FeelNet: A Lightweight Fast Fourier Transform EEG-based Emotion Recognition Network

Xueyao Wang[1,2] wangxueyao221@mails.ucas.ac.cn
Xiuding Cai[1,2] caixiuding20@mails.ucas.ac.cn
Yaoyao Zhu[1,2] zhuyaoyao19@mails.ucas.ac.cn
Yu Yao[1,2*] casitmed2022@163.com
[1] *Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu, China*
[2] *School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China*

Editors: Hung-yi Lee and Tongliang Liu

Abstract

Emotion recognition using Electroencephalography (EEG) is challenging due to its low signal-to-noise ratios and high-dimensional sparsity. We propose FeelNet, a novel Fast Fourier Transform (FFT)-based architecture that simultaneously extracts global and local features across joint frequency-time domains. FeelNet incorporates an adaptive Rhythm Spectral Block (RSB) for capturing key frequency patterns and filtering task-irrelevant noise through power spectral thresholding. Additionally, the Multi-scale Temporal Conv Block (MTCB) enhances the model's ability to decode complex temporal dynamics. Extensive evaluations on the DEAP and DREAMER datasets demonstrate that FeelNet outperforms existing state-of-the-art methods in accuracy and flexibility, even under noise-contaminated conditions. Owing to its computational efficiency and noise resilience, FeelNet provides an alternative perspective for EEG-based affective computing.

Keywords: Fast Fourier Transform; Convolutional Neural Networks; Emotion Recognition based EEG.

1. Introduction

Emotion recognition is crucial for understanding human behavior, as affective states dynamically modulate cognitive appraisal, behavioral responses, and physiological adaptation (Willett et al., 2021; Lu et al., 2023; Manral and Seeja, 2023; Lian et al., 2023). Electroencephalography (EEG) has emerged as a privileged modality for objective emotion assessment due to its millisecond-scale temporal resolution and direct neural correlates. Recent advances in deep learning have propelled EEG-based emotion recognition from hand-crafted feature paradigms (e.g., Differential Entropy (Duan et al., 2013), Power Spectrum Density features (Jia et al., 2020)) toward end-to-end representation learning, achieving state-of-the-art accuracy on benchmark datasets like SEED (Zheng and Lu, 2015) and DEAP (Koelstra et al., 2011). Nevertheless, EEG signals still challenge low signal-to-noise ratios and high-dimensional sparsity.

Deep learning technologies address these challenges through hierarchical abstraction of spatio-spectro-temporal patterns, demonstrating superior noise invariance. Among these,

[*] Corresponding author

Convolutional Neural Networks (CNNs) (Shen et al., 2022; Ding et al., 2023, 2024) have demonstrated particular promise in EEG emotion recognition by extracting local temporal features and enabling dimensionality reduction. TSception (Ding et al., 2023) employs multi-scale temporal convolutions to capture dynamic dependencies, while CLISA (Shen et al., 2022) leverages contrastive learning to enhance feature separability. However, the inherent local bias of CNNs impedes modeling of global neurodynamics, and critically fails to address spectral non-stationarity.

This limitation has spurred time-frequency integration paradigms, exemplified by SST-EmotionNet (Jia et al., 2020) which projects EEG signals into 2D spectral-temporal maps across five frequency bands. Notably, neural oscillations in the beta band (13-40 Hz) have been implicated in emotional processing and modulation (Zheng et al., 2019). However, existing methods still face significant challenges in spectral feature selection and denoising, often failing to fully capture the complexity and diversity of spectral information.

We propose **FeelNet**, a novel lightweight FFT-based EEG emotion recognition network that unifies temporal and spectral modeling. Inspired by spectral and hierarchical architectures such as TSLANet (Eldele et al., 2024), FeelNet incorporates neuroscientifically-grounded priors through five rhythm-specific spectral filters $(\delta, \theta, \alpha, \beta, \gamma)$ designed to attenuate task-irrelevant noise and amplify emotion-relevant spectral signatures. As illustrated in **Fig. 1**, FeelNet integrates two synergistic components: the Rhythm Spectral Block (RSB) and the Multi-scale Temporal Conv Block (MTCB). First, the RSB module transforms time-domain EEG signals into the Fourier domain, applying global filters and rhythm filters to capture spatial dependencies and emotion-relevant frequency features while suppressing noise. The filtered features are then returned to the time domain via Inverse Fast Fourier Transform (IFFT). Finally, the MTCB module integrates the multi-scale spatiotemporal features after the RSB module filtering, improving the model's ability to capture both local and global dependencies.

Extensive experiments demonstrate that FeelNet achieves state-of-the-art performance on DEAP and DREAMER datasets. By jointly modeling spectral and temporal dynamics, FeelNet significantly enhances both recognition accuracy and noise robustness. Furthermore, analysis methods like brain topography visualization and model ablations can extract neurophysiologically interpretable features from the FeelNet model. The following summarizes the key contributions of this work:

- We propose FeelNet, a novel lightweight EEG emotion recognition framework that integrates joint frequency-time processing to capture global spatial-spectral dependencies via Fast Fourier Transform.

- We propose the Rhythm Spectral Block (RSB) to filter task-irrelevant noise via power spectral thresholding, enabling the amplify emotion-relevant rhythms. Additionally, we propose the Multi-scale Temporal Conv Block (MTCB) to capture multi-scale spatiotemporal patterns after the RSB module filtering, facilitating deeper and more effective feature fusion with minimal parameters.

- Much smaller computational costs and higher recognition accuracy are found through experiments.

