Journal of Artificial Intelligence and Technology, 2026, 6, 337-348
https://doi.org/10.37965/jait.2026.0947

ISTP
RESEARCH ARTICLE

CAAST-Net: Causality-Aware Spiking Transformer with Flask API for EEG-Based Seizure Detection

V. Sonia Devi and R. Pallavi
Department of Computer Science and Engineering, Presidency University, University in Ittagallpura, Karnataka, India

(Received 25 September 2025; Revised 06 January 2026; Accepted 11 February 2026; Published online 09 March 2026)

Abstract: Epileptic seizure detection from electroencephalogram (EEG) signals is an essential task for real-time neurological monitoring. Traditional models face challenges with interpretability, energy efficacy, and capturing temporal causality in neural data. To address these drawbacks, this manuscript proposes a Causality-Aware Attention with Spiking Transformer Network (CAAST-Net). The frequency-domain features are extracted by fast Fourier transform (FFT) to acquire band power across five canonical EEG bands. Subsequently, the features are normalized using a Standard Scalar and transformed into spike trains by rate coding, which enables biologically inspired processing. The CAAST-Net model includes a linear projection layer, leaky integrate-and-fire (LIF) neurons and causality-aware attention module that ensures temporal consistency by facilitating signals from past to present. This model learns discriminative and temporally predictive EEG patterns with minimized computational overhead. The whole model is deployed through a Flask API for real-time seizure detection. The proposed CAAST-Net obtains a higher accuracy of 98.95% with a CHB-MIT dataset and 99.85% accuracy with a BONN dataset when compared with traditional models.

Keywords: causality-aware attention; epileptic seizure detection; fast Fourier transform; leaky integrate-and-fire and spiking transformer

I. INTRODUCTION

Epileptic seizure is a chronic nervous disorder classified through recurrent and uncontrolled seizures at all ages. Seizures lead to unregulated movements and loss of consciousness that severely impact a patient's mental and spiritual health [1–3]. Many patients suffered from various unpredictable symptoms of epilepsy, like depression, memory loss, and different psychiatric disorders [4,5]. Hence, precise and timely detection of epilepsy is essential for patients to provide appropriate medication and to minimize the risk of further epileptic relevance risks [6]. Electroencephalography (EEG) extracts difficult dynamic brain responses and utilizes them for detecting epilepsy types. It helps diagnose epileptic symptoms and supports the assessment of epilepsy recurrence risk [7–10]. Though epilepsy detection is performed by neurology specialists through manual interpretation of EEG files, it consumes more time for analysis [11]. Visual identification of a huge amount of EEG signals is ineffective and time-consuming, placing a huge problem on medical professionals. For improving the quality of life for epilepsy patients and reducing the workload of healthcare professionals [12,13], it is essential to develop a consistent seizure detection model.

Detection of epileptic and non-epileptic from EEG signals is a classification issue, which includes capturing meaningful attributes from EEG signals and categorizing them [14]. Hence, choosing suitable attributes from raw data is crucial [15]. Automatic epileptic detection models are generally divided into two types in accordance with generating signal features such as methods that depend on manual feature selection and models that depend on deep learning (DL) [16]. Recent EEG-based research has employed various models to process EEG signals, capture relevant attributes, and utilize machine learning (ML) or DL models for epileptic detection. The DL models suffer from generalization across patients and require handcrafted features [17,18]. In this manuscript, developed spiking transformers were developed and improved with causality-aware attention that enables much precise seizure detection through learning dynamic spatio-temporal dependencies from raw EEG signals while ensuring biological plausibility and energy efficacy. Among various transformer models like FocalNet, superior performance is achieved; however, spiking transformers enhance interpretability and minimize computational overhead by using spike-based attention and LIF neurons. For facilitating real-time deployment, the Flask API interface is integrated, thereby enabling seamless communication among users and trained model.

A. PROBLEM STATEMENT

The existing algorithms for EEG-based seizure detection struggle with less generalization across patients due to variability in EEG signal patterns and need extensive manual feature extraction. The traditional model cannot extract long-range dependencies and fails to capture dynamic and non-stationary nature of EEG patterns.

B. OBJECTIVE

The primary objective of this manuscript is to develop efficient and accurate seizure detection by combining Causality-Aware Attention with Spiking Transformer Network (CAAST-Net) model. This

Corresponding author: V. Sonia Devi (e-mail: sonia.20223cse0014@presidencyuniversity.in).

model captures fine-grained temporal dependencies in EEG signals while maintaining energy efficacy and interpretability. Moreover, the proposed model is deployed through Flask API to support real-time EEG classification, thereby enabling scalable and accessible seizure detection in healthcare environments.

C. CONTRIBUTION

The major contributions of the research are given below:

- Seizure detection model is developed that integrates CAAST-Net model, thereby facilitating the capture of temporal EEG dynamics with minimized computational complexity.
- Incorporation of leaky integrate-and-fire (LIF) neurons and spike-based attention allows the model to be biologically plausible, less-energy processing.
- The incorporation of causality-aware attention ensures that every EEG time step captures present and past data, which preserves temporal causality and improves interpretability.
- The model is processed by a lightweight Flask-based interface that enables real-time EEG seizure detection from raw signals, thereby supporting scalable and user-friendly clinical integration.

The manuscript is arranged in the following format: Section II analyzes existing models with their advantages and limitations. Section III provides details of the proposed model with the dataset description. Section IV shows results and comparison of the proposed model. Section V concludes a manuscript.

II. LITERATURE REVIEW

In this section, recent EEG-based seizure detection algorithms, focusing on their architectures, feature extraction models, and drawbacks, are described below.

Bingbing Yu *et al.* [19] presented the dual-channel DL model for classifying epileptic EEG signals into three classes, such as normal, ictal, and interictal states. Channel one combined Bidirectional Long Short-Term Memory (BiLSTM) with the Squeeze-and-Excitation (SE) ResNet attention model for adaptively extracting essential feature channels. Channel assigned dual-phase Convolutional Neural Network (CNN) for capturing deep and different attributes. However, the presented approach had limited cross-patient generalization because of variations in EEG patterns across individuals.

Jiahao Qin *et al.* [20] employed the Adaptive Dual-Modality Learning (ADML) method for epilepsy seizure detection by integrating time-series imaging into a transformer model. The employed model effectively extracted temporal dependencies and spatial relationship within EEG signals by an attention mechanism. The employed model determined strong generalization ability across datasets while maintaining computational efficacy. The employed model had a high computational cost and energy inefficiency of conventional DL models, thereby making it unsuitable for real-time clinical applications.

Bommala Silpa and Malaya Kumar Hota [21] developed a dual-stream DL model for extracting deep features through scalograms and time-series EEG signals. Initially, CNN captured spatial dimensional features from scalogram images, and SE algorithm improved relevant informative features by adjusting channel weights. Gated Recurrent Unit (GRU) was used to capture temporal features from time-series EEG signals, and a Confined Attention (CA) mechanism was incorporated to assign greater weights for essential features. Subsequently, extracted features were fused with deep features for the precise detection of seizures by the Support Vector Machine (SVM) classifier. For enhancing seizure detection rate, regression at the end of Variational Model Encryption (VME) algorithms were used in the preprocessing phase. Additionally, the performance of the developed dual-stream lightweight seizure network was named as DSLWNet. The employed model was unable to extract long-range temporal dependencies on EEG signals.

