TFSICNet: A Neural Feature-based Encoder for Visual and Motor Imagery EEG Data

Zonghan Du
Department of Computer
Sun Yat-sen University
Guangdong, China
Email: duzh9@mail2.sysu.edu.cn

Zhongyuan Lai
Ballsnow Technology (Shanghai)
Shanghai, China
Email: zhongyuan.lai@ballsnow.com

***Abstract*—Brain-computer interface (BCI) is a technology that enables direct connection and interaction of brain activity with external devices or systems. The encoding and decoding of neural signals play a crucial role in BCIs. The quality of such encodings are the key to robust and accurate information exchange and control between the brain and external devices. Currently, the limited capabilities of conventional brain signal processing is restricting a wider application of BCIs. In this paper, we propose a deep network encoder, Temporal-Frequency-Spatio-Importance Correlation Network (TFSICNet), to robustly and comprehensively encode the original electroencephalogram (EEG) data. The design of TFSICNet is based on an interpretable understanding of the brain's basic structural and connectivity features. The various submodules in TFSICNet were designed to ensure the different features were taken into account. We evaluated encoder performance on three motion imagery datasets and one picture stimulus dataset. The results show that our encoder performs better than traditional deep encoders and advanced deep neural network models that have excelled in extracting EEG features for classification in recent years.**

I. Introduction

In recent years, the concurrent rapid developments in artificial intelligence (AI) and neuroscience have spurred significant amount of research integrating both fields. In particular, the field of brain-computer interfaces (BCIs) has benefited greatly from this integration, as the quality of many BCI-based applications has been significantly improved and their robustness ensured. Such examples include the ability to decode EEG signals into speech [1], [2], [3], and images [4], [5], [6], or operate external hardware (such as a robotic arm [7], [8], [9]) via EEG control. An important prerequisite for such operations is the availability of high-quality encodings of raw EEG signals. The main reasons for their importance are clear: an accurate and representative encoding of neural EEG signals ensures that the correspondence between a specific thought or intention with the actual realization on the operational level is precise and one-to-one [10]. An illustrative example is shown in Figure 1. In this scenario, the encoding vectors corresponding to the distinct activity classes are clearly separated. For this to be the case, the embedding model needs to be able to capture the full range of information contained in the original input signal. We approach this problem by first considering the inverse problem: by focusing on the different characteristics of the human brain, and subsequently relating these to the form of EEG signals.

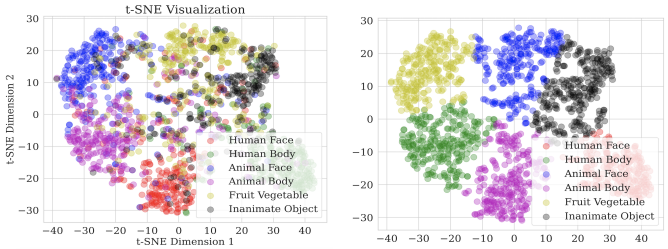

Fig. 1. EEG encoding data after TSNE and k-Means clustering. The figure on the left shows the sample points after dimensionality reduction by TSNE, and the figure on the right shows k-means clustering visualization. It is apparent that our model achieves excellent separability between the different classes

These characteristics can be classified into different classes:

- Connectomic information: It has been shown that connections between brain regions [11] play important roles in the functioning of the human brain; in particular, the brain connectome forms the direct basis for brain plasticity and learning [12], a symptom of neurological afflictions [13], [14], as well as an important topological indicator of brain structure. Therefore, it is reasonable to expect that connectomic information will lead to differences in the EEG signal content;
- Hierarchical information: The fact that the human brain is hierarchically organized is now relatively well-understood [15], [16]. Neural hierarchies is responsible for a multitude of brain functioning, including that of vision [17] and tasks associated with the central nervous system (CNS) as a whole. It has been shown [18] that this structure contributes to characteristics of the neural signals, as well.
- Functional network information: EEG signals have also been shown to reveal information about physical graph parameters such as connectivity and degree distribution [19]. Hence, the overall network structure of the brain is an important indicator of the informational content of EEG signals and their encoding should take such details into account as well.

From the points raised above, we see that structure and organization of the brain contributes significantly to the informational composition and structure of EEG signals. Assuming

that such a link is valid, we can now reformulate the question which actually interests us: how EEG signals help us infer information about the human brain, with emphasis on various cognitive processes and reasoning pathways. This is a decisive question for the efficacy of most BCI setups, since they are predicated on having a consistent and robustly generalizable encoding protocol that is able to capture the full spectrum of information. One apparent use case would be the encoding and subsequent decoding of visual signals [20]. This task typically entails the recording of EEG signals while under visual stimuli, and reproducing the stimuli via a encoding-decoding procedure. However, such a standardized encoding framework is non-trivial to design and implement, as the encoding model will need to consider the characteristics of EEG signal, which are significantly influenced by the different factors detailed above. In order to solve the problem of designing such a encoding framework, we propose a general embedding framework which is able to compute the optimal embedding vector for a wide-range of downstream use cases, hence unifying the different embedding models under a single framework. We numerically verified the robustness of our framework via testing on a wide spectrum of tasks and show that reliability is consistently high across all tasks. In summary, our contributions in this paper consists of the following:

