1 Introduction

A brain-computer interface (BCI) is a type a communication system that allows users to interact with computers through mental effort alone [1]. Implementing a BCI relies on techniques from biomedical engineering, signal processing, neuroscience, and machine learning. As BCIs allow users to interact with their environment without peripheral nervous system involvement, many medical applications have been developed: for example, in stroke and spinal-cord injury rehabilitation [2]. A BCI may guide a patient though a session of repetitive, goal-based rehabilitative practice by quantifying the extent to which they are engaged with a motor training task. The aim of this research is to expand the motor-related brain states a BCI can detect. Specifically, we wish to explore bimanual motor intentions. A BCI with bimanual control would distinguish whether a user is intending to move their left hand, right hand or both hands together.

2 Research objectives

1. Explore the potential of using a brain-computer interface to predict unimanual and bimanual motor imagery
2. Determine to what extent healthy users can accurately control a bimanual brain-computer interface compared to neurologically impaired people, i.e. stroke and spinal-cord injured patients

3 Background

In order to use a BCI, two phases are typically required: an offline training phase during which the system is calibrated, and an online phase in which the system can interpret brain signals in real time and send commands to a computer. An online BCI system is a closed-loop: the user performs a task that modulates their brain activity – such as the imagination of hand movement – while their EEG is measured. Then, the signals are pre-processed using various spatial and spectral filers, and 'features' – for example, band power – are extracted, allowing the EEG data to be represented in a more compact form. Features are then classified and a command is sent to a device, providing user feedback. This feedback could be visual, acoustic [3], or tactile, depending on the application. Figure 1 illustrates this typical BCI design.

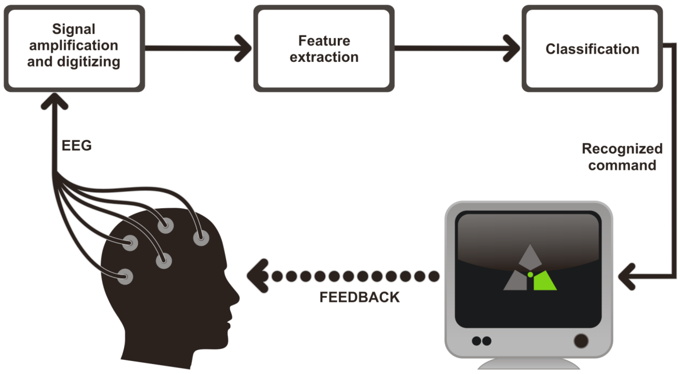


Figure 1: Classic BCI closed-loop: brain activity is measured with EEG and signals are represented as features (e.g. band power features), features relating to different mental states (e.g. left and right hand motor imagery) are labelled by a pre-trained classifier, which is translated into a type of feedback (e.g. visual, auditory, or tactile). Adapted from figure in [4].

4 Study: Classifying Uni- and Bimanual Motor Imagery as a Three-Class Problem

The research outlined in this section was conducted in the first year of the PhD, extending into the second. Increasingly the benefits of bimanual arm rehabilitation over unimanual training are being demonstrated [5]. The BCI community must reflect this trend by offering systems that can complement bimanual therapies. This study aims to take a step in this direction by evaluating the feasibility of classifying bimanual and unimanual motor imagery in healthy individuals and stroke patients.

4.1 Aims and Objectives

The aim of this study is to investigate the possibility of classifying three types of upper-limb motor imagery in a single BCI paradigm, the objectives include:

- Obtaining EEG data from healthy partcipants relating to left, right and bimanual upper-limb motor imagery
- Using labelled training data to create a BCI algorithm to accurately classify unseen test EEG data
- Verifying the system's performance by running an online session where new EEG data is classified in real time

4.2 Methods

4.2.1 Participants

With approval from the College of Science and Engineering Ethics Committee, healthy volunteers were recruited through personal contact. To date, 15 people have participated in this study (25.7±3.4 years old). Each session with a participant lasted 2-3 hours, with 40-90 minutes of setup and 50-60 minutes of EEG recording. The setup time was highly variable between participants as the impedance between the scalp and sensor depends to a large extent on hair thickness and scalp oiliness. In order to speed up the time spent reducing the impedance, participants were urged to wash their hair the day before, or the day of, the experiment.

