
DMNet: Self-comparison Driven Model for Subject-independent Seizure Detection

Shihao Tu
Zhejiang University
shihao.tu@zju.edu.cn

Linfeng Cao
The Ohio State University
cao.1378@osu.edu

Daoze Zhang
Zhejiang University
zhangdz@zju.edu.cn

Junru Chen
Zhejiang University
jrchen_cali@zju.edu.cn

Lvbin Ma
Zhejiang Huayun
Information Technology Co. Ltd
gmmfly@163.com

Yin Zhang
Zhejiang University
yinzhang@zju.edu.cn

Yang Yang[†]
Zhejiang University
yangya@zju.edu.cn

Abstract

Automated seizure detection (ASD) using intracranial electroencephalography (iEEG) is critical for effective epilepsy treatment. However, the significant domain shift of iEEG signals across subjects poses a major challenge, limiting their applicability in real-world clinical scenarios. In this paper, we address this issue by analyzing the primary cause behind the failure of existing iEEG models for *subject-independent* seizure detection, and identify a critical universal seizure pattern: seizure events consistently exhibit higher average amplitude compared to adjacent normal events. To mitigate the domain shifts and preserve the universal seizure patterns, we propose a novel *self-comparison* mechanism. This mechanism effectively aligns iEEG signals across subjects and time intervals. Based on these findings, we propose *Difference Matrix-based Neural Network* (DMNet), a subject-independent seizure detection model, which leverages self-comparison based on two constructed (*contextual, channel-level*) references to mitigate shifts of iEEG, and utilize a simple yet effective *difference matrix* to encode the universal seizure patterns. Extensive experiments show that DMNet significantly outperforms previous SOTAs while maintaining high efficiency on a real-world clinical dataset that we collected, as well as two public datasets for subject-independent seizure detection. Moreover, the visualization results demonstrate that the generated difference matrix can effectively capture the seizure activity changes throughout the seizure evolution process. Additionally, we deploy our method in an online diagnosis system to illustrate its effectiveness in real clinical applications.

1 Introduction

Epilepsy, a chronic neurological disorder, affects more than 65 million people around the world. Up to 70% of people with epilepsy can be free from seizure only if the seizure onset zone (SOZ) can be located and surgically removed [26]. To diagnose epilepsy, doctors rely on the assessment of

[†]Corresponding authors.

electrical activities that reflect the state and function of the subject’s brain. Electroencephalography (EEG) is a widely employed and cost-effective method to record these electrical activities by placing sensors on the scalp. However, as a non-invasive method, it is unable to accurately locate SOZ in the deep structures of the brain.

Nowadays, iEEG is widely employed to identify and locate SOZ. *Stereo-EEG*, one representative iEEG technique, involves the deep implantation of electrodes within the brain to record electrical activities. These electrodes contain multiple recording contacts, called *channels*, and are placed across different regions of the brain, which provide stereoscopic recordings of the brain from both cortical and subcortical structures simultaneously [30]. This fully developed technique has been proven to be both effective and safe [5].

Given a substantial volume of iEEG data, we present a pipeline tailored to real-world application scenarios for automated seizure detection (ASD), as depicted in Fig. 1 (g). Firstly, the ASD model is trained using data from accessible subjects. Subsequently, the trained model is applied to identify seizures in iEEG recordings from previously unseen subjects. Doctors can refer to the prediction results, allowing for a more accurate diagnosis and facilitating more effective treatment decisions.

However, most existing ASD methods are built on non-invasive EEG with a *subject-specific* setting [38, 31] and *subject-independent* setting [2, 44, 11]. However, these EEG-based methods are prone to failure when applied to iEEG data. This is primarily due to the significantly higher complexity of iEEG signals compared to EEG signals. iEEG signals exhibit a greater level of intricacy as a result of the structural and functional disparities in brain neural activities. Furthermore, there are variations in the number and placement of invasive electrodes (Fig. 1 (a,b)), leading to significant domain shifts across different subjects. However, Existing methods that employ domain adversarial training for *subject-independent* seizure detection on iEEG signals may encounter negative transfer effects [23, 27]. Although [41] proposed a method that utilizes a series of intricate pre-training strategies to learn the general pattern across subjects, it lacks efficiency. Consequently, developing an effective and efficient subject-independent ASD method using iEEG is crucial for clinical diagnosis. Here, we discuss the primary challenges associated with iEEG in ASD.

Challenges. The factors mentioned above cause a significant domain shift between subjects, posing an open question regarding the generalization of *subject-independent* epilepsy seizure detection using iEEG. This issue gives rise to challenges at both the subject level and the channel level:

(1) How to capture the general distinguishable representation for normal and seizures between different subjects and time intervals? Due to individual differences that exist among subjects, the inherent properties of iEEG recordings, such as amplitude, frequency, and others, are personalized to each subject. Even within the same individual, brain activities vary over time.

(2) How to reduce the inconsistency of the seizure patterns of different channels? The channels exhibit diverse patterns due to the iEEG records of various regions of the brain, potentially leading to conflicting patterns between subjects. For example, normal and seizure waves are indistinguishable between subjects or channels. As shown in Fig. 1, the normal wave in (e) is difficult to distinguish from the seizure waves in (c) and (d), while the seizure event in (f) is often mistaken for the normal one. Therefore, the second challenge is to personalize the representations of brain activity independent of channels, allowing the model to adapt to different subjects.

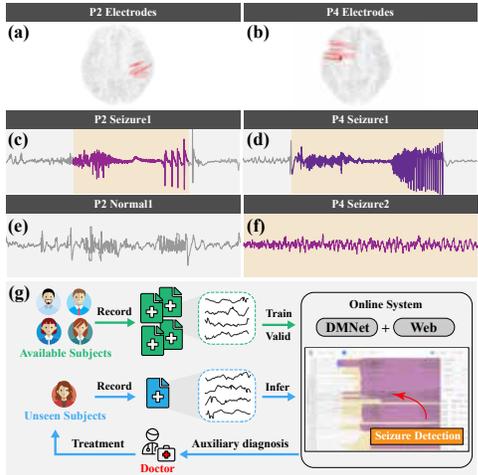


Figure 1: (a, b) Locations of iEEG depth electrode contacts (red circle) for subjects *P2* and *P4*. (c, d, e, f) Examples of seizure and normal iEEG recording activities of subjects *P2* and *P4*. (g) Application of our proposed method in real-world clinical scenarios.

Solution. To address these challenges, we first propose the mechanism of *self-comparison*: *comparing the target segment with adjacent normal segments*. Subsequently, we conduct a comprehensive observation study (Sec. 3) to demonstrate the effectiveness of *self-comparison*. Our findings indicate that *self-comparison* can obtain a general distinguishable representation of normal ones and seizures.

Drawing on these inspirations, we argue that the *self-comparison* mechanism is the key to easily and effectively capture subject-invariant patterns between subjects. To this end, we propose a novel model, namely *Difference Matrix based neural Network (DMNet)*, for subject-independent seizure detection. Specifically, considering that different seizure events would present different neural activities (local bias) within different recording channels (global bias), we therefore introduce two reference objects (i.e., *contextual reference* and *channel-level reference*). These references ensure that we can capture both local and global dependencies within data, which are the primary contributors of distribution shift and can be effectively mitigated through the self-comparison mechanism. Subsequently, we utilize a simple yet effective *fully differencing operation* to generate the difference matrix, which compares the target segment with its reference objects for *self-comparison* implementation. To effectively extract semantics from the difference matrix, we design a difference matrix encoder based on convolutional neural network (CNN) blocks to obtain the final representation of the detection segment. Our primary contributions are listed as follows:

- We investigate the problem of subject-independent ASD based on the iEEG. Through comprehensive analysis, we identify the *self-comparison* mechanism as a simple yet effective way to capture the general representation.
- We propose a novel model named **DMNet** for subject-independent ASD. The *fully differencing operation* based on *contextual reference* and *channel-level reference* for self-comparison can mitigate the local and global biases among subjects and channels, improving the generalizability of learned representations.
- Extensive experiments on clinical and public iEEG datasets show *DMNet* outperforms existing SOTAs. Moreover, the generated difference matrix effectively captures seizure activity changes during the seizure evolution process. Furthermore, *DMNet* outperforms existing SOTAs while maintaining the high efficiency. Building on these strengths, we deploy our method in an online system, enhancing clinical applications by assisting medical professionals in the diagnosis of epilepsy and by facilitating the provision of more effective treatment options for patients.

2 Problem Formulation

In this work, the iEEG recording is regarded as a set of time series $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=1}^C$, where C refers to the total number of channels. Each $\mathbf{x}^{(i)} \in \mathbb{R}^T$ corresponds to a channel, and T refers to the total timestamp. We sequentially test the channels one after another. Specifically, given *one channel* of iEEG time series $\mathbf{x}^{(i)} = \{x_1^{(i)}, \dots, x_T^{(i)}\}$ from a subject, we first divide the original recording data into segments for detection. For simplicity, we omit the channel index i in subsequent steps:

$$\{s_0, s_1, \dots, s_{m-1}\} \quad (1)$$

where $s_k = \{x_{\ell \times k + 1}, \dots, x_{\ell \times (k+1)}\}$, $0 \leq k \leq m - 1$, ℓ is the number of timestamp for each segment, m is the total number of segments. Each segment s_k has a corresponding label $y_k \in \{0, 1\}$, indicating whether the segment contains a seizure event. In this work, our aim is to predict y_k on each segment s_k for different subjects.