- We adopt a point-of-view on the problem of BCI effectiveness which emphasizes the importance of the encoding model in fully capturing information latent in EEG signals from visual stimuli;
- We design and implement a deep architecture for brain encoding which is able to take the three main points discussed above into account; thus making it suitable for a wide variety of visual tasks, ranging from decoding to classification and regression;
- We conduct extensive numerical experimentation on four motor imagery datasets, in order to conclusively show the efficacy of our encoding framework;

II. Related Work

A. Brain-Computer Interface Applications

The development of brain-computer interfaces has had a long history [21], [22], starting from the first recordings of neural electrical signals by Hans Berger [23]. Since then, the EEG has evolved into an essential tool for the study of cognitive functions and diagnosis of brain diseases. The original idea for a link between the human mind and external applications was put forward by Kamiya [24], in which he showed that EEG activity could be modulated by a human subject after some training. Such conscious modulations demonstrated that humans could control devices through brain activity. The term "BCI" was coined in [25] by Jacques Vidal, and conceptualization of this field could conceivably be traced back to this work. The first real-world applications of BCI were mainly in the area of neural control of external devices, such as the P300 speller [26], and conditioning based on neurofeedback [27]. The subject with which we are most concerned with in this work was first pursued in [28]. Here, Pfurtscheller and his colleagues tested on subjects which were requested to explicitly imagine movements of limbs, during which EEG signals were recorded. These signals were later analyzed using machine learning algorithms for (mainly) classification tasks. This was the original work introducing the idea of *motor-imagery* based BCIs. The modern variant on this seminal work extends the applicability of this paradigm, in the sense that we are able to perform *encoding* of this EEG signal, which enables us to retain this neural information for later downstream tasks.

B. Encoding of Brain EEG Data

In neuroscience, *encoding* is the process where stimulus features are used to predict patterns of brain activity [29]. In most practical workflows, encoding entails the process where raw neural signals are either: 1. projected onto a lower-dimensional space, and the projection vectors used for downstream tasks [30], or, 2. used directly in a learning framework for prediction [31]. In recent years, especially with the increasing prevalence of deep learning [32], the application of machine learning methods in the analysis and processing of neural information has increased dramatically. There are several reasons for this: deep learning's ability to handle high-dimensional and multimodal data [33] has seen its rapid adoption as the EEG encoder of choice. In addition, in contrast to conventional ways of encoding neural information, such as encoding according to time-, frequency- and space-domain features [34], deep learning frameworks offer a much more flexible and powerful tool for neural encoding, since these multi-domain features can all be captured via deep learning [32]. Finally, deep encoding models is able to accomodate a much more varied encoding strategy, which can be fine-tuned according to the different tasks requirements [35].

C. Decoding Tasks

Decoding is the inverse task of encoding; here patterns of brain activity is used to predict the stimulus features. This is the kind of tasks which has inspired the notion of "mind-reading" from popular culture [36], since the reconstruction of the original stimulus from neural activity can be directly interpreted as a reproduction of abstract human thought. In general, the field of neural decoding has been heavily influenced by the increasing use of deep learning, especially in recent years. We see this in context of the increasing sophistication of generative models [37], in particular those based on the diffusion mechanism [38]. By leveraging the training mechanism of a diffusion model with input EEG signals, recent works [38] have been able to show compelling results.

D. Relationship Between EEG Signals and Brain Structure and Topology

The idea that there exists a strong connection between EEG signals and brain topology has been variously explored in the past [39]. In particular, recent work employing topological

data analysis (TDA) to EEG analysis reveals specific disease signatures of several neurological afflictions: ADHD [40] and epilepsy [40]. In addition, brain structural information also provides a link with various learning [41] and general cognitive states of the brain during specific tasks. We note that our central assumption, that the topology and structure of the brain manifests itself in the EEG signal, has also been confirmed in several studies [42] [43]. This neuroscientific basis is the foundation upon which we conceived our reasoning and design of our deep EEG encoding model.

III. Methods

A. Dataset And Data Preprocessing

We used a total of four data sets to train and test our encoder, including three motion imagination data sets and one picture stimulus data set.

- **BNCI 2014-001 Motor Imagery dataset**: This data set consists of EEG data from 9 subjects. The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). We take 50% of the entire data set for training and the other 50% for testing. For this dataset, we used the raw, unfiltered data set to input an epoch into the neural network from 0.5 seconds before the start of the experiment to the end of the experiment, for a total of 4.5 seconds, 1125 sample data.
- **High-gamma dataset**: The "High-Gamma Dataset" is a 128-electrode dataset, which was later reduced to 44 sensors covering the motor cortex, obtained from 14 healthy subjects. The four classes of movements were movements of either the left hand, the right hand, both feet, and rest. For this dataset, we used the raw, unfiltered data set to input an epoch into the neural network from 0.5 seconds before the start of the experiment to the end of the experiment, for a total of 4.5 seconds with 2250 sample data.
- **BNCI 2015-001 Motor Imagery dataset**: In this dataset, the task for the user was to perform sustained right hand versus both feet movement imagery. We take 50% of the entire data set for training and the other 50% for testing. For this dataset, we used the raw, unfiltered data set to input an epoch into the neural network from 0.5 seconds before the start of the experiment to the end of the experiment, for a total of 5.5 seconds, 2816 sample data.
- **Object Dataset**: The dataset consists of 6 classes, each with 12 images that are shown to 10 participants, and EEG signals are recorded using a 128 channel device. The 6 classes include a human body (HB), human face (HF), animal body (AB), animal face (AF), fruit vegetable (FV), and an inanimate object (IO). After pre-processing, the final EEG data is of size 124×5184. We used the first 4000 experiments as the training set and the next 1184 experiments as the test set, in which 10% of the training set was used. For this data set, we use the original data set without filtering processing, from the beginning of the experiment to the end of the experiment as an epoch, a total of 0.5 seconds, 32 sample data into the neural network.