Zhentao Huang *et al.* [22] used the Spatio-Temporal Feature Fusion epilepsy EEG recognition framework with the Dual Attention (STFFDA) method as an EEG recognition framework. STFFDA was included in the multi-channel model, which directly understood epileptic states from actual EEG signals, thereby removing the requirement for extending data preprocessing and feature extraction. The model had a dependence on manual and domain-specific feature engineering, which limited its scalability and adaptability in seizure detection.

Zongpeng Zhang *et al.* [23] developed two aspects to solve the cross-patient generalization issue in DL models. Initially, they developed a data augmentation model to highly enhance generalization. They analyzed the statistical distribution of seizure EEG signals and developed Spatio-Temporal EEG Augmentation (STEA) for producing synthetic training seizure information to spatio-temporal dependency from EEG. Next, a Patient-Adversarial Neural Network (PANN) for learning patient invariance and processing adversarial optimization among feature extractor, identity discriminator, and seizure features was maintained. The developed model lacked a real-time model that processed and classified EEG signals dynamically without retaining model performance.

Ali Mehrabi *et al.* [24] introduced a hybrid spike-encoded spiking neural network for real-time EEG seizure detection. The model used a convolution-enhanced ConvSNN with depthwise-seperable convolutions and temporal self-attention. However, the usage of self-attention for the ConvSNN model limited the adaptability in modeling long-range temporal dependencies in EEG signals.

The literature review suggested that existing DL models for EEG-based seizure detection, such as Bi-LSTM, CNN, GRU, and different transformer-based models, face drawbacks in generalization, temporal modeling, and real-time applicability. While models like ADML and DSLWNet provide improvements by multimodal learning, dual-stream models still depend on handcrafted features. Moreover, traditional attention mechanism fail to capture temporal causality, and existing models are computationally expensive.

III. PROPOSED METHOD

In this manuscript, the CAAST-Net model is proposed for precise and effective EEG-based seizure detection. The proposed integrated CAAST-Net model aims to extract long-range temporal dependencies while maintaining biological plausibility and energy efficiency. It uses fast Fourier transform (FFT) to capture frequency-domain features by temporal encoding using LIF neurons. For enabling real-time deployment, the trained model is deployed in a Flask API interface that supports EEG input and dynamic seizure detection. Fig. 1 presents an overall process of epileptic seizure detection with the CHB-MIT and BONN dataset.

A. DATASET

This section presents EEG data sources, which are utilized for training and evaluation that include CHB-MIT and BONN

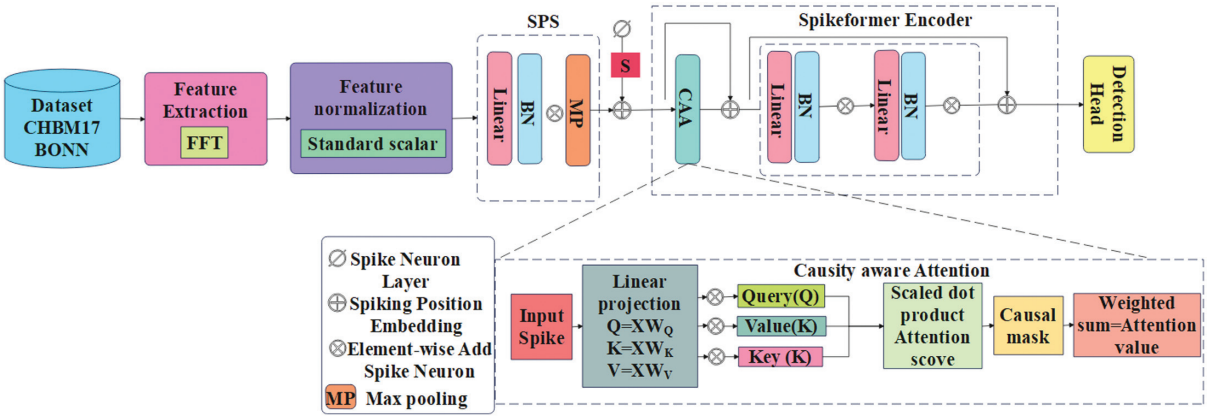

Fig. 1. Overall process of epileptic seizure detection with CHB-MIT and BONN dataset.

datasets. These datasets provide different seizure and non-seizure EEG recordings, thereby enabling robust performance evaluation of the proposed CAAST-Net model.

B. CHB-MIT DATASET

This dataset is created by the Massachusetts Institute of Technology (MIT) and Boston Children's Hospital (CHB) [25]. This includes scalp EEG records from several patients. The dataset employs a bipolar lead model dependent on 10 to 20 devices, extracting EEG signals from 22 electrodes at a sampling rate of 256 Hz and a precision of 16 bits. The dataset contains 23 EEG signal channels for certain causes, including 18 channels. Data from CHB01 to CHB21 are gathered from similar patients with a gap of 1.5 years. Every case approximately has 9 to 42 continuous EEG records, many of which include 1 hour of EEG records. EEG records contain 182 seizure records, each mapped with initial and end time. Fig. 2 presents the class distribution of the CHB-MIT dataset.

1). BONN DATASET. This dataset includes EEG records gathered from University of Bonn consisting of five classes characterized *Z*, *O*, *N*, *F*, and *S* by 100 single-channel EEG segments [26]. Every segment includes duration of 23.6 s with sampling rate of 173.61 Hz, resulting in 4097 data points per segment. Sets *Z* and *O* inclues EEG files from five healthy volunteers on the awake state

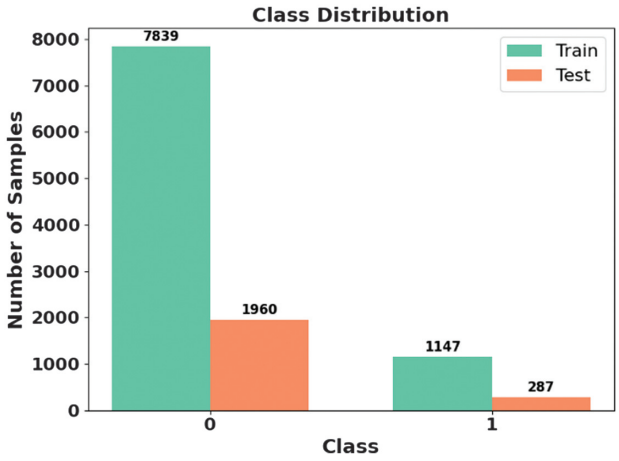

Fig. 2. Class distribution of CHB-MIT dataset.

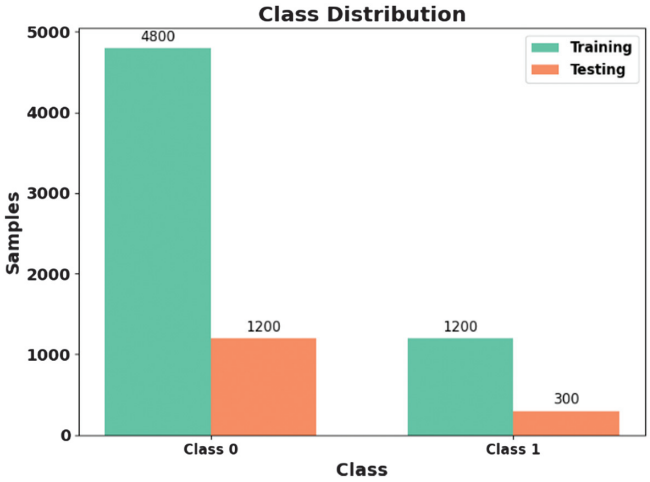

Fig. 3. Class distribution of BONN dataset.

with eyes open and closed. Sets *N* and *F* consist of intracranial records from five patients in seizure-free intervals, with set *N* obtained from hippocampal development and set *F* from epileptogenic zone. Set *S* includes seizure activity records from the whole records site, thereby providing ictal activity. EEG signals are recorded by 128-channel amplifier system to general mean reference. Fig. 3 presents class distribution of BONN dataset.