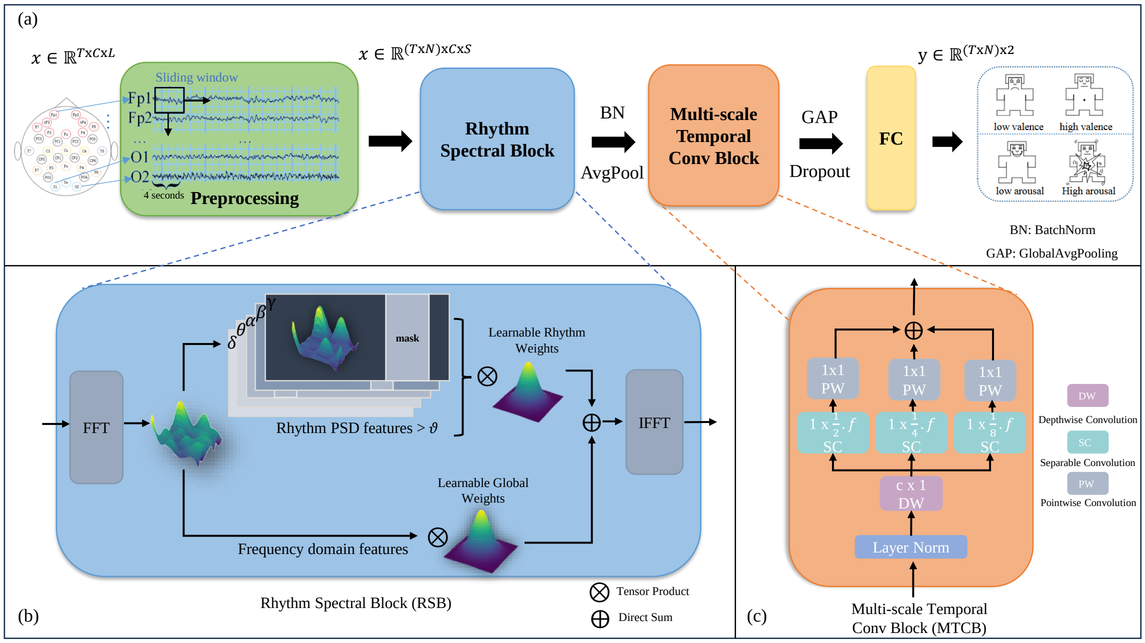

Figure 1: The proposed FeelNet structure. The input EEG signals are segmented into tokens and positional embeddings are added. Next, the preprocessed data are passed through the Rhythm Spectral Block that uses frequency domain representation for reliable feature extraction and adaptive rhythm threshold for noise reduction. The Multi-scale Temporal Conv Block is then used to integrate fine-grained frequency information and capture complex spatio-temporal patterns through convolutional operations. The final classification distinguishes between low/high arousal and valence through global pooling and fully connected layers.

2. Related Work

2.1. Spatiotemporal Feature Extraction

CNNs (Schirrmeister et al., 2017; Lawhern et al., 2018; Ingolfsson et al., 2020; Zhang et al., 2023; Huang et al., 2024; Liu et al., 2024) have showed significant effectiveness extracting spatiotemporal features from EEG signals, particularly in brain-computer interface applications. Lawhern et al. (2018) introduced EEGNet, a compact CNN architecture that combines deep and separable convolutional kernels for spatiotemporal feature extraction. Ding et al. (2023) proposed TSception, a multi-scale spatiotemporal CNN network to extract temporal dynamics and spatial asymmetry features in emotion classification based on EEG signals. Still, it is prone to overfitting on a small dataset due to the local bias. So, models like EEG-TCNet (Ingolfsson et al., 2020) and ERTNet (Liu et al., 2024) have been proposed, leveraging self-attention mechanisms (Tao et al., 2023) and dilated convolutions (Yu and Koltun, 2015) to capture complex temporal patterns and long-range dependencies. Nevertheless, these models suffer from high resource consumption and performance degradation as sequence length increases. To address these limitations, FeelNet reduces computational complexity by transforming the signal into the frequency domain without sacrificing model performance, thanks to the complexity of $\mathcal{O}(N \log N)$ of the FFT (Baxes, 1994).

2.2. Frequency Feature Extraction

The frequency domain (Jenke et al., 2014; Zheng et al., 2019; Jia et al., 2020; Rao et al., 2021, 2023; Tatsunami and Taki, 2024) has been extensively explored in EEG-based emotion recognition, as EEG signals with predefined frequency bands associated with specific cognitive and affective processes. While architectures like SST-EmotionNet (Jia et al., 2020) combine convolutional and recurrent layers to capture spatio-spectral features, their reliance on fixed frequency bands (e.g., rigid 4-8Hz θ band) fundamentally limits adaptability to neurobiological variability across individuals. To address this spectral rigidity, TSLANet (Eldele et al., 2024) pioneers adaptive Fourier analysis with data-driven thresholding, dynamically extracting multi-scale temporal dependencies. However, the global denoising paradigm unintentionally attenuates transient yet discriminative oscillatory patterns, such as gamma bursts greater than 30Hz (Zheng et al., 2019) correlating with emotional arousal. This has the effect of compromising feature fidelity in noisy environments. To preserve these behaviorally critical transients while maintaining denoising efficacy, FeelNet introduces the rhythm filtering and multi-scale feature extraction, thereby enabling robust emotion recognition with lower computational costs and enhanced noise suppression.

3. Methodology

3.1. Architecture Overview

The proposed FeelNet model comprises two core components: the Rhythm Spectral Block (RSB) and the Multi-scale Temporal Conv Block (MTCB) (**Fig. 1a**). The RSB module (**Fig. 1b**) processes EEG signals in the Fourier domain, enhancing emotion-relevant features through rhythmic spectral filtering. The MTCB module (**Fig. 1c**) then captures multi-scale spatiotemporal patterns from the reconstructed time-domain signal. This dual-domain approach enables comprehensive feature extraction for emotion recognition. Final classification distinguishes low/high arousal (calm/excited) and valence (unpleasant/pleasant) through global pooling and fully-connected layers.

