Universal EEG Epilepsy Detection via Evidential Multi-View De-Biasing

Ziqi Wen[1*], Cai Xu[1*], Wanqing Zhao[2], Jie Zhao[3], Wei Zhao[1†]

[1]School of Computer Science and Technology, Xidian University, Xi'an, China
[2]School of Information Science and Technology, Northwest University, Xi'an, China
[3]School of Data Science and Artificial Intelligence, Chang'an University, Xi'an, China
zqwenn@stu.xidian.edu.cn, cxu@xidian.edu.cn, zhaowq@nwu.edu.cn, jiezhao@chd.edu.cn, ywzhao@mail.xidian.edu.cn

Abstract

Epilepsy is a widespread neurological disorder characterized by highly patient-specific EEG patterns. Existing EEG-based seizure detection methods either train individualized models for each patient or adapt models pre-trained on known patients to new ones. However, when encountering previously unseen patients, these methods typically require retraining or fine-tuning, which limits their practical utility in clinical settings. This limitation can be linked to biases caused by patient-specific variations, which obscure the underlying pathological patterns of seizures. To address this, we propose an evidential multi-view framework that reinforces the learning of core epileptic features by promoting consistency across multiple views and reducing reliance on high-uncertainty, patient-specific segments. Specifically, we introduce Bias-guided Fisher-Evidential Multi-View Learning (BF-EML) to guide the model toward discovering intrinsic seizure patterns. BF-EML employs a two-stage training architecture: In Stage 1, we use the Fisher Information Matrix to reorder EEG segments by uncertainty and deliberately train a biased feature generator on low-evidence segments. In Stage 2, we design a dual-branch network where the biased and unbiased branches are alternately trained, encouraging the unbiased branch to reduce its reliance on patient-specific biases. Finally, we introduce a shift-calibrated fusion strategy to enhance the consistency of pathogenic feature integration. Extensive experiments on public datasets and a clinical dataset demonstrate that our method achieves superior performance in both single- and multi-patient scenarios. Importantly, it generalizes well to unseen patients without the need for retraining.

Code — https://github.com/Wednesque/BF-EML

1 Introduction

Epilepsy remains one of the most prevalent chronic neurological disorders globally. Recurrent seizures pose significant risks to cognitive function and mental health, and individuals with epilepsy face a threefold increased risk of premature death (Ding et al. 2006). Clinical evidence suggests that early detection and treatment can effectively control seizures in up to 70% of cases (Organization et al. 2019), highlighting the importance of rapid screening and monitoring in patient care. As a non-invasive and cost-effective neuroimaging technique, Electroencephalography (EEG) enables capturing neuronal discharge activities of the brain through scalp electrodes, making it a widely utilized tool in automated epilepsy detection systems.

Existing EEG-based seizure detection methods can be broadly classified into two categories: **Patient-Specific** and **Multi-Patient** approaches. **Patient-Specific** models are trained individually for each patient and can achieve high accuracy (typically $> 95\%$) (Dutta et al. 2024; Li et al. 2020). However, their performance degrades significantly when applied to new patients, due to the considerable variability in seizure manifestations. Epileptic seizures encompass a broad range of subtypes (e.g., absence, tonic, clonic) (Scheffer et al. 2017), each exhibiting highly individualized EEG signatures. To improve generalization, **Multi-Patient** models attempt to train a shared model across data from multiple patients (Dissanayake et al. 2021). While intuitively appealing, this approach often leads to cognitive bias (Bahng et al. 2020) during training: since patient-specific rhythms or artifacts are more statistically salient, models tend to rely on these features rather than learning the core pathological EEG patterns underlying seizure activity—namely, paroxysmal discharges caused by abnormal neuronal firing. As a result, these models often fail to generalize well to unseen patients, despite access to diverse training data.

Recent **Multi-Patient** approaches have introduced domain adaptation (DA) techniques to transfer knowledge from labeled source patients to an unlabeled target patient by aligning their features in a shared latent space. Representative methods include pseudo-labeling with distribution alignment (e.g., MMD) (Cui et al. 2023), prototype-based matching (Liang et al. 2023), and adversarial training for domain confusion (Peng et al. 2022). While effective within the standard DA setting—where the model is adapted from a fixed source domain to a single target domain—such approaches are ill-suited for real-world seizure detection, where each patient constitutes a distinct domain with unique EEG characteristics, and test-time patients are typically unseen during adaptation. This one-to-one adaptation paradigm fails to generalize: a model adapted from cats to tigers may work well for tigers, but seizure detection must also handle pandas—patients unlike both the source

*These authors contributed equally.
†Corresponding author.

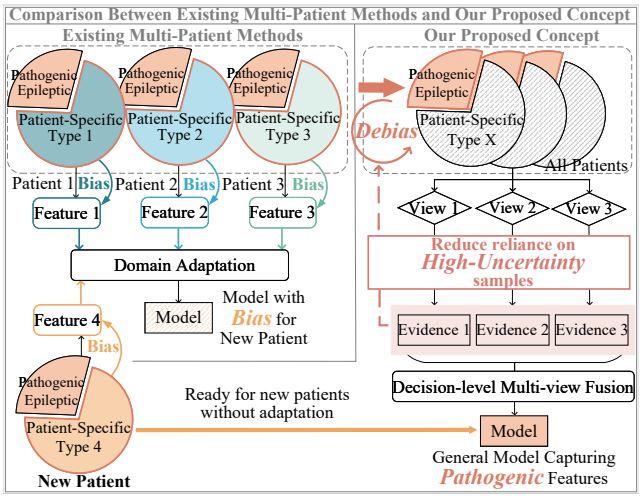

Figure 1: Conceptual comparison between conventional alignment-based methods and our BF-EML framework. Existing approaches often rely on patient-specific EEG traits, leading to poor generalization. BF-EML introduces decision-level multi-view fusion with bias elimination and cross-view consistency to encourage learning of patient-invariant epileptic patterns for universal seizure detection.

and the adapted target. More critically, the DA process encourages the model to exploit patient-specific distributional cues that are easy to align, rather than learning the true semantic patterns indicative of seizures (e.g., transient spikes or discharges). Since distribution alignment is the core DA objective, the model is incentivized to "shortcut" by matching surface-level traits that correlate with labels in the target, which can replicate the same kind of cognitive bias observed in prior multi-patient methods. Although this can yield high performance on the adapted patient, it fails to capture seizure semantics and necessitates retraining or fine-tuning for every new patient—making deployment impractical.

