TBEEG: A Two-Branch Manifold Domain Enhanced Transformer Algorithm for Learning EEG Decoding

Yanjun Qin, Wenqi Zhang, and Xiaoming Tao, *Member, IEEE*

***Abstract*— The electroencephalogram-based (EEG) brain-computer interface (BCI) has garnered significant attention in recent research. However, the practicality of EEG remains constrained by the lack of efficient EEG decoding technology. The challenge lies in effectively translating intricate EEG into meaningful, generalizable information. EEG signal decoding primarily relies on either time domain or frequency domain information. There lacks a method capable of simultaneously and effectively extracting both time and frequency domain features, as well as efficiently fuse these features. Addressing these limitations, a two-branch Manifold Domain enhanced transformer algorithm is designed to holistically capture EEG's spatio-temporal information. Our method projects the time-domain information of EEG signals into the Riemannian spaces to fully decode the time dependence of EEG signals. Using wavelet transform, the time domain information is converted into frequency domain information, and the spatial information contained in the frequency domain information of EEG signal is mined through the spectrogram. The effectiveness of the proposed TBEEG algorithm is validated on BCIC-IV-2a dataset and MAMEM-SSVEP-II datasets.**

***Index Terms*— EEG decoding, EEG signals, Riemannian spaces, spatial-temporal information.**

I. Introduction

ELECTROENCEPHALOGRAPHY (EEG) offers a non-invasive means to access neuronal dynamics with millisecond resolution, making it highly promising for brain-computer interfaces (BCIs). The field of EEG decoding seeks to uncover the perceptual, semantic, and cognitive information embedded within non-invasive measurements of brain activity [1]. As a result, the advancement of EEG decoding techniques have stimulated the emergence of representation learning methods applicable to various healthcare applications. Currently, the research on EEG decoding focuses primarily on time domain and frequency domain decoding.

There are two types of methods for using temporal domain information. One is based on manual feature extraction; the other is to use a deep learning algorithm to derive decoding features contained in temporal domain information. In traditional machine learning algorithms, feature information is derived from prior knowledge, lacking the capability to extract deep features. The temporal domain features incorporates various statistical measures, such as mean, extreme values, variance, skewness, and kurtosis [4], [5], [6], [7]. In the process of extracting features, it is possible for valuable information contained within the EEG signals to be inadvertently discarded or for complications such as excessive data smoothing to arise [9]. Traditional algorithms such as support vector machine (SVM) [10], random forests [11], Bayesian classifiers [12], LDA [13] were used in EEG Decoding studies. Deep learning algorithms [18], [19] have been widely used in EEG-based BCI and related domains [20]. The Hierarchical Convolution Neural Networks (HCNN) [21] is utilized for EEG decoding, where the calculation of Differential Entropy (DE) serves as the feature extraction method at specific time intervals for each channel. The LSTM-based (Long Short-Term Memory) is proposed to capture temporal features for EEG classification of drowsiness levels [22] and emotion recognition [23]. Fu et al. [24] use bidirectional recurrent neural network (BiRNN) encoding and decoding. MS-AMF [27] is proposed as a multi-scale fusion convolution neural network based on the attention mechanism. Attention-based algorithms demand extensive fine-tuning and heuristic adjustments, which may exacerbate the risk of overfitting.

In the domain of frequency analysis [32], [33], Singh and Malhotra [6] proposes a spectral feature-based two-layer LSTM network model for automatic prediction of epileptic seizures using long-term multichannel EEG signals. Chikhi et al. [34] conduct a analysis of spectral power in brain frequencies, specifically theta (4-8 Hz), alpha (8-12 Hz), and beta (12-30 Hz), to assess and monitor cognitive workload (CWL) during tasks. Zulliger et al. [36] present using EEG spectral power map correlations between local patterns of microstate feature activity and microstate classes. Férat et al. [38] study the spatio-temporal dynamics of EEG frequency bands.

Manuscript received 16 November 2023; revised 15 March 2024; accepted 17 March 2024. Date of publication 25 March 2024; date of current version 4 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant NSFC62306164, Grant NSFC62227801, and Grant 61925105; and in part by the Xplore Prize 2021 in information and electronics technologies. *(Corresponding author: Xiaoming Tao.)*

The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: qinyanjun@mail.tsinghua.edu.cn; zwq22@mails.tsinghua.edu.cn; taoxm@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3380595

Several other algorithms have been employed to explore the study of EEG decoding. A graph convolution neural network (M-GCN) proposed by [39], involves temporal-frequency processing performed through modified S-transform (MST). Sorkhi et al. [40] presented a Multi-scale FBCSP (MSF-BCSP) method and spatial patterns from multi-scaled data in different frequency bands that are learnt, and then, the temporal and frequency band information from projected signals is extracted. Reference [41] proposed a novel model called Hybrid-Scale Spatial-Temporal Dilated Convolution Network (HS-STDCN) for EEG-based imagined speech recognition. They used hybrid-scale rather than single-scale temporal filters on the input EEG data to learn the temporal frequency information at different levels. Reference [42] proposed an ensemble learning method based on temporal and spatial features and multi-scale filter banks which called TSMFBEL.

To address the limitations of the aforementioned methods in EEG signal processing and analysis, we introduce the TBEEG algorithm. TBEEG is a two-branch Manifold Domain enhanced transformer algorithm that effectively integrates information from both the time domain and frequency domain. The main contributions of this paper are summarized as follows:

1): We present TBEEG, an EEG decoding algorithm which leveraging a two-branch Manifold Domain enhanced transformer algorithm, and that integrates multi-dimensional features. TBEEG captures both temporal and frequency features simultaneously. In our approach, the temporal domain data is fed into an RNN-based transformer to capture temporal features, while the frequency domain data is processed by a vision-transformer to extract relevant features. These extracted features are then projected into a manifold domain to achieve enhanced robustness. Finally, an attention mechanism is employed to fuse the features effectively.

2): In particular, the frequency domain features undergo a transformation into spectrograms using Fast Fourier Transform (FFT) to enhance data stability. Subsequently, we apply the diffusion maps algorithm to perform nonlinear dimensional reduction on the EEG data, thereby representing the sampled data within an underlying manifold. This utilization of the diffusion mapping algorithm contributes to a reduction in computational complexity to a certain extent. By integrating local similarities across various scales, diffusion maps offer a comprehensive overview of the dataset, endowing the time and frequency domains with robust features.

3): By combining the strengths of both temporal and frequency domain features and utilizing advanced transformer-based architectures, TBEEG showcases promising results in EEG decoding tasks. The integration of multi-dimensional features and the attention-based fusion of information contribute to the algorithm's effectiveness.

II. Methods

A. Overview

Figure 1 illustrates the primary framework of our proposed TBEEG algorithm which is designed for general EEG decoding. To effectively address the complexities inherent in informative EEG features and account for the non-stationarity of brain dynamics, our approach focuses on exploring the data from two distinct angles: the temporal domain and the frequency domain. The objective is to capture and characterize the spatio-temporal representations of EEG data.

Initially, we transform the EEG data into the temporal-frequency domain to gain insights into its temporal and spectral components. For the temporal domain data, we leverage both RNN and transformer variant models to extract its features. Furthermore, we employ a diffusion projection algorithm to explore the interrelationships among temporal data points, enhancing the robustness of these features. To effectively incorporate the temporal information, we utilize a variant of the attention algorithm to assign weights and facilitate fusion. The frequency domain data, on the other hand, undergoes a distinct initial processing phase. Since spectrograms serve as input, we employ a vision-based transformer model to extract meaningful spectral features. Subsequently, we apply a spatial data processing pipeline akin to that of the temporal domain. Finally, we merge the extracted features from both domains using an attention mechanism, thereby completing the EEG decoding task.

