Manuscript accepted for publication in the 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2025, Copenhagen, Denmark

MI-CES: An explainable weak labelling approach to example selection for Motor Imagery BCI classification*

Alexander Thomas[1], Youngjun Cho[2], Hubin Zhao[1], Tom Carlson[1]

***Abstract*— Motor Imagery (MI) Brain Computer Interfaces (BCI) can be used to control assistive devices such as wheelchairs. These systems require a training period to get both the user and the machine to learn and adapt to each other, achieving an acceptable control accuracy. Previous systems have discovered that providing a form of feedback to the user about what the system thinks the user is thinking can increase the effect of training and increase both the control accuracy of the user and the classification accuracy of the BCI system. However, if this feedback is 'incorrect' due to the classifier behind the BCI system having a poor accuracy, this may cause the user to 'incorrectly' adapt to the feedback, providing the system with further poor examples of MI. In this paper, we propose MI-CES, an explainable 'example selection' approach based on the neuro-physiological principle of MI. We found that while using 2 classification techniques, we achieved a statistically significant increase in classification accuracy across 3 datasets that were comprised of both multi-participant and multi-session recordings.**

I. Introduction

EEG Motor Imagery (MI) brain computer interfaces (BCI) require a significant training period for both the user and the machine learning (ML) algorithms to adapt to each other [1], [2]. During this mutual learning cycle, the user learns how to control the system and the ML-based BCI learns to identify the user's unique set of mental instructions. This training often consists of multiple recording sessions of the user performing the required mental instructions after a corresponding cue (e.g. left hand or right hand) [3]. After one or multiple sessions, a form of feedback is introduced to the user representing either the classification of the previous epoch of the instruction or a measure of how close the user is to achieving the correct command [1]. The user then relies on this feedback to adjust their mental commands to learn to control the system correctly. However, if the classification accuracy of the system behind the feedback is poor, then the user could learn to perform the instructions 'incorrectly'. We hypothesize that by selecting examples in our dataset that contain the features most similar to the known neurological effects of MI, we can increase the classification accuracy, and by extension, the quality of the feedback. A few previous weak labelling BCI studies have been carried out, however, none of the methods have been physiologically explainable [4], [5]. In this paper, we propose *Motor imagery covariance example selection* (MI-CES), which aims to select examples for the ML training that are based on neurological features related to MI, which we hypothesize will improve classification accuracy.

*This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) for the University College London Doctoral Training Partnership. For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) license to any Accepted Manuscript version arising.

[1] A. Thomas, H. Zhao, T. Carlson are with UCL Aspire Create, Division of Surgery, UCL, London, United Kingdom t.carlson@ucl.ac.uk

[2]Y. Cho is with the Department of Computer Science, UCL, London, United Kingdom youngjun.cho@ucl.ac.uk

II. Related work

BCIs can facilitate communication between a person's brain and a robotic device or computer. This can be particularly useful for people suffering from disabilities or injuries that cause them to lose most of their motor function below the neck. EEG BCI systems to control assistive devices such as wheelchairs, spelling devices and tele-presence robots have been previously developed, and make use of different types of mental instructions [6]–[8]. Some devices are controlled using synchronous EEG paradigms, which rely on measuring the user response to a given stimulus [7]. Other devices rely on asynchronous BCI paradigms such as MI, which rely on classifying voluntary user commands [8], [9].

A. Motor Imagery

The MI paradigm is a popular asynchronous form of neurological communication between non-invasive EEG systems controlling assistive devices and the user. MI provides a relatively high classification accuracy and is suitable for users who suffer from tetraplegia due to spinal cord injuries due to the signals originating from a cortical area that is underutilized. MI functions by having the user think about moving a limb. This has the effect of activating the neurons across the motor sensory area associated with that limb. When a user is not thinking of moving a limb, the neurons in the motor cortex fire at their resting frequency within the 8-12Hz band, which is known as mu-rhythm. When the motor cortex is activated, the neurons begin to fire at different frequencies, desynchronizing, and reducing the magnitude of the 8-12Hz mu-rhythm component in the signal. This effect is known as event-related de-synchronization (ERD) [10]. MI has been used as a paradigm to control multiple different devices such as smart wheelchairs, telepresence robots and robotic arms [8], [9], [11].