In total, recording sessions took around 4 hours to complete. This includes pre-session setup (connecting amplifiers, fitting electrodes into cap; 30 minutes) and post-session clean up (putting equipment away, washing gel from electrodes and cap; 30 minutes).

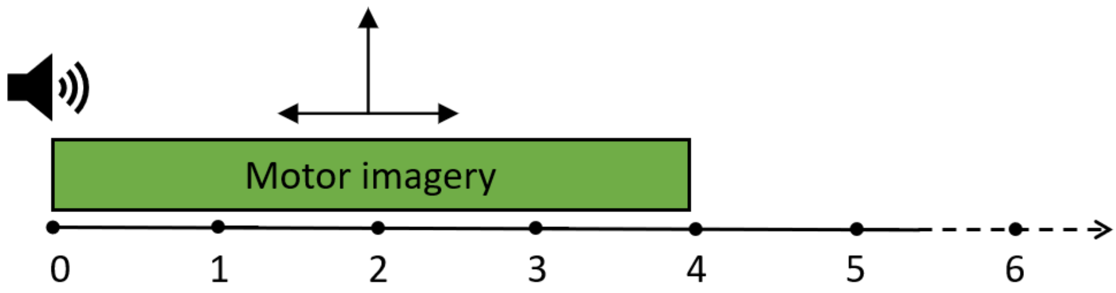


Figure 2: Trial timing. Each trial began with an audible beep and the appearance of an arrow pointing either left, right, or up, indicating to the subject to perform left hand, right hand or bimanual motor imagery, respectably. The arrow remained on the screen for 4 seconds and was followed by an inter-trial rest period of a random duration between 1.5 and 2.5 seconds.

4.2.2 Data recording

Subjects sat opposite a computer screen displaying a cue-based visual paradigm. Four g.USBamp amplifiers (g.tec, Graz, Austria) recorded EEG at 256 Hz from 64 electrodes positioned on the scalp. Abrasive gel was injected into each electrode to act as a conductive bridge between the scalp and sensor. Where the skin-electrode impedance was above 5 kΩ, it was abraded with a cotton bud until it fell to an appropriate level. Ground and reference electrodes were placed on the left and right earlobe, respectively.

The experiment was arranged into a series of trials, the timing of which is illustrated in Figure 2. Each trial began with an audible beep and the appearance of an arrow pointing either left, right, or up. A left arrow required subjects to imagine left arm movement. And vice-versa for the right arrow. An upwards pointing arrow required bimanual motor imagery, i.e. the combination of both left and right-upper-limb motor imagery. Subjects were instructed to sustain motor imagery for as long as the arrow remained on screen: 4 seconds. A 1.5-2.5 seconds inter-trail interval followed. To minimise subject fatigue, recording sessions were split into 7 runs of 45 trials with a short break in between. In total, each subject performed 315 trials relating to 105 trials of each condition.

4.3 Offline three-class classification and calibration

With the EEG recorded, we then explored signal processing and machine learning approaches for classification. These analyses were performed in Matlab. In general, two approaches were taken to classification: The first utilises state-of-the-art BCI techniques. Based on these results, and the algorithm's limitations, we devised a second classification algorithm.

Figure 3: EEG recording configuration. A participant performs a motor imagery task to a pace set by an onscreen cue-based paradigm whilst 64 channel EEG is recorded.