We propose a paradigm shift in the design of generalized seizure detection models, as illustrated in Fig. 1. Instead of allowing patient-specific biases to dominate model learning, we encourage the model to reduce its reliance on such idiosyncratic features, thereby facilitating the discovery of seizure-related pathological patterns. We observe that when trained on data from diverse patients, the model tends to assign higher confidence to patterns consistently associated with seizures, while expressing greater uncertainty toward patient-specific components (as validated in Sec. 4.4). Motivated by this, we introduce a two-stage training framework: the first stage identifies biased, uncertain components arising from patient-specific variations; the second stage explicitly suppresses reliance on these components to guide the model toward learning generalizable, pathology-centered representations. To further enhance this process, we incorporate an evidential multi-view learning strategy. By extracting view-wise seizure evidence—quantified via uncertainty—and fusing at the decision level, we leverage cross-view consistency to reinforce model attention on pathogenic epileptic signatures, while minimizing the influence of view-specific bias.

We then introduce the Bias-Guided Fisher-Evidential Multi-View Learning (BF-EML) framework, which leverages multi-view learning to consistently capture pathogenic epileptic features across EEG views (e.g., spike-wave complexes in the frequency domain) while mitigating non-pathogenic patient-specific biases. As illustrated in Fig. 2, BF-EML identifies learning biases and explicitly guides the model to overcome them. In **Stage 1**, we pre-train an evidence extraction model to quantify the evidential support for epilepsy detection from an uncertainty perspective. Specifically, we use the Fisher Information Matrix to evaluate the amount of evidence and re-rank samples accordingly. Samples with low evidence—those below a set threshold—are more likely to be noise or contain non-pathogenic individualized patterns, and are thus treated as potential bias sources. In **Stage 2**, we adopt a dual-branch training strategy: a biased branch is intentionally trained on low-evidence samples to capture patient-specific biases, while the unbiased branch is regularized using the Hilbert-Schmidt Independence Criterion (HSIC) to reduce reliance on those biased features and instead learn core, patient-invariant epileptic representations. To address inter-view distributional shifts introduced during bias mitigation, we further propose a cross-view bias calibration mechanism, which realigns the views and dynamically fuses their evidence at the decision level.

Our main contributions are summarized as follows: 1) We point out that the poor cross-patient generalization of existing EEG-based seizure detection models may partly result from their failure to capture seizure-indicative features and the cognitive biases introduced by patient-specific seizure patterns, and we address this by guiding the model to isolate and retain core pathological evidence; 2) We propose a novel bias-guided evidential learning framework that combines uncertainty-based filtering, dual-branch training, and intra-view debiasing to suppress non-generalizable components; 3) Extensive experiments on multiple benchmark datasets demonstrate that our BF-EML outperforms state-of-the-art patient-specific and multi-patient seizure detection methods. The results validate the effectiveness of our framework in promoting generalized seizure detection by enhancing the model's ability to capture essential seizure-related features.

2 Related Work

2.1 Patient-Specific Seizure Detection

Due to significant inter-patient variability in EEG signals, most early seizure detection methods were tailored to individual patients. Traditional statistical and machine learning approaches, such as thresholding (Li et al. 2006), SVM (Li et al. 2013), and Naive Bayes (Samiee, Kovacs, and Gabbouj 2014), offer simplicity but struggle with generalization. Deep learning models, including CNNs (Covert et al. 2019), RNNs, and attention-based BiLSTMs (Dutta et al. 2024), better capture temporal-frequency structures and have become the dominant paradigm.

To exploit the inherent multi-view nature of EEG, recent work has incorporated multi-view learning by treating EEG channels, domains, or extracted features as separate

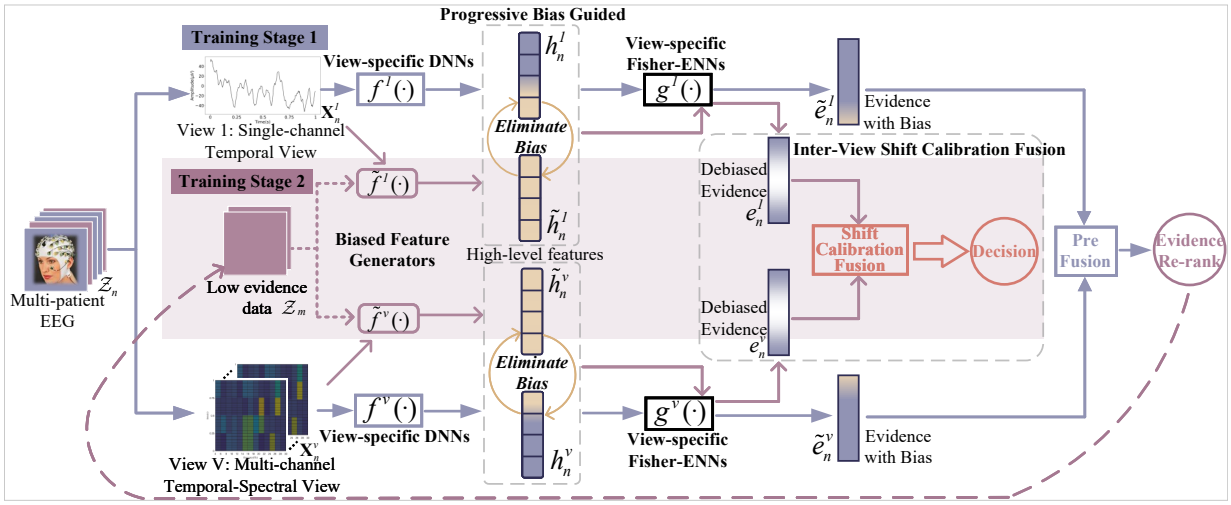

Figure 2: Overview of our two-stage training framework. In Stage 1, multi-view EEG signals are processed by view-specific DNNs to extract high-level features, which are passed to an evidential neural network with a Fisher Information Matrix to model uncertainty. Samples are re-ranked based on evidence, and low-evidence samples are selected as a biased subset. In Stage 2, a bias-prone network is trained on this subset to capture cognitive bias. All data are re-encoded to obtain bias-aware features, which are then debiased via HSIC. Finally, a calibration fusion mitigates inter-view semantic shifts caused by debiasing.

views (Tian et al. 2019; Liu et al. 2025). For instance, Cao et al. (Cao et al. 2019) proposed a stacked CNN with weighted fusion for seizure detection and preictal state classification. While effective within patients, these models tend to overfit and perform poorly when applied to unseen individuals.

2.2 Multi-patient Seizure Detection

To improve generalization, multi-patient models commonly adopt transfer learning or domain adaptation. Transfer-based approaches leverage pretrained CNNs to extract generalizable features from EEG inputs (Raghu et al. 2020; Cao et al. 2021; Chen, Han, and Debattista 2024), but often require fine-tuning for each new patient, limiting their scalability.

