JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-05, May (2026) pp 9-24 ISSN (Print) 0973-8975

A FREQUENCY-AWARE CNN–VISION TRANSFORMER WITH ADAPTIVE MULTI-STREAM FEATURE FUSION AND UNCERTAINTY ESTIMATION FOR EEG SEIZURE DETECTION

Sachin Chawla[1], Rajeev Ranjan[2], Yogendra Narayan[3]

[1,2,3] Department of Electronics and Communication Engineering, University Institute of Engineering, Chandigarh University, Mohali, 140413, Punjab, India

Email: [1]officialsachinchawla@gmail.com, [2]rajeevranjan1134@gmail.com, [3]narayan.yogendra1986@gmail.com

Corresponding Author: **Sachin Chawla**

https://doi.org/10.26782/jmcms.2026.05.00002

(Received: February 16, 2026; Revised: May 02, 2026; Accepted : May 11, 2026)

Abstract

Objective: Automated seizure detection from scalp electroencephalography(EEG) remains challenging because EEG signals are non-stationary, noisy, highly imbalanced, and vary substantially across patients. This study aimed to develop a robust deep learning framework for seizure detection under clinically relevant, leakage-controlled evaluation settings.

Methods: We proposed a frequency-aware CNN–Vision Transformer (FA-CNNViT) framework integrating deterministic dataset harmonization, subject-wise leakage-controlled cross-validation, split-specific preprocessing, and post-split window generation. The model combines convolutional encoding for local morphological features with transformer-based modeling of long-range dependencies. An adaptive multi-stream feature fusion module was used to preserve temporal, spectral, and spatial information. Asymmetric focal loss addressed class imbalance, and Monte Carlo dropout was used to estimate predictive uncertainty.

Results: On the CHB-MIT dataset, FA-CNNViT achieved 99.13% accuracy, 99.10% sensitivity, 99.47% specificity, 99.55% F1-score, and 99.74% ROC-AUC. In the cross-subject setting on the Turkish EEG dataset, it achieved 89.98% accuracy, 87.41% sensitivity, 88.61% specificity, 87.44% F1- score, and 88.79% ROC-AUC.

Conclusion: The proposed framework achieved strong subject-wise performance and competitive cross-subject performance under a leakage-controlled evaluation protocol. Further refinement of false-positive control and prospective validation is needed before real-time clinical deployment.

Keywords: Epileptic seizure detection, Electroencephalography (EEG) signals, Convolutional Neural Network, Vision Transformer, Adaptive Multi-Stream Feature Fusion

I. Introduction

Epilepsy is a major neurological disorder, and timely seizure detection remains important for diagnosis, treatment planning, and long-term patient management[I]. Electroencephalography (EEG) is the primary non-invasive tool for monitoring seizure-related brain activity, but automated seizure detection from scalp EEG remains challenging because EEG signals are highly non-stationary, noisy, and strongly affected by inter-subject and intra-subject variability [IV, XXI, XXII]. In addition, seizure events occupy only a small fraction of long-term recordings, creating severe class imbalance between seizure and non-seizure segments [XX, II]. These factors can reduce generalization performance and increase missed detections or false alarms in practical deployment [V].

Traditional seizure-detection systems generally rely on handcrafted time-frequency, or time–frequency-domain features, followed by conventional classifiers [IX]. Although such approaches offer some interpretability, their performance depends heavily on manually designed descriptors and often degrades under heterogeneous recording conditions [XIII, XIV]. Deep learning has therefore become increasingly attractive because it enables hierarchical representation learning directly from raw or minimally processed EEG signals [XI, XII]. In particular, convolutional neural networks (CNNs) are effective for capturing local seizure-relevant morphology, while recurrent and attention-based models are better suited for longer-range temporal dependency modelling [XVI, XIX, XXIII]. However, many existing methods still rely on a single representation paradigm and do not explicitly incorporate physiologically meaningful EEG frequency structure into the learning process [VI]. Recent work has shown that deep CNN, CNN-LSTM, residual recurrent, and hybrid CNN–Transformer architectures can improve seizure detection performance compared with traditional feature-based pipelines[XXII, XI, VII, XVIII, XXV]. Time-frequency representations, including wavelet-based features and spectrogram-derived inputs, have also proved useful for characterizing seizure-related spectral transients [XX, VII, XVII, XXIV]. In addition, recent reviews emphasize that clinically useful seizure-detection systems must address not only classification accuracy but also class imbalance, operating-threshold behavior, and robustness under heterogeneous acquisition settings [III, XV, VIII]. Hybrid CNN–Transformer models with multi-stream fusion have emerged as a promising direction because they combine local feature extraction with global contextual modeling [X, XXIV]. Nevertheless, current approaches often use fixed receptive fields, static fusion strategies, or generic tokenization schemes, and many do not explicitly encode frequency-band-aware EEG structure during feature learning [III, X, VIII, XXIV]. Motivated by these limitations, this study proposes a Frequency-Aware CNN-Vision Transformer (FA-CNNViT) framework for EEG seizure detection with uncertainty-aware classification. The proposed model combines CNN-based local morphological feature extraction with transformer-based contextual modeling, while explicitly incorporating frequency-aware processing through adaptive gating and

adaptive multi-stream feature fusion. To better handle seizure/non-seizure imbalance, asymmetric focal loss is used as the training objective, and Monte Carlo dropout is employed to estimate predictive uncertainty. Overall, the proposed framework aims to improve EEG-based seizure detection by jointly addressing three key challenges: physiologically informed representation learning, robust classification under extreme class imbalance, and uncertainty-aware inference. Experimental evaluation on benchmark EEG datasets demonstrates strong performance and consistent gains over ablated variants under leakage-controlled subject-wise evaluation settings.