4.3.1 Pre-processing

Using Matlab-based application *Brainstorm*, raw EEG data was segmented into epochs from -2 to 5 seconds, relative to motor imagery onset. To reduce the effects of muscle artifacts, EEG was first bandpass filtered between 1 and 40 Hz using a 5th order Butterworth filter, and powerline noise was removed with a 50 Hz notch filter. The common average reference was applied and noisy epochs were rejected through a process of visual inspection. Around 10% of epochs were removed per subject.

4.3.2 Algorithm 1

To assess the potential for predicting three-class motor imagery EEG, we began with a state-of-the-art BCI pipeline: spatial filtering followed by linear classification. The common spatial pattern (CSP) algorithm takes labelled training data of two classes (say, left- and right-hand motor imagery), and creates a filter such that when it filters new data it maximises the variance for one class and minimises it for the other. The variance of the filtered trials are then used to build a feature matrix to train a linear discriminant analysis (LDA) classifier. This technique has proven successful in many previous works [6]. We calculated our CSP filters using data from 1 second to 3 seconds, relative to the cue, as this was determined optimal through experimentation. Four pairs of spatial filters were used to create 4 features per filtered EEG.

To extend this pipeline to three-class classification we used a one-versus-rest approach [7]. This involved creating three binary CSP filters, each designed to maximise the variance of one class whilst minimising the variance of the other two, see Figure 4. These filters were used to yield features that trained three LDA classifiers. Hence, when introduced to test data, the classifier with the most confident prediction score assigned the class label—for instance, if the left-versus-rest classifier returned the highest score for left-hand motor imagery, compared to the right-versus-rest and bimanual-versus-rest classifier, then this data segment was labelled "left". The accuracy of the entire approach was defined as its ability to correctly label evaluation data.

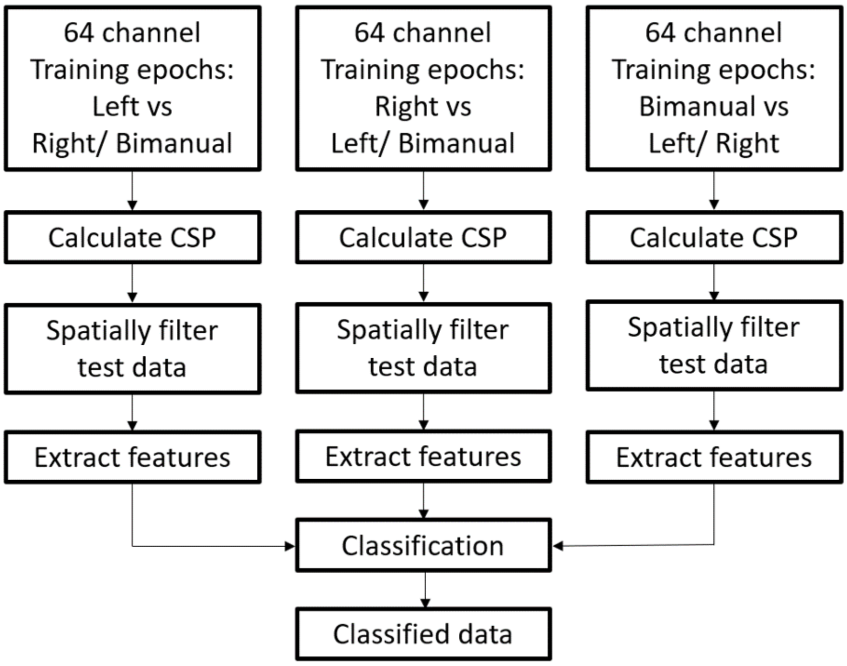


Figure 4: One-versus-rest approach to multiclass classification

As motor imagery tends to express itself within narrow frequency bands of the EEG's spectrum it was important to determine the most reactive band for each participant, rather than using a generic broadband (say, 8-30 Hz) [8]. The classification algorithm found the best frequency band by iterating over 18 different 4 Hz-wide bands from 1-40 Hz. The band that yielded the highest classification accuracy was used to evaluate the algorithm on unseen test data. The accuracy was measured using the 10-fold cross-validation score.