C. FEATURE EXTRACTION

FFT is a crucial method utilized to convert time-domain signals to frequency domain, which enables the extraction of meaningful features. Seizures generally have abnormal rhythmic discharges or energy spikes in particular frequency bands, which are efficiently extracted by analyzing the power spectral distribution of EEG segments using FFT. The process initializes with segmenting the EEG into short windows, while employing FFT on every segment and calculating the magnitude or power spectrum. Here, the EEG signals were segmented into a fixed-length windows of 2 s with 50% overlap to preserve temporal continuity. The FFT was applied to each window with the help of Hamming window, and the power spectral density was then computed. The band power features were obtained by integrating spectral energy within delta 0.5–4 Hz, theta (4–8 Hz), beta (13–30 Hz), and gamma (< 30 Hz) bands for each

Table I. Features that are extracted from the FFT model with frequency range and description

Features	Frequency range (Hz)	Description
Delta power	0.5–4	Deep sleep, high during seizures
Theta power	4–8	Drowsiness, raised in seizure onset
Alpha power	8–13	Relaxation, suppressed in seizures
Beta power	13–30	Alertness, maximizes in seizures
Gamma power	> 30	Cognitive spikes, high in ictal state

channel. From this, frequency-domain features like band-specific power, peak frequency, and spectral entropy are captured. These features minimize high-dimensional EEG into informative representations that are fed as input to the classifier. FFT is computationally effective $(O(N \log N)$, which facilitates better generalization across patients. The extracted features from the FFT are described in Table I. This feature vector (10–15 features per channel) is fed to a spiking transformer for seizure detection.

D. STANDARD SCALAR USING Z-SCORE NORMALIZATION

After feature extraction, Standard Scalar is employed to normalize the extracted EEG features before feeding them into the spike encoder. Standard Scalar transforms attributes by eliminating mean and scaling to unit variance, which ensures that every feature contributes equally to learning process. This is essential in EEG-based seizure detection, as the features captured from FFT vary in scale across frequency bands and patient recordings. Without normalization, the model tends to be biased toward high-magnitude features, thereby minimizing its generalization ability. Using Standard Scalar ensures the consistency between training and inference, thereby improving training stability and cross-patient robustness. Moreover, to avoid data leakage, the Standard Scalar was fitted on the training set and then applied to the validation and test sets using same learned parameters.

E. EPILEPTIC SEIZURE DETECTION

This section presents the proposed CAAST-Net model that integrates causality-aware attention with a spiking transformer for effectively capturing temporal dependencies in EEG signals. This model ensures that every time step attends to the relevance of past data while using spike-based calculation for energy-efficient and interpretable seizure detection.

2). SPIKING TRANSFORMER. The spiking transformer is inspired by BERT, using a similar tokenizer and embeddings, and then converts embedding vectors into spike trains by a spike encoding model. All transformer modules are redeveloped for SNN calculation, using spike train as a hidden state. After a transformer layer, task-specific detection head is employed to produce the results.

3). NEURAL MODEL. In SNNs, the primary unit, a LIF neuron module, is integrated into a spiking transformer. LIF module is suited for developing large-scale SNN models because of its simplicity and high efficacy. Postsynaptic neuron receives spike trains generated by neurons and gathers membrane potential that omits spike once the potential extends firing threshold through a membrane reset. Mathematical expression for dynamic of LIF model is given in Equations (1) and (2):

$$U_l^t = TU_l^{t-1}(1 - S_l^{t-1}) + WS_{l-1}^t \tag{1}$$

$$S_l^t = F(U_l^t - u_{th}) \tag{2}$$

In Equations (1) and (2), U_l^t is neurons' membrane potential in $l-th$ layer in time step t, S_l^t is the result spikes, W denotes the weights, u_{th} represents the firing threshold, and $F(u)$ represents the Heaviside stage process that provides 0, when $u < 0$ or else 1 otherwise. Since F is non-differentiable, surrogate gradients are determined, and their mathematical expression is given in Equation (3):

$$\frac{\partial F}{\partial u} = \exp(-|2u|) \tag{3}$$

Here, learnable membrane time constant τ is employed to ensure neuronal diversity, while maintaining constant firing threshold u_{th} to overcome potential variability in training.

4). SPIKE ENCODING. Rate coding is majorly utilized to encode image inputs in earlier SNN tasks, where pixel intensities are transformed into spike trains proportional to its values. Poisson coding is utilized for converting word embeddings into spike trains. Although rate coding is unable to use temporal data and has limited ability for word representation, it was selected due to its robustness to noise and computational simplicity, making it suitable for frequency-domain EEG features. Unlike precise temporal encoding, rate coding provides stable spike statistics across short windows. This mechanism improves training stability and real-time inference in seizure detection tasks. Here, a spike encoding model is developed, where at every time step a linear transformation is applied, followed by a firing function; its mathematical expression is given in Equation (4):

$$S^t = F(XW^t + b^t) \tag{4}$$

In Equation (4), $X \in R^{s\times d}$ represents actual embedding vector, $S^t \in \{0,1\}^{s\times d}$ represents produced spike in time step t, W^t and b^t represent learning parameters, s is the input sequence length, and d is model's hidden size. By assigning different parameters in various time steps, temporal dimension representation ability of spike trains is improved. Here, count of time steps is kept small, making the resulting parameter count manageable. In training, simple constraint is employed for controlling a firing rate of encoding and its mathematical expression is given in Equation (5):

$$L_{fr} = \left| \frac{1}{T} \sum_t^T S^t - fr \right| \tag{5}$$

In Equation (5), fr represents desired firing rate. The training stability in the spiking transformer ensures through the use of smooth surrogate gradients for a non-differentiable firing function. This prevents vanishing gradient during backpropagation. Rate-based spike encoding was selected because of its robustness and stable convergence when compared to temporally precise encodings that are more sensitive to noise and requires longer simulation horizons. This design choice balances training stability and computational efficiency while maintaining competitive performance.

5). SPIKE-BASED ATTENTION. The actual self-attention module utilizes Query (Q) and Key (K) to compute attention maps, which require floating-point matrix multiplication and a softmax function.