B. Encoder

The proposed deep encoder TFSICNet mainly consists of four blocks: frequency domain attention block (FA), graph Convolution block (GC), Convolution block (CV) and ATCNet block. This particular composition of our model reflects our model's design principles, which have been alluded to in the first section. The full diagrammatic representation of our model is shown in Figure 2

1) Frequency Attention Block: the **FA Block** is a deep module designed to enhance the feature representation of an input signal by applying attention mechanisms over the frequency domain. The module takes a multi-dimensional tensor x as input and converts it to the frequency domain through a series of operations, computes attention weights, performs weighting, and finally returns to the time domain via an inverse FFT transform.

- **Fourier transform to the frequency domain** The input signal x is converted to the frequency domain by a fast Fourier transform (FFT)

$$X = \text{FFT}(x) = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn} \tag{1}$$

  Where X is the frequency domain representation, $X[n]$is the time domain signal, N is the length of the signal, k is the frequency index, and $j$is the imaginary unit.
- **Calculation of the attention weight** We initialize three learnable parameter matrices: **Query**, **Key** and **Value**, all of which have dimensions of (C, F), where C is the number of channels and F is the size of the frequency dimension. We extend **Query**, **Key**, and **Value** to dimensions that match the **batch_size**. The attention logits are obtained via matrix multiplication of the transpose of **Query** and **Key** :

$$\text{Attn\_logits} = \text{Query} \cdot \text{Key}^\top \tag{2}$$

  The attention weights are obtained from applying the softmax function to Attn_logits :

$$\text{Attn\_weights}[i, j] = \frac{\exp(\text{Attn\_logits}[i, j])}{\sum_k \exp(\text{Attn\_logits}[i, k])} \tag{3}$$

  Where i and j represent indexes for channels and frequencies, respectively.
- **Attention weighting** Weighted **Value** with the calculated attention weight and multiplied with the original Fourier transform result X :

$$\text{Attn\_output} = \text{Attn\_weights} \cdot \text{Value} \odot X \tag{4}$$

- **Inverse Fourier transform back to time domain** Finally, the weighted frequency domain signal is converted

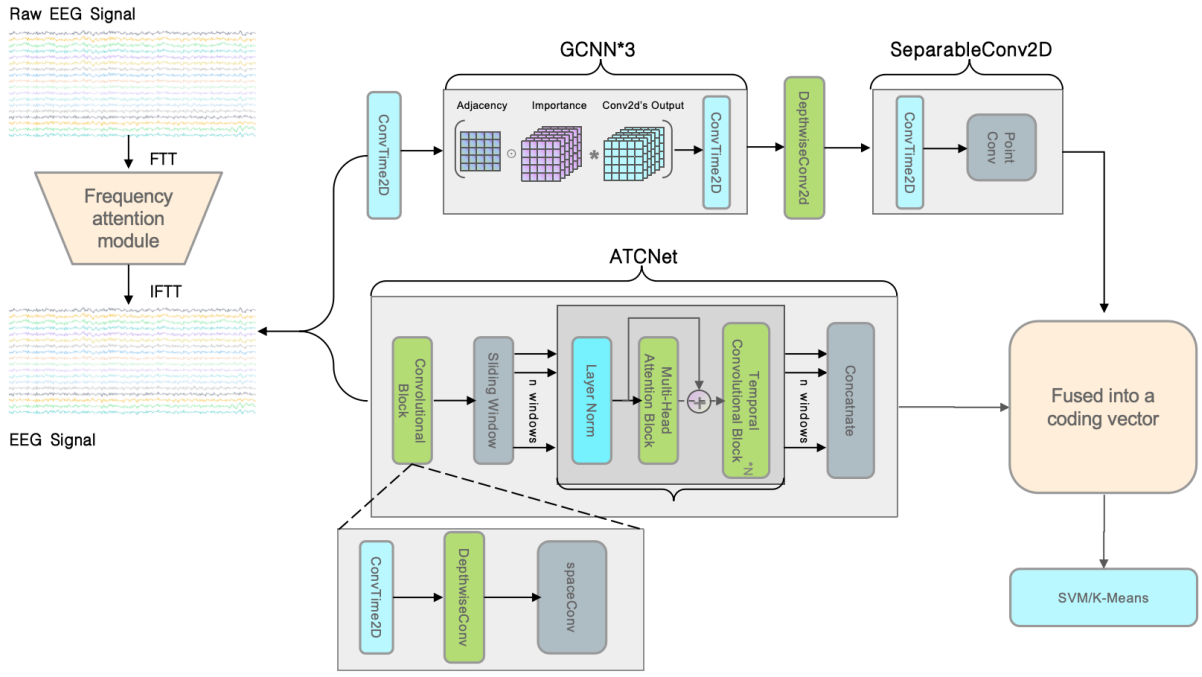

Fig. 2. Full setup of TFSICNet. The various consisting submodules are explained in detail in the text.

back to the time domain by inverse fast Fourier Transform (iFFT) :

$$x_{\text{out}} = \text{iFFT}(\text{Attn\_output}) = \frac{1}{N}\sum_{k=0}^{N-1} \text{Attn\_output}[k] e^{j\frac{2\pi}{N}kn} \tag{5}$$

A figure illustrating the frequency domain workflow is shown in Figure 3.

