Classification of Motor Imagery EEG Signals in Subjects with High Uncertainty using a Spectral Transformer

Jorge de la Rosa Padrón[1] and Alejandro del Río Torrado[1]

[1]*Polytechnic University of Madrid*

Abstract— **Motor Imagery (MI) Brain-Computer Interfaces (BCIs) play a pivotal role in facilitating communication and control for individuals through decoding brain signals. Traditional BCIs rely on Machine Learning (ML) algorithms, necessitating complex signal processing and feature engineering. Recent advancements in Deep Learning (DL) have shown promise in automating feature extraction from Electroencephalogram (EEG) signals. However, the efficacy of DL, especially with spectral transformers, remains underexplored in subject-dependent contexts, particularly in scenarios of low data and high uncertainty.**

This study addresses the challenges by investigating the applicability of a spectral transformer for classifying MI-BCI signals in subjects with limited data and high uncertainty. The proposed approach involves fine-tuning the transformer for individual patients, emphasizing subject-independent EEG classification. We compare the performance of the spectral transformer against state-of-the-art methods, shedding light on its potential advantages. The findings aim to contribute valuable insights into the viability of employing spectral transformers for enhancing the adaptability and efficiency of MI-BCI systems, with implications for a broader user base.

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) represent a cutting-edge technology that establishes a direct communication link between the human brain and external devices. These interfaces leverage the intrinsic electrical activity of the brain, captured through Electroencephalogram (EEG) signals [1], to interpret user intent and enable control over devices or computer systems. EEG signals can be use in various fields such as Neuroscience, Biometrics, Neuromarketing, etc. [2] Among the myriad applications of BCIs, one particularly promising area is Motor Imagery Brain-Computer Interfaces (MI-BCIs).

Motor imagery involves the mental simulation of movement without actual physical execution. In MI-BCIs, users engage in imagining movements, and the corresponding neural patterns are decoded from EEG signals. These are acquired through electrodes placed on the scalp (montage), measuring the electrical fluctuations resulting from synchronized neural activity. This non-invasive approach makes EEG an ideal candidate for real-time monitoring of brain function. This technology may be use to control drones, wheelchairs and Robots. [3]

E-mail address: j.dpadron@alumnos.upm.es (Jorge de la Rosa); alejandro.delrio@alumnos.upm.es (Alejandro del Río).

Traditionally, MI-BCIs have relied on conventional machine learning techniques to extract EEG features in the time, frequency and space domains. Most common techniques are Fast Fourier Transform (FFT), filter bank common spatial pattern (FBCSP) [4][5], Linear Discriminant Analysis (LDA) [6] and Multi-layer perceptrons (MLP) [7]. These methods, while effective in certain scenarios, face challenges in subject-independent applications, especially in cases of dealing with noisy datasets and accurately recognizing human intention from low signal-to-noise ratio and nonstatnionary brain signals fed with multiple sources of artifacts, such as biological artifacs, electronic equipement and environmental noise [8].

Recent advancements have witnessed the integration of deep learning models into the realm of MI-BCI. Some of these models are Convolutional Neural Networks (CNNs) [9], Long Short-Term Memory (LSTMs), Recurrent Neural Networks (RNN) [10] and Hybrid CNN [11], like CNN-LSTM or CNN-Autoencoders (AE). Since CNN model learns EEG features from different channels and both LSTMs and RNNs extract temporal information, a fusion between both of them have been demotrated that improve accuracy. These models offer the potential for automatic extraction of intricate spatio-temporal features from EEG signals, promising enhanced adaptability and performance in subject-independent contexts.

However, recent studies have shown that the Transformer model can be a good choice based on their attention mechanism. The first transformers in the field of EEG signal processing was introduce by Tao et al. with the Gated Transformer, using a gating mechanism instead of the residual connection [12]. Transformer-based models where introduced by Xie et al. in motor imagery EEG classification area [13]. Mahsa et al. also introduce an ensemble method which gathers two transfomers. A temporal transformer and a spectral transformer, used in this study [14]. This ensemble-based model aimed to get both spectral and temporal information. Hamdi et al. discussed a physics-informed attention temporal covolutional network for motor imagery classification where both Temporal and Convolutional Block were used too [8].

