

A Brain Wave Encodes a Thousand Tokens: Modeling Inter-Cortical Neural Interactions for Effective EEG-based Emotion Recognition

Nilay Kumar[†], Priyansh Bhandari, G. Maragatham

Department of Computational Intelligence,
SRM Institute of Science and Technology, KTR
{n19459, pr6479, maragatg}@srmist.edu.in

Abstract—Human emotions are difficult to convey through words and are often abstracted in the process; however, electroencephalogram (EEG) signals can offer a more direct lens into emotional brain activity. Recent studies show that deep learning models can process these signals to perform emotion recognition with high accuracy. However, many existing approaches overlook the dynamic interplay between distinct brain regions, which can be crucial to understanding how emotions unfold and evolve over time, potentially aiding in more accurate emotion recognition. To address this, we propose RBTransformer, a Transformer-based neural network architecture that models inter-cortical neural dynamics of the brain in latent space to better capture structured neural interactions for effective EEG-based emotion recognition. First, the EEG signals are converted into Band Differential Entropy (BDE) tokens, which are then passed through Electrode Identity embeddings to retain spatial provenance. These tokens are processed through successive inter-cortical multi-head attention blocks that construct an electrode $\times$ electrode attention matrix, allowing the model to learn the inter-cortical neural dependencies. The resulting features are then passed through a classification head to obtain the final prediction. We conducted extensive experiments, specifically under subject-dependent settings, on the SEED, DEAP, and DREAMER datasets, over all three dimensions, Valence, Arousal, and Dominance (for DEAP and DREAMER), under both binary and multi-class classification settings. The results demonstrate that the proposed RBTransformer outperforms all previous state-of-the-art methods across all three datasets, over all three dimensions under both classification settings. The source code is available at: <https://github.com/nnilayy/RBTransformer>.

Index Terms—Electroencephalography (EEG), Emotion Recognition, Transformers, Inter-Cortical Attention Mechanism, Brain-Computer Interface (BCI), Affective Computing

I. INTRODUCTION

Emotions are complex psychophysiological responses to internal or external stimuli and are deeply embedded in human cognition, shaping how individuals feel, form mental states, and react emotionally to their surroundings [1]–[3]. These responses can be expressed through either non-physiological modalities, such as facial expressions, gestures, speech, and language [4]–[6], which are often subject to voluntary control and may mask what a person is truly experiencing, or through

physiological modalities, including heart rate, skin conductance, and especially neural activity [7], [8], which provide a more direct and involuntary window into how people respond to stimuli and how emotions evolve internally over time.

Quantifying and recognizing emotions plays a crucial role in healthcare and clinical contexts, helping in the understanding of emotional dysregulation and impaired affective expression, which are important for the diagnosis and treatment of medically diagnosed conditions, as defined by the DSM-5 [9] and ICD-11 classification systems [10], including neuropsychiatric conditions (e.g., depression, anxiety) and broader neurodevelopmental and psychotic disorders (e.g., autism spectrum disorder, schizophrenia) [11], [12]. To quantify and enable such modelling, emotions can be modelled using either discrete or dimensional models. Discrete models can classify emotions into only a fixed set of labels (e.g., happiness, sadness, anger, fear, disgust, surprise) [13]. Dimensional models are more general and can describe any emotion by expressing them along three linear scales, Valence, Arousal, and Dominance (VAD) [14], [15], which aligns more closely with how emotions are naturally experienced and support representations that generalize across individuals and contexts [16]. As a result, any emotion represented using a discrete model can be plotted within the VAD dimensional model, shown in Figures 1 and 2. Valence represents the degree of pleasantness, Arousal represents the level of alertness or activation, and Dominance represents the sense of control over emotion. To capture raw and involuntary neural activity, various brain-computer interface (BCI) techniques are used like functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and near-infrared spectroscopy (NIRS) [17], [18]. However, electroencephalography (EEG) is still the most widely used technique because of its non-invasive nature, cost-effectiveness, and high temporal resolution, making it well-suited for real-time emotion monitoring.

Traditionally, emotion recognition using electroencephalographic (EEG) signals has relied on manually extracting features through domain-specific preprocessing and feature extraction techniques, subsequently passing them to shallow machine learning models. The extracted features are generally

[†]Corresponding author: Nilay Kumar (n19459@srmist.edu.in).

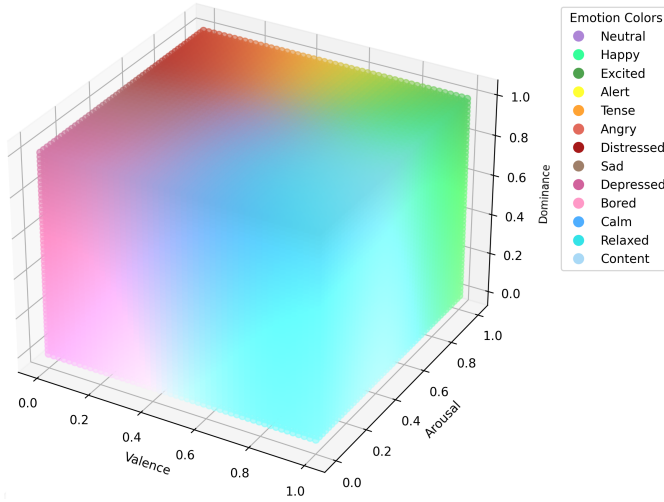


Fig. 1. Visual representation of a wide range of emotions in the three-dimensional Valence–Arousal–Dominance (VAD) space.

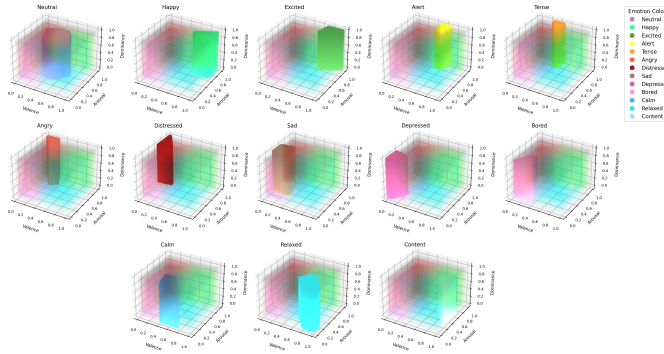


Fig. 2. Individual emotion representations within the three-dimensional Valence–Arousal–Dominance (VAD) space, showcasing the unique characteristics of each emotion.

derived from one of three primary domain representations: (i) time-domain (e.g., Hjorth parameters [19], Higher Order Crossing (HOC) [20]), (ii) frequency-domain (e.g., Differential Entropy (DE) [21], Power Spectral Density (PSD) [22]), and (iii) time–frequency domain (e.g., Discrete Wavelet Transform (DWT) [23], Continuous Wavelet Transform (CWT) [24]). Using these representations, features were manually handcrafted and fed into shallow classifiers. For example, Li et al. manually extracted 18 time-domain features (9 Hjorth parameters and 9 nonlinear features) from gamma-band EEG signals after applying Common Spatial Pattern (CSP) filtering and used a linear Support Vector Machine (SVM) for emotion recognition [25]. Patil et al. applied Empirical Mode Decomposition (EMD) to denoise EEG signals and extracted features using HOC for emotion recognition [26]. Shi et al. extracted features across five frequency bands using DE and passed them to SVM for final classification [21], and similarly, Duan et al. extracted features using DE from multichannel EEG data and passed them into a hybrid k-Nearest Neighbors (k-NN) and SVM classifier [27].

However, over the past decade, deep neural networks

(DNNs) have steadily outperformed traditional machine learning models across a range of fields, including computer vision [28], natural language processing [29], and human-computer interaction [8]. A similar progression was observed in the domain of EEG-based emotion recognition, when Zheng and Lu introduced Deep Belief Network (DBN) [30], one of the first deep learning models for affective emotion recognition. The DBN learned hierarchical representations from DE features and outperformed previous state-of-the-art machine learning approaches. In the following years, subsequent studies, such as those by Huang and Zhao [31] and Rudakov [32], made use of Convolutional Neural Networks (CNNs) [28] to capture spatial dependencies across EEG electrodes. By projecting signals into structured 2D topographies or frequency-based brain maps, CNNs enabled more effective modeling of spatial information. In parallel with the development of CNN-based approaches, the sequential and autoregressive nature of EEG signals led studies such as those by Zhang et al. [33] and Ma et al. [34] to explore Recurrent Neural Networks (RNNs) [35] and Long Short-Term Memory (LSTM) [36] architectures, as these were better suited for capturing temporal dependencies and tracking emotional state transitions over time. Naturally, to capture both the spatial and temporal richness of EEG signals, studies such as those by Shen et al. [37] and Yang et al. [38] introduced hybrid architectures combining CNNs and RNNs (or LSTMs). During the development of such models, studies such as those by Song et al. [39] and Yin et al. [40] made use of Graph Neural Networks (GNNs) [41] to more accurately ground the spatial structure of EEG data neurophysiologically, effectively modeling neural and inter-electrode dependencies as compared to the fixed-grid representations used in earlier approaches.

