TFSNet: EEG-based Emotion Recognition using Temporal and Frequency-Spatial Feature

Yeryeong Lee and Hyeryung Jang
Department of Artificial Intelligence Convergence, Computer Science & Artificial Intelligence
Dongguk University, Seoul, Korea
Emails: {yrlee040212, hyeryung.jang}@dgu.ac.kr

***Abstract*—Electroencephalography (EEG)-based Automatic Emotion Recognition (AER) has gained increasing attention as a reliable tool for affective computing. While prior studies have explored various temporal, frequency, and spatial-domain representations of EEG signals, few have effectively integrated these domains within a unified framework. In this paper, we propose TFSNet, a multi-domain deep learning model that combines temporal, frequency, and spatial features for robust emotion recognition. TFSNet consists of a dual-encoder architecture: a temporal encoder based on state-space modeling (S4D), and a frequency-spatial encoder that leverages CNNs, attention mechanisms, and graph filtering using a physiologically-informed adjacency matrix. These domain-specific embeddings are fused and passed through a classifier for final prediction. Experimental results on the DREAMER dataset demonstrate that TFSNet achieves superior performance across Valence, Arousal, and Dominance emotions, outperforming state-of-the-art models. The results highlight the effectiveness of combining domain-aware representations and spatial connectivity priors for EEG-based emotion recognition and its potential for real-time applications.**

I. Introduction

Emotions are fundamental psychological states that shape human cognition, decision-making, behavior, and social interactions [1]. With the rapid advancement of artificial intelligence (AI) and Human-Computer Interaction (HCI), Automatic Emotion Recognition (AER) has emerged as a promising technology that enables machines to understand and respond to human emotional states, opening up applications in emotion monitoring [2], [3], user-centered customized services [3], smart healthcare [2], and Brain-Computer Interface (BCI) [4].

Among physiological signals used for AER, such as electrocardiography (ECG) and electromyography (EMG), electroencephalography (EEG) holds distinct advantages [5]. Unlike peripheral signals, EEG captures direct, real-time electrical activity in the central nervous system, offering high temporal resolution and the ability to noninvasively monitor internal cognitive and emotional processes [2]. As a result, EEG-based AER has gained growing attention for its potential to reveal rich insights into human emotional dynamics.

Most prior EEG-based AER studies have relied on medical-grade datasets like DEAP [6] or SEED [7], collected under controlled laboratory conditions using expensive equipment. While these datasets ensure data precision, they limit the scalability and practical deployment of AER systems in real-world, non-medical settings. Addressing this challenge, we leverage the DREAMER [8] dataset, recorded using the affordable off-the-shelf Emotiv EPOC device, to explore robust AER under practical conditions. The EEG signals in DREAMER dataset are collected using wireless equipment and portable medical-grade devices, yet it yields comparable AER results comparable to those from other datasets from expensive medical equipment. In particular, the dataset includes 14-channel EEG recordings from 23 participants exposed to 18 audiovisual stimuli at 128 Hz, with self-reported 5-point ratings on Valence, Arousal, and Dominance dimensions.

Traditional AER approaches primarily employed hand-crafted feature extraction from physiological signals, including EEG, in the time or frequency domain, followed by machine learning classifiers like SVM or KNN [9]. However, due to the nonlinear and nonstationary nature of EEG, such methods often struggle to capture the complex patterns underlying emotional states. Deep learning methods, including convolutional [10] and recurrent neural networks [11], have shown promise by learning latent representations directly from raw data. Recent studies have begun to combine multiple domains - time, frequency, and/or spatial features - to further enhance model performance, but most existing studies still focus on single-domain representations.