We define our problem as an innovative study of *Domain Generalization (DG)* [45] in the context of epileptic diagnosis. In this study, we treat each subject’s data as a domain, and our goal is to utilize the data of available labeled subjects (source domains) to train a model that can be directly adopted to the subjects with unseen data (target domains).

3 Empirical Analysis

In this section, we first analyze the primary cause behind the failure of existing models for subject-independent seizure detection of the iEEG and the reasons of occurrence. Then we explore the possibility of a domain-consistent seizure pattern existing within iEEG data, taking into account the

domain shift issue. Finally, we validate whether *self-comparison* mechanism can mitigate distribution shifts in iEEG data and capture the potential domain consistent seizure patterns.

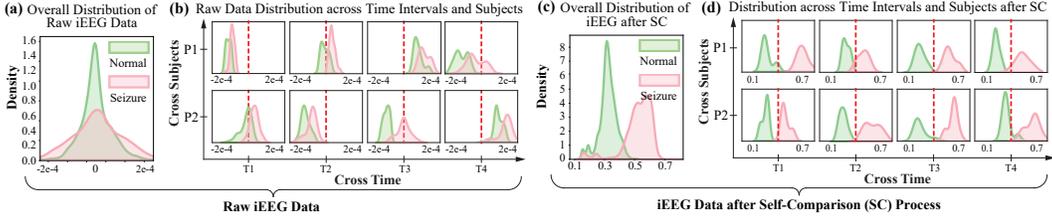


Figure 2: Observation results of the clinical iEEG dataset. (a) Overall distribution of raw iEEG signals (all subjects and channels), where normal and seizure events are indistinguishable. (b) Distribution of raw iEEG signals across different time intervals and subjects, where substantial domain shifts are evident in both distinct time intervals and among different subjects (red dashed line at 0 serves as reference line for domain shift). (c) Overall distribution of raw iEEG data after the *self-comparison* process. (d) Distribution of raw iEEG data across different time intervals and subjects after the *self-comparison* process. The *self-comparison* mechanism effectively mitigates distribution shifts across time intervals and subjects, thus enhancing the model’s ability to distinguish between seizure and normal events.

3.1 iEEG Domain Shift Issue and Domain Consistent Seizure Pattern

Most previous studies [10, 16] assert that vanilla detection models trained independently by each subject are prone to failure when applied to other subjects. To analyze the direct cause of failure, we merge raw iEEG signals from all subjects and channels in a clinical iEEG dataset (details in App. C), and analyze the distributions of seizure and normal events. As depicted in Fig. 2 (a), the results reveal nearly identical means and close variances for both seizure and normal signals in the merged data. This similarity leads to the indistinguishability between normal and seizure signals across subjects, which becomes a direct factor to the failure of subject-dependent models.

For a detailed analysis, we partition the iEEG signals into multiple intervals, each comprising 250 timestamps, for each subject. We then compute the distributions of seizure and normal events within each interval. Through empirical analysis, we observe significant domain shifts both across different subjects (inter-subject) and different time intervals for the same subject (intra-subject). Fig. 2 (b) presents the normal and seizure distributions within four intervals randomly sampled from subjects P1 and P2, where the distribution patterns of normal and seizure exhibit substantial variation across different time intervals and subjects. This observation aligns with findings from previous studies in neural science and medicine, which have consistently reported that brain signals exhibit high variability between subjects and sessions due to inherent background neural activities [33], seizure patterns [14], electrode locations [9], etc.

Despite the pronounced domain shift observed in iEEG signals across subjects and time intervals, there is a notable commonality among subjects. Specifically, seizure events consistently demonstrate a higher average amplitude in the frequency domain compared to their background signal (adjacent normal events), indicating more intense neural oscillatory activities. This finding is consistent with previous studies in the field [34, 21].

3.2 Self-Comparison can Help

Based on the commonality above, we propose a novel *self-comparison* mechanism, which compares the target segment with its adjacent normal segment, to mitigate domain shifts between subjects and time intervals. To verify the effectiveness of this mechanism, we conduct an empirical study. First, for each subject, we partition the contiguous iEEG data into small segments. Considering that the spectral signal of brain data can effectively track transient changes before and during seizures [6], we proceed to transform these segments into the frequency domain using Discrete Fourier Transform (DFT). Subsequently, we calculate the spectrum differences by subtracting the spectrum of the target segment from those of adjacent segments on both sides of the target segment. We then sum up the absolute values of these differences, generating a single value D for each target segment.

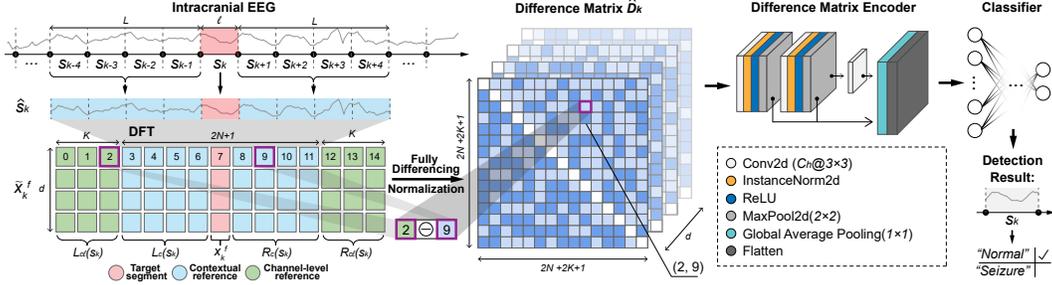


Figure 3: Overview of the proposed DMNet.

The overall distributions of D values for all seizure and normal segments are depicted in Figure 3 (c). Notably, the distribution of the normal and seizure segments becomes distinctly separated after the adoption of the self-comparison mechanism. Additionally, we analyze the distribution of D values for target segments within the same four intervals from P1 and P2, which is discussed in Section 3.1. The results are shown in Figure 3 (d). It is evident that the distribution of normal or seizures segments is well aligned across different subjects and time intervals, and the patterns of normal and seizure become more distinguishable. These results signify that our proposed self-comparison mechanism effectively mitigates the domain shift issue and preserves a domain consistent and distinguishable representation for normal and seizure segments.

4 Methodology

Inspired by the above observations, we propose a subject-independent seizure detection framework called **DMNet**, which leverages self-comparison to alleviate the distribution shift and preserve the domain consistency while distinguishing seizure patterns.

Overview. The overall framework is illustrated in Fig. 3. First, for each detection segment s_k (red box), we construct two reference segments (*contextual reference* $L_c(s_k) / R_c(s_k)$, *channel-level reference* $L_{cl}(s_k) / R_{cl}(s_k)$) to compare with the target segment (Fig. 3, left). Then we use a simple yet effective *fully differencing* operation with signed-min-max normalization for self-comparison implementation, and the compared information is encoded by a *difference matrix* (DM) (Fig. 3, middle). Then we employ a CNN-based difference matrix encoder to learn the latent representation of DM, and use a classifier for seizure detection (Fig. 3, right).

4.1 References for Self-Comparison

Contextual Reference. The structural and functional differences in brain neural activity lead to distribution shifts between subjects, and even across different time intervals within the same subject. Based on the insights from observation studies in Section 3.2, we propose to compare the seizure wave with its adjacent contexts for the identification of seizure activities, as it significantly mitigates the domain shifts between subjects and preserves a domain-consistent and distinguishable pattern of seizure events.

However, capturing the long dependencies between seizure events and their contexts is challenging with a single contextual segment. This is because a typical epileptic seizure phase consists of pre-seizure aura, seizure onset, and post-seizure periods, with varying duration (ranging from seconds to minutes or longer in the case of status epilepticus) [12]. To this end, we first extract $2 \times N$ temporal segments $\{s_i\}_{k-N \leq i \leq k-1}$ and $\{s_i\}_{k+1 \leq i \leq k+N}$ from both sides of s_k (each contains $N = N \times \ell$ timesteps in total, with each segment consisting of ℓ timesteps), ensuring that the context contain abundant normal information for comparison. Subsequently, we apply DFT to convert these segments into the frequency domain representation $\in \mathbb{R}^d$. Finally, the frequency domain segments obtained from the left and right sides are referred as the contextual references of s_k , denoted as $L_c(s_k) \in \mathbb{R}^{N \times d}$ and $R_c(s_k) \in \mathbb{R}^{N \times d}$ (blue box) respectively.

Channel-level Reference. Different physiological brain regions have variations in neural activities [37], leading to distribution shifts between channels, even within the same subject. Although

contextual reference can reduce local bias, it fails to address global bias. Moreover, for prolonged epileptic seizures, solely considering the adjacent contextual reference segments may not provide sufficient normal information as background for comparison. To address these issues, we introduce *channel-level* reference as representative features of channels. The aim is to personalize channels, alleviate global bias, and provide comprehensive global background information of normal events.

Specifically, we adopt the K-Means [24] algorithm to identify the most representative patterns within a channel. First, we divide the entire time series of each channel into segments with length ℓ . Then, similar with contextual reference, we use DFT to obtain the frequency domain representation $\in \mathbb{R}^d$ of all segments. Next, the K-Means clustering algorithm is applied to group all frequency domain representations into K clusters. Finally, we arrange the clusters in descending order according to the number of elements they contain (denoted as C_1, C_2, \dots, C_K , where $|C_1| \geq |C_2| \geq \dots \geq |C_K|$). The arrangement indicates that the higher the index of the cluster, the lower the frequency of occurrence of the corresponding general pattern in the respective channel.