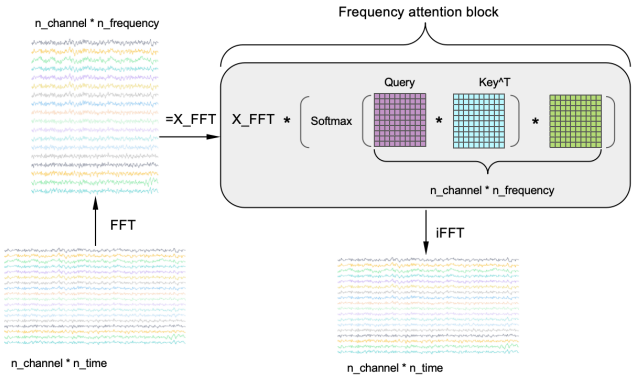

Fig. 3. Frequency domain attention block schematic

2) Graph Convolution Block: In the GCNN, given the graph $G = (V, E)$, where V represents the set of nodes, and E represents the set of edges between nodes in E, V data can be used on the matrix $X \in \mathbb{R}^{n\times d}$, where $n = |V|$ and d the input feature dimension. The edge set E can be used to construct the the weighted adjacency matrix $A \in \mathbb{R}^{n\times n}$, where $A_{ii} = 1$, $i = 1, 2, \ldots, n$. The data in other positions of the matrix are Pearson correlation coefficients. On the other hand, the GCNN learns elements of the matrix $I \in \mathbb{R}^{n\times n}$, namely the importance matrix. Between adjacent layers of GNN, the feature transformation can be written as:

$$Z^{l+1} = Z^l(A \odot I) \tag{6}$$

where $l = 0, 1, \ldots, l-1, l$ and $Z^0 = X$, $H^L = Z$, I to learn the importance of the matrix.

From this, we establish an correlation adjacency matrix with a size of $[N, N]$ as shown in Figure 4, where the value N represents the number of sensor channels. Each value of r_{ij} in the matrix represents the correlation between the first i and j EEG channels. Since we regard the EEG channels as an undirected graph, the adjacency matrix is symmetric with a unit diagonal. As shown in Eq. 6, I initializes a square coefficient matrix in the network that can be updated as the network learns, with its initial values randomly drawn from the standard normal distribution. The A is a learnable prefactor that adjusts the importance between different nodes so that the adjacency matrix becomes fully adaptive according to the input data. By making I a learnable parameter, the model adaptively adjusts the structure of the graph during training. An illustration on the relationship between brain connectome and the setup of our GC module (in form of the adjacency matrix) is depicted in Figure 4.

3) Convolutional And ATCNet Block: In the convolution block, we first use the convolution kernel of size $[N, 1]$ for convolution to encode the relationship between different channels of data, where N is less than the number of channels. After that, convolutional kernel with size $[1, M]$ is used for convolution to extract sequential sequence features for

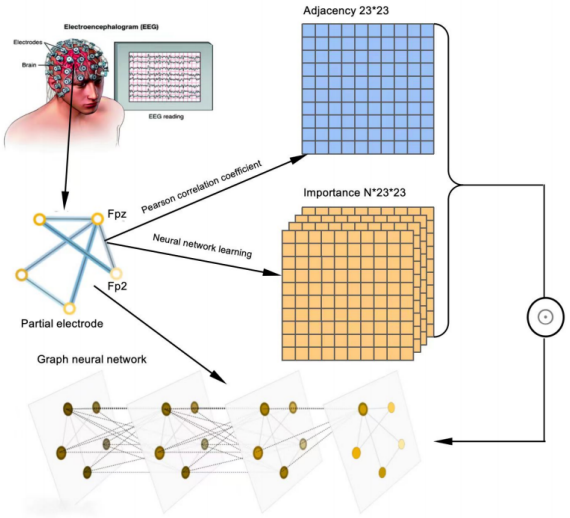

Fig. 4. The relationship between connectomic information from EEG traces and the adjacency matrix in TFSICNet

encoding. The size of M is different for different data sets, and is usually set to $1/4$ of the sampling rate, so that information above 4Hz can be extracted for encoding. Finally, separable convolution, consisting of a deep convolution and a point convolution, separates learning how to summarize individual feature maps in a timely manner (deep convolution) from learning how to best combine feature maps (point-to-point convolution).

For the ATCNet block, we remove the last classification layer and Softmax layer and carry out dimension transformation to better integrate with the encoding vector output from the other end.