However, these processes are not completely well suited to the computational process of SNNs and modify computation of an attention map. By using spike-form query and key, floating-point matrix multiplication is replaced with a logical AND process along with a mask acceleration process. Consider a spike-form input $X \in \{0,1\}^{s\times d}$. The mathematical expression for the self-attention mechanism is given in Equations (6) and (7):

$$Q = LIF(XW^Q),\ K = LIF(XW^K),\ V = XW^V \tag{6}$$

$$Attention = QK^TV/d \tag{7}$$

In Equations (6) and (7), W^Q,W^K, and W^V represent learnable parameters. To scale an attention map score between 0 and 1, it is divided by d. The order of calculation among Q, K, and V can be changed when using the softmax function. The binary dot product process is defined based on neuronal behavior, where an attention neuron fires a spike only when it receives simultaneous stimuli from a query and key neuron, equivalent to a logical AND process. Similarly, the mask accumulation process is defined by disinhibition neuron behavior, where primary and query neurons prevent signal broadcasting between input and output neurons. When spikes are fired, disinhibition is activated, which allows the input neuron to send signals to the output neuron. Generally, spiking attention is extended to $h-$ heads by separating $Q, K \in \{0,1\}^{s\times d}$,$V \in R^{N\times d}$ to $(q_1,\ldots,q_h)$,$(k_1,\ldots,k_h)$, $(v_1,\ldots,v_h)$, where q_i, $k_i \in \{0,1\}^{s\times d/h}$, $v_i \in R^{s\times d/h}$. After computing each attention head, the results are concatenated, and the mathematical expression is given in Equation (8):

$$MultiHeadAttention = LIF\left(\frac{Concat(q_1k_1^Tv_1,\ldots,q_kk_h^Tv_h}{d/h}\right)W^O \tag{8}$$

6). RESIDUAL CONNECTION. Transformers utilize a layer normalization for handling constancy on residual connections, but there is no equal process on SNNs. For addressing this, residual connection in transformer block is introduced, and the neurons of LIF are utilized to receive stimuli from direct input and substitute layer outputs. Its mathematical expression is given in Equation (9):

$$X' = LIF(X + \alpha Sublayer(X)) \tag{9}$$

In Equation (9), X, $X' \in \{0,1\}^{s\times d}$ represents spike form, when the outcomes of the sublayer act as present stimuli. When the firing threshold $u_{th} \le 1$, an identity mapping from X to X' is obtained. Surrogate gradients are utilized for LIF neurons to ensure that gradients are not fully shortened in residual connections. Here, the outcomes of sublayer are rescaled through a little factor α, allowing the network to train on conjunction to parameter initialization algorithm implemented later, thereby ensuring the model to process without normalization.

7). DETECTION HEAD. The normalized FFT feature vectors serve as direct inputs to the spike encoder, where each feature dimension is linearly projected and converted into the spike trains across multiple time steps. These spike sequences constitute the input tokens to the spiking transformer, which enables the temporal modeling of frequency-domain EEG dynamics. The detection head is a fully connected layer to LIF neurons that produces result spike trains. For the classification process with N classes, N output neurons are used, and its firing rates are measured as logit outputs of method and its mathematical expression is given in Equation (10):

$$z_i = \frac{1}{T}\sum_t^T s_i^t \tag{10}$$

In Equation (10), s_i^t represents the output spike of ith neuron at time step t. However, producing continuous value is not direct in SNNs. Instead, firing rates are used as the output, which leads to limited numerical representation and potential training variability. Therefore, an introduced model for regression tasks initially separates the continuous range of target scores into N discrete labels $\{L_1,\ldots,L_N\}$. Next, N output neurons are utilized, each predicting the probability of its respective label. Finally, a probability-weighted average term is measured, and its mathematical expression is given in Equation (11):

$$y = \frac{\sum_i^N \exp(kz_i)\cdot L_i}{\sum_i^N \exp(kz_i)} \tag{11}$$

In Equation (11), k is a scale factor that handles the smoothness of a softmax output.

8). CAUSALITY-AWARE ATTENTION. Causality-aware attention is a temporal attention mechanism that applies causal constraints in sequence modeling, every time step attends to itself and preceding time steps. This makes it suitable for EEG seizure detection, where the model is unable to access future information during inference time, which ensure that this is essential for maintaining interpretability and preventing data leakage. By enforcing the temporal causality, the attention mechanism prevents access to future EEG information during inference by improving interpretability while preventing information leakage to ensure suitability for real-time seizure detection.

• Input Spike Sequences. The spiking transformer receives spike-encoded representations of EEG features over T time steps: Input Shape: $[Batch, T, D]$, where T represents the number of time steps and D is the feature dimension. The $(Query\,Q, Key\,K, Value\,V)$ are computed from the spike sequence using liner layers $Q = X\cdot W_Q$, $K = X\cdot W_K$, $V = X\cdot W_V$, where $X \in R^{\{T\times D\}}$ represents the spiking sequence and W_Q,W_K,W_V are the learned weights.

Causal mask is employed for ensuring that every time step t attends to time step 0 by incorporating t. This develops less triangular mask M, and its mathematical expression is given in Equation (12):

$$M_{\{ij\}} = \begin{cases} 1 & if\ j \le i \\ \text{else} & -\infty \end{cases} \tag{12}$$

Mathematical expression for computing attention is given in Equation (13):

$$Attention\,scores = (Q\cdot K^T)/\sqrt{d} + M \tag{13}$$

Next, softmax is applied, and its mathematical expression is given in Equation (14):

$$A = softmax(Attention\,Scores) \tag{14}$$

The output of every time step is a weighted sum of past values, and the mathematical expression for this is given in Equation (15):

$$Output_t = \sum_j (A_{\{tj\}}\cdot V_j)\ \ for\ j \le t \tag{15}$$

The output is a time-causally consistent sequence with the similar shape $[Batch, T, D]$ and is transformed into a representation

at every time step, now encoding weighted contextual information. After the initial LIF layer produces sparse spike-encoded features at every time step, these features are processed by the attention mechanism. It allows a model to learn spiking signals from the earliest time points, thereby refining the understanding of temporal dependencies relevant to seizures and cognitive states. The resulting features are then pooled and fed to the last LIF layer and linear classifier.

9). FLASK API INTERFACE. Flask API acts as a deployment interface for the trained seizure detection method and enables real-time predictions on raw EEG files without retraining the model every time. API summarizes preprocessing, feature extraction, spike encoding, model interface, and formatting of output. Flask API integration enables a real-time interface of EEG-based seizure detection using a deployed spiking transformer model. Upon receiving raw EEG files through a POST request, Flask uses MNE to read and segment signals into overlapping windows. The preprocessing phases involve bipolar montage selection, microvolt μv scaling, and extraction of bandpower features through the FFT algorithm. These features are normalized by a prefitted standard scaler and encoded into spike trains by rate-based encoding. Spike sequences are passed to a spiking transformer that includes linear projection, LIF neurons, and a causality-aware multihead attention mechanism to extract temporal dynamics. The output layer generates class logits representing seizure or non-seizure. Flask shows the whole process and returns a JSON response with the prediction, thereby making the model lightweight, RESTful inference.