Domain adaptation methods attempt to bridge distribution gaps between source (labeled) and target (unlabeled) patients by learning shared feature spaces. Techniques include pseudo-labeling with conditional distribution alignment (e.g., via MMD) (Cui et al. 2023), prototype-based alignment using class centroids (Liang et al. 2023), and adversarial training with domain discriminators (Peng et al. 2022). These methods aim to produce domain-invariant yet seizure-discriminative representations. However, such strategies often prioritize aligning dominant global patterns, many of which reflect patient-specific rhythms. This focus may obscure rare yet clinically critical seizure signatures, such as transient spike discharges, which are vital for cross-patient generalization. Consequently, existing approaches struggle to capture the underlying pathological essence of seizures in a manner robust to patient heterogeneity.

3 Method

3.1 Problem Setting

Automatic epileptic seizure detection typically processes multi-channel EEG recordings by segmenting them into short samples for analysis. Let the dataset be denoted as $\boldsymbol{\mathcal{Z}} = \{\boldsymbol{Z}_n\}_{n=1}^{N}$, where each sample $\boldsymbol{Z}_n \in \mathbb{R}^{C\times T}$ represents a segment of C channels over T time points. The corresponding binary label $\boldsymbol{y}_n \in \{0, 1\}$ indicates whether the sample contains a seizure or not. A universal model is expected to generalize from the training set $\{\boldsymbol{Z}_n, \boldsymbol{y}_n\}_{n=1}^{N}$ to unseen patient data $\breve{\boldsymbol{\mathcal{Z}}}$, which may differ substantially in distribution—sometimes resembling a shift in domain. The model should accurately detect seizure activity $\breve{\boldsymbol{y}}_n$ in new patients without requiring retraining or fine-tuning.

3.2 Stage 1: Mult-view Fisher-Evidential Re-rank

Constructing Multi-View Features. The original multi-channel sequential EEG signals are temporal in nature. From these, we extract a range of features to construct multi-view representations, specifically including temporal, spectral, and temporal-spectral views. Then we design view-specific DNNs $f^v(\cdot)$ to obtain high-level multi-view features $\boldsymbol{h}_n^v$ from samples x_n^v, where v denotes the views:

$$\boldsymbol{h}_n^v = f^v(x_n^v). \tag{1}$$

View-Specific Evidential Learning. We apply Evidential Deep Learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) to the multi-view epilepsy detection setting. For the n-th sample in the v-th view, evidence is obtained via view-specific evidential DNNs $g^v(\cdot)$:

$$\boldsymbol{e}_n^v = g^v(\boldsymbol{h}_n^v). \tag{2}$$

Parameters $\boldsymbol{\alpha}^v$ are calculated as $\boldsymbol{\alpha}^v = \boldsymbol{e}^v + \mathbf{1}$. Following Subjective logic (Han et al. 2021), we define $(\boldsymbol{b}^v, u^v)$ as the opinion of views including the belief masses $\boldsymbol{b}^v$ and uncertainty masses u^v, where $\boldsymbol{b}^v = (\boldsymbol{\alpha}^v - 1)/S^v = \boldsymbol{e}^v/S^v = (b_1^v, \cdots, b_K^v)^\top$, $u^v = K/S^v$, and $S^v = \sum_{k=1}^{K} \alpha^v$ is the Dirichlet intensity.

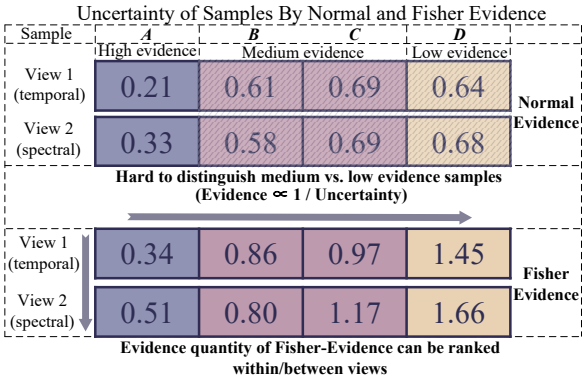

Figure 3: Comparison of uncertainty across two views for samples spanning high (A), medium (B/C), and low (D) evidence. Fisher Evidence captures relative uncertainty within and across views, while Normal Evidence over-penalizes high-uncertainty regions, leading to their blending.

Then we fuse the masses by Dempster's combination rule (Jøsang 2016a), where $C = \sum_{i \neq j} b_i^1 b_j^2$:

$$b_k^{1\diamondsuit 2} = \frac{1}{1-C}(b_k^1 b_k^2 + b_k^1 u^2 + b_k^2 u^1),\ u^{1\diamondsuit 2} = \frac{1}{1-C} u^1 u^2. \quad (3)$$

Fisher-Evidence Re-rank. To accurately identify low-evidence samples, precise quantification of evidence levels is essential. However, as shown in Fig. 3, standard EDL frameworks fail to differentiate between medium- and low-evidence samples, producing entangled evidence scores that lack discriminability both within and across views. This limitation arises from EDL's objective of aligning Dirichlet parameters with one-hot labels, which encourages the model to concentrate evidence on the ground-truth class while excessively suppressing evidence for others—even though these non-target classes may carry informative cues. In multi-view settings (Zhang et al. 2024), where each view may capture different ambiguous features, such over-penalization amplifies evidence loss across views. Consequently, the model becomes sensitive only to clearly distinguishable, high-evidence patterns, while failing to rank or exploit more ambiguous yet potentially informative samples (Xu et al. 2025).

Inspired by the work of (Deng et al. 2023), we incorporate Fisher Information (FI) into the training objective of the evidential network $g^v(\cdot)$ and propose a view-specific Fisher-evidential network to mitigate this issue. Since FI negatively correlates with evidence strength, it can serve as a natural regularizer for the model learning to preserve evidence. Additionally, multi-peaked outputs are encouraged by FI, thereby further preserving evidence across all categories.

We extend the FI matrix to multi-view learning to enable more affluent evidence learning across views and categories. Formally, the FI matrix of the Dirichlet distribution $\mathrm{Dir}(\boldsymbol{p}^v|\boldsymbol{\alpha}^v)$ is defined as:

$$\mathcal{I}(\boldsymbol{\alpha}^v) = \mathbb{E}[\frac{\partial \mathcal{D}}{\partial \boldsymbol{\alpha}^v}\frac{\partial \mathcal{D}}{\partial \boldsymbol{\alpha}^{vT}}] = \mathbb{E}[\frac{-\partial \mathcal{D}^2}{\partial \boldsymbol{\alpha}^v \boldsymbol{\alpha}^{vT}}], \quad (4)$$

where $\mathcal{D} = logDir(\boldsymbol{p}^v|\boldsymbol{\alpha}^v)$, and $\boldsymbol{p}^v = (p_1^v, \cdots, p_k^v)^\top$ is the probability that the instance is assigned to k-th category.