In recent years, studies have turned to Transformer and attention-based models [29] for their ability to capture long-range temporal dependencies and leverage a non-recurrent architecture that enables global context access, parallel computation, and more efficient training. Studies such as those by Tao et al. [42] apply spatial and temporal attention to identify informative electrodes and time steps for emotion classification. Xiao et al. [43] extend this by adding frequency-band attention to weigh spectral features. Liu et al. [44] employ dual channel attention within a 3D CNN to emphasize key spatio-temporal patterns. While these and other EEG-based emotion recognition models leverage attention mechanisms to highlight salient features within individual domains, such as spatial (electrode), spectral (frequency band), or temporal (time slice), they fundamentally treat each electrode’s signal as an independent input stream, and attention in these architectures primarily functions as a filtering mechanism, adaptively weighting channels or segments based on their relative informativeness for the task. However, none of the existing studies or model architectures take into account inter-electrode interactions, which are more neurologically grounded in how emotions emerge and are processed across cortical regions, and which can lead to more accurate and robust affective state recognition.

To explore and address this issue, we introduce RBTransformer, a Transformer-based neural network architecture that models these inter-cortical neural interactions in the latent space for EEG-based emotion recognition. RBTransformer first converts raw EEG signals into Band Differential Entropy (BDE) tokens, which are then passed through an Electrode Identity Embedding layer, which allows the model to retain awareness of each electrode’s unique identity and ordering. These representations are then passed through a stack of Inter-Cortical Multi-Head Attention Blocks, which allow each electrode to directly interact with every other electrode using an electrode $\times$ electrode attention matrix. And finally, these features are passed through a classification head to get the final class prediction. This architecture mimics the recurrent exchange of information across cortical regions, allowing RBTransformer to capture both inter-cortical dependencies and localized temporal dynamics, without relying on handcrafted features or sequential modeling at all. In brief, the primary contributions of this paper are as follows:

- 1) We introduce RBTransformer, a Transformer-based architecture that explicitly models inter-cortical neural interactions through a dedicated multi-head attention mechanism. By enabling structured communication between EEG electrodes and incorporating frequency-aware BDE tokens along with Electrode Identity Embeddings, the model captures both localized saliency and global inter-regional dependencies without relying on handcrafted features or explicit temporal modeling.
- 2) We conduct extensive experiments and demonstrate that RBTransformer achieves state-of-the-art performance across all three benchmark datasets, SEED, DEAP, and DREAMER, across all dimensions (Valence, Arousal, and Dominance for DEAP and DREAMER), under both binary and multi-class emotion classification settings.
- 3) We also present t-SNE visualizations and confusion matrices to demonstrate that the model effectively segregates emotional classes in the latent space and maintains consistent discriminative performance across classes.

The rest of the paper is organized as follows: Section II reviews related work. Section III describes data preprocessing steps and model architecture for RBTransformer. Section IV outlines datasets, evaluation metrics, and training configurations. Section V presents the results of extensive experiments and highlights the effectiveness of the proposed model. Finally, Section VI concludes the paper.

II. METHODOLOGY

In this section, we present the end-to-end pipeline used to preprocess the raw EEG data for RBTransformer and provide a detailed explanation of RBTransformer’s model architecture.

A. Dataset Preprocessing

We use the SEED [30], DEAP [8], and DREAMER [45] datasets for our experiments and apply a consistent preprocessing pipeline across all three datasets. The complete EEG

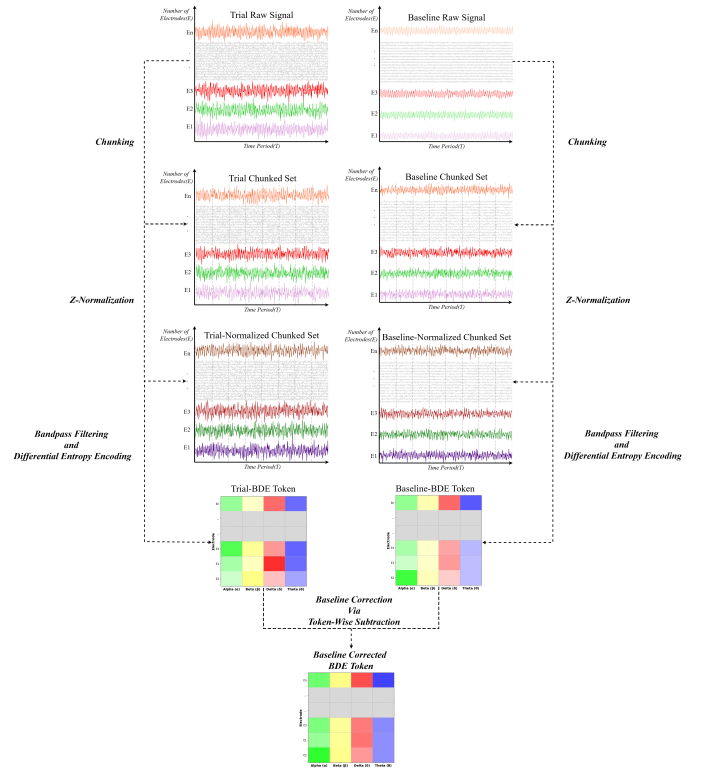


Fig. 3. Preprocessing pipeline applied across EEG datasets for RBTransformer.

preprocessing pipeline is presented in Fig. 3, and the entire workflow is explained as follows.

Each dataset X is made up of continuous multichannel EEG recordings, where each recording includes both a baseline segment X_B and a trial stimulus-induced response segment X_T , as shown in Eq. (1).

$$X = [X_B, X_T] \quad (1)$$

For each recording i , the pair (X_{Bi}, X_{Ti}) corresponds to the baseline segment $X_{Bi} \in \mathbb{R}^{C \times L_B}$ and the trial segment $X_{Ti} \in \mathbb{R}^{C \times L_T}$, where C is the number of EEG channels, and L_B, L_T are the respective time points in the baseline and trial segments.

Next, these baseline and trial signals are chunked into smaller windows. For the trial segments, a sliding window of size 512 with a stride of 117 is applied to each X_{Ti} . Each X_{Ti} is chunked into M overlapping windows of size 512, denoted as $X_{Ti} = \{X_{Ti}^{(1)}, X_{Ti}^{(2)}, \dots, X_{Ti}^{(M)}\}$, where each $X_{Ti}^{(j)} \in \mathbb{R}^{C \times 512}$. To chunk baseline segments, the average signal is computed per trial as shown in Eq. (2):

$$\bar{X}_{Bi} = \frac{1}{M} \sum_{j=1}^M X_{Bi}^{(j)} \quad (2)$$

Here, each X_{Bi} is chunked into M non-overlapping windows of size 128, denoted as $X_{Bi} = \{X_{Bi}^{(1)}, X_{Bi}^{(2)}, \dots, X_{Bi}^{(M)}\}$, where each $X_{Bi}^{(j)} \in \mathbb{R}^{C \times 128}$. This process yields a

set of averaged and chunked baseline signals $\bar{X}_B = \{\bar{X}_{B1}, \bar{X}_{B2}, \dots, \bar{X}_{Bn}\}$.