The proposed framework differs from prior CNN–ViT seizure detectors in three main ways. First, it uses a leakage-controlled subject-wise protocol in which splitting is completed before segmentation, normalization, or overlap generation. Second, it introduces adaptive multi-stream feature fusion using cross-branch aggregation, band-aware gating, and scale-aware weighting. TShird, it incorporates uncertainty-aware inference with epistemic–aleatoric decomposition and confidence-sensitive evaluation.

II. Proposed Methodology

Proposed Framework: FA-CNNViT

We propose FA-CNNViT, a Frequency-Aware CNN–Vision Transformer with Adaptive Multi-Stream Feature Fusion and Uncertainty Estimation for 3 EEG seizure detection, as illustrated in Fig. 1. The framework is designed to address two practical challenges in automated seizure analysis: robust representation learning from heterogeneous EEG recordings and strict prevention of information leakage during model development and evaluation.

The complete pipeline consists of deterministic dataset harmonization, leakage-controlled outer K-fold subject-wise cross-validation, split-specific preprocessing and post-split window generation, hybrid CNN–ViT feature learning, adaptive multi-stream feature fusion, uncertainty-aware classification, and held-out evaluation using discrimination-oriented and operational metrics. In contrast to conventional CNN–ViT fusion pipelines based on static concatenation, the proposed method emphasizes leakage control, adaptive multi-stream feature fusion, and confidence-aware prediction.

Problem Formulation

Let

$$D^{(d)} = \left(X_{(s,r)}^{(d)}, a_{(s,r)}^{(d)}\right) | s \in S^{(d)}, r \in R_s^{(d)} \tag{1}$$

denote dataset d, where $X_{(s,r)}^{(d)} \in R^{(d_c \times T_{(s,r)})}$ represents the r-th EEG recording of subject s, C_d is the number of channels, T_s, r is the recording length in samples, and $a_{(s,r)}^{(d)}$ is the corresponding seizure annotation vector. Each dataset is processed independently; therefore, the study evaluates within-dataset subject-wise generalization rather than explicit cross-dataset transfer performance.

Deterministic Dataset Harmonization

The proposed framework uses two public EEG datasets that differ in sampling rate, channel configuration, and recording characteristics. To improve structural comparability while preserving evaluation integrity, only deterministic harmonization operations are applied before fold assignment. These operations include sampling-rate alignment, channel intersection or masking, channel-set harmonization, missing-channel mapping, and metadata or format alignment. Since these operations do not estimate trainable or split-dependent statistics from the full dataset, they do not introduce leakage. Formally, the harmonized recording is written as

$$\widetilde{X_{(s,r)}^{(d)}} = H^{(d)}\left(X_{(s,r)}^{(d)}\right) \tag{2}$$

where $\mathcal{H}^{(d)}(\cdot)$ denotes the deterministic harmonization operator for dataset d.

During deterministic harmonization, three CHB-MIT recordings exhibited incompatible channel naming or montage conventions relative to the canonical harmonized channel configuration and were therefore excluded prior to fold-wise window generation. This conservative exclusion was adopted to avoid erroneous channel alignment and to preserve structural consistency across folds.

Leakage-Controlled Outer Subject-Wise Cross-Validation

To ensure a valid evaluation, FA-CNNViT adopts dataset-specific outer subject-wise cross-validation, using 8 outer folds for the CHB-MIT dataset and 10 outer folds for the Turkish EEG dataset. Unique subject identifiers are partitioned into K-folds and stratified, as far as possible, according to seizure presence or seizure burden. In each outer iteration, one fold is reserved as the held-out test set, validation subjects are selected only from the remaining folds, and the rest are used for training.

Let

$$S^{(d)} = \bigcup(k=1)^{(K_d)} F_k^{(d)}, F_i^{(d)} \cap F_j^{(d)} = \emptyset for i \neq j \tag{3}$$

where $F_k^{(d)}$ denotes the k-th subject fold for dataset d, and K_d is the number of outer folds used for that dataset. In this study, Kd = 8 for CHB-MIT and K_d = 10 for the