B. Temporal Sequence Data

Figure 2 presents a depiction of the temporal feature extraction process, continuing from the data pre-processing. The framework for this process has been segmented into three stages: Transformer fusion RNN module, Manifold Domain, and Attention Variant Model.

1) Transformer Fusion RNN Module: RNN algorithms have proven to be effective in processing time-series data due to their ability to capture contextual memory and exhibit powerful expressive capabilities; however, they are not without their limitations. Two such challenges are vanishing or exploding gradients, which hinder the training process and affect the classification model's performance. On the other hand, transformer models have gained prominence for their capacity to model global contextual information within input sequences using self-attention mechanisms. By learning the relationships between different positions in a sequence and assigning weights to each position, transformers can incorporate important information from the entire sequence effectively.

In order to fully utilize the temporal dependencies within EEG signals and simultaneously capture global contextual information, a combination of the advantages offered by both RNNs and transformers is necessary. RWKV [44] successfully combines the strengths of RNNs and transformers.

By combining the strengths of RNNs and transformers, RWKV can capture the crucial temporal information while preserving powerful representation capabilities. This leads to more robust features and improved accuracy, as demonstrated in the conducted ablation experiments. The advancements introduced by RWKV pave the way for further progress in EEG decoding and offer valuable insights into the potential of hybrid models that leverage the strengths of different architectures.

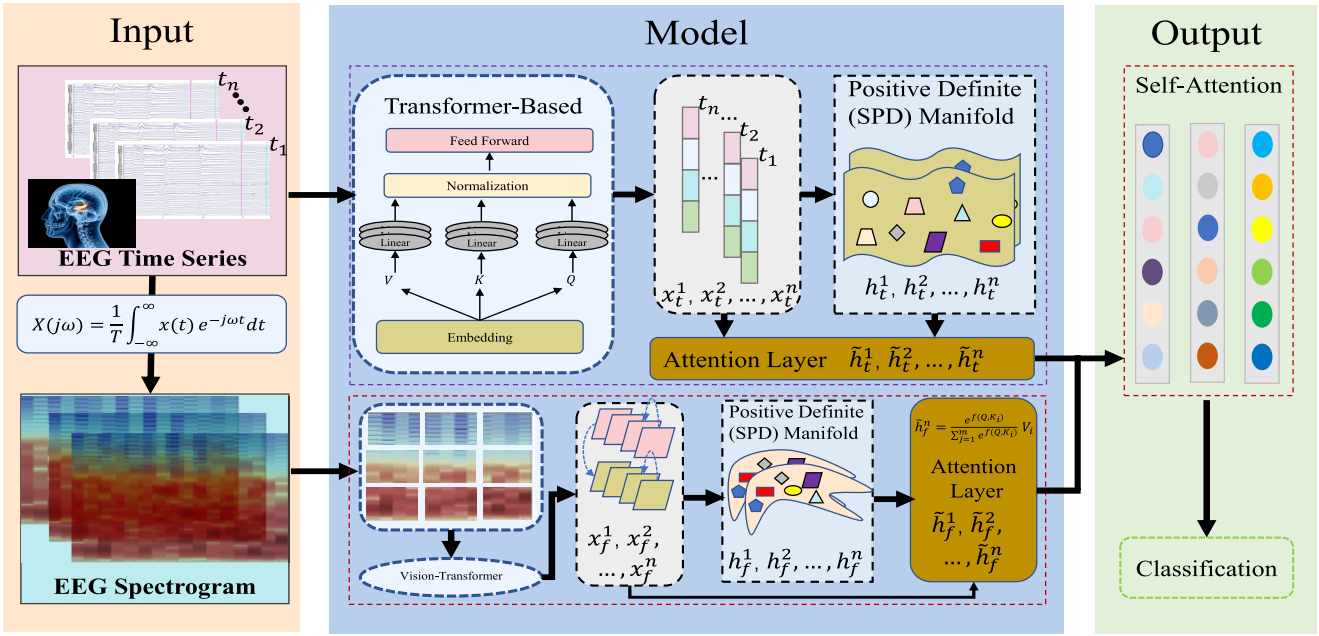

Fig. 1. The framework of TBEEG.

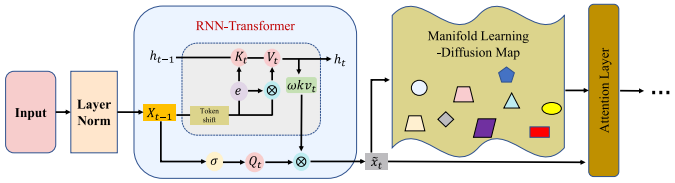

Fig. 2. The time-series data framework of TBEEG.

In our paper, the input of EEG is x_i, the matrix of EEG is $X_t = (x_0, \ldots, x_i, \ldots, x_t)$, μ is a padding operation using PyTorch (nn.ZeroPad2d((0,0,1,−1))), and the RNN-transformer formulations are (1-5):

$$Q_t = W_Q(\mu_Q \cdot x_t + (1-\mu_Q)\cdot x_t - 1) \tag{1}$$

$$K_t = W_K(\mu_K \cdot x_t + (1-\mu_K)\cdot x_t - 1) \tag{2}$$

$$V_t = W_V(\mu_V \cdot x_t + (1-\mu_V)\cdot x_t - 1) \tag{3}$$

$$\omega k V_t = \frac{\sum_{i=1}^{t-1} e^{(t-1-i)\omega + K_i} V_i + e^{\mu + K_t} V_t}{\sum_{i=1}^{t-1} e^{(t-1-i)\omega + K_i} + e^{\mu + K_t}} \tag{4}$$

$$\widetilde{x}_t = W_{X_t}(\sigma(Q_t) \odot \omega K V_t) \tag{5}$$

Q_t, K_t, V_t are the linear projection vectors which abtain linear combination of the current input and the previous input. W_Q, W_K, W_V are the weight decay vector which are trainable model parameters. μ_Q, μ_K, μ_V are the padding operation. The $i \in t$ donate time and $\omega \in R^{T\times T}$ is the learned pair-wise position biases. $\widetilde{x}_t$ donates the output.

2) Manifold Domain: Diffusion maps [45] leverage eigenvectors and eigenvalues of a diffusion operator applied to data to calculate coordinates, facilitating feature extraction. These algorithms belong to the broader category of nonlinear dimensionality reduction techniques, which aim to unveil the underlying manifold from which data points are sampled. Diffusion maps achieve a comprehensive portrayal of the dataset by amalgamating local similarities across various scales. Compared to other methods, they offer the advantage of being less susceptible to noise and more cost-effective in terms of computational resources.

We get the input data which is $\widetilde{X}_t = (\widetilde{x}_0, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_t)$ in the previous part. The process of defining diffusion maps can be divided into four distinct steps:

- Defining the similarity matrix L^α: The initial step involves constructing a similarity matrix that encapsulates the pairwise similarities or distances among data points. The selection of an appropriate similarity measure relies on the particular problem at hand and the characteristics of the data.

$$L^\alpha_{i,j} = \rho^\alpha(\widetilde{x}_i, \widetilde{x}_j) = \frac{L(i,j)}{(d(\widetilde{x}_i)d(\widetilde{x}_j))^\alpha}. \tag{6}$$

ρ denotes kernel function $\rho(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{\epsilon})$, α is to tune the influence of the data point density on the infinitesimal transition of the diffusion. $\widetilde{x}_i, \widetilde{x}_j \in \widetilde{X}_t$ are the input data. The $L(i,j)$ donates the connectivity, indicating the likelihood of transitioning from $\widetilde{x}_i$ to $\widetilde{x}_j$ in a single step within a random walk. The function $d(\widetilde{x}_i) = \int_{\widetilde{X}} \rho(x_i, x_j) d(x_j)$ is part of a procedure referred to as the normalized graph Laplacian construction, which establishes a reversible discrete-time Markov chain on the input data $\widetilde{X}$.