C. Experiments

For our numerical evaluations, we performed embedding of preprocessed EEG information from the four datasets considered. The obtained embeddings were then used in classification tasks, as well as visualization via TSNE and kMeans clustering. We compared the performance of

- ATCNet [44]: ATCNet is an attention-based framework for motor imagery classification. The framework consists of self-attention, temporal convolution submodules for feature extraction, and includes a convolutional-layer-based sliding window for data augmentation;
- EEGNet [45]: the EEGNet is a classic deep model for EEG tasks. The backbone of this model are convolutional layers which are organized serially in such a way to enhance generalizability across different BCI paradigms;
- EEG-ITNet [46]: this model is also based on the Inception framework, augmented by causal convolutions with dilations. The authors of this paper have also developed a visualization method for more effective EEG analysis result interpretation;
- EEGConformer [47]: the EEGConformer is a novel architecture based on convolutional transformer submodules, in which the convolutional layers learn the low-level local features, while the self-attention modules extracts the global correlations within the local features;
- LSTM: LSTM is a special type of recurrent neural network with the ability to remember long sequences and learn long-term dependencies. It is widely used for tasks dealing with sequential data, such as natural language processing and time series prediction;
- MUL-LSTM: MUL-LSTM is an improved version of LSTM, which enhances the model's ability to process multivariate time series data. It is suitable for multi-variable sequence analysis and improves the prediction accuracy and stability. MUL-LSTM performs well in complex time series prediction tasks;
- CNN: CNN is a basic convolutional neural network used for tasks such as image recognition. It uses convolutional check to extract input features and makes recognition prediction through a fully connected network. CNN is one of the most important models in the field of deep learning because of its high efficiency and flexibility;

IV. Results and Discussions

A. Performance Metrics

We evaluate the encoder proposed in this paper via the metrics accuracy (ACC) and the Calinski-Harabasz (CH) score. ACC is defined conventionally

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n} \mathrm{TP}_i / I_i}{n} \tag{8}$$

where TP_i is the true positive, i.e., the number of correctly predicted samples in class i, l_i is the number of samples in class i, and n indicates the number of classes. The CH score is a metric which quantifies the quality of clustering of a set of points into distinct clusters; a higher CH score indicates better clustering. The CH score is conventionally expressed in the form

$$CH = \frac{(BSS/(k-1))}{(WSS/(n-k))} \tag{9}$$

Here BSS measures the variance between clusters, representing the degree of dispersion between different clusters. WSS is the variance within a cluster, representing the degree of dispersion of data points within each cluster. Finally, k is the number of clusters, that is, the number of categories into which the clustering algorithm divides the data; and n is the total number of samples participating in the cluster analysis.

B. Training Procedure

The model was trained and tested on a single GPU (RTX 4090 24GB) using the Pytorch and *braindecode* frameworks. For all experiments, we used the following training configuration: the training was done using the Adam optimizer and the cross-entropy loss function with a learning rate of 0.0001, a batch size of 64, and training for 1000 epochs.

TABLE I
THE ACCURACY OF DATA ENCODING BY DIFFERENT ENCODERS AND TFSICNET AND CLASSIFICATION BY SVM ARE COMPARED IN THE TABLE

Dataset	Subject	TFSICNet	ATCNet	EEGNet	EEGITNet	EEGConformer	CNN	LSTM	Mul-LSTM
BNCI 2014-001	Avg	**79.51%**	74.68%	71.24%	73.64%	58.85%	41.68%	37.60%	40.91%
High-gamma	Avg	**91.88%**	88.75%	82.00%	82.50%	82.75%	59.75%	55.25%	56.87%
BNCI 2015-001	Avg	**81.00%**	77.04%	79.33%	65.95%	75.66%	56.75%	56.45%	56.00%
Object	Avg	**70.71%**	64.74%	64.40%	59.51%	52.20%	49.08%	39.05%	39.13%

TABLE II
THE CALINSKI-HARABASZ SCORES OBTAINED AFTER DATA ENCODING BY DIFFERENT ENCODERS AND TFSICNET, T-SNE DIMENSIONALITY REDUCTION AND K-MEANS CLUSTERING ARE COMPARED IN THE TABLE

Dataset	TFSICNet	ATCNet	EEGNet	EEGITNet	EEGConformer	CNN	LSTM	Mul-LSTM
BNCI 2014-001	**372**	351	313	312	338	295	280	329
High-gamma	**319**	306	298	269	191	256	250	253
BNCI 2015-001	1113	1046	844	780	**1342**	457	359	321
Object	**2104**	**2104**	1731	1945	1680	1471	1361	1514

TABLE III
ABLATION ANALYSIS OF TFSICNET WAS PERFORMED ON FOUR DATASETS

Dataset	Not Removed		Removed FA		Removed CV		Removed GC		Removed ATCNet	
	SVM	K-means	SVM	K-means	SVM	K-means	SVM	K-means	SVM	K-means
1-11 BNCI 2014-001	**79.51%**	**372**	77.51%	342	79.00%	352	78.01%	355	74.31%	264
High-gamma	**91.88%**	**319**	89.06%	304	90.64%	310	88.01%	288	84.72%	355
BNCI 2015-001	**81.00%**	**1113**	79.22%	1046	80.72%	1099	80.55%	1108	77.49%	784
Object	**70.71%**	**2104**	69.32%	1966	69.59%	2088	68.04%	1842	66.23%	1465

C. Comparison with Baseline Models

In Table I we show the averaged and SVM input accuracies of EEG data encoded by the proposed TFSICNet encoder using BNCI 2014-001, High-gamma, BNCI 2015-001 and Object datasets. We performed baseline comparisons with EEGNet, EEGITNet, ATCNet, EEGConformer, LSTM, Mul-LSTM, and CNN; some of these models are deep encoders and some are deep network models that perform well in classification tasks. The results of the replicated models are based on the hyperparameters identified in the original article, while the preprocessing, training, and evaluation follow the same procedures defined in this article. Table I shows that in all the four data sets, TFSICNet performs better than all the baseline models, with a significant lead, which indicates that the basis of our hypothesis on the important contributory factors to EEG information content is sound. The average confusion matrix of TFSICNet is shown in Figure 5.