In this work, we propose **TFSNet**, a novel multi-domain EEG-based emotion recognition framework that integrates *temporal and frequency-spatial* information. Specifically, TFSNet combines an S4D-based [12] temporal encoder for capturing time-domain patterns and a graph filtering-based [13] frequency-spatial encoder that leverages differential entropy and physical connectivity between EEG channels. By fusing these complementary representations, the model aims to robustly learn emotional dynamics across domains. The main contributions of this work are summarized as:

- We design a multi-domain EEG emotion recognition model combining an S4D Temporal encoder and Frequency-Spatial encoder with graph filtering.
- We integrate temporal and frequency-spatial embeddings for final emotion classification, validated on the DREAMER dataset.
- We demonstrate superior performance over state-of-the-art model [10], achieving improvements of 1.63%, 1.32%, and 1.34% in Valence, Arousal, and Dominance, respectively.

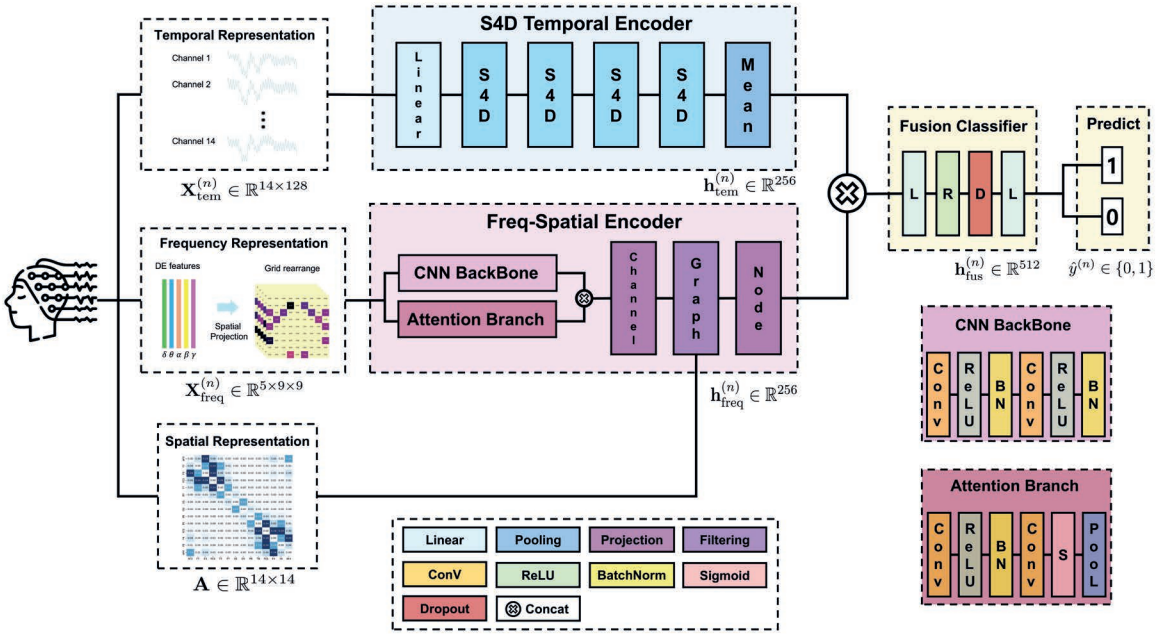

Fig. 1: Architecture of the proposed TFSNet. Temporal, frequency, and spatial EEG features are encoded via dual encoders (S4D and frequency-spatial encoders), fused, and classified for emotion prediction.

II. Related Work

Early EEG-based AER methods relied on manually extracting statistical or spectral features from time-domain (TD) or frequency-domain (FD) signals, followed by classification using traditional machine learning techniques such as support vector machines (SVM) [8], k-nearest neighbors (KNN) [9], decision trees [14], or Naive Bayes [14]. However, due to the susceptibility of EEG signals to noise and their nonlinear, nonstationary nature, these approaches often fell short in capturing the complexity of emotional states.

To address these limitations, deep learning models have been introduced, enabling automatic extraction of hierarchical and multi-domain features from EEG signals. Recent methods combine TD, FD, and time-frequency (TF) features to better capture the diverse neural signatures of emotion. TD features reflect temporal dynamics in brain [2], FD features capture spectral energy distribution (e.g., via power spectral density or differential entropy) that is converted from TD features into spectral space [2], [15], and TF features represent nonstationary patterns through techniques like wavelet transforms [16]. More recently, spatial-domain (SD) information, which models inter-channel structural relationships, has gained attention for improving recognition performance when combined with other domains [2].