$L(i,j)$ donates the connectivity which means the probability of walking from $\widetilde{x}_i$ to $\widetilde{x}_j$ in one step of the random walk. $d(\widetilde{x}_i)$ is a process known as the normalized graph Laplacian construction construct a reversible discrete-time Markov chain on input data $\widetilde{X}_t$.

$$L^\alpha = D^{-\alpha} L D^{-\alpha}. \tag{7}$$

Both equation (6) and equation (7) are equivalent in terms of their mathematical representation.D donates a diagonal matrix. $D_{i,i} = \sum_j (L_{i,j})$ donates a element of

the diagonal matrix D. α is a parameter to Normalize the matrix L^{α}.

- The normalization Laplacian M: The M matrix is commonly normalized to prevent any bias towards high-degree nodes during the diffusion process. This normalization step is crucial as it highlights the local structure of the data, enhancing the analysis and interpretation of the diffusion process.

$$M = (D^{\alpha})^{-1} L^{\alpha} = (\sum_{j} (L_{i,j}))^{\alpha - 1} \frac{L(i,j)}{(d(\widetilde{x}_i) d(\widetilde{x}_j))^{\alpha}}. \quad (8)$$

M is a normalization Laplacian metric. D^{α} is a diagonal matrix. $D_{i,i} = \sum_j L^{\alpha}_{i,j}$ donates a element of the diagonal matrix D^{α}.

- The eigenvalues and corresponding eigenvectors of M^t: The third step involves performing eigenvalue decomposition on the diffusion matrix, allowing us to extract the eigenvectors and eigenvalues. This process plays a crucial role in capturing the embedded structure and diffusion behavior of the data. Furthermore, as the chain is iterative advanced in temporal dimension (by taking increasingly higher powers of the diffusion matrix), it unveils the evolving geometric structure of the data at progressively larger scales, illustrating the diffusion process in formulation (9).

$$M^t_{i,j} = \sum_i \lambda_i^t \delta_i(\widetilde{x}_i) \psi_i(\widetilde{x}_j). \quad (9)$$

λ_i^t is the sequence of eigenvalues of M, δ and ψ the biorthogonal right and left eigenvectors respectively.

- Calculating the diffusion map embedding:

$$y_t(\widetilde{x}_t) = \begin{pmatrix} \lambda_0^t \delta_1(\widetilde{x}_0) \\ \lambda_0^t \delta_2(\widetilde{x}_1) \\ \cdots \\ \lambda_i^t \delta_i(\widetilde{x}_t) \end{pmatrix} \quad (10)$$

At last, the embedding of the diffusion map $y_t(\widetilde{x}_t)$ is obtained by combining the eigenvectors with their corresponding eigenvalues. The diffusion map algorithm typically uses a power function of the eigenvalues to control the influence of each eigenvector. This step effectively reduces the dimensional of the data while preserving the essential information and intrinsic geometry. Thus we get the diffusion map from the original data to a L-dimensional space which is embedded in the original space. The elements diffusion maps are defined as $\lambda_i^t \delta_i(\widetilde{x})$ and the embedding output is $y_t(\widetilde{x})$ in formulation (10).

3) Attention Variant Module: Considering that we simultaneously incorporate both the data features projected into the popular space and the original data features into the attention layer, it is crucial to recognize the underlying correlation between these two feature sets. Consequently, in the subsequent attention layer, we depart from a straightforward dot multiplication method and opt for cosine similarity instead. This choice allows us to effectively capture the local feature correlation between the two sets of features. By leveraging cosine similarity, we can gain insights into the inter-dependencies among these features and harness this knowledge to assign weights that reflect their relevance and importance within the attention mechanism.

We obtain two sets $Y_t = (y_1, \ldots, y_i, \ldots, y_t)$ which is the output of diffusion projection, and $\widetilde{X}_t = (\widetilde{x}_1, \ldots, \widetilde{x}_i, \ldots, \widetilde{x}_t)$ which is the output of Transformer fusion RNN layer. Through a linear network, we extract the query, key, and value representations from these inputs:

$$C_t = con(Y_t, \widetilde{X}_t) \quad (11)$$

$$q_i = Linear(c_i; W_q) \quad (12)$$

$$k_i = Linear(c_i; W_k) \quad (13)$$

$$v_i = Linear(c_i; W_v) \quad (14)$$

C_t donates the concatenated matrix of Y_t and $\widetilde{X}_t$ which can helpful to obtain data features of different dimensions. con is concatenation operation and $Linear$ is fully connected operation. $c_i \in C_t$ is element of matrix C_t. W_q, W_k and W_v are the weight.

To establish the relationship between q_i and k_i, we employ the cosine similarity measure. This enables us to quantify the similarity or dissimilarity is optional between the query (q_i) and key (k_i) vectors based on their respective orientations in the feature space. By computing the cosine similarity, we can determine the degree of correlation between the two vectors, allowing us to capture their relational information.

$$F(q_i, k_i) = Cosine(q_i, k_i) = \frac{\sum_1^t q_i k_i}{\sqrt{\sum_1^t {q_i}^2} \sqrt{\sum_1^t {k_i}^2}}. \quad (15)$$

Subsequently, we utilize the obtained similarity matrix $\mathcal{A}_t$ to compute the weighted output matrix. The weights in the matrix are determined by the attention mechanism, which assigns higher weights to more relevant or significant elements in the input. By multiplying the attention weights with the corresponding values in the input matrix, we generate the weighted output matrix. This process effectively emphasizes the important features and suppresses the less relevant ones. The most salient information to be capture and highlight in the output representation:

$$\mathcal{A}_t = softmax(F(q_i, k_i)) = \frac{exp(F(q_i, k_i))}{\sum_1^t exp(F(q_i, k_i))}. \quad (16)$$

$softmax$ is a softmax function to obtain the weights on the values. The final output of temporal dimension is O_t:

$$O_t = \mathcal{A} v_i = \frac{exp(F(q_i, k_i)) v_i}{\sum_1^t exp(F(q_i, k_i))}. \quad (17)$$

C. Frequency Domain Features

The entire process of acquiring frequency domain features is demonstrated in Figure 3. which involves four stages. The first stage involves transforming the raw data into frequency domain data using the fast Fourier transform (FFT) algorithm. The subsequent stages (2 to 4) follow a similar process to the acquisition of temporal dimension features in stages 1 to 3. Due to the inputs of frequency domain are spectral images, we use the methods used for processing temporal features are not applicable. Therefore, in the first stage of processing image data, a transformer model based on image processing techniques is employed.

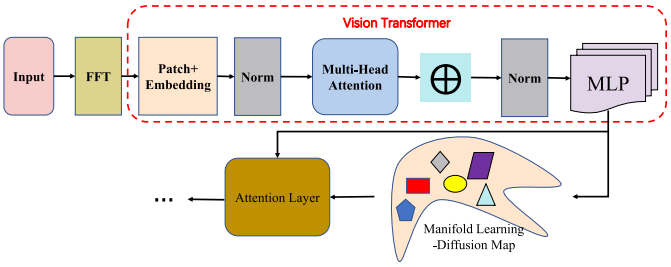

Fig. 3. Detailed description of the frequency domain data processing module in the TBEEG algorithm.

1) Shift Into Spectral Domain: The input is $X_t = (x_1, \ldots, x_i, \ldots, x_t)$, we transform into the frequency domain and the formulation is:

$$X_f = FFT(X_t) = \sum_{i=0}^{\frac{N}{2}-1} x_i(2i) W_{\frac{N}{2}}^{il}. \tag{18}$$

$l = (0, 1, \ldots, \frac{N}{2} - 1)$

$W_{\frac{N}{2}}$ is complex number which is on the unit circle in the complex plane and the argument is a positive and minimal.

