

Received 19 April 2026, accepted 13 May 2026, date of publication 19 May 2026, date of current version 27 May 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3694985

RESEARCH ARTICLE

EEG-Based Cross-Patient Epileptic Seizure Detection via State-Conditional Domain Adversarial Training and Temporal Modeling

RINA TAZAKI¹, (Graduate Student Member, IEEE), TOMOYUKI AKIYAMA²,
AND AKIRA FURUI¹, (Member, IEEE)

¹Graduate School of Advanced Science and Engineering, Hiroshima University, Higashihiroshima 739-8521, Japan

²Department of Pediatric Neurology, Okayama University Hospital, Okayama 700-8558, Japan

Corresponding authors: Rina Tazaki (rinatazaki@hiroshima-u.ac.jp) and Akira Furui (akirafurui@hiroshima-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP23K28128.

ABSTRACT Automated epileptic seizure detection from scalp electroencephalogram (EEG) remains challenging due to significant inter-patient variability. While domain adversarial training has emerged as a promising approach for patient-invariant features, existing methods typically rely on convolutional neural network (CNN)-based models that process each EEG window independently, failing to capture temporal progression patterns of seizures. Moreover, conventional domain adversarial training enforces patient invariance uniformly without distinguishing between seizure and non-seizure states, despite inter-patient variability differing across these states. To address these limitations, we propose a domain generalization framework that integrates state-conditional domain adversarial training with temporal modeling via a CNN and bidirectional long short-term memory (BiLSTM), using only source patient data without any target patient enrollment. The two objectives are decoupled into separate stages to avoid conflicting optimization. In the first stage, the CNN extracts state-adaptive patient-invariant features by conditional domain adversarial training. In the second stage, the BiLSTM captures temporal dependencies of these invariant features to model seizure evolution. We evaluated our method using leave-one-patient-out cross-validation on two focal epilepsy datasets: 14 patients from the Siena Scalp EEG Database (long-term recordings) and 20 from a clinical dataset (short-term recordings). The results demonstrate that both conditional domain adversarial training and temporal modeling contribute to performance improvement, with the benefit of state-adaptive invariant feature extraction particularly pronounced in long-term recordings containing diverse background EEG activities. This approach enables immediate application to unseen patients without any enrollment period, providing a clinically practical solution for automated seizure monitoring.

INDEX TERMS Electroencephalogram (EEG), epileptic seizure detection, domain adversarial training, domain generalization, cross-patient, temporal modeling.

I. INTRODUCTION

Epilepsy is a neurological disease caused by abnormal neuronal electrical activity, affecting approximately 50 million people worldwide [1]. This disease is characterized by transient and recurrent episodes called epileptic seizures. Since these seizures often involve convulsions and loss of

consciousness, they pose various risks, including serious injuries [2], [3]. Furthermore, epilepsy is known to cause cognitive decline [4], [5], severely affecting patients' quality of life. Given this background, accurate and rapid detection of epileptic seizures is crucial for appropriate therapeutic intervention and patient safety.

Electroencephalogram (EEG) recordings obtained from electrodes placed on the scalp surface are commonly used for epileptic seizure detection. However, since seizures are rarely

The associate editor coordinating the review of this manuscript and approving it for publication was Md Kafiul Islam¹.

captured during routine EEG examinations (typically 20–30 minutes), long-term video-EEG monitoring over several days is required for a definitive diagnosis [6], [7]. Visual inspection of these prolonged recordings by specialists is time-consuming and labor-intensive, presenting a significant burden on medical resources.

To reduce this clinical burden, machine learning-based automated seizure detection methods have been developed. These methods are broadly categorized into two approaches: patient-specific and cross-patient seizure detection. Patient-specific approaches train models individually using historical seizure data from each patient, achieving high detection accuracy [8], [9], [10], [11]. However, since sufficient seizure data must be collected in advance, these approaches have an inherent limitation: they cannot be applied to new patients or patients with low seizure frequency. In contrast, cross-patient approaches apply models trained on data from multiple known patients (source patients) to unknown new patients (target patients), offering high practicality due to their immediate clinical applicability [11], [12], [13], [14], [15].

In cross-patient seizure detection, domain shift caused by inter-patient variability in EEG patterns is a fundamental challenge. Significant individual differences exist in the amplitude, frequency characteristics, and spatial patterns of EEG signals during both seizure and non-seizure states. This variability causes models trained on source patients to exhibit degraded detection performance on target patients. To address the domain shift problem, methods applying domain adversarial training have recently attracted attention [16], [17], [18], [19], [20]. In particular, the domain generalization framework, which acquires patient-independent (patient-invariant) features solely from source patient data, eliminates the need for any target patient enrollment or data collection prior to deployment, making it the most clinically practical formulation for cross-patient seizure detection [18], [19], [20].

Existing domain generalization methods have two primary limitations. First, many approaches rely on convolutional neural network (CNN)-based models that process each EEG window independently, failing to capture the temporal evolution of seizures. Since epileptic seizures typically progress over tens of seconds to several minutes, this lack of temporal modeling inherently constrains detection performance. Second, conventional domain adversarial training enforces uniform patient invariance without distinguishing between seizure and non-seizure states. However, inter-patient variability may differ between these states. For example, non-seizure EEG is strongly influenced by background activity (e.g., alpha rhythms, sleep waveforms), whereas seizure EEG is dominated by abnormal discharge patterns, suggesting that patient characteristics influence the feature space differently depending on the state. Uniform patient invariance that ignores such state-dependent structures may hinder effective optimization of feature representations.

To address the first limitation, our previous work [21] proposed a preliminary framework that extended domain-adversarial neural networks (DANN) [22] to domain generalization and integrated it with temporal modeling. This method extracted local patient-invariant features through domain adversarial training applied to a CNN, and modeled their temporal dependencies using bidirectional long short-term memory (BiLSTM) [23], enabling the model to capture temporal patterns while maintaining patient invariance. However, this method acquired patient invariance uniformly without distinguishing between seizure and non-seizure states, leaving the second limitation unresolved.

This paper extends the preliminary framework in terms of both methodology and experimentation. Methodologically, we introduce the concept of conditional domain adversarial networks (CDAN) [24] to extract patient-invariant features while considering the characteristics of seizure and non-seizure states. By conditioning domain adversarial training on the predicted seizure state, the model can reduce patient-specific characteristics within each state independently, rather than forcing a single invariant representation across both states. Furthermore, entropy-based weighting suppresses the influence of ambiguous samples near the state boundary. Experimentally, in addition to the short-term recordings used in our previous study, we conduct evaluations using long-term recordings that reflect realistic clinical settings.

The main contributions of this study are as follows:

- Application of domain adversarial training conditioned on seizure and non-seizure states to cross-patient epileptic seizure detection. This enables the adaptive extraction of patient-invariant features specific to each state.
- Construction of a two-stage training framework that models temporal dependencies of patient-invariant features using BiLSTM. This allows the model to capture temporal progression patterns of seizures while maintaining patient invariance.
- Comprehensive evaluation using an open dataset (long-term) and clinical EEG recordings (short-term). The evaluation on long-term recordings encompasses entire recordings including sleep-wake cycles, assessing both window-based classification accuracy and seizure event-based detection performance.

II. PROPOSED METHOD

A. PROBLEM SETTING

In cross-patient epileptic seizure detection, a model is trained on data from multiple patients (source patients) to predict seizures in an unseen patient (target patient). EEG recordings are segmented into fixed-length windows of L samples, and each window is classified as seizure or non-seizure. The EEG segment at time t is denoted as $\mathbf{x}_t \in \mathbb{R}^{C \times L}$ (where C is the number of channels), and the corresponding class

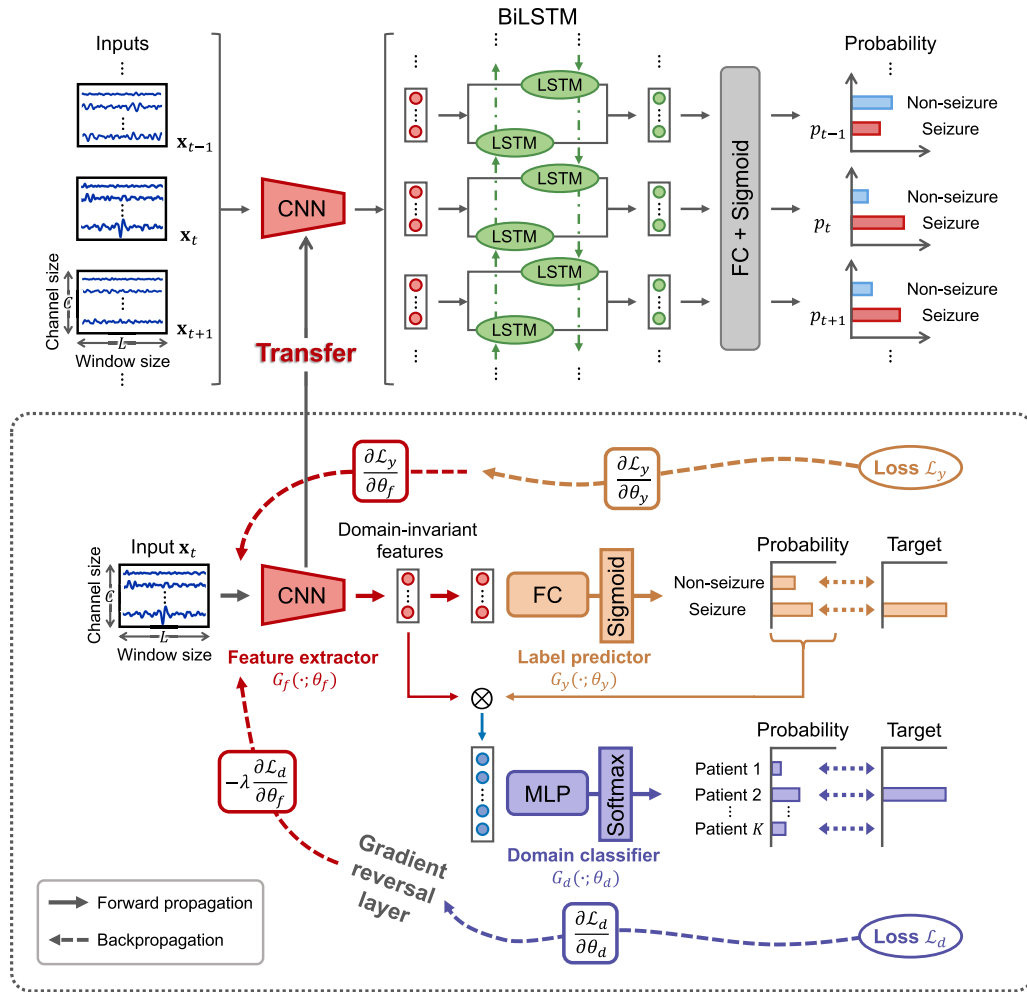


FIGURE 1. Schematic overview of the proposed method. In the first stage, a CNN extracts local patient-invariant features via conditional domain adversarial training. By conditioning on seizure prediction probabilities, the model learns to reduce patient-specific characteristics separately for seizure and non-seizure states. In the second stage, the CNN weights from the first stage are transferred to a CNN-BiLSTM, which captures temporal dependencies in the extracted features and models temporal progression of seizures. During inference, the CNN-BiLSTM outputs prediction probabilities for each EEG segment.

label is $y_t \in \{0, 1\}$ (0: non-seizure, 1: seizure). Let $k \in \{1, \dots, K\}$ denote the index of a source patient used for training, where patient k has $R^{(k)}$ EEG recordings. For the r -th recording ($r \in \{1, \dots, R^{(k)}\}$) with a time-series length of $T^{(k,r)}$, the corresponding dataset is defined as $\mathcal{D}^{(k,r)} = \{(\mathbf{x}_t^{(k,r)}, y_t^{(k,r)})\}_{t=1}^{T^{(k,r)}}$. Consequently, the complete dataset for patient k is $\mathcal{D}^{(k)} = \{\mathcal{D}^{(k,r)}\}_{r=1}^{R^{(k)}}$, and the dataset for all source patients is $\{\mathcal{D}^{(k)}\}_{k=1}^K$. The objective of this study is to predict the seizure labels $\{y_t^{\text{trg}}\}$ from the EEG time series $\{\mathbf{x}_t^{\text{trg}}\}$ of the target patient using a model trained on $\{\mathcal{D}^{(k)}\}_{k=1}^K$.