B. Machine learning

ML is a core component of most MI BCI systems [12]. The ability of ML algorithms to learn allows BCI systems to classify within each user's or recording session's specific data space (Fig. 1), which can differ due to factors such as cap position, cortical layout, peak band frequencies and a

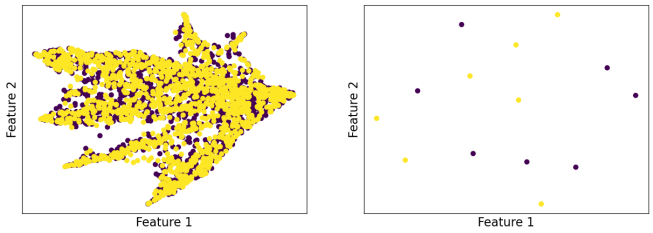

Fig. 1: Fourier basis frequency embedding of 2 instructions of all users in the PhysioNet data [18] set (left) compared to one (right). Data reduction was performed using uniform manifold approximation (UMAP) to maintain the multi-dimensional relationship, rather than extracting new features

difference in new features generated by prior pre-processing steps in the BCI pipeline (e.g. ICA, PCA).

The majority of ML methods used to classify EEG data have been linear models aiming to create a boundary between 2 areas in their feature spaces. The most popular methods used are Linear discriminant analysis (LDA), Support vector machines (SVM), Quadratic discriminant analysis (QDA), decision trees and Multi-layer perceptrons [13]. However, some sort of pre-processing is required to remove any artefacts or related features and embed the EEG signals into an appropriate feature space for ML methods to work [14]. Statistical classifiers such as naive bayes classifiers have also been used in BCI devices before [9].

Deep learning methods have emerged as a popular option to classify EEG signals. Lawhern et al. used a convolutional neural network (CNN) to classify 4 MI instructions using the BCI competition 4 dataset [15]. The shape of the CNN kernels was designed to extract features across the temporal and spectral domains separately, mimicking the filter bank common spatial pattern (CSP) processing method. On 4 classes they achieved an inter-participant accuracy of $\sim 60\%$. However, when this method was applied to a combination of all the participant's data the accuracy decreased.

Further developments were made to increase the accuracy of EEG BCI deep learning models by using 2D CNN kernels designed to learn combined temporal-spatial features, boosting their mean intra-participant accuracy to 76.2% [16]. Different types of deep learning layers have also been used in combination with CNNs, such as Long short-term memory layers (LSTMs), which combine the next input with the previous output of that layer in an aim to learn some causal features related to the signal. Models using LSTMs have managed to achieve higher accuracies than their pure CNN counterparts [17].

C. Limitations with previous methods

Even with further advances in ML, the development of more complex networks still struggle to achieve a level of accuracy that allows a user to control a wheelchair reliably in the long-term. Most ML papers focus on offline classification accuracy, which when focusing on the same dataset, can cause development to tend towards over-fitting that dataset rather than improving the generalization of models or their online application [12], [13], [19].

Another issue with most examples is the way data is extracted. Most MI datasets follow a broadly similar data collection process, where the user is presented with a prompt, and then asked to perform the MI task during a period of time [3], [18], [20]. This collection process may lead to a labelling issue that is not seen in other machine-learning applications. When instructions are extracted or segmented out of a larger recording, each one of these segments is assigned the label of the larger section is was taken from. A 10-second recording between a left cue and a stop cue can be split into 10 one-second recordings with all labelled as a left-hand instruction. EEG MI data is extremely difficult to visually identify as a *correct* example of one class, and the time it would take to do so for the number of examples generated in one training session would be extremely inefficient during a BCI training session. Segmented examples where the user may not have been performing the instruction correctly would then be introduced into the dataset as a labelled example of one class, even though it shows no similar trends to the rest of that class. This can make it not only difficult for the classifier to classify that example as well as generate a concise and generalized boundary for the other, but it can also make the classifier ineffective in future online classification, especially during online feedback supported training. Should this classifier provide poor quality feedback during training, it may then cause the user to constantly perform incorrect instructions, compounding the labelling problem. Therefore, this makes EEG MI classification from offline datasets a suitable "weak labelling problem".