Several notable models have advanced this field. For example, [8] constructs the DREAMER dataset and employs an SVM classifier with a TD-based RBF kernel. FLDNet [17] uses attention-based triple networks to integrate TD and SD features from multi-channel 3-second EEG segments. 3DFR [18] constructs three-dimensional representations combining TD and FD features to capture cross-channel and cross-frequency interactions, while LEDPatNet19 [19] applies frequency subband decomposition and fuses original and subband-specific features for each EEG channel. MTCA-CapsNet [20] integrates multi-task learning and capsule networks with channel attention mechanisms, focusing on TD features. ECNN-C [10] leverages supervised contrastive learning with efficient TD-based convolutional blocks, and GECNN [13] fuses FD and TF features, extracting local and global spatial features using CNNs and graph convolutional networks.

Despite these advances, most models underexplore the physical connectivity between EEG channels, an important consideration, as emotional processing emerges from distributed brain regions. While some methods include spatial relations, they often simplify adjacency matrices of EEG channels as binary connections, overlooking meaningful strengths of physical connectivity.

III. Method

A. Overview

In this paper, we propose **TFSNet**, a multi-domain EEG-based emotion recognition model that explicitly incorporates *temporal, frequency, and spatial* characteristics of EEG signals and integrates them for emotion classification. TFSNet consists of three main steps (see Fig 1): *(i)* extraction of temporal, frequency, and spatial representations from the input EEG signals; *(ii)* independent processing through two encoders - an S4D-based temporal encoder and a frequency-spatial encoder combining CNN, attention, and graph filtering; and *(iii)* fusion

of high-dimensional embeddings from both encoders, followed by a binary classifier trained with binary cross-entropy loss.

By combining complementary domain-specific representations, TFSNet robustly learns the complex emotional dynamics present in EEG signals. Notably, the frequency-spatial encoder integrates a graph-based adjacency matrix reflecting inter-channel connection strengths, computed via exponential distance-based weighting, which can also be set as trainable for further improvement. This is a key novelty, as most prior works either overlook this physiological aspect or simplify it using binary connections.

B. Data Preprocessing and Representation Extraction

We use the DREAMER dataset, which provides EEG and ECG signals; this study focuses solely on EEG. EEG channels were arranged according to the international $10-20$ system, and continuous EEG signals $\mathbf{X}$ were segmented into learnable units using a sliding window of $L = 128$ points (corresponding to 1 second). Each n-th segment $\mathbf{X}^{(n)}$ consists of $C = 14$ channels, represented as $\mathbf{X}^{(n)} = [\mathbf{X}_1^{(n)}, \mathbf{X}_2^{(n)}, \ldots, \mathbf{X}_C^{(n)}]^\top \in \mathbb{R}^{C\times L}$. To construct multi-domain representations, the EEG data is preprocessed and transformed in the following three ways.

Temporal representation. EEG signals reflect dynamic temporal patterns associated with emotional states [2]. To preserve these dynamics and ensure learning stability, we apply channel-wise Z-score normalization as follows:

$$\bar{x}_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i + \epsilon}, \tag{1}$$

where $x_i(t)$ is the raw signal at time t for channel i; μ_i and σ_i are the mean and standard deviation across the time series of channel i; and ϵ is a small constant to prevent division by zero. The normalized segment $\bar{\mathbf{X}}_i^{(n)}$ is then used to construct the temporal representation $\mathbf{X}_{\text{tem}}^{(n)}$ of the n-th segment:

$$\mathbf{X}_{\text{tem}}^{(n)} = \left[\bar{\mathbf{X}}_1^{(n)}(t), \bar{\mathbf{X}}_2^{(n)}(t), \ldots, \bar{\mathbf{X}}_C^{(n)}(t)\right]^\top. \tag{2}$$

We note that each segment-level temporal representation $\mathbf{X}_{\text{tem}}^{(n)}$, in the form of $\mathbb{R}^{C\times L}$, is used as input to the temporal encoder.