The primary objectives of this paper are to investigate the feasibility of applying a spectral transformer to classify EEG signals related to motor imagery in subjects characterized by high uncertainty and limited data availability. Specifically, we aim to assess the performance of traditional methods, such as Common Spatial Patterns (CSP) and Linear Discriminant Analysis, in subject-dependent classification scenarios. Additionally, we explore the application of CNN and hybrid models combining CSP with neural networks in an attempt to address the challenges posed by the cohesive nature of EEG data. Subsequently, our focus shifts to the evaluation of a spectral transformer's effectiveness in subject-dependent classification, comparing its performance with state-of-the-art models like ATCNet on benchmark datasets. Furthermore, we seek to refine the classification for subjects with high uncertainty by employing a fine-tuning approach with the spectral transformer. Through these objectives, we aim to contribute insights that advance the understanding and application of advanced neural models in the domain of motor imagery-based Brain-Computer Interfaces.

II. EXPERIMENTAL DATA

Our experimental investigation relies on the EEG Motor Movement/Imagery Database obtained from PhysioNet [15], comprising over 1500 EEG recordings collected from 109 volunteers. The dataset was designed to capture neural activity during various motor and imagery tasks, featuring a comprehensive experimental protocol.

*a) **Experimental Protocol**:* During the experimental sessions, subjects engaged in 14 runs, including baseline recordings and different motor/imagery tasks. Each subject performed one-minute baseline runs with eyes open and eyes closed, followed by two-minute runs of four distinct tasks. Notably, our focus centers on Task 4, where the subjects imagine performing open and close both fists or both feet. For this study, runs 6, 10 and 14 have been chosen, corresponding to the before-mentioned Task 4.

*b) **Montage**:* The EEG signals were recorded using a 64-channel montage, adhering to the international 10-10 system with specific electrode exclusions. Noteworthy omissions include electrodes Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10. The recorded EEG signals are sequentially numbered from 0 to 63, corresponding to the order of electrodes in the PhysioNet dataset.

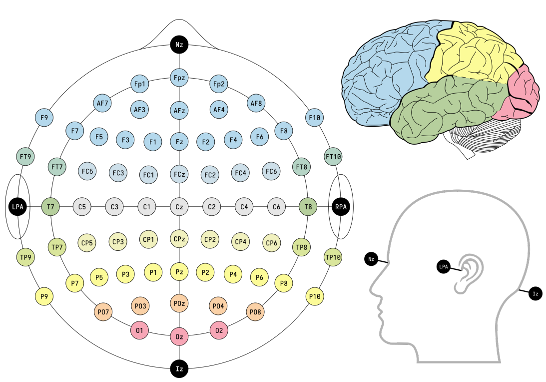


Fig. 1: International 10-10 system montage.

The EEG signals provided in EDF+ format were sampled at a sampling rate of 160 Hz, along with an annotation channel. Annotations include codes (T0, T1, or T2) denoting rest, onset of motion of the left fist or both fists, and onset of motion of the right fist or both feet. These annotations are usefull for task-related event identification in our analyses.

III. METHODOLOGY

A. Preproccesing

In the preprocessing stage, EEG data was preprocess using the MNE-Python package. First, raw EEG recordings from specific motor imagery tasks are loaded and concatenated across different experimental runs. Z-score normalization is used to allow for the comparison of EEG data across different recording sessions and subjects. EEG signals are non-stationary, meaning they can vary in each recording session, making it difficult to compare data directly. Z-score normalization helps to mitigate this issue by standardizing the data [16][17], enabling meaningful comparisons across different datasets and subjects .

To focus on the frequency range pertinent to neuroimaging, a bandpass filter is employed in order to retain just the frequencies from 7 Hz to 30 Hz while eliminating artifacts related to eye blink, heartbeat and muscle movement [18]. The alpha (8-12 Hz) and beta (18-28 Hz) frequency bands have been found to be related to motor activity and movement kinematics [19]. Notably, this filter is crucial for isolating frequency components associated with motor imagery tasks. Following the filtering process, event information is extracted from annotations, specifying conditions related to hands and feet motor imagery. The ensemble of signals is then visualized using Power Signal Density (PSD).

