2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Self-Selecting Semi-Supervised Transformer-Attention Convolutional Network for Four Class EEG-Based Motor Imagery Decoding

Han Wei Ng and Cuntai Guan
Faculty of Computer Science and Engineering, Nanyang Technological University
50 Nanyang Ave, 639798

Abstract— **Brain-computer interfaces (BCI) serve as an important tool in areas such as neurorehabilitation and constructing prostheses. Electroencephalogram (EEG) motor imagery (MI) signal is a common method used to communicate between the human brain and the computer interface. However, differentiating between multiple motor imagery signals may be challenging due to the presence of high noise-to-signal ratio and small dataset sizes. In this study, we propose a variational autoencoder and transformer-attention based convolutional neural network (SSTACNet) for multi-class EEG-based motor imagery classification. The SSTACNet model leverages upon variational autoencoders' ability to measure the contrastive distance between two sets of inputs to perform data self-selection. The model further utilizes multi-head self-attention as well as spatial and temporal convolutional filters to achieve superior extraction of signal features. The model additionally utilizes the variational autoencoder's ability to augment the dataset with feature-informed pseudo-data, achieving stronger classification results. The proposed model outperforms the current state-of-the-art techniques in the BCI Competition IV-2a dataset with an accuracy of 85.52% and 70.56% for the subject-dependent and subject-independent modes, respectively. Codes may be found at: https://github.com/NgHanWei/SSTACNet**

I. Introduction

Recent advances in brain-computer interfaces (BCI) have shown much promise in the areas of rehabilitation [1]. In brain-computer interfaces (BCI), electroencephalogram (EEG) signals are often used to communicate between the human brain and the computer due to its non-invasive nature. However, EEG signals suffer from high noise-to-signal ratio [2], causing the signals for different classes to be difficult to differentiate among one another. To overcome this challenge, many deep learning methodologies have been implemented to extract useful discriminative features that improve the classification accuracy in BCI applications [3], [4].

Previous work done by studies showed that the use of attention-based convolutional neural networks was able to achieve state-of-the-art performance for subject-dependent decoding of motor imagery [5].

However, the main limitations behind the existing methodologies lie in that the collected EEG trials used to train the neural networks contain significant amounts of noise [2]. Furthermore, given the abstract nature of the signals collected, it is difficult to determine whether the collected signals are valid. Thus, there could be anomalous trials among the training set that leads to poorer model outcomes.

Another limitation is that most conventional neural networks fail to consider the long-range correlations that the EEG signals might have across the entire trial period [6]. Common neural networks are sufficient in capturing the short-term dependencies, however certain information may be lost due to the lack of capturing these long-range dependencies which exist throughout the trial period from resting state to the execution of the motor imagery.

Thus, in this work we introduce the incorporation of transformer-based architecture [7] into spatial-temporal neural networks to fully utilize the effectiveness of capturing long-term features and dependencies in the EEG signal.

Additionally, neural networks are known to require large amounts of data for the classifier to become a good generalizer [3]. This is especially so in the case of EEG-based motor imagery classification, whereby due to significant inter- [8] and intra-subject variability [9], small datasets will suffer from performance deterioration of the network as the data populations have few feature overlaps with the target subject, resulting in confusion of the network. To overcome the issue of small datasets, different methodologies were suggested such as the usage of multiple different datasets [8], as well as augmentation techniques [10] to increase the number of datapoints used to train the classifier network.

Hence, we propose the use of self-selective semi-supervised transformer-attention convolutional network (SSTACNet) to tackle multi-class motor imagery decoding even with the lack of sufficient data. This is done by implementing variational autoencoder (VAE) self-selective semi-supervised learning into the neural network. The network is updated using a series of pseudo-labelled unlabelled trials. To further enhance the effectiveness of the pseudo-labelled dataset, the proposed framework utilizes a VAE network to discard EEG trials that do not fit well across the population of data while augmenting the dataset.

Our main contributions are as follows: (1) Introduction of a variational autoencoder to perform automatic data self-selection. (2) The combined use of semi-supervised learning to achieve a larger homogeneous dataset. (3) Application of transformer-attention mechanism on the features followed by the use of convolutional sliding windows on the transformed feature sequence.