2) The Transformer-Based Layer for Spectral Domain: The transformer encoder is structured with a series of layers, where each layer contains multi-headed self-attention and MLP blocks. Additionally, layer normalization (LN) is employed before each block to ensure optimal performance. The formula derivation process is as follows:

$$q_{fi} = Linear(x_{fi}; W_q) \tag{19}$$
$$k_{fi} = Linear(x_{fi}; W_k) \tag{20}$$
$$v_{fi} = Linear(x_{fi}; W_v) \tag{21}$$
$$\mathcal{A}_f = softmax(F(q_{fi}, k_{fi})) = \frac{exp(F(q_{fi}, k_{fi}))}{\sum_1^N exp(F(q_{fi}, k_{fi}))}. \tag{22}$$
$$O_f = \mathcal{A}_f v_{fi}. \tag{23}$$
$$\widetilde{x}_f = MLP(LN(\mathcal{M}O_f). \tag{24}$$

x_{fi} is a element of $X_f = (x_{f1}, \ldots, x_{fi}, \ldots, x_f)$. W_q, W_k and W_v are the weight. $\mathcal{A}_f$ donates similarity matrix of spectral domain. O_f donates the output of frequency domain. $\mathcal{M}$ is the multi-headed self-attention operation. LN is the normalization operation and MLP is fully connected operation.

D. Fusion

Finally, through the utilization of the self-attention layer, we are able to seamlessly combine the temporal and frequency domain features, allowing for their effective integration. This fusion of the two domains is paramount in successfully addressing the EEG decoding task at hand. By leveraging the self-attention mechanism, our model is able to capture and highlight the relevant patterns and correlations between the temporal and frequency components of the EEG data. This integration of information improves the overall decoding performance and provides a more robust and accurate representation of the underlying brain dynamics.

III. Experiments and Results

A. Results in Different Datasets

1) Implementation Details: Our method was implemented using the PyTorch library in Python 3.9.13, utilizing a GeForce 3090 GPU for efficient computation. We adopt PyTorch 1.12 as the default deep learning library. The model was trained using the Adam optimizer with a specific learning rate.

2) Dataset: **BCIC-IV-2a dataset** [46] is collected from 9 subjects. We allocated the first session of a subject to the training set, reserving one out of eight sessions for validation. The format is ".mat" and size is 284 MB. The preprocessing of the BCI dataset's 22-channel EEG signals includes several steps. Initially, the sampling rate was down-sampled from 256 Hz to 128 Hz. Subsequently, a bandpass filter constrained the data frequency to the 4-38 Hz range. Lastly, EEG signal is then segmented into 0.5 to 4-second intervals post-cue onset, yielding 438 timepoints per trial. The data was acquired using a cue-based BCI paradigm involving four different motor imagery tasks: imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Each run included 48 trials, resulting in a total of 288 trials per session. For EEG recording, a montage of 22 Ag/AgCl electrodes with an inter-electrode distance of 3.5cm was used. The left mastoid served as the reference electrode, and the right mastoid was used as the ground. The EOG channels are included specifically for artifact processing purposes and should not be utilized for classification analysis.

The **MAMEM-SSVEP-II** datasets [38] is captured from 11 subjects and the EEG signals with 256 channels. The size of MAMEM dataset is 24.4MB. The sampling rate is 250 Hz. The preprocessing for MAMEM dataset involved three steps: Firstly, a band-pass filter was applied within the range of 1-50 Hz. Secondly, eight specific channels located in the occipital area are selected where the visual cortex is situated. Finally, this produced 500 trials of 1-second, 8-channel SSVEP signals for each subject, with the EEG data's time duration set at 125.

3) Result: We have used three metrics to evaluate our algorithm named TBEEG which are accuracy, recall and F1-score. We have listed all the results of different subjects in two datasets.

The analysis of the data shows that the classification results for subjects 1, 3, 7, 8, and 9 are relatively better with high accuracy rates ranging from 74.31% to 88.54%. The recall and F1-Score for these subjects are 71.88% to 80.67% and 0.7192 to 0.8085, indicating that TBEEG is performing well in correctly identifying their driving behaviors. On the other hand, the classification results for subjects 2, 4, 5, and 6 are poor, with accuracy rates ranging from 45.12% to 54.27%. Recall is 53.12%, 53.12%, 52.78% and 39.58%,respectively. F1-Score is 0.5354, 0.5317, 0.5168, 0.3758, respectively. These findings indicate that the model's performance varies significantly across different subjects, suggesting that individuals' differences play a role in the classification accuracy.

By conducting a detailed analysis of the MAMEM dataset, we observed intriguing patterns in Figures 4 and 5.

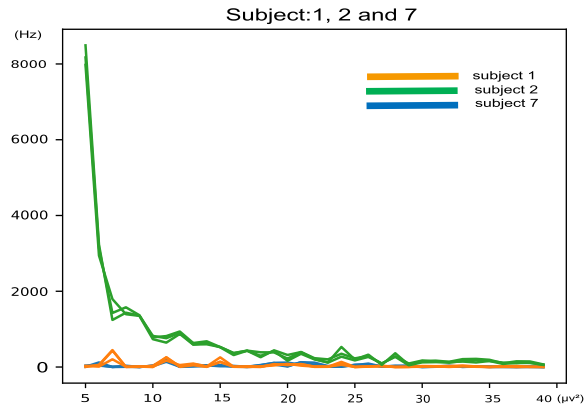

Fig. 4. The Raw data about Subject 1, 2 and 7.

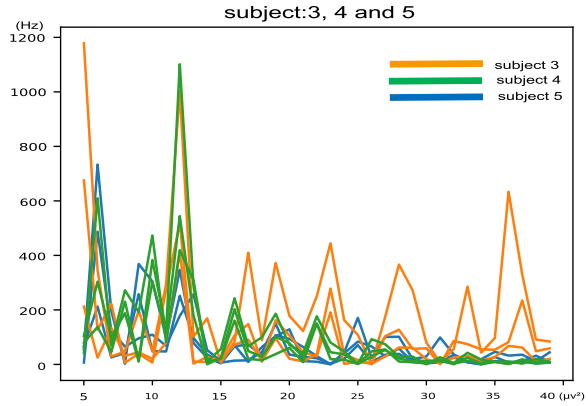

Fig. 5. The Raw data about Subject 3, 4 and 5.

Specifically, we noticed that the data for Subjects 1, 2, and 7 exhibited relatively smooth patterns, which posed challenges in extracting discriminative features. On the other hand, for Subjects 3, 4, and 5, the data exhibited more pronounced discrimination, enabling the extraction of discriminative features with relative ease. Subjects 1, 2, and 7 may require more sophisticated feature engineering techniques or advanced algorithms to reveal meaningful patterns, given their relatively smoother data distributions. In contrast, the discrimination in the raw data of Subjects 3, 4, and 5 may facilitating extract the identification features.

Overall, this study demonstrates that the model has promising potential for classifying driving behaviors based on physiological data. However, it also highlights the need for further refinement and optimization to achieve more consistent and accurate results across all subjects.

B. Comparison With State-of-the-Art Baselines

We validate the performance of TBEEG against other state-of-the-art methods in BCIC-IV-2a and MAMEM-SSVEP-I datasets, and the results show in Table III and Table IV.