3.2. Preprocessing

Given EEG signals as the input x, each signal $x \in \mathbb{R}^{T\times C\times L}$ has T trials, C channels, and a sample length L. First, the signal x is divided into N segments $S_i = \{S_1, S_2, ..., S_N\}$ along the L-axis (i.e., length) by the non-overlapping sliding window. Each segment has a length $l = G \times f$, where f is the sampling frequency and G is the time interval. Each segment $S_i \in \mathbb{R}^{T\times N\times C\times l}$ is normalized along the C-axis (i.e., channel). To maintain the order of the segments, positional embeddings E_i are added to each segment. The enhanced input is represented as $P_{SE_i} = S_i + E_i$ (Vaswani et al., 2017), enabling the model to capture spatiotemporal relationships in EEG data effectively.

3.3. Rhythm Spectral Block (RSB)

Inspired by TSLANet (Eldele et al., 2024), we propose the **Rhythm Spectral Block** (RSB) using Fourier analysis to capture the temporal and spectral charateristics of biological rhythms. The design of the RSB module incorporates a priori neuroscientific principles (Buzsáki and Draguhn, 2004) to implement cognitively-gated spectral filtering in **Fig. 1b**. In particular, it is applicable to rhythm-specific thresholds that are aligned with established affective neurodynamics. These are the delta band δ (0.5-4 Hz), theta band θ (4-8 Hz), alpha band α (8-13 Hz), beta band β (13-30 Hz), and gamma band γ (30-50 Hz) (Buzsáki and Draguhn, 2004). The suppression of extraneous noise is achieved while selectively refining emotion-related spectral components. This biologically constrained design amplifies discriminative oscillatory signatures, thereby enhancing feature separability for emotion recognition.

We apply an FFT operation along the spatial dimension to the position-embedded segmented EEG signal P_{SE} to obtain its frequency domain representation:

$$F = \mathcal{F}[P_{SE}] \in \mathbb{C}^{T\times N\times C\times l'}, \tag{1}$$

where $\mathcal{F}[\cdot]$ represents a one-dimensional FFT operation and l' is the transformed sample length in the frequency domain.

Then, we propose an adaptive frequency-specific masking mechanism to amplify emotion-discriminative spectral components while suppressing noise. For each rhythm band $F_i \in \{\delta, \theta, \alpha, \beta, \gamma\}$, we firstly compute the Power Spectral Density (PSD): $P_i = |F_i|^2$ and apply the learnable thresholds ϑ_i optimized by gradient backpropagation to generate filtered data $F_{filtered}$:

$$F_{filtered} = \sum_{i=1}^{n} F_i \odot (P_i > \vartheta_i) \in \mathbb{C}^{T\times N\times C\times l'}, \tag{2}$$

where $\odot$ is the element-wise multiplication (Hadamard product). The way to reduce noise is through adaptively masking unrelated-frequency components by $P_i > \vartheta_i$.

The RSB module employs two parallel learnable filters: a global filter W_G that processes the full-band frequency data and a rhythm filter W_L that handles the adaptively filtered data F_G and F_L. $F_{integrated}$ is then integrated by using learnable scalar parameters μ and ρ to balance their contributions:

$$F_G = W_G \odot F, \tag{3}$$

$$F_L = W_L \odot F_{filtered}, \tag{4}$$

$$F_{integrated} = F_G \times \mu + F_L \times \rho. \tag{5}$$

Finally, we utilize the IFFT operation to return the integrated frequency data to the time domain. The resulting time-domain signal, designated as S', is expressed as follows:

$$S' = \mathcal{F}^{-1}[F_{integrated}] \in \mathbb{C}^{T\times N\times C\times L'}. \tag{6}$$

After IFFT, the filtered features maintain the same dimensions as the input. FeelNet using FFT avoids the emergence of the attention mechanism with $\mathcal{O}(N^2)$ complexity, making it efficient and suitable for EEG signal processing.

3.4. Multi-scale Temporal Conv Block (MTCB)

After filtering, EEG signals become two-dimensional time series with spatial (channels) and temporal (time points) dimensions. To better integrate rhythm-filtered features, FeelNet introduces the **Multi-scale Temporal Conv Block** (MTCB) in **Fig. 1c**, which combines spatial depthwise convolution Z_{depth} and multi-scale separable convolution Z_{separ} based on depthwise separable convolution (Lawhern et al., 2018).

First, the spatial depthwise convolution applies a 1D convolutional kernel k_D of size $(c, 1)$ to learn spatial relationships across EEG channels c, resulting in $Z_{depth} \in \mathbb{R}^{n\times d\times c\times f'}$:

$$Z_{depth} = AP(\Phi_{ELU}(Conv2D(S', k_D))), \tag{7}$$

where S' is the output of the RSB module, $Conv2D(\cdot)$ is the 2D convolution operation with the k_D kernel size, and $\Phi_{ELU(\cdot)}$ is the $ELU(\cdot)$ activation function.

Subsequently, a temporal convolution $Z_{separ} \in \mathbb{R}^{n\times s\times c\times f'_i}$ followed by multi-scale separable convolutions with kernels $k_S^i = (1, \eta^i \cdot f), i \in [1, 2, 3]$ captures diverse temporal features:

$$Z^i_{separ} = AP(\Phi_{ELU}(Conv2D(Z_{depth}, k^i_S)), \tag{8}$$

where the ratio coefficient is set at $[0.5, 0.25, 0.125]$ with $\eta = 0.5$, enabling the network to learn multi-scale temporal representations as supported by Lawhern et al. (2018); Ding et al. (2023).

The outputs from each separable convolution level are concatenated and passed through Global Average Pooling (GAP) and a squeeze operation $\Gamma(\cdot)$, followed by a softmax activation to produce the final prediction:

$$\hat{y} = \Phi_{softmax}(\Gamma(GAP([Z^1_{separ}, ..., Z^j_{separ}]))), j \in [1, 2, 3]. \tag{9}$$

The learning objective of this structure is to optimize the gap between predicted and real categories and the loss function uses a cross-entropy loss as:

$$\mathcal{L} = -\left(w_1 \cdot y \cdot \log(\hat{y}) + w_0 \cdot (1 - y) \cdot \log(1 - \hat{y})\right), \tag{10}$$

where w_1 and w_0 are the weights for the positive class and the negative class for the class imbalance problem. For more details, please refer to the Algorithm 1.