Algorithm 1. Process of CAAST-Net with Flask API to epileptic seizure detection

Input – CHB-MIT and BONN datasets
Output – Detected epileptic and non-epileptic
Feature extraction using FFT
    Fft_result = FFT (segment)
    Features = [*delta_power*, $theta_{power}$, $alpha_{power}$, $beta_{power}$, *gamma_power*, *spectral entropy and peak frequency*]
    Return features
Feature Normalization
    Function Normalize_Features (features):
    Normalized_features = Zscore_normalization(features)
    Return normalized_features
Spike Encoding
    Function Spike_Encode (features):
        For each timestep *t*:
            Linear_proj = Linear Transform (features[t])
            Spike [t] = Firing function (linear_proj, threshold)
        Return spike_sequence
Spiking Transformer Inference
Function Spiking_transformer (spike-sequence)
    For every transformer layer:
        Q, K, V = Generate spike attention inputs (spike_sequence)
        Causal_mask = Crete Causal mask (Q.shape)
        Attention_output = Spike Attention (Q, K, V, mask=causal_mask)
        Spike_sequence = LIF neuron (attention_output)
    Output spikes = Detection Head (spike_sequence)
    Output probabilities = Softmax (FiringRates (output_spikes))
    Return output_probabilities
Detection
    Predicted_label = ArgMax (probabilities)
    Return predicted_label
Flask API
    Input: raw_EEG_file
    Preprocessed = Preprocess_EEG (EEg_data)
    Features = Extract_FFT_features (features)
    Normalized = Normalize_features (features)
    Spikes = Spike_encode (normalized)
    Probabilites = Spiking_Transformer (spikes)
    Prediction = Classify_EEG (probabilities)
    Return JSON_Response (prediction)

IV. EXPERIMENTAL RESULTS

The proposed CAAST-Net is simulated with Python 3.7 environment, and the required configurations are 12th Gen Intel (R) Core (TM) i7-12700K (3.60 GHz), 64 GB RAM, 64-bit OS, and Windows 11 Pro with version 24H2. The metrics like accuracy, precision, recall, and F1-score are considered to validate the performance of a proposed CAAST-Net method.

Table II presents a training configuration utilized for the proposed CAAST-Net model. The model is trained by the Adam optimizer with a learning rate of 0.0001, which ensures stable and effective convergence. The batch size of 64 and training epochs of 100 are utilized to allow sufficient learning across the data. The loss function is a sparse categorical cross-entropy, which is utilized for multi-class classification with integer labels. A softmax activation function is used on the output layer to generate class probabilities. The train-test split of 80:20 is applied to maintain a balanced and generalized evaluation setup.

Table II. Parameters and training configuration used for CAAST-Net model

Parameter	Value
Loss function	Sparse_categorical_crossentropy
Epoch	100
Batch size	64
Learning rate	0.0001
Optimizer	Adam
Activation function	Softmax
Train-test ratio	80:20

Table III. Class-wise performance of the proposed CAAST-Net using CHB-MIT dataset

Classes	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Class 0	98.94	98.84	98.99	98.92
Class 1	98.95	98.85	98.98	98.91
Average	98.95	98.85	98.99	98.92

Table IV. Class-wise performance of the proposed CAAST-Net using BONN dataset

Classes	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Class 0	99.85	99.73	99.87	99.81
Class 1	99.84	99.74	99.88	99.80
Average	99.85	99.74	99.88	99.81

Tables III and IV show the class-wise performance of the proposed CAAST-Net method using the CHB-MIT and BONN datasets to classify epileptic seizures. In the CHB-MIT dataset, the proposed CAAST-Net achieved an accuracy of 98.95%, and in the BONN dataset, it achieved 99.85% accuracy, which shows the model generalization across datasets.

Table V presents the performance of a proposed CAAST-Net using the CHB-MIT and BONN dataset with traditional transformer models. In the CHB-MIT dataset, the proposed CAAST-Net acquired an accuracy of 98.95%, and in the BONN dataset, it obtained 99.85% accuracy, while comparing with other models. This enhancement is because of their integration of causality-aware attention and spiking transformer, which captures temporal dependencies in EEG signals while maintaining high performance and interpretability. These outcomes show the model's robustness and suitability for epileptic seizure detection.

Table VI represents k-fold cross-validation performance of the proposed CAAST-net model by CHB-MIT and BONN datasets with 2-, 3-, and 5-fold. In $K = 5$, the proposed model consistently outperforms other models, thereby showing high generalization and stability. In the CHB-MIT dataset, the proposed CAAST-Net acquired an accuracy of 98.95%, and in the BONN dataset, it obtained 99.85% accuracy, when $K = 5$. These results validate the effectiveness and scalability of a proposed CAAST-Net on various EEG scenarios.

Tables VII and VIII show cross-dataset validation of the proposed CAAST-Net for generalization ability when compared to existing transformers. When a method is trained on CHB-MIT dataset and tested on BONN dataset, the proposed model obtained 97.55% accuracy while outperforming ViT and FocalNet. When a method is trained on BONN dataset and tested on CHB-MIT dataset, it obtains an accuracy of 97.88%, while showing high robustness across different data distributions. The result presented in tables shows the model's effectiveness in cross-patient and cross-dataset seizure detection.

Table IX presents the statistical analysis of the model performance using p-values, confidence interval (CI), and accuracy of CHB-MIT and BONN dataset. The proposed model obtained p-value of < 0.05, representing that the difference in performance is statistically significant. The proposed CAAST-Net obtains a high accuracy of 98.95% on CHB-MIT dataset and 99.85% on BONN dataset to narrow confidence intervals, while representing consistent and reliable performance. Compared to other transformer

Table V. Performance of proposed CAAST-Net using CHB-MIT and BONN datasets

Models	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ViT	CHB-MIT	95.98	95.87	95.74	95.68
	BONN	95.74	95.84	95.64	95.74
Swin transformer	CHB-MIT	96.74	96.54	96.87	96.35
	BONN	96.87	96.68	96.57	96.55
FocalNet	CHB-MIT	97.24	97.52	97.65	97.58
	BONN	97.64	97.58	97.65	97.74
Proposed CAAST-Net model	CHB-MIT	98.95	98.85	98.99	98.92
	BONN	99.85	99.74	99.88	99.81

Table VI. K-fold cross-validation of a proposed CAAST-Net model using CHB-MIT dataset and BONN dataset

Dataset	*K*-fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CHB-MIT	2	97.45	97.55	97.65	97.74
	3	97.89	97.81	97.64	97.54
	5	98.95	98.85	98.99	98.92
BONN	2	96.89	96.87	96.88	96.54
	3	96.82	96.77	96.57	96.66
	5	99.85	99.74	99.88	99.81

Table VII. Cross-dataset validation with training on CHB-MIT dataset and testing on BONN dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ViT	97.22	97.14	97.35	97.45
Swin transformer	97.65	97.85	97.68	97.98
FocalNet	97.54	97.24	97.28	97.42
Proposed CAAST-Net model	97.55	97.65	97.45	97.44

Table VIII. Cross-dataset validation with training on BONN dataset and testing on CHB-MIT dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ViT	97.75	97.58	97.84	97.68
Swin transformer	97.84	97.54	97.65	97.84
FocalNet	97.54	97.84	97.92	97.95
Proposed CAAST-Net model	97.88	97.68	97.84	97.98

models, the proposed model shows superior generalization and minimized variability. The proposed model obtains high performance and maintains statistically robust and stable detection.