Other common classifiers were explored, namely the support vector machine (SVM) and *k*-nearest neighbour (*k*-NN) techniques. The results are outlined in an abstract submitted to the 8th International BCI Meeting 2020 (postponed to 2021 due to COVID-19), see Appendix A.

4.3.3 Algorithm 2

Algorithm 2 attempts to resolve some of the limitations of the first. EEG is a highly subject-specific biosignal, hence BCIs should be tailored to the user as much as possible [9]. Hence, we created a more sophisticated version of the frequency band selection method outlined above. Again, CSP-derived variance features were extracted from 18 different frequency bands. This time, however, we concatenated these features into a filter bank and found the mutual information between the features and their class label [10]. This allowed us to sort the features and choose only the most discriminatory. In general, this algorithm found the most important spectral components of EEG, whereas the previous algorithm ignored potentially useful information.

The algorithm also found the best number of features to use for classification. By iterating from 1 to N features, where N is the total number of features, the algorithm would find the fewest number of features that would minimise classification error.

4.3.4 Evaluation measures

The effectiveness of classification was evaluated by computing the number of correctly recognised class examples (true positives), the number of correctly classified examples that do not belong to the class (true negatives), and examples that either were incorrectly assigned to the class (false positives) or that were not recognises as class examples (false negatives). Accuracy describes the overall effectiveness of the classifier by relating the proportion of true positives and true negatives to the sum of all classified examples. This measure was extended for this three-class approach by averaging over each classifier.

$$AverageAccuracy = \frac{\sum_{i=1}^{3} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{3} \tag{1}$$

Precision is the number of true positives, tp, over the number of true positive plus the number of false positives, fp.

$$Precision = \frac{\sum_{i=1}^{3} \frac{tp_i}{tp_i + fp_i}}{3} \tag{2}$$

Recall is the number of true positives over the number of true positives and false negatives, fn. It describes the effectiveness of the classifier to identify positive labels.

$$Recall(Sensitivity) = \frac{\sum_{i=1}^{3} \frac{tp_i}{tp_i + fn_i}}{3} \tag{3}$$

A system with high recall but low precision returns many results, but most of its predicted labels are incorrect when compared to the training labels. A system with high precision but low recall is just the opposite, returning very few results, but most of its predicted labels are correct when compared to the training labels. An ideal system with high precision and high recall will return many results, with all results labeled correctly [11].

4.4 Results and discussion

Table 1: Results for Algorithm 1 and 2: Maximum accuracy (mean $\pm$ standard deviation) across all trials with the most discriminative frequency band as chosen by the feature selection algorithm. Best scores are highlighted in bold.

	Algorithm 1: CSP + LDA		Algorithm 2: FBCSP + LDA			
Subject	Accuracy (%)	BB (Hz)	Accuracy (%)	Precision (%)	Recall (%)	BB (Hz)
1	$\mathbf{78 \pm 7}$	6-10	83 ± 8	83 ± 4	84 ± 4	4-8
2	46 ± 11	7-11	$\mathbf{87 \pm 7}$	$\mathbf{87 \pm 4}$	$\mathbf{88 \pm 4}$	4-8
3	77 ± 10	2-6	75 ± 10	75 ± 4	76 ± 5	2-6
4	72 ± 9	4-8	67 ± 7	67 ± 6	68 ± 6	4-8
5	62 ± 7	3-7	64 ± 11	64 ± 6	63 ± 6	4-8
6	64 ± 11	8-12	76 ± 8	76 ± 6	77 ± 6	10-14
7	68 ± 8	3-7	71 ± 2	71 ± 3	71 ± 3	4-8
8	68 ± 8	1-5	64 ± 8	64 ± 2	65 ± 2	2-6
9	50 ± 10	4-8	49 ± 8	49 ± 4	49 ± 3	4-8
10	57 ± 9	2-6	60 ± 6	59 ± 6	59 ± 7	2-6
11	54 ± 6	5-9	47 ± 10	47 ± 3	49 ± 5	2-6
12	49 ± 8	15-19	71 ± 4	71 ± 4	71 ± 4	4-8
13	49 ± 5	1-5	51 ± 4	51 ± 4	50 ± 4	2-6
14	70 ± 11	1-5	76 ± 7	78 ± 4	78 ± 4	2-6
15	75 ± 8	7-11	80 ± 8	83 ± 4	84 ± 4	6-10
Mean	62 ± 8	5-9	$\mathbf{68 \pm 11}$	68 ± 12	68 ± 12	4-8