To obtain the general patterns of the channel, we use the centroid of each cluster, resulting in the final representation denoted as $\mu_k \in \mathbb{R}^d, k = 1, \dots, K$. These μ_k values are then concatenated to construct the left side channel-level reference $L_{cl}(s_k) \in \mathbb{R}^{K \times d}$. Empirically, the right side $R_{cl}(s_k)$ is formed by reversing the order of $L_{cl}(s_k)$.

4.2 Difference Matrix

In this subsection, we present the self-comparison implementation based on the constructed references and target segments through a simple yet effective approach of *fully differencing operation*. The comparison information is encoded using a *difference matrix* (DM).

We first obtain the frequency domain representation x_k^f (□) of target segment s_k via DFT, and then concatenate x_k^f with the constructed contextual (□) and channel-level (□) references to form the augmented segment \tilde{x}_k^f (as shown in Figure 3, bottom-left):

$$\tilde{x}_k^f = L_{cl}(s_k) || L_c(s_k) || x_k^f || R_c(s_k) || R_{cl}(s_k), \quad (2)$$

where $||$ is a concatenate operation and $\tilde{x}_k^f \in \mathbb{R}^{(2N+2K+1) \times d}$.

Fully Differencing Operation. To implement self-comparison of target segment and references, we introduce a fully differencing operation. Unlike traditional first-order differencing [13] that considers only adjacent points, the fully differencing operation makes a pairwise comparison across all segments in \tilde{x}_k^f (described in Equation 3 and Fig. 3, middle), which can more effectively capture the essential seizure patterns of brain activities. Moreover, since there are inherent scale differences in the vanilla difference matrix \mathbf{D}_k caused by varying magnitudes between seizure and normal iEEG signals across subjects, channels and time intervals, min-max normalization is adopted to address the scale difference issue. After these two operations, a synthetic difference matrix $\hat{\mathbf{D}}_k$ can be obtained:

$$\mathbf{D}_k[i, j] = \tilde{x}_k^f[i] - \tilde{x}_k^f[j], 1 \leq i, j \leq 2N + 2K + 1, \quad (3)$$

$$\hat{\mathbf{D}}_k = \text{Min-Max-Norm}(\mathbf{D}_k) \in \mathbb{R}^{(2N+2K+1) \times (2N+2K+1) \times d} \quad (4)$$

The generated difference matrix $\hat{\mathbf{D}}_k$ contains rich semantic information about the evolution of seizures. We provide a more detailed discussion on each component of the difference matrix $\hat{\mathbf{D}}_k$ and the corresponding semantic properties in App B.

Difference Matrix Encoder. To well capture and learn these essential differences, we adopt the CNNs as the DM encoder (Fig. 3, right), which have been proven to be powerful in learning representations from 2D matrices [35, 42]. The output of the CNNs will be concatenated as a representation $\hat{\mathbf{Z}}$ of the DM. Finally, $\hat{\mathbf{Z}}$ will be fed into a linear classifier to obtain the detection result.

5 Experiments

In this section, we conduct extensive experiments on public and clinical iEEG datasets to address three primary research questions: **RQ1.** How does the proposed *DMNet* model perform in subject-independent seizure detection compared to other methods? **RQ2.** Do the proposed contextual

Table 1: Average performance of subject-independent seizure detection tasks on clinical & public datasets. The \mathbf{v} indicates the first rank in a column and \underline{v} indicates the second. The performance with standard deviation is given in App. G.

Model \ Dataset	Clinical				MAYO				FNUSA			
	Pre.	Rec.	F1	F2	Pre.	Rec.	F1	F2	Pre.	Rec.	F1	F2
SelfReg	51.60	48.74	<u>51.24</u>	48.63	60.40	32.13	36.13	32.12	62.54	48.19	49.20	47.73
GroupDRO	47.60	44.74	45.15	46.33	48.31	35.00	27.82	28.04	53.48	71.44	60.47	66.38
MTL	20.46	52.33	28.59	39.13	46.87	22.08	15.68	16.31	60.04	52.64	53.90	52.83
CORAL	38.70	49.20	42.01	47.66	62.17	29.86	20.41	22.01	65.13	53.23	55.93	53.88
CDANN	33.43	40.41	35.72	37.58	36.79	79.55	45.49	56.60	<u>64.37</u>	54.85	54.35	53.86
SD	18.69	<u>54.40</u>	28.78	40.81	47.73	55.59	46.97	50.35	<u>56.99</u>	57.97	55.42	56.45
IB-IRM	29.19	49.75	37.91	42.64	47.57	57.17	46.86	50.71	54.22	63.26	55.96	59.47
VREx	44.80	32.45	36.33	35.34	51.21	59.95	51.19	54.85	54.74	60.15	54.64	57.12
IB-ERM	40.30	37.40	37.59	37.19	46.29	57.36	47.21	51.44	54.64	55.26	52.68	53.70
TRM	34.03	42.74	38.93	41.58	47.55	58.97	43.96	47.87	60.68	58.46	56.00	56.74
Abou-Abbas et al.	43.24	45.84	43.15	46.95	48.47	51.40	50.69	48.86	49.83	56.90	52.33	57.53
Zhao et al.	30.17	49.44	36.16	42.65	37.07	56.06	26.17	38.48	41.64	44.20	40.62	42.12
Dissanayake et al.	40.12	39.29	38.30	40.82	50.39	68.99	57.69	64.82	63.85	76.01	63.75	67.94
SICR	46.27	43.91	45.65	43.86	79.01	63.29	69.88	66.17	63.78	66.77	64.25	65.10
SEEGNet	44.89	47.70	46.25	47.11	71.82	60.50	64.87	63.15	62.23	72.35	<u>66.81</u>	68.12
PPi	<u>51.72</u>	49.70	49.78	51.12	<u>74.49</u>	70.21	<u>72.28</u>	71.02	59.53	75.42	<u>65.83</u>	71.59
DMNet	59.58	55.24	54.49	55.93	68.82	90.06	73.08	81.54	62.30	85.39	67.80	75.15
DMNet w/o L_{cl}	48.25	53.30	49.62	<u>51.20</u>	47.10	<u>89.15</u>	62.63	<u>76.43</u>	52.57	78.49	60.99	70.75
DMNet w/o L_c	51.39	47.43	47.32	47.15	58.34	76.73	64.28	71.79	49.28	73.48	58.12	65.23
DMNet w/o DM	43.58	45.79	46.72	43.42	49.67	71.54	60.63	66.57	46.98	66.89	56.21	62.18

reference, channel-level reference and difference matrix contribute to seizure detection? **RQ3.** How does the difference matrix reflect seizure activity changes during seizure evolution process?

5.1 Experimental Setup

Datasets. To evaluate the performance of our *DMNet* model, we conduct experiments on both the public benchmark dataset, which includes MAYO and FNUSA [25], and the private clinical dataset (details refer to App. C).

Evaluation Metrics. For fair comparison, we use precision, recall, F1-score and F2-score as evaluation metrics. Typically, F2-score is particularly emphasized in practical clinical studies [15, 43] since overlooking any seizure can be costly in terms of diagnosis. Therefore, in our study, the F2-score serves as the primary metric for comparison.

Settings. To conduct the experiment under the domain generalization settings, we divide the subjects in the datasets into multiple groups and assign different groups as source and target domains for model training, validation and testing. A more detailed description of experimental setup that includes *DMNet* hyperparameters and setup on clinical and public datasets, can be found in App. D.

Baselines. We compare the proposed *DMNet* with state-of-the-art subject-independent seizure detection algorithms for both iEEG-based methods and EEG-based methods. For iEEG-based methods, we compared against SICR [17], SEEG-Net [39], and PPi [41]. For EEG-based methods, we compared against Abou-Abbas et al. [2], Zhao et al. [44], and Dissanayake et al. [11]. Additionally, we compare the performance with the domain generalization (DG) algorithms in other areas like Self-Reg [18], GroupDRO [32], MTL [7], CORAL [36], CDANN [22], SD [28], IB-IRM [3], VREx [19], IB-ERM [4], TRM [40]. More details of the baselines are shown in App. A.

5.2 Overall Performance Comparison (RQ1)

The overall performance of our proposed model *DMNet* and other baselines for subject-independent seizure detection are presented in Tab. 1. From the results, we can see that our proposed *DMNet* significantly outperforms other SOTA subject-independent seizure detection algorithms and the latest domain generalization methods, with an average improvement of 9.41%, 14.81% and 4.97% in terms of F2 score on clinical and public datasets (MAYO and FNUSA) respectively. These results highlight the superior generalization ability of *DMNet*. Compared to DG baselines, our model exhibits a

substantial margin in all evaluation metrics, suggesting that general DG baselines fail to capture the diverse and evolving distribution patterns of iEEG over time. Furthermore, the EEG-based methods proposed by [2], [44] and [11] may not be adept at handling complex iEEG signals, which lead to subpar performance. Although the iEEG-based methods [41], SEEG-Net [39] and SICR [17] outperform most EEG-based methods and DG methods on F2-score, they still fall short of the performance compared to our model. This may be attributed to the fact that they do not explicitly capture the general pattern of difference between normal and seizure signals. In contrast, we utilize a difference matrix to achieve a distinguishable representation of normal and seizure.

5.3 Ablation Study of *DMNet* (RQ2)

To validate the contribution of each component of our proposed *DMNet*, we conduct ablation studies on key components (**w/o** means without and **w/** means with). The results of the ablation evaluation can be seen in Tab. 1, from which we can see full *DMNet* significantly outperforms other ablated models on F2-score. These results of the ablation experiment highlight the effectiveness of the following components:

Channel-level Reference. Removing channel-level reference causes poorer results compared to the full version. It indicates that utilizing the channel-level reference with representative characteristics of each channel for global dependencies modeling can improve the performance.