To reduce inter-subject and inter-channel variability, z-score normalization is applied independently to both the trial chunks $X_{Ti}^{(j)} \in \mathbb{R}^{C \times 512}$ and the averaged baseline segments $\bar{X}_{Bi} \in \mathbb{R}^{C \times 128}$. Let $M \in \mathbb{R}^{C \times L}$ denote either a trial chunk $X_{Ti}^{(j)}$ or a baseline segment $\bar{X}_{Bi}$, where $L = 512$ for trial chunks and $L = 128$ for baseline segments. The normalization is performed per channel $c \in \{1, 2, \dots, C\}$, shown in equations Eqs. (3)–(5):

$$\mu_c = \frac{1}{L} \sum_{t=1}^L M[c, t] \quad (3)$$

$$\sigma_c = \sqrt{\frac{1}{L} \sum_{t=1}^L (M[c, t] - \mu_c)^2} \quad (4)$$

$$\hat{M}[c, t] = \frac{M[c, t] - \mu_c}{\sigma_c}, \quad \forall t \in \{1, \dots, L\} \quad (5)$$

Here, μ_c and σ_c are the mean and standard deviation across the L time points in channel c , and $\hat{M} \in \mathbb{R}^{C \times L}$ denotes the normalized matrix with zero mean and unit variance per channel. This process yields the set of normalized trial chunks $\hat{X}_T = \{\hat{X}_{T1}^{(1)}, \hat{X}_{T1}^{(2)}, \dots, \hat{X}_{Tn}^{(M)}\}$ and the set of normalized baseline segments $\hat{X}_B = \{\hat{X}_{B1}, \hat{X}_{B2}, \dots, \hat{X}_{Bn}\}$, which serve as standardized inputs for the subsequent feature extraction stage.

Band Differential Entropy (BDE) [30] is then calculated for each normalized trial chunk and the corresponding normalized baseline segment. Bandpass filters are used to extract sub-band signals corresponding to four EEG frequency bands: Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–45 Hz). The differential entropy for each frequency band b and channel c is computed using the following formulation, shown in Eq. (6):

$$\text{BDE}_c^{(b)} = \frac{1}{2} \log_2 \left(2\pi e \cdot \text{Var} \left(\hat{S}^{(b)}[c] \right) \right) \quad (6)$$

Here, $\hat{S}^{(b)}[c]$ denotes the filtered signal in frequency band b for channel c , which may correspond to either a trial segment $\hat{X}_{Ti}^{(j,b)}[c]$ or a baseline segment $\hat{X}_{Bi}^{(b)}[c]$. $\text{Var}(\cdot)$ represents the sample variance of the sub-band signal, capturing the frequency-specific energy distribution across EEG channels. This yields trial BDE feature tokens $F_i^{(j)} \in \mathbb{R}^{C \times 4}$ and baseline BDE feature tokens $\bar{F}_i \in \mathbb{R}^{C \times 4}$. The full sets are denoted as $\mathcal{F} = \{F_1^{(1)}, F_1^{(2)}, \dots, F_n^{(M)}\}$ and $\bar{\mathcal{F}} = \{\bar{F}_1, \bar{F}_2, \dots, \bar{F}_n\}$, respectively.

Finally, to eliminate subject-specific trends, remove static background activity, and isolate emotion-related neural activity, baseline correction is applied by subtracting the baseline BDE feature token $\bar{F}_i \in \mathbb{R}^{C \times 4}$ from the corresponding trial BDE feature token $F_i^{(j)} \in \mathbb{R}^{C \times 4}$, as shown in Eq. (7):

$$\tilde{F}_i^{(j)} = F_i^{(j)} - \bar{F}_i \quad (7)$$

Here, $\tilde{F}_i^{(j)} \in \mathbb{R}^{C \times 4}$ denotes the final baseline-corrected BDE token for the j^{th} chunk of the i^{th} trial. Collectively, this yields the fully preprocessed dataset $\tilde{\mathcal{F}} = \{\tilde{F}_1^{(1)}, \tilde{F}_1^{(2)}, \dots, \tilde{F}_n^{(M)}\}$, which serves as the input for training and inference for RBTransformer.

B. Model Architecture

The RBTransformer model is an attention-based architecture tailored for EEG-based emotion recognition. It operates on the baseline-normalized Band Differential Entropy (BDE) tokens, where each token encodes the deviation of each electrode's spectral energy from its baseline across canonical EEG bands. These inputs are passed through the following sequence of processing components, depicted in Figure 4, which illustrates the complete architecture layout from BDE token input to classification output.

(1) BDE Feature Projection Layer: A batch B of preprocessed BDE tokens $\tilde{F} \in \mathbb{R}^{B \times C \times 4}$ is passed through a shared linear projection layer that maps each 4-dimensional BDE vector into a d_{model} -dimensional embedding vector for all electrodes C , as shown in Eq. (8):

$$Z = \tilde{F} W_{\text{proj}} + b_{\text{proj}} \quad (8)$$

Here, a dot product is computed between each 4-dimensional BDE token and the shared projection matrix $W_{\text{proj}} \in \mathbb{R}^{4 \times d_{\text{model}}}$, followed by the addition of a learnable bias vector $b_{\text{proj}} \in \mathbb{R}^{d_{\text{model}}}$, which results in each BDE token being transformed into a richer representational space of dimension d_{model} .

(2) Electrode Identity Embedding Layer: To support inter-cortical modeling, the attention mechanism relies on a fixed electrode-to-electrode structure, which is achieved by assigning a unique learnable embedding to each electrode to retain its spatial information. Accordingly, after BDE projection, a learnable identity embedding $E_{\text{identity}} \in \mathbb{R}^{1 \times C \times d_{\text{model}}}$ is added to the projected tensor Z , broadcast across the batch, as shown in Eq. (9):

$$Z' = Z + E_{\text{identity}} \quad (9)$$

which is then followed by a Dropout for regularization, shown in Eq. (10):

$$\tilde{Z} = \text{Dropout}(Z') \quad (10)$$

(3) Inter-Cortical Attention Mechanism: Following the addition of electrode identity embeddings, the intermediate features then enter the model's core module, the Inter-Cortical Attention Mechanism. For this, RBTransformer uses a stack of attention blocks, each made up of a Multi-Head Self-Attention module and a Feedforward Network, both of which are explained in detail in the following sections.

(I) Multi-Head Self-Attention (MHSA): The Electrode Identity Embedded tokens are then passed through the Multi-Head Self-Attention (MHSA) module, which serves as the

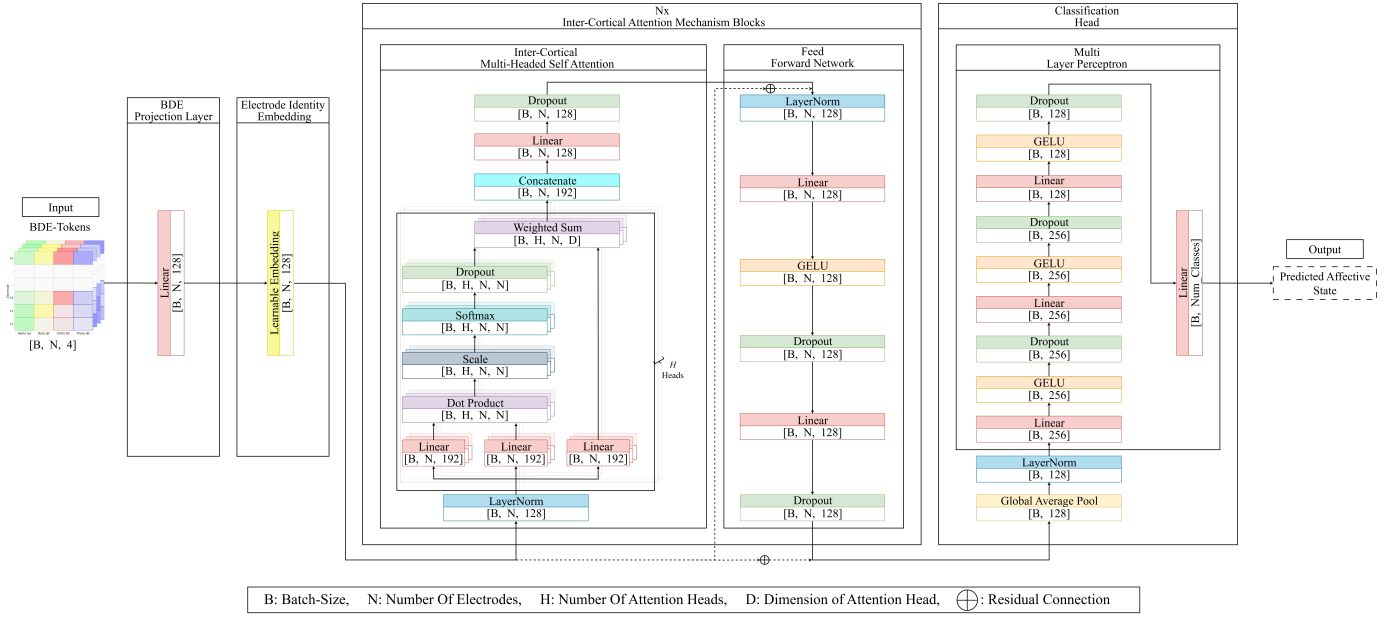


Fig. 4. Schematic architecture diagram of RBTransformer implementing inter-cortical attention.