Instead of predicting a deterministic category label, we sample $\boldsymbol{p}^v \sim \mathrm{Dir}(\boldsymbol{\alpha}^v)$ to generate a category probability vector. For the observed one-hot label $\boldsymbol{y}$, it is assumed to follow a multi-variate Gaussian distribution with mean $\boldsymbol{p}^v$ and covariance matrix scaled by the inverse of the FI matrix, i.e., $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{p}^v, (\sigma^v)^2 \mathcal{I}(\boldsymbol{\alpha}^v)^{-1})$. This dynamic covariance matrix enables categories with lower certainty (higher FI) to be modeled with larger variance, allowing each view-specific network to retain as much category evidence as possible to support more accurate evidence-based ranking in subsequent steps. The learning objective is defined as:

$$\min_{\boldsymbol{\theta}^v} \mathbb{E}[-log\mathcal{N}(\boldsymbol{y}|\boldsymbol{p}^v, (\sigma^v)^2 \mathcal{I}((\boldsymbol{\alpha}^v)^{-1}))], \quad (5)$$

where $\boldsymbol{\theta}^v$ are parameters of $g^v(\cdot)$.

For sample $\{x_n^v\}_{v=1}^V$, we employ a Fisher evidence loss which is derived from the FI matrix (i.e., Eq. 5):

$$\mathcal{L}_F(\boldsymbol{\alpha}_n^v) = \sum_{k=1}^{K} \left[(y_{nk} - \frac{\alpha_{nk}^v}{S_n^v})^2 + \frac{\alpha_{nk}^v (S_n^v - \alpha_{nk}^v)}{(S_n^v)^2 (S_n^v + 1)} \right] \psi^{(1)}(\alpha_{nk}^v) - \lambda \log |\mathcal{I}(\boldsymbol{\alpha}_n^v)|, \quad (6)$$

where $|\cdot|$ indicates the determinant of a matrix, which measures the overall information volume in the distribution. The trigamma function $\psi^{(1)}(x) = d^2 \ln \Gamma(x)/dx^2$ is a monotonic decreasing function for $x > 0$ (where $\Gamma(\cdot)$ is the gamma function), which means that higher-evidence categories receive less penalty and more evidence are preserved.

3.3 Stage 2: Bias Mitigation

Dual-Branch Alternating Debiasing Strategy. Our objective is to guide the model toward learning pathogenic epileptic features. As previously discussed, a major source of bias in epilepsy detection arises when the network over-relies on patient-specific features, which obscures the underlying pathological patterns. Drawing inspiration from the concept of "bias correction via bias amplification" (Bahng et al. 2020), we design a tailored debiasing strategy that aligns with the nature of bias in epileptic EEGs, where unreliable regions often correspond to patient-specific signals.

To this end, we introduce a debiasing method that reduces reliance on low-evidence samples. We have employed a FI-based evidential network to assess the reliability (i.e., evidence strength) of each sample. Since the network is trained on data from all patients, features that appear consistently across patients yield higher evidence, while patient-specific patterns typically induce higher uncertainty and lower evidence. Leveraging this property, we identify low-evidence samples and use them to train a deliberately biased network $\tilde{f}(\cdot)$, which is expected to encode patient-specific characteristics. This network serves as the biased branch to generate features $\tilde{\boldsymbol{h}}_n^v$ for all inputs $\{x_n^v\}_{v=1}^V$.

To mitigate cognitive bias arising from patient-specific patterns, we enforce statistical independence between the outputs of the biased network $\tilde{f}(\cdot)$ and the unbiased network $f(\cdot)$. We adopt the Hilbert-Schmidt Independence Criterion (HSIC) to quantify and minimize their dependency. Unlike conventional correlation-based metrics, HSIC captures all

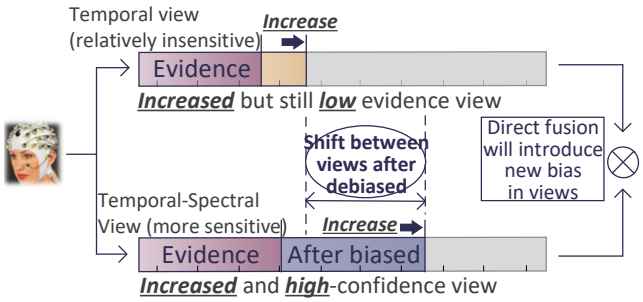

Figure 4: Illustration of Inter-view Shift. Different EEG views may naturally express varied opinions. Bias guidance amplifies this divergence: the Temporal view, limited by modality, remains low-evidence, while the T-S view, enriched by learned pathological features, becomes high-evidence. This leads to a semantic shift as decisions diverge.

forms of dependence—linear and nonlinear—and guarantees $H(U,V)=0$ if and only if U and V are statistically independent. This constraint ensures that the unbiased network focuses on patient-invariant, disease-relevant features, without inadvertently reusing biased patterns.

To ensure effective debiasing, we adopt a min-max optimization strategy between the two branches. In the unbiased branch, we freeze the biased network $\tilde{f}(\cdot)$ and train the unbiased network $f(\cdot)$ to minimize HSIC, thereby reducing its dependency on biased features:

$$\mathcal{L}_H(x_n^v, \tilde{\boldsymbol{h}}_n^v)=\mathcal{L}_F(g^v(f^v(x_n^v))+1)+\mu_1 H(f^v(x_n^v), \tilde{\boldsymbol{h}}_n^v). \quad (7)$$

However, if optimized unilaterally, the biased network may gradually degrade—outputting low-variance, non-informative features that fail to retain patient-specific bias. This weakens its role as a counterfactual reference. Therefore, in the biased branch, we reverse the setup: we freeze the unbiased network $f(\cdot)$ and update $\tilde{f}(\cdot)$ to maximize HSIC, encouraging it to maintain distinguishable biased features:

$$\tilde{\mathcal{L}}_H(x_n^v, \boldsymbol{h}_n^v)=\mathcal{L}_F(g^v(\tilde{f}^v(x_n^v))+1)-\mu_2 H(\tilde{f}^v(x_n^v), \boldsymbol{h}_n^v). \quad (8)$$

Through alternating optimization of both branches, we progressively decouple biased and unbiased representations, enabling the unbiased branch to better capture patient-invariant features essential for generalizable detection.