Contextual Reference. Removing contextual reference results in a noticeable performance drop. It illustrates that introducing the informative contextual reference for self-comparison to capture long-term dependencies and complete patterns of seizure improves the performances.

Difference Matrix. Removing DM leads to a significant drop in model performance, indicating the effectiveness of a fully differencing operation for the implementation of self-comparison.

5.4 Case Study (RQ3)

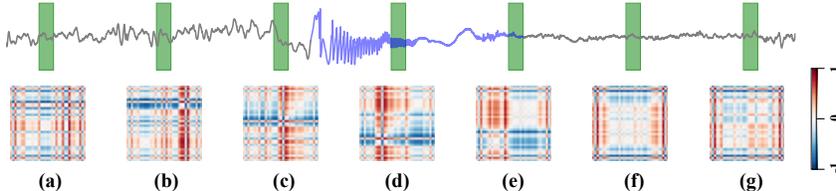


Figure 4: Case study.

To provide a more intuitive demonstration of *DMNet*, we present the visualization results of difference matrix throughout the seizure process in Fig. 4 (A full visualization can be found in App. E). The upper figure shows the raw brain signal containing a full seizure process, with the gray wave representing the normal signal and the purple wave representing the seizure. The green masked blocks indicate the segments for detection. Notably, there are clear distinctions between seizure and normal difference matrices during the seizure evolution. Segments being closer to seizure events show rougher difference matrices (e.g., segments c, d, and e), while those further away appear smoother (e.g., segments a, b, f, and g). This case clearly illustrates how the difference matrix captures seizure activity changes and demonstrates the effectiveness of *DMNet*.

5.5 Hyperparameters Analysis of *DMNet*

Number of Segments N . As described in Sec. 4.1, we utilize $2 \times N$ segments to form the contextual reference. Increasing the value of N results in a longer contextual reference, indicating the inclusion of a greater amount of contextual information. As depicted in Fig. 5(a), the evaluation scores generally increase as N increases from 8 to 12. This trend is attributed to the fact that a longer contextual reference can provide more comprehensive information throughout the entire seizure phase.

Segment Length ℓ . We investigate the effects of segment length by varying the segment length ℓ of contextual reference. The performance of *DMNet* with different segment lengths (100, 150, 200, 250, 300) is shown in Fig. 5(b). As ℓ increases, the model precision decreases.

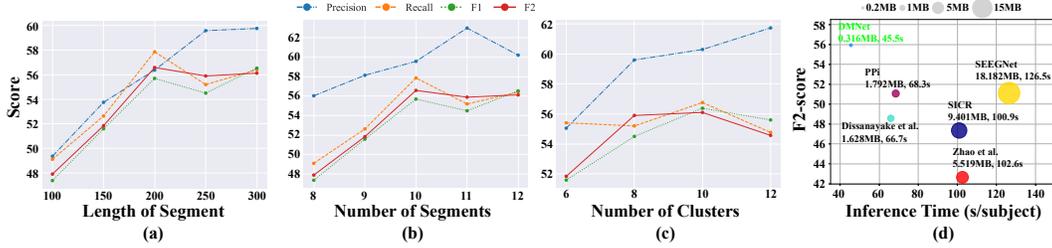


Figure 5: Hyperparameters Analysis: (a, b, c) and Computational Efficiency Analysis (d).

In contrast, recall, F1 score and F2-score initially increase when ℓ ranges from 100 to 200 but then decline. This suggests that shorter segment lengths (ℓ) fail to provide representative semantic information, preventing the model from capturing long-term dependencies in the brain signal. Conversely, excessively long segment lengths can result in coarse-grained temporal representations, leading to the loss of fine-grained patterns and details.

Number of Clusters K . We vary the number of clusters K in channel-level reference, which controls the number of generated channel representative features of a specific channel. As we can see in Fig. 5(c), the evaluation metrics (recall, F1 score, and F2-score) demonstrate an initial increase followed by a decrease trend as K varies from 6 to 12. However, precision shows an upward trend. This indicates that introducing global information can reduce false positive samples but it also affects recall. Therefore, it is necessary to consider trade-offs when selecting the value of K .

5.6 Generalization Ability Analysis

To further assess the generalization capability of DMNet on a broader range of subjects with greater heterogeneity, we evaluated the model on data of 179 previously unseen subjects from the large TUSZ EEG dataset [1]. Additional details about this evaluation study are provided in App.F.

5.7 Application Scenario

Model Efficiency. We compare the model efficiency of *DMNet* and several benchmark models in terms of parameter count and average inference time per patient file. As shown in Fig. 5 (d), the parameter count of *DMNet* is only 19.4% of the model with the smallest parameter count, but its performance is 109.4% of the best-performing model. Additionally, *DMNet* has an average inference time of 45.5 seconds per subject file, which is also the fastest among all models. Overall, *DMNet* demonstrates significant advantages in both parameter count and inference time. Therefore, *DMNet* is an ideal choice, providing a reliable and high-performance solution for real application scenario.

Online Deployment. *DMNet* has been deployed on an online system (Fig.6), which illustrates the effectiveness of our method in a real clinical application. This system serves as an auxiliary tool for expert doctors, significantly enhancing the accuracy and efficiency of the diagnostic process. The system comprises two important pages: the overview page and the detail page. The overview page (Fig.6 (top)) offers a comprehensive view of the 12-hour patient file. Each square on the page represents a 1-minute iEEG signal segment and is color coded to indicate various states, including no epileptic waves (gray), correct predictions (green), incorrect predictions (blue), and missing predictions (red) made by our model. By clicking on a square, doctors can access the detail page (Fig.6 (bottom)), where they can change the presented time period using the top toolbar. The page includes a data operation panel and seizure events displayed on the right



Figure 6: Online auxiliary diagnosis system.

side. In the center of the page, the purple section represents the true seizure annotations provided by doctors, while the yellow section showcases our model’s predictions. As depicted in the figure, the predictions of our model align well with the actual seizures.

6 Related Work

Epileptic Seizure Detection for EEG. Data-driven methods for seizure detection have gained attention in clinical medicine. [38] combine CNN and SVM, [31] leverage frequency components and CNN. However, these methods are subject-specific and not suitable for real-world application. Several efforts have been made in the study of subject-independent methods in EEG. [2] proposed and investigates a subject-independent seizure detection model that uses stable EEG-based features obtained by comparing multiple feature selection methods. [44] proposed two subject independent deep learning architectures with different learning strategies that can learn a global function utilizing data from multiple subjects. [11] proposed IBA (Information Bottleneck Attribution), a subject-independent seizure detection model that utilizes multi-view information. It employs adversarial deep learning to learn seizure-specific feature representations directly from raw EEG data. However, these methods are based on non-invasive EEG recordings.

Epileptic Seizure Detection for iEEG. iEEG, through the placement of electrodes inside or on the surface of the brain, offers higher temporal and spatial resolution, enabling more accurate capture and analysis of brain activity in specific regions, including subtle changes in electrophysiological signals. Consequently, this has spurred research into epilepsy detection based on iEEG [26]. There are several seizure detection methods being designed in subject-independent settings, which are more applicable to real-world scenarios. SEEG-Net [39] and SICR [17] proposed adversarial training to learn subject-invariant features on iEEG recordings. However, their experiments were carried out on small, manually denoised datasets with a balanced positive-negative sample ratio, resulting in a significant data bias that deviates from real clinical requirements. Moreover, [41] proposed a method that utilizes a series of intricate pre-training strategies to learn general pattern cross subjects, which is lack of efficiency.

General Domain Generalization. Our work focuses on detecting seizure events in unseen subjects, which can be framed as the domain generalization (DG) problem [45]. Existing DG studies can be categorized into two groups: *Invariant representation based method* [36, 22, 4, 18, 3, 7] and *Learning strategy based method* [20, 32, 19, 28]. Invariant representation methods aim to learn domain-invariant representations. CORAL [36] aligns covariance in feature layers, enhancing the extraction of domain-invariant features. CDANN [22] introduces adversarial training to encourage robust and domain-invariant feature learning. IB-ERM [4] minimizes empirical risk across domains to improve generalization. SelfReg [18] uses self-supervised learning and contrastive loss to capture invariant information across domains. Learning strategy methods aim to enhance generalization capability through various learning strategies. GroupDRO [32] achieves higher worst-group accuracy by coupling robust optimization models with increased regularization. VREx [19] penalizes the variance of training risks to improve domain extrapolation. Despite prior efforts, DG problems in time series, like iEEG signals, continue to be a relatively unexplored area and poses considerable challenges due to the diverse and evolving distribution patterns with time.

7 Conclusion

In this paper, we present a novel seizure detection framework called DMNet. Our model addresses the challenge of generalization across different subjects by incorporating a self-comparison mechanism to capture the subject-invariant representation. Extensive experiments conducted on clinical intracranial EEG dataset and public dataset demonstrate the effectiveness of our model in subject-independent seizure detection tasks. Moreover, our generated difference matrix effectively captures seizure activity changes during the seizure evolution process, which is valuable for clinicians to better understand the seizure event and develop more effective treatment. Furthermore, *DMNet* outperforms existing SOTAs while maintaining the high efficiency. Based on these, we deploy our method in an online system, enhancing clinical applications by assisting physicians in the diagnosis of epilepsy and in offering optimal treatment strategies. We hope that this work will shed light on the development of a more robust subject-independent seizure detection system.