B. INTER-PATIENT VARIABILITY IN SEIZURE AND NON-SEIZURE EEG PATTERNS

The neurophysiological properties of EEG differ between seizure and non-seizure states. Nonlinear analyses have revealed that interictal EEG exhibits a higher degree of freedom in neural activity and greater complexity compared

to ictal EEG [25]. This high degree of freedom suggests that background EEG has a substantial capacity to reflect diverse patient-specific characteristics. Indeed, background EEG is known to be strongly influenced by age and sleep-wake states [26], [27], [28], [29], with individual differences observed in the peak frequency and amplitude asymmetry of alpha activity (8–13 Hz), which predominates in adults [30]. In patients with epilepsy, in addition to these general individual differences, the frequency and distribution of interictal epileptiform discharges vary across patients [6], [7], [30], further increasing inter-patient variability during non-seizure periods. In contrast, seizure EEG arises from a common neural mechanism of hypersynchronous neuronal discharges and tends to exhibit more similar patterns across patients compared to non-seizure EEG; however, patient-specific features related to seizure type and the localization of the epileptogenic zone also exist [30], [31]. These findings indicate that the sources and characteristics of inter-patient

variability differ between seizure and non-seizure states, and this difference should be considered when extracting patient-invariant features.

C. OVERVIEW OF THE PROPOSED METHOD

Fig. 1 illustrates an overview of the proposed method. In this approach, a CNN extracts informative features from EEG recordings at specific time points, and a BiLSTM captures the temporal dependencies of these features for seizure detection. However, this architecture alone cannot explicitly address inter-patient variability in EEG patterns, thereby limiting generalization to unseen patients.

To address this challenge, we introduce domain adversarial training. Conventional domain adversarial training uniformly aligns the entire feature space to achieve patient invariance; however, as discussed in Section II-B, the sources and characteristics of inter-patient variability differ between seizure and non-seizure states. Therefore, conditional domain adversarial training is necessary to acquire patient-invariant features independently for each state.

Nevertheless, since acquiring patient-invariant features and capturing temporal patterns are distinct optimization objectives, simultaneous training may cause mutual interference. This motivates a two-stage training framework that explicitly separates these two objectives. In the first stage, we apply conditional domain adversarial networks with entropy conditioning (CDAN+E) [24] to the CNN serving as the feature extractor. Specifically, in addition to the feature extractor and a label predictor, we introduce a domain classifier and train the feature extractor to deceive the domain classifier. By using the outer product of the feature extractor output and the label prediction probability as input to the domain classifier, patient-invariant features are acquired separately for seizure and non-seizure states. In the second stage, we fine-tune the entire CNN-BiLSTM model using the CNN weights learned in the first stage as initialization. Since seizures exhibit temporal continuity lasting tens of seconds, patient-invariant features extracted from individual time points alone are insufficient for accurate seizure detection. The BiLSTM captures temporal dependencies between consecutive windows, enabling modeling of the dynamic progression of seizures. This two-stage training enables patient-invariant seizure detection that accounts for temporal context.

D. CONDITIONAL DOMAIN ADVERSARIAL TRAINING FOR CNN

In the first stage, we apply domain adversarial training conditioned on the seizure state to the CNN to capture patient-invariant features. For this training, we construct the training dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i, d_i)\}_{i=1}^N$ from the source dataset $\{\mathcal{D}^{(k)}\}_{k=1}^K$, where $d_i \in \{1, \dots, K\}$ indicates the patient (domain) to which each sample belongs, and N denotes the total number of training samples. During training, mini-batches of size N_b are generated from this dataset, each containing data from different patients and time points.

Our method employs three components: a feature extractor G_f (implemented as a CNN), a label predictor G_y , and a domain classifier G_d . The parameters of these components are denoted as θ_f , θ_y , and θ_d , respectively. The label predictor G_y consists of a fully connected (FC) layer with a sigmoid function, while the domain classifier G_d consists of a multilayer perceptron (MLP) followed by a softmax function.

For label prediction, we use the weighted binary cross-entropy loss $\mathcal{L}_y$:

$$\mathcal{L}_y(\theta_f, \theta_y) = -\frac{1}{N_b} \sum_{i=1}^{N_b} w_{y_i} [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)], \quad (1)$$

where $p_i = G_y(G_f(\mathbf{x}_i; \theta_f); \theta_y)$ denotes the seizure probability predicted by the label predictor G_y based on features extracted by the feature extractor G_f . To address class imbalance, the weight coefficients w_{y_i} are defined as:

$$w_0 = \frac{N}{2 \sum_{i=1}^N (1 - y_i)}, \quad w_1 = \frac{N}{2 \sum_{i=1}^N y_i}. \quad (2)$$

A key characteristic of our method is the use of conditional feature representations as input to the domain classifier G_d . Specifically, let $\mathbf{f}_i = G_f(\mathbf{x}_i; \theta_f) \in \mathbb{R}^h$ denote the feature vector extracted by the feature extractor G_f , where h is the feature dimensionality. We compute the outer product of $\mathbf{f}_i$ and the seizure prediction probability vector $\mathbf{g}_i = [p_i, 1 - p_i]^T \in \mathbb{R}^2$ output by the label predictor G_y , and vectorize the result to obtain $\phi_i = \text{vec}(\mathbf{f}_i \otimes \mathbf{g}_i) \in \mathbb{R}^{2h}$, which serves as the input to the domain classifier. Through this outer product, the feature vector $\mathbf{f}_i$ is weighted according to the prediction probability p_i , such that features from seizure-like samples ($p_i \approx 1$) and non-seizure-like samples ($p_i \approx 0$) are separated before being fed into the domain classifier. This enables the reduction of patient-specific characteristics for each state while preserving discriminative features important for seizure detection.

For domain classification, we use the weighted cross-entropy loss $\mathcal{L}_d$:

$$\mathcal{L}_d(\theta_f, \theta_d) = -\frac{1}{N_b} \sum_{i=1}^{N_b} w_{d_i} \mu_i \sum_{k=1}^K \mathbb{1}_{[k=d_i]} \ln G_d^{(k)}(\phi_i; \theta_d). \quad (3)$$

Here, $G_d^{(k)}(\phi_i; \theta_d)$ denotes the k -th element of the softmax output of the domain classifier (i.e., the probability of belonging to patient k), and $\mathbb{1}_{[k=d_i]}$ is the indicator function (1 if $k = d_i$, 0 otherwise). To address imbalance across domains, the weight coefficient w_{d_i} is defined as:

$$w_{d_i} = \frac{N}{K \cdot N_{d_i}}, \quad (4)$$

where N_{d_i} denotes the number of samples belonging to domain d_i . Furthermore, μ_i is a weight coefficient based on prediction uncertainty, defined as $\mu_i = 1 + \exp(-H(p_i))$, where the entropy of the label prediction is

$H(p_i) = -[p_i \ln p_i + (1 - p_i) \ln(1 - p_i)]$. Since the entropy $H(p_i)$ is maximized when $p_i \approx 0.5$, samples with uncertain predictions have smaller μ_i values, thereby reducing their contribution to the domain classification loss. This mechanism prevents overfitting to ambiguous samples near the class boundary.

The overall objective function is formulated as the following minimax problem:

$$\min_{\theta_f, \theta_y} \max_{\theta_d} [\mathcal{L}_y(\theta_f, \theta_y) - \lambda \mathcal{L}_d(\theta_f, \theta_d)], \quad (5)$$

where λ is a hyperparameter controlling the weight of the domain classification loss. In implementation, a gradient reversal layer (GRL) is placed between the feature extractor G_f and the domain classifier G_d , which multiplies the gradient of $\mathcal{L}_d$ by $-\lambda$ during backpropagation. Through this optimization, the features generated by the feature extractor G_f are adjusted to degrade the classification accuracy of the domain classifier G_d . In contrast, the domain classifier G_d is trained to improve its domain classification accuracy. This adversarial training process enables the feature extractor G_f to learn domain-independent, i.e., patient-invariant, features effectively.

E. JOINT TRAINING OF CNN AND BiLSTM

In the second stage, we connect a BiLSTM and an FC layer to the CNN trained with conditional domain adversarial training in the first stage and train the entire network to capture temporal patterns. By using the CNN weights learned in the first stage as initialization, the network can learn temporal dependencies while preserving patient invariance.

The BiLSTM integrates contextual information from both past and future at each time point by combining a forward LSTM and a backward LSTM. The forward LSTM captures the progression from seizure onset, while the backward LSTM captures context by tracing back from the post-seizure state. This bidirectional processing improves the detection accuracy of seizure onset and offset, and effectively models temporal dependencies between consecutive windows.

In this stage, we extract subsequences consisting of W consecutive windows from the source data $\{\mathcal{D}^{(k,r)}\}$. Specifically, the window sequence from time t to $t + W - 1$, denoted as $\{\mathbf{x}_t^{(k,r)}, \dots, \mathbf{x}_{t+W-1}^{(k,r)}\}$, together with the corresponding labels $\{y_t^{(k,r)}, \dots, y_{t+W-1}^{(k,r)}\}$, constitutes a single training unit. Here, W represents the length of temporal context required for seizure detection.

During training, the CNN first extracts a feature vector from each window, and the BiLSTM then sequentially processes the resulting sequence of W feature vectors to output a hidden state at each time step. Finally, the seizure probability at each time step is computed by an FC layer followed by a sigmoid function. Mini-batches are constructed by randomly selecting subsequences from different patients, recordings, and time points. The model is trained to minimize

the following weighted binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N'_b W} \sum_{i=1}^{N'_b} \sum_{\ell=0}^{W-1} w_{y_\ell^{(i)}} \times [y_\ell^{(i)} \ln p_\ell^{(i)} + (1 - y_\ell^{(i)}) \ln(1 - p_\ell^{(i)})]. \quad (6)$$

Here, N'_b denotes the batch size of subsequences, and ℓ denotes the offset within a subsequence (corresponding to time $t + \ell$). The terms $y_\ell^{(i)}$ and $p_\ell^{(i)}$ represent the ground truth label and predicted probability at the ℓ -th time point in the i -th subsequence, respectively. The class weight $w_{y_\ell^{(i)}}$ is computed in the same manner as in Eq. (2).

F. INFERENCE ALGORITHM

During inference, the EEG time series of a target patient $\{\mathbf{x}_t^{\text{trg}}\}$ is processed in subsequences of W consecutive windows, as in training. Unlike training, subsequences are extracted with a fixed stride in a sliding window manner to improve temporal resolution.

To integrate predictions from overlapping subsequences, only the central segment of each prediction sequence is retained. This is because the BiLSTM has limited access to contextual information at boundaries of a subsequence, whereas it can effectively integrate bidirectional information near the center. The final seizure prediction probabilities $\{p_t^{\text{trg}}\}$ are obtained by concatenating the retained central segments in temporal order.

