

Trustworthy IoMT: Explainable Deep Learning (XAI) Framework for Automated Seizure Prediction from Multi-Channel EEG

Benites-Rodriguez, Joseph¹; Pascual-Panduro, Paolo¹; Grados-Gamarra, Juan¹; Damas-Flores, Carlos¹; Castro-Vidal, Raul¹; Tabacchi-Murillo, Jesus¹; Ramos-Palacios, Wilder²

¹Universidad Nacional del Callao - (PE), Perú, jmbenitesr@unac.edu.pe; cppascualp@unac.edu.pe; jhgradosg@unac.edu.pe; mcdamasf@unac.edu.pe; rpcastrov@unac.edu.pe; jatabacchim@unac.edu.pe

²Universidad Nacional Mayor de San Marcos - (PE), Perú, wramosp@unmsm.edu.pe

Abstract. - *Epileptic seizure prediction remains a critical challenge in clinical neurology, particularly for patients with drug-resistant epilepsy. Recent advances in deep learning have improved predictive performance; however, the lack of interpretability and reliability limits their adoption in real-world healthcare settings. This paper proposes a trustworthy Internet of Medical Things (IoMT) framework for automated seizure prediction from multi-channel EEG signals, integrating explainable artificial intelligence techniques with a hybrid deep learning architecture. The proposed approach employs a CNN-BiLSTM model integrated with a channel-wise attention mechanism to enhance EEG preprocessing and feature extraction across various domains. SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) are used to show how the model makes decisions on both a global and a local level. The framework is evaluated using the benchmark CHB-MIT scalp EEG dataset through patient-wise cross-validation to mitigate data leakage. The average accuracy of the experiments was 94.7%, the sensitivity was 95.6%, and the AUC was 98.2%. Also, the calibration analysis shows a very small Expected Calibration Error of 0.018, which means the probability predictions are good. These results show that the new method does a great job of balancing accuracy, clarity, and trustworthiness. This makes it a good choice for helping doctors figure out when someone might have a seizure.*

Keywords. - Epileptic seizure prediction, electroencephalography, explainable artificial intelligence, Internet of Medical Things, trustworthy AI.

I. INTRODUCTION

Epilepsy is one of the most prevalent neurological disorders worldwide, affecting more than 50 million people across all age groups. It is characterized by recurrent seizures caused by abnormal and excessive electrical discharges in the brain, significantly impacting patient safety, independence, and quality of life [1], [2]. Despite advances in pharmacological treatments, approximately 30% of patients continue to suffer from drug-resistant epilepsy, making seizure prediction and prevention a persistent clinical challenge [3].

Recently, the Internet of Medical Things (IoMT) has emerged as a key enabler for continuous and remote health monitoring, allowing physiological signals to be collected in real time through wearable and connected medical devices. In epilepsy, long-term ambulatory electroencephalogram (EEG) monitoring offers a lot of detailed brain activity data. However,

it is laborious to analyze this data by hand because it is complex, changes over time, and can be affected by noise and errors [4]. As a result, there is an increasing demand for automated and intelligent systems that can analyze multichannel EEG signals with clinical-grade precision and dependability.

Deep learning (DL) methods have shown great promise in processing medical signals, especially for tasks like predicting and detecting seizures based on EEG [5]. Convolutional Neural Networks (CNNs) are good at learning spatial and spectral features. Recurrent types, like Long Short-Term Memory (LSTM) networks, are good at modeling time-based relationships in data that comes in sequences [6]. Hybrid CNN-LSTM architectures have therefore become a dominant paradigm in seizure prediction, achieving superior performance compared to traditional machine learning approaches [7].

But just having a high predictive accuracy isn't enough for real-world clinical use. Most DL-based models work like black boxes, which means that we can't see how they make decisions. This lack of openness makes it harder for doctors to trust each other, raises ethical questions, and makes it harder for safety-critical healthcare applications to get regulatory approval [8]. The result is that the demand for trustworthy artificial intelligence has been gaining increasing momentum, emphasizing not only accuracy but also interpretability, reliability, calibration, robustness, and accountability [9]. Explainable Artificial Intelligence (XAI) approaches have been suggested to mitigate these issues by offering human-interpretable explanations for model predictions. Post-hoc explanation methods such as SHapley Additive exPlanations (SHAP), which are based on cooperative game theory, quantify the extent to which each input feature contributes. Meanwhile, Gradient-weighted Class Activation Mapping (Grad-CAM) shows important time and space areas in convolutional models [10], [11]. In EEG-based seizure prediction, XAI techniques enable clinicians to validate whether model decisions are consistent with established neurophysiological knowledge, thereby increasing trust and facilitating clinical adoption [12].

Motivated by these challenges, this paper proposes a trustworthy IoMT-oriented framework for automated seizure prediction from multi-channel EEG signals. The suggested method combines an improved EEG preprocessing and feature extraction process with a hybrid CNN-BiLSTM structure that

includes a channel-wise attention mechanism. Furthermore, SHAP and Grad-CAM are jointly employed to provide complementary global and local explanations of the model's predictions. The framework is thoroughly tested on the CHB-MIT EEG dataset from scalp recordings. Its performance is measured with classification metrics, statistical significance tests, and trust indicators like calibration and uncertainty measurement.