4. Experiments

4.1. Experimental Setup

We evaluated the FeelNet model on two widely used EEG datasets: DEAP (Koelstra et al., 2011) and DREAMER (Katsigiannis and Ramzan, 2018), following established preprocessing protocols Ding et al. (2023); Liu et al. (2024). Their details are summarized in **Table 1**. For the DEAP dataset, a 3-second pre-stimulus baseline segment was removed. The data were subsequently downsampled to 128 Hz, and non-EEG channels were excluded. Arousal and valence labels were binarized by applying a threshold of 5 to the Self-Assessment Manikin scores. The trials were then segmented into non-overlapping 4-second epochs. Class weights w_1 and w_0 were both set to 1 for this dataset.

Algorithm 1 FeelNet Algorithm

Input: EEG data $x \in \mathbb{R}^{T\times C\times L}$, ground truth label y
Output: Prediction $\hat{y}$ of FeelNet

```
do in sequential
    // The RSB module processing
    Convert x^n to frequency domain via Eq. 1
    Compute F_integrated using Eqs. 2-5
    Obtain S' via inverse transform via Eq. 6

    // The MTCB module processing
    Calculate Z_depth via Eq. 7 // spatial depthwise convolution
    // multi-scale separable convolution
    for i ← 1 to 3 do
        Determine i-th kernel size via Eq. 8
        Compute Z^i_separ via Eq. 8
    end
end
Predict ŷ via Eq. 9
return ŷ
```

For the DREAMER dataset, the last 60 seconds of each film clip (sampled at 128 Hz) were extracted. These data were segmented into 2-second non-overlapping epochs. Arousal and valence labels were similarly binarized using a threshold of 3. To mitigate class imbalance, class weights were adjusted to $w_1 = 0.55$ and $w_0 = 0.45$. Model performance was evaluated using trial-wise 10-fold cross-validation. Hyperparameters were selected through systematic Bayesian search over defined spaces: learning rate $\in [1 \times e^{-4}, 1 \times e^{-2}]$, batch size $\in \{32, 64, 128\}$, dropout $\in [0.3, 0.7]$. The FeelNet model was implemented in PyTorch and trained using the Adam optimizer with optimal values ($1 \times e^{-3}$, 64, 0.5). A fixed random seed of 2021 was set for reproducibility. Training proceeded for 100 epochs, and performance was assessed using Accuracy (ACC) and F1 score (F1).

Table 1: The dataset content summary.

Dataset	**Subjects (male/female)**	**Age**	**Trials**	**Channels**	**Rating value**
DEAP	32 (16/16)	26.9	40	32	1-9
DREAMER	23 (14/9)	26.6	18	14	1-5

4.2. Results and Analysis

4.2.1. Performance Comparison

Quantitative results demonstrate FeelNet's superiority for EEG-driven emotion classification on the DEAP dataset and DREAMER dataset in **Table 2** and **Table 3**. This validates our hypothesis that hand-crafted features constrain traditional algorithms' performance. FeelNet outperforms six established deep learning benchmarks, including standard convolutional networks and Transformer-based models, while also demonstrating superior performance among lightweight models. When compared specifically to competitive lightweight architectures such as EEGNet and TSLANet, FeelNet achieves optimal results. Notably, it outperforms the strong contender TSception by a significant margin: achieving a 0.48% higher ACC on the arousal dimension and a 1.01% higher ACC on the valence dimension.

On the DREAMER dataset, **Table 3** validates the robust cross-dataset applicability of FeelNet. Despite challenges posed by the dataset's smaller scale and inherent class imbalance, our method achieves competitive performance with ACC reaching 57.59% for arousal and 54.43% for valence. FeelNet consistently outperforms comparative models in both prediction tasks. Crucially, it demonstrates significant improvements for high-recognition subjects. Nevertheless, its efficacy for subjects with low recognition remains comparatively limited in comparison with TSception. Collectively, these results suggest that larger training datasets substantially enhance model generalizability and recognition performance, particularly for challenging subject groups.

Table 2: The Trial-wise 10-fold cross-validation classification results (%) of our proposed method with SOTA methods on the DEAP dataset at 4s segment length.

Methods	**Arousal**		**Valence**	
Metric	ACC	F1	ACC	F1
SVM (Mehmood and Lee, 2015)	$60.37_{12.25}$	$57.33_{26.61}$	$55.19_{6.97}$	$57.87_{11.36}$
KNN (Mehmood and Lee, 2015)	$59.48_{12.34}$	$57.49_{24.96}$	$53.03_{9.14}$	$55.12_{16.27}$
SCN (Schirrmeister et al., 2017)	$60.68_{10.12}$	$61.04_{19.25}$	$59.43_{7.27}$	$62.59_{10.41}$
DCN (Schirrmeister et al., 2017)	$61.57_{9.47}$	$62.50_{18.71}$	$\underline{59.96}_{8.50}$	$60.92_{10.56}$
EEGNet (Lawhern et al., 2018)	$59.76_{9.34}$	$61.78_{14.93}$	$58.23_{8.38}$	$61.19_{9.31}$
ETCNet (Ingolfsson et al., 2020)	$61.70_{13.54}$	$60.65_{22.40}$	$57.78_{8.82}$	$60.85_{9.60}$
TSception (Ding et al., 2023)	$61.74_{10.35}$	$\underline{63.03}_{17.11}$	$59.69_{7.43}$	$\underline{62.69}_{9.24}$
TSception (Ding et al., 2023)($\star$)	$\underline{61.80}_{10.65}$	$62.97_{17.20}$	$59.21_{6.94}$	$62.12_{9.42}$
ERTNet (Liu et al., 2024)	$60.32_{12.19}$	$60.81_{18.61}$	$57.34_{9.63}$	$60.44_{10.33}$
TSLANet (Eldele et al., 2024)	$58.43_{11.09}$	$57.57_{21.54}$	$51.68_{6.01}$	$53.76_{13.64}$
FeelNet	$\mathbf{62.22}_{10.10}$	$\mathbf{63.19}_{18.40}$	$\mathbf{60.70}_{8.99}$	$\mathbf{64.11}_{10.59}$

Note: $\star$ refers to the channel number of TSception; without $\star$ channel number = 28, with $\star$ channel number = 32; others use 32 channels.
SCN: ShallowConvNet, DCN: DeepConvNet, ETCNet: EEG-TCNet.
Bold: best results, Underline: second best.