The seizure decision at time t is made as follows:

$$\hat{y}_t^{\text{trg}} = \begin{cases} 1 \text{ (seizure)} & \text{if } p_t^{\text{trg}} \geq \tau, \\ 0 \text{ (non-seizure)} & \text{otherwise,} \end{cases} \quad (7)$$

where τ is a predefined decision threshold ($0 < \tau < 1$).

III. EXPERIMENTS

A. DATASET

To comprehensively evaluate the generalization performance of the proposed method, we employed two EEG datasets from patients with focal epilepsy, recorded under different conditions. The long-term recordings comprise extended monitoring data including sleep-wake cycles and were primarily used to evaluate performance in realistic clinical settings. In contrast, the short-term recordings consist of brief segments centered around seizure events, making them suitable for evaluating detection accuracy at seizure boundaries. Both datasets were recorded using electrode placements based on the international 10–20 system. For each recording, a longitudinal bipolar montage was applied to the 19 electrodes (excluding the earlobes A1 and A2), yielding 18-channel signals ($C = 18$).

For the long-term recordings, we used the publicly available Siena Scalp EEG Database [32], [33], [34]. This dataset comprises EEG recordings from 14 patients (aged 20–71 years) sampled at 512 Hz. Seizure onset and offset times were extracted from the accompanying annotation files. Recordings PN10-2 and PN10-4.5.6 were excluded

from the analysis due to incomplete annotations. Table 1 summarizes the patient characteristics, including sex, age, total data length, seizure duration, and number of seizures. The mean total recording duration was 9.6 ± 6.4 h, and the mean total seizure duration was 195 ± 85.9 s. Regarding seizure types, patient PN10 exhibited focal to bilateral tonic-clonic seizures, PN14 showed focal onset without impaired awareness, and the remaining patients presented with focal onset impaired awareness. Epilepsy diagnosis and seizure classification were performed by an expert clinician according to the criteria of the International League Against Epilepsy. Because this dataset includes sleep-wake cycles and diverse background EEG activities that closely reflect real-world clinical conditions, we designated it as the primary dataset for evaluating the generalization performance of the proposed method.

For the short-term recordings, we used a private dataset collected in a previous study [35]. This dataset contains EEG recordings from 20 patients (aged 0.5–41 years) sampled at 500 Hz. Table 2 summarizes the patient characteristics, including age, sex, total data length, and seizure duration. Each patient contributed a single EEG recording containing one seizure event. The mean recording and seizure durations were 319.5 ± 44.5 s and 47.4 ± 22.7 s, respectively; each recording captures approximately 300 s of data centered around the seizure. Seizure onset and offset times were annotated at the sample level by an expert epileptologist. Written informed consent was obtained from all patients. This study was conducted with approval from the Okayama University Ethics Committee (approval No: 1706-019).

TABLE 1. Long-term recordings (Siena Scalp EEG Database): patient conditions.

Patient	Sex	Age (year)	Total data length (m)	Seizure duration (s)	Number of seizures
PN00	Male	55	193	325	5
PN01	Male	46	809	128	2
PN03	Male	54	1453	244	2
PN05	Female	51	361	104	3
PN06	Male	36	723	282	5
PN07	Female	20	524	62	1
PN09	Female	27	410	203	3
PN10	Male	25	703	227	6
PN11	Female	58	145	55	1
PN12	Male	71	366	290	4
PN13	Female	34	520	264	3
PN14	Male	49	1228	163	4
PN16	Female	41	293	230	2
PN17	Male	42	307	153	2

B. PREPROCESSING

To remove artifacts caused by body movements and muscle activity, a 1.6–30 Hz bandpass filter was applied following previous work [13]. Different standardization strategies were adopted for each dataset. A previous study reported that standardization can degrade performance by causing loss of amplitude information during seizures, although it can be beneficial when batch sizes are small [36]. For the

TABLE 2. Short-term recordings: patient conditions.

Patient	Sex	Age (year)	Total data length (s)	Seizure duration (s)
A	Male	2	288	70
B	Male	23	300	52
C	Female	4	301	30
D	Male	4	377	92
E	Male	0.5	319	38
F	Male	41	300	32
G	Female	3	240	52
H	Male	19	391	34
I	Male	0.8	388	78
J	Male	20	362	14
K	Male	36	300	22
L	Male	9	300	42
M	Male	13	300	42
N	Male	15	300	47
O	Male	8	300	15
P	Female	19	300	69
Q	Male	27	424	61
R	Male	38	300	65
S	Male	17	300	17
T	Male	19	300	75

long-term recordings, standardization was not applied to preserve amplitude information that contributes to seizure detection. For the short-term recordings, where recording durations are brief and batch sizes are limited, channel-wise standardization (zero mean, unit variance) was applied.

After preprocessing, each recording was segmented into non-overlapping one-second windows. The window length corresponded to the sampling frequency: $L = 512$ for the long-term recordings and $L = 500$ for the short-term recordings. The class label for each window was assigned based on the majority class (seizure or non-seizure) within that window.

C. DATA PARTITIONING

In the first training stage for the long-term recordings, a sampling strategy was applied to address the severe class imbalance between seizure and non-seizure segments. Specifically, all seizure windows were retained, while non-seizure windows were randomly undersampled to achieve a seizure-to-non-seizure ratio of 1 : 10. This ratio was chosen to balance the risk of information loss from excessive undersampling against training bias caused by class imbalance. For the short-term recordings, all windows were used without sampling, given the brief recording duration and high seizure density.

In the second training stage and during inference, subsequences of W consecutive windows were constructed as inputs to the BiLSTM. For the short-term recordings, the entire time series of each patient was treated as a single subsequence, with W equal to the total number of windows per patient. For the long-term recordings, non-overlapping subsequences of $W = 90$ windows (corresponding to 90 seconds) were extracted during training. This sequence length was chosen to balance capturing sufficient temporal context against minimizing detection delay. During inference

on the long-term recordings, overlapping subsequences were processed with a stride of 30 windows (30 seconds), and predictions from the central 30-second segment of each subsequence were retained, following the procedure described in Section II-F.

D. IMPLEMENTATION DETAILS

1) NETWORK ARCHITECTURE

Fig. 2 illustrates the architecture of the CNN used in the proposed method. The CNN consists of four blocks. Each block applies a sequence of a one-dimensional convolutional layer (kernel size 3, stride 1, padding 1), batch normalization, and Swish activation twice, followed by a squeeze-and-excitation (SE) module [37]. The SE module adaptively recalibrates channel-wise feature responses and is effective for capturing inter-channel interactions in EEG signals. The reduction ratios for the SE modules were set to 2, 2, 4, and 4 for the four blocks, respectively. A max pooling layer (kernel size 2, stride 2) at the end of each block halves the temporal dimension. The number of output channels progressively increases to 8, 16, 32, and 64, and global average pooling yields a 64-dimensional feature vector ($h = 64$).

The label predictor consists of a single FC layer, while the domain classifier is a three-layer MLP with output dimensions of 64, 32, and K , where batch normalization and Swish activation are applied to the intermediate layers. The BiLSTM consists of two layers with a hidden size of 32, and its output is transformed by a single FC layer to produce the final prediction.

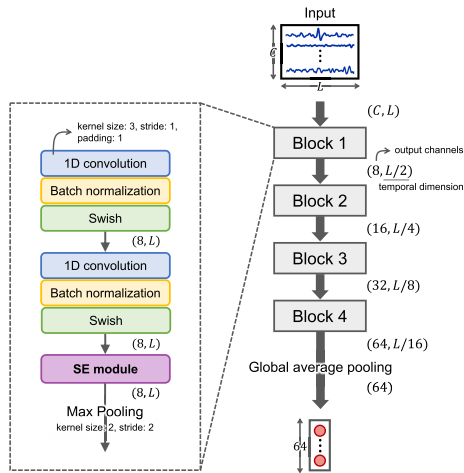


FIGURE 2. CNN architecture in the proposed method. The CNN consists of four blocks. In each block, a sequence of 1D convolution, batch normalization, and Swish activation is applied twice, followed by an SE module and a max pooling layer. The numbers alongside each arrow indicate the output dimensions in the format (number of channels, temporal length). The number of channels progressively increases across block, and global average pooling generates the final 64-dimensional feature vector.

2) TRAINING SETTINGS

In the first stage, the feature extractor, label predictor, and domain classifier were trained jointly. The Adam optimizer

TABLE 3. Hyperparameter settings for the second-stage training on the long-term and short-term recordings.

Parameter	Long-term	Short-term
Batch size N'_b	32	4
Learning rate		
CNN	5×10^{-4}	5×10^{-4}
BiLSTM, FC layer	5×10^{-4}	10^{-3}
Weight decay	5×10^{-5}	0
Number of epochs	5	10

was used, and training was conducted for 25 epochs. The learning rates were set to 0.01 for the feature extractor and label predictor, and 0.005 for the domain classifier; the lower rate for the domain classifier was intended to stabilize multi-domain adversarial training. The batch sizes were set to $N_b = 128$ for the long-term recordings and $N_b = 32$ for the short-term recordings. Following previous work [22], the hyperparameter λ in Eq. (5) was scheduled as:

$$\lambda = \frac{2}{1 + \exp(-10s)} - 1, \quad (8)$$

where $s \in [0, 1]$ denotes the training progress, computed as the ratio of the current training step to the total number of training steps. This schedule gradually increases λ from 0 to 1, prioritizing seizure detection accuracy in the early stages while progressively enforcing patient invariance in later stages.

In the second stage, the CNN weights learned in the first stage were used to initialize the model, and the entire network including the BiLSTM and FC layer was fine-tuned. The Adam optimizer was used, and the training parameters for each dataset are summarized in Table 3.

The model was implemented in PyTorch 2.8.0 and trained on an NVIDIA RTX 4500 Ada Generation GPU. The source code is publicly available at https://github.com/rntche/EEG_CDAN

E. EVALUATION METHODOLOGY

To statistically account for both inter-patient variability and the stochastic nature of the training process, leave-one-patient-out cross-validation was conducted 15 times with different random seeds. Of the $K + 1$ patients in each dataset, K patients served as source patients for training, and the remaining one served as the target patient for testing ($K = 13$ for the long-term recordings, $K = 19$ for the short-term recordings). This procedure was repeated $K + 1$ times so that each patient served as the target exactly once. The mean values of the following evaluation metrics were calculated across all repetitions.

For window-based evaluation, seizure detection performance was assessed at the individual window level. The evaluation metrics included sensitivity, specificity, the Matthews correlation coefficient (MCC), the area under the receiver operating characteristic curve (AUC-ROC), the area under the precision-recall curve (AUC-PR), and the partial area under the curve (pAUC). The pAUC was calculated as the AUC-ROC restricted to false positive rates below

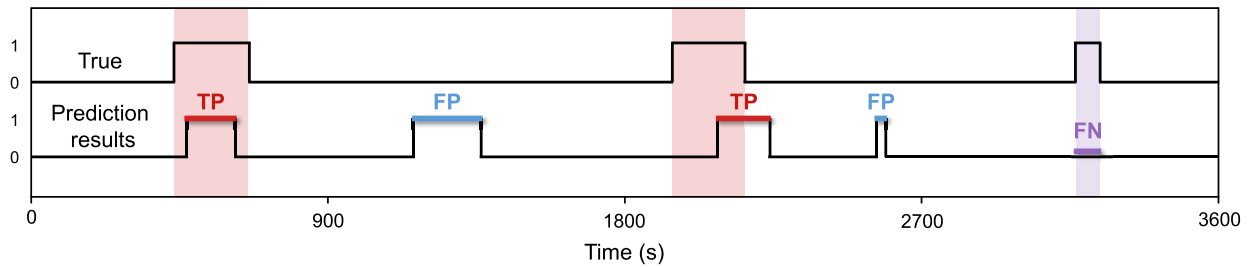


FIGURE 3. Definition of event-based evaluation metrics for seizure detection. The top trace shows true labels (0: non-seizure, 1: seizure) based on expert annotations. The bottom trace shows prediction results obtained by binarizing prediction probabilities with a threshold. A detected event, defined as a contiguous interval of predicted seizure, is classified as a true positive (TP, red) if it overlaps with a ground-truth seizure interval by at least one second. A detected event without such overlap is classified as a false positive (FP, blue). A ground-truth seizure interval without any overlapping detected event is classified as a false negative (FN, purple). Sensitivity is calculated as $TP/(TP + FN)$, and FPs/h is calculated as the total number of FPs divided by the total recording duration in hours.