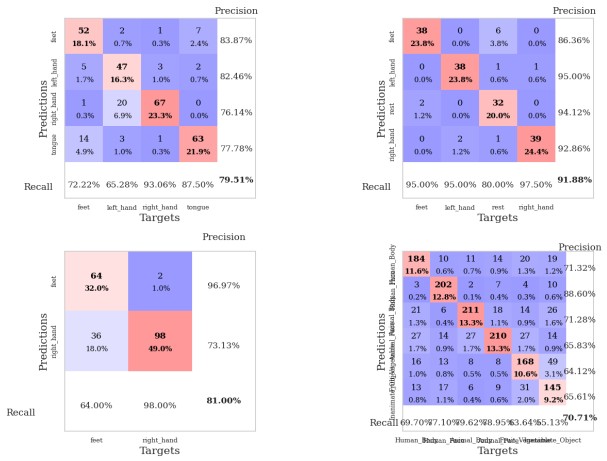

Fig. 5. From left to right and from top to bottom are the accuracy of BNCI 2014-001, High-gamma, BNCI 2015-001 and Object data sets after being encoded by TFSICNet encoder and input SVM for classification

D. TSNE dimension reduction and K-means clustering

Table II summarizes of the CH scores for TFSICNet as compared to the baselines. This experiment was performed on all considered datasets pairs. The CH scores were obtained by first encoding the EEG data and performing TSNE dimensionality reduction, followed by a KMeans procedure. As mentioned above, a higher CH score implied a higher-quality clustering, which is essential for high-quality reconstructions using the computed embeddings. Again, we compared with all our reproduced baselines, the results for which are based on the hyperparameters identified in the original articles, while the preprocessing, training, and evaluation follow the same procedures defined in this article. Table II shows that TFSICNet scored a lower CH than EEGConformer on the BNCI 2015-001 data set, but otherwise builds a large lead over the rest of the baseline models, which again validates our design principles. A full visualization of the TSNE and

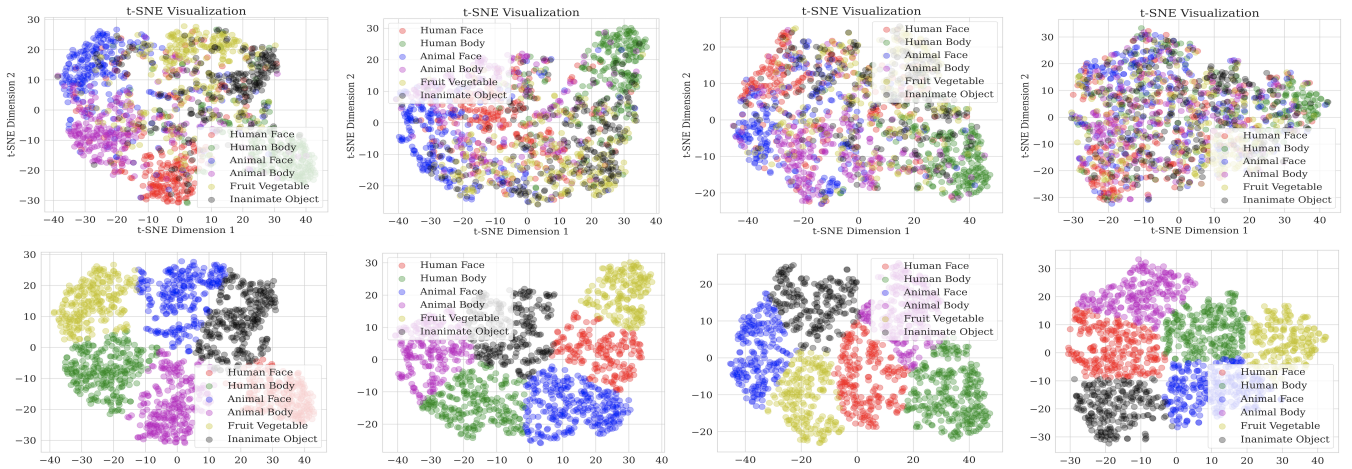

Fig. 6. The first row, from left to right, shows TSNE results from TFSICNet, EEGITnet, EEGNet and MUL-LSTM on the Object dataset. The next row, from left to right, shows additional K-means visualization on the TSNE figures.

k-Means results are shown in Figure 6. From the four visual datasets in use in this work, we selected the Object dataset for full visualization. Evaluation on all models were done and the corresponding results plotted.

E. Ablation analysis

We conducted ablation studies to highlight the role of different modules in achieving the classification results we reported. The overall ablation results reported in Table III showed a decrease in both accuracy and CH scores when we respectively removed four different submodules from our framework; the overall effect strongly indicates the correctness of our model design principles when translated into practical results. We noted that the EEG channels of BNCI 2014-001, High-gamma, BNCI 2015-001 and Object data in the four datasets were 22,128,13 and 124, respectively, and the performance of the more EEG channels after the deletion of GC blocks was more affected. Deletion of the ATCNet block had the greatest effect on performance, indicating that the importance of time series and correlation play an important role in EEG data.