Frequency representation. To capture spectral characteristics, the normalized temporal representation $\mathbf{X}_{\text{tem}}^{(n)}$ is transformed into the frequency domain across five standard EEG bands [2]: δ ($1-4$ Hz), θ ($4-8$ Hz), α ($8-13$ Hz), β ($13-30$ Hz), and γ ($30-45$ Hz). For each channel i, differential entropy (DE) values are computed and arranged into 9×9 grids $\mathbf{G}^{(b)} \in \mathbb{R}^{9\times 9}$, with $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$ being EEG band, aligned with the physical electrode layout. This results in a 3D tensor per segment:

$$\mathbf{X}_{\text{freq}}^{(n)} = \left[\mathbf{G}^{(\delta)}, \mathbf{G}^{(\theta)}, \mathbf{G}^{(\alpha)}, \mathbf{G}^{(\beta)}, \mathbf{G}^{(\gamma)}\right]^\top \in \mathbb{R}^{5\times 9\times 9}. \tag{3}$$

The frequency representation $\mathbf{X}_{\text{freq}}^{(n)}$ of the n-th segment serves as input to the frequency-spatial encoder.

Spatial representation. EEG signals are recorded from multiple brain channels, each having spatial properties determined by their anatomical locations and inter-regional connections.

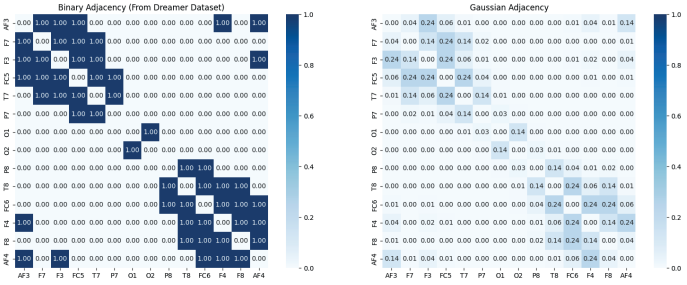

Fig. 2: Comparison of spatial adjacency matrices $\mathbf{A}$ used in the frequency-spatial encoder. Binary (left) and Gaussian (right) adjacency matrices for EEG channels, obtained directly from the DREAMER dataset and reflecting distance-weighted spatial connectivity, respectively.

Spatial feature analysis is thus crucial for understanding how regional interactions influence emotions [2]. To model spatial relationships, we construct an adjacency matrix $\mathbf{A} \in \mathbb{R}^{C\times C}$ using the physical electrode coordinates (x_i, y_i) of each channel i. The Euclidean distance d_{ij} between channels i and j is computed, and the corresponding connection strength A_{ij} is defined using an exponential distance-based weighting as follows:

$$A_{ij} = \begin{cases} \exp(-d_{ij}), & \text{if } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

This adjacency matrix captures spatial continuity, assigning higher weights to closer channels, and is later used in the graph filtering module of the frequency-spatial encoder. Fig. 2 illustrates an example of a binary adjacency matrix from the DREAMER dataset and our spatial representation matrix $\mathbf{A}$ obtained from (4).

Label assignment. Finally, for binary emotion classification, self-reported emotion ratings below 3 are assigned a label of 0, while ratings above 3 are assigned a label 1. All domain-specific representations of the n-th segment $\left(\mathbf{X}_{\text{tem}}^{(n)}, \mathbf{X}_{\text{freq}}^{(n)}\right)$ are linked to the same binary label $y^{(n)} \in \{0, 1\}$.