Table X presents the computational analysis of a proposed CAAST-Net model by CHB-MIT and BONN datasets with the metrics of training time, memory usage, inference time, FLOPs, and model parameters. The proposed CAAST-Net shows that superior efficacy obtains less FLOPs and has a small parameter count. Moreover, CAAST-Net maintains inference time and minimizes the training time on both datasets while showing its speed and adaptability. When its memory usage is high, the model uses spike-based temporal encoding and attention mechanism, which improves accuracy and temporal awareness. The proposed CAAST-Net shows better balance between performance and efficacy, which makes it suitable for scalable EEG seizure detection. To quantitatively assess interpretability, the attention weight

Table IX. Statistical analysis of a proposed CAAST-Net model by CHB-MIT dataset and BONN dataset

Methods	Dataset	*p*-Value from t-test	Confidence interval (CI)	CI (±)	Accuracy (%)
ViT	CHB-MIT	0.0311	[95.95–95.79]	± 0.08	95.87
	BONN	0.0275	[96.06–95.92]	± 0.07	95.99
Swin transformer	CHB-MIT	0.0413	[96.33–96.15]	± 0.09	96.24
	BONN	0.0317	[96.62–96.46]	± 0.08	96.54
FocalNet	CHB-MIT	0.0241	[97.30–97.18]	± 0.06	97.24
	BONN	0.0342	[97.54–97.36]	± 0.09	97.45
Proposed CAAST-Net model	CHB-MIT	0.0351	[98.55–98.39]	± 0.08	98.95
	BONN	0.0411	[98.18–98.06]	± 0.06	99.85

Table X. Computational analysis of a proposed CAAST-Net model by CHB-MIT dataset and BONN dataset

Methods	Dataset	Training time (sec)	Memory usage (MB)	Inference time (sec)	Flops (G)	Parameter (M)
ViT	CHB-MIT	1021	2785.28	0.010	17.5	86
	BONN	610	1804	0.014		
Swin transformer	CHB-MIT	1043	2865.14	0.008	4.5	29
	BONN	630	1850	0.007		
FocalNet	CHB-MIT	1054	2687.14	0.009	5.5	30
	BONN	640	1824	0.008		
Proposed CAAST-Net model	CHB-MIT	912	46,873.6	0.008	2	8.98
	BONN	112	2,785.28	0.009		

distributions across time steps were analyzed. The causality-aware attention consistently assigns higher weights to earlier EEG segments preceding seizure onset while highlighting clinically relevant temporal patterns. This behavior confirms that the model focuses on meaningful historical EEG rather than spurious future context by supporting interpretability through transparent temporal attribution.

A. QUALITATIVE ANALYSIS USING CHB-MIT DATASET

The figures present the performance of a proposed CAAST-Net model on CHB-MIT dataset. Fig. 4 shows accuracy vs epochs, while Fig. 5 represents loss vs epochs. Both figures show a consistent increase in training and validation accuracy with a decrease in loss, which represents stable convergence without overfitting. The confusion matrix presented in Fig. 6 shows a high classification rate with less misclassifications, which determines the model's high discriminative ability. The ROC curve presented in Fig. 7 shows effective sensitivity and specificity. These results validate the reliability and robustness of CAAST-Net for precise epileptic seizure detection.

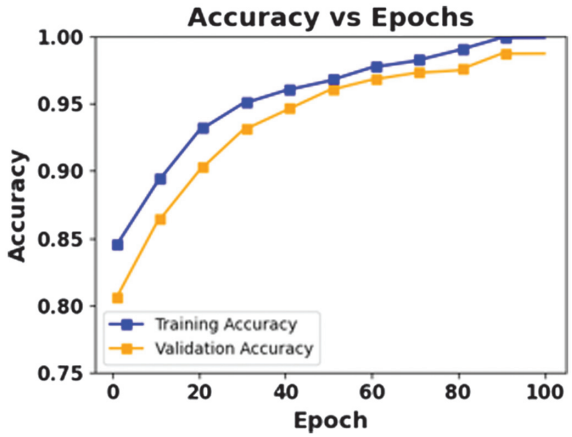

Fig. 4. Accuracy vs Epochs for CHB-MIT.

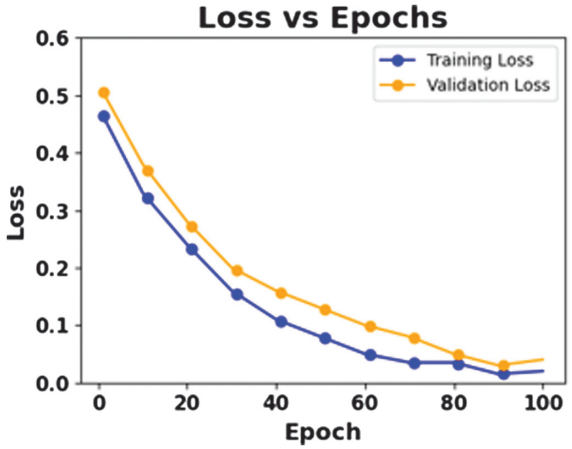

Fig. 5. Loss vs Epochs for CHB-MIT.

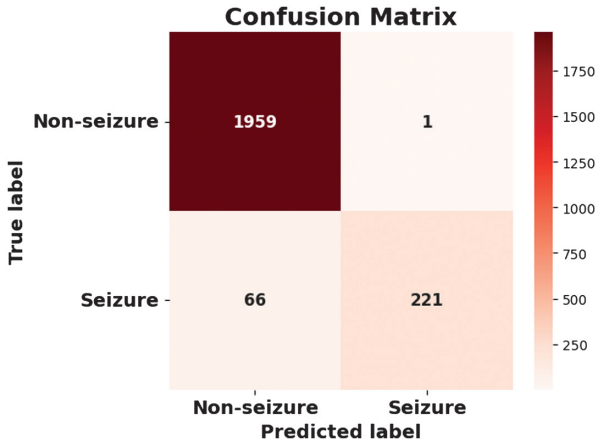

Fig. 6. Confusion matrix for CHB-MIT.

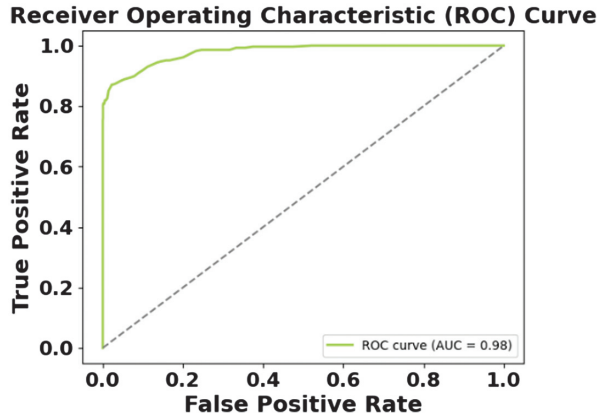

Fig. 7. ROC curve for CHB-MIT.

B. QUALITATIVE ANALYSIS USING THE BONN DATASET

The figures 8–11 show the effectiveness of a proposed CAAST-Net model in BONN dataset. Accuracy graph presented in Fig. 8 and loss graph presented in Fig. 9 across 100 epochs show consistent

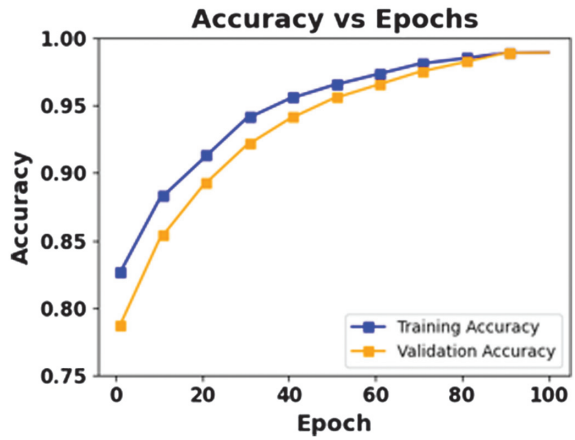

Fig. 8. Accuracy vs Epochs for BONN.