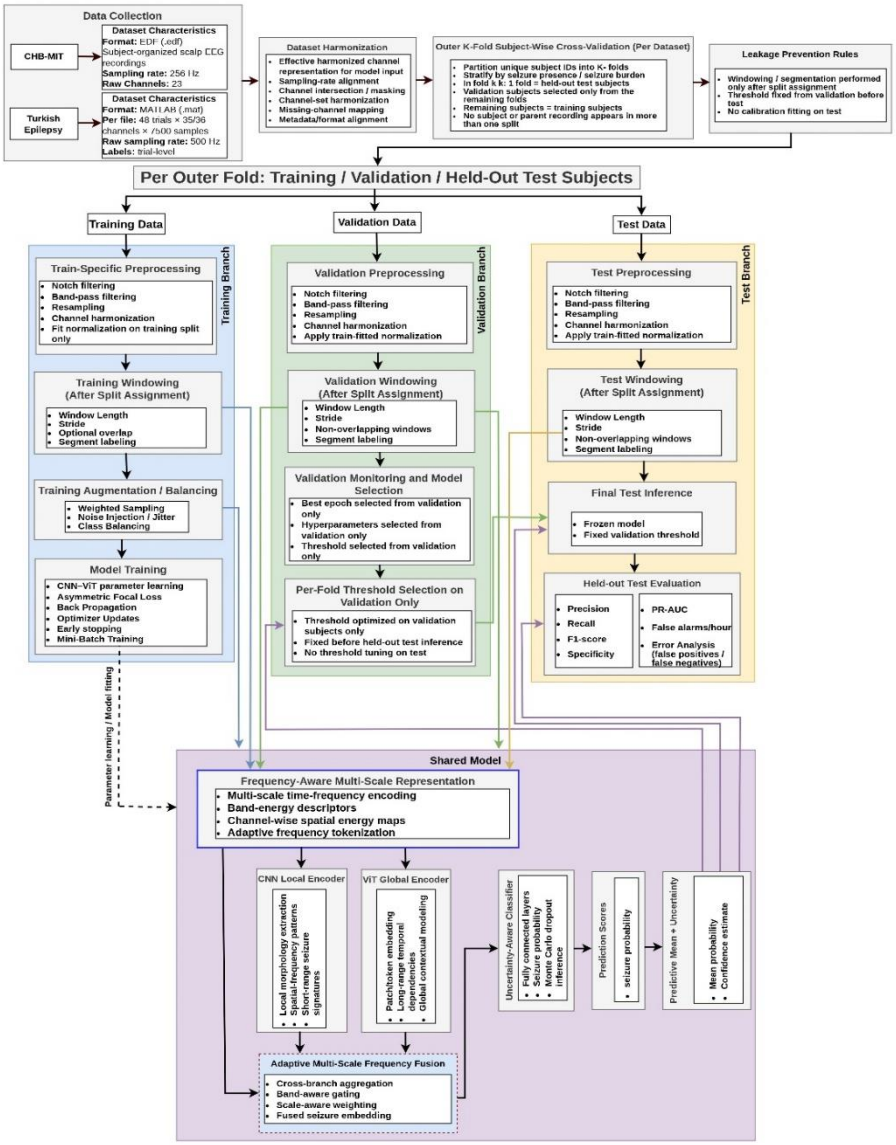

Fig. 1. Leakage-controlled FA-CNNViT framework for EEG seizure detection

Turkish EEG dataset. For the k-th outer iteration, the held-out test subjects are defined as

$$S_t e^{(d,k)} = F_k^{(d)} \tag{4}$$

From the remaining folds, 20% of subjects were selected as the validation set using a stratified subject-level split based on seizure presence and seizure burden, and the remaining subjects were assigned to training. The protocol enforces the following leakage-prevention rules: (i) split assignment is completed before any segmentation or

overlap generation, (ii) no subject or parent recording appears in more than one split within a fold, (iii) no normalization statistic was estimated from validation or test recordings at any stage of model development, (iv) overlapping windows are permitted only in the training branch, whereas validation and test windows are strictly non-overlapping, (v) threshold selection is performed on validation data only and fixed before test inference, and (vi) no calibration fitting is performed on the test branch.

Split-Specific Preprocessing and Post-Split Windowing

After the fold assignment, the training, validation, and test branches are processed independently. In the training branch, EEG signals undergo notch filtering, band-pass filtering, resampling, and channel harmonization, followed by normalization fitted only on the training split. In the validation and test branches, the same deterministic preprocessing steps are applied, but the normalization parameters are not re-estimated; instead, the training-fitted parameters are reused unchanged. Window generation is also performed only after split assignment. The training branch uses a fixed window length and a stride with optional overlap to enrich the number of training samples. By contrast, both validation and test branches use strictly non-overlapping windows to avoid near-duplicate temporal segments. Each window is assigned a seizure or non-seizure label according to the annotation of its parent recording.

Training Branch: Augmentation, Balancing, and Optimization

To mitigate class imbalance and improve robustness, the training branch includes weighted sampling, noise injection, or temporal jitter, and class balancing. These operations are restricted to the training split only. Model parameters are optimized using mini-batch training with backpropagation and iterative optimizer updates, while early stopping is used to reduce overfitting.

The official training objective is Asymmetric Focal Loss, defined as

$$\mathcal{L}_{\mathcal{AFL}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\alpha y_i (1-p_i)^{\gamma}\log(p_i+\epsilon) + (1-\alpha)(1-y_i)p_i^{\gamma}\log(1-p_i+\epsilon)\right] \quad (5)$$

where $y_i \in \{0, 1\}$ is the ground-truth label, p_i is the predicted seizure probability, α is the class-weighting factor, $\gamma+$ and $\gamma-$ are focusing parameters for positive and negative samples, and ϵ is a small numerical constant. This loss is adopted as the sole optimization objective throughout the manuscript and should be described consistently in Methods, tables, and Results.