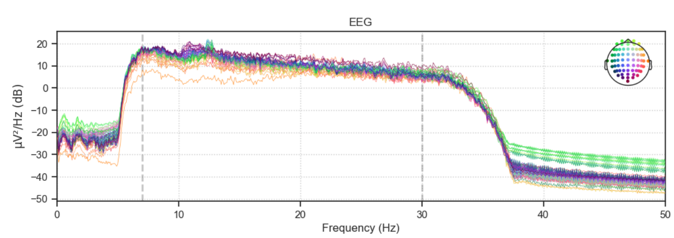


Fig. 2: Representation of the filtering within the 7 Hz to 30 Hz range using PSD for Subject No. 1.

Subsequently, epochs are defined by segmenting the data around the events of interest, providing a temporal window for analysis. In this case, epochs are extracted from 1.0 to 2.0 seconds post-event, capturing the critical period for motor imagery.

In order to enhance data visualization, we utilize Common Spatial Patterns to extract the four principal components from the EEG data. CSP is a mathematical technique commonly applied in EEG signal processing to decompose a multivariate signal into additive subcomponents. It transforms EEG signals into a low-dimensional spatial subspace, facilitating the extraction of pertinent features for subsequent classification and analysis. This method allows for the visualization of the four extracted components, providing valuable insights into regions of heightened activity associated with the assigned task for Subject No. 1.

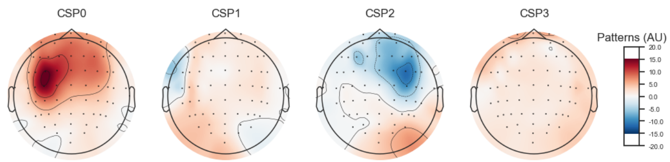


Fig. 3: Representation of the four components using CSP for Subject No. 1.

B. Subject-Independent Approach

Initially, a subject-independent approach was attempted, wherein data from all subjects were aggregated. Each subject contributed a total of 45 relevant samples, resulting in a dataset comprising 4747 samples. Four subjects with observational errors were excluded from the analysis. A training-test split was performed with a ratio of 90/10, respectively. Two distinct methodologies were employed for classification.

1) Machine Learning: The first involved classical techniques, specifically Common Spatial Pattern followed by Linear Discriminant Analysis (CSP-LDA). The combination of CSP for feature extraction and LDA for classification has been widely adopted in EEG signal processing due to its ability to effectively capture task-related information from EEG signals [20].

2) Deep Learning: Additionally, deep learning techniques were explored, categorized as follows:

- Convolutional Neural Network (CNN): Employed with and without standardized data, the CNN treated EEG channel data as an image, capitalizing on its inherent spatial structure. The decision to use CNN for multiple temporal series is motivated by its effectiveness in capturing spatial dependencies in data.
- CSP-Neural Network (CSP-NN): Introduced in the second category, this technique involved passing the four components derived from CSP to a neural network instead of LDA [21]. This approach aimed to enhance classification robustness by transcending linear boundaries. Similar to the CNN, both standardized and non-standardized data were utilized for experimentation.

Despite the variety of models explored, the results have been relatively suboptimal. As seen in Table II, the broad spectrum of models did not yield accuracy significantly superior to 60%, considering a baseline accuracy of 50%. This challenges the efficacy of the selected methodologies for subject-independent EEG classification in the current experimental setup.

These results can be attributed to the following reason. Upon visualizing the components extracted using CSP from samples of all subjects simultaneously, shown in Fig. 4, it becomes evident that correctly classifying the samples is challenging due to strong cohesion between data points belonging to the two categories. However, the components extracted through CSP from an individual subject exhibit linear separability. Consequently, our subsequent approach entails classifying data for each subject individually.