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (No. 62322606, No. 62441605).

References

- [1] The temple university hospital seizure detection corpus. *Frontiers in Neuroinformatics*, page 83, 2018.
- [2] Lina Abou-Abbas, Khadidja Henni, Imene Jemal, Amar Mitiche, and Neila Mezghani. Patient-independent epileptic seizure detection by stable feature selection. *Expert Systems with Applications*, 232:120585, December 2023.
- [3] Kartik Ahuja, Ethan Caballero, Dinghui Zhang, Jean-Christophe Gagnon-Audet, Yoshua Bengio, Ioannis Mitliagkas, and Irina Rish. Invariance Principle Meets Information Bottleneck for Out-of-Distribution Generalization. In *Advances in Neural Information Processing Systems*, volume 34, pages 3438–3450, 2021.
- [4] Kartik Ahuja, Ethan Caballero, Dinghui Zhang, Jean-Christophe Gagnon-Audet, Yoshua Bengio, Ioannis Mitliagkas, and Irina Rish. Invariance principle meets information bottleneck for out-of-distribution generalization. *Advances in Neural Information Processing Systems*, 2021.
- [5] Abdu Soha Alomar, Jaes Jones, Andres Maldonado, and Jorge Gonzalez-Martinez. The stereo-electroencephalography methodology. *Neurosurgery Clinics of North America*, pages 83–95, 2016.
- [6] Mojtaba Bandarabadi, César A Teixeira, Jalil Rasekhi, and António Dourado. Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, 126(2):237–248, 2015.
- [7] Gilles Blanchard and Aniket Anand Deshmukh. Domain Generalization by Marginal Transfer Learning.
- [8] Gilles Blanchard, Aniket Anand Deshmukh, Urun Dogan, Gyemin Lee, and Clayton Scott. Domain generalization by marginal transfer learning. *Journal of Machine Learning Research*, pages 1–55, 2021.
- [9] Stephan Chabardes, Taylor J Abel, Francesco Cardinale, and Philippe Kahane. Commentary: understanding stereoelectroencephalography: what’s next? *Neurosurgery*, pages E15–E16, 2018.
- [10] Xin Chai, Qisong Wang, Yongping Zhao, Xin Liu, Ou Bai, and Yongqiang Li. Unsupervised domain adaptation techniques based on auto-encoder for non-stationary eeg-based emotion recognition. *Computers in biology and medicine*, 79:205–214, 2016.
- [11] Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Deep Learning for Patient-Independent Epileptic Seizure Prediction Using Scalp EEG Signals. *IEEE Sensors Journal*, 21(7):9377–9388, April 2021.
- [12] Arjan Hillebrand, Niall Holmes, Ndedi Sijsma, George C. O’Neill, Tim M. Tierney, Niels Liberton, Anine H. Stam, Nicole van Klink, Cornelis J. Stam, Richard Bowtell, Matthew J. Brookes, and Gareth R. Barnes. Non-invasive measurements of ictal and interictal epileptiform activity using optically pumped magnetometers. *Scientific Reports*, 2023.
- [13] S. L. Ho and M. Xie. The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 1998.
- [14] M Shamim Hossain, Syed Umar Amin, Mansour Alsulaiman, and Ghulam Muhammad. Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Transactions on Multimedia Computing, Communications, and Applications*, pages 1–17, 2019.

- [15] Yingying Hu, Ruijia Chen, Haibing Gao, Haitao Lin, Jinye Wang, Xiaowei Wang, Jingfeng Liu, and Yongyi Zeng. Explainable machine learning model for predicting spontaneous bacterial peritonitis in cirrhotic patients with ascites. *Scientific Reports*, 2021.
- [16] Eunjin Jeon, Wonjun Ko, and Heung-II Suk. Domain adaptation with source selection for motor-imagery based bci. In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, 2019.
- [17] Eunjin Jeon, Wonjun Ko, Jee Seok Yoon, and Heung-II Suk. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [18] Daehee Kim, Youngjun Yoo, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. Selfreg: Self-supervised contrastive regularization for domain generalization. In *Proceedings of the International Conference on Computer Vision*, pages 9619–9628, 2021.
- [19] David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, and Aaron Courville. Out-of-distribution generalization via risk extrapolation (rex). In *International Conference on Machine Learning*. PMLR, 2021.
- [20] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, 2018.
- [21] Guangye Li, Shize Jiang, Sivylla E Paraskevopoulou, Meng Wang, Yang Xu, Zehan Wu, Liang Chen, Dingguo Zhang, and Gerwin Schalk. Optimal referencing for stereo-electroencephalographic (seeg) recordings. *NeuroImage*, pages 327–335, 2018.
- [22] Ya Li, Xinmei Tian, Mingming Gong, Yajing Liu, Tongliang Liu, Kun Zhang, and Dacheng Tao. Deep domain generalization via conditional invariant adversarial networks. In *Proceedings of the European Conference on Computer Vision*, pages 624–639, 2018.
- [23] Hong Liu, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable Adversarial Training: A General Approach to Adapting Deep Classifiers. In *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019.
- [24] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [25] Petr Nejedly, Vaclav Kremen, Vladimir Sladky, Jan Cimbalnik, Petr Klimes, Filip Plesinger, Filip Mivalt, Vojtech Travnicek, Ivo Viscor, Martin Pail, Josef Halamek, Benjamin H. Brinkmann, Milan Brazdil, Pavel Jurak, and Gregory Worrell. Multicenter intracranial EEG dataset for classification of graphoelements and artifactual signals. *Scientific Data*, 7(1):179, June 2020.
- [26] Min Ni, Bushra Afroz, Chao Xing, Chunxiao Pan, Yanqiu Shao, Ling Cai, Brandi L Cantarel, Jimin Pei, Nick V Grishin, Stacy Hewson, et al. A pathogenic *ufsp2* variant in an autosomal recessive form of pediatric neurodevelopmental anomalies and epilepsy. *Genetics in Medicine*, pages 1–9, 2021.
- [27] Xingchao Peng, Zijun Huang, Ximeng Sun, and Kate Saenko. Domain Agnostic Learning with Disentangled Representations. In *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019.
- [28] Mohammad Pezeshki, Oumar Kaba, Yoshua Bengio, Aaron C Courville, Doina Precup, and Guillaume Lajoie. Gradient Starvation: A Learning Proclivity in Neural Networks. In *Advances in Neural Information Processing Systems*, volume 34, pages 1256–1272, 2021.
- [29] Mohammad Pezeshki, Sékou-Oumar Kaba, Yoshua Bengio, Aaron Courville, Doina Precup, and Guillaume Lajoie. Gradient starvation: A learning proclivity in neural networks. *arXiv e-prints*, pages arXiv–2011, 2020.

- [30] Timothée Proix, Viktor K. Jirsa, Fabrice Bartolomei, Maxime Guye, and Wilson Truccolo. Predicting the spatiotemporal diversity of seizure propagation and termination in human focal epilepsy. *Nature Communications*, 9(1):1088, March 2018.
- [31] Md Rashed-Al-Mahfuz, Mohammad Ali Moni, Shahadat Uddin, Salem A Alyami, Matthew A Summers, and Valsamma Eapen. A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (eeg) data. *IEEE Journal of Translational Engineering in Health and Medicine*, 2021.
- [32] Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731*, 2019.
- [33] Wojciech Samek, Frank C Meinecke, and Klaus-Robert Müller. Transferring subspaces between subjects in brain–computer interfacing. *IEEE Transactions on Biomedical Engineering*, pages 2289–2298, 2013.
- [34] Catherine A Schevon, Shennan A Weiss, Guy McKhann Jr, Robert R Goodman, Rafael Yuste, Ronald G Emerson, and Andrew J Trevelyan. Evidence of an inhibitory restraint of seizure activity in humans. *Nature communications*, 2012.
- [35] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, 2014.
- [36] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision – ECCV 2016 Workshops*. Springer, 2016.
- [37] James X. Tao, Shasha Wu, Maureen Lacy, Sandra Rose, Naoum P. Issa, Carina W. Yang, Katherine E. Dorociak, Maria Bruzzone, Jisoon Kim, Ahmad Daif, Jason Choi, Vernon L. Towle, and Peter C. Warnke. Stereotactic EEG-guided laser interstitial thermal therapy for mesial temporal lobe epilepsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 2018.
- [38] Syed Muhammad Usman, Shehzad Khalid, and Muhammad Haseeb Aslam. Epileptic seizures prediction using deep learning techniques. *IEEE Access*, pages 39998–40007, 2020.
- [39] Yiping Wang, Yanfeng Yang, Gongpeng Cao, Jinjie Guo, Penghu Wei, Tao Feng, Yang Dai, Jinguo Huang, Guixia Kang, and Guoguang Zhao. Seeg-net: An explainable and deep learning-based cross-subject pathological activity detection method for drug-resistant epilepsy. *Computers in Biology and Medicine*, page 105703, 2022.
- [40] Yilun Xu and Tommi Jaakkola. Learning representations that support robust transfer of predictors. *arXiv e-prints*, pages arXiv–2110, 2021.
- [41] Zhizhang Yuan, Daoze Zhang, Yang Yang, Junru Chen, and Yafeng Li. PPI: Pretraining brain signal model for patient-independent seizure detection. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [42] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. In *Computer Vision – ECCV 2014*. Springer International Publishing, 2014.
- [43] Kevin Zhang and Dina Demner-Fushman. Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations. *Journal of the American Medical Informatics Association : JAMIA*, 2017.
- [44] Yanna Zhao, Gaobo Zhang, Yongfeng Zhang, Tiantian Xiao, Ziwei Wang, Fangzhou Xu, and Yuanjie Zheng. Multi-view cross-subject seizure detection with information bottleneck attribution. *J. Neural Eng.*, 19(4):046011, August 2022.
- [45] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain Generalization: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

A Baselines

Firstly, we compare our model to other iEEG-based subject-independent epilepsy detection models. The Details of these baseline models are given here:

- SICR [17]: a framework that learns class-relevant and subject- invariant feature representations, which shows a promising performance in non-invasive brain-computer interface.
- SEEG-Net [39]: a model that can address the problems of sample imbalance, cross-subject domain shift, and poor interpretability and realizes high-sensitivity SEEG pathological activity detection. The source code of SEEG-Net is not released, so we implement it by ourselves to conduct the experiments.
- PPI [39]: proposed a method that utilizes a series of pre-training strategies to extract rich information from iEEG data while preserving the unique characteristics between brain signals recorded from different brain areas.