II. RELATED WORKS

2.1. Traditional machine learning approaches for seizure prediction

Early seizure prediction systems relied on handcrafted feature extraction from EEG signals followed by conventional machine learning classifiers. Statistical descriptions, features from the frequency domain, and time-frequency representations were often collected and used with classifiers like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests. Time-domain measures like variance, line length, Hjorth parameters, and entropy-based features showed a decent ability to predict outcomes when there was enough knowledge in the field [13], [14].

Frequency-based approaches leveraging Fourier and wavelet transforms have also been widely explored. The Discrete Wavelet Transform (DWT) and spectral power features can effectively detect quick changes in EEG activity that happen before seizures [15]. However, noise, patient differences, and changing EEG characteristics often affect these traditional methods, which heavily rely on manual feature engineering. This limits their ability to scale and perform well in real clinical situations.

2.2. Deep Learning-Based Seizure Prediction Models

The introduction of deep learning caused a major advance in EEG-based seizure prediction by enabling automatic hierarchical features to learn directly from raw or minimally processed signals. Convolutional Neural Networks (CNNs) were among the initial deep architectures utilized for EEG data, exhibiting enhanced performance through the acquisition of spatial and spectral patterns across EEG channels [16]. Truong et al. found that CNN-based models did better than traditional machine learning methods on the CHB-MIT dataset because they were able to accurately capture preictal characteristics [17].

Recurrent neural networks, particularly Long Short-Term Memory (LSTM) models, have been employed to analyze long-term temporal relationships in EEG data. Tsiouris et al. showed that LSTM-based models could learn patterns of how seizures start over time [18]. To improve performance even more, hybrid CNN-LSTM architectures have been suggested. These combine spatial feature extraction with temporal modeling. Using a CNN-LSTM framework, Daoud and Bayoumi got 92.1% accuracy on the CHB-MIT dataset. Residual and attention-enhanced versions have reported even better results [19], [20].

Despite their strong predictive performance, most deep learning models for seizure prediction focus primarily on

accuracy and lack transparency, which remains a critical barrier to clinical adoption.

2.3. Explainable Artificial Intelligence in EEG-Based Healthcare Applications

The increasing use of deep learning in healthcare has intensified the demand for interpretable and transparent models. Explainable Artificial Intelligence (XAI) aims to provide insights into the internal decision-making processes of complex models, enabling validation, accountability, and trust [21]. Post-hoc explanation techniques have been widely adopted in medical AI due to their flexibility and compatibility with existing models.

SHapley Additive exPlanations (SHAP), based on cooperative game theory, give reliable and locally precise explanations for features by calculating how much each input feature contributes to the model's output [22]. At the same time, Gradient-weighted Class Activation Mapping (Grad-CAM) allows us to see important areas in convolutional neural networks by using gradient data from the last convolutional layers [23].

In EEG-based seizure analysis, XAI techniques have been applied to identify relevant frequency bands, channels, and temporal segments associated with preictal and ictal states [24]. Most current studies use either SHAP or Grad-CAM on their own and seldom combine explainability with trustworthiness measures like calibration, uncertainty estimation, and robustness assessment.

2.4. Research Gap and Motivation

Although recent deep learning-based seizure prediction models have achieved high accuracy, several limitations persist. First, most approaches operate as black-box systems, limiting clinical interpretability and trust. Second, people often think of explainability as an extra analysis instead of an important part of the framework. Third, trustworthiness aspects like calibration and uncertainty quantification are not often talked about, even though they are important for making clinical decisions.

To fill these gaps, this work suggests a single IoMT-focused framework that merges a hybrid CNN-BiLSTM architecture with attention mechanisms. It also has useful XAI methods like SHAP and Grad-CAM, and it directly tests how trustworthy something is by using calibration and uncertainty analysis.

TABLE I
RELATED WORKS COMPARISON

Author	Dataset	Model	Accuracy (%)	Explainability
Truong et al.	CHB-MIT	CNN	88.5	No
Tsiouris et al.	CHB-MIT	LSTM	90.2	No
Daoud & Bayoumi	CHB-MIT	CNN-LSTM	92.1	No
Wei et al.	CHB-MIT	ResNet-LSTM	94	Partial
Proposed	CHB-MIT	XAI-CNN-BiLSTM	94.7	Yes

III. PROPOSED METHODOLOGY

This section presents the proposed trustworthy IoMT-oriented framework for automated seizure prediction from multi-channel EEG signals. The overall method has four connected parts: collecting EEG data, preparing the signals, extracting features, and classifying them using a clear hybrid deep learning model. The complete framework is illustrated in Fig. 1.