D. Weak Labelling

Weak labelling is a concept in ML that focuses on learning from data whose corresponding labels may be incorrect, ambiguous, only exist for a fraction of the dataset, or provide partial information about the concept it's designed to learn [21]. Semi-supervised learning can be useful in incomplete labeling when you have a large dataset which would be extremely costly to label manually or requires an expert to label. A small subset of data can be used to train an initial classifier to identify the rest of the dataset. This method however is not suited to the problem in BCI MI studies where the high variance across participants makes reliable expert labeling and label estimation impractical (Fig. 1). The issue of BCI MI learning is more suited to inexact labelling, where we have a small amount of information about a group of data examples, but not enough to identify where in the data or which segments are correct. The common method for solving these problems is multi-instance learning.

E. Multi-instance learning

One key concept of weak labelling is multi-instance learning (MIL). MIL (Fig. 2) aims to learn a concept from the differences between or areas of high data concentration in bags of data [22]. A bag is a group of examples, where each bag is either a positive or negative bag. Negative bags must

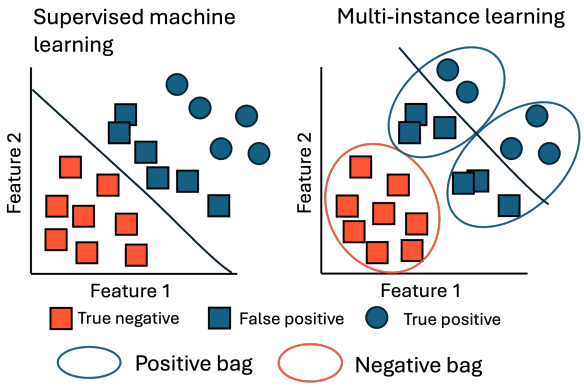

Fig. 2: A comparison of supervised learning (left) and multi-instance learning (right). Multi-instance learning aims to learn only the true positive examples inside the positive bags.

not contain any positive examples, whereas a positive bag must contain at least one positive example. This technique has been particularly effective at image segmentation, where the aim of the ML algorithm is to learn a concept from multiple different images which contain an example in comparison to images which do not. It can also be very effective in classifying and locating images which are too large or have relativity small discriminant features that a CNN architecture would be unable to extract such as high-resolution cancer slide images [23] and identifying objects in video format which combines a temporal-spatial elements [24]

Sadatnejad et al. used MIL to classify a mental health-related data EEG dataset [4]. MIL was used to account for biological and external artefacts. After applying different bandpass filters and data segmentation, they produced a bag of concepts for each example. After this, the covariance matrix of each concept was then embedded across a symmetric positive definite manifold. MIL was then used to learn the bag level difference between the classes.

Caicedo-Acosta et al. used a MIL framework for both instance and feature selection [5]. Each EEG trail was converted into a bag using a sliding window and short-time Fourier transform. Instance selection was based on the sparse representation classifier. A set of weights is learned through a linear regressive process to create a vector of scalar coefficients. Each instance is then sparsely reconstructed using these learned weights and all other examples. Any example where the difference between the actual signal and the reconstructed signal greater than a threshold is removed.

Both MIL methods function based on learning a concept generated within their feature space. However, the high variability and artifact-prone nature of EEG data can mean that this learned concept can become detached from the data produced by the MI task and makes the assumption that the majority of the data provided represents the true concept you are aiming to classify. In the case of a classifier used to feedback and to direct the user during training, the classifier can overlearn a concept specific to that session and not related to correct MI. For this reason, we aimed to design an MI example selection algorithm that is based of the neurological principles of MI.

III. Method

In this section, we describe the method and reasoning behind MI-CES. Unlike the previous methods used for motor imagery example selection [4], [5], our method uses a novel approach by using the known neurological principles of MI in its reasoning for selecting correct examples by comparing each example to an similarly embedded 'Ideal example'. This aims to constrain the example selection algorithm within an area of the search space we believe to be neurologically relevant.