Inter-view Shift Calibration Fusion. We have introduced a bias-guiding mechanism to steer each view toward learning pathogenic epileptic patterns. However, due to heterogeneous learning difficulties, view-specific networks converge at different rates, resulting in unequal shifts across views.

This imbalance leads to inter-view conflicts. As shown in Fig. 4, one view may successfully capture epileptic signatures, producing strong evidence for a class, while another still provides weak support. Since uncertainty u is inversely related to evidence e (Jøsang 2016b), the resulting uncertainty gap causes low-uncertainty views to dominate during naive fusion, compromising multi-view complementarity.

Moreover, asymmetric evidence growth may cause class preference shifts in one view while others remain stable, introducing semantic inconsistency beyond covariate shift. These semantic shifts hinder effective fusion and may degrade overall performance. To mitigate this, we adopt a variant of Averaging Belief Fusion (Chen et al. 2025):

$$b^{1\underline{\Diamond}2}=\frac{b^1u^2+b^2u^1}{u^1+u^2}, \quad u^{1\underline{\Diamond}2}=\frac{2u^1u^2}{u^1+u^2}, \quad (9)$$

unlike traditional fusion schemes (Liang et al. 2021) where uncertainty monotonically decreases, this formulation penalizes unreliable views, yielding more balanced results.

To further calibrate inter-view shift, we quantify the inter-view offset $\mathcal{M}$ using two metrics derived from the Dirichlet distribution: the base misalignment $\mathcal{M}_B$, which measures class-wise support inconsistency; and the scale alignment term $\mathcal{M}_S$, which captures concentration mismatch:

$$\mathcal{M}_B(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)=\sum_{k=1}^{K}\left(\frac{\alpha_k^1}{S^1}-\frac{\alpha_k^2}{S^2}\right)^2, \quad (10)$$

where α_k/S represents the normalized belief mass for class k, $\mathcal{M}_B$ quantifies the discrepancy in belief allocation between views.

$$\mathcal{M}_S(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)=1-\frac{K}{S^1}-\frac{K}{S^2}+\frac{K^2}{S^1S^2}, \quad (11)$$

where larger S indicates higher confidence (i.e., greater concentration). This joint term penalizes fusion between uncertain views and favors alignment among confident ones.

We define $\mathcal{M}=\mathcal{M}_B\mathcal{M}_S$ and minimize the average pairwise offset across all views ($i\neq j$):

$$\mathcal{L}_M=\frac{1}{V-1}\sum_{i=1}^{V}\sum_{j=1}^{V}\mathcal{M}_B(\boldsymbol{\alpha}^i, \boldsymbol{\alpha}^j)\mathcal{M}_S(\boldsymbol{\alpha}^i, \boldsymbol{\alpha}^j). \quad (12)$$

Training Objectives. We convert the traditional DNN into an evidential DNN by replacing the softmax with a non-negative activation (e.g., ReLU), treating its output as evidence. However, the loss in Eq. 6 does not explicitly suppress evidence for incorrect labels. To address this, we introduce a KL divergence-based regularizer:

$$\mathcal{L}_{KL}(\boldsymbol{\alpha}^v)=KL\left[D(\boldsymbol{p}^v|\tilde{\boldsymbol{\alpha}}^v)\parallel D(\boldsymbol{p}^v|\mathbf{1})\right], \quad (13)$$

where $\tilde{\boldsymbol{\alpha}}^v=\boldsymbol{y}_n+(\mathbf{1}-\boldsymbol{y}_n)\odot\boldsymbol{\alpha}_n^v$ is the Dirichlet parameters after the removal of non-misleading evidence from predicted parameters $\boldsymbol{\alpha}_n$, and $Dir(\boldsymbol{p}^v|\mathbf{1})$ is the uniform distribution.

Finally, the overall loss can be calculated as:

$$\mathcal{L}(\boldsymbol{\alpha}_n^v)=\mathcal{L}_H(x_n^v, \tilde{\boldsymbol{h}}_n^v)+\gamma\mathcal{L}_{KL}(\boldsymbol{\alpha}_n^v). \quad (14)$$

In summary, we employ a multitasking strategy to supervise both view-specific and aggregated opinions:

$$\mathcal{L}=\sum_{v=1}^{V}\mathcal{L}(\boldsymbol{\alpha}_n^v)+\mathcal{L}(\boldsymbol{\alpha}_n^{(v_1\Diamond v_2\Diamond\cdots\Diamond v_V)})+\beta\mathcal{L}_M. \quad (15)$$

4 Experiments

4.1 Experimental Setup

Datasets[1][2]**. CHB-MIT dataset** is a widely used benchmark from Boston Children's Hospital, containing EEG