Acronyms: BB - best band

Table 1 shows the results of both algorithms. The first algorithm, which used a CSP and LDA trained on EEG from a single frequency band, was able to predict correct motor imagery with an accuracy of 62%. All subjects returned accuracies greater than the threshold for chance level classification: 33%. Five subjects scored above 70%, suggesting they would be able to use a BCI without the need for further training. This is an encouraging score for a first attempt at a three-class problem with a bimanual condition.

Algorithm 2, which found the best features to use across multiple frequency bands, yielded 68% accuracy, a statistically significant increase upon Algorithm 1 ($p<0.05$ with a paired *t*-test). A 2018 study reported similar classification accuracies with a two-class BCI (right versus bimanual motor imagery) [12], suggesting the inclusion of a third class comes at no additional cost in terms of accuracy. The precision and recall are the same for Algorithm 2, $64 \pm 12\%$, implying the algorithm is as effective at predicting positives as it is at predicting negatives.

The difference between Algorithm 1 and 2 is even more stark within individual subjects: Subject 2, for example, shows a 37% performance increase with the new method, a vast improvement. Some subjects do not show significant improvement, but the general trend is clear: a filter-bank CSP (FBCSP) approach results in greater classification accuracy.

Figure 5 shows a more detailed breakdown of the algorithm's ability to discriminate between classes. Averaged over all subjects, the left class appears to have the most misclassifications with a true positive rate of 65%. The bimanual and right classes suggest equal disciminability with 67% and 67%, respectively.

Figure 5: Algorithm 2: Group level confusion matrix

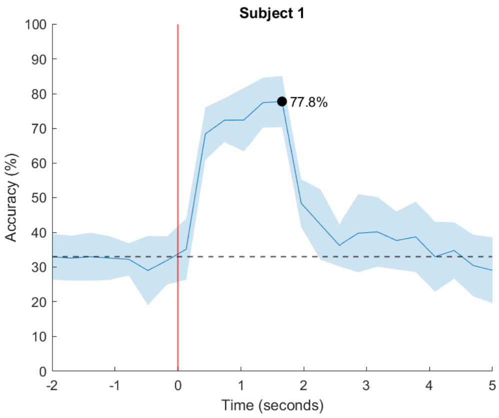
BCIs are difficult to evaluate as their performance is often determined by factors beyond the system's classification ability. The mental effort exerted by the user, the properties of their head (skull/skin thickness), and brain topography all contribute to the final score, amongst many others. In fact, it is widely reported that around 15-30% of participants are unable to control a BCI with high accuracy due to a phenomenon called BCI illiteracy [13]. A subject-wise average is not always representative of a BCI's true performance.

Figure 6 shows the classification accuracy of the best and worst performing participants averaged across trials. The cue to begin motor imagery begins at zero seconds, notice the prediction accuracy is at chance level before this. After the cue, the discriminability immediately rises to a maximum classification accuracy at around 1-2 seconds post-cue. Subject 1 achieved an excellent score with a 82.6% prediction rate, indicating encouraging potential for a bimanual BCI. Subject 13, the worst performing subject with Algorithm 2, peaked at 50.9%, which, although well above chance level, would prove frustrating in practice. This subject would benefit from additional training sessions in order to improve their ability to modulate the brain rhythms [14].