1) **Conformer** [48]: The EEG Conformer is a compact convolutional transformer model. The convolution unit utilizes one-dimensional temporal and spatial convolution layers to detect low-level local attributes. Meanwhile, the self-attention unit is responsible for deriving global correlations out of the localized temporal characteristics.
2) **Trans** [31]:Trans is a fusion model integrates CNN and transformer modules. The CNN extracts features, considering spatial and temporal aspects, which are then fed into the transformer's attention layer. Both spatial and temporal models utilize CNN-derived features enriched with positional information. The transformed features pass through fully connected layers for classification.
3) **SGLNet** [49] SGLNet is an spiking neural network (SNN) model designed for EEG-based brain-computer interfaces (BCIs). SNNs neglect the spatial topology of EEG channels and the temporal dependencies within spikes.
4) **DeepSleepNet** [50]: The MBSTCNN-ECA-LightGBM is an end-to-end deep learning model for MI-EEG decoding. It employs a multi-branch CNN module to capture spectral-temporal features and incorporates an efficient channel attention mechanism for enhanced discriminative features.
5) **CS-CNN** [51]: This study introduces a shallow network architecture with three convolutional layers to extract implicit global spatial features. A 1D convolution kernel is utilized to match the input size, equivalent to the CSP spatial filter. The number of convolution kernels gradually increases in complexity (8, 16, and 32).
6) **EEG-inception** [9]: EEG-inception fuse a CNN module with two layers, an attention mechanism layer, two global pooling layers, three fully connected layers with concatenation, and a final softmax layer. This architecture enables effective feature extraction and classification by capturing intricate patterns in the input data. The attention mechanism focuses on relevant features, while the global pooling layers aggregate spatial information. The fully connected layers facilitate rich interactions, and the softmax layer provides classification probabilities.
7) **TA-MFFNet** [52]: The T-A-MFFNet, a multi-feature fusion network, integrates temporal domain and attention networks. Comprised of TNet for temporal series information extraction, CANet and SANet for channel and spatial feature fusion, and MFFNet for multi-dimensional feature merging.

In Tables III, our proposed algorithm TBEEG shows excellent performance in all subjects, exceeding the accuracy achieved in most of the baseline experiments. Notably, the algorithm achieved the highest improvement of 45.47% compared to the baselines, indicating its remarkable effectiveness. In Tables VI, the accuracy of the TBEEG algorithm is higher than all the baseline experiments. We can be seen from the table that the performance of the algorithm SGLNet algorithm is relatively poor. The average accuracy are 42.17% and 27.51%. This algorithm constructs the adjacency matrix for the graph network using electrode positions. However, EEG signal spatial information is not solely captured by electrode proximity; signals from distant electrodes can more accurately represent coherent characteristics. Consequently, this approach does not fully capture the spatial information

TABLE I
PERFORMANCE ON BCIC-IV-2A DATASET

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg±Std
ACC	81.54%	54.27%	88.54%	53.84%	51.73%	45.12%	74.31%	72.10%	78.13%	66.57±14.69
Recall	80.67%	53.12%	80.55%	53.12%	52.78%	39.58%	73.26%	71.88%	78.47%	64.82±14.38
F1-score	0.8085	0.5354	0.8031	0.5317	0.5168	0.3758	0.7333	0.7192	0.7802	66.49±14.79

TABLE II
PERFORMANCE ON MAMEM-SSVEP-I DATASET

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Avg±Std
ACC	81.39%	94.48%	63.64%	42.32%	33.24%	84.71%	74.17%	38.16%	91.75%	71.47%	92.91%	69.93±21.48
Recall	80.58%	94.25%	63.16%	41.50%	32.67%	82.83%	66.58%	38.16%	91.99%	91.17%	90.00%	70.26±22.30
F1-score	0.8078	0.9431	0.6266	0.4082	0.3005	0.8334	0.6694	0.2265	0.9195	0.7109	0.9159	66.92±24.27

TABLE III
COMPARISONS WITH STATE-OF-THE-ART METHODS ON BCIC-IV-2A DATASET

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg±Std
Conformer	80.37%	52.53%	87.85%	65.97%	35.98%	35.83%	81.94%	79.86%	72.22%	65.83±18.74
Trans	72.92%	46.18%	79.86%	65.97%	49.31%	52.43%	83.68%	79.86%	79.51%	67.75±14.81
SGLNet	54.86%	43.33%	47.57%	38.39%	24.18%	32.99%	39.93%	38.89%	59.38%	42.17±10.76
DeepSleepNet	48.96%	29.81%	59.38%	39.58%	25.35%	39.93%	50.35%	49.65%	64.24%	45.25±12.82
CS-CNN	51.39%	30.90%	59.38%	43.40%	27.78%	39.58%	59.03%	57.64%	65.28%	48.26±13.45
EEG-inception	76.74%	43.90%	89.93%	67.01%	55.90%	52.53%	80.21%	79.51%	84.03%	71.31±13.83
TA-MFFNet	46.18%	30.21%	60.76%	37.15%	26.04%	30.56%	54.86%	36.11%	59.03%	42.32±13.26
TBEEG	81.54%	54.27%	88.54%	53.84%	51.73%	45.12%	74.31%	72.10%	78.13%	66.57±14.69

TABLE IV
COMPARISONS WITH STATE-OF-THE-ART METHODS ON MAMEM DATASET

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Avg±Std
Conformer	78.79%	69.70%	57.58%	23.46%	22.45%	76.77%	63.64%	26.26%	90.91%	67.68%	89.90%	60.65±25.56
Trans	69.70%	46.46%	63.64%	26.53%	25.51%	80.81%	67.68%	27.55%	89.90%	60.61%	91.92%	59.12±24.60
SGLNet	36.73%	43.33%	24.49%	27.99%	24.18%	23.47%	24.49%	24.49%	26.53%	20.41%	26.53%	27.51±6.65
DeepSleepNet	62.24%	25.51%	36.73%	27.55%	22.45%	35.71%	48.98%	23.47%	56.12%	32.65%	41.84%	37.57±13.43
CS-CNN	59.18%	27.55%	35.35%	22.45%	29.59%	24.49%	34.83%	27.55%	48.98%	26.53%	42.70%	34.47±11.47
EEG-inception	73.74%	27.27%	58.59%	27.27%	26.53%	76.77%	72.73%	27.55%	90.91%	61.62%	90.91%	57.63±26.08
TA-MFFNet	70.71%	20.41%	24.49%	24.49%	25.51%	23.47%	26.53%	26.00%	45.92%	25.00%	36.73%	31.75±14.80
TBEEG	81.39%	94.48%	63.64%	43.32%	33.24%	84.70%	74.17%	38.16%	91.75%	71.47%	92.91%	69.93±21.48

of EEG data. Additionally, the use of spike-based LSTM units to analyze temporal dependencies introduces cumulative errors. The Conformer, Trans, and TA-MFFNet are based on Transformer and Attention algorithms. The average accuracy are 42.32% and 31.75%. The first two algorithms have relatively high accuracy, which is mainly for the sufficient information mining of the event dimension. However, the TA-MFFNet algorithm performs worse, mainly because it uses CNN network and LSTM network to obtain shallow feature information in the time dimension and space dimension, which will lead to some local features being smoothed. TA-MFFNet algorithm only uses the attention module in the last layer for spatial dimension features. DeepSleepNet, CS-CNN, and EEG-inception all utilize CNN networks to extract feature information. There are certain defects in feature acquisition, such as potential limitations in capturing complex temporal and spatial dependencies inherent in the data. These models, while effective in certain contexts, may not fully leverage the depth and richness of the dataset, potentially impacting the overall performance and accuracy of the algorithms. These results highlight the superiority of our approach in accurately classifying driving behaviors based on physiological data.

To ensure the fairness of our laboratory comparison, all baseline experiments were conducted using the same dataset. However, the dataset employed in the original studies for baseline algorithms like SGLNet, DeepSleepNet, CS-CNN, and TA-MFFNet differs from ours. This discrepancy may result in lower accuracy in our verification compared to the original findings, primarily due to: (1) limited generalization capability, as the baseline models focus on analyzing EEG signals' time-domain data, which can significantly vary and be easily influenced by various noises; and (2) our replication of the models as described in these studies, where differences in data preprocessing and optimization techniques from the original work could introduce variations in the outcomes.