C. Domain-specific Encoders

S4D temporal encoder. EEG signals exhibit time-varying patterns that reflect changes in emotional states. To model these temporal dynamics effectively, we adopt a time-series encoder based on the State Space Model (SSM) [21], formulated as:

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \tag{5}$$

where $u(t)$ is the input, $x(t)$ is the latent state, $y(t)$ is the output, and A, B, C, D are learnable parameter matrices. Specifically, we employ the S4D architecture [12], which simplifies A to a diagonal matrix, improving modeling efficiency for long sequences. As shown in Fig. 1, the input temporal segment $\mathbf{X}_{\text{tem}}^{(n)} \in \mathbb{R}^{C\times L}$ in (2) is first projected linearly to a latent space of dimension d_{model}, passed through four stacked S4D layers, and then aggregated using average pooling to obtain a fixed-length embedding $\mathbf{h}_{\text{tem}}^{(n)} \in \mathbb{R}^{d_{\text{model}}}$.

Frequency-Spatial encoder. Given that EEG channels have physically structured scalp positions and contain distinct frequency band patterns, we design a frequency-spatial encoder to jointly model these aspects. In particular, the input frequency segment $\mathbf{X}_{\text{freq}}^{(n)} \in \mathbb{R}^{5\times9\times9}$ in (III-B) is processed by a CNN-based backbone, complemented by an attention module that highlights salient regions. The extracted features are fused via element-wise multiplication, flattened, and projected onto $C = 14$ EEG channel nodes, resulting in node-wise feature vectors of dimension $d = 16$. We then apply *graph filtering* [22] using the normalized adjacency matrix $\hat{\mathbf{A}}$, defined as:

$$\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}, \quad \mathbf{H} = \hat{\mathbf{A}} \cdot \mathbf{X}, \tag{6}$$

where $\mathbf{D}$ is the degree matrix, and $\mathbf{X} \in \mathbb{R}^{b\times C\times d}$ is the CNN feature tensor, where $b = 5$ (number of frequency bands), $C = 14$ (number of EEG channels), and $d = 16$ (node feature dimension). The filtered node features are then projected into high-dimensional embeddings (d_{model}) via a fully connected layer and averaged across nodes to yield the final frequency–spatial embedding $\mathbf{h}_{\text{freq}}^{(n)} \in \mathbb{R}^{d_{\text{model}}}$. See Fig. 1 for details.

Fusion classifier. After obtaining the temporal $\mathbf{h}_{\text{tem}}^{(n)}$ and frequency-spatial $\mathbf{h}_{\text{freq}}^{(n)}$ embeddings from two encoders, they are concatenated:

$$\mathbf{h}_{\text{fus}}^{(n)} = \left[\mathbf{h}_{\text{tem}}^{(n)}; \mathbf{h}_{\text{freq}}^{(n)}\right] \in \mathbb{R}^{2d_{\text{model}}}. \tag{7}$$

This fused vector $\mathbf{h}_{\text{fus}}^{(n)}$ of the n-th segment is fed into a classifier $f : \mathbb{R}^{2d_{\text{model}}} \to \mathbb{R}$ to predict the emotion logit $\hat{y}^{(n)} = f(\mathbf{h}_{\text{fus}}^{(n)})$. This parallel encoding-based fusion structure of TFSNet allows for preserving domain-specific features while leveraging complementary temporal and frequency-spatial information, enhancing robustness in emotion recognition.

D. Training TFSNet

Our model TFSNet is trained for three independent binary emotion classifications across the Valence, Arousal, and Dominance. The model is implemented with a latent embedding of size $d_{\text{model}} = 256$. The adjacency matrix $\mathbf{A}$ used in the frequency-spatial encoder is either fixed (pre-computed based on the physical distances of EEG channels) or set as a learnable parameter, allowing the model to adapt to a more meaningful graph structure during training. For each emotion dimension, binary labels $y^{(n)} \in 0, 1$ and predicted logits $\hat{y}^{(n)}$ are used to compute the Binary Cross-Entropy (BCE) loss:

$$\begin{aligned}\mathcal{L}_{\text{BCE}}(y^{(n)}, \hat{y}^{(n)}) = &-y^{(n)} \log(\sigma(\hat{y}^{(n)})) \\ &- (1 - y^{(n)}) \log(1 - \sigma(\hat{y}^{(n)})),\end{aligned} \tag{8}$$

where $\sigma(\cdot)$ is the sigmoid function.