4.2.2. Ablation Study

Quantitative ablation studies demonstrate the critical contribution of each core component of FeelNet in **Table 4**. Specifically, removing the local rhythm filter module (w/o RSB-L) from the RSB module leads to significant performance degradation, with ACC dropping

Table 3: The Trial-wise 10-fold cross-validation classification results (%) of our proposed method with SOTA methods on the DREAMER dataset at 2s segment length.

Methods	**Arousal**		**Valence**	
Metric	ACC	F1	ACC	F1
SCN	$52.29_{8.97}$	$39.25_{20.02}$	$51.29_{11.78}$	$\mathbf{32.96}_{14.56}$
DCN	$54.43_{10.46}$	$40.69_{20.49}$	$\underline{52.83}_{12.75}$	$31.36_{15.46}$
EEGNet	$55.06_{9.11}$	$39.73_{21.42}$	$50.88_{11.42}$	$32.04_{13.28}$
ETCNet	$55.19_{11.53}$	$33.43_{26.47}$	$49.47_{10.26}$	$19.19_{17.99}$
TSception($\star$)	$56.01_{9.23}$	$\underline{40.81}_{22.71}$	$50.98_{10.33}$	$\underline{32.71}_{14.88}$
ERTNet	$\underline{56.28}_{12.66}$	$30.99_{28.57}$	$52.52_{12.18}$	$31.09_{14.52}$
TSLANet	$53.94_{7.65}$	$39.99_{19.69}$	$50.75_{9.97}$	$30.63_{13.48}$
FeelNet	$\mathbf{57.59}_{9.61}$	$\mathbf{42.29}_{21.78}$	$\mathbf{54.43}_{11.47}$	$30.42_{16.96}$

to 61.28% (arousal) and 56.90% (valence) on the DEAP dataset and 57.42% (arousal) and 54.41% (valence) on the DREAMER dataset. This result underscores the critical role of the local rhythm filter module in the task: it effectively removes extraneous noise and extracts emotionally relevant features. Removing the entire RSB module also causes a slight performance degradation. However, the degradation is less severe compared to removing the MTCB module, as we employed a depthwise separable convolution module as its replacement. This comparative analysis further validates the impact of multi-scale temporal feature extraction on classification accuracy. Additionally, we explore data segmentation lengths in **Table 5**. A 4-second length achieves optimal performance on the DEAP dataset, consistent with prior studies Ding et al. (2023). A 2-second length is chosen for the DREAMER dataset to increase the sample size and enchance training.

4.2.3. Efficiency-Performance Trade-off Analysis

We compare the complexity of FeelNet with state-of-the-art models, including ConvNet, EEGNet, TSception, EEG-TCNet, ERNet and TSLANet, based on parameters, FLOPs, Accuracy and Inference Time in **Table 6**. FeelNet achieves the highest ACC of 62.22%, surpassing the second-best model (TSception) by 0.48%. Notably, FeelNet requires only 18.99M FLOPs and 7.78K parameters, which are 13.6% and 44.1% lower than TSception (21.98M FLOPs and 13.91K parameters). Furthermore, FeelNet demonstrates a significant speed advantage with an inference time of 8.03 ms, which is 38.3% faster than TSception (13.01 ms) and nearly 9 times faster than EEGNet and EEG-TCNet (70 ms). Compared to the most efficient model TSLANet (10.51M FLOPs and 1.34K parameters), FeelNet improves accuracy by 3.79% with a moderate increase in computation and parameters, while still being faster (8.03 ms vs. 9.01 ms).

To facilitate clearer comparison of model performance, **Fig. 2** visualizes the results. Here, the x-axis denotes model parameter count, the y-axis represents prediction accuracy, and the circle size corresponds to computational cost (FLOPs). Models positioned closer to the top-left corner (indicating higher ACC with fewer parameters) and associated

Table 4: Ablation analysis of components on the DEAP and DREAMER dataset

Variant	Arousal		Valence	
	ACC	F1	ACC	F1
w/o R	61.20	61.58	60.54	63.93
w/o R(L)	61.28	61.41	56.90	60.74
w/o M	61.51	60.68	59.67	62.77
DEAP	**62.22**	**63.19**	**60.70**	**64.11**
w/o R	56.71	39.79	54.13	29.73
w/o R(L)	57.42	40.82	54.41	28.30
w/o M	56.87	37.12	53.92	19.44
DREAMER	**57.59**	**42.29**	**54.43**	**30.42**

Note: R: Rhythm Spectral Block (RSB);
R(L): local rhythm filter in RSB;
M: Multi-scale Temporal Conv Block;
w/o: Without the component.

Table 5: Ablation analysis of segment lengths in FeelNet

Length	Arousal		Valence	
	ACC	F1	ACC	F1
DEAP dataset				
1s	59.05	59.08	55.05	57.65
2s	59.00	59.29	57.18	59.75
3s	58.95	59.38	57.78	60.52
4s	**62.22**	**63.19**	**60.70**	**64.11**
5s	60.93	61.88	57.16	60.08
6s	61.13	61.72	57.50	60.05
DREAMER dataset				
1s	57.08	38.62	55.09	28.97
2s	**57.59**	**42.29**	54.43	**30.42**
3s	57.11	38.55	55.26	27.71
4s	56.92	36.81	55.39	26.39
5s	56.76	40.75	56.72	28.56
6s	57.15	40.46	**56.98**	27.78

Table 6: Model efficiency and performance comparison on EEG emotion recognition.