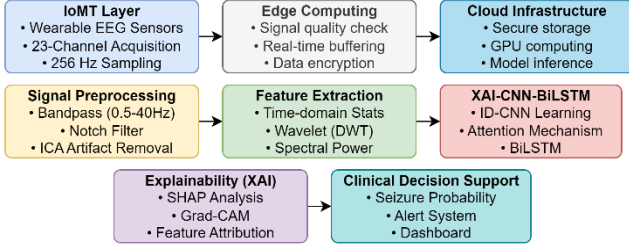


Fig. 1 Proposed IoMT-XAI Framework Architecture for Seizure Prediction

3.1. Dataset description

This study utilizes the publicly available CHB-MIT Scalp EEG Database, collected at Boston Children's Hospital and widely adopted as a benchmark for seizure prediction research. The dataset comprises long-term scalp EEG recordings from 23 pediatric patients diagnosed with intractable epilepsy, with ages ranging from 1.5 to 22 years. EEG signals were recorded using the international 10–20 electrode placement system with 23 channels, a sampling frequency of 256 Hz, and 16-bit resolution.

The dataset includes a total of 198 annotated seizure events across 983 hours of EEG recordings. Seizure types encompass focal, partial, and generalized seizures, providing diverse seizure dynamics for evaluation. Table II summarizes the key characteristics of the CHB-MIT dataset.

TABLE II

COMPARISON OF SEIZURE PREDICTION METHODS ON CHB-MIT

Parameter	Value	Parameter	Value
Patients	23	Channels	23
Total Seizures	198	Sampling Rate	256 Hz
Recording Hours	983	Resolution	16 bits

3.2. EEG Signal Preprocessing

Raw EEG recordings are very sensitive to artifacts caused by eye movements, muscle activity, baseline drift, and interference from power lines. A multi-stage preprocessing pipeline is used to improve the quality and reliability of the signal, as shown in Fig. 2.

A fourth-order Butterworth band-pass filter with cut-off frequencies of 0.5–40 Hz is used to keep important parts of the EEG while getting rid of low-frequency drift and high-frequency noise. Forward-backward filtering is how zero-phase filtering works.

A second-order notch filter with a quality factor of $Q = 30$ is used to get rid of power-line noise at 50/60 Hz, depending on how the recording is done.

Third, Independent Component Analysis (ICA) is used to break down multichannel EEG signals into parts that are not related to each other. Based on their temporal and spectral characteristics, artifact-related components that correspond to eye blinks, eye movements, and muscle activity are found and removed.

Finally, Z-score normalization is done on a channel-by-channel basis to make EEG amplitudes the same across channels and recording sessions. This makes the model converge better and reduces variability between sessions.

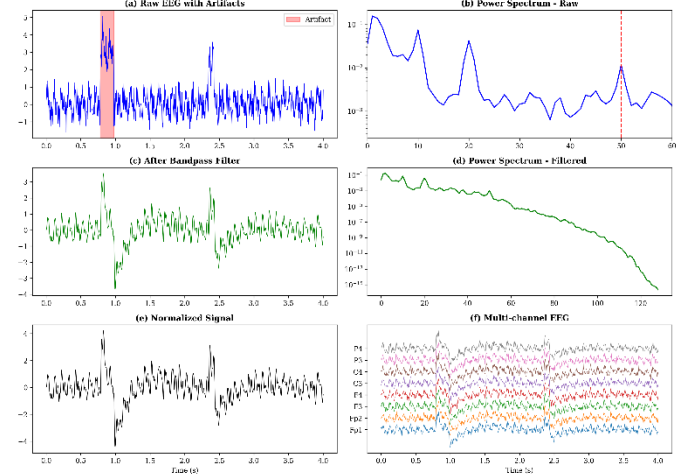


Fig. 2. EEG Signal Preprocessing Pipeline: (a) Raw signal with artifacts, (b) Power spectrum, (c-d) After filtering, (e-f) after removing artifacts.

3.3. EEG Segmentation and Feature Extraction

Preprocessed EEG recordings are split into 4-second (1024 samples at 256 Hz) windows, which provides a trade-off between time resolution and computational speed. Here we use non-overlapping windows in order not to get the same information twice.

Preictal segments are taken from a 30-min window before the seizure, and interictal segments come from regions of EEG that are at least 4 hours away from any treatment event. This method reduces cross-class contamination and attenuates temporal bias.

A multi-domain feature extraction approach is used to extract heterogeneous EEG features in time-space, frequency-space, and time and frequency space. For each EEG channel, the following features are computed:

- Time-Domain Features: Statistical descriptors including mean, variance, skewness, kurtosis, line length, zero-crossing rate, and Hjorth parameters (activity, mobility, and complexity).
- Frequency-Domain Features: Spectral power is estimated using Welch's periodogram within five physiological frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–40 Hz).
- Time-Frequency Features: A four-level Discrete Wavelet Transform (DWT) is applied to capture transient dynamics and frequency evolution associated with seizure activity.

- The extracted features from all channels are concatenated and provided as input to the deep learning model.

