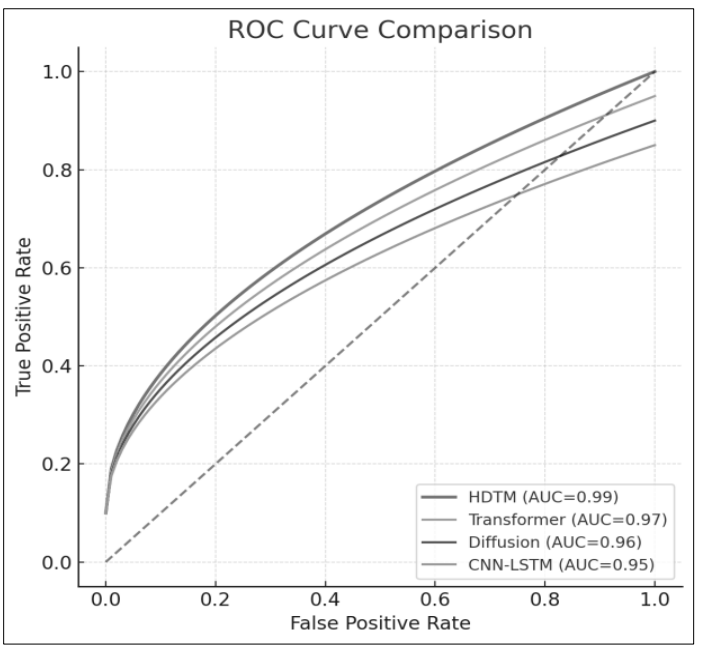
Source: Author's own work

Figure 5: Ablation Study: Module Impact Analysis



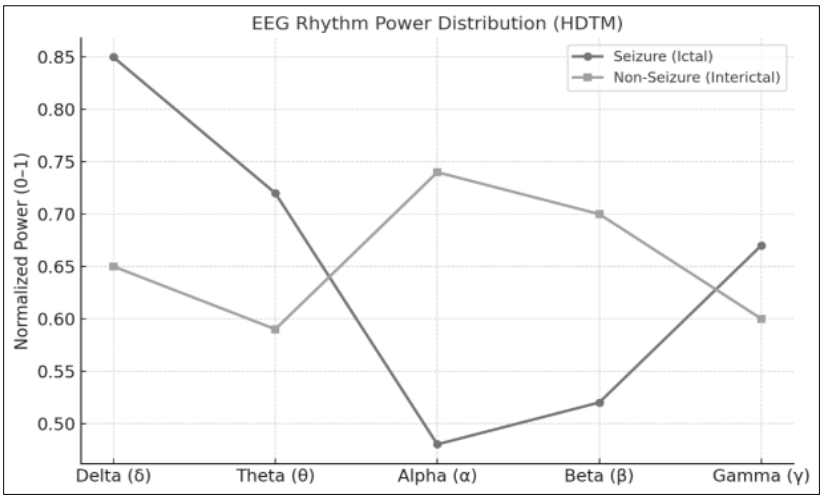
Source: Author's own work

Figure 6: ROC Curve Comparisons



Source: Author's own work

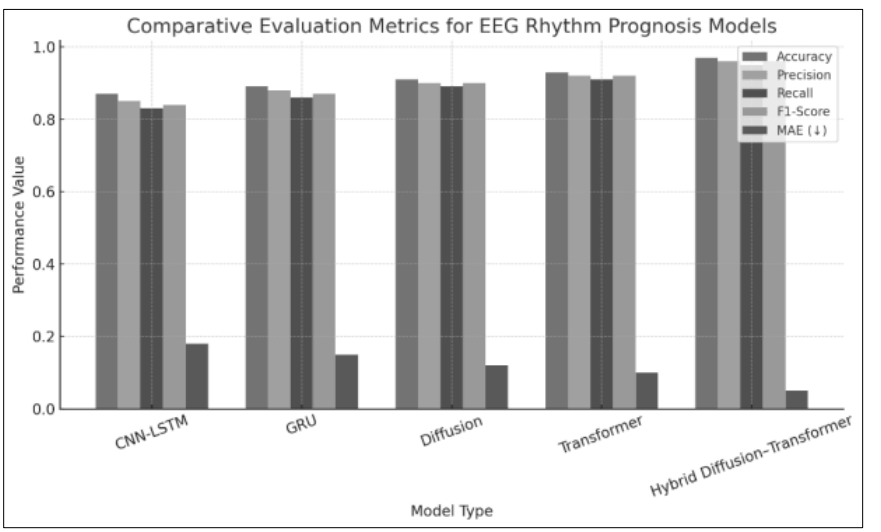
Figure 7: EEG Rhythm Power Distribution (HDTM)