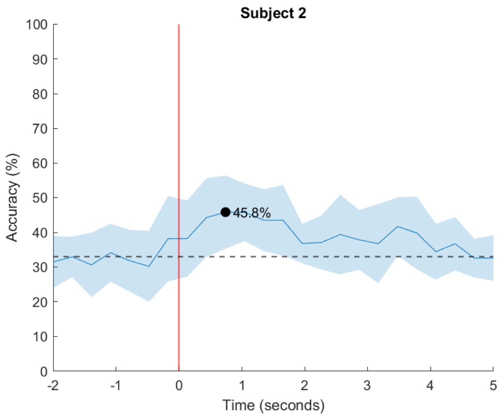
The effectiveness of classification of individual classes is illustrated in Figure 7. This figure presents the true positives for each class on the diagonal, and the mislabelled instances on the off-diagonal values. For Subject 1, the right class appears the most discriminable, with a 87% true positive rate. The left and bimanual classes were roughly as discriminable as each other with 81%, and 82%, respectively. Compared with Subject 13, misclassificaitons are more abundant, with much greater values in the off-diagonals than for Subject 1. Bimanual classification is favoured for Subject 13, 57%, compared to 51% and 53% for the left and right classes, respectively. The difference in distrubution between the classes for Subject 1 and 13 may be explained by their neurophysiology: perhaps the differences in spatial patterns detected by EEG is more stark between the mental states for Subject 1 than for Subject 13, aiding separability. This could be confirmed by performing a time-frequency analysis for each subject.

Another benefit of Algorithm 2 is that the feature selection method allows greater insight into the neural mechanisms underpinning the three types of motor imagery—that is, the spectral components generated by the task. Figure 8 displays the spectral components most selected by the feature selection stage for each motor imagery condition. Further analysis is required to draw firm conclusions from these results, but in general it appears that lower frequency ranges (2-14 Hz) contain more relevant information than the higher bands (12-22 Hz). Taking Subject 6 as an example, it may also be the case that frequencies in the traditional μ band (8-12 Hz) are more important to left and right hand motor imagery than bimanual motor imagery, which appears to involve slower rhythms (2-6 and 4-8 Hz) to a greater extent.

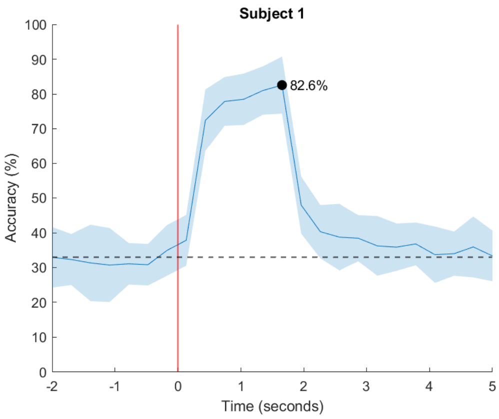
These novel results suggest that a BCI could be used to distinguish between unimanual and bimanual motor imagery with a high degree of accuracy in a subset of participants.