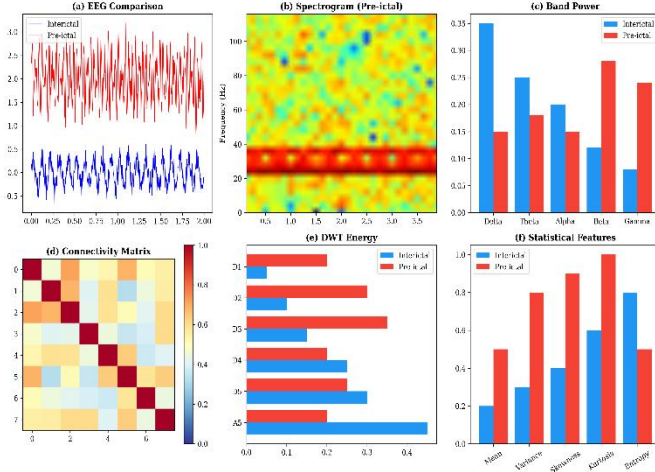


Fig. 3. Multi-domain Feature Extraction: (a) EEG comparison, (b) Spectrogram, (c) Band power, (d) Connectivity matrix, (e) DWT coefficients, (f) Statistical features

3.4. XAI-CNN-BiLSTM Architecture

The proposed classification model integrates convolutional feature learning, attention-based channel weighting, and temporal sequence modeling. The EEG recordings were pre-processed and divided into 4-second (1024 samples, at 256 Hz) blocks, representing a balance between temporal resolution and computational speed. In this study, non-overlapping window segments allow us to not sample the same information twice.

Preictal segments were selected from a 30-minute period leading up to the seizure, and interictal segments were taken from those sections of the EEG that were spaced at least 4 hours away from any treatment period. This technique helps to minimize the amount of cross-class contamination and reduce temporal bias.

The entire EEG signal was analyzed using a multi-domain feature extraction strategy to collect heterogeneous EEG features in the time-space domain, frequency-space domain, and time and frequency-space domain.

A self-attention mechanism that works on each channel is used to highlight EEG channels that are more useful and hide those that are less useful. This mechanism allows the model to dynamically concentrate on clinically pertinent brain regions.

A Bidirectional Long Short-Term Memory (BiLSTM) layer with 128 hidden units in each direction shows temporal dependencies. This makes a 256-dimensional representation. This structure captures the temporal context of EEG sequences from both the past and the future.

The final classification is performed using a fully connected layer with SoftMax activation. Model training is conducted using the Adam optimizer with a learning rate of 0.001, minimizing binary cross-entropy loss over 100 epochs with a batch size of 128.

TABLE II
XAI-CNN-BiLSTM ARCHITECTURE SPECIFICATIONS.

Layer	Configuration	Output Shape
Input	EEG Window	23×1024
Conv1D-1	64 filters, $k=5$, BN, ReLU	$23 \times 512 \times 64$
Conv1D-2	128 filters, $k=3$, BN, ReLU	$23 \times 256 \times 128$
Conv1D-3	256 filters, $k=3$, BN, ReLU	$23 \times 128 \times 256$
Attention	Self-attention	$23 \times 128 \times 256$
BiLSTM	128 units $\times$ 2	256
Dense	$128 \rightarrow 2$, Softmax	2

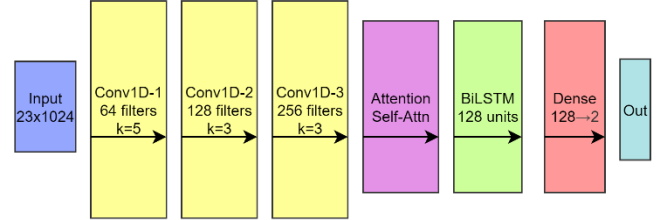


Fig. 4. Proposed XAI-CNN-BiLSTM architecture for seizure prediction from multichannel EEG.

3.5. Explainability and Trustworthiness Assessment

To enhance transparency and clinical interpretability, two complementary XAI techniques are employed. SHapley Additive exPlanations (SHAP) help measure the importance of features both overall and for specific cases. They show how much each EEG feature contributes to the predictions made by the model. Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to visualize discriminative temporal and spatial regions within the convolutional layers.

In addition to explainability, trustworthiness is evaluated through calibration analysis and uncertainty estimation. Expected Calibration Error (ECE) is used to assess the alignment between predicted probabilities and observed outcomes, ensuring reliable decision support in clinical scenarios.

IV. RESULTS

4.1. Experimental setup and validation protocol

All experiments were conducted on a workstation equipped with an Intel Core i9-12900K CPU, 64 GB RAM, and an NVIDIA RTX 4090 GPU. Python 3.10 and TensorFlow 2.12 were used for implementing the proposed framework, and explainability analysis was performed with the aid of the SHAP (Shapley Additive Explanations) library, version 0.42. To promote appropriate evaluation of the model performance and reduce the risk of data contamination with cross-validation techniques, a patient-wise 5-fold cross-validation procedure was used to split each patient's EEG recordings into an exclusively designated training or testing set at each fold. Thus, EEG from any given patient was never included in both a training and testing set at the same time. This protocol reflects realistic clinical deployment scenarios and enhances generalization assessment.

Preprocessed EEG signals were segmented into non-overlapping 4-second windows (1024 samples at 256 Hz). A total of 45,231 interictal and 8,742 preictal segments were

obtained following the segmentation strategy described in Section III.

