

## Reviewer 3s8L

Weaknesses:

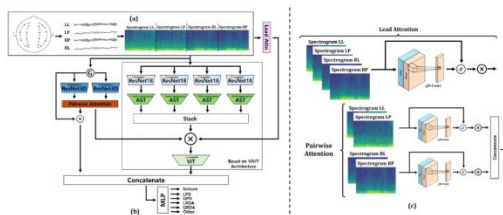
- Fig. 1 is difficult to understand. Due to multiple reasons:
  1. Several symbols are used but not introduced, e.g. G, S
  2. Fire emoji usually means that certain layers are trainable. Do you have frozen parameters? What does the emoji mean in your Figure?
  3. ImageNet, Audio Spectrogram, Kinetics400v1 indicates that the layers were pre-trained? If so, I would probably leave it away as this is too detailed.
  4. In the left (a, b), you use boxes to highlight specific parts, but in the right (c, d), you use curly braces. I suggest standardizing it.
- I do not understand the lead and pairwise attention parts. Are the inputs condensed to just 4 values?
- What is the time frame of each sample?

Feedback:

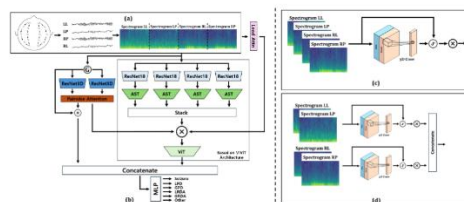
- Shorten the running title

!!!!

### 1. Figure 1 Editing



**Fig. 1.** Architecture of ESCAViT. (a) EEG preprocessing: 20 raw leads compressed into four key leads (LL, LP, RL, RP) and converted to Mel-Spectrograms. (b) ESCAViT structure: integration of ViViT-based inter-channel modeling and hemisphere-specific 3D ResNet pathways. (c) Lead-Attention and Pairwise Attention mechanisms for modeling inter-lead relationships and enhancing spatial-spectral feature extraction.



**Fig. 1.** Architecture of ESCAViT. (a) EEG preprocessing: 20 raw leads compressed into four key leads (LL, LP, RL, RP) and converted to Mel-Spectrograms. (b) ESCAViT structure: integration of ViViT-based inter-channel modeling and hemisphere-specific 3D ResNet pathways. (c) Lead-Attention and (d) Pairwise Attention mechanisms for modeling inter-lead relationships and enhancing spatial-spectral feature extraction.

<Left: before editing; Right: after editing>

### 2. I do not understand the lead and pairwise attention parts. Are the inputs condensed to just 4 values?

#### 3.2 Feature Extraction with Lead Attention

Seizure EEG signals are characterized by the sudden appearance of distinct spectral patterns at specific time points. To effectively capture these temporal and spectral fluctuations, we propose a Lead Attention mechanism based on spatial attention [23]. Unlike CBAM, which uses 2D spatial pooling, Lead Attention explicitly models inter-lead dependencies and EEG-specific time-frequency variations. It also preserves temporal information via 3D convolutions.

#### 3.2 Feature Extraction with Lead Attention

Seizure EEG signals are characterized by the sudden appearance of distinct spectral patterns at specific time points. To effectively capture these temporal and spectral fluctuations, we propose a Lead Attention mechanism based on spatial attention [23]. Unlike CBAM, which employs 2D spatial pooling, Lead Attention explicitly models inter-lead dependencies and EEG-specific time-frequency variations while preserving temporal information through 3D convolutions.

Lead Attention dynamically learns the importance of four leads at each time frame. By extracting mean and maximum values from the time-frequency representations of each lead and generating attention weights through 3D convolutions, the mechanism can selectively focus on specific leads exhibiting seizure activity. Pairwise Attention groups left hemisphere leads (LL, LP) and right hemisphere leads (RL, RP) to explicitly model inter-hemispheric symmetry.

<Left: before editing; Right: after editing>

### 3. What is the time frame of each sample?

#### 3.1 Overview

The preprocessing pipeline of ESCAViT consists of two stages, as illustrated in Fig. 1(a). First, the Banana Montage technique [16] reduces the original 20 EEG leads into four key leads (LL, LP, RL, RP), simplifying complexity while maintaining spatial relationships. Each lead is subsequently converted into a Mel-Spectrogram [25] to facilitate temporal-frequency analysis. In the lead notation, the first letter denotes the Left/Right hemisphere, while the second letter indicates the Lateral/Parasagittal position.

#### 3.1 Overview

The preprocessing pipeline of ESCAViT comprises two stages, as illustrated in Fig. 1(a). Initially, the Banana Montage technique [16] reduces the original 20 EEG leads to four key leads (LL, LP, RL, RP), thereby reducing computational complexity while preserving spatial relationships. Subsequently, each lead is transformed into a Mel-Spectrogram [25] with a temporal axis of 256 seconds to facilitate time-frequency analysis. In the lead notation, the first letter denotes the Left/Right hemisphere, while the second letter indicates the Lateral/Parasagittal position.