II. Dataset

For this study, we use the BCI Competition IV 2a dataset [11] to examine the efficacy of the proposed method in discriminating multi-class motor imagery EEG signals. The

979-8-3503-7769-9/24/$31.00 ©2024 IEEE

dataset consists of 9 healthy individual 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). Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

III. Methodology

The proposed SSTACNet model framework (Fig. 1) consists of two main network blocks, a contrastive VAE model as well as a transformer based spatial-temporal convolutional neural network. The two blocks are first optimized separately, but are subsequently joined together to form a single pipeline to achieve motor imagery classification.

The first block consists of the neural network's contrastive learning ability as defined by the variational autoencoder network (Fig. 2). The benefit of using the autoencoder architecture lies in that it is able to extract latent features of the data without the requirement of target labels. Due to this property, autoencoders have shown great success in areas of unsupervised learning, semi-supervised learning and supervised learning. In this work, we utilize the VAE network to compare the distributions between the various datapoints, removing any EEG trial that may be too distant from the general population of data.

The second block comprises the sections responsible for motor imagery signal classification. Spatial and temporal convolutional filters are utilized similar to conventional methods such as DeepConvNet and EEGNet. In addition to these filters, we further propose the addition of transformer-attention sliding-window layers to increase the model's ability to capture signal correlations across longer time frames.

A. Data Self-Selection via Constrastive Learning

In this block (Fig. 2), a variational autoencoder consisting of spatial and temporal convolutional filters are used to extract the relevant features across the trials. The VAE model is first used for extracting the relevant features of each individual trial to be compared against the known EEG data population.

Traditional VAEs consist of probabilistic encoder-decoder pairs. For a given input x, the encoder is an inference model with weights and biases θ which gives the hidden latent variables as output z. The inference model is thus given by $q_\Theta(z|x)$, a Gaussian probability distribution. For the same VAE, a decoder model with weights and biases ϕ is given by a joint probability $p_\Phi(x,z) = p_\Phi(x|z)p(z)$. During training, the encoder and decoders are trained simultaneously by finding the parameters that best optimise the variational lower bound of the likelihood $p_\phi(x) = \int p_\phi(x,z)dz$.

Thus, the effectiveness of the VAE in reconstructing the original input is given by the reconstruction log-likelihood $logp_\Phi(x|z)$. The reconstruction loss function is therefore given by the expected negative log-likelihood $-\mathbb{E}_{q_\Theta(z|x)}[\log\ p_\Phi(x|z)]$ computed with respect to the distribution of the latent features by the encoder.

In addition to the reconstruction loss, VAEs also take into account a regularisation term given by the Kullback-Leibler (KL) divergence between two continuous distributions [12], the encoder's variational posterior $q_\Theta(z|x)$ and the prior $p(z)$ where the latent variables are sampled from. The divergence measures how close the two distributions q and p are to each other and is given by $\mathbb{KL}(q_\Theta(z|x)||p(z))$. Therefore, the overall loss function L_i for an input datapoint x_i is:

$$L_i = -\mathbb{E}_{q_\Theta(z|x_i)}[\log\ p_\Phi(x_i|z)] + \mathbb{KL}(q_\Theta(z|x_i)||p(z)) \quad (1)$$

The network is trained upon the loss function as defined in equation 1, encouraging the model to optimize towards parameters that maximizes the network's likelihood of regenerating the original input from the latent features computed. Through this, the encoder portion of the network learns to encode useful information from the trials.

As such, when the trial is fed into the model, the network attempts to extract and encode the trial into the latent features based on the hyperparameters that was trained using the rest of the remaining training set. Since the network was trained to reconstruct the training set, signals that deviate further away from the population of the training set will likely display significantly higher loss values and poorer reconstruction outcomes. The trained VAE network is thus able to act as a discriminator to determine whether the incoming signal belongs to the existing population. Therefore, we leverage upon the properties of the trained VAE network to eliminate potential trials that would confuse the motor imagery classifier model. To achieve this, the VAE network is first trained upon the known training data. Following which, the individual EEG trial samples may be fed through the trained model. The contrastive loss of each of the signal may be recorded. The trials with the furthest contrastive loss are finally discarded from the training set.

B. Transformer-Attention Sliding-Window

Transformer-attentions were introduced to overcome the challenge of retaining information across long sequences. Furthermore, transformer-attention are non-sequential, allowing the network to avoid recursion by processing the entire input sequence and learning the relationship between different time points via the use of self-attention heads and positional embeddings. This gives transformer-attention networks an advantage over convolutional networks which are unable to take into consideration the different possible relations between time points due to the practical limitation in the number of kernels and kernel dimensions used.