core component for modeling the inter-cortical neural interactions in the latent space, simulating how different regions of the brain exchange information. In this, input tokens are first passed through a normalization layer, which helps to stabilize the training by ensuring that the features are scaled consistently across the embedding dimension. This is shown in Eq. (11).

$$\hat{Z} = \text{LayerNorm}(\tilde{Z}) \quad (11)$$

The normalized tokens $\tilde{Z} \in \mathbb{R}^{B \times C \times d_{\text{model}}}$ are then linearly projected into three learnable representations, queries (Q), keys (K), and values (V). This operation is performed independently across H attention heads, each of dimensionality d_{head} . Each head maintains its own set of QKV projections, $Q, K, V \in \mathbb{R}^{B \times H \times C \times d_{\text{head}}}$, allowing the model to capture multi-perspective inter-regional dependencies from the same neural signal simultaneously. This linear projection is computed via a dot product between the normalized input tokens and a learnable weight matrix $W_{qkv} \in \mathbb{R}^{d_{\text{model}} \times 3Hd_{\text{head}}}$, as shown in Eq. (12).

$$[Q, K, V] = \hat{Z} \cdot W_{qkv} \quad (12)$$

After this comes the inter-cortical attention mechanism, which computes an attention matrix of shape (electrodes $\times$ electrodes). This attention matrix $\text{Attention} \in \mathbb{R}^{B \times H \times C \times C}$ enables each electrode to attend to every other electrode, effectively capturing inter-regional neural interactions. It is calculated by taking the dot product between the Queries (Q) and Keys (K) matrices, scaled by the square root of the head dimension $\sqrt{d_{\text{head}}}$, as shown in Eq. (13).

$$\text{Attention} = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_{\text{head}}}} \right) \quad (13)$$

These attention weights are applied to the values (V) to obtain the head-specific outputs. The outputs from all heads are concatenated and projected back to the original embedding dimension using a linear transformation $W_{\text{out}} \in \mathbb{R}^{Hd_{\text{head}} \times d_{\text{model}}}$, as shown in Eq. (14):

$$Z_{\text{attn}} = \text{Concat}(\text{Attention} \cdot V) \cdot W_{\text{out}} \quad (14)$$

Finally, a residual connection is applied, along with dropout regularization, to produce the final output of the MHSA module $Z_{\text{MHSA}} \in \mathbb{R}^{B \times C \times d_{\text{model}}}$, as given in Eq. (15):

$$Z_{\text{MHSA}} = \tilde{Z} + \text{Dropout}(Z_{\text{attn}}) \quad (15)$$

(II) Feedforward Network (FFN): The output of the MHSA block Z_{MHSA} , is then passed through a Feedforward Network (FFN), which enhances the model's ability to capture complex local patterns and non-linear relationships within each signal. Specifically, the FFN applies a stack of linear transformations and dropout layers to project the input representation and promote regularization, while a GELU activation introduces non-linearity to better model complex local dependencies. This sequence refines the representation at each electrode. Finally, a residual connection adds the FFN output back to the original input Z_{MHSA} , helping preserve the original information flow and mitigate vanishing gradients during training, as shown in Eqs. (16)–(17).

$$Z_{\text{FFN}} = \text{Dropout}(W_{\text{out}} \cdot \text{Dropout}(\text{GELU}(W_{\text{in}} \cdot Z_{\text{MHSA}} + b_{\text{in}})) + b_{\text{out}}) \quad (16)$$

$$Z_{\text{AttnBlock}} = Z_{\text{MHSA}} + Z_{\text{FFN}} \quad (17)$$

Here, $Z_{\text{AttnBlock}} \in \mathbb{R}^{B \times C \times \text{embed_dim}}$ represents the output of one attention block. $W_{\text{in}} \in \mathbb{R}^{\text{embed_dim} \times \text{hidden_dim}}$ and

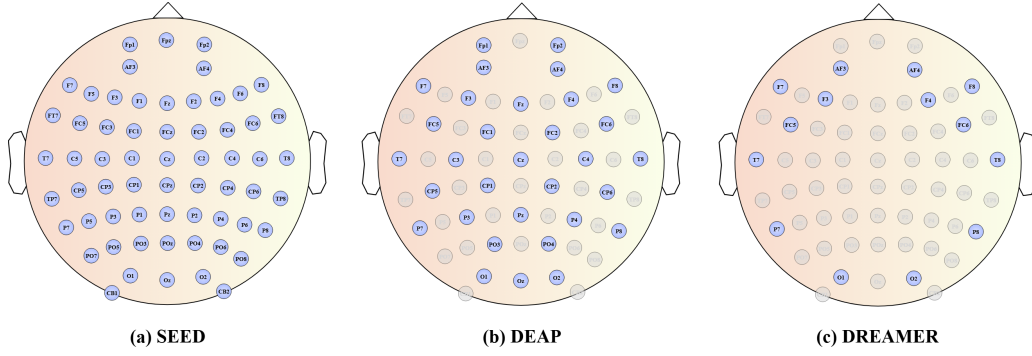


Fig. 5. Electrode layouts for (a) SEED [30] with 62 electrodes, (b) DEAP [8] with 32 electrodes, and (c) DREAMER [45] with 14 electrodes. Blue circles indicate active electrodes used in each dataset; Grey circles denote unused electrodes.

$W_{\text{out}} \in \mathbb{R}^{\text{hidden_dim} \times \text{embed_dim}}$ are learnable weight matrices, while $b_{\text{in}} \in \mathbb{R}^{\text{hidden_dim}}$ and $b_{\text{out}} \in \mathbb{R}^{\text{embed_dim}}$ are the corresponding learnable bias vectors.

(4) Classification Head: Finally, after the input is processed through multiple attention blocks, it enters the classification head, which converts the contextualized electrode representations into emotion predictions. The output tensor $Z_{\text{AttnBlockFinal}}$ is first aggregated using Global Average Pooling (GAP) across the electrode dimension. The resulting pooled vector, $Z_{\text{GAP}} \in \mathbb{R}^{B \times \text{embed_dim}}$, is then normalized via LayerNorm (LN) to obtain $Z_{\text{LN}} \in \mathbb{R}^{B \times \text{embed_dim}}$, which is subsequently passed through a linear layer to produce the final class logits, $y_{\text{logits}} \in \mathbb{R}^{B \times \text{num_classes}}$, as shown in Eqs. (18)–(20):

$$Z_{\text{GAP}} = \frac{1}{C} \sum_{i=1}^C Z_{\text{AttnBlock}}[:, i, :] \quad (18)$$

$$Z_{\text{LN}} = \text{LayerNorm}(Z_{\text{GAP}}) \quad (19)$$

$$y_{\text{logits}} = W_{\text{logits}} \cdot Z_{\text{LN}} + b_{\text{logits}} \quad (20)$$

Here, $W_{\text{logits}} \in \mathbb{R}^{\text{embed_dim} \times \text{num_classes}}$ and $b_{\text{logits}} \in \mathbb{R}^{\text{num_classes}}$ are the learnable weights and bias matrices of the final classification head.

III. EXPERIMENTAL SETUP

In this section, we describe the datasets used for training and benchmarking, the metrics used for performance evaluation, and the training configuration for RBTransformer.

A. Datasets

We perform experiments on three standardized EEG emotion recognition benchmarks, SEED [30], DREAMER [45], and DEAP [8]. The details of each dataset are explained as follows.

(1) SEED: The SEED dataset [30] contains EEG recordings from 15 subjects (7 male, 8 female) with an average age of 23.3 years. Each participant completed three sessions spaced at least one week apart to assess intra-subject stability. During each session, they viewed 15 emotionally evocative movie clips, five each for positive, neutral, and negative emotions,

and labeled their emotional responses accordingly. EEG signals were recorded using a 62-channel setup following the international 10–20 system, originally sampled at 1000 Hz and later downsampled to 128 Hz. Band Differential Entropy (BDE) features were extracted across five canonical frequency bands: delta, theta, alpha, beta, and gamma. A detailed electrode layout used in SEED is shown in Figure 5.(a).