[1]https://physionet.org/content/chbmit/1.0.0/

[2]https://physionet.org/content/siena-scalp-eeg/1.0.0/

Datasets	Metric	Methods						
		SVM	CABLNet	AMV-DFL	Hybrid-Trans	MIP-TRL-FS	JPDDA	BF-EML
CHB-MIT *Universal-New*	Accuracy	50.39±8.35	66.82±6.31	70.93±4.03	89.10±4.75	90.14±1.48	93.79±2.11	**97.65±1.43**
	Sensitivity	56.13±9.11	69.24±6.57	60.99±7.37	87.79±4.00	81.57±5.28	90.91±3.46	**95.76±2.38**
	Specificity	52.44±4.36	78.08±3.82	82.62±4.25	90.49±3.93	95.06±2.21	94.38±2.33	**98.66±2.14**
	F1-Score	47.16±4.36	66.77±4.69	64.28±5.36	86.24±3.60	86.18±4.62	90.84±2.18	**96.67±1.67**
Siena *Universal-New*	Accuracy	49.64±8.07	64.53±5.13	66.24±3.56	86.38±2.64	89.85±2.31	88.41±3.03	**93.74±1.69**
	Sensitivity	53.37±6.45	65.28±4.68	61.72±6.09	85.29±3.01	84.56±3.93	86.60±2.85	**90.48±2.71**
	Specificity	55.24±5.91	72.47±3.75	77.55±4.81	90.88±3.34	90.64±2.82	88.73±2.46	**93.29±2.55**
	F1-Score	46.59±4.45	61.42±3.78	61.93±3.12	85.11±2.01	84.33±2.75	84.81±2.89	**89.56±1.94**
CHB-MIT *Universal-Old*	Accuracy	60.48±5.15	82.36±3.08	88.07±2.59	95.65±1.99	97.48±1.06	97.12±0.62	**99.02±0.51**
	Sensitivity	65.57±6.22	86.05±2.83	83.48±4.10	89.76±2.43	93.52±3.13	92.46±1.84	**98.15±1.33**
	Specificity	68.15±4.63	84.28±3.79	88.25±2.80	97.28±1.75	96.23±1.47	94.67±1.79	**99.28±0.42**
	F1-Score	60.38±4.90	81.75±3.14	83.15±2.96	92.49±1.51	93.83±1.35	92.07±1.14	**98.43±0.69**
Siena *Universal-Old*	Accuracy	57.36±3.42	82.67±2.96	84.39±3.13	92.20±1.33	94.08±1.85	95.74±2.48	**97.81±1.03**
	Sensitivity	63.27±4.82	84.43±2.34	81.58±3.39	88.47±3.48	92.62±2.82	91.65±2.05	**95.24±1.92**
	Specificity	62.08±3.65	82.72±3.11	86.50±2.70	93.06±1.13	91.33±0.60	92.49±1.34	**98.39±0.75**
	F1-Score	56.10±3.02	78.88±3.18	80.08±2.67	88.32±1.35	89.29±1.03	89.66±1.92	**96.23±1.57**
CHB-MIT *Single*	Accuracy	72.57±3.24	98.83±0.22	99.01±0.15	91.12±1.13	95.34±2.69	92.72±3.18	**99.73±0.19**
	Sensitivity	66.74±4.90	98.77±0.18	97.39±1.04	86.63±2.28	90.91±0.95	89.51±1.61	**99.32±0.45**
	Specificity	75.43±3.26	99.05±0.38	99.21±0.08	94.45±1.47	92.03±1.52	91.39±1.27	**99.87±0.11**
	F1-Score	64.60±3.75	98.57±0.29	98.02±0.61	88.36±1.60	89.15±1.37	88.05±1.56	**99.54±0.23**
Siena *Single*	Accuracy	70.94±3.16	98.03±0.93	98.46±0.58	89.94±2.21	94.18±1.87	89.13±1.48	**98.81±0.77**
	Sensitivity	63.37±5.29	94.67±1.44	94.03±1.56	86.38±2.07	88.62±2.35	89.36±1.74	**99.32±0.45**
	Specificity	71.88±4.43	98.12±0.69	97.82±0.55	90.30±1.92	92.91±1.85	90.57±1.82	**98.50±0.94**
	F1-Score	60.43±2.98	95.10±0.87	95.81±1.12	85.26±2.01	88.25±1.45	86.97±1.31	**98.38±0.78**

Table 1: Classification accuracy(%) of BF-EML and baseline methods on the datasets in Universal-New, Universal-old, and Single Settings . The best and the second best results are highlighted by **boldface** and underlined respectively.

Datasets	Methods			Metrics			
	$\mathcal{L}_F$	$\mathcal{L}_H$	$\mathcal{L}_M$	Accuracy	Sensitivity	Specificity	F1-Score
Universal *-new*	✓	-	-	83.76±3.82	79.25±3.38	85.85±2.90	84.73±2.28
	-	✓	-	89.08±3.96	85.49±2.87	89.53±3.09	84.71±3.20
	✓	✓	-	94.13±1.83	92.84±2.61	95.08±2.57	92.35±2.08
	✓	✓	✓	**97.65±1.43**	**95.76±2.38**	**98.66±2.14**	**96.67±1.67**
Universal *-old*	✓	-	-	91.14±1.48	89.03±2.67	93.43±1.85	89.00±2.58
	-	✓	-	92.35±1.48	90.86±1.89	92.80±1.14	90.08±1.75
	✓	✓	-	96.58±0.96	95.62±1.90	98.05±1.32	96.09±1.88
	✓	✓	✓	**99.02±0.51**	**98.15±1.33**	**99.28±0.42**	**98.43±0.69**
Single	✓	-	-	97.39±0.64	95.81±1.31	98.13±0.74	96.27±1.48
	-	✓	-	91.06±1.01	88.95±1.72	92.87±0.99	88.69±1.22
	✓	✓	-	99.06±0.37	98.58±0.64	99.19±0.28	98.59±0.59
	✓	✓	✓	**99.73±0.19**	**99.32±0.45**	**99.87±0.11**	**99.54±0.23**

Table 2: Ablation performance across settings on CHB-MIT.

recordings from 24 patients. We used 23 common channels across patients. **Siena dataset** comprises 14 patients with specific epilepsy types, supporting evaluation across seizure categories. A consistent set of 29 standard channels was used. **Clinical dataset** was collected with partner hospitals, comprising under-5-minute segments from 523 patients.

Universal Model Building Strategy. For CHB-MIT, we randomly selected 18 patients for training and used their held-out segments to test performance on seen patients, while the remaining 5 evaluated generalization to unseen patients. Five such splits were constructed. For Siena, we used 11 for training and 3 as unseen patients, forming four splits.

Compared Methods. A. **Patient-Specific Methods**. (1) SVM (Cortes 1995): classical machine learning baseline using hyperplane-based classification. (2) CABLNet (Dutta et al. 2024): single-view BiLSTM with multi-head attention to enhance temporal EEG analysis. (3) AMV-DFL (Ahmad et al. 2024): multi-view model combining rule-based and deep learning components. B. **Multi-Patient Methods**. (1) MIP-TRL-FS (Li et al. 2023): uses multi-view fuzzy transfer learning to reduce domain discrepancy. (2) Hybrid-Trans (Hu et al. 2023): combines Hybrid Transformer with fine-tuning for cross-patient adaptation. (3) JPDDA (Cui et al. 2023): aligns joint distributions of source and pseudo-labeled target domains.

4.2 Experimental Results

Performance Comparison. We evaluate three types of methods under three experimental settings: **Method types:** (1) *Single-patient models*, trained and tested on data from the same patient; (2) *Multi-patient models*, trained on multiple patients, including both inference-based and transfer-based methods; and (3) *Our proposed method*, which is an inference-based multi-patient model requiring no patient-specific fine-tuning. **Evaluation settings:** *Universal-New* evaluates generalization to unseen patients: models are trained on multiple patients and tested on entirely new patients excluded from training. For fairness, transfer-based models are allowed to fine-tune a separate model for each target patient using their unlabeled EEG data, while inference-based methods (including ours) use a single model trained on the training set and applied directly to all test patients. For CHB-MIT, we report the mean performance over five random patient groupings; for Siena, we average results across four constructed datasets. *Universal-Old* assesses model performance on patients seen during training, using disjoint data segments for training and testing. *Single* follows the traditional patient-specific setting, where each model is trained and tested using data from the same individual, with no overlap in data segments. All models are evaluated using consistent protocols within each setting.