<Above : before editing; Below: after editing>

### 3. Shorten the running title

ESCAViT: EEG Symmetry-Aware Training with  
Domain-Adaptive Contrastive Learning for IIIC  
Pattern Recognition

ESCAViT: Symmetry-Aware EEG Classification

<Above : before editing; Below: after editing>

4. I am not sure if this is a common concept in EEG, but I am missing some introduction into why lead symmetry is so important. Also, the occlusion sensitivity experiment falls a bit short.

1 Introduction

Electroencephalogram (EEG) monitoring plays a vital role in detecting and managing neurological injuries in intensive care units (ICUs) [17]. Among various EEG patterns, IIC patterns are frequently observed in critically ill patients. These patterns, which include Seizure, Lateralized Periodic Discharges (LPD), Generalized Periodic Discharges (GPD), Lateralized Rhythmic Delta Activity (LRDA), and Generalized Rhythmic Delta Activity (GRDA), provide crucial diagnostic insights into subclinical seizures and seizure-like electrical events, aiding early neurological injury detection [7].

However, IIC pattern classification remains challenging due to two key factors:

1. Inter-lead relationships & spatial dependencies—Existing methods fail to effectively model symmetrical correlations between EEG leads, which are crucial for seizure characterization [4, 5].

1 Introduction

Electroencephalogram (EEG) monitoring plays a vital role in detecting and managing neurological injuries in intensive care units (ICUs) [17]. Among various EEG patterns, IIC patterns are frequently observed in critically ill patients. These patterns, which include Seizure, Lateralized Periodic Discharges (LPD), Generalized Periodic Discharges (GPD), Lateralized Rhythmic Delta Activity (LRDA), and Generalized Rhythmic Delta Activity (GRDA), provide crucial diagnostic insights into subclinical seizures and seizure-like electrical events, aiding early neurological injury detection [7].

However, IIC pattern classification remains challenging due to two key factors:

1. Inter-lead relationships & spatial dependencies—In IIC classification, Lateralized patterns are confined to one hemisphere, while Generalized patterns manifest symmetrically across both hemispheres. This hemispheric symmetry is crucial for distinguishing seizures from non-ictal activity, yet existing methods fail to effectively model these inter-lead correlations [4, 5].

<Left: before editing; Right: after editing>

Reviewer aviv

Weaknesses:

1. The Method section emphasizes the strengths of ESCAViT but lacks rigorous explanation of how the proposed components concretely address the stated problems. Rather than clearly linking AES-Mix and LIGCL to their corresponding challenges (e.g., class imbalance, data ambiguity), the paper presents an all-encompassing claim that ESCAViT addresses class imbalance, OOD, and data ambiguity. These should be explained in a one-to-one manner.
2. The claim that ESCAViT handles OOD scenarios is unsubstantiated. There is no definition of what constitutes “out-of-distribution” in the context of EEG in this study. In medical imaging, OOD typically refers to differences in data distribution arising from scanner variation or shifts in patient demographics—none of which are discussed or tested in this work.

/////

1. The Method section emphasizes the strengths of ESCAViT but lacks rigorous explanation of how the proposed components concretely address the stated problems. Rather than clearly linking AES-Mix and LIGCL to their corresponding challenges (e.g., class imbalance, data ambiguity), the paper presents an all-encompassing claim that ESCAViT addresses class imbalance, OOD, and data ambiguity. These should be explained in a one-to-one manner.

3.4 Domain Robust Technique

ESCAViT integrates two domain-adaptive learning strategies, AES-Mix and LIGCL, to address data ambiguity, out-of-distribution (OOD) issues, and class imbalance in EEG classification. AES-Mix addresses the failure to learn unique features when applying Mixup to minority classes [26] by first applying random left-right flipping and oversampling of LPD, GRDA, and LRDA classes. RDA is a rare EEG pattern in critically ill patients. To preserve minority class characteristics such as LRDA and GRDA, mixing operations are selectively applied to specific frequency bands while standard Mixup is used for remaining classes [27].

LIGCL enhances inter-class feature separation through Mixup ratio-based weighting and establishes clear class boundaries via cosine similarity normalization. The complementary strengths of both techniques help mitigate each other’s limitations when integrated. LIGCL makes the model overly confident about cluster boundaries for ambiguous data, degrading minority class accuracy. AES-Mix dilutes features, reducing majority class performance, but these limitations are mutually compensated when both methods are integrated. Consequently, ESCAViT achieves superior performance in IIC pattern classification of imbalanced and ambiguous EEG signals.

3.4 Domain Robust Technique

ESCAViT integrates two domain-adaptive learning strategies, AES-Mix and LIGCL, to address data ambiguity, HOC issues, and class imbalance in EEG classification. Each technique targets specific challenges through complementary mechanisms.