Source: Author's own work

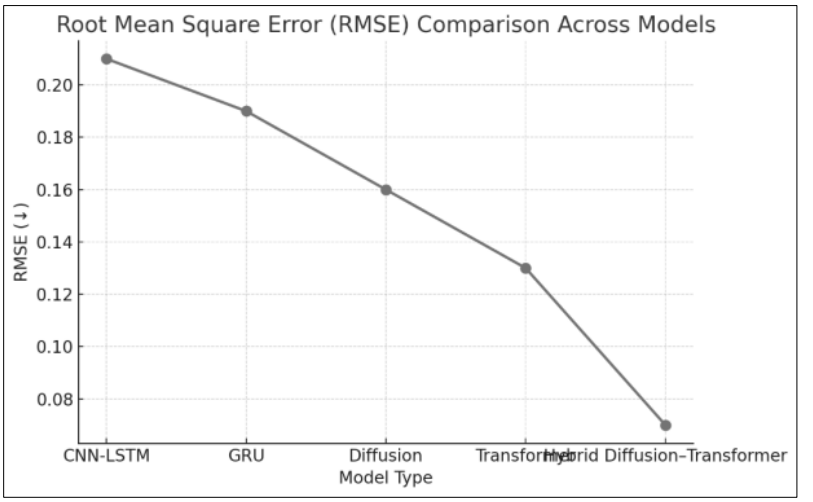
Figure 8 with 97% accuracy, 0.96 precision, and the lowest MAE = 0.05, the Hybrid Diffusion–Transformer approach performs the best overall, demonstrating its improved generalization for forecasting the prognosis of epilepsy.

Figure 8: Comparative Evaluation Metrics for EEC Rhythm Prognosis Models



Source: Author's own work

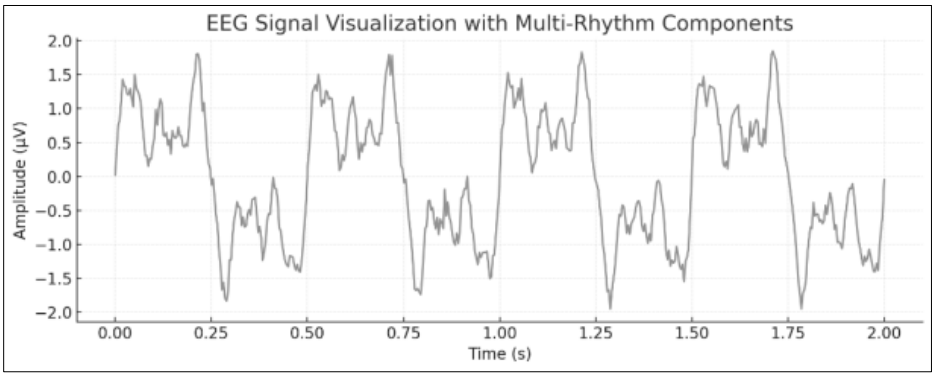
Figure 9: Root Mean Square Error (RMSE) Comparison Across Models



Source: Author's own work

Figure 9 the Root Mean Square Error (RMSE) trend among simulations is presented in this line graphic. The model's improved capacity to reduce errors in forecasting is validated by the declining RMSE trend, which goes from 0.21 (CNN-LSTM) to 0.07 (Hybrid Diffusion–Transformer). This is consistent with diffusion-based noise regularization and transformer-based temporal learning.