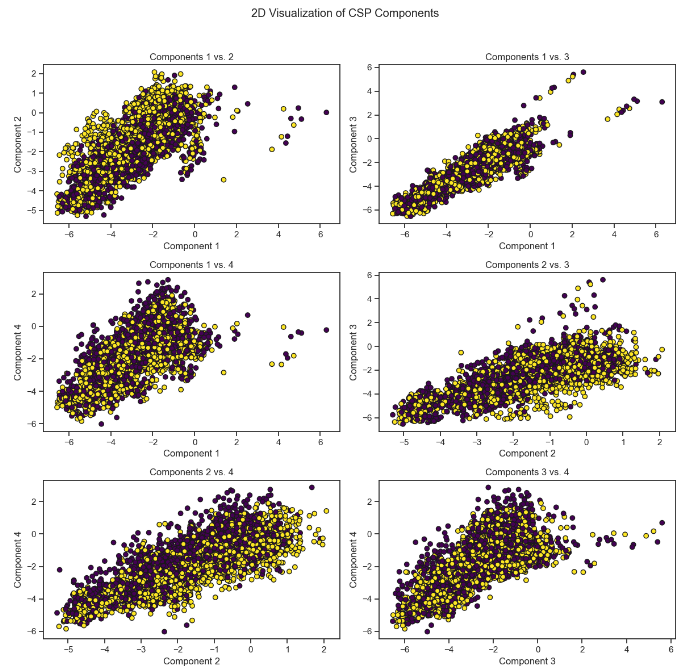


Fig. 4: Two-dimensional representation of components obtained through CSP for all subjects.

C. Subject-Dependent Approach

In this scenario, our objective is to classify EEG signals independently for each subject, meaning we train an individual model for each participant. With only 45 data points per subject, we confront a challenge of limited data for training and evaluation. Two distinct approaches have been employed for this purpose in order to see which one performs best.

On the one hand, we applied a Common Spatial Pattern followed by Linear Discriminant Analysis for each subject. On the other hand, we utilized the CSP-Neural Network (CSP-NN) model. The distribution of the maximum scores obtained by these models for each subject is illustrated in Fig. 5.

For more than half of the subjects, scores exceeding 80% were achieved. That is better than the results obtained in section III-B. However, some subjects yielded accuracy scores below the baseline 50%. A closer examination, as shown in Fig. 6, reveals high cohesion among data points for these subjects.

To address the challenge of accurately classifying these subjects exhibiting high cohesion in their data, we intend to employ a spectral transformer. Our approach involves training the spectral transformer initially with data from all

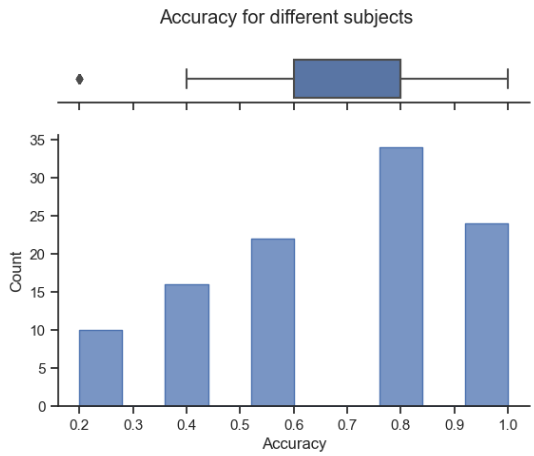


Fig. 5: Maximum score distribution using CSP-LDA and CSP-NN for all subjects.