IV. Experimental Results

A. Experimental setup

Training configurations. To assess our model TFSNet, we train the model with DREAMER dataset, using subject-dependent 10-fold cross-validation, accounting for inter-subject variability in EEG signals. For each subject, average performance is computed across folds, and final results are averaged across all subjects. For training, the AdamW optimizer is used with a learning rate and weight decay of 1×10^{-4}, a batch size of 64, and 50 training epochs.

Evaluation metrics. We primarily use accuracy to evaluate model performance on AER, alongside precision, recall, specificity, and F1-score for a comprehensive assessment. These are computed as:

$$\begin{aligned}\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN}, \quad \text{P} = \frac{TP}{TP+FP}, \\ \text{R} = \frac{TP}{TP+FN}, \quad \text{Spec} = \frac{TN}{TN+FP}, \quad \text{F1} = \frac{2\cdot\text{P}\cdot\text{R}}{\text{P}+\text{R}},\end{aligned} \tag{9}$$

where TP, TN, FP and FN represent true positives, true negatives, false positives, and false negatives, respectively. Given the clinical nature of AER, metrics such as recall and F1-score are particularly critical, as failing to correctly detect positive cases may lead to adverse medical consequences.

Model Variant	Valence	Arousal	Dominance
S4D Temporal encoder	86.30	87.88	86.84
Freq-Spatial encoder (fixed $\mathbf{A}$)	98.57	98.11	98.28
Freq-Spatial encoder (learnable $\mathbf{A}$)	98.56	98.10	98.30
TFSNet (fixed $\mathbf{A}$)	**98.61**	98.14	98.32
TFSNet (learnable $\mathbf{A}$)	98.59	**98.17**	**98.34**

TABLE I: Ablation study TFSNet (10-fold average accuracy). The full model with both encoders consistently outperforms single-domain variants. Learnable adjacency yields marginal gains over fixed adjacency.

B. Ablation Study

In Table I, we report the 10-fold average accuracy across Valence, Arousal, and Dominance dimensions for different module configurations. TFSNet, which integrates the S4D temporal encoder and frequency-spatial encoder in parallel, consistently achieved the highest performance across all emotions. While the S4D temporal encoder alone shows relatively lower accuracy (Valence: 86.30%, Arousal: 87.88%, Dominance: 86.84%), combining it with the frequency-spatial encoder boosted accuracy beyond 98%. Since our experiments are conducted within the subject-dependent, 10-fold cross-validation settings, where only a few segments are available for training, S4D temporal encoder struggles from insufficient temporal patterns, resulting in poor performance compared to the frequency-spatial encoder that uses global information. The results obtained in our experiments suggest that short-length sequences alone may be unable to capture the complex temporal changes in emotional states, yet TFSNet trained with both encoders still improves performance.

Table II summarizes the quantitative performance of TFSNet with fixed and learnable adjacency matrices $\mathbf{A}$, respectively. Both configurations achieve consistent and robust

Metric	Fixed $\mathbf{A}$			Learnable $\mathbf{A}$		
	Valence	Arousal	Dominance	Valence	Arousal	Dominance
Accuracy	**98.61 ± 1.32**	98.14 ± 1.87	98.32 ± 1.74	98.59 ± 1.35	**98.17 ± 1.84**	**98.34 ± 1.69**
Precision	**98.48 ± 1.98**	97.96 ± 2.74	98.12 ± 2.36	98.45 ± 2.00	**97.97 ± 2.74**	**98.15 ± 2.28**
Recall	**98.10 ± 2.38**	97.85 ± 2.70	98.68 ± 1.58	98.06 ± 2.48	**97.92 ± 2.64**	**98.71 ± 1.62**
Specificity	**98.92 ± 1.43**	98.08 ± 2.41	97.59 ± 3.35	98.90 ± 1.46	**98.11 ± 2.33**	**97.60 ± 3.21**
F1-Score	**98.27 ± 1.72**	97.88 ± 2.20	98.38 ± 1.56	98.23 ± 1.77	**97.92 ± 2.18**	**98.41 ± 1.53**

TABLE II: Performance (accuracy) comparison of TFSNet with fixed and learnable adjacency matrices. Learnable adjacency (right) consistently outperforms fixed adjacency (left) across multiple metrics. Bold numbers indicate the better value for each metric.