We further break down the decoding network block into three main parts, the convolutional sliding-window, the transformer-attention architecture and the temporal convolutional network.

The transformer-attention module is applied onto the outputs of the convolutional filters which serves to an produce an output vector for the features sequence with encoded

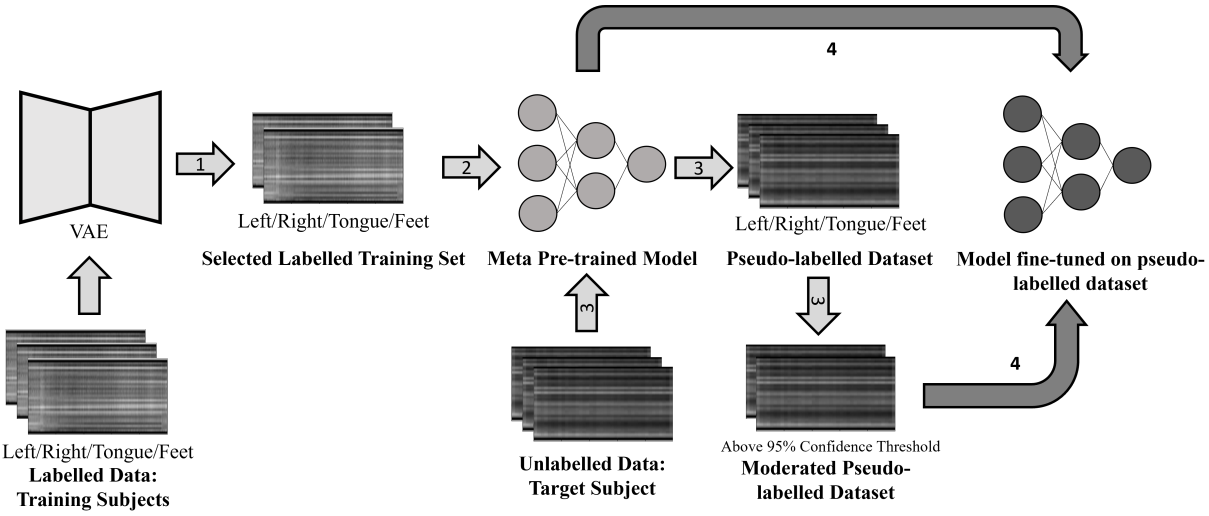

Fig. 1. Model architecture of the proposed SSTACNet. The model consists of three main blocks (1) The variational autoencoder for self-selection of data, (2) Spatial-temporal convolutional filters, (3) Transformer-attention convolutional sliding windows and finally the use of (4) Semi-supervised learning through the self-selected data.

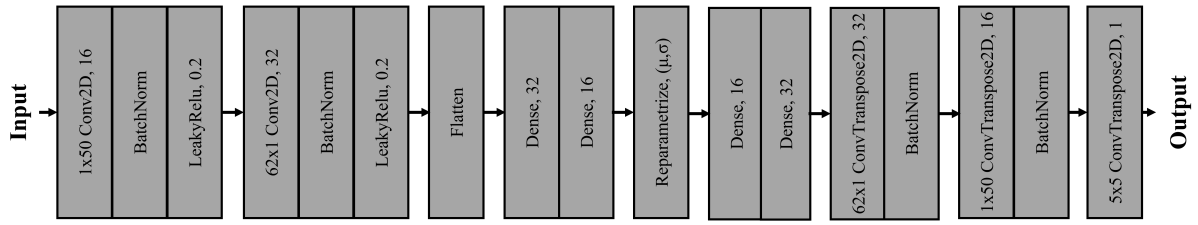

Fig. 2. Overall variational autoencoder model architecture used to extract the relevant features of the data in order to determine the overall distribution of the data and to remove data that would likely confuse the classifier network.