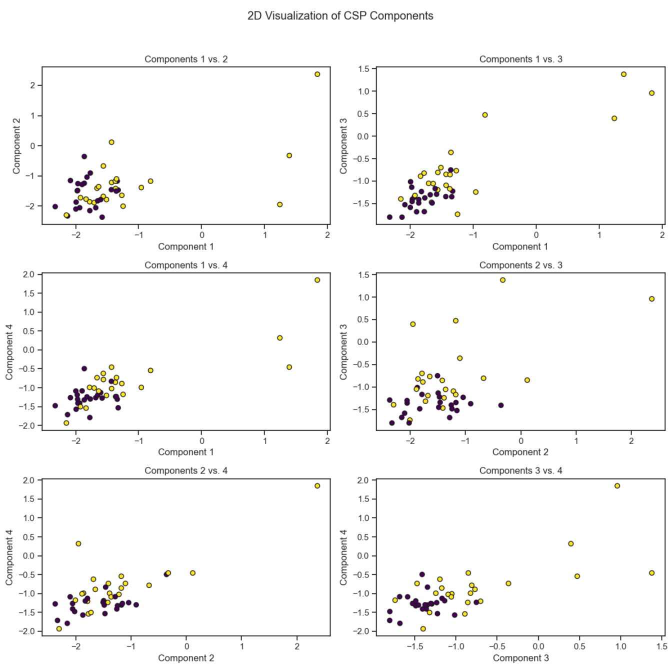


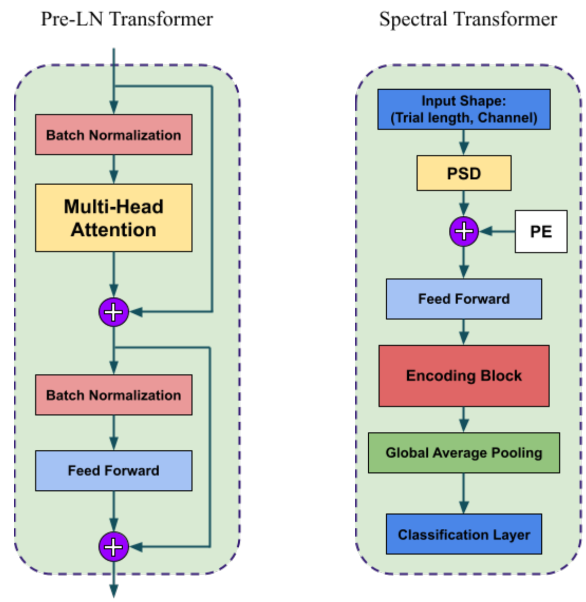
Fig. 6: Two-dimensional representation of components obtained through CSP for a subject with a suboptimal classification score.

subjects to learn their general characteristics. Subsequently, fine-tuning will be performed specifically on subjects presenting classification difficulties. This hybrid approach aims to combine aspects of both subject-dependent and subject-independent strategies, leveraging the strengths of each for enhanced classification.

D. Spectral Transformer

A signal can be represented in two distinct domains: the time domain and the frequency domain. The spectral transformer proposed by Mahsa et al. exclusively operates within the frequency domain, allowing us to explore the efficacy of leveraging frequency components in EEG signal classification [14].

The architecture of the spectral transformer shown in Fig. 7b consists of two primary blocks: encoding and frequency dependency extraction. The encoding block is crucial in the transformer architecture, and its implementation varies between two vanilla transformer variations, namely the Post-LN Transformer and the Pre-LN Transformer.



(a) Architecture of a Pre-LN Transformer.

(b) Final architecture of a Spectral Transformer.

Fig. 7: Diagram of architectures used for spectral transformer modeling.

In the Post-LN Transformer, a normalization layer is performed between the residual blocks, leading to challenges such as large gradients close to the output layer. To mitigate this issue, a learning rate warmup stage is necessary. Alternatively, the Pre-LN Transformer, as suggested in prior literature [12] [22] [23], incorporates a normalization layer within the residual blocks. This configuration ensures proper behavior of gradients during initialization, resulting in normalized input for the feed-forward layer and self-attention, functioning as a regularization. Given these advantages, we employ a Pre-LN architecture in this study (Fig. 7a).

To extract temporal dependencies from EEG data, features along the channel axis at the same time point were considered, and the correlation between different time points was calculated. This is added through the positional encoding.

For the extraction of frequency dependencies, the power spectrum density for each channel was computed. Subsequently, the data along the channel axis, after incorporating positional encoding, were treated as features and utilized as input for a fully connected layer. The output from the PSD layer undergoes further processing by a fully connected layer, aligning it optimally for the subsequent attention layer. Ultimately, the output from the encoding block undergoes further processing through a global average pooling layer. This pooled output is subsequently linked to a fully connected layer, where the number of output units aligns with the count of class labels.

In this paper, the configuration of the spectral transformer includes four Transformer modules (Encoding Blocks), with a head size of 128 and 1 head in each multi-head attention layer. This design choice is made to ensure effective extraction and utilization of frequency-dependent features for robust EEG signal classification.

- *Comparison with State-of-the-Art Model:*

In order to assess the performance of our proposed model, we conducted a comprehensive comparison with a state-of-the-art model using the BCI Competition IV-2a (BCI-2a) dataset. The BCI-2a dataset is widely recognized for its use in evaluating and comparing models in the domain of motor imagery neuroimaging [24]. The dataset includes EEG data from 9 subjects, each participating in two sessions, featuring a cue-based BCI paradigm with four distinct 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 session comprises 6 runs, with 48 trials per run (12 for each class), resulting in a total of 288 trials per session. At the onset of each session, a 5-minute recording was conducted to estimate EOG influence, consisting of three blocks: two minutes with eyes open, one minute with eyes closed, and one minute with eye movements. EEG data were recorded using 22 electrodes corresponding to the international 10-20 system, with a sampling rate of 250 Hz and a bandpass filter between 0.5 Hz and 100 Hz. The signals were monopolarly recorded, with the left mastoid as reference and the right mastoid as ground. Additionally, three monopolar EOG channels were recorded for artifact processing, and trials with artifacts were appropriately marked.