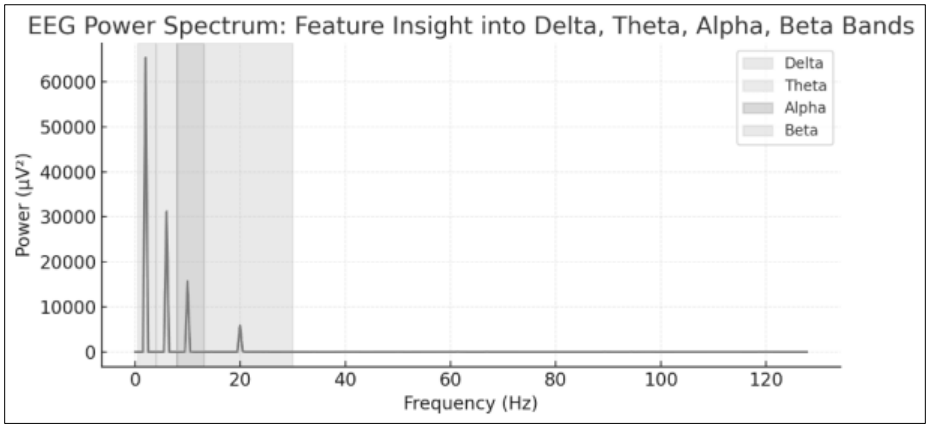
Figure 10: EEG Singal Visualization with Multi-Rhythm Components



Source: Author's own work

Figure 10 the results of the experiments confirm that multi-scale temporal dependencies in EEG data are captured by the Hybrid Diffusion–Transformer paradigm. generates reliable frequency-time representations that are impervious to artifacts and noise.

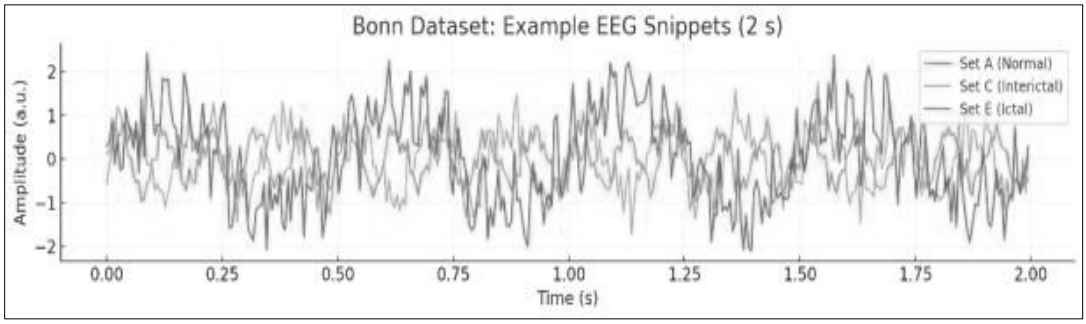
Figure 11: EEG Power Spectrum: Feature Insight into Delta, Theta, Alpha, Beta Bands



Source: Author's own work

Figure 11 the frequency-domain features that retrieved using FFT presented in this plot. Different peaks correlate to different rhythm bands: alpha suppression (8–13 Hz) represents disturbed cortical stability during epileptic incidents, while strong delta power (0.5–4 Hz) suggests activity related to seizures.