Model	Representations			Accuracy		
	Time	**Freq.**	**Spatial**	**Valence**	**Arousal**	**Dominance**
SVM [8]	✓			62.49	62.17	61.84
FLDNET [23]	✓			89.91	87.67	90.28
3DFR [18]		✓		93.15	91.30	92.04
LEDPatNet19 [19]		✓		94.44	94.58	92.86
MTCA-CapsNet [20]	✓			95.54	94.96	95.52
ECNN-C [10]	✓			97.03	96.89	97.04
TFSNet (fixed A)	✓	✓	✓	**98.61**	<u>98.14</u>	<u>98.32</u>
TFSNet (learnable A)	✓	✓	✓	<u>98.59</u>	**98.17**	**98.34**

TABLE III: Evaluation of single- and multi-domain emotion recognition models on the DREAMER dataset (average accuracy). TFSNet variants outperform prior methods across all emotion types. Best results are in **bold**, second-best are <u>underlined</u>.

classification across all metrics, with mean accuracies above 98% for all emotion classes. Notably, precision, recall, specificity, and F1-scores remain high even under class imbalance, confirming the robustness of our model TFSNet. The minor performance difference between using fixed versus learnable adjacency matrices $\mathbf{A}$ in (4) suggests that initial physical connectivity already encodes meaningful domain knowledge, but additional learning of adjacency weights brings marginal gains in emotion recognition from EEG signals.

C. Comparison with Baselines

We compare TFSNet against several state-of-the-art models under the same experimental conditions, i.e., subject-dependent 10-fold cross-validation, 1-second sliding window, binary emotion classification. As shown in Table III, TFSNet outperforms both traditional machine learning models (e.g., SVM [8]) and recent deep learning approaches (e.g., FLDNet [23], 3DFR [18], LEDPatNet19 [19], MTCA-CapsNet [20], and ECNN-C [10]), achieving the highest accuracy across all emotion classes (Valence: 98.61%, Arousal: 98.17%, Dominance: 98.34%). In particular, TFSNet shows over 5% improvement compared to SVM and substantial gains over multi-domain models like 3DFR, demonstrating the strength of integrating temporal, frequency, and spatial features with graph-based connectivity modeling. This superior performance stems not merely from structural complexity but from TFSNet's ability to comprehensively analyze the multifaceted characteristics of emotional states. Unlike models that focus solely on a single aspect of EEG signals, this approach integrates information from three mutually complementary domains: temporal, frequency, and spatial.

V. Conclusion

This paper introduces TFSNet, a dual-encoder neural architecture that integrates temporal, frequency, and spatial information for EEG-based emotion recognition. The model leverages a state-space (S4D) temporal encoder and a frequency–spatial encoder that includes CNNs, attention mechanisms, and graph filtering based on learnable inter-channel distances. Experimental results on the DREAMER dataset demonstrated that TFSNet consistently outperforms existing approaches across three emotional dimensions. The results validate the benefit of combining complementary domain features and incorporating physiologically-grounded spatial priors for the precision of EEG-based emotion recognition. Future work includes extending this framework to subject-independent settings and exploring its ability to generalize across larger EEG emotion datasets.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2020-II201789), and the Artificial Intelligence Convergence Innovation Human Resources Development(IITP-2025-RS-2023-00254592) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Elizabeth A Phelps. Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol.*, 57(1):27–53, 2006.

[2] Yiming Wang, Bin Zhang, and Lamei Di. Research progress of eeg-based emotion recognition: a survey. *ACM Computing Surveys*, 56(11):1–49, 2024.