As a representative of the state-of-the-art, we chose to compare our model against the ATCNet. The ATCNet is an attention-based temporal convolutional network designed for EEG-based motor imagery classification [8]. The ATCNet incorporates multiple techniques to enhance MI classification performance with a relatively small number of parameters. This includes scientific machine learning to design a domain-specific DL model with interpretable and explainable features, multi-head self-attention to highlight the most valuable features in MI-EEG data, temporal convolutional network (TCN) to extract high-level temporal features, and a convolutional-based sliding window to efficiently augment the MI-EEG data.

E. *Fine-Tuning*

In this phase, we aim to classify EEG signals from subjects with previously identified poor classification performance. Our approach involves a two-step process. Initially, we train the spectral transformer using a subject-independent approach, utilizing data from all selected subjects. The objective here is to allow the model to learn general patterns and features inherent in the overall dataset.

Subsequently, we employ a fine-tuning strategy, where the learning rate is significantly reduced for a specific subject chosen based on their suboptimal classification results. This fine-tuning step is crucial as it enables the model to adapt and learn subject-specific characteristics. By adjusting the learning rate for the selected subject, we guide the model to focus on refining its understanding of the unique aspects of that individual's EEG signals.

The rationale behind this approach is to provide the model with a foundation of generalized knowledge acquired from the subject-independent training. Following this, the fine-tuning step allows the model to specialize and adapt its classification capabilities to the condition of the selected subject. The expectation is that this dual-step process empowers the model to build upon a foundational understanding of EEG signal patterns and subsequently refine its classification performance for specific subjects.

This fine-tuning methodology is grounded in the notion that a hybrid approach, combining subject-independent learning with targeted subject-specific adjustments, enhances the model's overall adaptability and classification accuracy.

IV. RESULTS

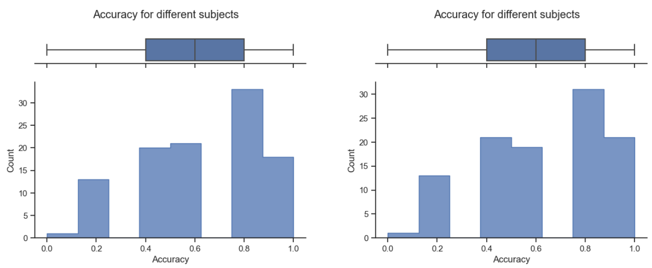
In this section, we present the outcomes of our experiments on the classification of motor imagery EEG signals. The evaluation is conducted based on both subject-independent and subject-dependent approaches. We begin by presenting the results of the Subject-Independent Approach obtained in Section III-B.

TABLE I: Subject-Independent Models Results

Method	Accuracy (%)	Precision (%)	Recall (%)
CSP-LDA	**65.1**	**65.1**	**65.1**
CNN (S)	48.9	48.6	48.7
CNN (N-S)	48.9	48.7	49.1
CSP-NN (S)	64.6	63.7	64.6
CSP-NN (N-S)	62.1	61.9	62.0

From the results, we observe that the CSP-LDA model achieved the highest accuracy, precision, and recall among the subject-independent models. The CNN models, whether using standardized (S) or non-standardized (N-S) data, showed lower performance. The CSP-NN model with standardized data performed better than the non-standardized version but slightly below the CSP-LDA model in terms of accuracy and recall. These findings suggest that, in this subject-independent setting, the CSP-LDA model outperforms the considered deep learning models for motor imagery classification.

In the Subject-Dependent Approach presented in Section III-C, we showcase the results using CSP-LDA in Fig. 8a and neural networks in Fig. 8b.