Furthermore, the consistent performance across different subjects indicates the robustness and generalizability of our proposed algorithm. Overall, the results obtained from our algorithm provide strong evidence of its effectiveness in accurately classifying driving behaviors. This performance advantage positions our approach as a promising solution for enhancing the safety and efficiency of autonomous vehicles and intelligent driver assistance systems. Continued research and optimization can further enhance the algorithm's potential

TABLE V
ABLATION RESULTS ON TWO DATASETS

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Avg±Std
Conformer	78.79%	69.70%	57.58%	23.46%	22.45%	76.77%	63.64%	26.26%	90.91%	67.68%	89.90%	60.65±25.56
Trans	69.70%	46.46%	63.64%	26.53%	25.51%	80.81%	67.68%	27.55%	89.90%	60.61%	91.92%	59.12±24.60
SGLNet	36.73%	43.33%	24.49%	27.99%	24.18%	23.47%	24.49%	24.49%	26.53%	20.41%	26.53%	27.51±6.65
DeepSleepNet	62.24%	25.51%	36.73%	27.55%	22.45%	35.71%	48.98%	23.47%	56.12%	32.65%	41.84%	37.57±13.43
CS-CNN	59.18%	27.55%	35.35%	22.45%	29.59%	24.49%	34.83%	27.55%	48.98%	26.53%	42.70%	34.47±11.47
EEG-inception	73.74%	27.27%	58.59%	27.27%	26.53%	76.77%	72.73%	27.55%	90.91%	61.62%	90.91%	57.63±26.08
TA-MFFNet	70.71%	20.41%	24.49%	24.49%	25.51%	23.47%	26.53%	26.00%	45.92%	25.00%	36.73%	31.75±14.80
TBEEG	81.39%	94.48%	63.64%	43.32%	33.24%	84.70%	74.17%	38.16%	91.75%	71.47%	92.91%	69.93±21.48

and contribute to advancing the field of driving behavior analysis and classification.

C. Ablation Results

In our study, we proposed an algorithm called TBEEG, which comprises five components aimed at improving the classification of EEG data. To evaluate the effectiveness of each component, we conducted ablation experiments on two diverse datasets, providing valuable insights into their contributions. The results summarized in Table V, shed light on the impact of various components on the overall performance of the algorithm.

(1) We examined the accuracy rates when isolating the temporal domain data or the frequency domain data. It was observed that both scenarios yielded lower accuracy rates compared to the fused data. This clearly indicates the importance of integrating information from dimensions to enhance classification accuracy. By combining the data in various dimensions, we effectively captured complementary features and achieved improved results. The Manifold Domain is characterized by the property that the Euclidean distance between points in the embedded space is equivalent to the "diffusion distance" between the probability distributions centered on these points. This integration of local similarities at various scales enables the diffusion map to offer a representation of the dataset at a global level.

(2) We conducted experiments to investigate the role of the transformer-based model within the temporal-frequency domain network. Remarkably, when the transformer-based model was removed from the architecture, a noticeable decrease in accuracy was observed, amounting to 23.12% and 37.69%. when the transformer-based model was removed from the architecture, a noticeable decrease in accuracy was observed, amounting to 23.12% and 37.69%. These findings highlight the significance of employing different transformer-based models to effectively capture the complex spatial-temporal feature relationships present in EEG data.

(3) We explored the impact of the attention layer in the feature fusion process. By replacing the attention layer with addition and splicing methods, the accuracy of the algorithm decreased 7.56% and 18.58%. This underscores the crucial role played by the attention layer in fusing and weighting features, indicating that effective feature fusion techniques are vital for achieving optimal classification performance.

D. Parameter Sensitivity

The parameter experiments in our study encompassed a range of settings to thoroughly explore their impact on the algorithm's performance. We systematically varied multiple parameters to assess their effects on the classification accuracy of the model.

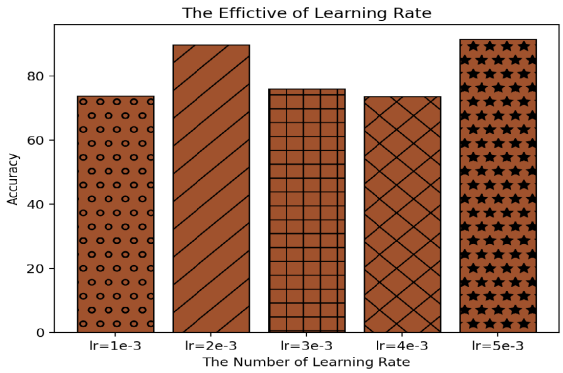

Fig. 6. The effective of learning rate on classification accuracy.

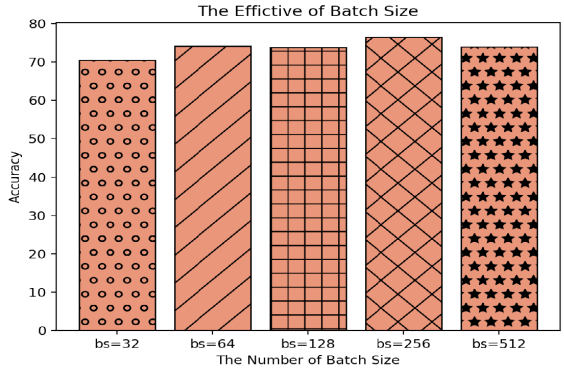

Fig. 7. The effective of batch size on classification accuracy.

Specifically, we examined the influence of parameters such as learning rate, batch size and number of muti-heads. Each parameter was individually adjusted while keeping other settings constant, allowing us to isolate their effects on the algorithm's performance.

The Figure 6 presents the effect of different learning rates on the classification accuracy. The results are displayed in Figure 4, where the x-axis represents the values of the adjusted learning rates, and the y-axis indicates the corresponding accuracy percentages. The learning rates examined are 1e-3, 2e-3, 3e-3, 4e-3, and 5e-3, resulting in accuracy rates of 73.77%, 89.78%, 75.93%, 73.60%, and 91.39%, respectively. From the analysis, we can observe that the learning rate significantly influences the performance of the classification model. When the learning rate is set to 5e-3, the accuracy achieves the highest value of 91.39%, indicating that a larger learning rate in this context leads to better classification results. However, when the learning rate decreases to 1e-3, the accuracy drops to 73.77%, indicating that a very small learning rate may hinder

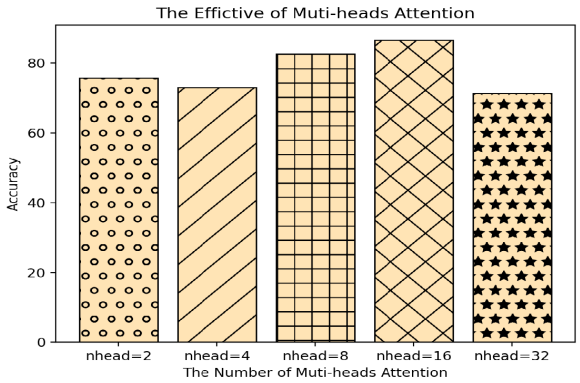

Fig. 8. The effective of muti-heads attention on classification accuracy.

the convergence of the model during training and reduce its effectiveness.

Figure 7 illustrates the influence of batch size on the classification accuracy. The x-axis represents the different batch sizes used in the experiments, while the y-axis displays the corresponding accuracy percentages. The batch sizes examined in this study include 32, 64, 128, 256, and 512, resulting in accuracy rates of 70.38%, 74.08%, 73.77%, 76.35%, and 73.94%, respectively. The analysis of batch size's impact on classification accuracy reveals interesting findings. As observed in the results, the accuracy varies with different batch sizes. When the batch size is set to 256, the model attains the highest accuracy of 76.35%. This indicates that a moderate batch size can lead to improved classification performance. However, excessively small batch sizes, such as 32 and 64, result in relatively lower accuracy values of 70.38% and 74.08%, respectively. This is because small batch sizes may hinder the model's ability to generalize well, leading to sub-optimal results. On the other hand, very large batch sizes, such as 512, yield an accuracy of 73.94%. This suggests that extremely large batch sizes might not effectively exploit the data information during the training process, which can affect the model's generalization capabilities.