(2) DEAP: The DEAP dataset [8] comprises EEG and peripheral physiological recordings from 32 participants (17 male, 15 female) aged between 19 and 37 years. Each participant watched 40 one-minute music video clips and rated their emotional responses along four dimensions: Valence, Arousal, Dominance, and liking. EEG signals were recorded using a 32-channel BioSemi ActiveTwo system and downsampled to 128 Hz after preprocessing. The electrode layout used in DEAP is shown in Figure 5.(b).

(3) DREAMER: The DREAMER dataset [45] comprises EEG and ECG recordings from 23 participants (14 male, 9 female) aged between 22 and 33 years. Participants watched 18 audio-visual movie clips and provided self-assessments of their emotional states using continuous ratings for Valence, Arousal, and Dominance. EEG signals were recorded using a 14-channel Emotiv EPOC device at a sampling rate of 128 Hz. Band Differential Entropy (BDE) features were extracted from overlapping time windows to capture temporal dynamics. The electrode layout used in DREAMER is shown in Figure 5.(c).

B. Metrics

To evaluate the performance of RBTransformer, we report the following classification metrics: Accuracy, Precision, Recall, and F1-score [46].

(1) Accuracy: Accuracy is the proportion of correct predictions across all samples, denoted in Eq. 21:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

Here, TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectively.

(2) Precision: Precision [46] is the proportion of true positives out of all predicted positives, denoted in Eq. 22:

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

(3) **Recall:** Recall [46], also known as Sensitivity or True Positive Rate, is the proportion of actual positives correctly identified by the model, denoted in Eq. 23:

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

(4) **F1-score:** The F1-score [46] is the harmonic mean of Precision and Recall, balancing both in a single metric:

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (24)$$

C. Training Configurations

RBTransformer was trained using the PyTorch framework [47] on an NVIDIA P100 GPU. We used the AdamW optimizer [48] for a total of 300 epochs with an L2 regularization (weight decay) of 1×10^{-3} . A cosine annealing learning rate scheduler [49] dynamically adjusted the learning rate from 1×10^{-3} to 1×10^{-6} throughout the training. The model was trained using Cross-Entropy loss with label-smoothing [50] set to 0.12. Batch size was set to 256 for the first 150 epochs and reduced to 64 for the remaining epochs. To handle class imbalance, SMOTE [51] was applied only to the training folds. A 5 fold cross-validation was carried out to evaluate the performance across all experiments which were tracked using Weights & Biases (WandB) [52].

IV. RESULTS AND DISCUSSION

In this section, RBTransformer is extensively evaluated on the SEED, DEAP, and DREAMER datasets under a subject-dependent setting for both binary and multi-class classification tasks, and results are compared with existing state-of-the-art models. An extended metric evaluation is carried out for model performance validation, and to analyze class-wise predictions and visualize RBTransformer's ability to distinguish and segregate class clusters in the latent space, confusion matrices and t-SNE plots are presented, respectively, across all datasets and dimensions for both classification settings. Finally, an ablation study is carried out to assess the impact of individual model components.

A. Subject-Dependent Emotion Recognition Evaluation

For performance evaluation, RBTransformer is assessed under a subject-dependent setting, in which the model is trained and evaluated separately on data from the same subjects. The entire dataset, which comprises EEG recordings across all subjects and trials is first preprocessed, and then is divided into training and validation sets using an 80–20 train–val split. A 5 fold cross-validation is then applied, where each fold maintains the subject-dependent setting. This means that within each fold, data from the same subjects can appear in both the training and validation splits, but always as distinct subsets. Same procedure is applied consistently across all three datasets, along all three dimensions, for both binary and multi-class classification tasks.

B. Binary-Class Classification Evaluation Results

RBTransformer is first evaluated under a binary-class classification setting across DEAP and DREAMER, along their three affective dimensions: Valence, Arousal, and Dominance. For binary classification, the labels are converted as follows: DEAP labels (1–9) are split into "High" (above 5) and "Low" (5 or below), while DREAMER labels (1–5) are split into "High" (above 3) and "Low" (3 or below). The performance of RBTransformer is compared against the existing state-of-the-art models, and the detailed results for both datasets under a binary-class classification setting are listed below.

TABLE I
BINARY-CLASS CLASSIFICATION PERFORMANCE COMPARISON ON DEAP DATASET

DEAP Dataset			
Model	Valence	Arousal	Dominance
ACRNN	93.72 ± 3.21	93.38 ± 3.73	–
GANSER	93.86 ± –	94.38 ± –	–
4D-CRNN	94.22 ± 2.61	94.58 ± 3.69	–
BiDCNN	94.38 ± 2.61	94.72 ± 2.56	–
CLDTA	94.58 ± 1.40	94.11 ± 2.10	–
DFCN	94.59 ± –	95.32 ± –	94.78 ± –
RACNN	96.65 ± 2.65	97.11 ± 2.01	–
4D-ANN	96.90 ± 1.65	97.39 ± 1.75	–
TRPO-NET	97.87 ± 1.89	98.08 ± 1.83	98.33 ± 1.55
TDMNN	98.08 ± 2.13	98.25 ± 2.85	–
RBTransformer (Ours)	99.84 ± 0.02	99.83 ± 0.05	99.82 ± 0.06

TABLE II
BINARY-CLASS CLASSIFICATION PERFORMANCE COMPARISON ON DREAMER DATASET

DREAMER Dataset			
Model	Valence	Arousal	Dominance
GANSER	85.28 ± –	84.16 ± –	–
DGCNN	86.23 ± 12.29	84.54 ± 10.18	85.02 ± 10.25
DFCN	93.15 ± –	91.30 ± –	92.04 ± –
RACNN	95.55 ± 2.18	97.01 ± 2.74	–
ACRNN	97.93 ± 1.73	97.98 ± 1.92	98.23 ± 1.42
BiDCNN	98.35 ± 0.87	98.66 ± 1.46	99.01 ± 0.96
TRPO-NET	98.86 ± 0.57	98.97 ± 0.49	98.93 ± 0.69
TDMNN	99.45 ± 0.91	99.51 ± 0.79	–
RBTransformer (Ours)	99.61 ± 0.05	99.74 ± 0.06	99.79 ± 0.04

Experimental results under the binary-class classification setting are reported in Table I and Table II for the DEAP and DREAMER datasets, respectively, and are reported as mean accuracy ± standard deviation (ACC ± STD). On DEAP dataset, RBTransformer achieves ACC ± STD values of 99.84% ± 0.02, 99.83% ± 0.05, and 99.82% ± 0.06 along its Valence, Arousal, and Dominance dimensions, improving the previous state-of-the-art results by 1.76% and 1.58% on Valence and Arousal, respectively. Similarly, on DREAMER dataset, RBTransformer achieves ACC ± STD values of 99.61% ± 0.05, 99.74% ± 0.06, and 99.79% ± 0.04 along

it's Valence, Arousal, and Dominance dimensions, improving the previous state-of-the-art results by 0.16% and 0.23% on Valence and Arousal, respectively.

C. Multi-Class Classification Evaluation Results

Similarly, RBTransformer is also assessed on multi-class classification tasks across SEED, DEAP, and DREAMER. SEED is inherently a multi-class dataset with three emotion classes: Positive, Neutral, and Negative. For DEAP and DREAMER, the original continuous emotion ratings across three primary emotional dimensions, Valence, Arousal, and Dominance, are converted into discrete classes. Specifically, DEAP is converted into nine classes (1 to 9) for each dimension, while DREAMER is converted into five classes (1 to 5). RBTransformer's performance is directly compared against the existing state-of-the-art models on the task of multi-class classification, and the detailed results for each dataset are listed below.