Validation Branch: Model Selection and Threshold Determination

The validation branch serves two functions: model selection and threshold determination. First, it is used to select the best-performing epoch and validate hyperparameter choices without exposing the held-out test fold. Second, it supports per-fold threshold optimization using validation subjects only. The final decision threshold was determined on the validation branch by optimizing a predefined validation objective reflecting the trade-off between seizure sensitivity and false alarms per hour. Once selected, the validation-derived threshold was fixed and applied unchanged to the held-out test fold.

Test Branch: Frozen Inference and Held-Out Evaluation

The test branch constitutes the final and fully unseen phase of evaluation. Test data undergo the same deterministic preprocessing pipeline, using normalization parameters derived exclusively from the training split and nonoverlapping window segmentation. The trained model is then evaluated in a fixed inference setting, without any further parameter updates or threshold recalibration. As a result, the test branch provides an unbiased held-out assessment of model performance under conditions that more closely resemble real-world deployment.

Shared FA-CNNViT Architecture

The shared model is built around a frequency-aware representation module and an Adaptive Multi-Stream Feature Fusion (MSFF) module designed to capture complementary seizure-related patterns from EEG windows. It integrates multi-scale time-frequency representations, band-energy features, channel-wise spatial energy maps, and adaptive frequency tokenization. Collectively, these components retain fine spectral details while also preserving broader channel-dependent characteristics associated with ictal activity.

The resulting representation is then passed through two parallel encoders. The CNN-based local encoder focuses on extracting morphological details, spatial-frequency patterns, and short-duration seizure signatures. In parallel, the ViT-based global encoder learns token-level interactions, long-range temporal relationships, and broader contextual dependencies across the input. This two-branch design enables FA-CNNViT to capture both detailed local patterns and wider temporal structure within EEG signals. The features produced by the CNN and ViT branches are subsequently integrated through the Adaptive Multi-Stream Feature Fusion module. Instead of relying on simple static concatenation, this fusion stage performs cross-branch aggregation, band-aware gating, and scale-aware weighting to form a unified seizure embedding. In doing so, the model can dynamically emphasize the most informative frequency scales and contextual cues for each EEG segment, which improves robustness to inter-subject differences and dataset-specific variability.

Epistemic and Aleatoric Uncertainty Decomposition

The uncertainty-aware classifier decomposes predictive uncertainty into epistemic and aleatoric components. Epistemic uncertainty was estimated using Monte Carlo dropout with stochastic forward passes during inference, 8 while aleatoric uncertainty was estimated using a heteroscedastic output head that predicts an input-dependent log-variance term. The total uncertainty was computed as the sum of epistemic and aleatoric components. This decomposition helps distinguish uncertainty caused by limited model knowledge from uncertainty caused by noisy, artifact-contaminated, or clinically ambiguous EEG segments. Detailed equations are provided in Supplementary Section S3.

Evaluation Metrics

Held-out test performance was evaluated using both discrimination-based and clinically meaningful measures. At the segment level, model performance was

quantified using precision, recall, F1-score, specificity, and PR-AUC. To reflect practical operating behavior, false alarms per hour were also reported, and false-positive as well as false-negative predictions were examined qualitatively to identify common error patterns. Taken together, these metrics provide a more informative assessment than accuracy alone, as they capture not only discriminative performance but also operational timeliness and potential clinical utility. Since validation and held-out test windows were not resampled or class-balanced, the primary reported discrimination metrics were already computed on the natural evaluation distribution. Prior correction and importance-weighted risk estimation were therefore used as additional safeguards for probability-based uncertainty assessment and risk summaries under training– evaluation class-prior mismatch.

Summary of the Proposed Pipeline

In summary, FA-CNNViT combines a leakage-controlled subject-wise evaluation protocol with a hybrid deep architecture tailored for seizure detection. Deterministic harmonization improves structural comparability without using split-dependent information. Subject-wise partitioning is completed before segmentation, ensuring strict isolation between training, validation, and held-out testing. The model then learns complementary local and global representations from frequency-aware multi-stream inputs, adaptively fuses them across branches and scales, and outputs both seizure probability and predictive uncertainty. Together, these components aim to improve not only classification performance but also the reliability and interpretability of automated EEG seizure detection.

III. Experimental Study

Experimental Environment and Reproducibility

All experiments were implemented in Python using TensorFlow/Keras with GPU acceleration. To improve reproducibility, fixed random seeds were used for Python, NumPy, and TensorFlow, and a structured experiment manifest was maintained to record software versions, hardware configuration, random seeds, and key hyperparameters. Experiments were executed on an ASUS Zephyrus G16 laptop with TensorFlow GPU support enabled.