0.05, evaluating discriminative performance in the low false-positive region. Among these, MCC and AUC-ROC were designated as the primary metrics for statistical analysis: MCC provides a threshold-dependent measure that is robust to class imbalance, while AUC-ROC offers a threshold-independent assessment of overall discriminative ability. The threshold τ for converting seizure prediction probabilities to binary decisions was set to 0.5.

For event-based evaluation, detection performance was assessed on a per-seizure basis using the long-term recordings (Fig. 3). This evaluation reflects clinical priorities, where minimizing both missed seizures and false alarms is essential. A detected event was defined as a contiguous interval during which the predicted seizure probability exceeded the threshold τ . A detected event was classified as a true positive (TP) if it overlapped with a ground-truth seizure interval by at least one second; otherwise, it was classified as a false positive (FP). Ground-truth seizures without any overlapping detected events were classified as false negatives (FN). Sensitivity (the proportion of ground-truth seizures detected) and false positives per hour (FPs/h) were used as evaluation metrics.

F. COMPARISON METHODS

1) COMPONENT-WISE EVALUATION

To evaluate the effectiveness of each proposed component, we compared four configurations based on the presence or absence of CDAN+E and the BiLSTM: CNN alone, CNN with CDAN+E, CNN-BiLSTM without CDAN+E, and CNN-BiLSTM with CDAN+E (proposed). Models without CDAN+E were trained using only the label prediction loss, with the domain classifier removed in the first stage. All other settings were identical to those of the proposed method.

To statistically analyze the effects of CDAN+E and the BiLSTM, a linear mixed-effects model was fitted. The method (i.e., experimental condition) was treated as a fixed effect, while random seed and patient were modeled as random effects. Using the primary metrics (MCC and AUC-ROC) as response variables, the following three contrasts were pre-specified: (1) the effect of CDAN+E on CNN, (2) the effect of the BiLSTM (without CDAN+E), and

(3) the effect of CDAN+E on CNN-BiLSTM. The Holm method was applied for multiple comparison correction at a significance level of 5%. For each contrast, the estimated difference and the 95% confidence interval (CI) were computed; the estimated difference represents the mean improvement in the evaluation metric attributable to the component.

Furthermore, to quantitatively evaluate the degree of patient invariance achieved under each ablation condition, we applied linear probing via logistic regression to the CNN feature representations from the training data, computing patient prediction accuracy. Note that even for configurations incorporating the BiLSTM, patient prediction accuracy was computed from the CNN features. By evaluating the training data that CDAN+E targets for optimization, we directly measured the extent to which patient-specific information was removed through conditional domain adversarial training. For both datasets, the first-stage training data was used across all configurations to ensure a consistent comparison. A lower prediction accuracy indicates better alignment of feature representations across patients.

2) COMPARISON OF DOMAIN ADVERSARIAL TRAINING METHODS

To examine the effect of different domain adversarial training strategies, we compared the following two methods:

- DANN [22]: Only the feature extractor output was used as input to the domain classifier, without conditioning on seizure state.
- CDAN [24]: Conditioning was performed identically to the proposed method, but without entropy weighting. This comparison assessed the contribution of uncertainty-based weighting.

All methods used identical network architecture and training settings.

3) COMPARISON OF TRAINING STRATEGIES

To assess the impact of training strategy on the CNN-BiLSTM model with CDAN+E, we compared the following two alternatives:

- One-step learning: CDAN+E was applied to the entire CNN-BiLSTM without dividing training into stages. The Adam optimizer was used with learning rates of 0.01 for the feature extractor and label predictor, and 0.005 for the domain classifier. Training was conducted for 25 epochs with batch size of $N'_b = 32$ for the long-term recordings and $N'_b = 4$ for the short-term recordings.
- Two-step learning without fine-tuning: The two-stage training procedure was followed as in the proposed method; however, the CNN weights from the first stage were frozen during the second stage. This configuration evaluated the contribution of fine-tuning. All training parameters were identical to those of the proposed method.

4) COMPARISON OF TEMPORAL MODELING ARCHITECTURES

To evaluate alternative temporal modeling approaches, we compared the following three methods:

- CNN-LSTM: The BiLSTM was replaced with a unidirectional LSTM. This configuration evaluated the contribution of bidirectional processing.
- CNN-Transformer (shallow): The BiLSTM was replaced with a Transformer encoder consisting of two layers, four attention heads, and a feed-forward dimension of 256.
- CNN-Transformer (deep): The BiLSTM was replaced with a deeper variant of the Transformer encoder, consisting of four layers and eight attention heads, with the same feed-forward dimension.

All three methods incorporated CDAN+E, and all other settings were identical to those of the proposed method.

IV. RESULTS

A. ANALYSIS OF TRAINING PROCESS

Fig. 4 shows *t*-SNE [38] visualizations of feature representations extracted from the training data of the long-term recordings. Here, PN03 served as the target patient, with the remaining patients used as source patients for first-stage training; data points are colored by patient. For the baseline using signal variance of each window as features, patient-specific clusters emerged in the feature space (Fig. 4(a)). With the CNN alone, data points from different patients partially overlapped; however, patient-dependent distributional biases remained evident (Fig. 4(b)). In contrast, incorporating CDAN+E eliminated these patient-specific clusters, with data points distributed throughout the feature space (Fig. 4(c)).

Fig. 5 shows the loss progression during the first-stage training. Results were obtained using a fixed random seed, with mean and standard deviation calculated across patients at each epoch. The label predictor loss decreased monotonically as training progressed. Conversely, the domain classifier loss increased during the early training phase and then tended to converge.

B. EVALUATION OF PROPOSED COMPONENTS

Table 4 presents the performance comparison across different model configurations. Applying CDAN+E to the CNN alone did not yield clear performance improvements. Conversely, incorporating BiLSTM improved all metrics on both datasets. The proposed method, which integrates CDAN+E and BiLSTM, achieved the highest values for all metrics except sensitivity on the long-term recordings. On the short-term recordings, it achieved the highest values for specificity and MCC, and the second-highest values for the remaining four metrics.

Table 5 summarizes the statistical analysis results using a linear mixed-effects model. Incorporating BiLSTM significantly improved MCC on both datasets ($p < 0.001$). AUC-ROC also increased significantly on both the long-term recordings ($p = 0.015$) and the short-term recordings ($p < 0.001$). The estimated improvement in MCC was 0.215 (95% CI: 0.179–0.252) for the long-term recordings and 0.131 (95% CI: 0.103–0.160) for the short-term recordings. Regarding the effect of applying CDAN+E to the CNN-BiLSTM, significant improvements were observed in both MCC (estimated difference: 0.034, $p = 0.047$) and AUC-ROC (estimated difference: 0.055, $p < 0.001$) on the long-term recordings. On the short-term recordings, only MCC showed a significant improvement (estimated difference: 0.035, $p = 0.006$). When CDAN+E was applied to the CNN alone, AUC-ROC decreased significantly on the short-term recordings (estimated difference: -0.040 , $p < 0.001$).

Table 6 presents patient prediction accuracy computed from the feature representations of the training data. A lower value indicates that samples from different patients are less linearly separable. Applying CDAN+E reduced patient prediction accuracy across both datasets and architectures. For the long-term recordings, accuracy decreased from 32.493% to 23.998% with the CNN alone, and from 33.936% to 25.405% with the CNN-BiLSTM (chance level: 7.692%). For the short-term recordings, a similar trend was observed: accuracy decreased from 22.152% to 13.802% and from 23.350% to 14.579%, respectively (chance level: 5.263%). In contrast, incorporating BiLSTM had little effect on patient invariance, with patient prediction accuracy increasing by only approximately 1% regardless of whether CDAN+E was applied.

C. EVALUATION OF CLINICAL EFFECTIVENESS

Fig. 6 compares patient-specific event-based metrics using the long-term recordings for the CNN-BiLSTM with and without CDAN+E. Higher sensitivity (closer to 1) and lower FPs/h (closer to 0) indicate better model performance. The mean and standard deviation across patients for sensitivity and FPs/h were 0.711 ± 0.367 and 2.503 ± 3.281 without CDAN+E, and 0.712 ± 0.369 and 1.627 ± 2.994 with CDAN+E, respectively. Sensitivity was comparable or improved with CDAN+E in 12 out of 14 patients, whereas it decreased in only 2 patients. FPs/h improved in 7 out of

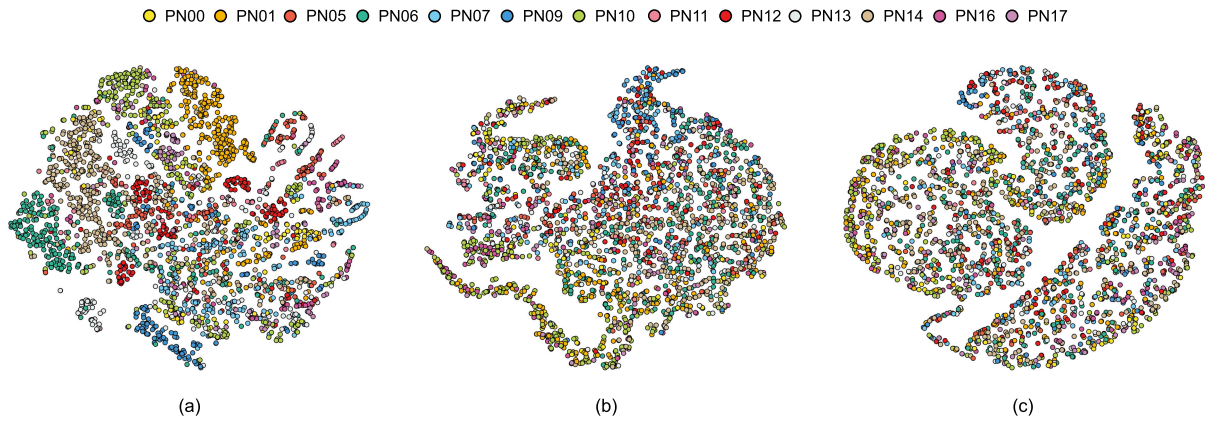


FIGURE 4. t-SNE visualization of feature representations for the long-term recordings. (a) Signal variance of each preprocessed EEG window, used as a baseline feature. Features extracted by the first-stage CNN: (b) CNN alone and (c) CNN with CDAN+E (ours). The target patient is PN03 (remaining patients used as source patients), with data points colored by patient. For visualization, 10% of the training data was randomly sampled.

TABLE 4. Performance comparison of model configurations: mean $\pm$ standard deviation across patients. The ‘✓’ symbol denotes the presence of each component. The best and second-best results for each dataset are indicated in bold and underlined, respectively.