Figure 8 depicts the analysis of how the number of Multi-heads affects the classification accuracy. The x-axis represents the different numbers of Multi-heads used in the experiments, while the y-axis shows the corresponding accuracy percentages. The evaluated numbers of Multi-heads are 2, 4, 8, 16, and 32, resulting in accuracy rates of 75.73%, 73.07%, 82.57%, 86.48%, and 71.24%, respectively. The analysis of the impact of Multi-heads on classification accuracy reveals interesting patterns. As observed in the results, the accuracy varies with different numbers of Multi-heads. When the number of Multi-heads is set to 16, the model achieves the highest accuracy of 86.48%. This indicates that an intermediate number of Multi-heads can lead to improved classification performance, as it allows the model to effectively capture and integrate diverse features from different attention heads. However, very low numbers of Multi-heads, such as 2 and 4, result in relatively lower accuracy values of 75.73% and 73.07%, respectively. This is because a limited number of attention heads might not fully exploit the potential feature interactions in the data, leading to suboptimal results. On the other hand, excessively high numbers of Multi-heads, such as 32, yield an accuracy of 71.24%. This suggests that an excessive number of attention heads can introduce noise and redundancy, impacting the model's ability to learn meaningful representations.

By systematically exploring and adjusting these parameters, we gained insights into their effects on the algorithm's performance and identified the optimal configuration for achieving the highest classification accuracy.

IV. Discussion

This paper introduces a novel dual-layer model called TBEEG, which aims to decode EEG signals by extracting features from both the temporal domain and frequency domains. In the temporal domain network, the focus is on capturing the intricate temporal relationships within EEG data. To achieve this, we employ an RNN-transformer-based algorithm that effectively explores the dependencies among different temporal dimensions of the EEG features. By leveraging this approach, we are able to uncover valuable temporal patterns and dependencies that contribute to the overall decoding performance.

In the frequency domain, our methodology begins with a transformation step using the Fast Fourier Transform (FFT). This allows us to convert the temporal domain EEG data into the frequency domain, providing us with a different perspective on the underlying signals. Next, we utilize the vision-transformer algorithm, which has proven to be effective in capturing complex patterns and relationships in visual data. By adapting this algorithm to the EEG domain, we are able to extract informative features from the frequency domain.

To further enhance the robustness and stability of the spatial dimension features, we apply the diffusion maps algorithm. The diffusion maps allows us to project the temporal and frequency domain features onto a high-dimensional space, where more discriminative and resilient spatial features can be extracted. The effectiveness of the TBEEG algorithm heavily relies on the specific choice of algorithms used for feature extraction, such as RNN-transformer and vision-transformer. The generalization of these algorithms to different EEG datasets or decoding tasks may vary, and their performance may not be consistent across different scenarios.

Finally, an attention mechanism is incorporated to assign weights to the features from different dimensions. This attention-based fusion ensures that each dimension contributes appropriately to the decoding task, allowing for accurate and efficient EEG decoding. The proposed TBEEG model is extensively evaluated through experiments conducted on various datasets. Comparative analyses with eight state-of-the-art baseline methods, ablation experiments, and visualizations further validate the effectiveness of our approach in EEG decoding.

There are some potential drawbacks or limitations of the TBEEG algorithm: (1) Complexity: The proposed dual-layer model with feature extraction from both the temporal and frequency domains introduces additional complexity to the algorithm. This increased complexity may result in higher computational requirements and longer processing times, limiting its real-time applicability; (2) Sensitivity to Parameters:

The TBEEG algorithm involves the use of various parameters, such as the number of layers, diffusion map parameters, and attention mechanism weights. The performance of the algorithm could be volatile, depending on the selection and fine-tuning of these parameters; this indicates a requirement for careful optimization for different datasets and applications. (3) Data Requirements: The TBEEG algorithm assumes the availability of sufficient training data for effective learning. Limited or imbalanced datasets may affect the performance and generalized of the algorithm, potentially leading to over-fitting or biased results.

The limitations of our proposed TBEEG algorithm are as follows: 1) The computational complexity is relatively high in our TBEEG algorithm. While projection into the Manifold Domain is utilized for down-sampling to reduce complexity, the use of a self-attention algorithm during the feature fusion stage leads to an increase in computational complexity. In future research, we hope to identify more effective methods to decrease computational complexity. 2) There is a certain degree of variance among different subject samples, making it challenging to achieve high accuracy for specific individuals. In future research, we aim to obtain better accuracy across all subjects, thereby enhancing the algorithm's generalization capability.

V. Conclusion

In this paper, we introduce TBEEG, a novel dual-layer model designed for EEG decoding that is capable of harnessing features from both the temporal and frequency domains. In the temporal domain network, temporal relationships among different temporal dimensions are captured using an RNN-transformer-based algorithm, followed by projecting the obtained features into a high-dimensional space using the diffusion map algorithm to achieve robustness and scalability. Temporal domain features are extracted using the vision-transformer algorithm; further frequency domain data is then extracted from the temporal domain using FFT transformation. Processing of these features using the diffusion map algorithm enhances stability and robustness in the spatial dimension. An attention mechanism is then utilized to assign weights to features from different dimensions, facilitating accurate and efficient EEG decoding. The generalization across different datasets and the sensitivity of performance to parameter settings should be considered. Additionally, the algorithm's reliance on sufficient training data and the need for careful parameter optimization should be taken into account. Experimental validation on various datasets, comparisons with baseline methods, ablation experiments, and visualizations support the algorithm's effectiveness.

References

[1] E. Su, S. Cai, L. Xie, H. Li, and T. Schultz, "STANet: A spatiotemporal attention network for decoding auditory spatial attention from EEG," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2233–2242, Jul. 2022.

[2] S. M. Pani, M. Fraschini, M. Figorilli, L. Tamburrino, R. Ferri, and M. Puligheddu, "Sleep-related hypermotor epilepsy and non-rapid eye movement parasomnias: Differences in the periodic and aperiodic component of the electroencephalographic power spectra," *J. Sleep Res.*, vol. 30, no. 5, Oct. 2021, Art. no. e13339.

[3] M. B. Sterman, "Physiological origins and functional correlates of EEG rhythmic activities: Implications for self-regulation," *Biofeedback Self-Regulation*, vol. 21, no. 1, pp. 3–33, Mar. 1996.

[4] N. Deo, A. Rangesh, and M. M. Trivedi, "How would surround vehicles move? A unified framework for maneuver classification and motion prediction," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 2, pp. 129–140, Jun. 2018.

[5] K.-T. Kim, H.-I. Suk, and S.-W. Lee, "Commanding a brain-controlled wheelchair using steady-state somatosensory evoked potentials," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 654–665, Mar. 2018.

[6] G. Vanhollebeke, S. De Smet, R. De Raedt, C. Baeken, P. van Mierlo, and M.-A. Vanderhasselt, "The neural correlates of psychosocial stress: A systematic review and meta-analysis of spectral analysis EEG studies," *Neurobiol. Stress*, vol. 18, May 2022, Art. no. 100452.

[7] T. Kaufmann, A. Herweg, and A. Kübler, "Toward brain–computer interface based wheelchair control utilizing tactually-evoked event-related potentials," *J. NeuroEng. Rehabil.*, vol. 11, no. 1, p. 7, 2014.

[8] Y. He and X. He, "Facial expression recognition based on multi-feature fusion and HOSVD," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Mar. 2019, pp. 638–643.