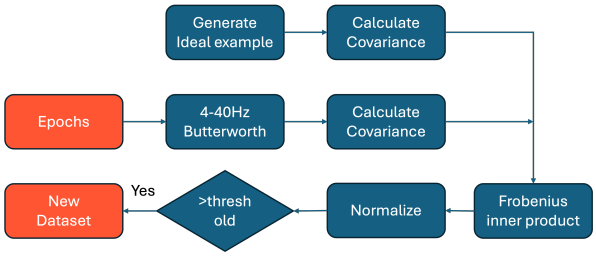

Fig. 3: Flow chart describing the MI-CES method

A. Datasets

Three MI datasets were selected for classification. Firstly, the BCI-competition iV 2a dataset [20] set was selected for its common use in a large number of MI BCI classification papers. The PhysioNet BCI 2000 dataset was also selected for its common use as well as it having over a hundred participants with less data, allowing us to test over a wider population [18]. The BCI-VR dataset recorded by Thomas et al. was also selected because a feedback-based classifier was used throughout the training sessions [3]. Due to the longitudinal style of the dataset (multiple separate sessions over different days per participant), each session was individually tested instead of testing between participants.

B. Channel selection and instructions

In order to fairly assess the effect of MI-CES over the 3 datasets with different electrode layouts, 15 common electrodes across the motor cortex were selected from all 3 datasets. The electrodes are positioned in the 10-20 system positions as shown in Fig. 4.

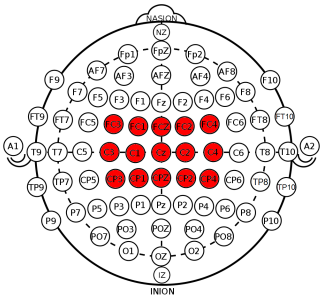

Fig. 4: A topographical layout of the international 10-20 system. Electrodes highlighted in red are the electrodes selected across all datasets for this study

Since MI-CES extracts spatial features, we have opted to test the viability of MI-CES on left and right-hand MI due to the neuro-physiological effects of the MI instructions system having a greater spatial separation across the motor cortex. This measure is taken to avoid possible component contamination by foot or tongue examples captured by electrodes that may have been misplaced between the key areas of the motor-sensory area.

IV. Segmentation and filtering

For each dataset, the time between instruction cues was extracted and segmented into separate epochs using a 1-second sliding window with a 0.5 second overlap between examples. Each epoch was then filtered using a 4-40Hz 4th order Butterworth filter to extract the frequencies that are neuro-physiologically relevant.

A. Covariance combination

To create a general method that derives its logic from neurological principles, that can be applied without any other information, the method should be able to judge an example based on consistent neurological features across the human population. A temporal method would not be appropriate for this due to the high variability in peak alpha power across the population. Due to the structural layout of the brain in healthy participants being consistent, we designed a method that assesses the spatial difference of the participant's neural information. By taking the covariance of each epoch, we can assess the similarity and difference between each channel, where channels that have the greatest difference in frequency have low covariance and channels that are similar having a high covariance. When an MI mental instruction causes ERD across a certain part of the motor cortex, the covariance of that channel and all of the others should decrease (Fig. 6). Unlike other connectivity measures such as coherence which aims to measure which areas of the brain are firing at the same specific frequency [25], low covariance is a measure of difference in frequency across the spectrum and is therefore less sensitive to user-specific peak alpha and ERD beta frequency.

B. Ideal example generation

In order to assess an example, we need to generate a signal for comparison. We generate a matrix p where each vector along the temporal axis is a sinusoid of a specific frequency. each sinusoid is generated using the equation:

$$X_n = \frac{2\pi n f}{f_s} \tag{1}$$

Where X is the vector representing the signal, n is the step of the signal, f is the desired frequency and f_s is the sampling frequency of our signal. The vectors in the matrix associated with the channels located over the expected cortical area affected by the MI instruction are sinusoids at 20Hz to mimic the increase in beta activity in that area that occurs during MI [10]. For left-hand instructions, these are the vectors linked to the 10-20 channels FC4 & C4 and for the right-hand, the channels selected are FC3 & C3 (Fig. 4). All channels that are not associated with the instruction are set to 10Hz to mimic the alpha activity (Fig. 5). Due to covariance being a relative measure, representing a relationship between two signals, adjustments to these frequencies do not need to be made for individual users, sessions or the non-stationary nature of EEG.