Secondly, we also compare our model to several EEG-based subject-independent epilepsy detection models. The Details of these baseline models are given here:

- [2]: they propose and investigates a patient-independent seizure detection model that uses stable EEG-based features obtained by comparing multiple feature selection methods.
- [44]: they propose two subject independent deep learning architectures with different learning strategies that can learn a global function utilizing data from multiple subjects.
- [11]: they propose a subject-independent seizure detection model, called IBA (Information Bottleneck Attribution), that utilizes multi-view information. By employing adversarial deep learning, the model learns seizure-specific feature representations directly from raw EEG data.

Moreover, we offer more details on comparisons to other latest domain generalization methods utilized in this paper:

- CDANN [22]: an end-to-end conditional invariant deep DG approach by leveraging deep neural networks for domain-invariant representation learning.
- CORAL [36]: an unsupervised domain adaptation method that aligns the second-order statistics of the source and target distributions with a linear transformation.
- GroupDRO [32]: a model coupling group DRO models with increased regularization, where DRO allows to learn models that instead minimize the worst-case training loss over a set of groups.
- MTL [8]: a representative framework for DG, which augments the original feature space with the marginal distribution of feature vectors.
- SD [29]: a regularization method aimed at decoupling feature learning dynamics, improving accuracy and robustness in cases hindered by gradient starvation.
- SelfReg [18]: a regularization method for DG based on contrastive learning, self-supervised contrastive regularization.
- TRM [40]: a robust estimation criterion that is specifically geared towards optimizing transfer to new environments.
- VREx [19]: a penalty on the variance of training risks as a simpler variant based on a form of robust optimization over a perturbation set of extrapolated domains.
- IB-ERM [4]: a DG method that improve generalization via minimizes the empirical risk over multiple domains.
- IB-IRM [4]: a DG method that improve generalization via minimizes the invariant risk over multiple domains.

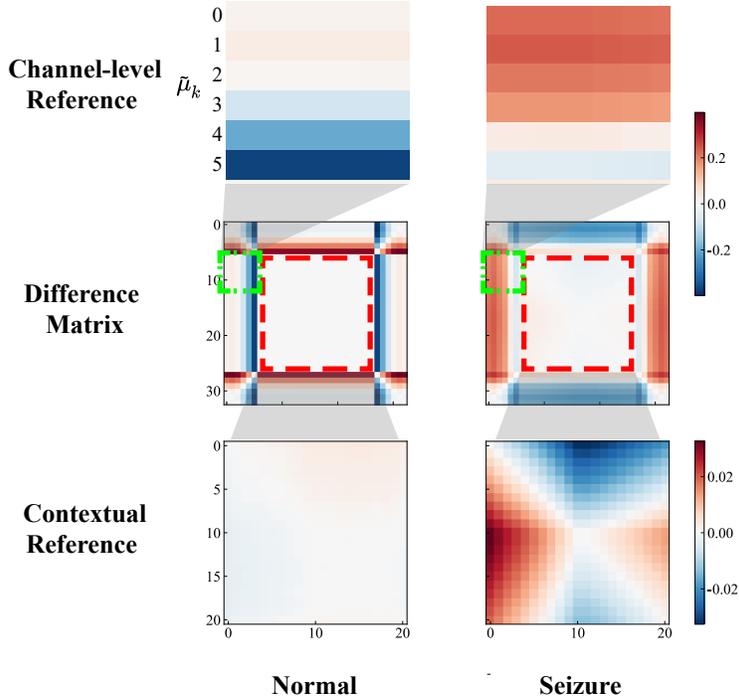


Figure 7: Visualization of difference matrix.

B Indepth Visualization of Difference Matrix

In this section, we present visualizations of the difference matrix (DM) constructed in Section 4.2, for providing an in-depth analysis of the DM properties. Initially, the difference matrix is constructed separately for normal and seizure segments of all subjects using the method mentioned in Section 4. Next, we average all difference matrices of normal and seizure segments to obtain the averaged 2-dimensional matrices. These two averaged 2D matrices are shown in the second row of Fig. 7. The green square with dashed line represents the difference between the channel-level reference and target segment x_k^f . A zoom-in visualization is provided in the first row of the Fig. 7. The left (normal) shows little difference of x_k^f with most of the channel patterns (small cluster index), but significant difference with a small portion of channel patterns (large cluster index). Conversely, the right one (seizure) exhibits the opposite pattern, validating the effectiveness of channel-level reference. Furthermore, the red square with the dashed line represents the difference within x_k^f (only containing contextual reference). The third row of the figure presents a zoom-in visualization, showing a smooth 2-dimensional matrix for normal event and a rough 2-dimensional matrix for seizure event. This demonstrates the effectiveness of introducing contextual reference.

C Datasets

To evaluate the performance of our model, we conduct extensive experiments on the following two datasets.

Clinical Dataset. The clinical dataset in our study is provided by a first-class hospital, and the intracranial EEG electrode implantation surgery and data collection are approved by the official ethics committee. For each subject, 4 to 10 invasive electrodes with 52 to 126 channels are implanted according to clinical needs to obtain brain signals. The dataset is quite massive due to the high frequency and multiple channels used to record intracranial EEG data, with more than 738 recording hours and 877.3GB total size. Data are labeled by professional neurologists at point level. Moreover, the positive sample (seizure event) ratio of a single subject in our dataset is around 0.003 on average, which is extremely imbalanced. More details can be found in Table 3.

Public Dataset. The public dataset MAYO and FNUSA [25] used in our paper, is collected from St. Anne’s University Hospital (Brno, Czech Republic). This dataset is composed of intracranial EEG data collected in an awake resting state from 38 diagnosed subjects. Specifically, the dataset comprises a total of 348,300 segments. Each segment spans a duration of 3 seconds and consists of 15,000 data points, with a sampling frequency of 5000Hz. These segments are labeled into 4 categories, including physiological activity, pathological activity (seizure event), artifacts, and power line noise. We follow the setup in [39] of omitting invalid data from one subject while retaining the data from 30 subjects for analysis. More details can be found in Table 3.

Table 2: Basic hyperparameters of DMNet.

Parameter	Clinical dataset	Public dataset
Length of Segment ℓ	250	500
Number of Segments N	10	7
Number of Clusters K	8	8
Base Filter Number C_h	64	8
Batch Size	24	24
Optimizer	Adam	Adam
Learning Rate	3×10^{-4}	3×10^{-4}
Max Epoch	20	20
Valid Metric	F1-score	F1-score

Table 3: Details information of the clinical and public dataset.

Dataset	Subject ID	Time (hours)	Sample frequency(Hz)	#Electrodes	#Channels	Positive sample ratio	#Samples	gender
Clinical	P0	72	1000	10	126	0.003	8,113,760	male
	P1	24	512	7	93	0.0002	4,176,200	female
	P2	6	512	5	69	0.004	3,698,920	male
	P3	21	1000	4	52	0.004	3,460,280	female
	P4	114	1000	10	126	0.001	9,664,920	female
	P5	36	512	5	63	0.002	3,818,240	female
	P6	24	512	7	89	0.009	6,801,240	female
	P7	36	1024	4	52	0.001	1,789,800	male
MAYO	P0	2.2	5000	-	-	0	2648	-
	P1	7.9	5000	-	-	0.1020	9536	-
	P2	2.3	5000	-	-	2.2231	2788	-
	P3	5.6	5000	-	-	0	6693	-
	P4	2.4	5000	-	-	0	2853	-
	P5	6.3	5000	-	-	0	7585	-
	P6	10.7	5000	-	-	0	12873	-
	P8	2.3	5000	-	-	1	2816	-
	P9	0.6	5000	-	-	0	740	-
	P14	3.2	5000	-	-	6.8795	3924	-
	P16	3.2	5000	-	-	0	3876	-
	P17	8.5	5000	-	-	0	10194	-
	P18	4.0	5000	-	-	0	4826	-
	P19	4.7	5000	-	-	0	5613	-
P20	2.3	5000	-	-	0	2702	-	
P21	2.9	5000	-	-	59.1724	3490	-	
P23	3.46	5000	-	-	1.96	4152	-	
FNUSA	P0	1.6	5000	-	-	1	1912	-
	P1	10.3	5000	-	-	0.1341	12358	-
	P2	6.7	5000	-	-	0.9985	8088	-
	P3	7.1	5000	-	-	0	8463	-
	P4	10.0	5000	-	-	0.1268	12038	-
	P5	2.1	5000	-	-	0.6176	2516	-
	P6	13.2	5000	-	-	0.4884	15843	-
	P7	18.9	5000	-	-	0.0834	22774	-
	P8	5.6	5000	-	-	1	6750	-
	P9	9.7	5000	-	-	0.3675	11591	-
	P10	8.6	5000	-	-	0.3953	10301	-
	P11	39.3	5000	-	-	0.1631	47270	-
P12	16.4	5000	-	-	0.2708	19635	-	

Table 4: Group information of clinical and public datasets.