Dataset	CNN	BiLSTM	CDAN+E	Sensitivity	Specificity	MCC	AUC-ROC	AUC-PR	pAUC
Long-term	✓			0.571 ± 0.239	0.937 ± 0.049	0.218 ± 0.136	0.841 ± 0.079	0.278 ± 0.226	0.719 ± 0.115
	✓		✓	<u>0.596 ± 0.234</u>	0.940 ± 0.055	0.241 ± 0.158	0.861 ± 0.083	0.264 ± 0.225	0.719 ± 0.119
	✓	✓		0.569 ± 0.316	<u>0.974 ± 0.040</u>	<u>0.433 ± 0.304</u>	<u>0.869 ± 0.130</u>	<u>0.493 ± 0.344</u>	<u>0.799 ± 0.145</u>
	✓	✓	✓	<u>0.590 ± 0.313</u>	<u>0.982 ± 0.031</u>	<u>0.468 ± 0.338</u>	<u>0.924 ± 0.093</u>	<u>0.511 ± 0.358</u>	<u>0.830 ± 0.137</u>
Short-term	✓			0.613 ± 0.252	0.856 ± 0.110	0.429 ± 0.220	0.800 ± 0.142	0.532 ± 0.253	0.664 ± 0.131
	✓		✓	0.560 ± 0.276	<u>0.891 ± 0.067</u>	0.445 ± 0.251	0.760 ± 0.169	0.521 ± 0.260	0.660 ± 0.124
	✓	✓		<u>0.734 ± 0.325</u>	0.866 ± 0.190	<u>0.560 ± 0.303</u>	<u>0.902 ± 0.150</u>	<u>0.755 ± 0.279</u>	<u>0.811 ± 0.160</u>
	✓	✓	✓	<u>0.733 ± 0.316</u>	<u>0.907 ± 0.100</u>	<u>0.595 ± 0.300</u>	<u>0.890 ± 0.189</u>	<u>0.754 ± 0.300</u>	<u>0.805 ± 0.162</u>

TABLE 5. Results of a linear mixed-effects model analysis comparing the effects of CDAN+E and BiLSTM. Estimated differences (Est. diff.) and 95% confidence intervals (CI) indicate the mean improvement attributed to each component. The experimental condition was modeled as a fixed effect; random seed and patient were modeled as random effects. The p -values were corrected using the Holm method at a significance level of 5%.

Dataset	Tested effect	MCC			AUC-ROC		
		Est. diff.	95% CI	p -value	Est. diff.	95% CI	p -value
Long-term	Effect of CDAN+E (w/o BiLSTM)	0.022	$[-0.014, 0.059]$	0.140	0.021	$[-0.005, 0.046]$	0.051
	Effect of BiLSTM (w/o CDAN+E)	0.215	$[0.179, 0.252]$	$<.0001$	0.028	$[0.003, 0.053]$	0.015
	Effect of CDAN+E (w/ BiLSTM)	0.034	$[-0.002, 0.071]$	0.047	0.055	$[0.030, 0.080]$	$<.0001$
Short-term	Effect of CDAN+E (w/o BiLSTM)	0.016	$[-0.013, 0.045]$	0.187	-0.040	$[-0.059, -0.022]$	$<.0001$
	Effect of BiLSTM (w/o CDAN+E)	0.131	$[0.103, 0.160]$	$<.0001$	0.102	$[0.084, 0.120]$	$<.0001$
	Effect of CDAN+E (w/ BiLSTM)	0.035	$[0.007, 0.064]$	0.006	-0.013	$[-0.031, 0.005]$	0.091

14 patients, including 4 patients (PN00, PN01, PN16, and PN17) for whom FPs/h decreased from over 1.5 to below 0.25.

D. COMPARISON OF DESIGN CHOICES

1) DOMAIN ADVERSARIAL TRAINING STRATEGIES

Table 7 presents the effects of different domain adversarial training strategies on the CNN-BiLSTM. On the long-term recordings, CDAN+E achieved the highest values for four out of six metrics (sensitivity, MCC, AUC-ROC, and pAUC). CDAN showed the best results for specificity and AUC-PR, but the differences from CDAN+E were only 0.001 and 0.002, respectively. On the short-term recordings, performance differences among methods were small, with

the best-performing method varying across metrics; however, DANN and CDAN+E exhibited relatively high performance on most metrics.

2) TRAINING STRATEGIES

Fig. 7 shows the impact of different training strategies on performance. The two-step learning approaches consistently outperformed the one-step approach on both metrics across both datasets. The effect of fine-tuning within the two-step framework differed between datasets. On the long-term recordings, fine-tuning improved MCC by 0.087 while decreasing AUC-ROC by 0.036 (Fig. 7(a)). On the short-term recordings, the differences with and without fine-tuning were 0.001 for MCC and 0.005 for AUC-ROC (Fig. 7(b)).

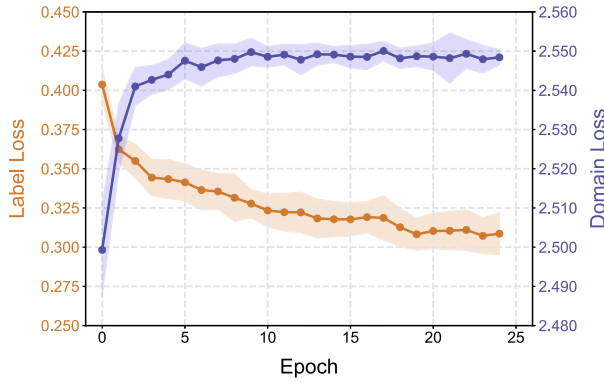


FIGURE 5. Averaged loss curves during first-stage training as a function of epochs for the long-term recordings. Results were obtained using a single random seed. Circles indicate the mean across patients, and shaded regions represent the standard deviation. Label loss corresponds to seizure prediction, and domain loss to patient classification.

TABLE 6. Patient prediction accuracy (%) on training data across ablation conditions: mean $\pm$ standard deviation across patients. A lower value indicates that patient-specific information is more effectively removed from the feature space, reflecting stronger patient invariance. Chance levels are 7.692% for the long-term recordings and 5.263% for the short-term recordings. The ‘✓’ symbol denotes the presence of each component. The best and second-best results for each dataset are indicated in bold and underlined, respectively.

Dataset	CNN	BiLSTM	CDAN+E	Patient Pred. Acc. (%)
Long-term	✓			32.493 ± 1.259
	✓		✓	23.998 ± 1.209
	✓	✓		33.936 ± 1.368
	✓	✓	✓	<u>25.405 ± 1.268</u>
Short-term	✓			22.152 ± 1.070
	✓		✓	13.802 ± 0.860
	✓	✓		23.350 ± 1.160
	✓	✓	✓	<u>14.579 ± 0.775</u>

3) TEMPORAL MODELING APPROACHES

Fig. 8 presents seizure prediction probabilities for different temporal modeling approaches for Patient T in the short-term recordings. Gray regions indicate seizure periods annotated by an epileptologist. The CNN-BiLSTM demonstrated clear class separation, with prediction probabilities around 0.1 during the non-seizure periods and nearly 1.0 during the seizure period, except near seizure onset and offset (Fig. 8(a)). For the CNN-LSTM, prediction probabilities remained around 0.25 throughout the recording, with a 15-second delay from seizure onset until the probability exceeded 0.5 (Fig. 8(b)). The CNN-Transformer models maintained probabilities around 0.9 during most of the seizure period; however, the probability exceeded 0.5 at multiple points during the non-seizure periods (Fig. 8(c), (d)).

Fig. 9 compares the performance of temporal modeling approaches using window-based metrics. On the long-term recordings, the CNN-BiLSTM outperformed the other approaches in both metrics (Fig. 9(a)). On the short-term recordings, the CNN-BiLSTM achieved the highest MCC, whereas the CNN-Transformer (shallow) achieved the highest AUC-ROC (Fig. 9(b)). Among the CNN-Transformer

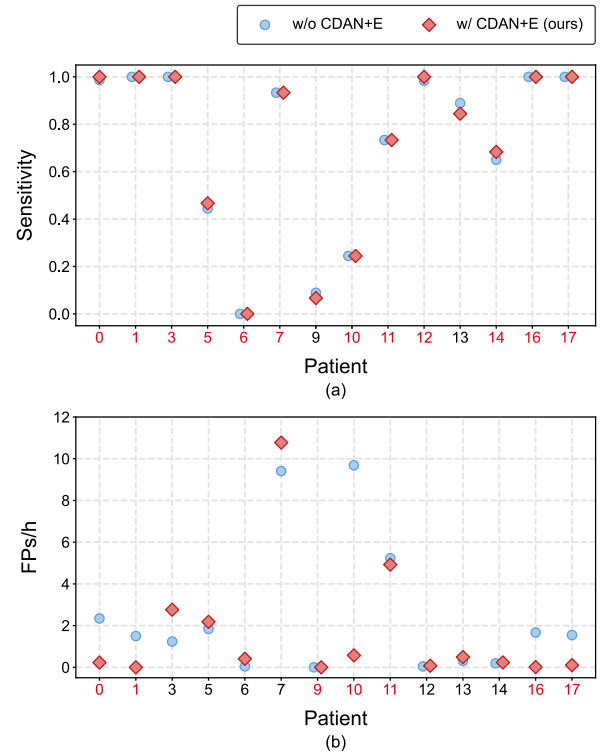


FIGURE 6. Comparison of patient-specific event-based metrics for the CNN-BiLSTM with and without CDAN+E: (a) Sensitivity and (b) FPs/h. Higher sensitivity and lower FPs/h indicate better model performance. Patients highlighted in red achieved equal or superior performance with CDAN+E compared to without it.

configurations, the shallow architecture outperformed the deep architecture across both datasets (Fig. 9).

Fig. 10 shows the comparison of event-based metrics for the long-term recordings. The CNN-LSTM achieved the highest sensitivity, although the differences among the four methods were small (Fig. 10(a)). The CNN-BiLSTM yielded the lowest FPs/h, with the CNN-LSTM following at 7.226, whereas the CNN-Transformer showed relatively higher values of 15.126 and 15.263 for the shallow and deep configurations, respectively (Fig. 10(b)). The two CNN-Transformer configurations performed comparably, with no improvement from increasing the number of layers or attention heads.

4) SUBSEQUENCE LENGTH DEPENDENCY

Fig. 11 presents the effect of subsequence length (consisting of W consecutive windows input to the BiLSTM) on the performance of the proposed method for the long-term recordings. During inference, both the stride and the central segment were set to $W/3$, following the procedure described in Section III-C. For window-based evaluation, sensitivity improved consistently with increasing subsequence length W (Fig. 11(a)). AUC-ROC and pAUC also showed an increasing trend, peaking at $W = 210$. While specificity remained around 0.98, a slight decline was observed as W increased. In contrast, event-based evaluation (Fig. 11(b))

revealed non-monotonic behavior for both sensitivity and FPs/h. Sensitivity was lowest at $W = 60$ with a value of 0.697, then increased with W , and peaked at 0.788 when $W = 300$. Meanwhile, FPs/h peaked at 3.831 at $W = 15$, remained near its minimum (between 1.627 and 2.185) for W ranging from 60 to 150, and rose slightly from $W = 210$ onward.

E. QUALITATIVE EVALUATION

Fig. 12 presents comparative examples of seizure prediction probabilities and detection results for Patients T, C, and L in the short-term recordings. For Patient T, prediction probabilities exceeded the 0.5 threshold during the seizure and remained lower during the non-seizure periods, regardless of whether domain adversarial training was applied (Fig. 12(a)). In contrast, Patient C exhibited false positives when domain adversarial training was not applied (Fig. 12(b)). The prediction probability surpassed 0.5 around 250–300 s during the non-seizure period. Applying DANN or CDAN+E successfully suppressed these values below 0.5. For Patient L, while no false positives occurred during the non-seizure period with any method, differences were observed in detection coverage within the seizure period (Fig. 12(c)). Without domain adversarial training, the prediction probability increased to around 0.8 immediately after seizure onset but decreased within a few seconds, resulting in detection of only the initial portion of the seizure. Although DANN and CDAN+E showed similar trends, CDAN+E maintained probabilities around 0.9 for a longer duration than DANN, yielding higher detection coverage within the seizure period.