[9] K. Barmpas, Y. Panagakis, S. Bakas, D. A. Adamos, N. Laskaris, and S. Zafeiriou, "Improving generalization of CNN-based motor-imagery EEG decoders via dynamic convolutions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1997–2005, 2023.

[10] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Autonomous Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[11] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, Oct. 2012.

[12] M. Miao, H. Zeng, A. Wang, C. Zhao, and F. Liu, "Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: An sparse regression and weighted Naïve Bayesian classifier-based approach," *J. Neurosci. Methods*, vol. 278, pp. 13–24, Feb. 2017.

[13] F. Lotte et al., "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[14] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Jan. 2012.

[15] M. Tavakolan, Z. Frehlick, X. Yong, and C. Menon, "Classifying three imaginary states of the same upper extremity using time-domain features," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0174161.

[16] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. N. Tibarewala, "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data," in *Proc. Int. Conf. Syst. Med. Biol.*, Dec. 2010, pp. 126–131.

[17] W. Fadel, C. Kollod, M. Wahdow, Y. Ibrahim, and I. Ulbert, "Multi-class classification of motor imagery EEG signals using image-based deep recurrent convolutional neural network," in *Proc. 8th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2020, pp. 1–4.

[18] C. Liu et al., "SincNet-based hybrid neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 540–549, 2022.

[19] E. Santamaría-Vázquez, V. Martínez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2773–2782, Dec. 2020.

[20] S. U. Amin, G. Muhammad, W. Abdul, M. Bencherif, and M. Alsulaiman, "Multi-CNN feature fusion for efficient EEG classification," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6.

[21] J. Li, Z. Zhang, and H. He, "Implementation of EEG emotion recognition system based on hierarchical convolutional neural networks," in *Proc. Advances in Brain Inspired Cognitive Systems*, Beijing, China. Cham, Switzerland: Springer, 2016, pp. 22–33.

[22] D. Das Chakladar, S. Dey, P. P. Roy, and D. P. Dogra, "EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm," *Biomed. Signal Process. Control*, vol. 60, Jul. 2020, Art. no. 101989.

[23] X. Du et al., "An efficient LSTM network for emotion recognition from multichannel EEG signals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1528–1540, Jul. 2022.

[24] Z. Fu, C. Huang, L. Zhang, S. Wang, and Y. Zhang, "Deep learning model of sleep EEG signal by using bidirectional recurrent neural network encoding and decoding," *Electronics*, vol. 11, no. 17, p. 2644, Aug. 2022.

[25] G. Xu, W. Guo, and Y. Wang, "Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-conv architecture," *Med. Biol. Eng. Comput.*, vol. 61, no. 1, pp. 61–73, Jan. 2023.

[26] G.-Y. Bae and S. J. Luck, "Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials," *J. Neurosci.*, vol. 38, no. 2, pp. 409–422, Jan. 2018.

[27] D. Li, J. Xu, J. Wang, X. Fang, and J. Ying, "A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, pp. 2615–2626, 2020.

[28] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018.

[29] H. Raza, A. Chowdhury, S. Bhattacharyya, and S. Samothrakis, "Single-trial EEG classification with EEGNet and neural structured learning for improving BCI performance," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[30] X. Che, Y. Zheng, X. Chen, S. Song, and S. Li, "Decoding color visual working memory from EEG signals using graph convolutional neural networks," *Int. J. Neural Syst.*, vol. 32, no. 2, Feb. 2022, Art. no. 2250003.

[31] J. Xie et al., "A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2126–2136, 2022.

[32] K. Singh and J. Malhotra, "Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2405–2418, Jun. 2022.

[33] J. Lanzone et al., "EEG spectral exponent as a synthetic index for the longitudinal assessment of stroke recovery," *Clin. Neurophysiol.*, vol. 137, pp. 92–101, May 2022.

[34] S. Chikhi, N. Matton, and S. Blanchet, "EEGpower spectral measures of cognitive workload: A meta-analysis," *Psychophysiology*, vol. 59, no. 6, Jun. 2022, Art. no. e14009.

[35] L. Ding, W. Duan, Y. Wang, and X. Lei, "Test-retest reproducibility comparison in resting and the mental task states: A sensor and source-level EEG spectral analysis," *Int. J. Psychophysiol.*, vol. 173, pp. 20–28, Mar. 2022.

[36] J. Zulliger, L. D. Hernandez, and T. Koenig, "Within and between subject spectral fingerprints of EEG-microstate parameters," *Brain Topography*, vol. 35, no. 3, pp. 277–281, May 2022.

[37] C. S. Guay et al., "Postoperative delirium severity and recovery correlate with electroencephalogram spectral features," *Anesthesia Analgesia*, vol. 136, no. 1, pp. 140–151, Jan. 2023.

[38] V. Férat, M. Seeber, C. M. Michel, and T. Ros, "Beyond broadband: Towards a spectral decomposition of electroencephalography microstates," *Hum. Brain Mapping*, vol. 43, no. 10, pp. 3047–3061, Jul. 2022.

[39] F. Xu et al., "EEG decoding method based on multi-feature information fusion for spinal cord injury," *Neural Netw.*, vol. 156, pp. 135–151, Dec. 2022.

[40] M. Sorkhi, M. R. Jahed-Motlagh, B. Minaei-Bidgoli, and M. Reza Daliri, "Learning temporal-frequency features of physionet EEG signals using deep convolutional neural network," *Int. J. Modern Phys. C*, vol. 34, no. 4, Apr. 2023, Art. no. 2350047.

[41] F. Li et al., "Decoding imagined speech from EEG signals using hybrid-scale spatial–temporal dilated convolution network," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460c4.

[42] L. Zheng et al., "Ensemble learning method based on temporal, spatial features with multi-scale filter banks for motor imagery EEG classification," *Biomed. Signal Process. Control*, vol. 76, Jul. 2022, Art. no. 103634.

[43] B. Wu, Y. Wang, W. Chen, and X. Zheng, "Time–frequency optimized spatial patterns for movement-related EEG decoding," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2011, pp. 84–87.

[44] B. Peng et al., "RWKV: Reinventing RNNs for the transformer era," 2023, *arXiv:2305.13048*.

[45] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, Jul. 2006.

[46] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," Inst. Knowl. Discov. Graz Univ. Technol., Graz, Austria, Tech. Rep., 16, 2008, pp. 1–6.

[47] I. Lazarou et al., "EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21 st Century," *Frontiers in Human Neuroscience*, vol. 12, no. 14, 2018.

[48] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *Proc. IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2022.

[49] P. Gong, P. Wang, Y. Zhou, and D. Zhang, "A spiking neural network with adaptive graph convolution and LSTM for EEG-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1440–1450, 2023.

[50] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh, and X. Li, "Self-supervised learning for label-efficient sleep stage classification: A comprehensive evaluation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1333–1342, 2023.

[51] Y. Hu et al., "A cross-space CNN with customized characteristics for motor imagery EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1554–1565, 2023.

[52] B. Peng, Y. Zhang, M. Wang, J. Chen, and D. Gao, "T-A-MFFNet: Multi-feature fusion network for EEG analysis and driving fatigue detection based on time domain network and attention network," *Comput. Biol. Chem.*, vol. 104, Jun. 2023, Art. no. 107863.

[53] H. Jia et al., "A model combining multi branch spectral-temporal CNN, efficient channel attention, and LightGBM for MI-BCI classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1311–1320, 2023.

[54] Y. Hou et al., "Deep feature mining via attention-based BiLSTM-GCN for human motor imagery recognition," 2020, *arXiv:2005.00777*.

[55] H. B. Dirik, A. Darendeli, and H. Ertan, "The new wireless EEG device mentalab explore is a valid and reliable system for the measurement of resting state EEG spectral features," *Brain Res.*, vol. 1798, Jan. 2023, Art. no. 148164.