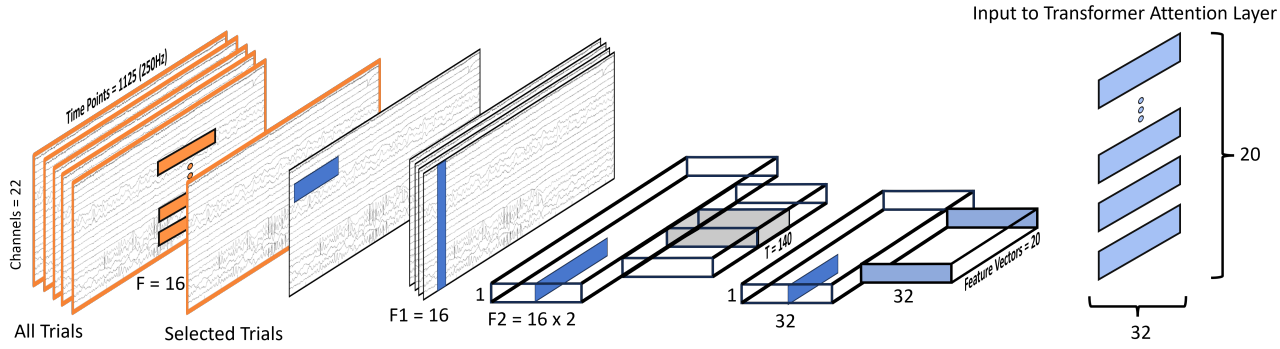

Fig. 3. Overall architecture of the spatial-temporal transformer-attention convolutional network. The network utilizes a variational autoencoder for extraction of meaningful trial data. This is followed by spatial-temporal filters and the transformer-attention mechanism to extract meaningful features. The features are subsequently fed into a convolutional sliding window and residual blocks to eventually obtain the output.

information on how each feature in the sequence should relate to each of the remaining features.

The usage of transformer-attention allows the network to capture the long-range dependencies between the features extracted by the spatial filter. This contrasts with the conventional networks which often rely on local information between features [5], allowing the proposed network to take into account the relationship between features across time. This is especially useful in the instance of motor imagery, given that the users are tasked to imagine the movement across a period of time after the visual stimulation is given. As such, the usage of transformer-attention allows the network to take into consideration how the start, end and relations between different timepoints of the motor imagery may influence the nature of the imagination that the user is current executing.

The transformer-attention architecture utilizes both the scaled dot-product and multi-head attention to improve effectiveness in time-series classification tasks, whereby the Scaled Dot-Product attention is given by:

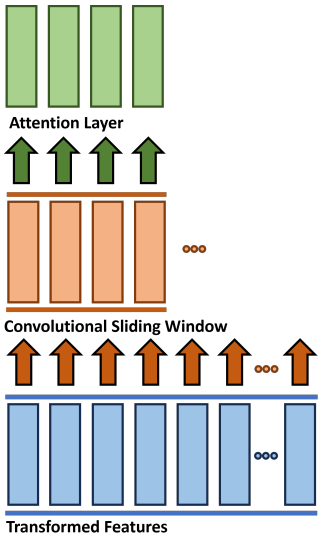

Fig. 4. The usage of convolutional sliding windows on the transformed feature vectors to augment the model network as well as improve model focus on important segments of the feature sequence.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (2)$$

where $\sqrt{d_k}$ is the dimension of the key vector k and query vector q .

The multi-head attention is an extension of the scaled dot-product attention, whereby instead of using a single set of attention weights, it employs multiple attention mechanisms in parallel and is given by:

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O \quad (3)$$

where

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

In addition, transformer attention is highly parallelizable, which makes it computationally efficient and scalable. This is in contrast to recurrent neural networks (RNNs), which process sequences sequentially and can be computationally expensive when training networks to decode for long sequences.

The convolutional-based sliding-window was first suggested by Schirrmeister et. al. [13] and used in the proposed DeepConvNet framework which is widely used to benchmark most EEG-based classification tasks [14], [15]. The convolutional-based sliding-window serves two purposes. The first is to help augment the data [16], given that EEG-based datasets often suffer from a lack of sufficient datapoints. The sliding window helps to increase the number of permutations of data that the model can learn from. Secondly, the sliding window serves to isolate parts of the entire feature sequence, allowing the model to better extract information [17] from only the relevant parts of the sequence while reducing weights in the parts that do not have a strong influence in classification ability.

The sliding window is defined to have a length of T_w with a step stride of one, thus dividing the input temporal feature sequence F_t into multiple window sequences of F_t^w whereby w denotes the window index and $w = 1, ..., n$ with n representing the total number of windows given the input sequence. Thus, the window length T_w may be defined as:

$$T_w = T_f - n + 1, T_f > n \geq 1 \quad (5)$$

Where T_f is the number of output feature vectors, which is influenced by the original input sequence length, as well as the size of the 2 average pooling lengths.