V. DISCUSSION

A. EFFECTIVENESS OF THE PROPOSED METHOD

Most existing domain generalization methods rely on CNN-based models that process each time step independently, thereby limiting the effective use of time-series information [18], [19], [20]. In this study, by integrating conditional domain adversarial training with temporal modeling, we established a framework that captures the temporal evolution of seizures while maintaining patient invariance.

Fig. 4(a) shows that patient-specific features dominate the EEG signals. Although the CNN alone acquires some degree of invariance while improving seizure detection accuracy, the separation among patients remains evident (Fig. 4(b)). In contrast, applying CDAN+E results in overlapping patient distributions, forming a patient-invariant feature space (Fig. 4(c)). This change is further supported by the loss progression in Fig. 5. The monotonic decrease in label prediction loss, combined with the increase and subsequent convergence of the domain classification loss, indicates a stable adversarial interplay between the feature extractor and the domain classifier. Through this adversarial training process, the feature extractor learns discriminative features useful for seizure detection while simultaneously acquiring patient invariance.

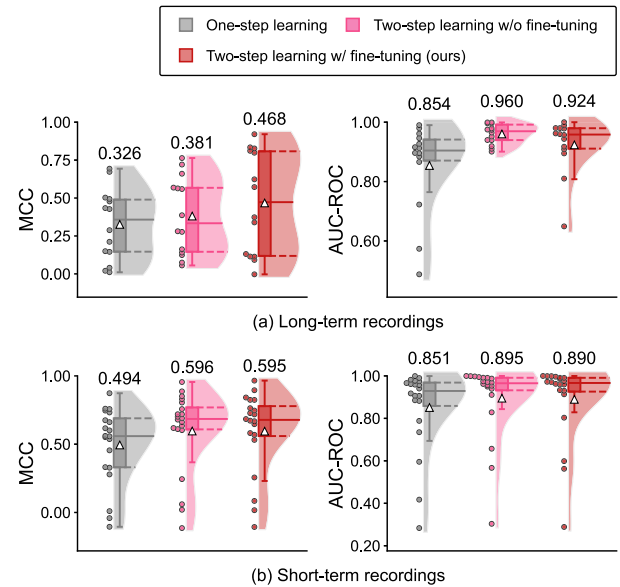


FIGURE 7. Comparison of MCC and AUC-ROC across training strategies: (a) Long-term recordings and (b) Short-term recordings. The plots summarize the distribution of metrics across patients using violin plots (density) and box plots, with individual data points. In the violin plots, solid horizontal lines represent the median, while dashed lines indicate the first and third quartiles. The ‘ Δ ’ symbol indicates the mean value, corresponding to the number shown above each distribution.

The proposed method separates the roles of each component through the two-stage training. In the first stage, the CNN with CDAN+E acquires patient-invariant feature representations while maintaining seizure classification accuracy. In the second stage, the BiLSTM captures the temporal patterns of the acquired invariant features. As shown in Table 5, the performance improvement from applying CDAN+E to the CNN alone is limited, indicating that accurate classification remains difficult without time-series information even when patient-invariant features are acquired. These findings are further supported by patient prediction accuracy, as shown in Table 6. Applying CDAN+E substantially reduced patient prediction accuracy in both the CNN alone and CNN-BiLSTM, confirming that patient-invariant features are acquired through the first-stage training. Nevertheless, incorporating the BiLSTM had little effect on this accuracy, suggesting that the BiLSTM contributes not to acquiring patient invariance but to temporally leveraging the learned invariant features.

Notably, on the short-term recordings, applying CDAN+E to the CNN alone significantly decreased AUC-ROC. This suggests that enforcing patient invariance without temporal context may remove discriminative features necessary for accurate classification, particularly when seizure boundaries are ambiguous. In contrast, incorporating the BiLSTM yielded significant performance improvements on both datasets. Furthermore, when these components were trained jointly in a single stage, seizure detection accuracy decreased substantially (Fig. 7). This decrease can be attributed

TABLE 7. Results on the effectiveness of domain adversarial training methods: mean $\pm$ standard deviation across patients. The best and second-best results for each dataset are indicated in bold and underlined, respectively.

Dataset	Method	Sensitivity	Specificity	MCC	AUC-ROC	AUC-PR	pAUC
Long-term	DANN	0.574 $\pm$ 0.309	0.983 $\pm$ 0.030	0.462 $\pm$ 0.344	0.910 $\pm$ 0.108	0.507 $\pm$ 0.361	0.823 $\pm$ 0.141
	CDAN	0.571 $\pm$ 0.318	0.983 $\pm$ 0.031	0.461 $\pm$ 0.334	0.917 $\pm$ 0.077	0.513 $\pm$ 0.360	0.821 $\pm$ 0.146
	CDAN+E (ours)	0.590 $\pm$ 0.313	<u>0.982</u> $\pm$ 0.031	0.468 $\pm$ 0.338	0.924 $\pm$ 0.093	<u>0.511</u> $\pm$ 0.358	0.830 $\pm$ 0.137
Short-term	DANN	0.743 $\pm$ 0.306	<u>0.899</u> $\pm$ 0.102	0.598 $\pm$ 0.305	<u>0.895</u> $\pm$ 0.183	<u>0.753</u> $\pm$ 0.295	0.805 $\pm$ 0.162
	CDAN	0.732 $\pm$ 0.307	0.895 $\pm$ 0.104	0.582 $\pm$ 0.292	0.899 $\pm$ 0.169	0.753 $\pm$ 0.279	<u>0.802</u> $\pm$ 0.156
	CDAN+E (ours)	<u>0.733</u> $\pm$ 0.316	0.907 $\pm$ 0.100	<u>0.595</u> $\pm$ 0.300	0.890 $\pm$ 0.189	0.754 $\pm$ 0.300	0.805 $\pm$ 0.162

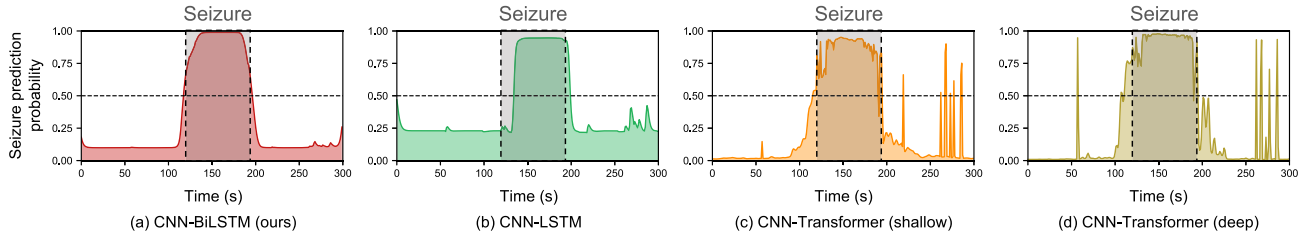


FIGURE 8. Seizure prediction probabilities for Patient T using different temporal modeling approaches in the short-term recordings. Models compared: (a) CNN-BiLSTM (ours), (b) CNN-LSTM, (c) CNN-Transformer (shallow), and (d) CNN-Transformer (deep). Gray shaded regions indicate seizure periods annotated by an epileptologist. Horizontal dashed lines represent the detection threshold, fixed at 0.5.

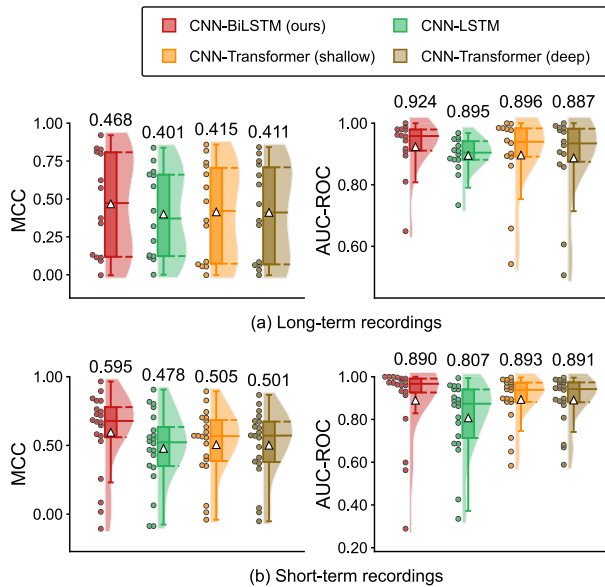


FIGURE 9. Comparison of MCC and AUC-ROC across temporal modeling architectures: (a) Long-term recordings and (b) Short-term recordings. The plots summarize the distribution of metrics across patients using violin plots (density) and box plots, with individual data points. In the violin plots, solid horizontal lines represent the median, while dashed lines indicate the first and third quartiles. The 'Δ' symbol indicates the mean value, corresponding to the number shown above each distribution.

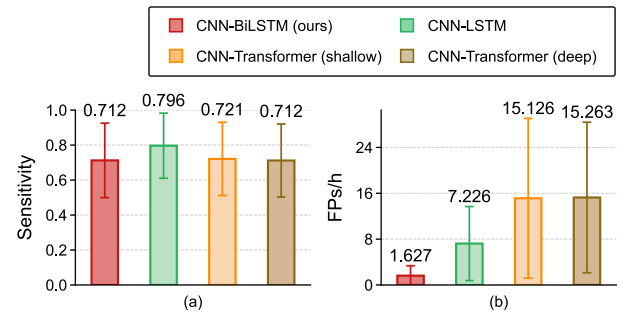


FIGURE 10. Comparison of event-based metrics: (a) Sensitivity and (b) FPs/h. Higher sensitivity and lower FPs/h indicate better model performance. Error bars represent 95% confidence intervals, and the values above the bars indicate the mean.

to the difficulty of simultaneously optimizing for patient invariance and temporal patterns. These results demonstrate the effectiveness of the role separation via two-stage training, in which CDAN+E serves to acquire patient invariance and the BiLSTM focuses on leveraging temporal context. Notably, one-step and two-step learning share the same inference architecture, indicating that the observed improvement reflects the effectiveness of the learning strategy rather than increased model capacity.