TABLE III
MULTI-CLASS CLASSIFICATION PERFORMANCE COMPARISON ON SEED DATASET

SEED Dataset	
Model	Accuracy
DGCNN	90.40 $\pm$ 8.49
4D-CRNN	94.74 $\pm$ 2.32
CLDTA	95.09 $\pm$ 4.48
SST-EmotionNet	96.02 $\pm$ 2.17
4D-ANN	96.25 $\pm$ 1.86
TDMNN	97.20 $\pm$ 1.57
3DCANN	97.35 $\pm$ -
GANSER	97.71 $\pm$ -
RBTransformer (Ours)	99.51 $\pm$ 0.02

TABLE IV
MULTI-CLASS CLASSIFICATION PERFORMANCE COMPARISON ON DEAP DATASET

DEAP Dataset			
Model	Valence	Arousal	Dominance
TRPO-NET	97.63 $\pm$ 2.38	97.74 $\pm$ 2.26	97.88 $\pm$ 2.24
RBTransformer (Ours)	99.87 $\pm$ 0.04	99.84 $\pm$ 0.04	99.87 $\pm$ 0.05

TABLE V
MULTI-CLASS CLASSIFICATION PERFORMANCE COMPARISON ON DREAMER DATASET

DREAMER Dataset			
Model	Valence	Arousal	Dominance
TRPO-NET	98.18 $\pm$ 0.97	98.37 $\pm$ 0.93	98.40 $\pm$ 0.80
RBTransformer (Ours)	99.54 $\pm$ 0.06	99.55 $\pm$ 0.04	99.60 $\pm$ 0.05

Experimental results under the multi-class classification setting are reported in Table III, Table IV, and Table V for

the SEED, DEAP, and DREAMER datasets, respectively, and are reported as mean accuracy $\pm$ standard deviation (ACC $\pm$ STD). On SEED dataset, RBTransformer achieves an ACC $\pm$ STD of 99.51% $\pm$ 0.02, improving the previous state-of-the-art results by 1.80%. On DEAP dataset, it achieves ACC $\pm$ STD values of 99.87% $\pm$ 0.04, 99.84% $\pm$ 0.04, and 99.87% $\pm$ 0.05 along it's Valence, Arousal, and Dominance dimensions, respectively, improving the previous state-of-the-art results by 2.10%, 2.24%, and 1.99%. Similarly, on DREAMER dataset, RBTransformer achieves ACC $\pm$ STD values of 99.54% $\pm$ 0.06, 99.55% $\pm$ 0.04, and 99.60% $\pm$ 0.05 along it's Valence, Arousal, and Dominance dimensions, improving the previous state-of-the-art results by 1.18%, 1.36%, and 1.20%, respectively.

D. Ablation Study

To evaluate the contribution of different architectural components, training configurations, and regularization choices for RBTransformer, five ablation experiments are carried out on the DREAMER dataset along the Arousal dimension for the binary classification task. Results are reported as the mean and standard deviation over five cross-validation folds. In the first ablation, the Inter-Cortical Attention mechanism is removed from the model to assess its impact.

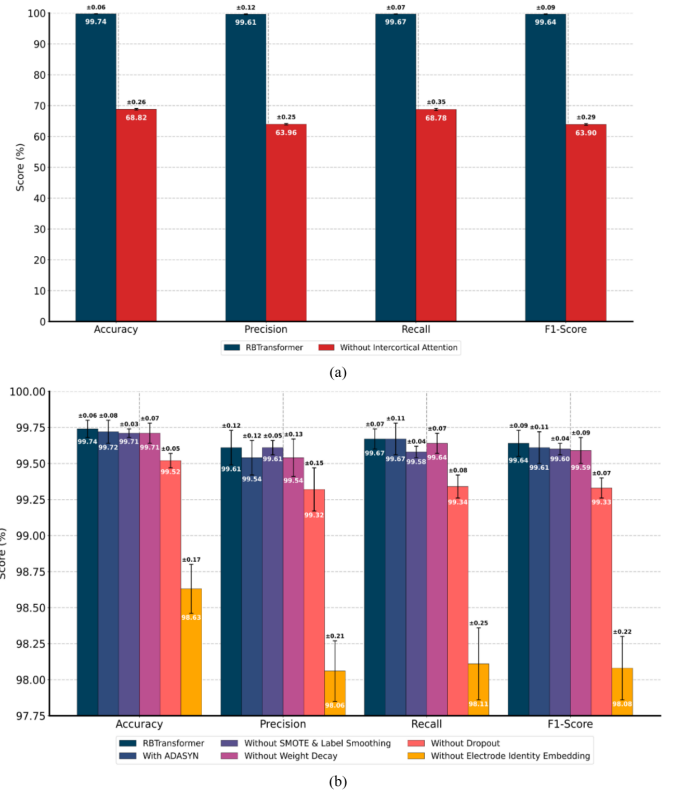


Fig. 6. Ablation study for binary classification on the DREAMER dataset (Arousal dimension). Top: Impact of inter-cortical attention. Bottom: Impact of training and regularization components including ADASYN, SMOTE with label smoothing, weight decay, and dropout.

As shown in Figure 6.(a), this leads to a sharp drop of around 30% across all four metrics, highlighting the critical

TABLE VI
EXTENDED METRIC EVALUATION (ACCURACY, PRECISION, RECALL, F1-SCORE), MEAN $\pm$ SD ON SEED, DEAP, AND DREAMER DATASETS (BINARY AND MULTI-CLASS CLASSIFICATION)

Metric	SEED	DEAP						DREAMER					
		Binary-Class Classification			Multi-Class Classification			Binary-Class Classification			Multi-Class Classification		
		Valence	Arousal	Dominance	Valence	Arousal	Dominance	Valence	Arousal	Dominance	Valence	Arousal	Dominance
Accuracy (%)	99.51 $\pm$ 0.02	99.84 $\pm$ 0.02	99.83 $\pm$ 0.05	99.82 $\pm$ 0.06	99.87 $\pm$ 0.04	99.84 $\pm$ 0.04	99.87 $\pm$ 0.05	99.61 $\pm$ 0.05	99.74 $\pm$ 0.06	99.79 $\pm$ 0.04	99.54 $\pm$ 0.06	99.55 $\pm$ 0.04	99.60 $\pm$ 0.05
Precision (%)	99.51 $\pm$ 0.02	99.84 $\pm$ 0.02	99.83 $\pm$ 0.06	99.81 $\pm$ 0.07	99.87 $\pm$ 0.06	99.84 $\pm$ 0.05	99.86 $\pm$ 0.06	99.58 $\pm$ 0.05	99.61 $\pm$ 0.12	99.63 $\pm$ 0.06	99.54 $\pm$ 0.06	99.49 $\pm$ 0.06	99.56 $\pm$ 0.05
Recall (%)	99.51 $\pm$ 0.02	99.83 $\pm$ 0.02	99.83 $\pm$ 0.06	99.81 $\pm$ 0.06	99.86 $\pm$ 0.03	99.84 $\pm$ 0.05	99.87 $\pm$ 0.04	99.60 $\pm$ 0.06	99.67 $\pm$ 0.07	99.72 $\pm$ 0.09	99.54 $\pm$ 0.05	99.46 $\pm$ 0.09	99.58 $\pm$ 0.14
F1-score (%)	99.51 $\pm$ 0.02	99.83 $\pm$ 0.02	99.83 $\pm$ 0.06	99.81 $\pm$ 0.07	99.86 $\pm$ 0.04	99.84 $\pm$ 0.04	99.87 $\pm$ 0.05	99.59 $\pm$ 0.06	99.64 $\pm$ 0.09	99.68 $\pm$ 0.06	99.54 $\pm$ 0.05	99.47 $\pm$ 0.07	99.57 $\pm$ 0.09