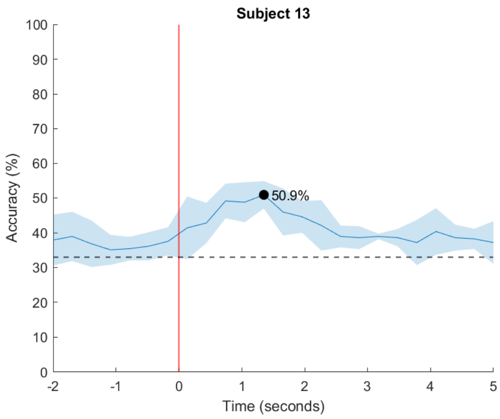
(a) Algorithm 1: Highest score



(b) Algorithm 1: Lowest score



(c) Algorithm 2: Highest score



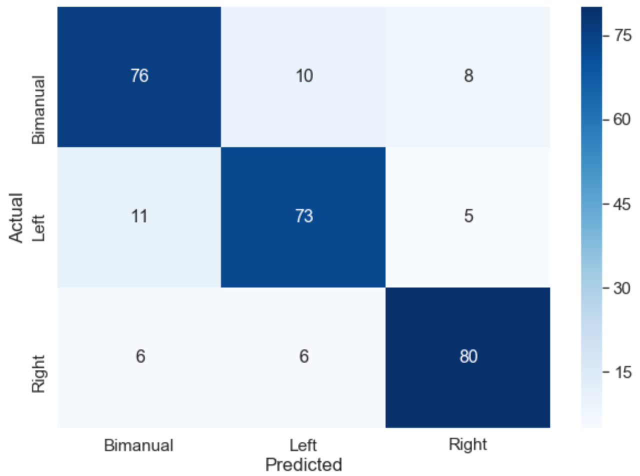
(d) Algorithm 2: Lowest score

Figure 6: Classification accuracy over time for the best and worst performing subjects. The horizontal line marks the beginning of motor imagery, and the dashed line indicates chance level. (a) and (b) relate to Algorithm 1, and (c) and (d) relate to Algorithm 2.

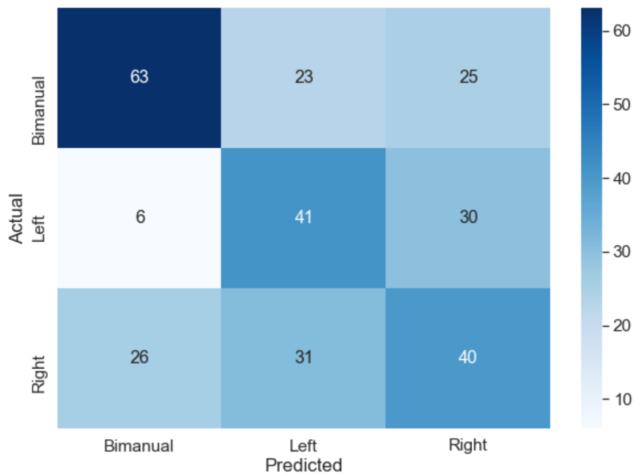
4.5 Going forward

The next step in this study is to run an *online* analysis. Can the above accuracies be replicated in real-time? Online BCIs offer performance-dependent feedback, allowing users to adapt to the system, often resulting in improved performance with repeated sessions as individuals learn how to modulate their brain rhythms [14].

Another future goal is to test the BCI on potential end users: stroke patients. We will recruit patients from community support groups in the Greater Glasgow and Clyde region. Our initial contact with these groups have proved encouraging, with many individuals expressing an interest in participating. University ethical approval has been granted but



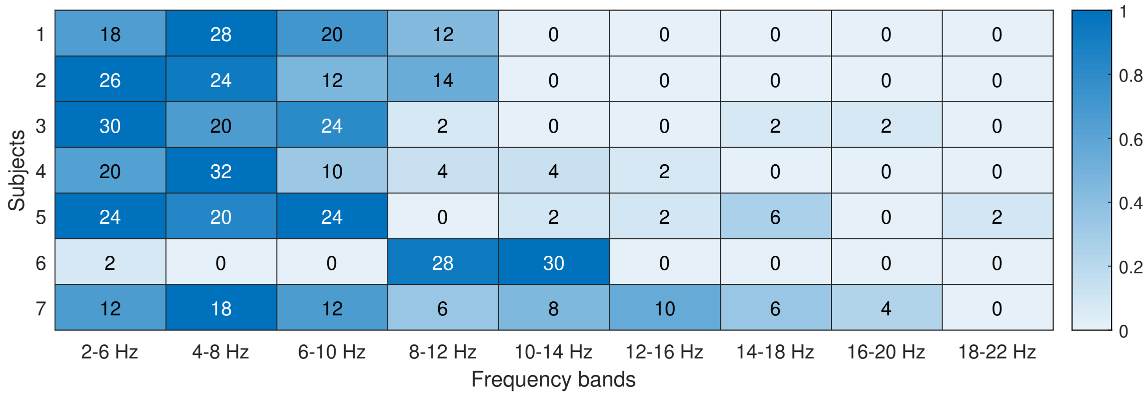
(a) Algorithm 2: Subject 1 confusion matrix