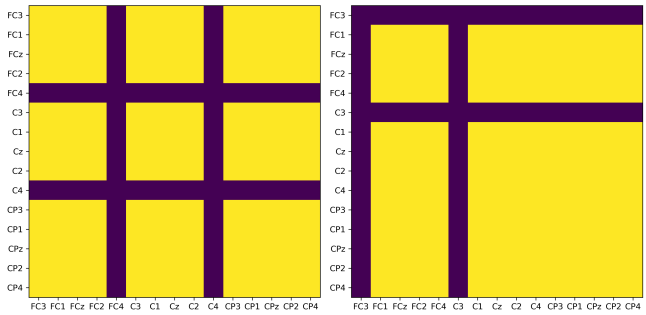

Fig. 5: Heat maps of left-hand (left) instruction and right-hand (right) instruction "Ideal example" covariance matrices. These examples are generated by measuring the covariance of the generated "Ideal example" of both a left-hand and right-hand motor imagery task

C. Comparison and selection

To compare the covariance matrices from a user example to an "ideal example" (Fig. 6), we take the Frobenius inner product, which allows us to compare 2 matrices, rather than having to take the inner product or similar measures with flattened vector representations of the matrix. The Frobenius inner product is expressed as:

$$\langle A, B \rangle_F = tr(\overline{A^T} B) \tag{2}$$

For each session or participant (dataset dependent) we normalize the inner products for each instruction group of epochs between 0 and 1. All values above a threshold are then selected to be the new dataset.

D. Classification

Two proven classification techniques will be used to assess the effect of the feature selection method.

The first technique is a 4-component common spatial patterns (CSP) transformation followed by a class-balanced SVM using a linear kernel. The simple nature of this classifier was chosen instead of a more advanced prepossessing and ML method to not bias our results and to effectively assess MI-CES. A second frequency-based method was chosen, consisting of a 6-20Hz Butterworth filter, followed by a fast Fourier transform followed by an RBF kernel class balanced SVM. This was chosen to assess the effect of an embedding method that does not extract its basis vectors from the data it is provided, instead embedding the data into a common pre-defined frequency space.

Since the method provided does not maintain a fixed number of examples and with no prior knowledge of which

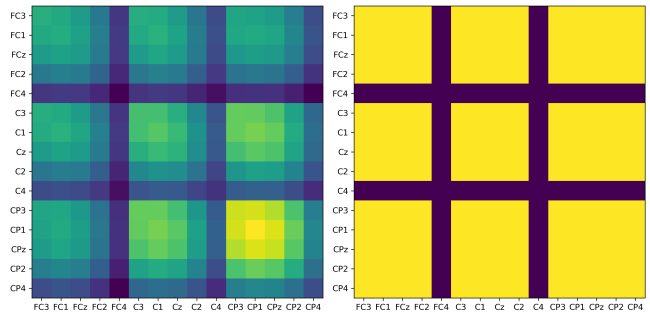

Fig. 6: Left-hand example covariance matrix (left) and the covariance matrix of an "ideal example" for a left-hand example (right). Both show a similar grid like area of love covariance, indicating that the left-hand human example (left) is a good example of of left hand motor imagery example.

examples inside our data are correct, we decided not to test any of the deep learning methods developed in previous literature as they would most likely over-fit the small and possibly imbalanced dataset provided by the example selection algorithm.

V. Results

A. Same session classification accuracy

For all sessions and results below, a normalized score threshold of > 0.5 was used. Statistical significance was measured using a Wilcoxon signed rank test. A red '*' is used to indicate a statistically significant difference between the 5-fold cross-validation accuracies between the standard BCI pipeline attempting to classify all examples and the same pipeline classifying the examples selected using the MI-CES method.

1) BCI Comp IV 2a dataset:

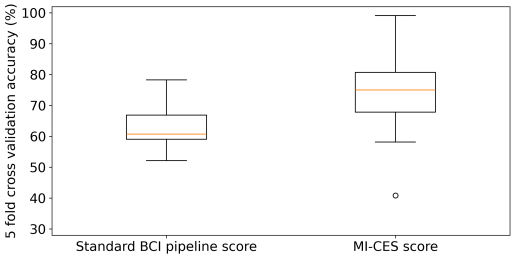

Fig. 7: Box plot of accuracies per participant using a CSP linear-SVM classier on the BCI Comp IV 2a dataset

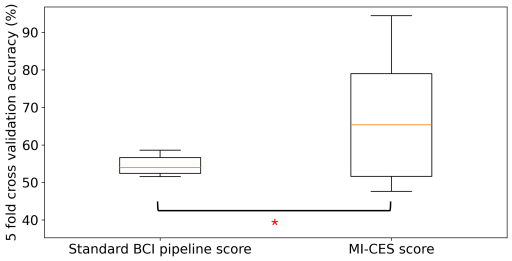

Fig. 8: Box plot of accuracies per participant using an FFT-SVM classier on the BCI Comp IV 2a dataset