Analysis. Table 1 reveals the following key observations: (1) Single-patient methods excel in personalized settings, while multi-patient models perform better under universal

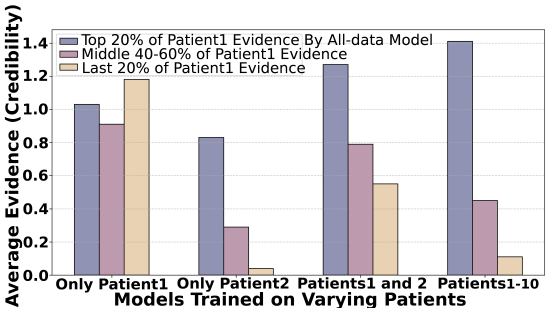

Figure 5: The data for Patient 1 is grouped based on evidence assessed by the model from all patients. Each group is then evaluated for evidence by models trained on different patients, with the mean evidence calculated for each group.

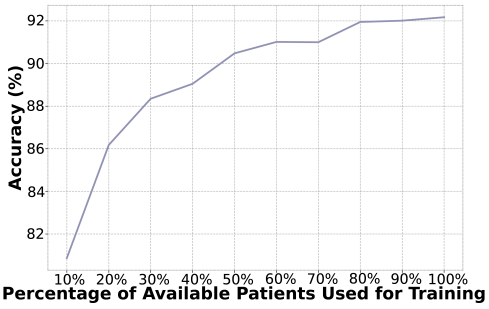

Figure 6: Accuracy trends as 10% of the 400 training patients are incrementally added each round.

conditions with diverse data; (2) Single-patient models suffer significant drops in performance under universal settings, especially in sensitivity, even when test patients are seen during training; (3) Overall performance is lower on the Siena dataset than CHB-MIT, likely due to greater clinical heterogeneity and patient variability; (4) Our proposed BF-EML consistently performs well across all settings. Notably, in the challenging universal-new scenario, it maintains strong accuracy and sensitivity, crucial for minimizing missed diagnoses and enabling early detection.

4.3 Ablation Study

We perform ablation studies to assess the contribution of each component in BF-EML (Table 2): (1) Using only $\mathcal{L}_F$ denotes a multi-view network trained solely with the fusion strategy in Eq. 3. (2) Using only $\mathcal{L}_H$ removes the multi-view structure, retaining only the temporal-spectral view with the two-stage evidence filtering and bias elimination. (3) Combining $\mathcal{L}_F$ and $\mathcal{L}_H$ disables the shift alignment module $\mathcal{L}_M$, while the full BF-EML includes all three components.

The results show: (1) $\mathcal{L}_F$ captures complementary features across views and performs well in single-patient settings, but generalization to unseen patients remains limited. (2) $\mathcal{L}_H$, though based on a single view, improves robustness to new patients by extracting more generalizable, pathology-relevant features, albeit with weaker performance on single-patient tasks. (3) Adding $\mathcal{L}_M$ further stabilizes performance and enhances reliability across all evaluation settings.

4.4 Evidence-Based Group Analysis

To examine the relationship between uncertainty and pathological relevance, we performed an analysis on the CHB-MIT dataset (Fig. 5). A stage-1 evidence-ranking model was first pretrained using data from all 24 patients. We then input EEG segments from Patient 1 into this model and ranked them by the output evidence scores. Based on the ranking, we selected three groups: the top 20% (purple; high evidence, low uncertainty), the middle 40% (pink; moderate uncertainty), and the bottom 20% (yellow; low evidence, high uncertainty). To evaluate the generalizability of these segments, we used four models trained on increasingly diverse patient data: from Patient 1 only, Patient 2 only, Patients 1 and 2, and Patients-1–10. High-evidence segments consistently yielded high scores across all models, suggesting the presence of shared seizure-related patterns. In contrast, low-evidence segments retained high scores only in the Patient 1-specific model and exhibited substantial performance drops in all others, reflecting strong patient specificity.

These findings suggest that high-evidence segments encode universal epileptic features, likely reflecting core neurophysiological mechanisms, while low-evidence segments primarily capture non-generalizable or noisy patterns. Emphasizing high-evidence regions during learning thus supports the development of more robust and transferable EEG-based seizure detection.

4.5 Clinical Dataset

Our clinical dataset features a large number of patients, each contributing only a short seizure segment—reflecting real-world scenarios with high patient diversity and limited individual data. To assess scalability, we split the dataset into 400 training patients (200 seizure, 200 non-seizure) and 123 test patients (77 seizure, 46 non-seizure). We performed ten training rounds, incrementally adding 40 new patients (20 seizure, 20 non-seizure) per round and tracking accuracy. As shown in Fig. 6, performance consistently improved with more patients, confirming the scalability of our method.

Unlike single-patient models that fail to generalize, and conventional multi-patient methods requiring source-target adaptation, our method scales to large patient populations without retraining, making it practical for deployment.

5 Conclusion

In this work, we identify the poor cross-patient generalization of existing EEG-based epilepsy detection methods as a result of cognitive bias caused by patient-specific idiosyncrasies, which obscure the learning of core seizure patterns. To address this, we propose BF-EML, a bias-guided evidential multi-view framework that quantifies uncertainty via Fisher Information, detects and filters low-evidence (biased) segments, and employs an alternating dual-branch training strategy to isolate and suppress patient-specific features. Extensive experiments show that BF-EML achieves competitive performance on seen patients and significantly improves generalization to unseen ones, advancing toward universal EEG-based seizure detection.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62425605, 62273275, 62133012, 62502051, 62572375 and 62472340, in part by the Key Research and Development Program of Shaanxi under Grants 2025GH-YBXM-018, 2025CY-YBXM-041, 2024GXYBXM-122, and 2022ZDLGY01-10, and in part by Natural Science Basic Research Program of Shaanxi(Program Nos. 2025JC-QYXQ-040, 2025JC-YBQN-900), and in part by Xidian University Specially Funded Project for Interdisciplinary Exploration (TZJHF202506), and in part by Postdoctoral Fellowship Program of CPSF under Grant Number GZC20251101.

References

Ahmad, I.; Liu, Z.; Li, L.; Ullah, I.; Aboyeji, S. T.; Wang, X.; Samuel, O. W.; Li, G.; Tao, Y.; Chen, Y.; et al. 2024. Robust Epileptic Seizure Detection Based on Biomedical Signals Using an Advanced Multi-View Deep Feature Learning Approach. *IEEE Journal of Biomedical and Health Informatics*.

Bahng, H.; Chun, S.; Yun, S.; Choo, J.; and Oh, S. J. 2020. Learning de-biased representations with biased representations. In *International Conference on Machine Learning*, 528–539. PMLR.