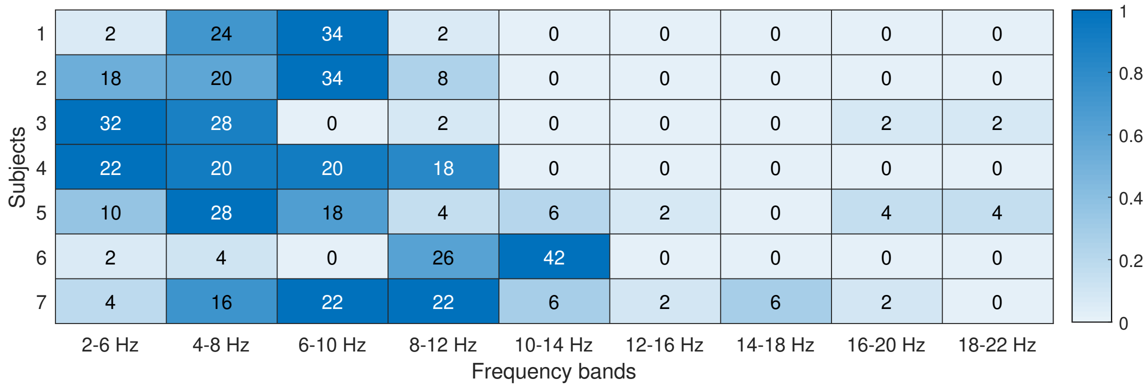
(b) Algorithm 2: Subject 13 confusion matrix

Figure 7: Confusion matrices for the best and worst performing subjects using Algorithm 2. The diagonal elements represent the number of examples which were correctly classified, and the off-diagonals were mislabelled by the algorithm.

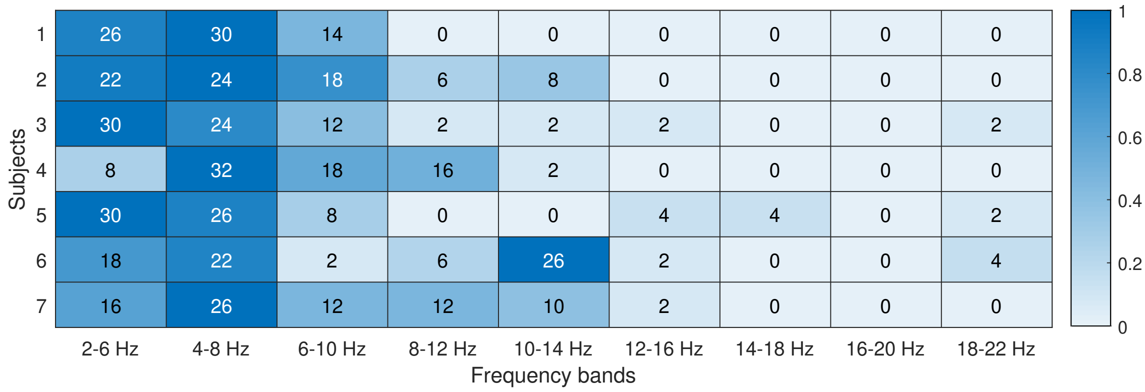
given the current situation regarding COVID-19 it is unclear when we will begin recording sessions with stroke patients.



(a) Left



(b) Right



(c) Bimanual

Figure 8: Histogram of features selected from specific filter bands for each classifier based on the OVR approach using all the training data. The vertical axis in each chart represents Subjects 1 to 7 from top to bottom. The horizontal axis represents the filter banks.