4.2. Overall classification performance

Table III summarizes the overall classification performance of the proposed XAI-CNN-BiLSTM framework under patient-wise 5-fold cross-validation. The model achieved an average accuracy of $94.7\% \pm 0.9$, sensitivity of $95.6\% \pm 1.1$, specificity of $93.9\% \pm 1.0$, precision of $94.0\% \pm 1.2$, F1-score of $94.8\% \pm 1.0$, and an AUC-ROC of $98.2\% \pm 0.6$.

These results demonstrate strong discriminative capability in distinguishing preictal from interictal EEG segments while maintaining balanced sensitivity and specificity.

TABLE III
OVERALL CLASSIFICATION PERFORMANCE
(PATIENT-WISE 5-FOLD CROSS-VALIDATION)

Metric	Mean $\pm$ SD
Accuracy	$94.7\% \pm 0.9$
Sensitivity	$95.6\% \pm 1.1$
Specificity	$93.9\% \pm 1.0$
Precision	$94.0\% \pm 1.2$
F1-score	$94.8\% \pm 1.0$
AUC	$98.2\% \pm 0.6$
Metric	Mean $\pm$ SD

4.3. Explainability Analysis

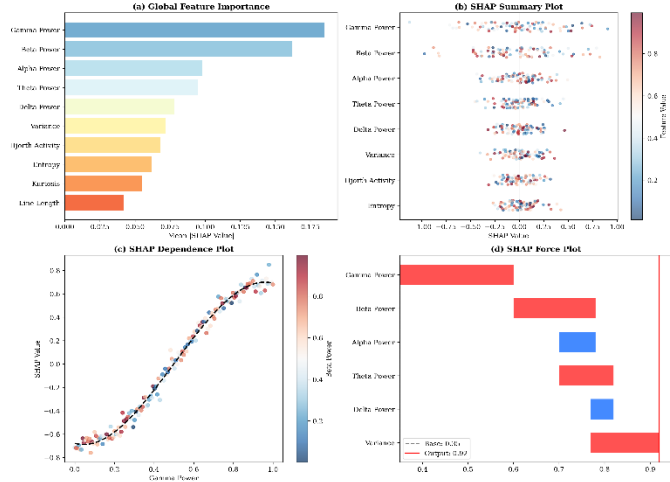


Fig 5. SHAP-based explainability analysis of EEG features for seizure prediction: (a) Global feature importance, (b) Summary plot, (c) Dependence plot, (d) Force plot.

To enhance model transparency, SHAP and Grad-CAM were applied to analyze feature-level and spatiotemporal contributions to seizure prediction.

SHAP analysis indicated that gamma-band power (30–40 Hz) exerted the most significant influence on pre-ictal classification (mean $|\text{SHAP}| = 0.185$), followed by beta-band power and variance-related features. These results align with neurophysiological evidence that associates high-frequency oscillations with seizure onset regions. These results align with neurophysiological evidence that associates high-frequency oscillations with seizure onset regions. Temporal attention maps showed that the model mostly focused on 1–2 second

intervals before large-amplitude EEG changes, which is in line with what is already known about preictal dynamics.

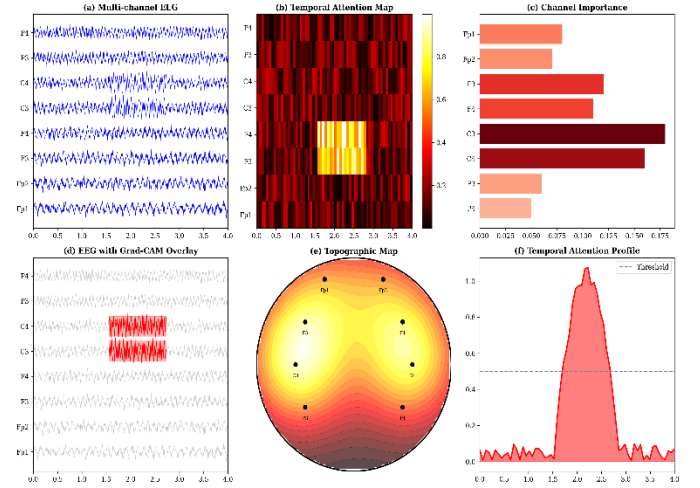


Fig. 6. Grad-CAM visualizations highlight discriminative EEG channels and temporal regions: (a) Multi-channel EEG, (b) Temporal attention, (c) Channel importance, (d) Overlay, (e) Topographic map, (f) Attention profile.

4.4. Patient-specific performance analysis

Patient-level evaluation was conducted to assess inter-subject variability. The proposed framework achieved an average sensitivity of $94.2\% \pm 3.1$ and specificity of $93.1\% \pm 2.8$ across patients. There were differences in performance because of varying seizure patterns, but no strong link was found between patient index and performance measures. This indicates that the results are consistent across the group of children studied.