[3] Runfang Guo, Hongfei Guo, Liwen Wang, Mengmeng Chen, Dong Yang, and Bin Li. Development and application of emotion recognition technology—a systematic literature review. *BMC psychology*, 12(1):95, 2024.

[4] Essam H Houssein, Asmaa Hammad, and Abdelmgeid A Ali. Human emotion recognition from eeg-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications*, 34(15):12527–12557, 2022.

[5] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7):2074, 2018.

[6] 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.

[7] 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.

[8] 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(1):98–107, 2018.

[9] Omid Bazgir, Zeynab Mohammadi, and Seyed Amir Hassan Habibi. Emotion recognition with machine learning using eeg signals. In *2018 25th national and 3rd international iranian conference on biomedical engineering (ICBME)*, pages 1–5. IEEE, 2018.

[10] Chang Li, Xuejuan Lin, Yu Liu, Rencheng Song, Juan Cheng, and Xun Chen. Eeg-based emotion recognition via efficient convolutional neural network and contrastive learning. *IEEE Sensors Journal*, 22(20):19608–19619, 2022.

[11] Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi. Emotion recognition based on eeg using lstm recurrent neural network. *International Journal of Advanced Computer Science and Applications*, 8(10), 2017.

[12] Albert Gu, Karan Goel, Ankit Gupta, and Christopher Ré. On the parameterization and initialization of diagonal state space models. *Advances in Neural Information Processing Systems*, 35:35971–35983, 2022.

[13] Tengfei Song, Wenming Zheng, Suyuan Liu, Yuan Zong, Zhen Cui, and Yang Li. Graph-embedded convolutional neural network for image-based eeg emotion recognition. *IEEE Transactions on Emerging Topics in Computing*, 10(3):1399–1413, 2021.

[14] Didar Dadebayev, Wei Wei Goh, and Ee Xion Tan. Eeg-based emotion recognition: Review of commercial eeg devices and machine learning techniques. *Journal of King Saud University-Computer and Information Sciences*, 34(7):4385–4401, 2022.

[15] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Eeg-based emotion recognition using frequency domain features and support vector machines. In *International conference on neural information processing*, pages 734–743. Springer, 2011.

[16] Raveendrababu Vempati and Lakhan Dev Sharma. A systematic review on automated human emotion recognition using electroencephalogram signals and artificial intelligence. *Results in Engineering*, 18:101027, 2023.

[17] Zhe Wang, Tianhao Gu, Yiwen Zhu, Dongdong Li, Hai Yang, and Wenli Du. Fldnet: Frame-level distilling neural network for eeg emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 25(7):2533–2544, 2021.

[18] Dongdong Li, Bing Chai, Zhe Wang, Hai Yang, and Wenli Du. Eeg emotion recognition based on 3-d feature representation and dilated fully convolutional networks. *IEEE Transactions on Cognitive and Developmental Systems*, 13(4):885–897, 2021.

[19] Turker Tuncer, Sengul Dogan, and Abdulhamit Subasi. Ledpatnet19: Automated emotion recognition model based on nonlinear led pattern feature extraction function using eeg signals. *Cognitive Neurodynamics*, 16(4):779–790, 2022.

[20] Chang Li, Bin Wang, Silin Zhang, Yu Liu, Rencheng Song, Juan Cheng, and Xun Chen. Emotion recognition from eeg based on multi-task learning with capsule network and attention mechanism. *Computers in biology and medicine*, 143:105303, 2022.

[21] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

[22] Siheng Chen, Aliaksei Sandryhaila, José MF Moura, and Jelena Kovacevic. Signal denoising on graphs via graph filtering. In *2014 ieee global conference on signal and information processing (globalsip)*, pages 872–876. IEEE, 2014.

[23] Shuaiqi Liu, Yingying Zhao, Yanling An, Jie Zhao, Shui-Hua Wang, and Jingwen Yan. Glfanet: A global to local feature aggregation network for eeg emotion recognition. *Biomedical Signal Processing and Control*, 85:104799, 2023.