In the comparison of temporal modeling approaches (Figs. 8 and 9), the unidirectional LSTM tended to output high prediction probabilities during non-seizure periods (Fig. 8(b)), suggesting difficulty in determining seizure offset based solely on past information. Both Transformer configurations showed unstable prediction probabilities during non-seizure periods (Fig. 8(c), (d)). Since this instability persisted despite increasing the number of layers and attention heads, it can be attributed to overfitting given the limited amount of data, rather than insufficient model capacity. In contrast, the BiLSTM achieved stable prediction by effectively capturing seizure onset and offset through bidirectional temporal integration. This difference in prediction stability across methods was particularly evident in the clinically important event-based metrics (Fig. 10(b)). While sensitivity did not differ substantially among the methods, the CNN-BiLSTM achieved notably lower FPs/h than the other methods. In particular, both Transformer configurations exhibited approximately nine times higher

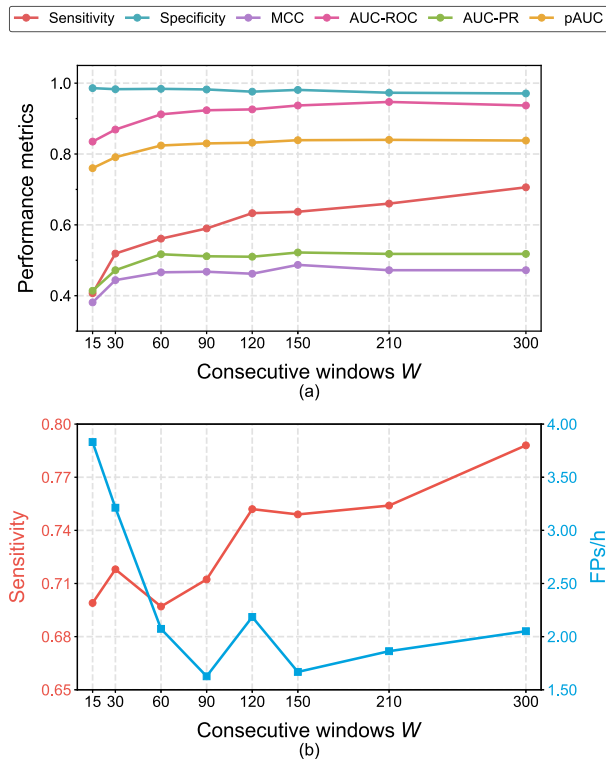


FIGURE 11. Averaged metrics for the CNN-BiLSTM with CDAN+E (ours) as a function of the number of consecutive windows W for the long-term recordings: (a) Window-based metrics and (b) Event-based metrics. Higher sensitivity and lower FPs/h indicate better model performance.

FPs/h than the CNN-BiLSTM, demonstrating that unstable prediction probabilities during non-seizure periods directly led to increased false positives at the event level. For EEG signals, temporal continuity is crucial for stable seizure detection. Consequently, under limited data conditions, integrating bidirectional context while preserving sequential order, as in the BiLSTM, appears more effective than the parallel attention mechanism of the Transformer.

B. RELATIONSHIP BETWEEN DATASET CHARACTERISTICS AND PERFORMANCE

A clear difference in performance trends was observed between the long-term and short-term recordings used in this study. This difference is attributable to the nature of non-seizure periods in each dataset. Both datasets consist of real clinical data; however, the long-term recordings are continuous, spanning several hours to days, encompassing sleep-wake cycles and diverse background EEG activities, and inherently include real-world clinical noise such as motion artifacts and electrode displacement. In contrast, the short-term recordings consist of approximately 300 s segments extracted around seizures, with non-seizure periods primarily comprising peri-ictal transition states.

In the long-term recordings, CDAN+E outperformed other domain adversarial training methods (Table 7). This result can be attributed to the inclusion of diverse non-seizure

states, which increases the distributional difference between seizure and non-seizure segments, thereby allowing the conditioning to be effective. Specifically, this conditioning enables the extraction of common seizure features while reducing cross-patient variability within the seizure class, as well as the learning of features common to diverse states such as sleep and wakefulness within the non-seizure class. In addition to achieving high threshold-independent metrics, the proposed method enabled stable seizure detection even with a fixed threshold (Table 4). These results have important implications for application in real-world clinical settings involving long-term recordings.

In the short-term recordings, the proposed method achieved the highest values for specificity and MCC (Table 4). This suggests that accounting for individual differences in EEG patterns enables the model to detect seizure periods more accurately, even in ambiguous states near seizure boundaries. On the other hand, the limited diversity of non-seizure periods in the approximately 300 s recordings restricted the effect of conditioning, resulting in a small performance gap between DANN and CDAN+E (Table 7). Nevertheless, CDAN+E outperformed CDAN, confirming the effectiveness of entropy weighting in emphasizing samples with high prediction confidence.

Performance improvements through domain adversarial training were observed in both datasets, demonstrating the importance of accounting for individual differences in EEG patterns. The effect of conditioning depended on dataset characteristics, with a greater impact observed in the long-term recordings due to the higher diversity of non-seizure periods.

C. IMPLICATIONS FOR CLINICAL APPLICATION

In this study, we evaluated event-based detection performance using entire long-term recordings. Many previous studies have relied on segments extracted around seizures or random sampling from long-term recordings [18], [19], which may overestimate performance in real-world scenarios [39]. Our evaluation setting, which includes sleep-wake cycles and prolonged non-seizure periods, provides a more realistic assessment for clinical application.

The primary clinical scenario targeted in this study is review support for long-term video-EEG monitoring. The manual review of these recordings, which span several days, imposes a heavy burden on specialists [6], [7]. In such settings, the ability to accurately detect seizure events is crucial. The event-based evaluation used in this study addresses this need, where clinical utility is determined by the balance between minimizing missed seizures (high sensitivity) and reducing false alarms (low FPs/h).

The proposed method with CDAN+E reduced FPs/h while maintaining sensitivity compared to the method without CDAN+E (Fig. 6). The mean FPs/h across patients decreased from 2.503 to 1.627, and four patients (PN00, PN01, PN16, and PN17) achieved ideal detection performance with a

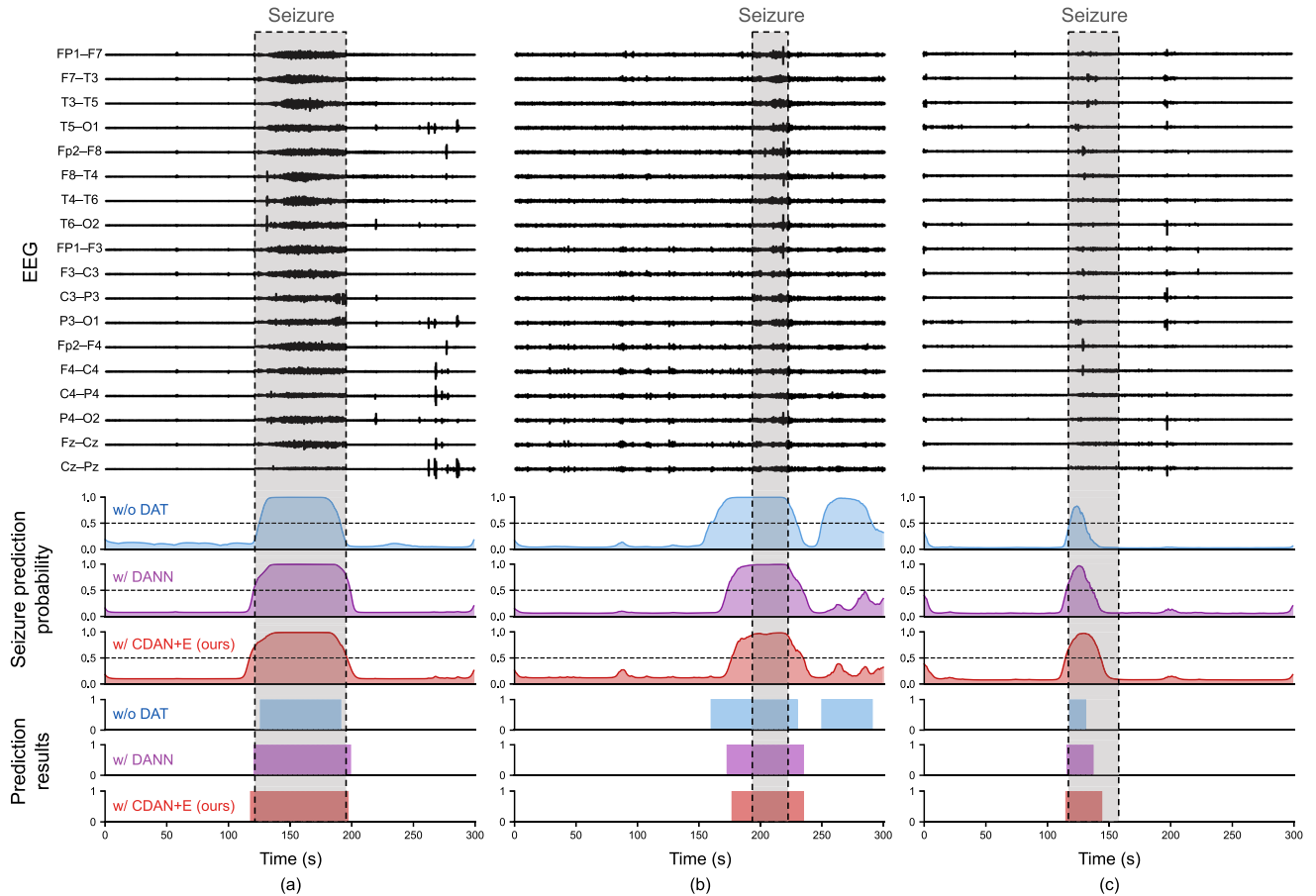


FIGURE 12. Preprocessed EEG signals and seizure detection results in the short-term recordings for (a) Patient T, (b) Patient C, and (c) Patient L. Results are shown for the CNN-BiLSTM without domain adversarial training (DAT), with DANN, and with CDAN+E (ours). Gray shaded regions indicate seizure periods annotated by an epileptologist. Horizontal dashed lines in the prediction plots represent the detection threshold, fixed at 0.5.

sensitivity of 1.0 and FPs/h near 0. Although sensitivity was high for these patients even without CDAN+E, the reduction in FPs/h was more pronounced with its incorporation. Reducing false alarms is crucial for clinical application. When physicians must verify each detected seizure, false alarms directly increase their workload. Furthermore, frequent false alarms can erode trust in the system, potentially leading to inadequate responses to actual seizures.

A practical strength of the proposed method is that it does not require any data from new patients for training. Unlike methods based on unsupervised domain adaptation, which require EEG recordings of new patients during training [16], [17], the proposed method is trained solely on existing patient data, enabling immediate application to new patients. Additionally, the network during inference consists solely of the CNN-BiLSTM and the FC layer, as the domain classifier is used only during training; thus, it is equivalent in computational cost to that of a model without domain adversarial training. These practical advantages lower the barrier to clinical adoption, such as automatic extraction of candidate seizure segments during long-term video-EEG

monitoring. Furthermore, as shown in Table 4, the high threshold-independent evaluation metrics indicate that the trade-off between sensitivity and FPs/h can be flexibly adjusted by tuning the threshold to suit clinical needs.

The selection of the subsequence length W also allows for adjustments based on clinical requirements (Fig. 11). Overall, increasing W tended to improve model sensitivity, a trend particularly pronounced in event-based metrics. However, a larger W leads to an increased detection delay and higher computational cost. Although $W = 90$ offers a reasonable trade-off between FPs/h and detection delay, a larger W can be selected if preventing missed seizures is prioritized.

D. LIMITATIONS

This study aims to methodologically evaluate the proposed framework integrating conditional domain adversarial training and temporal modeling. We limited the seizure type to focal seizures and adopted datasets with comparable sampling frequencies (500–512 Hz) to reduce the influence of confounding factors. While these design choices ensure the clarity of our evaluation, several aspects remain for

future work. First, applicability to other seizure types, such as generalized seizures, has not been evaluated. Because seizure manifestations differ substantially across types (e.g., generalized seizures involve bilateral hemispheric activity, unlike the localized onset patterns typical of focal seizures), the patient-invariant features learned from focal seizure data may not generalize to other seizure types. In clinical settings, seizure types can often be estimated through medical interviews; thus, constructing type-specific models is a possible approach. However, considering cases where seizure types cannot be identified in advance or patients present with multiple seizure types, robustness to diverse seizure types remains an important issue. Future evaluations using datasets containing diverse seizure types are necessary to assess the model's generalizability.

Additionally, the datasets used in this study comprise only 14 patients for the long-term recordings and 20 patients for the short-term recordings. Patient-invariant features extracted from a small number of patients may reflect biases inherent to the target population, and generalization performance to patient populations with more diverse age groups or clinical backgrounds is not guaranteed. Future work should expand the number of patients to evaluate the robustness of the extracted invariant features.