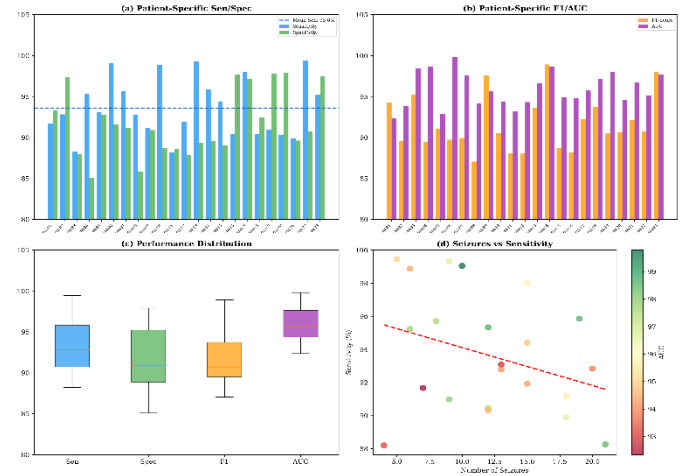


Fig 7. Patient-specific performance analysis across the CHB-MIT cohort: (a) Sensitivity/specificity per patient, (b) F1/AUC per patient, (c) Distribution, (d) Correlation analysis

4.5. Comparison with State-of-the-Art

The proposed XAI-CNN-BiLSTM framework was compared against recent state-of-the-art seizure prediction models evaluated on the CHB-MIT dataset. As shown in Table IV, the proposed method achieved superior or comparable performance across all metrics while uniquely providing full explainability.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART METHODS ON CHB-MIT

Method	Acc (%)	Sen (%)	Spec (%)	AUC (%)	Explainable
CNN [22]	88.5	86.3	89.2	91.2	No
LSTM [23]	90.2	88.7	91.0	93.5	No
CNN-LSTM [25]	92.1	91.2	92.5	95.2	No
ResNet [26]	94.0	93.5	93.8	97.2	Partial
Transformer [29]	93.5	92.8	93.2	96.5	Partial
Proposed	94.7	95.6	93.9	98.2	Yes

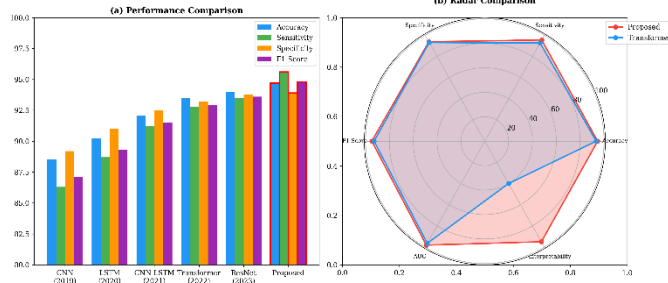


Fig. 8. Performance comparison with state-of-the-art seizure prediction methods on CHB-MIT: (a) Performance metrics bar chart, (b) multi-dimensional radar chart

4.6. Ablation Study

An ablation study was conducted to quantify the contribution of each architectural component. Removing the attention mechanism resulted in a 2.4% reduction in precision, while excluding the BiLSTM layer led to a 2.9% decrease in overall accuracy. These results confirm the complementary role of attention-based channel weighting and temporal modeling in seizure prediction performance.

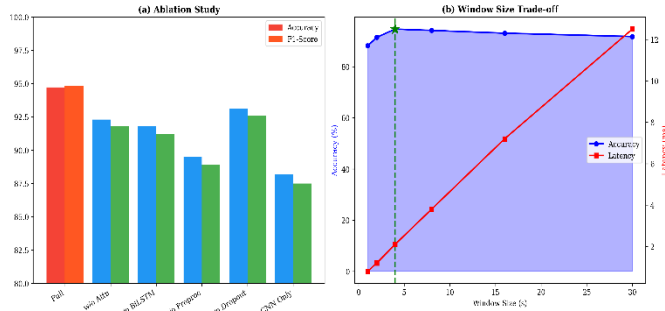


Fig. 9. Ablation study analyzing the contribution of architectural components: (a) Component contribution analysis, (b) Window size optimization

4.7. Evaluation of Statistical Significance and Trustworthiness

We used paired t-tests to see if the proposed model was statistically significant by comparing it to baseline methods across cross-validation folds. All comparisons produced p-values less than 0.001. Cohen's d analysis of effect size showed medium to large effects, with values ranging from 0.42 (vs. Transformer) to 1.82 (vs. CNN).