Datasets and Cohort Characteristics

The proposed framework was evaluated on two publicly accessible EEG datasets: the CHB-MIT scalp EEG dataset and the Turkish Epilepsy EEG dataset. These datasets were selected because they differ in storage format, annotation structure, and cohort organization, thereby providing complementary subject-wise evaluation settings for seizure detection. After exclusion of non-data system artifacts, the cleaned CHB-MIT repository contained 686 primary EEG recordings in .edf format together with seizure annotation and auxiliary metadata files, whereas the Turkish EEG dataset contained 121 primary EEG files in MATLAB .mat format. The two datasets were processed independently and were not pooled during evaluation. Supplementary Table S2 summarizes the cleaned repository composition. The CHB-MIT dataset is organized mainly as subject-wise .edf recordings with separate seizure annotations, whereas the Turkish EEG dataset is distributed as .mat files with structural information embedded

in the dataset, as shown in Supplementary Fig. S1. This structural heterogeneity motivated the deterministic harmonization stage applied before subject-wise split assignment. File-level inspection confirmed consistent data and label fields in the Turkish EEG repository. For CHB-MIT, header-level harmonization auditing revealed non-uniform montage representations and channel-label variants. Using a canonical harmonized bipolar channel configuration, 655 recordings were retained, and 31 were excluded prior to fold-wise preprocessing, as illustrated in Supplementary Fig. S2. Overall, these differences justify the use of deterministic harmonization and leakage-controlled subject-wise evaluation.

Leakage-Controlled Evaluation Protocol

All experiments were performed using subject-wise outer cross-validation applied independently to each dataset. The resulting cohort consisted of 24 CHB-MIT subject units and 121 Turkish EEG subject- or file-level units. At the subject level, CHB-MIT contained only seizure-positive units, whereas the Turkish EEG dataset contained 50 seizure-positive and 71 seizure-negative units. For CHB-MIT, outer-fold construction used burden-aware subject-level stratification, resulting in 8 outer folds. For the Turkish dataset, 10 outer folds were used with subject-level stratification by seizure presence. In each outer iteration, one fold was reserved as the held-out test set, 20% of the remaining subject units were assigned to validation, and the rest were used for training. Split sizes remained stable across folds, and no subject overlap occurred across training, validation, and test subsets within any outer fold. Supplementary Fig. S3 and S4 illustrate the resulting subject allocation, while Table 1 summarizes the dataset-specific protocol.

Table-1: Summary of the leakage-controlled evaluation protocol used for each dataset

Dataset	Split unit	Outer folds	Outer stratification strategy	Typical train / val / test allocation
CHB-MIT	Subject	8	Subject-wise burden-aware Stratification using 3 equally sized seizureburden strata	16 / 5 / 3 subjects per fold
Turkish EEG	Subject- or file-level unit	10	Subject-wise stratification by seizure presence	≈86–87 / 22 / 12–13 units per fold

Preprocessing and Window Generation Settings

Preprocessing was applied only after subject-wise split assignment to preserve the leakage-controlled protocol. For each branch, EEG recordings underwent deterministic conditioning, including notch filtering, band-pass filtering, resampling, and channel harmonization. CHB-MIT retained recordings were stored as EDF signals sampled at 256 Hz, whereas the Turkish EEG dataset required reconstruction of multichannel EEG trials from MATLAB object-array storage. After preprocessing, both datasets were mapped to a common target sampling rate of 256 Hz for downstream analysis. Because the raw repositories differed in channel count and montage convention, downstream model development used dataset-specific harmonized channel representations. For CHB-MIT, the effective representation used for normalization and model input

contained 18 channels after deterministic compatibility filtering. For the Turkish EEG dataset, reconstructed 36-channel trials were mapped to 14 channels aligned to the shared bipolar target space. Normalization parameters were estimated exclusively on the training split using channel-wise z-score normalization and then reused unchanged for validation and test data. Supplementary Fig. S5, S6, and S7 show representative examples before and after preprocessing and normalization. Window generation was also performed only after split assignment. The training branch used 4 s windows with optional overlap, whereas validation and test branches used strictly non-overlapping windows. This design enriches training samples while maintaining a conservative held-out evaluation. Supplementary Fig. S8 and S9 illustrate the adopted windowing policy, while Supplementary Tables S3 and S4 summarize the corresponding preprocessing and harmonization settings. To quantify the distribution shift introduced by training-only sampling, original, training-adjusted, and validation/test class ratios were recorded.

Sampling, weighted optimization, and class-balanced adjustment were restricted to the training branch, while validation and held-out test distributions remained non-resampled and naturally imbalanced. The detailed class-ratio comparison is provided in Supplementary Table S10.

Implementation Details and Hyperparameter Configuration

The proposed FA-CNNViT framework was implemented in Python using TensorFlow/Keras as a modular architecture composed of a frequency-aware multi-stream representation block, a CNN-based local encoder, a ViT-based global encoder, an adaptive fusion module, and a dropout-enabled classifier head for uncertainty-aware inference. The same core design was retained across datasets, while the input channel dimension was adjusted according to the harmonized EEG representation of each dataset. The CHB-MIT configuration used an input shape of 18×1024×1, whereas the Turkish EEG dataset configuration used 14 × 1024 × 1. The frequency-aware representation stage employed parallel temporal convolutions with kernel sizes of 7, 15, and 31. The CNN branch used two convolutional blocks with 32 and 64 filters, while the ViT branch used a patch size of 16, embedding dimension 128, 4 attention heads, and 2 encoder blocks with an intermediate MLP dimension of 256. The outputs of both branches were projected into a common latent space of dimension 128 and combined through adaptive fusion before classification. Optimization used Adam with a learning rate of 10−4, batch size 32, maximum 50 epochs, and early stopping with patience 10. Asymmetric Focal Loss was used as the sole training objective, and uncertainty estimation was performed using Monte Carlo dropout with 20 stochastic forward passes. Supplementary Tables S5 and S6 summarize the implementation settings, input shapes, and parameter counts of the two dataset-specific variants. The uncertainty-aware classifier used two output heads: one head predicted the seizure logit, while the second heteroscedastic head predicted an input-dependent log-variance term for aleatoric uncertainty estimation. Epistemic uncertainty was estimated using 20 Monte Carlo dropout forward passes, and total uncertainty was computed as the sum of epistemic and aleatoric components.