Model	**FLOPs** (M)	**Params** (K)	**ACC** (%)	**Inference Time** (ms)
SCNet	81.89	54.56	60.68	10.00
DCNet	32.12	153.75	61.57	**5.99**
EEGNet	17.79	2.23	59.76	70.00
ETCNet	22.33	56.85	61.70	70.00
TSception	21.98	13.91	61.74	13.01
ERTNet	17.94	3.07	60.32	**5.99**
TSLANet	**10.51**	**1.34**	58.43	9.01
FeelNet	18.99	7.78	**62.22**	8.03

with smaller circles (indicating lower FLOPs) represent more computationally efficient and high-performing architectures. The FeelNet model outperformed the lightweight TSLANet, achieving superior recognition accuracy with competitive parameter efficiency, thereby validating its high performance.

4.2.4. Noise Robustness and Interpretability

To further validate the robustness of FeelNet, we systematically injected varying levels of Gaussion noise based on the signal amplitude into the input data to evaluate its impact on recognition performance. FeelNet with the rhythm filter consistently outperforms TSception in arousal dimension, especially as noise levels exceed 20 std in **Fig. 3**. While noise levels below 20 std have minimal impact on the signal, higher levels cause a rapid decline in

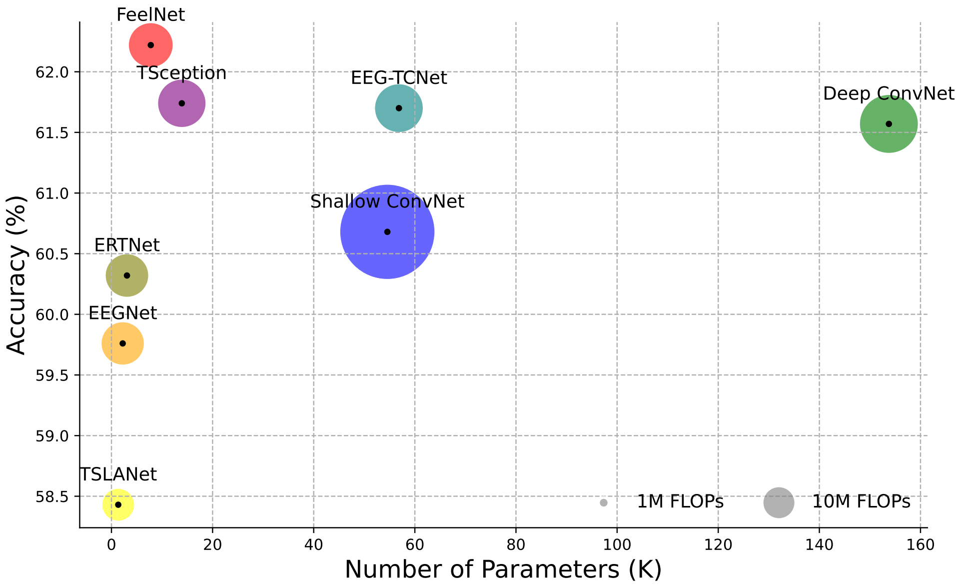

Figure 2: Performance-Parameter trade-off comparison between FeelNet and baseline models on DEAP dataset

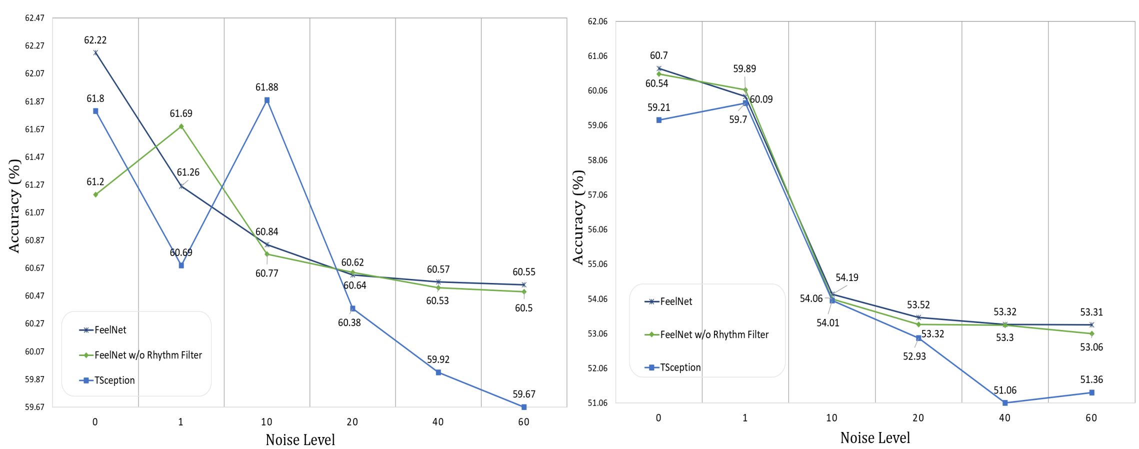

Figure 3: Robustness against noise levels in Arousal (left) and Valence (right)

TSception's performance. In contrast, FeelNet maintains stable accuracy, with the rhythm filter improving its noise resilience. Notably, all models exhibit marginal ACC gains ($0.8\% \sim 1.5\%$) with injected Gaussian noise ($\sigma < 20$). This aligns with the intrinsic signal variability ($\sigma = 18.2_{2.4}$ in raw EEG), suggesting that sub-threshold noise acts as a data augmentation mechanism that enhances feature discriminability.