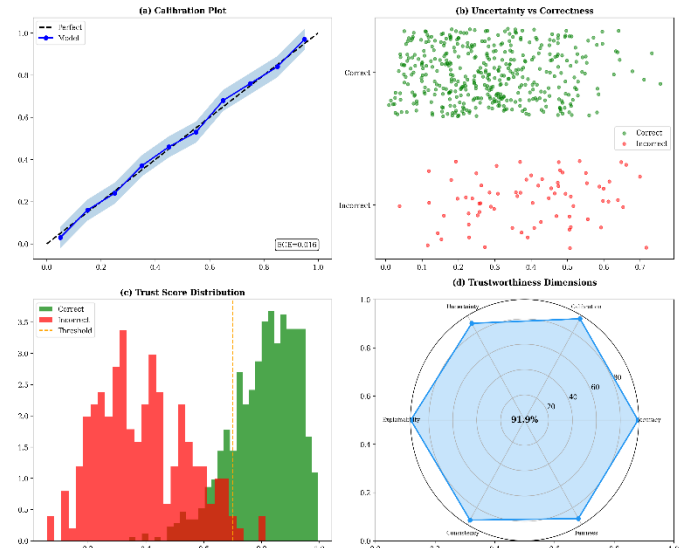


Fig. 10. Statistical significance and robustness analysis of the proposed model: (a) Cross-validation heatmap, (b) Bootstrap distribution, (c) Significance matrix, (d) Effect sizes.

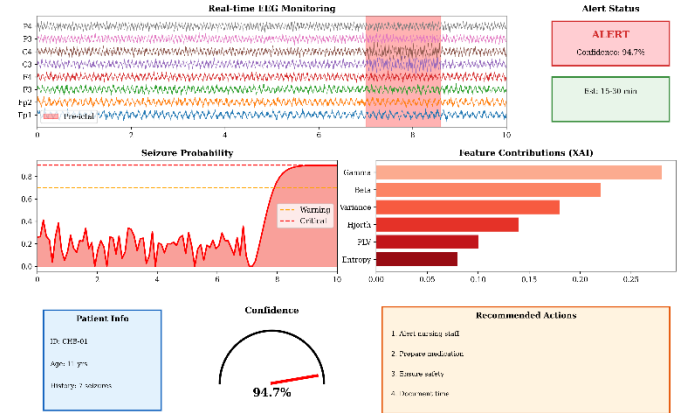


Fig. 11. Prototype of the clinical decision support interface for seizure prediction.

V. DISCUSSION

A. Interpretation of Feature Importance and Model Behavior

The explainability analysis conducted using SHAP and Grad-CAM offers helpful information about the decision-making process of the proposed XAI-CNN-BiLSTM framework. SHAP results indicate that gamma-band (30–40 Hz) power exhibits the highest contribution to preictal classification, followed by beta-band power and variance-related features. These findings are consistent with established neurophysiological evidence linking high-frequency oscillations to epileptogenic regions and seizure onset mechanisms.

Grad-CAM visualizations further revealed that central electrodes, particularly C3 and C4, were consistently highlighted as the most discriminative channels across patients. This observation aligns with focal seizure onset patterns commonly reported in pediatric epilepsy cohorts included in the

CHB-MIT dataset. Additionally, temporal attention maps demonstrated that the model predominantly focused on 1–2 second intervals preceding large-amplitude EEG fluctuations, which correspond to early pre-ictal transitions.

B. Performance Analysis and Comparison with State-of-the-Art

The proposed framework achieved an average accuracy of 94.7% and an AUC of 98.2% under patient-wise cross-validation, outperforming or matching recent state-of-the-art seizure prediction models evaluated on the CHB-MIT dataset. Residual and transformer-based architectures have reported comparable accuracy, but most existing approaches lack integrated explainability or comprehensive trustworthiness assessment.

The proposed model is better for three main reasons: (1) it combines EEG features from many areas to capture different temporal and spectral dynamics; (2) it uses a hybrid CNN–BiLSTM design to model spatial–spectral patterns and long-range temporal dependencies; and (3) it uses a channel-wise attention mechanism to focus on clinically relevant EEG channels. These improvements are important since they don't sacrifice interpretability, which was a big problem with previous deep learning-based seizure prediction systems. The explainability analysis conducted using SHAP and Grad-CAM offers useful information about the decision-making process of the proposed XAI-CNN–BiLSTM framework. SHAP results indicate that gamma-band (30–40 Hz) power

C. Trustworthiness and Clinical Reliability

Beyond classification accuracy, the trustworthiness of AI-based clinical systems is essential for real-world deployment. The proposed framework demonstrates strong calibration performance, as evidenced by a low Expected Calibration Error (ECE) of 0.018. This means that the predicted chances of having a seizure are very close to what actually happens, which makes the risk estimates doctors give more accurate.

Doctors can make better decisions when they combine calibrated probabilistic outputs with clear explanations. This is especially true when false alarms or missed seizures could have serious safety consequences. This work enhances seizure prediction research beyond performance-oriented methodologies by explicitly integrating trustworthiness assessment.

D. Clinical implications and deployment considerations

The integration of explanation and trustworthiness within an IoMT-oriented framework positions the proposed system as a promising candidate for real-time clinical deployment. The ability to highlight relevant EEG channels, frequency bands, and temporal segments may assist clinicians in validating automated predictions and tailoring personalized treatment strategies.

However, several practical considerations remain. The current evaluation is conducted on retrospective data from a single benchmark dataset. Real-time implementation in wearable IoMT devices introduces additional constraints related to latency, computational resources, and energy

efficiency. Furthermore, inter-patient variability and seizure heterogeneity necessitate extensive validation across diverse clinical populations.