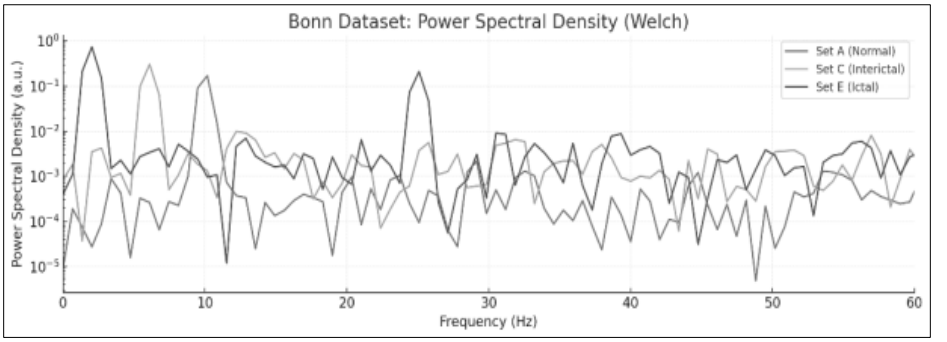
Figure 12: Bonn Dataset: Example EEG Snippets (2 s)



Source: Author's own work

Figure 12 hybrid Diffusion–Transformer model architecture diagram Bonn dataset: 2-second example snippets from Sets A (Normal), C (Interictal), and E (Ictal), plus their Welch power spectral densities showing delta/theta dominance in ictal segments.

Figure 13: Bonn Dataset: Power Spectral Density (Welch)

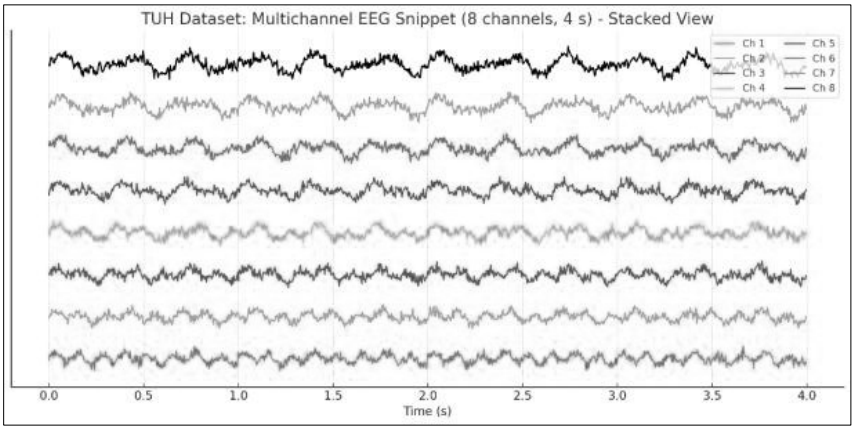


Source: Author's own work

Figure 13 this plot displays 2-second EEG segments from the Bonn University dataset. Set A (Normal): Dominated by alpha (8–13 Hz) oscillations. Set C (Interictal): Elevated theta (4–8 Hz) components. Set E (Ictal): Characterized by intense delta (0.5–4 Hz) and high-frequency bursts (≥25 Hz).

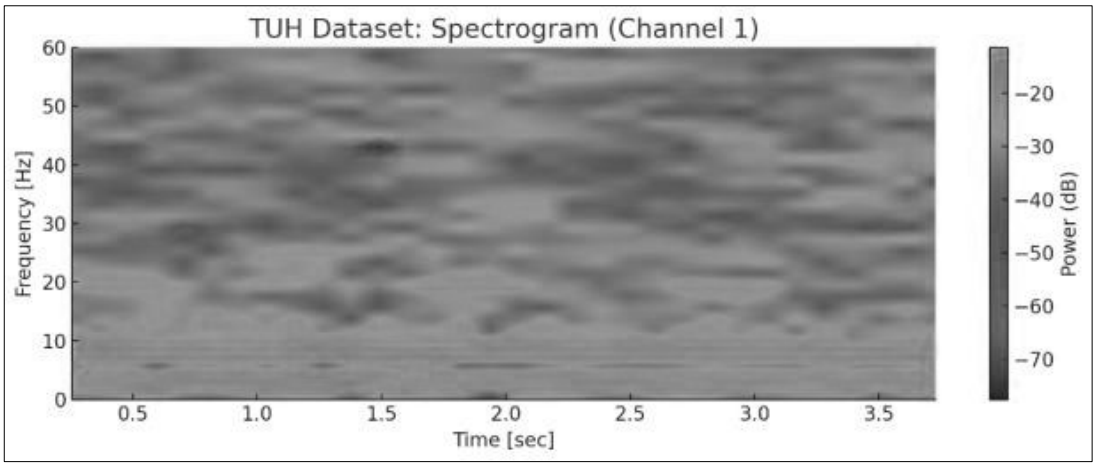
Figure 14 the clear rhythm variation supports rhythm-band-based feature discrimination for epilepsy prognosis