V. Conclusion

The quality and informativeness of encoded EEG data strongly affects the viability of brain-computer interfaces (BCIs). EEG (electroencephalogram) is a type of electrical brain signal collected by special electrodes from the activity of neurons in the cerebral cortex, which can reflect the activity pattern of the brain. In BCIs, EEG data plays a key role as a bridge between the human brain and external devices. Deep learning methods have gained widespread application in many areas of scientific research in recent years, in part because of its ability to automatically extract the most relevant features and provide extensive multimodal datasets for almost all fields. In this work, we take advantage of this characteristic learning ability of deep learning and incorporate important prior knowledge regarding structural and connectivity information of the human brain into our design principles. Based on these considerations, we designed a new model, TFSICNet, to be a robust, complete and informative encoder of EEG signals. It extracts the importance and correlation features of EEG signals in time, frequency and spatial domains, respectively. The comparison with the baseline model shows SOTA performance against four current models and three traditional models, each based on a variety of different learning structures. We evaluate the performance of our encoder on conventional classification task and the intrinsic separability of the encoded vectors via the Calinski-Harabasz (CH) score obtained by K-means clustering after dimensionality reduction of coding vectors by TSNE. Finally, the ablation experiments show the validity of our design principles from different angles. This work hints at the possibility of building robust and accurate models of motor imagery and visual stimulus, based on sound design principles, built upon knowledge and understanding of prior medical information and data.

References

[1] X. Chen, R. Wang, A. Khalilian-Gourtani, L. Yu, P. Dugan, D. Friedman, W. Doyle, O. Devinsky, Y. Wang, and A. Flinker, "A neural speech decoding framework leveraging deep learning and speech synthesis," *Nature Machine Intelligence*, vol. 6, no. 4, pp. 467–480, apr 2024. [Online]. Available: https://doi.org/10.1038/s42256-024-00824-8

[2] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, apr 2019. [Online]. Available: https://doi.org/10.1038/s41586-019-1119-1

[3] Y. Yang, Y. Duan, Q. Zhang, H. Jo, J. Zhou, W. H. Lee, R. Xu, and H. Xiong, "Neuspeech: Decode neural signal as speech," 2024.

[4] N. Koide-Majima, S. Nishimoto, and K. Majima, "Mental image reconstruction from human brain activity: Neural decoding of mental imagery via deep neural network-based bayesian estimation," *Neural Networks*, vol. 170, pp. 349–363, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608023006470

[5] Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, "Decoding natural images from eeg for object recognition," 2024.

[6] T. Dado, Y. Güçlütürk, L. Ambrogioni, G. Ras, S. Bosch, M. van Gerven, and U. Güçlü, "Hyperrealistic neural decoding for reconstructing faces from fmri activations via the gan latent space," *Scientific Reports*, vol. 12, no. 1, p. 141, jan 2022. [Online]. Available: https://doi.org/10.1038/s41598-021-03938-w

[7] M. Kim, M. S. Choi, G. R. Jang, J. H. Bae, and H. S. Park, "Eeg-controlled tele-grasping for undefined objects," *Frontiers in Neurorobotics*, vol. 17, p. 1293878, dec 2023.

[8] O. P. Idowu, P. Fang, X. Li, Z. Xia, J. Xiong, and G. Li, "Towards control of eeg-based robotic arm using deep learning via stacked sparse autoencoder," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018, pp. 1053–1057.

[9] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Scientific Reports*, vol. 6, p. 38565, dec 2016. [Online]. Available: https://doi.org/10.1038/srep38565

[10] J. C. Roddey, B. Girish, and J. P. Miller, "Assessing the performance of neural encoding models in the presence of noise," *Journal of Computational Neuroscience*, vol. 8, no. 2, pp. 95–112, 2000.

[11] H. Gast and Y. Assaf, "Weighting the structural connectome: Exploring its impact on network properties and predicting cognitive performance in the human brain," *Network Neuroscience*, vol. 8, no. 1, pp. 119–137, 04 2024. [Online]. Available: https://doi.org/10.1162/netn_a_00342

[12] S. H. Bennett, A. J. Kirby, and G. T. Finnerty, "Rewiring the connectome: Evidence and effects," *Neuroscience Biobehavioral Reviews*, vol. 88, pp. 51–62, May 2018, epub 2018 Mar 11.

[13] M. P. van den Heuvel and O. Sporns, "A cross-disorder connectome landscape of brain dysconnectivity," *Nature Reviews Neuroscience*, vol. 20, no. 7, pp. 435–446, Jul 2019.

[14] M. Kaiser, "The potential of the human connectome as a biomarker of brain disease," *Frontiers in Human Neuroscience*, vol. 7, 2013. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2013.00484

[15] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in Neuroinformatics*, vol. 3, 2009. [Online]. Available: https://www.frontiersin.org/articles/10.3389/neuro.11.037.2009

[16] E. T. Rolls, "40Hierarchical organization," in *Cerebral Cortex: Principles of Operation*. Oxford University Press, 08 2016. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780198784852.003.0002

[17] T. Serre, *Hierarchical Models of the Visual System*. New York, NY: Springer New York, 2013, pp. 1–12.