To further improve the effectiveness of the sliding-window mechanism, the transformed sequence of feature vectors are fed into the sliding window as opposed to the raw feature vectors. After the transformer-attention layer has been applied to the feature vectors, an additional layer of attention heads are then applied to the output of the sliding window (Fig. 4) to extract further relevant information pertaining towards the features visible within the sliding window.

C. Temporal Convolutional

The temporal convolutional neural network was first proposed by Ingolfsson et al. [18] as a method for accurate decoding for embedded motor-imagery BCIs. The network utilizes a series of causal convolutions, dilated convolutions and residual blocks to increase the receptive field size of the network to capture longer sequences. The main benefit behind this module lies in its low computational complexity. However, a key point to note is that the module relies on causal convolutions to extract the relationship between one time point to another, which accounts only for the short-term relationship between the inputs. Therefore, the introduction of the transformer-layer network beforehand can immensely benefit the network to recall and draw relationships between the significant features from the previous block.

D. Self-Selected Semi-Supervised Learning

In addition to the proposed network, we introduce the usage of semi-supervised learning of the trained neural network during the inference phase. In semi-supervised learning, the network typically learns and updates its hyperparameters from the additional knowledge obtained from the newly seen unlabelled target trials. Semi-supervised learning takes place by first training a neural network with the existing labelled data. Following which, the network evaluates a series of unlabelled data, creating pairs of associated pseudo-labels. Finally, the network parameters may be further updated by fine-tuning the model on the newly obtained pseudo-labels and respective datapoints.

In the case of EEG-based classification tasks, this gives the network the advantage to reduce the effect of inter-subject variability. Since the EEG signals collected between subjects can have very high deviation from one another, semi-supervised learning offers the benefit to encourage the network to converge towards the new unseen subject. This is

TABLE I

AVERAGE CLASSIFICATION ACCURACY OF VARIOUS METHODOLOGIES FOR FOUR-CLASS EEG-BASED MOTOR IMAGERY CLASSIFICATION. * INDICATES THE SIGNIFICANT DIFFERENCE BETWEEN THE PROPOSED MODEL AGAINST THE OTHER MODELS IN SUBJECT PAIRWISE COMPARISON ($P<0.05$).

Methodology	Accuracy (%)
SCSSP [19]	65.05
Dual Attention Relation [20]	65.65
FBCSP [21]	67.75
BO [22]	68.13
ISMDA [23]	69.51
FBCSP-SVM [24]	71.18
CW-CNN [24]	73.07
DFFN [25]	76.44
MI-DAGSC [26]	76.81
ATCNet [5]	81.10
Semi-Supervised Transformer-Attention Convolutions (Ours)*	**85.52**

especially beneficial towards EEG datasets which suffer from a lack of data. Typically, semi-supervised learning attempts to fine-tune the network to better understand the data patterns of the target data without relying on any explicit data labels or annotations. The algorithm automatically generates supervisory signals from the data, making it a more cost-effective and scalable approach compared to traditional supervised learning methods.

In the case of motor imagery paradigm, we follow the assumption that for each motor imagery, the key features of each class are separable into distinct clusters. The neural network learns the features for each of these clusters and differentiates the signals based on the learnt features. As such, we continue to utilize this assumption when implementing semi-supervised learning to the network.

IV. RESULTS

In this section, we detail the results obtained for both subject-dependent and subject-independent paradigms when implementing the proposed network in this study. When comparing against the previous known work by Altaheri et al. [5], it is noted that the previous study obtained the final accuracy for each subject by using the evaluation set as the validation set and selecting the model which displayed the highest validation accuracy. This results in a data leakage issue which would overstate the performance of the model. Therefore, in this work we re-implement the work by allowing the model to select based on the highest training accuracy, thus removing the issue of data leakage and allowing for realistic and fair comparison.

A. Subject-Dependent

In the subject-dependent paradigm, the data is split into a training and testing set similar to the original BCI Competition IV 2a division. The first session of each subject is used for training, while the second session is used for evaluation. The proposed model is further trained and evaluated using a cross-fold validation strategy, with the number of folds equal to the number of subjects used for training and validation. In using the LOSO training paradigm, the model's performance is evaluated on a subject not seen by the model.