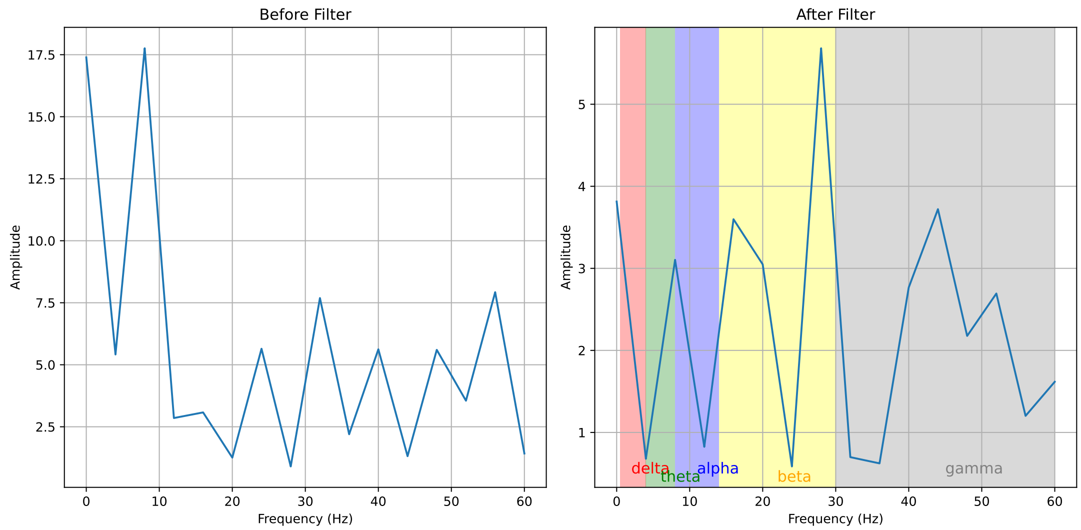

Figure 4: Feature comparison before and after rhythm filtering.

Building upon the demonstrated noise robustness, **Fig. 4** reveals that the rhythm filter effectively suppresses noise spikes while selectively amplifying spectral signatures within 20-40 Hz. Crucially, enhanced oscillatory patterns emerge in the Beta (12-30 Hz) and Gamma (30-40 Hz) bands. This spectral sharpening suggests higher frequency neural oscillations may serve as biomarkers for emotional processing, corroborating neurophysiological studies linking gamma synchrony to affective states (Zheng et al., 2019).

Saliency maps (Zhang et al., 2023) provide a readily interpretable tool for EEG-based emotion recognition, improving model optimization and classification accuracy. The saliency maps from a subset of subjects was selected to visualize their mean saliency across the time dimension in **Fig. 5**. The top three saliency maps (a)-(c) correspond to subjects with the highest F1 and ACC (top 10%), while the lower three maps (d)-(f) represent subjects with the lowest performance (bottom 10%). The results indicate that the frontal, temporal, parietal, and occipital regions of the brain hold more informational value. For high-performing subjects, channels like Fp1, Fp2, F4, O1, T8, AF3, and FC5 show higher saliency, suggesting greater emotional responses to audio-visual stimuli. In contrast, lower-performing subjects show saliency in regions such as O1, AF3, and FC5, indicating weaker model learning. This pattern aligns with previous findings Zheng et al. (2019); Gao et al. (2021).

5. Conclusion

In this paper, we propose a lightweight EEG-based emotion recognition framework named FeelNet. FeelNet's innovation lies in its key design: the Rhythm Spectral Block (RSB) for adaptive rhythm-specific spectral filtering and the Multi-scale Temporal Conv Block (MTCB) for hierarchical temporal feature extraction. Extensive evaluations on DEAP and DREAMER datasets demonstrate that FeelNet achieves state-of-the-art Accuracy while maintaining deployment-friendly efficiency. Crucially, it exhibits exceptional noise robust-

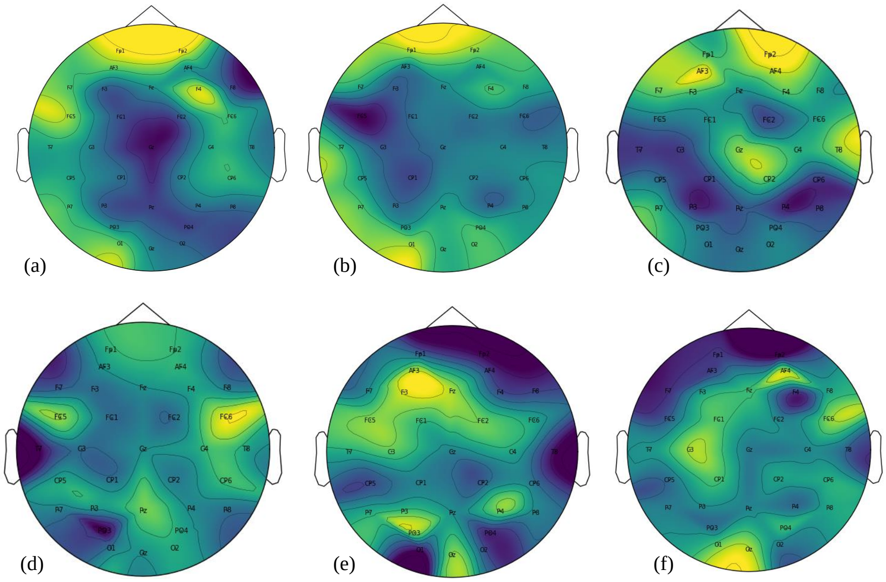

Figure 5: The following presents the mean saliency maps using the DEAP dataset. The three superior saliency maps (a) - (c) are derived from the subject pool comprising the top 10% in terms of F1 scores and Accuracy. The remaining three lower saliency maps (d) - (f) are derived from the subject pool comprising those in the lowest 10% about valence and arousal. The regions identified by the neural network as being the most informative are the frontal, temporal, and parietal regions, as well as those associated with subjects with a high F1 score.

ness and cross-dataset generalizability. In summary, the proposed FeelNet has some academic value and practical application value.