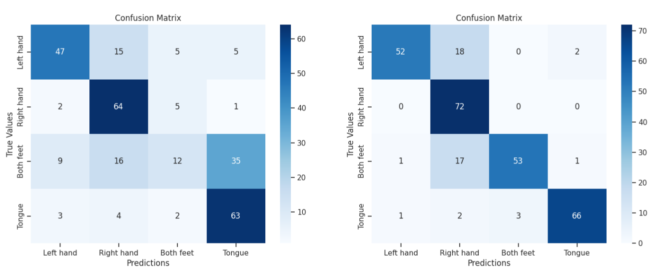
(a) Classification using CSP-LDA. **(b)** Classification using CSP-NN.

Fig. 8: Subject-Dependent classification approach.

The results from both graphs are quite similar since the only difference lies in the methods employed for class separation. Now, let's delve into the comparison results between the spectral transformer and ATCNet on the BCI-2a dataset.

TABLE II: BCI Competition IV-2a Model Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)
Spectral Transformer	**64.6**	**63.1**	**64.6**
ATCNet	84.4	88.2	84.4



(a) Spectral transformer. **(b)** ATCNet.

Fig. 9: Confusion matrices.

Upon close inspection of the confusion matrix in Fig. 9a, it is evident that the spectral transformer tends to struggle with accurate classification, particularly in distinguishing the "Both Feet" class, often misclassifying it as the "Tongue" class.

Finally, specific results for an individual subject, subject no. 14, have been obtained, yielding an accuracy of 80%. This marks a substantial improvement, surpassing the 20% accuracy achieved using CSP-LDA and CSP-NN. A recall of 80% and a precision of 86.7% were obtained for this subject.

V. Conclusions

Within the scope of the current investigation, an extensive evaluation of diverse machine learning and deep learning algorithms was systematically conducted, focusing on their application in the classification of Motor Imagery Brain-Computer Interfaces. The dataset employed for these analyses originated from the EEG Motor Movement/Imagery Database, accessed through PhysioNet. It is noteworthy that the dataset's dimensions proved inadequate for comprehensive training and evaluation of state-of-the-art algorithms, thereby constraining the robustness of the comparative analysis among the proposed models. This disparity becomes apparent when contrasting the performance of machine learning algorithms, which exhibit suboptimal proficiency in learning the inherent data and consequently generalizing, against the superior capabilities of deep learning algorithms.

The observed parity in the results between machine learning and deep learning algorithms underscores the need for a more expansive exploration of model efficacy. Future investigations must recognize the methodological integrity of the ongoing project while acknowledging the inherent limitation posed by insufficient data volume. To address this constraint, the experiment was replicated employing the BCI-2a dataset, a widely acknowledged dataset established for validating state-of-the-art propositions. This dataset's accessibility ensures transparency and facilitates result reproducibility, augmenting the credibility and applicability of the findings.

Despite challenges and inconclusive results, we advocate for fine-tuning tailored to individual differences as a key factor in improving outcomes. The inherent diversity in each person's brain structure makes the creation of a universal model impractical. Customizing models for specific individuals is essential for meaningful performance enhancements, recognizing the limitations of a one-size-fits-all approach.

Looking ahead, upon confirming the adequacy of the chosen dataset for generating robust outcomes, our future research will move beyond replication. We plan to advance the design of an improved spectral transformer, especially considering the favorable comparisons with ATCNet. Our goal is to introduce strategic modifications aimed at enhancing the performance of state-of-the-art algorithms. These enhancements will be carefully crafted and systematically implemented, contributing to the iterative refinement of existing methodologies. This progressive approach not only ensures result reproducibility but also actively contributes to the evolutionary trajectory of state-of-the-art techniques in the field of MI-BCI classification.

VI. Source Code

The source code for the experiments conducted with TensorFlow is available on GitHub at the following repository: `https://github.com/Mixnikon108/EEG-Signal-classification`.

References

[1] X. Zheng and W. Chen, "An attention-based bi-lstm method for visual object classification via eeg," *Biomedical Signal Processing and Control*, vol. 63, p. 102174, 2021.