2) PhysioNet dataset:

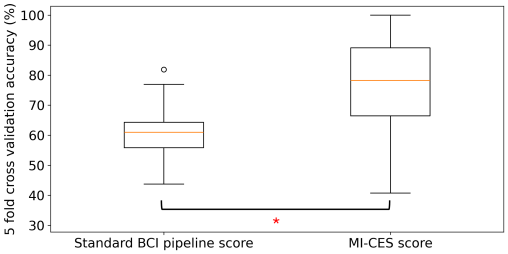

Fig. 9: Box plot of accuracies per participant using a CSP linear-SVM classier on the PhysioNet dataset

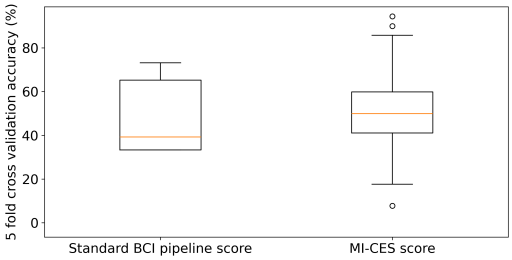

Fig. 10: Box plot of accuracies per participant using an FFT-SVM classier on the PhysioNet dataset

3) BCI-VR dataset dataset:

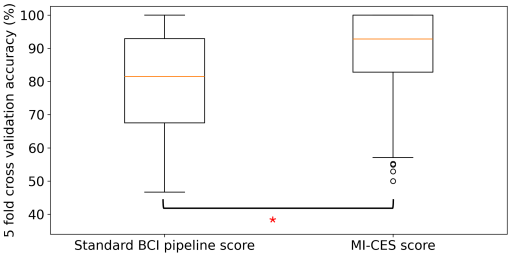

Fig. 11: Box plot of accuracies per session using a CSP linear-SVM classier on the BCI-VR dataset

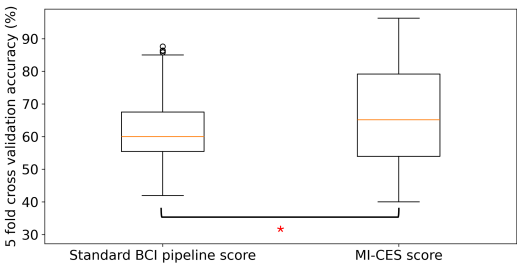

Fig. 12: Box plot of accuracies per session using an FFT-SVM classier on the BCI-VR dataset

4) Example loss to accuracy correlation:

In order to assess the relationship between the amount of examples lost and the accuracy, scatter plots are generated showing the number of examples lost in comparison to the accuracy achieved. This provides information to see how 'destructive' the method is and if high accuracies are only achieved when most of the data is lost. The least mean squares method was used to measure the correlation. This was only done using the CSP-SVM classifier and was not performed on the BCI comp IV 2a dataset as only 9 tests were performed.

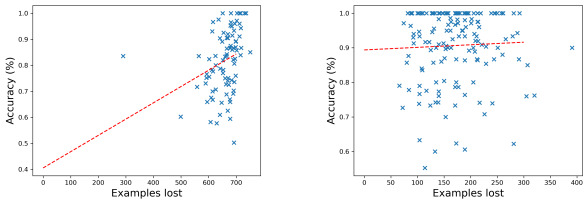

Fig. 13: Scatter plots of the amount of examples lost during the selection process compared to the 5-fold cross-validation accuracy achieved by the classifier after MI-CES example selection. Left is the scatter plot of each participant in the PhysioNet BCI 2000 dataset. Right is the accuracy of each session in the BCI-VR dataset

	PhysioNet	BCI-VR
R^2	0.22	$8.45 * 10^{-4}$

TABLE I: R^2 values of the correlations between examples lost and accuracy achieved on both datasets (Fig. I)

VI. Discussion

A. BCI Comp IV 2a analysis

Both classifiers while using example selection mostly achieved a higher accuracy than when classifying all examples (Fig. 7, Fig. 8), with the FFT-SVM classifier achieving a statistically significant increase ($p < 0.05$) in classification accuracy per participant. Although the mean achieved by the CSP-SVM classifier had a higher mean accuracy, the MI-CES method produced a wider distribution, also achieving a lower accuracy in some cases (Fig. 11). With the BCI-comp IV 2a dataset only containing 9 participants, any examples at the tail ends of the distribution can significantly harm the results.