However, the fixed frequency band divisions in the current model fail to capture inter-subject variability, limiting its generalizability. Future work will focus on developing personalized frequency band learning mechanisms to address inter-subject variability and exploring cross-subject transfer learning for unseen population generalization.

Acknowledgments

This work was partially supported by the Department of Science and Technology of Sichuan Province (RZHZ2022008).

References

Gregory A Baxes. *Digital image processing: principles and applications.* John Wiley & Sons, Inc., 1994.

György Buzsáki and Andreas Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, 2004. doi: 10.1126/science.1099745. URL https://www.science.org/doi/abs/10.1126/science.1099745.

Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, July 2023. ISSN 2371-9850. doi: 10.1109/taffc.2022.3169001. URL http://dx.doi.org/10.1109/TAFFC.2022.3169001.

Yi Ding, Su Zhang, Chuangao Tang, and Cuntai Guan. Masa-tcn: Multi-anchor space-aware temporal convolutional neural networks for continuous and discrete eeg emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 28(7):3953–3964, 2024. doi: 10.1109/JBHI.2024.3392564.

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84, 2013. doi: 10.1109/NER.2013.6695876.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: Rethinking transformers for time series representation learning. *ArXiv*, abs/2404.08472, 2024. URL https://api.semanticscholar.org/CorpusID:269137300.

Yunyuan Gao, Zhen Cao, Jia Liu, and Jianhai Zhang. A novel dynamic brain network in arousal for brain states and emotion analysis. *Mathematical Biosciences and Engineering*, 18(6):7440–7463, 2021. ISSN 1551-0018. doi: 10.3934/mbe.2021368. URL https://www.aimspress.com/article/doi/10.3934/mbe.2021368.

Weichen Huang, Wenlong Wang, Yuanqing Li, and Wei Wu. Fbstcnet: A spatio-temporal convolutional network integrating power and connectivity features for eeg-based emotion decoding. *IEEE Transactions on Affective Computing*, pages 1–13, 2024. doi: 10.1109/TAFFC.2024.3385651.

Thorir Mar Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2958–2965, 2020. doi: 10.1109/SMC42975.2020.9283028.

Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 5:327–339, 2014. URL https://api.semanticscholar.org/CorpusID:10882119.

Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, and Jing Wang. Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*, page 2909–2917, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379885. doi: 10.1145/3394171.3413724. URL https://doi.org/10.1145/3394171.3413724.

Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22:98–107, 2018. URL https://api.semanticscholar.org/CorpusID:23477696.

Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, July 2018. ISSN 1741-2552. doi: 10.1088/1741-2552/aace8c. URL http://dx.doi.org/10.1088/1741-2552/aace8c.

Hailun Lian, Cheng Lu, Sunan Li, Yan Zhao, Chuangao Tang, and Yuan Zong. A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. *Entropy*, 25(10), 2023. ISSN 1099-4300. doi: 10.3390/e25101440. URL https://www.mdpi.com/1099-4300/25/10/1440.

Ruixiang Liu, Yihu Chao, Xuerui Ma, Xianzheng Sha, Limin Sun, Shuo Li, and Shijie Chang. Ertnet: an interpretable transformer-based framework for eeg emotion recognition. *Frontiers in Neuroscience*, 18, 2024. URL https://api.semanticscholar.org/CorpusID:267048723.

Cheng Lu, Wenming Zheng, Hailun Lian, Yuan Zong, Chuangao Tang, Sunan Li, and Yan Zhao. Speech emotion recognition via an attentive time–frequency neural network. *IEEE Transactions on Computational Social Systems*, 10(6):3159–3168, 2023. doi: 10.1109/TCSS.2022.3219825.

Pooja Manral and K. R. Seeja. Emotion recognition from eeg signals: A survey. In Tomonobu Senjyu, Chakchai So-In, and Amit Joshi, editors, *Smart Trends in Computing and Communications*, pages 635–651, Singapore, 2023. Springer Nature Singapore. ISBN 978-981-99-0838-7.

Raja Majid Mehmood and Hyo Jong Lee. Emotion classification of eeg brain signal using svm and knn. In *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–5, 2015. doi: 10.1109/ICMEW.2015.7169786.

Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. *Advances in neural information processing systems*, 34:980–993, 2021.

Yongming Rao, Wenliang Zhao, Zheng Zhu, Jie Zhou, and Jiwen Lu. Gfnet: Global filter networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10960–10973, 2023. doi: 10.1109/TPAMI.2023.3263824.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, August 2017. ISSN 1097-0193. doi: 10.1002/hbm.23730. URL http://dx.doi.org/10.1002/hbm.23730.

Xinke Shen, Xianggen Liu, Xin Hu, Dan Zhang, and Sen Song. Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2496–2511, 2022.

Wei Tao, Chang Li, Rencheng Song, Juan Cheng, Yu Liu, Feng Wan, and Xun Chen. Eeg-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing*, 14(1):382–393, 2023. doi: 10.1109/TAFFC.2020.3025777.

Yuki Tatsunami and Masato Taki. Fft-based dynamic token mixer for vision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15328–15336, 2024.

Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing Systems*, 2017. URL https://api.semanticscholar.org/CorpusID:13756489.

Francis R Willett, Donald T Avansino, Leigh R Hochberg, Jaimie M Henderson, and Krishna V Shenoy. High-performance brain-to-text communication via handwriting. *Nature*, 593(7858):249–254, 2021.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *CoRR*, abs/1511.07122, 2015. URL https://api.semanticscholar.org/CorpusID:17127188.

Ying Zhang, Chen Cui, and Shenghua Zhong. Eeg-based emotion recognition via knowledge-integrated interpretable method. *Mathematics*, 11(6), 2023. ISSN 2227-7390. doi: 10.3390/math11061424. URL https://www.mdpi.com/2227-7390/11/6/1424.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 10(3):417–429, 2019. doi: 10.1109/TAFFC.2017.2712143.