Dataset	Group					
	A	B	C	D	E	F
Clinical	0,7	5,6	2,4	1,3	-	-
MAYO	0,18,21	1,9,19	2,5,16	3,4,23	6,8	14,17,20
FNUSA	0,4	1,8	2,3,11	5,6	7,9	10,12

Table 5: Detailed information of dataset setup in the experiment.

Dataset	Exp. id	Training	Validation	Test
Clinical	0	C+D	B	A
	1	B+D	C	A
	2	B+C	D	A
	3	C+D	A	B
	4	A+D	C	B
	5	A+C	D	B
	6	B+D	A	C
	7	A+D	B	C
	8	A+B	D	C
	9	B+C	A	D
	10	A+C	B	D
	11	A+B	C	D
Public	0	A+B+C+D	E	F
	1	A+B+C+F	D	E
	2	A+B+E+F	C	D
	3	A+D+E+F	B	C
	4	C+D+E+F	A	B
	5	B+C+D+E	F	A

D Experimental Setup

All experiments were run on a Linux system with 2 CPUs (AMD EPYC 7H12 64-Core Processor) and 4 GPUs (NVIDIA GeForce RTX 3090).

Setup on Clinical Dataset. To conduct the experiment under domain generalization setting, we divide these subjects in a clinical dataset into 4 groups, each group contains 2 subjects (detailed grouping information is listed in Table 4). For model training and testing, we adopt a “2-1-1 setting” for the division of training, validation, and test sets. Specifically, we assign two groups to the training set and one group to the validation set, collectively forming the source domains. Additionally, we assign another group as the target domain for testing. For a comprehensive experiment, we conduct the experiments under 12 different grouping combinations (detailed settings in Table 5). To reduce computational load, we downsample the data to 250Hz. The length of the patch ℓ is set to 1 second (250 data points). More experimental details, including model and optimization settings, are listed in Table 2.

Setup on Public Dataset. For the public dataset, we employ the grouping strategy mentioned in [39]. Specifically, we divide these subjects into 6 groups. We adopt a “4-1-1 setting” for model training, validation, and testing, respectively: randomly choose 5 groups (4 testing groups and 1 validation group) as the source domain, while the other one group as the target domain for testing. For the public dataset, we conduct the experiments under 6 different grouping combinations (detailed setups in Table 5). Moreover, because the public dataset is sampled data, there is no complete channel data, so for DMNet with public data, we remove the channel-level reference. More experimental details, including model and optimization settings are listed in Table 2.

E Full Visualization of Seizure Evolution Process

To additional illustrate the effectiveness of our proposed *DMNet*, we present the visualization results of the difference matrix throughout the entire seizure process in Fig. 8. The upper figure shows the

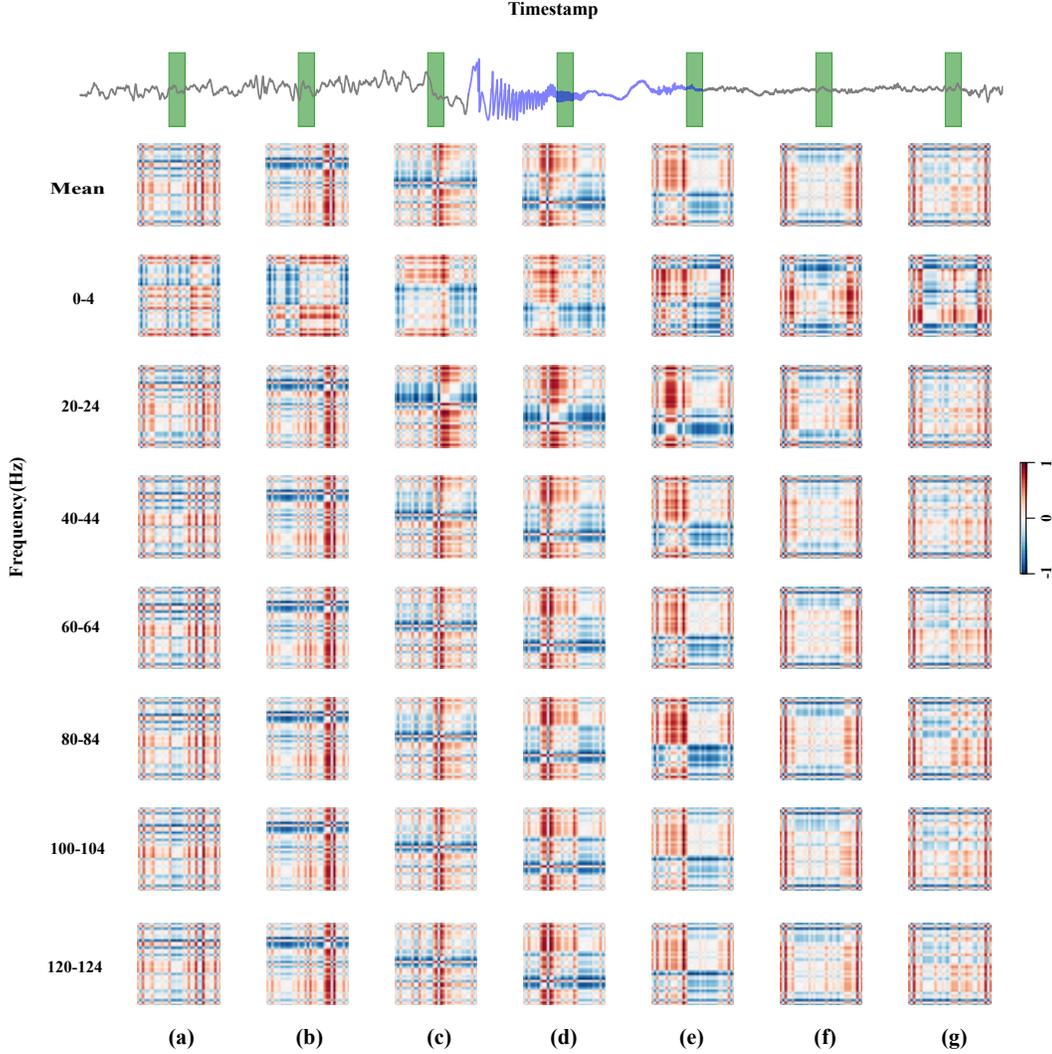


Figure 8: Illustration of how difference matrices reflect seizure activity changes during seizure evolution process. Purple line refers to seizure waves, gray line refers to normal waves.

raw iEEG signal that contains a complete seizure process, where the gray wave represents the normal signal and the purple wave represents the seizure. The green masked blocks indicate the segments for detection. Notably, distinctions between seizure and normal matrices are evident during the evolution of the seizure. For segments that are temporally closer to the seizure events, the difference matrices tend to be rougher (e.g., segments c, d, and e) and vice versa (e.g., segments a, b, f, and g). In summary, this case clearly illustrates how the difference matrix captures the seizure activity changes in different frequency bands, and indicates the effectiveness of our proposed method.

F Generalization Ability Analysis

To further evaluate the generalization capability of DMNet across a wider range of subjects with greater heterogeneity, we assessed the model using data from the extensive TUSZ EEG dataset [1] which encompasses numerous subjects. After data preprocessing, we retained data from 179 subjects, dividing them into training, validation, and testing sets in a 6:2:2 ratio, with distinct subjects in each split. Please refer to Tab. 6 for the results of the experiment. The results indicate that DMNet

Table 6: Performance of DMNet on TUSZ Dataset.

Model \ Dataset	TUSZ			
	Pre.	Rec.	F1	F2
Abou-Abbas et al.	25.03 ± 3.881	57.20 ± 3.382	36.27 ± 4.528	43.88 ± 3.181
Zhao et al.	23.74 ± 4.502	49.73 ± 3.467	30.28 ± 4.022	42.41 ± 3.285
Dissanayake et al.	28.67 ± 3.294	54.91 ± 4.492	36.70 ± 2.587	49.15 ± 5.128
SICR	32.54 ± 4.621	61.32 ± 5.062	40.71 ± 3.296	50.45 ± 4.957
SEEGNet	34.16 ± 4.381	58.20 ± 5.124	45.34 ± 4.213	52.63 ± 5.648
PPi	36.49 ± 5.682	62.72 ± 4.482	49.69 ± 5.107	56.09 ± 4.784
DMNet	48.87 ± 2.520	67.59 ± 4.834	56.56 ± 3.347	65.86 ± 2.972

Table 7: Full average performance with standard deviation of of subject-independent seizure detection tasks on clinical dataset. The \underline{v} indicates the first in a column and \bar{v} indicates the second.