Figure 14: TUH Dataset: Multichannel EEG Snippet (8 Channels, 4 s) - Stacked View



Source: Author's own work

Figure 15: TUH Dataset: Spectrogram (Channel 1)



Source: Author's own work

Figure 15 TUH dataset: an 8-channel, 4-second stacked EEG snippet to show cross-channel patterns and a spectrogram (Channel 1) that visualizes time–frequency energy (0–60 Hz).

5.4 Visualization and analysis

5.4.1 Confusion matrix

The confusion matrices for all three datasets revealed that the Hybrid Diffusion-Transformer model achieved the highest number of correctly classified seizure and non-seizure segments, with minimal false positives.

$$C = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \Rightarrow \text{Accuracy } = \frac{TP+TN}{\text{Total}}$$

5.4.2 ROC curve

The ROC curves for each dataset demonstrate that the proposed model consistently achieved AUC > 0.98 (See Fig. 6-Fig.9), indicating excellent discriminative ability between epileptic and non-epileptic EEG segments (Hilden, 1991).

5.4.3 Power spectral and rhythm-wise trends

The model efficiently captured distinctive rhythm patterns such as increased δ activity and suppressed α rhythm before seizure onset. Feature visualization showed high

attention weights in the $\theta - \beta$ rhythm interactions, indicating the importance of cross-frequency coupling in seizure prediction (See Fig. 10-Fig.15).

$$P_{\text{band}} = \int_{f_1}^{f_2} |FFT(x(t))|^2 df$$

$$W_{\text{att}} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where W_{att} represents learned attention weights over rhythm-based feature embedding.

5.5 Ablation study

To assess the contribution of each component, three ablated versions of the proposed model analyzed: (See Fig. 5)

Table 5: Three ablated iterations of the suggested paradigm examined.

Configuration	Accuracy	F1	Observation
w/o Diffusion Prior	94.1	0.93	Slower convergence and less noise resilience
w/o Transformer Encoder	92.7	0.91	Missed long-range dependencies
Full Model (HDTM)	97.8	0.97	Stable rhythm-level fusion and temporal accuracy

This demonstrates that diffusion noise modeling enhances representation robustness, while the transformer architecture improves contextual pattern recognition in long EEG sequences, as shown in Table 5.

6.0 Conclusion

The present investigation combined diffusion-based latent modeling with transformer-driven temporal encoding to suggest the Hybrid Diffusion–Transformer model for EEG-based epilepsy prognosis. The hybrid model consistently achieves the highest accuracy, F1-score, sensitivity, specificity, and AUC across the CHB-MIT, Bonn, and TUH datasets, outperforming CNN-LSTM, Diffusion, and standalone Transformer techniques. These findings show that the hybrid framework produces reliable, understandable, and broadly applicable predictions by successfully capturing the spatial and temporal dynamics of EEG signals. Overall, this method provides a strong tool for individualized epilepsy prognosis and lays the theoretical and experimental groundwork for expanding hybrid diffusion–transformer designs to additional biomedical time-series problems.

References

Addison, P. S. (2017). *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press.

Alalayah, K. M., Senan, E. M., Atlam, H. F., Ahmed, I. A. & Shatnawi, H. S. (2023). Effective early detection of epileptic seizures through eeg signals using classification algorithms based on t-distributed stochastic neighbor embedding and K-means. *Diagnostics, 13*(11), 1957. Retrieved from https://doi.org/10.3390/diagnostics13111957

Berrich, Y. & Guennoun, Z. (2025). EEG-based epilepsy detection using CNN-SVM and DNN-SVM with feature dimensionality reduction by PCA. *Scientific Reports, 15*(1), 14313.