AES-Mix addresses class imbalance by selectively augmenting minority classes (LPD, GRDA, LRDA) to resolve feature learning failures in underrepresented patterns [17]. Since RDA exhibit diagnostic features in 1-4Hz band, mixing is restricted to this range to preserve critical characteristics [18].

LIGCL targets data ambiguity from low inter-rater agreement through adaptive contrastive learning. It uses mixup ratio  $\lambda$  as weights—higher for original-like samples to maintain boundaries, lower for mixed samples to control ambiguity.

Their synergistic integration overcomes individual limitations: AES-Mix alone dilutes majority class features while LIGCL alone over-sharpens minority class boundaries. Combined, they enable robust performance on imbalanced and ambiguous IIC patterns.

<Left : before editing; Right: after editing>

2. The claim that ESCAViT handles OOD scenarios is unsubstantiated. There is no definition of what constitutes “out-of-distribution” in the context of EEG in this study. In medical imaging, OOD typically refers to differences in data distribution arising from scanner variation or shifts in patient demographics—none of which are discussed or tested in this work.

out-of-distribution (OOD) → heterogeneous Other class (HOC)

---

## Reviewer 5Xq9

### Weaknesses:

- Unclear presentation of class imbalance: The abstract and introduction mention class imbalance, but details on its extent and impact are missing upfront. Key statistics only appear later in Section 4.1, weakening the early motivation.
- Unclear architectural specifications and contributions: The paper provides little justification for architectural choices such as the use of ViViT, its depth, patch size, or input dimensions. Beyond the ablation of AES-Mix and LIGCL, the impact of other components is not clearly isolated or systematically assessed.

""""

1. Unclear presentation of class imbalance: The abstract and introduction mention class imbalance, but details on its extent and impact are missing upfront. Key statistics only appear later in Section 4.1, weakening the early motivation.

2      Anonymized Author et al.

2. Data ambiguity & class imbalance—Expert disagreement and class imbalance introduce significant classification bias [2,13].

2      Anonymized Author et al.

2. Data ambiguity & class imbalance—Expert disagreement and severe class imbalance (Other: 7,205 samples vs. LRDA: 936 samples) introduce significant classification bias [2,13].

<Above : before editing; Below: after editing>

2. (2) Several important implementation details are either omitted or unclear - specify architectural parameters - examples: ViViT layers, patch size, input resolution, positional encoding type, etc.



As illustrated in Fig. 1(c), Lead Attention extracts mean and maximum values along the frequency and time axes to generate attention weights using a 3D convolutional network. The extracted Lead Attention weights refine Audio Spectrogram Transformer (AST)-based spectrogram representations, enhancing the model’s ability to detect localized seizure patterns. Unlike standard AST models, ESCAViT incorporates Overlapping Convolutional Projection to improve local feature extraction, thereby overcoming the limitations of ViT-based models in seizure pattern modeling [8, 24].

For lead-wise feature extraction, the AST architecture employs DeiT-Base (12 layers, 768 dimensions, 12 attention heads) applied to each lead, dividing mel-spectrograms into  $16 \times 16$  patches. Unlike standard AST models, ESCAViT incorporates Overlapping Convolutional Projection to overcome the limitations of ViT-based models in capturing fine-grained seizure morphology [8, 24]. As illustrated in Fig. 1(c), Lead Attention extracts mean and maximum values along the frequency and time axes and generates attention weights through a 3D convolutional network. These weights refine the AST-based representations to enhance localized seizure pattern detection.

<Above : before editing; Below: after editing>

### 3.3 Feature Integration

Extracted features from each lead are integrated using a Global Feature Transformer (Fig. 1(c)), which is a ViT-Base model pretrained on ImageNet. The integration leverages Convolutional Projection [24] for enhanced local feature encoding and overlapping patch embeddings to maintain critical long-range dependencies. Position embedding helps preserve lead-specific information, which is essential for analyzing the temporal progression of seizure patterns.

Pairwise Attention (Fig. 1(c)) is employed to model hemispheric relationships, improving seizure pattern detection by distinguishing left-right asymmetries. This ensures that the ESCAViT framework effectively integrates spatial and spectral EEG features, outperforming traditional methods in capturing inter-lead dependencies.

### 3.3 Feature Integration

Extracted features from each lead are integrated using a Global Feature Transformer (Fig. 1(c)), which is a ViT-Base model pretrained on ImageNet. The Global Feature Transformer integrates four leads as  $2 \times 1$  patches and employs learnable absolute position embeddings at all stages. This integration leverages Convolutional Projection [24] for enhanced local feature encoding and overlapping patch embeddings to maintain critical long-range dependencies.

Pairwise Attention (Fig. 1(c)) models hemispheric relationships by distinguishing left-right asymmetries, thereby improving seizure pattern detection. Through these mechanisms, ESCAViT effectively integrates spatial and spectral EEG features, outperforming traditional methods in capturing inter-lead dependencies.

<Above : before editing; Below: after editing>