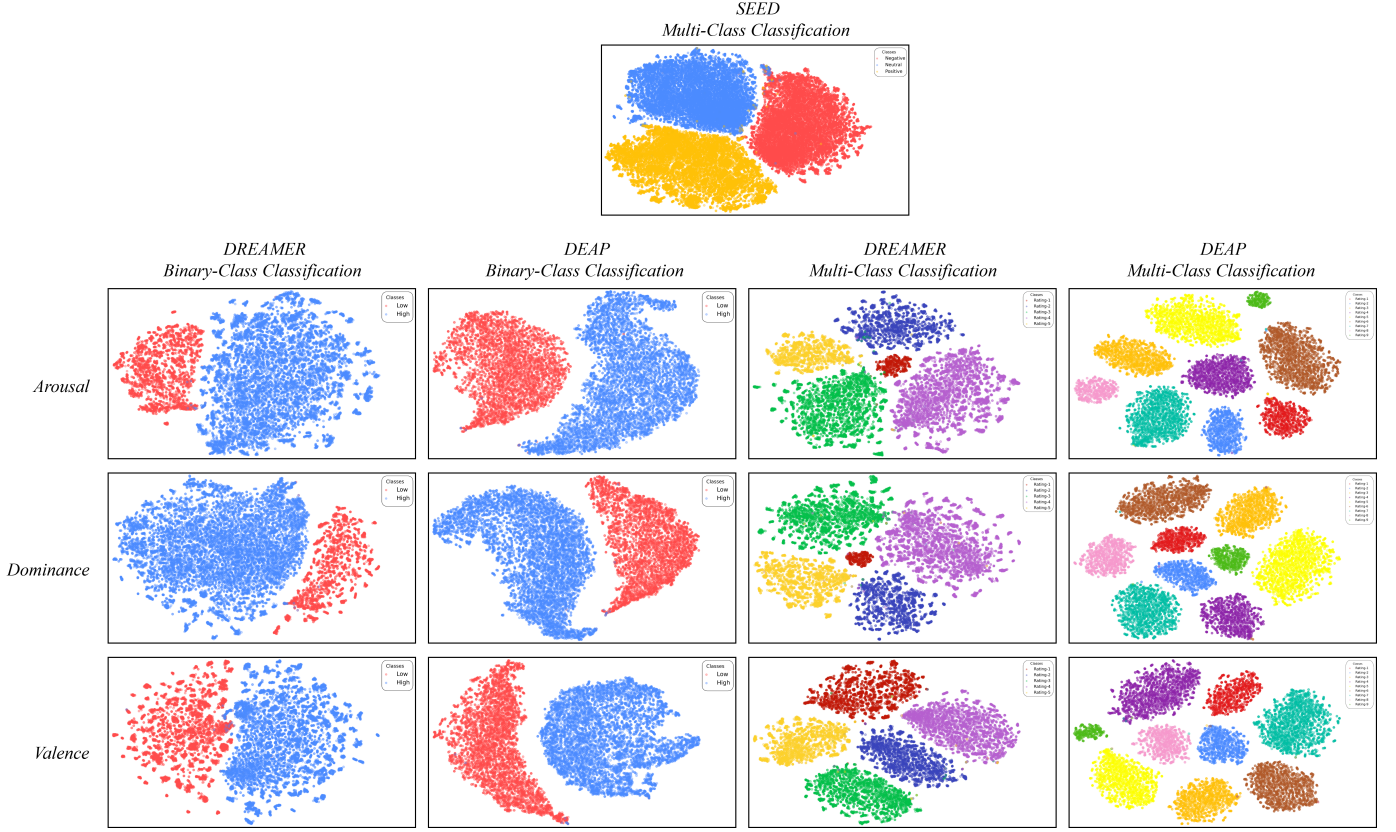


Fig. 7. t-SNE visualization of RBTransformer’s learned feature representations on SEED, DEAP, and DREAMER for both binary and multi-class classification tasks across all dimensions (Valence, Arousal, and Dominance).

importance of the Inter-Cortical Attention mechanism. Figure 6.(b) presents the remaining four ablations: the second ablation replaces SMOTE with ADASYN; the third disables both SMOTE and label smoothing; the fourth removes weight decay; and the fifth removes dropout. In all cases, the performance ranks lower than the final RBTransformer configuration, demonstrating the effectiveness and necessity of the chosen architectural and training design decisions.

E. Extended Metric Evaluation for Performance Validation

While accuracy provides a general indication of performance, it may overlook model’s class-wise prediction capability. So, to assess the robustness and generalizability of RBTransformer in depth, an extended evaluation of metrics, including Precision, Recall, and F1-Score is carried out. Precision reflects the model’s ability to avoid false positives, Recall captures how well it retrieves relevant instances, and

F1-Score balances the two. Evaluating these metrics helps assess whether the model performs reliably across all emotional classes, including less frequent ones. The extended metric evaluation is carried out on the SEED, DEAP, and DREAMER datasets across the three affective dimensions, Valence, Arousal, and Dominance (for DEAP and DREAMER), for both binary and multi-class classification tasks. Results are reported as the mean and standard deviation over five cross-validation folds and include Accuracy, Precision, Recall, and F1-Score. As shown in Table 1, RBTransformer demonstrates balanced, stable, and consistently high performance across all metrics, datasets, and corresponding dimensions, supporting its overall accuracy and ensuring that the observed accuracy is not driven by biased learning or class imbalance.

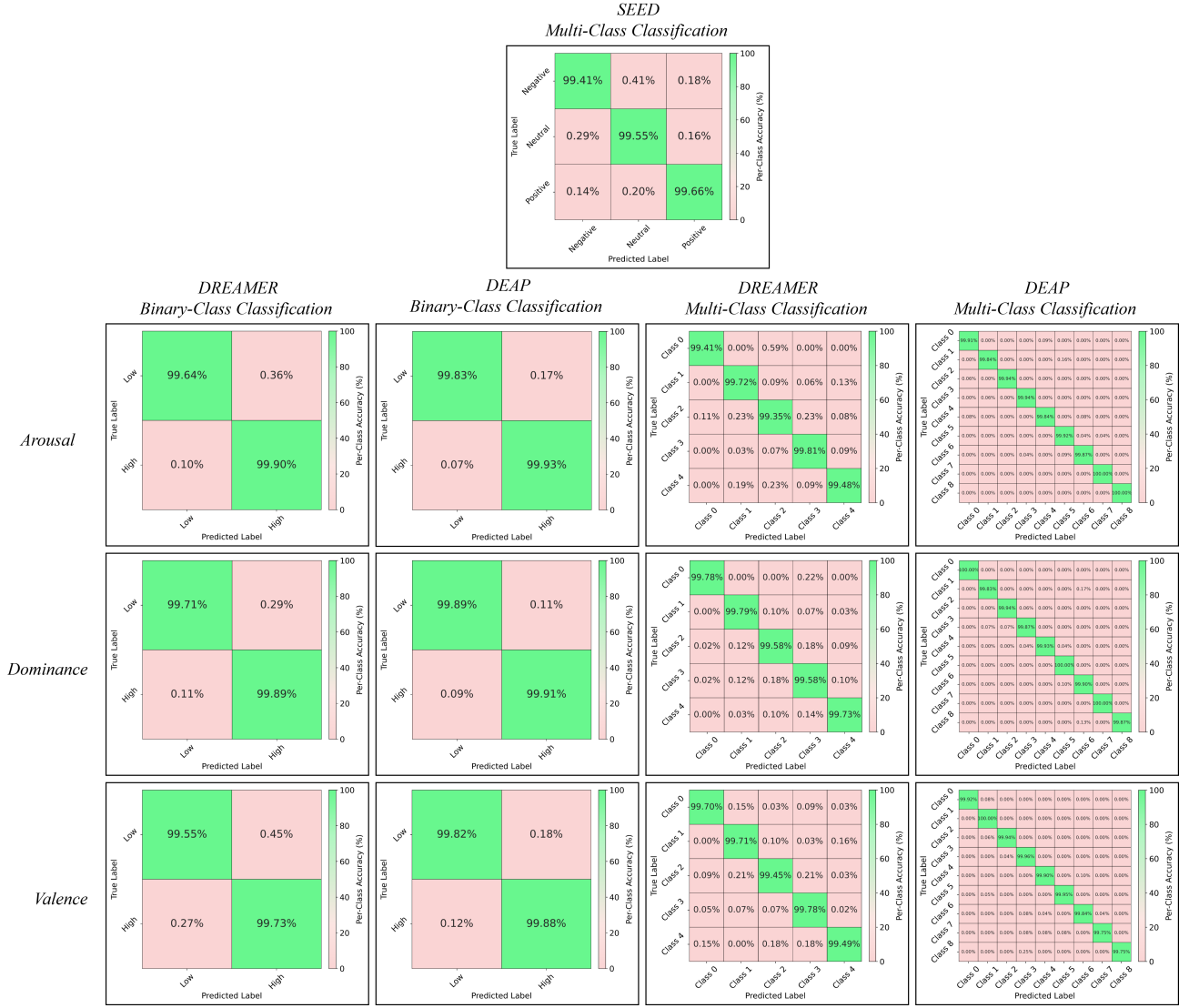


Fig. 8. Confusion matrices of RBTransformer predictions on SEED, DEAP, and DREAMER for both binary and multi-class classification across all dimensions (Valence, Arousal, and Dominance).

F. Visualization of Learned Feature Representations by RBTransformer

To visualize RBTransformer’s ability to classify unseen EEG signals, its high-dimensional operational feature space is projected into two dimensions using the t-SNE (t-Distributed Stochastic Neighbor Embedding) algorithm. As observable from Figure 7, RBTransformer demonstrates strong performance across all datasets: SEED shows clear separation of Positive, Neutral, and Negative classes; in binary classification, DEAP and DREAMER form distinct groupings for High and Low classes across all dimensions, Valence, Arousal, and Dominance; and in multi-class settings, DEAP’s nine and DREAMER’s five emotion classes also form well-segregated clusters across these same dimensions, validating the model’s ability to capture and differentiate emotional states in the learned feature space.