B. Subject-Independent

In the subject-independent paradigm, the neural network is first trained on seven of the nine subjects, with one subject used for validation and the selected subject for evaluation.

In the case of subject-independent performance, this dataset suffers strongly from a lack of sufficient data which is needed to obtain robust decoding performance. Due to the small amount of data, the neural network is unable to extract features that is generalizable towards unseen data.

V. DISCUSSION

A. Overcoming Inter-Session Variability

Between subjects, there exist significant session-to-session variability in the underlying brain state of each subject. This challenge persists across various other EEG datasets [27], causing difficulties in training a good decoding model. Conventionally, to overcome this variability a large amount of prior training data is utilized to optimize the neural network hyperparameters to be able to extract features that are useful for discrimination across a number of subjects. The concept lies in that given a large number of subjects, the neural network aims to maximize the decoding performance across all of the subjects. Any new unseen subject should ideally lie within the prior population of the training set. However, as the BCI Competition IV 2a dataset [11] contains only a relatively small number of subjects, the neural network is unable to fully leverage upon the subject-independent paradigm to train a generalizable motor imagery decoder. The unseen subject in each case would therefore be likely to fall outside of the training data population that the network was optimized on, leading to further degradation of performance.

Furthermore, we perform a qualitative study to display the presence of intra-subject and inter-session variability using the extracted features from the variation autoencoder during the data self-selection phase. The obtained feature values are subsequently plotted across each trial. The resulting figure (Fig. 5) displayed that there exists visible intra-subject variation within each session, with certain trials within each session showing feature values that deviate far from the population norm. In addition, comparing between the two populations of data across sessions (Fig. 5), it is observed that there exists significant inter-sessions variability where each of the population of the features in each session show little overlap. It may be seen that there exists significant differences between the first 40 trials of the second EEG recording session in this particular subject. The deviation is only reduced in the later trials of the session. From this, it may be hypothesized that a discriminator trained on the first session of data would therefore perform poorly during the initial portion of the second session due to inter-session variations, but perform better in the later remaining trials.

B. Subject-wise Analysis

1) Subject-Dependent: Certain subjects such as subject 2 of the dataset showed significantly poorer performance in

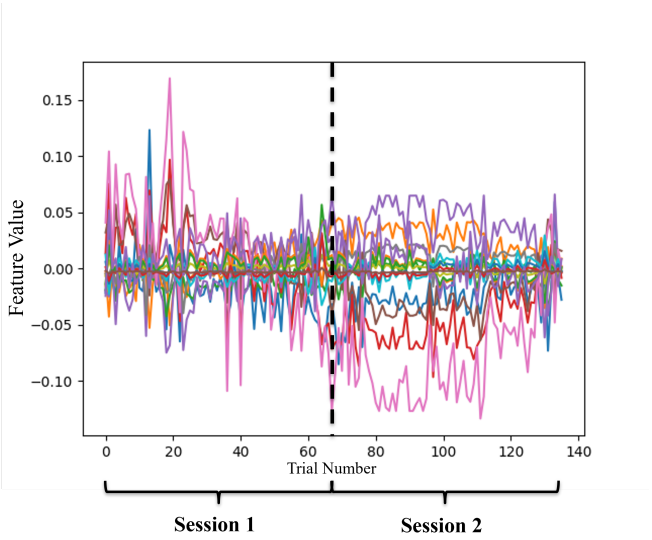

Fig. 5. Learned latent features from a specific subject from the BCI Competition IV 2a Dataset. Differing trends in the features between the first and second sessions indicate strong inter-session variability and potentially poor adaptation results. Each color represents a single feature variable.

Subject	Subject-Dependent Accuracy (%)	Subject-Independent (%)
1	89.17	76.32
2	75.09	55.33
3	94.22	81.95
4	88.09	64.86
5	80.14	65.73
6	73.65	59.83
7	91.34	77.09
8	85.92	78.44
9	92.06	75.49
Average	**85.52**	**70.56**

motor imagery decoding. Observing the confusion matrix, it can be seen that while the decoder was successfully able to discriminate the foot and tongue motor imagery well, the performance was drastically lower in the left and right-hand motor imagery. This suggests that for certain subjects, the activation patterns of each sides of the hand are similar, leading to the decoder being unable to learn accurately key discriminating features between the two hand motor imagery signals. This is observed mainly in subjects 2 and subject 6 of the BCI Competition dataset [11].