B. PhysioNet dataset analysis

Using the PhysioNet dataset, we achieved a statistically significant increase in accuracy ($p < 0.05$) across participants using MI-CES in combination with the CSP-SVM classifier (Fig. 9). We however did not achieve a statistically significant increase using the FFT-SVM classifier, with the example selection algorithm producing a far wider distribution of accuracies (Fig. 10).

Although both classifiers achieved a higher mean accuracy when trained on the chosen examples, the distribution of these accuracies is wider than when trained on all examples (Fig. 9, Fig. 10). This effect could be caused by the channels with the greatest difference for those participants not matching the channels chosen in the examples, therefore causing the system to select poor examples or their data containing a small number of considerably poor examples relative to the rest of the dataset causing the distribution of normalized scores to positively skewed.

C. BCI-VR dataset

The use of the feature selection algorithm increased the performance of both classifiers (Fig. 11, Fig. 12) and the inclusion of the MI-CES produced a statistically significant increase in accuracy using both classifiers.

D. Example loss to accuracy correlation

Although there is a positive correlation seen between the number of examples lost and the accuracy (Fig. 13), the application of both datasets produce different correlations.

For the PhysioNet dataset, the application of the MI-CES can be observed to be equally destructive across all examples, yet have a high variance in accuracy between examples within that range. This combined with the increase in accuracy seen across participants (Fig. 9) suggests that the feature selection method can be beneficial to some participants or have little to no effect. This can be caused by consistent artifacts not being removed, and the very low scores provided by noisy examples increasing the low scores of the poorer quality clear instructions during the normalization phase. These trends may also be produced by the lack of any filter personalization, where cap misalignment or differences in cognitive layout relative to the cap position produce examples where the spatial ERD/ERS elements are miss-aligned with the ideal features generated for the comparison (Fig. 6).

The correlation between accuracy and example loss for the BCI-VR session is only slightly positive (Table: I). With a large number of sessions achieving a 100% accuracy across a wide range of examples lost, this suggests that MI-CES works well across this dataset. This result may lend to the experiment process used to collect the BCI-VR dataset, where feedback was provided from the second training session onward to each participant possibly introducing incorrect feedback from very early on in the training period [3].

Without testing MI-CES on participants training in a real time, the real effect of a high example loss remains unknown, as the reason for the high rate of example loss could be attributed to many factors such as the accuracy of feedback, the users BCI literacy or number of 'poor examples' included within the dataset.

E. Effects on future training

With the statistically significant increase in accuracy that MI-CES achieved, we expect that the use of the MI-CES method could improve the effectiveness of a feedback-assisted MI BCI training paradigm. Due to MI-CES selecting examples based on the neurophysiology principles of MI, we believe that the feedback provided will more accurately represent how close the user is to performing the task accurately rather than how close the user is to performing the same instructions from the previous sessions. In the future, to test this hypothesis an online multiple-session test across a group of participants would be required, analyzing their classification accuracy through the training period, any changes in the number of results they got correct and the change in response to the feedback provided.

VII. Conclusion

In this paper, we set out to develop an explainable example selection algorithm designed to assess a potential weak labelling problem that occurs during EEG MI BCI data collection which introduces the possibility of providing bad feedback to the user during online training. We designed a method that compares the covariance matrix of a synthesized 'ideal example' for a specific instruction and compared it to the covariance matrix of a real MI example. Any example with a normalized score above 0.5 was selected. We found a statistically significant increase in the classification accuracies of 2 BCI pipelines with different embeddings across 3 different datasets, with our highest scores being seen in the BCI VR dataset, which when collected did include real-time online feedback. The wide correlations seen when applying MI-CES to different datasets suggest that although MI-CES is beneficial to some participants, a method of filter personalization may help provide the benefit to all the participants. Further testing is required to assess the effect of the MI-CES in real-time online scenarios.

References

[1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[2] R. Liu, S. Kumar, H. Alawieh, E. Carnahan, and J. D. R. Millán, "On Transfer Learning for Naive Brain Computer Interface Users," in *International IEEE/EMBS Conference on Neural Engineering, NER*, vol. 2023-April. IEEE Computer Society, 2023.