G. Confusion Matrix Analysis

In addition to plotting t-SNE plots for visualizing RBTransformer’s ability to distinguish between different class clusters in the latent space, we plot confusion matrices for all the datasets to further analyze the class-wise prediction capability of RBTransformer in detail. The confusion matrices are generated for all three datasets, SEED, DEAP, and DREAMER, across all three affective dimensions, Valence, Arousal, and Dominance (for DEAP and DREAMER), under both binary and multi-class classification tasks, shown in Figure 8. As observable from the plots, RBTransformer achieves a very high alignment between predicted and true labels, with most values concentrated along the principal diagonal. On average, the model achieves class-wise accuracy of around 99.50%, with only minor misclassifications appearing in off-principal-diagonal entries, providing a quantitative and visual confirma-

tion of RBTransformer’s strong class-wise performance across all datasets and affective dimensions under both classification settings.

V. CONCLUSION

In this paper, we propose RBTransformer, a Transformer-based neural network architecture that simulates and captures inter-cortical neural interactions in latent space for effective EEG-based emotion recognition. RBTransformer first converts raw EEG signals into Band Differential Entropy (BDE) tokens that are then passed into an Electrode-Identity Embedding, which allows the model to retain spatial awareness of each electrode’s position and ordering. These intermediate features are then passed through an Inter-Cortical Attention module, which is made up of successive Inter-Cortical Multi-Head Attention blocks, each comprised of electrode $\times$ electrode attention matrix, allowing each electrode to interact directly with every other electrode in the latent space, and a feedforward network. These attention blocks are stacked on top of one another, which mimics the recurrent exchange of signals between distinct brain regions, allowing RBTransformer to capture both inter-regional spatial dependencies and localized temporal dependencies, without relying on any handcrafted features or explicit temporal sequence modeling. Experimental results demonstrate that RBTransformer achieves state-of-the-art performance, outperforming all existing models, across all benchmarks, DEAP, DREAMER, and SEED, across all dimensions, Valence, Arousal, and Dominance (for DEAP and DREAMER), under both binary and multi-class EEG-based emotion recognition tasks.

REFERENCES

- [1] W. James, “What is an emotion?” *Mind*, vol. 9, no. 34, pp. 188–205, 1884.
- [2] S. Schachter and J. E. Singer, “Cognitive, social, and physiological determinants of emotional state,” *Psychol. Rev.*, vol. 69, no. 5, pp. 379–399, 1962.
- [3] C. E. Izard, J. Kagan, and R. B. Zajonc, Eds., *Emotion, Cognition, and Behavior*. Cambridge, UK: Cambridge University Press, 1984.
- [4] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.
- [5] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” *Inf. Fusion*, vol. 37, pp. 98–125, 2017.
- [6] S. K. D’Mello and J. Kory, “A review and meta-analysis of multimodal affect detection systems,” *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–36, 2015.
- [7] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010.
- [8] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “Deap: A database for emotion analysis using physiological signals,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2011.
- [9] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Washington, DC: American Psychiatric Association, 2013.
- [10] World Health Organization, *International Classification of Diseases, 11th Revision (ICD-11)*, 11th ed. Geneva, Switzerland: World Health Organization, 2018.
- [11] P. A. Kragel and K. S. LaBar, “Decoding the nature of emotion in the brain,” *Trends Cognit. Sci.*, vol. 20, no. 6, pp. 444–455, 2016.
- [12] R. Adolphs, “How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences,” *Soc. Cogn. Affect. Neurosci.*, vol. 12, no. 1, pp. 24–31, 2017.
- [13] P. Ekman, “An argument for basic emotions,” *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [14] J. A. Russell, “A circumplex model of affect,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [15] A. Mehrabian, “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament,” *Curr. Psychol.*, vol. 14, no. 4, pp. 261–292, 1996.
- [16] L. F. Barrett and J. A. Russell, “The structure of current affect: Controversies and emerging consensus,” *Curr. Dir. Psychol. Sci.*, vol. 8, no. 1, pp. 10–14, 1999.
- [17] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, “The brain basis of emotion: A meta-analytic review,” *Behav. Brain Sci.*, vol. 35, no. 3, pp. 121–143, 2012.
- [18] H. Doi, S. Nishitani, and K. Shinohara, “Nirs as a tool for assaying emotional function in the prefrontal cortex,” *Front. Hum. Neurosci.*, vol. 7, p. 770, 2013.
- [19] B. Hjorth, “Eeg analysis based on time domain properties,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.
- [20] P. C. Petrantonakis and L. J. Hadjileontiadis, “Emotion recognition from eeg using higher order crossings,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, 2009.
- [21] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, “Differential entropy feature for eeg-based vigilance estimation,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 6627–6630.
- [22] S. M. Redwan, M. P. Uddin, A. Ulhaq, M. I. Sharif, and G. Krishnamoorthy, “Power spectral density-based resting-state eeg classification of first-episode psychosis,” *Sci. Rep.*, vol. 14, no. 1, p. 15154, 2024.
- [23] N. Kumar, K. Alam, and A. H. Siddiqi, “Wavelet transform for classification of eeg signal using svm and ann,” *Biomed. Pharmacol. J.*, vol. 10, no. 4, pp. 2061–2069, 2017.
- [24] M. Ende, A. K. Louis, P. Maass, and G. Mayer-Kress, “Eeg signal analysis by continuous wavelet transform techniques,” in *Nonlinear Analysis of Physiological Data*, H. Kantz, J. Kurths, and G. Mayer-Kress, Eds. Berlin, Heidelberg: Springer, 1998, pp. 213–219.
- [25] K.-E. Ko, H.-C. Yang, and K.-B. Sim, “Emotion recognition using eeg signals with relative power values and bayesian network,” *International Journal of Control, Automation and Systems*, vol. 7, no. 5, pp. 865–870, 2009.
- [26] P. C. Petrantonakis and L. J. Hadjileontiadis, “Emotion recognition from eeg using higher order crossings,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.
- [27] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, “Differential entropy feature for eeg-based emotion classification,” in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.
- [28] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [30] W. Zheng and B. Lu, “Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks,” *IEEE Trans. Auton. Mental Dev.*, vol. 7, no. 3, pp. 162–175, 2015.
- [31] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, and D. Jiang, “Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for eeg emotion recognition,” *Neurocomputing*, vol. 448, pp. 140–151, 2021.
- [32] E. Rudakov, L. Laurent, V. Cousin, A. Roshdi, R. Fournier, A. Nait-Ali, T. Beyrouthy, and S. A. Kork, “Multi-task cnn model for emotion recognition from eeg brain maps,” in *Proc. 4th Int. Conf. Bio-Engineering for Smart Technologies (BioSMART)*, 2021, pp. 1–4.
- [33] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal recurrent neural network for emotion recognition,” *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, 2018.
- [34] J. Ma, H. Tang, W.-L. Zheng, and B.-L. Lu, “Emotion recognition using multimodal residual LSTM network,” in *Proc. 27th ACM Int. Conf. Multimedia (MM)*, 2019, pp. 176–183.
- [35] J. L. Elman, “Finding structure in time,” *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, "Eeg-based emotion recognition using 4d convolutional recurrent neural network," *Cogn. Neurodyn.*, vol. 14, pp. 815–828, 2020.
- [38] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–7.
- [39] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, 2018.
- [40] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm," *Appl. Soft Comput.*, vol. 100, p. 106954, 2021.
- [41] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, 2008.
- [42] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "Eeg-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 382–393, 2020.
- [43] X. Xiao, X. Chen, Y. Wang, T. Ma, X. Zhang, and X. Chen, "Multi-frequency band and temporal-spatial based attention neural network for eeg emotion recognition," *Sensors*, vol. 20, no. 21, p. 5986, 2020.
- [44] X. Liu, X. Yang, Y. Du, W. Zeng, Q. Chen, Y. Zhou, C. Li, and Y. Xu, "Emotion recognition from multichannel eeg data through dual channel 3d cnn model," *Frontiers in Neuroscience*, vol. 16, 2022.
- [45] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, 2017.
- [46] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.
- [47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.
- [48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.
- [49] —, "SGDR: Stochastic gradient descent with warm restarts," in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017.
- [50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
- [51] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [52] L. Biewald, "Experiment tracking with weights and biases," 2020, software available from wandb.com. [Online]. Available: <https://www.wandb.com/>