Furthermore, interpretability and noise robustness remain challenges for clinical application. Since the localization of the epileptogenic zone in focal seizures varies across patients, identifying the EEG channels that contribute to detection is crucial from a clinical perspective. Moreover, although the real-world clinical EEG data used in this study inherently include motion and electrode-related artifacts, a systematic evaluation of robustness against specific artifacts remains a technical challenge.

Finally, for clinical application, performance that is independent of recording conditions and facility-specific characteristics is essential. Although leave-one-patient-out cross-validation was conducted within each dataset, cross-dataset evaluation, where a model trained on one dataset is applied to the other, was not performed. While this study treats each patient as a distinct domain, cross-dataset settings involve hierarchical domain structures, where factors such as clinical facilities, equipment, and protocols serve as higher-level sources of variance. Addressing such structures would require revisiting the domain definition, which remains an important challenge for real-world clinical deployment.

VI. CONCLUSION

In this paper, we proposed a two-stage learning framework incorporating conditional domain adversarial training for patient-invariant time-series epileptic seizure detection. In the first stage, CDAN+E is applied to a CNN to acquire patient-invariant features adaptive to seizure and non-seizure states. In the second stage, a BiLSTM is connected to capture the temporal patterns of the acquired patient-invariant features.

We evaluated the effectiveness of the proposed method from multiple perspectives using an open dataset (Siena Scalp EEG Database, long-term recordings) and clinical EEG recordings (short-term recordings). The long-term recordings were used to evaluate performance in realistic clinical settings, while the short-term recordings were used to assess classification accuracy at seizure boundaries. The results demonstrated that both conditional domain adversarial training and temporal modeling contribute to improving model performance. In particular, the effect of extracting patient-invariant features conditioned on each state was notable in the long-term recordings, which include sleep-wake cycles and diverse background EEG activities. Furthermore, in seizure event-based evaluation, which is important in clinical practice, the proposed method achieved a sensitivity of 0.712 and 1.627 FPs/h.

This study treated each patient as a distinct domain; however, EEG patterns exhibit temporal variability resulting from sleep-wake cycles and long-term pathological changes. In future work, we aim to further improve detection performance by expanding domain definitions to account for such intra-patient variability and by integrating the proposed approach with patient-specific seizure detection methods.

REFERENCES

- [1] World Health Organization. Epilepsy Fact Sheet. [Online]. Available: <http://www.who.int/news-room/fact-sheets/detail/epilepsy>
- [2] C. Camfield and P. Camfield, "Injuries from seizures are a serious, persistent problem in childhood onset epilepsy: A population-based study," *Seizure*, vol. 27, pp. 80–83, Apr. 2015.
- [3] B. de Almeida Fischer and G. M. A. S. Tedrus, "Seizure-related injuries in adults: A prospective case-controlled study on risk factors, seizure severity and quality of life," *Epilepsy Behav.*, vol. 134, Sep. 2022, Art. no. 108849.
- [4] J. Taylor, R. Kolamunnage-Dona, A. G. Marson, P. E. M. Smith, A. P. Aldenkamp, and G. A. Baker, "Patients with epilepsy: Cognitively compromised before the start of antiepileptic drug treatment?" *Epilepsia*, vol. 51, no. 1, pp. 48–56, Jan. 2010.
- [5] A. Novak, K. Vizjak, and M. Rakusa, "Cognitive impairment in people with epilepsy," *J. Clin. Med.*, vol. 11, no. 1, Jan. 2022, Art. no. 267.
- [6] C.-H. Lee, S.-N. Lim, F. Lien, and T. Wu, "Duration of electroencephalographic recordings in patients with epilepsy," *Seizure*, vol. 22, no. 6, pp. 438–442, Jul. 2013.
- [7] S. R. Benbadis, S. Beniczky, E. Bertram, S. MacIver, and S. L. Moshé, "The role of EEG in patients with suspected epilepsy," *Epileptic Disorders*, vol. 22, no. 2, pp. 143–155, Apr. 2020.
- [8] B. Hunyadi, M. Signoretto, W. Van Paesschen, J. A. K. Suykens, S. Van Huffel, and M. De Vos, "Incorporating structural information from the multichannel EEG improves patient-specific seizure detection," *Clin. Neurophysiol.*, vol. 123, no. 12, pp. 2352–2361, Dec. 2012.
- [9] M. Kaleem, A. Guergachi, and S. Krishnan, "Patient-specific seizure detection in long-term EEG using wavelet decomposition," *Biomed. Signal Process. Control*, vol. 46, pp. 157–165, Sep. 2018.
- [10] Y. Zhang, S. Yao, R. Yang, X. Liu, W. Qiu, L. Han, W. Zhou, and W. Shang, "Epileptic seizure detection based on bidirectional gated recurrent unit network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 135–145, Jan. 2022.
- [11] P. Zhu, W. Zhou, C. Cao, G. Liu, Z. Liu, and W. Shang, "A novel SE-TCN-BiGRU hybrid network for automatic seizure detection," *IEEE Access*, vol. 12, pp. 127328–127340, May 2024.
- [12] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthc. Conf. (MLHC)*, vol. 56, Aug. 2016, pp. 178–190.

- [13] J. Craley, E. Johnson, C. Jouny, and A. Venkataraman, "Automated inter-patient seizure detection using multichannel convolutional and recurrent neural networks," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102360.
- [14] Y. Zhu and M. D. Wang, "Automated seizure detection using transformer models on multi-channel EEGs," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Oct. 2023, pp. 1–6.
- [15] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Effective early detection of epileptic seizures through EEG signals using classification algorithms based on t-distributed stochastic neighbor embedding and K-means," *Diagnostics*, vol. 13, no. 11, Jun. 2023, Art. no. 1957.
- [16] P. Peng, L. Xie, K. Zhang, J. Zhang, L. Yang, and H. Wei, "Domain adaptation for epileptic EEG classification using adversarial learning and Riemannian manifold," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103555.
- [17] H. Feng, S. Wang, H. Lv, C. Nie, W. Feng, H. Peng, and Y. Zhao, "Cross-subject seizure detection with vision transformer and unsupervised domain adaptation," *Biomed. Signal Process. Control*, vol. 111, Jan. 2026, Art. no. 108341.
- [18] X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, and Y. Li, "Adversarial representation learning for robust patient-independent epileptic seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2852–2859, Oct. 2020.
- [19] Y. Yang, F. Li, X. Qin, H. Wen, X. Lin, and D. Huang, "Feature separation and adversarial training for the patient-independent detection of epileptic seizures," *Frontiers Comput. Neurosci.*, vol. 17, Jul. 2023, Art. no. 1195334.
- [20] Z. Zhang, T. Ji, M. Xiao, W. Wang, G. Yu, T. Lin, Y. Jiang, X. Zhou, and Z. Lin, "Cross-patient automatic epileptic seizure detection using patient-adversarial neural networks with spatio-temporal EEG augmentation," *Biomed. Signal Process. Control*, vol. 89, Mar. 2024, Art. no. 105664.
- [21] R. Tazaki, T. Akiyama, and A. Furui, "EEG-based inter-patient epileptic seizure detection combining domain adversarial training with CNN-BiLSTM network," in *Proc. 47th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2025, pp. 1–6.
- [22] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Jul. 2015, pp. 1180–1189.
- [23] A. Graves and J. Schmidhuber, "Frame-wise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.
- [24] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, Dec. 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/hash/ab88b15733f543179858600245108dd8-Abstract.html
- [25] H. Jing and M. Takigawa, "Comparison of human ictal, interictal and normal non-linear component analyses," *Clin. Neurophysiol.*, vol. 111, no. 7, pp. 1282–1292, Jul. 2000.
- [26] H.-P. Landolt and A. A. Borbély, "Age-dependent changes in sleep EEG topography," *Clin. Neurophysiol.*, vol. 112, no. 2, pp. 369–377, Feb. 2001.
- [27] P. J. Marshall, Y. Bar-Haim, and N. A. Fox, "Development of the EEG from 5 months to 4 years of age," *Clin. Neurophysiol.*, vol. 113, no. 8, pp. 1199–1208, Aug. 2002.
- [28] U. R. Acharya, O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," *Comput. Methods Programs Biomed.*, vol. 80, no. 1, pp. 37–45, Oct. 2005.
- [29] Z. A. Gaál, R. Boha, C. J. Stam, and M. Molnár, "Age-dependent features of EEG-reactivity—Spectral, complexity, and network characteristics," *Neurosci. Lett.*, vol. 479, no. 1, pp. 79–84, Jul. 2010.
- [30] J. L. Alcalá-Zermeno, R. Katyal, B. Frauscher, D. Schomer, M. R. Sperling, R. Strowd, W. O. Tatum, E. Wirrell, S. Beniczky, and F. A. Nascimento, "Seminar in epileptology: Normal awake and sleep patterns, interictal abnormalities, and ictal patterns on scalp EEG," *Epileptic Disorders*, vol. 27, no. 5, pp. 803–866, Oct. 2025.
- [31] P. van Mierlo, B. J. Vorderwülbecke, W. Staljanse, M. Seeck, and S. Vulliemoz, "Ictal EEG source localization in focal epilepsy: Review and future perspectives," *Clin. Neurophysiol.*, vol. 131, no. 11, pp. 2600–2616, Nov. 2020.
- [32] P. Detti, "Siena scalp EEG database (version 1.0.0)," PhysioNet, RRID:SCR_007345, 2020. [Online]. Available: <https://doi.org/10.13026/5d4a-j060> and <https://physionet.org/content/siena-scalp-eeeg/1.0.0/>
- [33] P. Detti, G. Vatti, and G. Z. M. de Lara, "EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings," *Processes*, vol. 8, no. 7, p. 846, Jul. 2020.
- [34] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
- [35] A. Furui, R. Onishi, A. Takeuchi, T. Akiyama, and T. Tsuji, "Non-Gaussianity detection of EEG signals based on a multivariate scale mixture model for diagnosis of epileptic seizures," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 515–525, Feb. 2021.
- [36] D. H. Agustsson and S. Gudmundsson, "Convolutional neural networks for seizure detection: A study on training strategies," in *Proc. 47th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2025, pp. 1–5.
- [37] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.
- [38] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res. (JMLR)*, vol. 9, no. 86, pp. 2579–2605, 2008.
- [39] E. Ali, M. Angelova, and C. Karmakar, "Epileptic seizure detection using CHB-MIT dataset: The overlooked perspectives," *Roy. Soc. Open Sci.*, vol. 11, no. 5, May 2024, Art. no. 230601.



RINA TAZAKI (Graduate Student Member, IEEE) received the B.S. degree from Hiroshima University, Hiroshima, Japan, in 2025, where she is currently pursuing the M.S. degree with the Graduate School of Advanced Science and Engineering. Her research interests include deep learning, epilepsy, and domain generalization.



TOMOYUKI AKIYAMA received the M.D. and Ph.D. degrees from Okayama University, Okayama, Japan, in 1995 and 1999, respectively. He is currently an Associate Professor with the Department of Pediatric Neurology, Okayama University Hospital.



AKIRA FURUI (Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees from Hiroshima University, Hiroshima, Japan, in 2016, 2018, and 2019, respectively. From 2018 to 2019, he was a Research Fellow with Japan Society for the Promotion of Science. He is currently an Associate Professor with the Informatics and Data Science Program, Graduate School of Advanced Science and Engineering, Hiroshima University. His research interests include biosignal processing, human interface, machine learning, medical image processing, and stochastic modeling.

...