[2] M. Soufineyestani, D. Dowling, and A. Khan, “Electroencephalography (eeg) technology applications and available devices,” *Applied Sciences*, vol. 10, no. 21, 2020.
[3] H. Altaheri, G. Muhammad, M. Alsulaiman, and et al., “Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: a review,” *Neural Computing and Applications*, vol. 35, pp. 14681–14722, 2023.
[4] P. S. López, H. K. Iversen, and S. Puthusserypady, “An efficient multi-class mi based bci scheme using statistical fusion techniques of classifiers,” in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pp. 378–382, 2019.
[5] A. S. Al-Fahoum and A. A. Al-Fraihat, “Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains,” *International Scholarly Research Notices*, vol. 2014, p. 730218, 2014.
[6] R. Fu, Y. Tian, T. Bao, and et al., “Improvement motor imagery eeg classification based on regularized linear discriminant analysis,” *J Med Syst*, vol. 43, no. 6, p. 169, 2019.
[7] O. O. Samuel, Y. Geng, X. Li, and et al., “Towards efficient decoding of multiple classes of motor imagery limb movements based on eeg spectral and time domain descriptors,” *J Med Syst*, vol. 41, no. 12, p. 194, 2017.
[8] 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, pp. 2249–2258, Feb 2023.
[9] Z. Tang, C. Li, and S. Sun, “Single-trial eeg classification of motor imagery using deep convolutional neural networks,” *Optik*, vol. 130, pp. 11–18, 2017.
[10] N. F. Güler, E. D. Übeyli, and Güler, “Recurrent neural networks employing lyapunov exponents for eeg signals classification,” *Expert Systems with Applications*, vol. 29, no. 3, pp. 506–514, 2005.
[11] Y. R. Tabar and U. Halici, “A novel deep learning approach for classification of eeg motor imagery signals,” *Journal of Neural Engineering*, vol. 14, p. 016003, nov 2016.
[12] Y. Tao, T. Sun, A. Muhamed, S. Genc, D. Jackson, A. Arsanjani, S. Yaddanapudi, L. Li, and P. Kumar, “Gated transformer for decoding human brain eeg signals,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 125–130, IEEE Engineering in Medicine and Biology Society, 2021.
[13] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, and Y. Zhan, “A transformer-based approach combining deep learning network and spatial-temporal information for raw eeg classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2126–2136, 2022.
[14] M. Zeynali, H. Seyedarabi, and R. Afrouzian, “Classification of eeg signals using transformer based deep learning and ensemble models,” *Biomedical Signal Processing and Control*, vol. 86, p. 105130, 2023.
[15] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, “Bci2000: A general-purpose brain-computer interface (bci) system,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
[16] R. Nicolas, L. Jean-Marc, C. M. John, B. Fabrice, and B. Christian-George, “Time-frequency strategies for increasing high-frequency oscillation detectability in intracerebral eeg,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 12, pp. 2595–2606, 2016.
[17] B. Kaliraman, S. Nain, R. Verma, Y. Dhankhar, and P. B. Hari, “Pre-processing of eeg signal using independent component analysis,” pp. 1–5, 2022.
[18] N. Yahya, H. Musa, Z. Y. Ong, and I. Elamvazuthi, “Classification of motor functions from electroencephalogram (eeg) signals based on an integrated method comprised of common spatial pattern and wavelet transform framework,” *Sensors (Basel, Switzerland)*, vol. 19, no. 22, p. 4878, 2019.
[19] H. Yuan, C. Perdoni, and B. He, “Relationship between speed and eeg activity during imagined and executed hand movements,” *Journal of Neural Engineering*, vol. 7, no. 2, p. 026001, 2010.
[20] M. Aljalal, R. Djemal, K. AlSharabi, and S. Ibrahim, “Feature extraction of eeg based motor imagery using csp based on logarithmic band power, entropy and energy,” in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–6, 2018.
[21] D. Maryanovsky, M. Mousavi, N. Moreno, and V. R. de Sa, “Csp-nn: a convolutional neural network implementation of common spatial patterns,” in *Graz Brain-Computer Interface Conference*, 2017.
[22] Q. Wang *et al.*, “Learning deep transformer models for machine translation,” *arXiv Preprint arXiv:1906.01787*, 2019.
[23] A. Baevski and M. Auli, “Adaptive input representations for neural language modeling,” *arXiv Preprint arXiv:1809.10853*, 2018.
[24] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, “Bci competition 2008–graz data set a,” *Inst. Knowl. Discov. Graz Univ. Technol.*, vol. 16, pp. 1–6, 2008.