Evaluation Metrics

Performance was assessed using both threshold-dependent and threshold-independent metrics. Following common practice in automated seizure detection, we report accuracy, sensitivity, specificity, precision, F1-score, ROCAUC, and PR-AUC, while operational behavior was characterized using false alarms per hour as illustrated in Supplementary Table S7. Threshold selection was performed on the validation split only. To avoid optimistic bias, no threshold was tuned directly on the test set. In preliminary experiments, heavily penalizing false alarms tended to produce overly conservative thresholds with very low recall. Therefore, the final operating threshold was selected from the validation sweep using an F1-priority policy, with recall and specificity used as secondary tie-breakers. This yielded a more balanced operating point for seizure screening scenarios. The operating threshold was selected independently on the validation split of each outer fold and then fixed before inference on the corresponding held-out test fold; therefore, all threshold-dependent metrics were computed using fold-specific validation-derived thresholds rather than a single global cutoff.

Prior Correction and Importance-Weighted Risk Estimation

Because the effective training class distribution differed from the natural held-out evaluation distribution, prior correction and importance-weighted risk estimation were used as safeguards for probability-based and risk-based summaries. Model probabilities were adjusted using the ratio between the effective training prior and the unmodified evaluation prior, and importance-weighted risk was computed using class-specific training–evaluation prior ratios. The full formulation and class-prior values are provided in Supplementary Section S1. Since validation and held-out test windows were not resampled or class-balanced, the primary discrimination metrics were computed on the natural non-resampled evaluation distribution. Statistical Analysis Statistical comparison between FA-CNNViT and ablated or competing models was performed using subject-level bootstrap significance testing. Subjects from held-out test predictions were sampled with replacement, and paired performance differences were recomputed to obtain empirical confidence intervals and bootstrap p-values. Because multiple models and metrics were compared, Bonferroni and Benjamini–Hochberg false-discovery-rate corrections were applied. A comparison was considered statistically reliable only when the corrected p-value was below 0.05 and the 95% bootstrap confidence interval excluded zero. Full statistical formulations are provided in Supplementary Section S2.

Main Quantitative Results

All main performance results were computed on the original imbalanced held-out test distribution. Although training used sampling-related adjustments, weighted sampling, overlap-based augmentation, and class-balanced optimization to mitigate class imbalance, validation and held-out test windows were generated without resampling, without overlap-based augmentation, and without class balancing. Therefore, the reported accuracy, sensitivity, specificity, precision, F1-score, ROC-AUC, PR-AUC, and false-alarm rates reflect performance on the natural non-resampled evaluation distribution. The subject-wise results in Supplementary Table S8

show highly consistent performance on the CHB-MIT dataset. Mean accuracy, sensitivity, specificity, F1-score, and ROC-AUC reached 99.13%, 99.10%, 99.47%, 99.55%, and 99.74%, respectively, with low standard deviations, indicating limited performance fluctuation across subjects. Threshold-dependent metrics in Table 2 were computed using fold-specific thresholds selected on validation subjects and fixed before held-out test inference.

The average confusion matrices further confirmed balanced classification behavior, with dominant diagonal entries and limited off-diagonal errors. These matrices are provided in Supplementary Fig. S10.

Table 2: Overall held-out test performance of FA-CNNViT on CHB-MIT and Turkish EEG

Dataset	Acc	Sens	Spec	Prec	F1	PR-AUC	ROC-AUC	FA/h
CHB-MIT	99.13	99.10	99.47	99.32	99.55	99.61	99.74	0.06
Turkish EEG	89.98	87.41	88.61	87.51	87.44	88.11	88.79	0.34

Ablation Studies

Statistical comparisons between the proposed model and ablated variants were performed using subject-level bootstrap significance testing with multiple-comparison correction. Only comparisons with corrected $p < 0.05$ and a 95% bootstrap confidence interval excluding zero were interpreted as statistically significant. Standalone CNN and ViT ablations confirmed that neither local convolutional modeling nor transformer-based global modeling alone was sufficient to match FA-CNNViT. The detailed single-branch ablation results are reported in Supplementary Table S11. Sampling-rate-dependent experiments showed stable performance under reasonable temporal downsampling, with detailed results provided in Supplementary Table S9. The proposed MSFF module outperformed Serial-Concat and Channel- Gate fusion under the same backbone, training protocol, and data partitions as shown in Supplementary Table S16. The advantage was most visible in the Turkish EEG and reduced-data settings, supporting the value of the proposed Adaptive Multi-Stream Feature Fusion module.