Cao, J.; Hu, D.; Wang, Y.; Wang, J.; and Lei, B. 2021. Epileptic classification with deep-transfer-learning-based feature fusion algorithm. *IEEE transactions on cognitive and developmental systems*, 14(2): 684–695.

Cao, J.; Zhu, J.; Hu, W.; and Kummert, A. 2019. Epileptic signal classification with deep EEG features by stacked CNNs. *IEEE Transactions on Cognitive and Developmental Systems*, 12(4): 709–722.

Chen, C.; Han, J.; and Debattista, K. 2024. Virtual category learning: A semi-supervised learning method for dense prediction with extremely limited labels. volume 46, 5595–5611. IEEE.

Chen, H.; Xu, C.; Guan, Z.; Zhao, W.; and Liu, J. 2025. Biased Incomplete Multi-View Learning. 39(15): 15767–15775.

Cortes, C. 1995. Support-Vector Networks. *Machine Learning*.

Covert, I. C.; Krishnan, B.; Najm, I.; Zhan, J.; Shore, M.; Hixson, J.; and Po, M. J. 2019. Temporal graph convolutional networks for automatic seizure detection. In *Machine learning for healthcare conference*, 160–180. PMLR.

Cui, X.; Wang, T.; Lai, X.; Jiang, T.; Gao, F.; and Cao, J. 2023. Cross-Subject Seizure Detection by Joint-Probability-Discrepancy-Based Domain Adaptation. *IEEE Transactions on Instrumentation and Measurement*, 72: 1–13.

Deng, D.; Chen, G.; Yu, Y.; Liu, F.; and Heng, P.-A. 2023. Uncertainty estimation by fisher information-based evidential deep learning. In *International Conference on Machine Learning*, 7596–7616. PMLR.

Ding, D.; Wang, W.; Wu, J.; Ma, G.; Dai, X.; Yang, B.; Wang, T.; Yuan, C.; Hong, Z.; de Boer, H. M.; et al. 2006. Premature mortality in people with epilepsy in rural China: a prospective study. *The Lancet Neurology*, 5(10): 823–827.

Dissanayake, T.; Fernando, T.; Denman, S.; Sridharan, S.; and Fookes, C. 2021. Geometric deep learning for subject independent epileptic seizure prediction using scalp EEG signals. *IEEE Journal of Biomedical and Health Informatics*, 26(2): 527–538.

Dutta, A. K.; Raparthi, M.; Alsaadi, M.; Bhatt, M. W.; Dodda, S. B.; Sandhu, M.; and Patni, J. C. 2024. Deep learning-based multi-head self-attention model for human epilepsy identification from EEG signal for biomedical traits. *Multimedia Tools and Applications*, 1–23.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2021. Trusted Multi-View Classification. In *International Conference on Learning Representations*.

Hu, S.; Liu, J.; Yang, R.; Wang, Y.; Wang, A.; Li, K.; Liu, W.; and Yang, C. 2023. Exploring the applicability of transfer learning and feature engineering in epilepsy prediction using hybrid transformer model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 1321–1332.

Jøsang, A. 2016a. A formalism for reasoning under uncertainty. *Artificial Intelligence: Foundations, Theory, and Algorithms, Springer*, 10: 978–3.

Jøsang, A. 2016b. *Subjective logic*, volume 3. Springer.

Li, A.; Deng, Z.; Zhang, W.; Xiao, Z.; Choi, K.-S.; Liu, Y.; Hu, S.; and Wang, S. 2023. Multiview Transfer Representation Learning with TSK Fuzzy System for EEG Epilepsy Detection. *IEEE Transactions on Fuzzy Systems*, 32(1): 38–52.

Li, S.; Zhou, W.; Yuan, Q.; Geng, S.; and Cai, D. 2013. Feature extraction and recognition of ictal EEG using EMD and SVM. *Computers in biology and medicine*, 43(7): 807–816.

Li, X.; Chu, M.; Qiu, T.; and Bao, H. 2006. A method of epileptic EEG spike detection based on time-frequency analysis. *Chinese Journal of Biomedical Engineering*, 25(6): 678.

Li, Y.; Liu, Y.; Cui, W.-G.; Guo, Y.-Z.; Huang, H.; and Hu, Z.-Y. 2020. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(4): 782–794.

Liang, D.; Liu, A.; Gao, Y.; Li, C.; Qian, R.; and Chen, X. 2023. Semi-supervised domain-adaptive seizure prediction via feature alignment and consistency regularization. *IEEE Transactions on Instrumentation and Measurement*, 72: 1–12.

Liang, X.; Qian, Y.; Guo, Q.; Cheng, H.; and Liang, J. 2021. AF: An association-based fusion method for multi-modal classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9236–9254.

Liu, Y.; Xu, C.; Wen, Z.; and Dong, Y. 2025. Trust EEG epileptic seizure detection via evidential multi-view learning. *Information Sciences*, 694: 121699.

Organization, W. H.; et al. 2019. *Epilepsy: a public health imperative*. World Health Organization.

Peng, P.; Xie, L.; Zhang, K.; Zhang, J.; Yang, L.; and Wei, H. 2022. Domain adaptation for epileptic EEG classification using adversarial learning and Riemannian manifold. *Biomedical Signal Processing and Control*, 75: 103555.

Raghu, S.; Sriraam, N.; Temel, Y.; Rao, S. V.; and Kubben, P. L. 2020. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Networks*, 124: 202–212.

Samiee, K.; Kovacs, P.; and Gabbouj, M. 2014. Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform. *IEEE transactions on Biomedical Engineering*, 62(2): 541–552.

Scheffer, I. E.; Berkovic, S.; Capovilla, G.; Connolly, M. B.; French, J.; Guilhoto, L.; Hirsch, E.; Jain, S.; Mathern, G. W.; Moshé, S. L.; et al. 2017. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*, 58(4): 512–521.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31.

Tian, X.; Deng, Z.; Ying, W.; Choi, K.-S.; Wu, D.; Qin, B.; Wang, J.; Shen, H.; and Wang, S. 2019. Deep multi-view feature learning for EEG-based epileptic seizure detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(10): 1962–1972.

Xu, C.; Wen, Z.; Zhao, J.; Zhao, W.; Yu, J.; Chen, H.; Guan, Z.; and Zhao, W. 2025. Beyond Equal Views: Strength-Adaptive Evidential Multi-View Learning. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 1278–1287.

Zhang, Q.; Wei, Y.; Han, Z.; Fu, H.; Peng, X.; Deng, C.; Hu, Q.; Xu, C.; Wen, J.; Hu, D.; et al. 2024. Multimodal fusion on low-quality data: A comprehensive survey. *arXiv preprint arXiv:2404.18947*.