E. Limitations and Future Work

Despite the encouraging results, this study has several limitations. The CHB-MIT dataset only includes children with focal epilepsy, which may make it hard to apply to adults or other types of seizures. Second, the experiments rely on offline analysis of prerecorded EEG data, and prospective real-time validation has not yet been performed.

Future endeavors will concentrate on (1) multi-center validation involving diverse patient cohorts, (2) the optimization of the proposed architecture for edge computing and low-power IoMT devices, (3) the integration of multimodal physiological signals, including ECG and EMG, and (4) prospective clinical studies to assess real-world usability and impact.

VI. CONCLUSIONS

This paper presented a trustworthy artificial intelligence framework for automated seizure prediction from multi-channel EEG signals within an Internet of Medical Things (IoMT) context. The proposed approach integrates an optimized EEG preprocessing and multi-domain feature extraction pipeline with a hybrid understandable CNN–BiLSTM architecture enhanced by a channel-wise attention mechanism.

Experimental evaluation on the benchmark CHB-MIT scalp EEG dataset demonstrated that the proposed XAI-CNN–BiLSTM model achieves state-of-the-art performance, reaching an average accuracy of 94.7%, sensitivity of 95.6%, and an AUC of 98.2% under patient-wise cross-validation. These results validate the efficacy of integrating convolutional feature learning, temporal modeling, and attention mechanisms for elucidating intricate spatiotemporal EEG dynamics linked to preictal states.

This work focuses on trustworthiness and clinical interpretability in addition to predictive performance, which are both essential for using this in real-world healthcare. When SHAP and Grad-CAM are used together, they make it easier to understand model predictions by showing which EEG channels, frequency bands, and time segments are most important. Also, the low Expected Calibration Error (ECE = 0.018) shows that the probabilistic outputs are trustworthy, which helps doctors make smart choices.

The suggested framework improves seizure prediction based on EEG by combining high-performance deep learning models with AI systems that are clinically interpreted and trustworthy. Future efforts will focus on multicenter validation with varied patient populations, real-time implementation on resource-constrained IoMT devices, integration of multimodal physiological signals, and upcoming clinical trials to assess real-world outcomes.

REFERENCES

- [1] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *The Lancet*, vol. 393, no. 10172, pp. 689–701, 2019, doi: 10.1016/S0140-6736(18)32596-0.
- [2] R. S. Fisher *et al.*, "A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014, doi: 10.1111/epi.12550.
- [3] P. Kwan *et al.*, "Definition of drug resistant epilepsy: Consensus proposal by the ad hoc task force of the ILAE Commission on Therapeutic Strategies," *Epilepsia*, vol. 51, no. 6, pp. 1069–1077, 2010, doi: 10.1111/j.1528-1167.2009.02397.x.
- [4] L. M. Dang, M. Piran, D. Han, K. Min, and H. Moon, "A survey on Internet of Things and cloud computing for healthcare," *Electronics*, vol. 8, no. 7, p. 768, 2019, doi: 10.3390/electronics8070768.
- [5] N. Truong, A. D. Nguyen, L. Kuhlmann, M. J. Cook, and A. M. K. Law, "Convolutional neural networks for seizure prediction using intracranial and scalp EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 958–967, 2018, doi: 10.1109/TNSRE.2018.2812160.
- [6] P. Tsiouris, M. G. Pezoulas, D. Z. Koutsouris, and D. I. Fotiadis, "Long short-term memory networks for seizure prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1533–1544, 2018, doi: 10.1109/JBHI.2018.2853730.
- [7] H. Daoud and M. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 2019, doi: 10.1109/TBCAS.2019.2929053.
- [8] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, e1312, 2019, doi: 10.1002/widm.1312.
- [9] D. Gunning *et al.*, "XAI—Explainable artificial intelligence," *Defense Advanced Research Projects Agency (DARPA)*, Tech. Rep., 2019.
- [10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [11] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [12] A. Samek, W. Samek, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing, and interpreting deep learning models," *arXiv preprint*, arXiv:1708.08296, 2017.
- [13] S. Shueb and J. Gutttag, "Application of machine learning to epileptic seizure detection," in *Proc. 27th Int. Conf. Machine Learning*, 2010, pp. 975–982.
- [14] U. R. Acharya *et al.*, "Automated EEG-based screening of epilepsy using wavelet-based entropy and machine learning," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401–409, 2012.
- [15] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.
- [16] Y. Yuan, G. Xun, Q. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 83–94, 2019.
- [17] A. Antoniadis *et al.*, "Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2285–2294, 2017.
- [18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [19] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, 2017, pp. 1321–1330.
- [20] E. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, New York, NY, USA: Basic Books, 2019.
- [21] A. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
- [22] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proc. AAAI Conf. Artificial Intelligence*, New Orleans, LA, USA, 2018, pp. 1527–1535.
- [23] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.
- [24] A. Roy, J. R. H. Wu, and H. W. Deng, "Deep learning-based electroencephalography analysis: A systematic review," *IEEE Access*, vol. 7, pp. 117159–117170, 2019, doi: 10.1109/ACCESS.2019.2936624.