Model \ Dataset	Clinical			
	Pre.	Rec.	F1	F2
SelfReg	51.60 ± 3.234	48.74 ± 4.185	<u>51.24</u> ± 3.138	48.63 ± 2.980
GroupDRO	47.60 ± 1.760	44.74 ± 3.123	45.15 ± 1.920	46.33 ± 1.990
MTL	20.46 ± 3.350	52.33 ± 4.040	28.59 ± 3.128	39.13 ± 2.963
CORAL	38.70 ± 2.455	49.20 ± 3.893	42.01 ± 2.620	47.66 ± 2.680
CDANN	33.43 ± 5.783	40.41 ± 5.050	35.72 ± 5.013	37.58 ± 4.668
SD	18.69 ± 2.475	<u>54.40</u> ± 3.823	28.78 ± 2.600	40.81 ± 2.645
IB_IRM	29.19 ± 1.948	49.75 ± 3.785	37.91 ± 2.035	42.64 ± 2.208
VREx	44.80 ± 2.213	32.45 ± 3.745	36.33 ± 2.450	35.34 ± 2.613
IB_ERM	40.30 ± 1.668	37.40 ± 3.338	37.59 ± 1.880	37.19 ± 2.125
TRM	34.03 ± 2.023	42.74 ± 4.285	38.93 ± 2.448	41.58 ± 3.003
Abou-Abbas et al.	43.24 ± 5.313	45.84 ± 4.433	43.15 ± 4.718	46.95 ± 4.550
Zhao et al.	30.17 ± 3.630	49.44 ± 5.893	36.16 ± 4.168	42.65 ± 4.943
Dissanayake et al.	40.12 ± 4.853	39.29 ± 6.493	38.30 ± 5.783	40.82 ± 6.160
SICR	46.27 ± 5.655	43.91 ± 6.253	45.65 ± 5.380	43.86 ± 5.695
SEEGNet	44.89 ± 7.073	47.70 ± 4.355	46.25 ± 6.432	47.11 ± 4.570
PPi	<u>51.72</u> ± 6.847	49.70 ± 5.712	49.78 ± 4.050	51.12 ± 4.293
DMNet	59.58 ± 4.648	55.24 ± 5.120	54.49 ± 3.915	55.93 ± 3.751
DMNet w/o L_{cl}	48.25 ± 4.110	53.30 ± 3.859	49.62 ± 3.858	<u>51.20</u> ± 3.096
DMNet w/o L_c	51.39 ± 3.820	47.43 ± 4.107	47.32 ± 4.953	47.15 ± 4.120
DMNet w/o DM	43.58 ± 4.562	45.79 ± 3.870	46.72 ± 3.205	43.42 ± 4.823

consistently outperforms existing SOTA models, demonstrating its effectiveness in seizure detection on EEG dataset with numerous subjects.

G Full Result

H Limitations

There are two main limitations to my approach. Firstly, it lacks a theoretical foundation, which may call into question the effectiveness and reliability of the method. It becomes challenging to explain and justify the underlying principles behind the approach. Secondly, although the method is efficient, it suffers from a limited number of parameters. This limitation can potentially impact the generalizability of the method. Therefore, while the method may show promising results in specific contexts, caution should be exercised when applying it to broader or unfamiliar situations.

Table 8: Full average performance with standard deviation of of subject-independent seizure detection tasks on MAYO. The \mathbf{v} indicates the first in a column and $\underline{\mathbf{v}}$ indicates the second.

Model \ Dataset	MAYO			
	Pre.	Rec.	F1	F2
SelfReg	60.40 \pm 5.788	32.13 \pm 4.320	36.13 \pm 4.423	32.12 \pm 3.988
GroupDRO	48.31 \pm 7.228	35.00 \pm 9.053	27.82 \pm 6.870	28.04 \pm 6.823
MTL	46.87 \pm 6.840	22.08 \pm 7.813	15.68 \pm 4.915	16.31 \pm 5.003
CORAL	62.17 \pm 7.843	29.86 \pm 9.835	20.41 \pm 7.335	22.01 \pm 7.430
CDANN	36.79 \pm 5.453	79.55 \pm 4.335	45.49 \pm 5.578	56.60 \pm 4.843
SD	47.73 \pm 6.518	55.59 \pm 4.735	46.97 \pm 4.825	50.35 \pm 4.485
IB_IRM	47.57 \pm 7.173	57.17 \pm 5.315	46.86 \pm 6.093	50.71 \pm 5.490
VREx	51.21 \pm 6.155	59.95 \pm 5.323	51.19 \pm 5.738	54.85 \pm 5.503
IB_ERM	46.29 \pm 6.553	57.36 \pm 4.405	47.21 \pm 5.303	51.44 \pm 4.725
TRM	47.55 \pm 7.498	58.97 \pm 4.785	43.96 \pm 4.788	47.87 \pm 3.473
Abou-Abbas et al.	48.47 \pm 4.898	51.40 \pm 5.690	50.69 \pm 4.985	48.86 \pm 5.290
Zhao et al.	37.07 \pm 4.344	56.06 \pm 4.686	26.17 \pm 4.367	38.48 \pm 4.870
Dissanayake et al.	50.39 \pm 6.033	68.99 \pm 4.323	57.69 \pm 3.838	64.82 \pm 3.173
SICR	79.01 \pm 3.980	63.29 \pm 3.058	69.88 \pm 2.208	66.17 \pm 2.623
SEEGNet	71.82 \pm 7.341	60.50 \pm 3.091	64.87 \pm 3.421	63.15 \pm 3.340
PPI	<u>74.49</u> \pm 8.550	70.21 \pm 2.725	<u>72.28</u> \pm 3.915	71.02 \pm 3.070
DMNet	68.82 \pm 7.168	90.06 \pm 1.270	73.08 \pm 3.738	81.54 \pm 3.283
DMNet w/o L_{cl}	47.10 \pm 4.055	<u>89.15</u> \pm 3.688	62.63 \pm 2.243	<u>76.43</u> \pm 5.543
DMNet w/o L_c	58.34 \pm 4.825	76.73 \pm 4.608	64.28 \pm 4.553	71.79 \pm 4.823
DMNet w/o DM	49.67 \pm 5.303	71.54 \pm 4.553	60.63 \pm 4.825	66.57 \pm 2.808

Table 9: Average performance with standard deviation of of subject-independent seizure detection tasks on FNUSA. The \mathbf{v} indicates the first in a column and $\underline{\mathbf{v}}$ indicates the second.

Model \ Dataset	FNUSA			
	Pre.	Rec.	F1	F2
SelfReg	62.54 \pm 4.373	48.19 \pm 6.663	49.20 \pm 4.413	47.73 \pm 5.630
GroupDRO	53.48 \pm 4.318	71.44 \pm 3.943	60.47 \pm 3.728	66.38 \pm 3.665
MTL	60.04 \pm 4.088	52.64 \pm 5.173	53.90 \pm 4.298	52.83 \pm 4.780
CORAL	65.13 \pm 3.615	53.23 \pm 5.395	55.93 \pm 4.283	53.88 \pm 4.930
CDANN	<u>64.37</u> \pm 5.103	54.85 \pm 5.558	54.35 \pm 4.608	53.86 \pm 4.520
SD	56.99 \pm 4.583	57.97 \pm 4.970	55.42 \pm 3.688	56.45 \pm 4.240
IB_IRM	54.22 \pm 4.985	63.26 \pm 4.608	55.96 \pm 3.503	59.47 \pm 3.768
VREx	54.74 \pm 4.923	60.15 \pm 5.458	54.64 \pm 4.010	57.12 \pm 4.553
IB_ERM	54.64 \pm 4.948	55.26 \pm 5.363	52.68 \pm 4.178	53.70 \pm 4.690
TRM	60.68 \pm 4.240	58.46 \pm 6.150	56.00 \pm 4.325	56.74 \pm 5.213
Abou-Abbas et al.	49.83 \pm 4.215	56.90 \pm 4.368	52.33 \pm 3.923	57.53 \pm 4.123
Zhao et al.	41.64 \pm 3.050	44.20 \pm 2.168	40.62 \pm 0.798	42.12 \pm 1.123
Dissanayake et al.	63.85 \pm 6.980	76.01 \pm 3.873	63.75 \pm 4.368	67.94 \pm 2.598
SICR	63.78 \pm 3.238	66.77 \pm 4.078	64.25 \pm 3.363	65.10 \pm 3.748
SEEGNet	62.23 \pm 4.070	72.35 \pm 2.620	<u>66.81</u> \pm 4.337	68.92 \pm 3.832
PPI	59.53 \pm 4.823	75.42 \pm 3.021	65.83 \pm 4.981	<u>71.59</u> \pm 4.515
DMNet	62.30 \pm 5.475	85.39 \pm 2.623	67.80 \pm 3.795	75.15 \pm 2.460
DMNet w/o L_{cl}	52.57 \pm 6.623	<u>78.49</u> \pm 5.323	60.99 \pm 3.995	70.75 \pm 3.908
DMNet w/o L_c	49.28 \pm 6.155	73.48 \pm 4.418	58.12 \pm 3.723	65.23 \pm 4.686
DMNet w/o DM	46.98 \pm 6.225	66.89 \pm 4.206	56.21 \pm 4.195	62.18 \pm 4.846

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: In this paper, abstract and introduction accurately reflect the paper's contributions and scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The limitations of the work are discussed in Appendix H.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not include theoretical results

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The private dataset so one may cannot reproduce the experiments on the private dataset. However, we also use another two public datasets, and the code of our work is fully provided too, which could help you understand our work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The private dataset so one may cannot reproduce the experiments on the private dataset. However, we also use another two public datasets, and the code of our work is fully provided too, which could help you understand our work.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We give a detailed description of our experiment setting in Appendix D. We also do a hyperparameter analysis in the Section 5.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The standard derivation is shown in the Appendix G.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide information about Experiments Compute Resources such as CPU, GPU, etc. in the Appendix D.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: All authors reviewed and conducted the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: There is no social impact of the work performed.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [No]

Justification: This paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All the existing assets are properly referenced.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: Our code is provided as a supplement.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [No]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.