Despite having poorer discrimination for the hand movements, good decoding ability for foot and tongue motor imagery was still observed in these subjects. This outcome is expected, given that there are significant differences in the brain activation patterns between the hands against the foot and tongue activation patterns. If the two hand classes were to be combined into a single class, we observe that the accuracy for successfully discriminating a hand movement rises to 71%.

2) Subject-Independent: In the subject-independent case, we observe a lower accuracy in the BCI Competition IV 2a Dataset as compared to the subject-dependent scenario. This is to be expected due to the relatively small size of the dataset. The performance of the subject-independent classifier depends heavily upon the closeness of the distribution

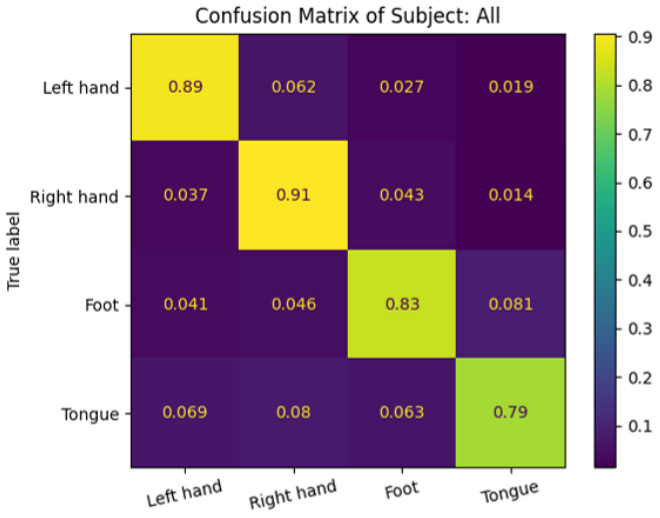

Fig. 6. Confusion matrix across all subjects in the four-class subject-dependent motor imagery dataset for subject-dependent classification. The overall decoding accuracy across all the subjects showed good discrimination across the four classes, with the hand motor imagery performing slightly worse in comparison to the foot and tongue movements.

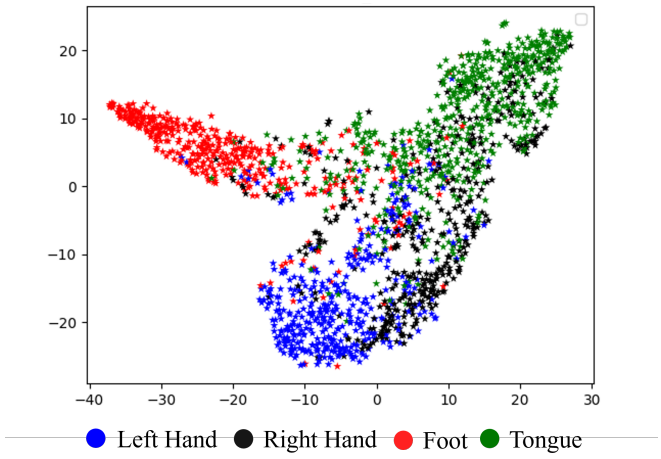

Fig. 7. t-SNE plot of the model's learnt class distribution for a single subject. The colors indicate the true class labels. The right hand class shows the highest confusion, with many of the datapoints being wrongly classified into other classes. The left and right hand classes show a tighter overlapped distribution as compared to the foot and tongue which has more distinct individual clusters.

across each of the individual subjects' brain signals. Given the smaller number of subjects, the disparity is likely to be larger between subjects and any subject with sufficiently different activation patterns would cause a larger confusion to the network as opposed to larger datasets which are more robust to anomalous trials. Furthermore, if the number of trials is too small in comparison to the network complexity, the network is likely to overfit towards the training set, failing to learn good generalizable features. As such, the network is more likely to be confused by the trials as opposed to benefiting due to the apparent domain gap between the target and the remaining training subjects.

C. Limitations

The main limitations in this network is that the proposed model is only able to capture the dependencies between different timepoints up to the same length as the input used to train the model. If given an input with a longer time length than the original training data, the model will be unable to appropriately capture the long-term dependency between time points and another point that occurs more than the training input size later. One possible methodology that can be used to alleviate this challenge is to use a sliding window of the same length as the training set, followed by utilizing the hidden states of the encoded sliding input to be leveraged in the subsequent window.