Comparative Experiments

Table 3 compares FA-CNNViT with four widely used deep learning baselines: LSTM, Transformer, CNN-LSTM, and CNN-Transformer. The comparison was designed to assess relative performance in both subject-specific and cross-subject seizure classification. For comparative experiments, bootstrap-based significance testing was used to compare the proposed model with competing approaches, and Bonferroni and Benjamini–Hochberg false–discovery–rate corrections were applied to account for multiple comparisons. For fairness, all comparative baselines (LSTM, Transformer, CNN-LSTM, and CNN-Transformer) were evaluated using the same deterministic harmonization pipeline, subject-wise fold assignments, split-specific preprocessing, training-only normalization, and validation-based thresholding protocol as FA-

CNNViT, while architecture-specific settings were tuned within the same validation framework and comparable optimization budget.

Table-3: Comparison results of FA-CNNViT, LSTM, Transformer, CNN-LSTM, and CNNTransformer on the CHB-MIT and Turkish EEG datasets

Databases	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC (%)
CHB-MIT	FA-CNNViT	99.13	99.10	99.47	99.55	99.74
	LSTM	86.41	84.22	82.76	83.94	86.15
	Transformer	92.39	90.53	91.22	90.97	91.31
	CNN-LSTM	93.21	93.22	92.42	93.64	92.44
	CNN-Transformer	97.71	96.16	96.73	96.90	97.37
Turkish EEG	FA-CNNViT	89.98	87.41	88.61	87.44	88.79
	LSTM	72.96	73.44	73.72	71.15	71.64
	Transformer	82.24	81.75	80.85	82.09	81.16
	CNN-LSTM	84.98	81.64	82.59	84.31	83.07
	CNN-Transformer	87.41	85.70	86.14	84.39	85.64

Across both datasets, FA-CNNViT achieved the strongest numerical performance among the evaluated baselines. These results support the benefit of combining convolutional local feature extraction, transformer-based contextual modeling, and adaptive multi-stream feature fusion under the proposed leakage-controlled protocol. Subject-level bootstrap testing with multiple-comparison correction confirmed that FA-CNNViT achieved statistically reliable improvements over competing baselines. Detailed confidence intervals and corrected p-values are reported in Supplementary Table S12. Uncertainty analysis was performed using held-out test predictions pooled across outer folds. Correct and high-confidence predictions showed lower uncertainty, whereas incorrect and low-confidence predictions showed higher epistemic uncertainty. Noisy or ambiguous EEG windows showed higher aleatoric uncertainty, supporting the usefulness of the heteroscedastic uncertainty head. Subject-level predictive variance was positively associated with misclassification rate, and risk decreased after rejecting high-uncertainty predictions. Calibration quality was summarized using the ECE computed with 10 equal-width confidence bins. Detailed uncertainty decomposition, subject-level correlation, and risk–coverage/ECE summaries are reported in Supplementary Tables S13–S15.

Conclusion and Future Work

This study presented FA-CNNViT, a frequency-aware CNN–Vision Transformer framework for EEG-based seizure detection that combines convolutional local feature extraction, transformer-based contextual modeling, adaptive fusion, and uncertainty-aware prediction. By jointly modeling local morphology and longer-range dependencies, the proposed approach was designed to better capture seizure-relevant temporal, spectral, and spatial EEG characteristics. Experimental results on the CHB-MIT and Turkish EEG datasets showed that the framework achieved strong subject-

wise performance and competitive cross-subject performance under a leakage-controlled evaluation protocol. Comparative and ablation studies further indicated that combining convolutional and transformer-based modeling with the proposed fusion strategy contributed positively to overall performance. Despite these encouraging results, several limitations remain. Evaluation on two public datasets cannot fully represent the variability encountered in real-world clinical environments. In addition, the hybrid CNN–Transformer design increases computational cost, which may limit deployment in real-time or resource-constrained settings. Further validation is therefore needed across acquisition protocols, devices, and unseen clinical populations. Future work will focus on patient-specific adaptation, cross-dataset domain generalization, computational optimization for real-time deployment, and explainable artificial intelligence methods to improve clinical interpretability and support clinician trust.

Compliance with Ethical Standards

This study used publicly available, de-identified datasets; therefore, ethics approval and participant consent were not required. The authors declare no 17 conflict of interest and received no specific funding. The CHB-MIT and Turkish EEG datasets are publicly available from their original sources. Source code is available from the corresponding author upon reasonable request. All authors contributed to the study design, validation, writing, and approval of the final manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the article.

References

I. Chen, Wenna, et al. "An automated detection of epileptic seizures EEG using CNN classifier based on feature fusion with high accuracy." BMC Medical informatics and Decision making 23.1 (2023): 96. 10.1186/s12911-023-02180-w

II. Christou, Vasileios, et al. "Evaluating the window size's role in automatic EEG epilepsy detection." Sensors 22.23 (2022): 9233. 10.3390/s22239233

III. Ein Shoka, Athar A., et al. "EEG seizure detection: concepts, techniques, challenges, and future trends." Multimedia tools and applications 82.27 (2023): 42021-42051. 10.1007/s11042-023-15052-2

IV. Feng, Lufeng, et al. "A multi-view neural framework with attention for epileptic seizure classification." Journal of Neural Engineering 23.1 (2026): 016018. 10.1088/1741-2552/ae33f8