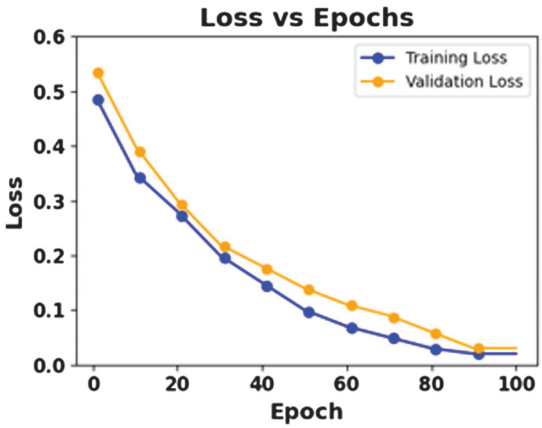

Fig. 9. Loss vs Epochs for BONN.

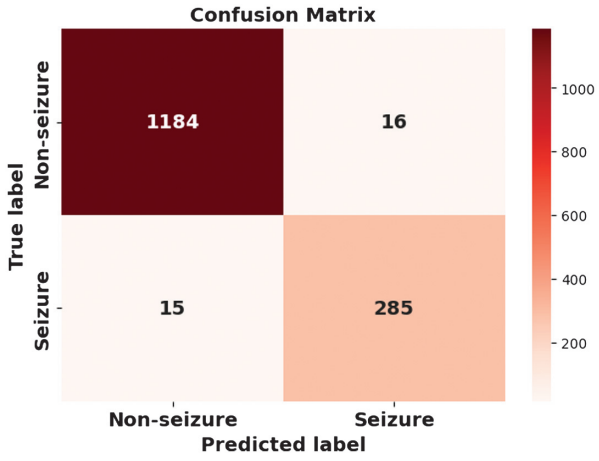

Fig. 10. Confusion matrix for BONN.

improvements in training and validation, thereby representing strong learning stability and minimizing the risk of overfitting. The confusion matrix presented in Fig. 10 shows high classification accuracy, with fewer misclassified samples. At last, the ROC curve in Fig. 11 represents the model's discriminative ability among seizure and non-seizure classes while representing its robust generalization and efficient temporal modeling.

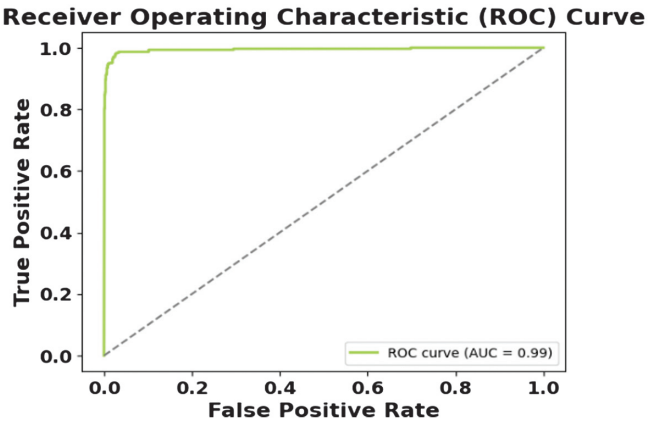

Fig. 11. ROC curve for BONN.

C. COMPARATIVE ANALYSIS

Table XI presents the comparative analysis of a proposed CAAST-Net method against existing models like BiLSTM with AE [19], ADML [20], DSLWNet [21], and STFFDA [22] using CHB-MIT dataset and BONN dataset. The proposed CAAST-Net obtains high accuracy of 98.95% on CHB-MIT dataset and 99.85% on BONN dataset. The proposed CAAST-Net integrates causality-aware attention with a spiking transformer, enabling it to extract fine-grained temporal dependencies in EEG signals while maintaining computational efficacy. Moreover, it provides better generalization and cross-patient robustness, as well as superiority over existing models. Although, this comparative analysis included a ConvSNN [24] model for detecting seizures using the CHB-MIT dataset. But this model did not use any attention mechanism, which limited the capability in modeling long-range temporal dependencies in EEG signals. The performance improvement observed in the proposed CAAST-Net is primarily attributed to the integration of causality-aware attention within the spiking transformer framework. This enhances temporal modeling beyond conventional SNN architectures.

D. RESEARCH IMPLICATIONS

The proposed model integrates causality-aware attention with a spiking transformer framework for EEG-based seizure detection and deploys it in a real-time Flask API, thereby providing essential research implications.

Table XI. Comparative analysis of a proposed CAAST-Net model against existing algorithms

Methods	Dataset	Accuracy (%)	Precision (%)	Recall/sensitivity (%)	F1-score (%)
BiLSTM with AE [19]	CHB-MIT	98.65	97.06	NA	NA
ADML [20]	CHB-MIT	98.7	NA	98.3	NA
	BONN	99.7	NA	98.9	NA
DSLWNet [21]	CHB-MIT	98.67	98.45	99.12	98.66
	BONN	99.5	99.61	99.77	99.69
STFFDA [22]	CHB-MIT	95.18	95.16	95.16	95.16
	BONN	77.65	77.52	77.67	77.47
ConvSNN [24]	CHB-MIT	94.70	NA	NA	89.30
Proposed CAAST-Net model	CHB-MIT	98.95	98.85	98.99	98.92
	BONN	99.85	99.74	99.88	99.81

- By using SNNs and LIF neurons, this manuscript determines that biologically inspired models facilitate less energy, event-driven processing suitable for embedded medical devices.
- The incorporation of causality-aware attention provides a novel mechanism to extract temporal and spatial EEG patterns, which are essential for seizures. This has broad implications for explainable neuroscience models in understanding brain-state transitions.
- The integration of FFT-based feature extraction, spike generation, and transformer provides a hybrid model that imitates the neurophysiological data process.
- The incorporation of a Flask-based API acts as a research prototype for clinically deployable seizure detection. This represents the path from model deployment to real-time inference, thereby highlighting reproducibility and accessibility.

E. DISCUSSION

The proposed CAAST-Net model integrates causality-aware attention with a spiking transformer, which determines significant enhancements over existing EEG-based seizure detection models by accuracy, generalization, computational efficacy, and real-time applicability. The experimental results across datasets like CHB-MIT and BONN show that the model consistently obtains high performance. The primary advantage of CAAST-Net lies in its spike-based model. By using LIF neurons and a biologically plausible spike encoding mechanism, the model imitates the temporal firing dynamics of real neurons. This process ensures energy-efficient computation, which is essential for real-time clinical applications, and improves interpretability by positioning to human brain functioning. The enhancement of the causality-aware attention mechanism that provides every EEG time step allows for its present and previous values. This preserves the temporal integrity of seizure patterns, prevents data leakage, and improves the robustness of the model in inference-time scenarios. Compared to a conventional attention mechanism that relies on future context, this model is more suitable for medical signal processing. Moreover, the design of CAAST-Net based on LIF neurons and sparse spike communication reduces redundant computations and supports low-latency inference. This mechanism makes the proposed framework suitable for deployment in real-time clinical environments like demonstrated through the flask-based inference API. The proposed model is evaluated in cross-dataset and k-fold validations while showing its generalization ability across various patient populations and data distributions. The statistical analysis shows fewer p-values and narrower confidence intervals, which show the model's robustness and consistent performance. Computational analysis shows that the proposed model efficiently balances performance and efficiency. Memory usage is high, but spike-based multi-head attention and temporal encoding layer maintain less inference time, minimize FLOPs, and have fewer parameters compared to traditional transformers. Moreover, the reported memory usage for the CHB-MIT dataset reflects peak host-side memory consumption during training, which includes buffered EEG recordings. This spike tensor unrolls across time steps and intermediate attention states. During the inference, the model operates with a significantly reduced memory footprint due to sparse spike activations, fixed temporal windows, and the absence of gradient storage. Consequently, the runtime memory requirements are substantially lower than training time measurements that support deployment on resource-constrained devices. Finally, the integration of Flask API for real-time deployment is an essential phase for model development and clinical application by enabling a RESTful interface on raw EEG data by facilitating accessible and user-friendly seizure detection in healthcare.