VI. Conclusion

In summary, this work proposes the introduction of a semi-supervised transformer-attention convolutional network (SSTACNet) that builds upon previous foundations of attention-based spatial temporal convolutional networks [5]. We introduce the use of a VAE network architecture to perform self-selection of useful training data. Furthermore, the proposed classifier utilizes spatial-temporal filters to extract features from the original input data. Subsequently, transformer-attention modules are used to capture the long-term correlations between the features. finally, a sliding window alongside residual temporal convolutions are used to capture the local relationships between the features.

Acknowledgment

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-PhD-2021-08-021) and the RIE2020 AME Programmatic Fund, Singapore (No. A20G8b0102).

References

[1] E. López-Larraz, A. Sarasola-Sanz, N. Irastorza-Landa, N. Birbaumer, and A. Ramos-Murguialday, "Brain-machine interfaces for rehabilitation in stroke: a review," *NeuroRehabilitation*, vol. 43, no. 1, pp. 77–97, 2018.

[2] C. Q. Lai, H. Ibrahim, M. Z. Abdullah, J. M. Abdullah, S. A. Suandi, and A. Azman, "Artifacts and noise removal for electroencephalogram (eeg): A literature review," in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2018, pp. 326–332.

[3] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network," *Neural Networks*, vol. 136, pp. 1–10, 2021.

[4] H. W. Ng and C. Guan, "Efficient representation learning for inner speech domain generalization," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2023, pp. 131–141.

[5] 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.

[6] M. Wairagkar, Y. Hayashi, and S. J. Nasuto, "Dynamics of long-range temporal correlations in broadband eeg during different motor execution and imagery tasks," *Frontiers in neuroscience*, vol. 15, p. 660032, 2021.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[8] S. Pérez-Velasco, E. Santamaría-Vázquez, V. Martínez-Cagigal, D. Marcos-Martínez, and R. Hornero, "Eegsym: Overcoming inter-subject variability in motor imagery based bcis with deep learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1766–1775, 2022.

[9] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: a brief review," *Journal of neuroscience methods*, vol. 243, pp. 103–110, 2015.

[10] O. George, R. Smith, P. Madiraju, N. Yahyasoltani, and S. I. Ahamed, "Data augmentation strategies for eeg-based motor imagery decoding," *Heliyon*, vol. 8, no. 8, 2022.

[11] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[12] F. Pérez-Cruz, "Kullback-leibler divergence estimation of continuous distributions," in *2008 IEEE international symposium on information theory*. IEEE, 2008, pp. 1666–1670.

[13] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[14] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "Eeg datasets for motor imagery brain–computer interface," *GigaScience*, vol. 6, no. 7, p. gix034, 2017.

[15] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "Eeg dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.

[16] P. Saideepthi, P. Gaur, A. Chowdhury, K. Nyati, T. Dwivedi, S. Sharma, and R. B. Pachori, "Sliding window along with eegnet based prediction of eeg motor imagery," *IEEE Sensors Journal*, 2023.

[17] S. Cososchi, R. Strungaru, A. Ungureanu, and M. Ungureanu, "Eeg features extraction for motor imagery," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 1142–1145.

[18] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 2958–2965.

[19] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery bci systems," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 15–29, 2015.

[20] S. An, S. Kim, P. Chikontwe, and S. H. Park, "Dual attention relation network with fine-tuning for few-shot eeg motor imagery classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[21] 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 in neuroscience*, vol. 6, p. 39, 2012.

[22] H. Bashashati, R. K. Ward, and A. Bashashati, "User-customized brain computer interfaces using bayesian optimization," *Journal of neural engineering*, vol. 13, no. 2, p. 026001, 2016.

[23] H. Wang, P. Chen, M. Zhang, J. Zhang, X. Sun, M. Li, X. Yang, and Z. Gao, "Eeg-based motor imagery recognition framework via multisubject dynamic transfer and iterative self-training," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[24] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5619–5629, 2018.

[25] D. Li, J. Wang, J. Xu, and X. Fang, "Densely feature fusion based on convolutional neural networks for motor imagery eeg classification," *IEEE Access*, vol. 7, pp. 132 720–132 730, 2019.

[26] D. Zhang, H. Li, J. Xie, and D. Li, "Mi-dagsc: A domain adaptation approach incorporating comprehensive information from mi-eeg signals," *Neural Networks*, vol. 167, pp. 183–198, 2023.

[27] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in eeg dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 371–382, 2017.