V. Guhdar, Mohammed, Ramadhan J. Mstafa, and Abdulhakeem O. Mohammed. "A multimodal temporal attention network for seizure classification via one-dimensional convolutional neural architecture." Biomedical Signal Processing and Control 112 (2026): 108495. 10.1016/j.bspc.2025.108495

VI. Islam, Md Rabiul, et al. "Epileptic seizure focus detection from interictal electroencephalogram: a survey." Cognitive neurodynamics 17.1 (2023): 1-23. 10.1007/s11571-022-09816-z

VII. Kashefi Amiri, Homa, Masoud Zarei, and Mohammad Reza Daliri. "Epileptic seizure detection from electroencephalogram signals based on 1D CNN-LSTM deep learning model using discrete wavelet transform." Scientific Reports 15.1 (2025): 32820. 10.1038/s41598-025-18479-9

VIII. Kode, Hepseeba, Khaled Elleithy, and Laiali Almazaydeh. "Epileptic seizure detection in EEG signals using machine learning and deep learning techniques." IEEE access 12 (2024): 80657-80668. 10.1109/ACCESS.2024.3409581

IX. Lebal, Abdelhamid, Abdelouahab Moussaoui, and Abdelmounaam Rezgui. "Epilepsy-Net: attention-based 1D-inception network model for epilepsy detection using one-channel and multi-channel EEG signals." Multimedia tools and applications 82.11 (2023): 17391-17413. 10.1007/s11042-022-13947-0

X. Li, Qi, Wei Cao, and Anyuan Zhang. "Multi-stream feature fusion of vision transformer and CNN for precise epileptic seizure detection from EEG signals." Journal of Translational Medicine 23.1 (2025): 871. 10.1186/s12967-025-06862-z

XI. Li, Yang, et al. "Automatic seizure detection using fully convolutional nested LSTM." International journal of neural systems 30.04 (2020): 2050019. 10.1142/S0129065720500197

XII. Luo, Weitao, et al. "EEG-Based Brain-Computer Interface: Fundamentals, Methods, Applications, and Challenges." IEEE Internet of Things Journal (2025). 10.1109/JIOT.2025.3625060

XIII. Ma, Mengnan, et al. "Research on epileptic EEG recognition based on improved residual networks of 1-D CNN and indRNN." BMC Medical Informatics and Decision Making 21.Suppl 2 (2021): 100. 10.1186/s12911-021-01438-5

XIV. Pan, Yayan, et al. "Downsampling of EEG signals for deep learning-based epilepsy detection." IEEE Sensors Letters 7.12 (2023): 1-4. 10.1109/LSENS.2023.3332392

XV. Ren, Juntao, et al. "An Event-Based Filtering and Weighted Enhanced Deep Learning Epileptic Seizure Prediction Method." Neural Networks (2025): 108424. 10.1016/j.neunet.2025.108424

XVI. Salini, G. Indu, I. Sowmy, and T. K. Sreeja. "FrAdadelta-CSA: Fractional Adadelta Chameleon Swarm Algorithm-based feature selection with SpikeGoogle-DenseNet for epileptic seizure detection." Computational Biology and Chemistry 119 (2025): 108550. 10.1016/j.compbiolchem.2025.108550

XVII. Shah, Syed Yaseen, et al. "Epileptic seizure classification based on random neural networks using discrete wavelet transform for electroencephalogram signal decomposition." Applied Sciences 14.2 (2024): 599. 10.3390/app14020599

XVIII. Shanmugam, Shalini, and Selvathi Dharmar. "A CNN-LSTM hybrid network for automatic seizure detection in EEG signals." Neural Computing and Applications 35.28 (2023): 20605-20617. 10.1007/s00521-023-08832-2

XIX. Srinivasan, Saravanan, et al. "Detection and classification of adult epilepsy using hybrid deep learning approach." Scientific reports 13.1 (2023): 17574. 10.1038/s41598-023-44763-7

XX. Sunaryono, Dwi, Riyanarto Sarno, and Joko Siswantoro. "Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features." Journal of King Saud University-Computer and Information Sciences 34.10 (2022): 9591-9607. 10.1016/j.jksuci.2021.11.015

XXI. Tasci, Irem, et al. "Epilepsy detection in 121 patient populations using hypercube pattern from EEG signals." Information Fusion 96 (2023): 252-268. 10.1016/j.inffus.2023.03.022

XXII. Tawhid, Md Nurul Ahad, Siuly Siuly, and Tianning Li. "A convolutional long short-term memory-based neural network for epilepsy detection from EEG." IEEE Transactions on Instrumentation and Measurement 71 (2022): 1-11. 10.1109/TIM.2022.3217515

XXIII. Wang, Quanhong, et al. "A hybrid SVM and kernel function-based sparse representation classification for automated epilepsy detection in EEG signals." Neurocomputing 562 (2023): 126874. 10.1016/j.neucom.2023.126874

XXIV. Wang, Zhuohan, et al. "EEG-based seizure detection using dual-branch CNN-ViT network integrating phase and power spectrograms." Brain sciences 15.5 (2025): 509. 10.3390/brainsci15050509

XXV. Xu, Gaowei, et al. "A one-dimensional CNN-LSTM model for epileptic seizure recognition using EEG signal analysis." Frontiers in neuroscience 14 (2020): 578126. 10.3389/fnins.2020.578126