[18] G. K. Steinke and R. F. Galán, "Brain rhythms reveal a hierarchical network organization," *PLOS Computational Biology*, vol. 7, no. 10, pp. 1–15, 10 2011. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1002207

[19] G. Chiarion, L. Sparacino, Y. Antonacci, L. Faes, and L. Mesin, "Connectivity analysis in eeg data: A tutorial review of the state of the art and emerging trends," *Bioengineering*, vol. 10, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2306-5354/10/3/372

[20] K. Seeliger, M. Fritsche, U. Güçlü, S. Schoenmakers, J.-M. Schoffelen, S. E. Bosch, and M. Van Gerven, "Convolutional neural network-based encoding and decoding of visual object recognition in space and time," *NeuroImage*, vol. 180, pp. 253–266, 2018.

[21] S. Chandrasekaran, M. Fifer, S. Bickel, L. Osborn, J. Herrero, Y. Kim, G. Watrous, D. McMullen, J. Shupe, Z. S. Tan, O. Gonzalez, S. N. Flesher, M. L. Boninger, J. L. Collinger, R. A. Gaunt, S. J. Bensmaia, M. Afzal, A. Rastogi, J. S. Jennings, and D. W. Moran, "Historical perspectives, challenges, and future directions of implantable brain-computer interfaces for sensorimotor applications," *Bioelectronic Medicine*, vol. 7, no. 1, p. 14, 2021.

[22] A. Kawala-Sterniuk, N. Browarska, A. Al-Bakri, M. Pelc, J. Zygarlicki, M. Sidikova, R. Martinek, and E. J. Gorzelanczyk, "Summary of over fifty years with brain-computer interfaces-a review," *Brain Sciences*, vol. 11, no. 1, p. 43, 2021.

[23] H. Berger, "Über das elektrenkephalogramm des menschen," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, 1929.

[24] J. Kamiya, "Conscious control of brain waves," *Psychology Today*, vol. 1, pp. 56–60, 1968.

[25] J. J. Vidal, "Toward direct brain-computer communication," *Annual Review of Biophysics and Bioengineering*, vol. 2, pp. 157–180, 1973.

[26] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[27] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An eeg-based brain-computer interface for cursor control," *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 3, pp. 252–259, 1991.

[28] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interface: a new communication device for handicapped persons," *Journal of Microcomputer Applications*, vol. 16, pp. 293–299, 1993. [Online]. Available: https://api.semanticscholar.org/CorpusID:62582338

[29] A. J. Anderson, B. D. Zinszer, and R. D. Raizada, "Representational similarity encoding for fmri: Pattern-based synthesis to predict brain activity using stimulus-model-similarities," *NeuroImage*, vol. 128, pp. 44–53, 2016.

[30] E. Hernandez and J. Andreas, "The low-dimensional linear geometry of contextualized word representations," *arXiv preprint arXiv:2105.07109*, 2021.

[31] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," *npj Computational Materials*, vol. 2, no. 1, pp. 1–7, 2016.

[32] F. Farahnakian and J. Heikkonen, "A deep auto-encoder based approach for intrusion detection system," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2018, pp. 178–183.

[33] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[34] B. Ghoraani and S. Krishnan, "Time–frequency matrix feature extraction and classification of environmental audio signals," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 7, pp. 2197–2209, 2011.

[35] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 31–40, 2016.

[36] A. Realo, J. Allik, A. Nõlvak, R. Valk, T. Ruus, M. Schmidt, and T. Eilola, "Mind-reading ability: Beliefs and performance," *Journal of Research in Personality*, vol. 37, no. 5, pp. 420–445, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092656603000217

[37] Q. Zhou, C. Du, D. Li, H. Wang, J. K. Liu, and H. He, "Neural encoding and decoding with a flow-based invertible generative model," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.

[38] Y. Takagi and S. Nishimoto, "High-resolution image reconstruction with latent diffusion models from human brain activity," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 453–14 463.

[39] F. Altındiş, B. Yılmaz, S. Borisenok, and K. İçöz, "Parameter investigation of topological data analysis for eeg signals," *Biomedical Signal Processing and Control*, vol. 63, p. 102196, 2021.

[40] A. B. El-Yaagoubi, M. K. Chung, and H. Ombao, "Topological data analysis for multivariate time series data," *Entropy*, vol. 25, no. 11, p. 1509, 2023.

[41] H.-J. Park and K. Friston, "Structural and functional brain networks: from connections to cognition," *Science*, vol. 342, no. 6158, p. 1238411, 2013.

[42] E. C. van Straaten and C. J. Stam, "Structure out of chaos: functional brain network analysis with eeg, meg, and functional mri," *European Neuropsychopharmacology*, vol. 23, no. 1, pp. 7–18, 2013.

[43] K. Smith and J. Escudero, "The complex hierarchical topology of eeg functional connectivity," *Journal of Neuroscience Methods*, vol. 276, pp. 1–12, 2017.

[44] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for eeg-based motor imagery classification," *IEEE transactions on industrial informatics*, vol. 19, no. 2, pp. 2249–2258, 2022.

[45] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[46] A. Salami, J. Andreu-Perez, and H. Gillmeister, "Eeg-itnet: An explainable inception temporal convolutional network for motor imagery classification," *IEEE Access*, vol. 10, pp. 36 672–36 685, 2022.

[47] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2022.