[3] A. Thomas, J. Chen, A. Hella-Szabo, M. Kelly, and T. Carlson, "High stimuli virtual reality training for a brain controlled robotic wheelchair," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 11 305–11 311.

[4] K. Sadatnejad, M. Rahmati, R. Rostami, R. Kazemi, S. S. Ghidary, A. Müller, and F. Alimardani, "EEG representation using multi-instance framework on the manifold of symmetric positive definite matrices," *Journal of Neural Engineering*, vol. 16, no. 3, 2019.

[5] J. Caicedo-Acosta, L. Velasquez-Martinez, D. Cárdenas-Peña, and G. Castellanos-Dominguez, "Multiple Instance Learning Selecting Time-Frequency Features for Brain Computing Interfaces," in *Progress in Artificial Intelligence and Pattern Recognition*. Springer International Publishing, 2018, pp. 326–333.

[6] J. Leaman and H. M. La, "A Comprehensive Review of Smart Wheelchairs: Past, Present, and Future," pp. 486–489, 8 2017.

[7] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain-computer interface," in *2004 IEEE International Workshop on Biomedical Circuits and Systems*, 2004.

[8] R. Leeb, L. Tonin, M. Rohm, L. Desideri, T. Carlson, and J. D. R. Millán, "Towards independence: A BCI telepresence robot for people with severe motor disabilities," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 969–982, 6 2015.

[9] T. Carlson and J. Del R. Millan, "Brain-controlled wheelchairs: A robotic architecture," *IEEE Robotics and Automation Magazine*, vol. 20, no. 1, pp. 65–73, 2013.

[10] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, 2006.

[11] G. Gillini, P. D. Lillo, and F. Arrichiello, "An Assistive Shared Control Architecture for a Robotic Arm Using EEG-Based BCI with Motor Imagery," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 4132–4137.

[12] N. Sharma, M. Sharma, A. Singhal, R. Vyas, H. Malik, A. Afthanorhan, and M. A. Hossaini, "Recent Trends in EEG-Based Motor Imagery Signal Analysis and Recognition: A Comprehensive Review," *IEEE Access*, vol. 11, pp. 80 518–80 542, 2023.

[13] A. Saibene, M. Caglioni, S. Corchs, and F. Gasparini, "EEG-Based BCIs on Motor Imagery Paradigm Using Wearable Technologies: A Systematic Review," *Sensors*, vol. 23, no. 5, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/5/2798

[14] S. Aggarwal and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: A review," *Array*, vol. 1-2, 1 2019.

[15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," 11 2016. [Online]. Available: http://arxiv.org/abs/1611.08024 http://dx.doi.org/10.1088/1741-2552/aace8c

[16] M. Riyad, M. Khalil, and A. Adib, "A novel multi-scale convolutional neural network for motor imagery classification," *Biomedical Signal Processing and Control*, vol. 68, 7 2021.

[17] H. Li, M. Ding, R. Zhang, and C. Xiu, "Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network," *Biomedical Signal Processing and Control*, vol. 72, 2 2022.

[18] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[19] A. W. D. Boernama, N. A. Setiawan, and O. Wahyunggoro, "Multiclass Classification of Brain-Computer Interface Motor Imagery System: A Systematic Literature Review," in *AIMS 2021 - International Conference on Artificial Intelligence and Mechatronics Systems*. Institute of Electrical and Electronics Engineers Inc., 4 2021.

[20] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008-Graz data set A Experimental paradigm," Tech. Rep., 2008. [Online]. Available: https://www.bbci.de/competition/iv/

[21] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 1 2018. [Online]. Available: https://doi.org/10.1093/nsr/nwx106

[22] M. A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, 5 2018.

[23] Y. Xu, J.-Y. Zhu, E. I.-C. Chang, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Medical Image Analysis*, vol. 18, no. 3, pp. 591–604, 2014.

[24] J. Wang, B. Li, W. Hu, and O. Wu, "Horror video scene recognition via Multiple-Instance learning," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1325–1328.

[25] M. A. Guevara and M. Corsi-Cabrera, "EEG coherence or EEG correlation?" *International Journal of Psychophysiology*, vol. 23, no. 3, pp. 145–153, 1996.