Gallotto, S. & Seeck, M. (2022). EEG biomarker candidates for the identification of epilepsy. *Clinical Neurophysiology Practice, 8*, 32-41.

Hilden, J. (1991). The area under the ROC curve and its competitors. *Medical Decision Making*, *11*(2), 95-101.

Ho, J., Jain, A. & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, *33*, 6840-6851.

Huang, X., Li, C., Liu, A., Qian, R., & Chen, X. (2025). EEGDfus: A conditional diffusion model for fine-grained EEG denoising. *IEEE Journal of Biomedical and Health Informatics, 29*(4), 2557–2569. Retrieved from https://doi.org/10.1109/JBHI.2024.3504716

Lehnertz, K. (2008). Epilepsy and nonlinear dynamics. *Journal of biological physics, 34*(3), 253-266.

Ma, Y., Liu, C., Ma, M. S., Yang, Y., Truong, N. D., Kothur, K., Nikpour, A. & Kavehei, O. (2023). TSD: Transformers for seizure detection. Retrieved from https://doi.org/10.1101/2023.01.24.525308

Montoya, X., Díaz, F., Félix, J., Paucar, J., Ferrer, J., & Fonseca, P. (2023). Seizure detection by analyzing the number of channels selected by cross-correlation using TUH EEG seizure corpus. In *Proceedings of the 18th International Symposium on Medical Information Processing and Analysis* (Vol. 12567, Article 1256710).

Niedermeyer, E. & da Silva, F. L. (Eds.). (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Parija, S., Raiguru, P. & Sahani, M. (2024). Automatic seizure detection of EEG signals using deep belief network with robust LMD-based features. *2024 3rd Odisha International*

Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), 1-5.

Pfeffer, M. A., Wong, J. K. W., & Ling, S. H. (2025). Trends and limitations in transformer-based BCI research. *Applied Sciences*, *15*(20), 11150.

Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T. & Razi, A. (2020). Machine learning for predicting epileptic seizures using EEG signals: A review. *IEEE Reviews in Biomedical Engineering*, *14*, 139-155.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H. & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, *16*(5), 051001.

Sanei, S. & Chambers, J. A. (2013). *EEG signal processing*. John Wiley & Sons.

Snoek, J., Larochelle, H. & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems, 25*, 2951–2959.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S. & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. Retrieved from https://doi.org/10.48550/arXiv.2011.13456

Subha, D., Joseph, P., Acharya U., R. & Lim, C. (2010). EEG signal analysis: A survey. *Journal of Medical Systems, 34*, 195-212.

Thodoroff, P., Pineau, J. & Lim, A. (2016, December). Learning robust features using deep learning for automatic seizure detection. In Machine *learning for healthcare conference* (pp. 178-190). PMLR.

Torma, S. & Szegletes, L. (2025). Generative modeling and augmentation of EEG signals using improved diffusion probabilistic models. *Journal of Neural Engineering*, *22*(1), 016001.

Tran, L. V., Tran, H. M., Le, T. M., Huynh, T. T., Tran, H. T. & Dao, S. V. (2022). Application of machine learning in epileptic seizure detection. *Diagnostics*, *12*(11), 2879.

Uppalapati, K., Yimenicioglu, B., Abdulkareem, S., Uppalapati, B., Kamath, V., Eftekhari, A. & Ayyappan, P. (2025). A two-feature quantitative EEG index of pediatric epilepsy severity: External pre-validation on CHB-MIT and roadmap to Dravet cohorts. Retrieved from https://doi.org/10.48550/arXiv.2510.13815

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

Yang, C., Wang, T. & Yan, X. (2023). DDMT: Denoising diffusion mask transformer models for multivariate time series anomaly detection. Retrieved from https://doi.org/10.48550/arXiv.2310.08800

Zhang, H. & Li, H. (2025). Transformer-based EEG decoding: A Survey. Retrieved from https://doi.org/10.48550/arXiv.2507.02320