V. CONCLUSION

This manuscript developed the CAAST-Net model for EEG seizure detection, which integrated frequency-based feature extraction with spiking neural dynamics and causality-aware attention. By extracting spectral band power features using FFT and translating that into temporally encoded spike trains, the model obtained biologically inspired efficacy. The incorporation of causality-aware attention ensured that temporal dependencies were learned in a forward-consistent manner while improving interpretability and detection stability. The LIF spiking layer minimized computational cost and energy consumption. The proposed CAAST-Net obtained a high accuracy of 98.95% with CHB-MIT dataset and 99.85% accuracy with BONN dataset when compared to existing and state-of-the-art methods. The experimental validation showed that the model obtained high accuracy and sensitivity while maintaining less latency and computational overhead.

A. FUTURE WORK

As future work, we plan to integrate adaptive learning to personalize seizure detection per patient and extend the model to multi-modal signals (ECG and EMG) to improve robustness. Moreover, we can refine causality-aware attention by graph-based temporal reasoning to further improve interpretability and accuracy.

CONFLICT OF INTEREST STATEMENT

The author(s) declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] T. Yan *et al.*, "EEG opto-processor: Epileptic seizure detection using diffractive photonic computing units," *Engineering*, vol. 35, pp. 56–66, 2024.

[2] N. Ikizler and G. Ekim, "Investigating the effects of Gaussian noise on epileptic seizure detection: The role of spectral flatness, bandwidth, and entropy," *Eng. Sci. Technol. Int. J.*, vol. 64, p. 102005, 2025.

[3] F. A. Khan *et al.*, "Explainable AI for epileptic seizure detection in internet of medical things," *Digital Commun. Netw.*, vol. 11, no. 3, pp. 587–593, 2025.

[4] Y. Tang *et al.*, "Epileptic seizure detection based on path signature and Bi-LSTM network with attention mechanism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 304–313, 2024.

[5] H. Kode, K. Elleithy, and L. Almazaydeh, "Epileptic seizure detection in EEG signals using machine learning and deep learning techniques," *IEEE Access*, vol. 12, pp. 80657–80668, 2024.

[6] G. Amrani, A. Adadi, and M. Berrada, "An attention mechanism-based interpretable model for epileptic seizure detection and localization with self-supervised pre-training," *IEEE Access*, vol. 13, pp. 60213–60232, 2025.

[7] M. Sadiq *et al.*, "Novel EEG feature selection based on hellinger distance for epileptic seizure detection," *Smart Health*, vol. 35, p. 100536, 2025.

[8] M. S. Nafea and Z. H. Ismail, "GT-STAFG: Graph transformer with spatiotemporal attention fusion gate for epileptic seizure detection in imbalanced EEG data," *AI*, vol. 6, no. 6, p. 120, 2025.

[9] P. Busia *et al.*, "Wearable epilepsy seizure detection on FPGA with spiking neural networks," *IEEE Trans. Biomed. Circuits Syst.*, pp. 1–11, 2025.

[10] H. F. Atlam, G. E. Aderibigbe, and M. S. Nadeem, "Effective epileptic seizure detection with hybrid feature selection and SMOTE-based data balancing using SVM classifier," *Appl. Sci.*, vol. 15, no. 9, p. 4690, 2025.

[11] S. A. Karthik *et al.*, "Enhanced EEG signal processing for accurate epileptic seizure detection," *SN Comput. Sci.*, vol. 6, no. 6, p. 608, 2025.

[12] A. K. Samantaray and A. D. Rahulkar, "A novel method for epileptic seizure detection using separable gabor wavelets," *IEEE Access*, vol. 13, pp. 116158–116169, 2025.

[13] C. Nouboue *et al.*, "Heart rate variability-based detection of epileptic seizures: Machine learning analysis and characterization of discriminant metrics," *Clin. Neurophysiol.*, vol. 177, p. 2110793, 2025.

[14] T. Shawly and A. A. Alsheikhy, "Eeg-based detection of epileptic seizures in patients with disabilities using a novel attention-driven deep learning framework with SHAP interpretability," *Egypt. Inf. J.*, vol. 31, p. 100734, 2025.

[15] S. Das *et al.*, "Epileptic seizure detection from decomposed EEG signal through 1D and 2D feature representation and convolutional neural network," Information, vol. 15, no. 5, p. 256, 2025.

[16] X. Cao *et al.*, "A hybrid CNN-Bi-LSTM model with feature fusion for accurate epilepsy seizure detection," *BMC Med. Inf. Decis. Mak*, vol. 25, p. 6, 2025.

[17] P. Kunekar, M. K. Gupta, and P. Gaur, "Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques," *J. Eng. Appl. Sci.*, vol. 71, p. 21, 2024.

[18] V. V. Grubov *et al.*, "Two-stage approach with combination of outlier detection method and deep learning enhances automatic epileptic seizure detection," *IEEE Access*, vol. 12, pp. 122168–122182, 2024.

[19] B. Yu, M. Zuo, and L. Sui, "Deep learning with dual-channel feature fusion for epileptic EEG signal classification," Eng, vol. 6, no. 7, p. 150, 2025.

[20] J. Qin *et al.*, "Dual-modality transformer with time series imaging for robust epileptic seizure prediction," *Appl. Sci.*, vol. 15, no. 3, p. 1538, 2025.

[21] B. Silpa and M. K. Hota, "DSLWNet: A dual-stream lightweight deep learning network for the detection of epileptic seizures using EEG signals," *Connect. Sci.*, vol. 37, no. 1, p. 2518985, 2025.

[22] Z. Huang *et al.*, "EEG detection and recognition model for epilepsy based on dual attention mechanism," *Sci. Rep.*, vol. 15, p. 9404, 2025.

[23] Z. Zhang *et al.*, "Cross-patient automatic epileptic seizure detection using patient-adversarial neural networks with spatio-temporal EEG augmentation," *Biomed. Signal Process. Control*, vol. 89, p. 105664, 2024.

[24] A. Mehrabi *et al.*, "Hybrid spike-encoded spiking neural networks for real-time EEG seizure detection: A comparative benchmark", *Biomimetics*, vol. 11, no. 1, p. 75, 2026.

[25] CHB-MIT dataset, Available: https://physionet.org/content/chbmit/1.0.0/

[26] BONN dataset, Available: https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlinear-deterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi.