

Multi-segment majority voting decision fusion for MI EEG brain-computer interfacing

Correspondence Address: Centre for Signal and Image Processing, University of Strathclyde, Glasgow, G1 1XW, UK.

Authors: Natasha Padfield, Jinchang Ren, Paul Murray, Huimin Zhao

Addresses:

Natasha Padfield, Jinchang Ren and Paul Murray are with the Centre for Signal and Image Processing, University of Strathclyde, Glasgow, G1 1XW, UK.

H. Zhao is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China.

Corresponding Author: Jinchang Ren. *Email address:* jinchang.ren@strath.ac.uk. *Telephone:* +44(0)141-5482384.

Structured Abstract

Background/Introduction: Brain-computer interfaces (BCIs) based on electroencephalogram (EEG) generated during motor imagery (MI) have the potential to be used in brain-controlled prosthetics, neurorehabilitation, and gaming. Many MI EEG classification systems segment EEG into windows for classification. However, a comprehensive analysis of decision fusion based on segmented EEG data, within the context of different classifiers, has not been carried out.

Methods: This study presents a multi-segment majority voting (MSMV) decision fusion approach in which an EEG trial is segmented using overlapping windows, segments are labelled, and then a final classification label for the trial is derived through majority voting. Common spatial pattern (CSP) features were used. The impact of the MSMV approach on the classification accuracy of six classifiers was investigated. The effects of window size and overlap were analysed. Results were generated using five different subsets of EEG channels and channel subsets for static EEG analysis are proposed. The BCI Competition III dataset IVa was used.

Results: The MSMV decision fusion approach was found to significantly improve classification accuracy for linear discriminant analysis (LDA), support vector machine (SVM), naïve-Bayes (NB) and random forest (RF) classifiers. Classification accuracy was improved by 5.02%, 4.41%, 1.25% and 3.62% for the SVM, LDA, NB and RF classifiers, respectively. The channel analysis indicated the importance of central-parietal and central-frontal electrode regions for MI EEG classification.

Conclusions: MSMV decision fusion improved MI EEG classification performance and could be considered for future studies, particularly in online systems that deal with buffered data.

Keywords: channel montage, decision fusion, electroencephalography, motor imagery

Introduction

The process of imagining movements, known as motor-imagery (MI), is an important aspect of neurorehabilitation and be used to intuitively control prosthetic technology. Research has been focused on the modelling and classification of signals generated during MI tasks, with the aim of developing novel brain-computer interfaces [1].

During a MI task, characteristic changes occur within the *mu* and *beta* frequency bands of the EEG signal, which are the (8-12) Hz and (12Hz-30Hz) bands, respectively [2]. The onset of a MI event is characterised by a decrease in power within the *mu* and *beta* bands, known as event-related desynchronization [2]. MI activity originates within the sensorimotor area of the brain and shows most strongly on the central EEG electrodes, although significant changes in frequency content have also been observed in the parietal and frontal electrode regions [3].

Real-time BCIs can use buffered EEG data to make decisions [4, 5]. The buffered data is a window of EEG data from which features can be extracted and used for classification. This buffering can be replicated in a windowing-based offline system. This process incrementally moves a window over the data in an EEG trial, and at each position of the window a feature vector is extracted and used for classification. The windowing approach thus presents two design choices: window size and increment size.

Many offline processing systems use windows of arbitrary length, and there is high variability in approaches [6, 7, 8, 9,10,11,12]. Some studies use the entire EEG trial, the length of which varies depending on the dataset, but 3.5 second and 4-second-long trials are common [10, 11]. Other studies use maximally overlapped windows [12] and some use increments of milliseconds [9].

Some approaches segment EEG data only to perform data augmentation due to the limited number of EEG trials available in datasets [9]. Samuel et al. [9] found that using a 100ms window with 25ms increments gave the best results for data augmentation in an FFT-based system, resulting in 99.79% accuracy. The data segmentation was carried out prior to division of the data into training and test sets, possibly leading to parts of trials existing within both the training and testing sets, leading to a possible bias.

Other approaches use segmented EEG data to facilitate post-processing based on the labels obtained for each temporal segment of an EEG trial. Using a CSP-LDA pipeline, Asenio-Cubero et al. [7, 8] segmented EEG trails and obtained a classification label for each trial. The final label for the EEG trial was obtained through a majority vote. Asensio-Cubero et al. [7] found that segmenting the data resulted in better classification when the window size was over 2s and concluded that using overlapped windows and a majority voting approach led to the best classification accuracy. However, there is a lack of research into the relationship between accuracy, windowing approach and classifier type for majority voting-based MI EEG classification. Furthermore, a recent study on electromyography (EMG) signal classification [13] found that a majority voting-based post-processing system had the potential to improve the classification accuracy of various conventional classifiers. EMG signals

are created by electrical signals within muscles [13] and, like EEG signals, are noisy and non-stationary biosignals. The main aim of this paper is to investigate the possible benefits of using majority voting decision fusion post-processing for MI EEG classification within the context of various classifiers.

Many studies compare the performance of different classifiers used for MI EEG classification. Frequently, support-vector machines (SVM) have been used [12, 14, 15] and have been found to outperform linear discriminant analysis (LDA), naïve Bayes (NB) and random forest (RF) classifiers [12, 14, 16]. However, comparative studies of classifiers have not analysed performance within the context of majority voting decision fusion for MI EEG classification. This paper fills this gap in the literature.

The subsets of EEG channels chosen for MI EEG classification within the literature are diverse. A whole field of study has been devoted to deriving channel selection algorithms that aim to choose the optimum subset of EEG channels for a particular subject [17] or MI task [18, 19]. Different approaches have been used to select channels, including genetic algorithm (GA)-based methods [17], iterative weight-based techniques [18], ranking using features [19] and analysing signal properties [20]. However, many studies still use a static, arbitrarily chosen set of channels for signal processing and focus on presenting novel signal processing techniques [14, 10, 22]. These arbitrarily chosen groups are varied, with some using all channels available within a dataset [14, 21], others depending only on central channels C3, Cz and C4 [10] or some other subset of EEG channels [22]. There is conflicting evidence as to whether increasing the number of EEG channels used improves results [22, 23].

Different EEG scalp regions are related to different mental tasks. The central region has been linked to imagined and executed motor activity and is widely used for MI EEG analysis. The frontal region is associated with the control of voluntary behaviour [24] and the parietal region has been associated with focused attention [25]. In this paper, the proposed system was assessed using different EEG channel subsets, which were constructed from different combinations of electrodes from the central, central-parietal, and central-frontal regions.

This paper presents an EEG classification system based on common spatial pattern (CSP) feature extraction [3, 17, 18, 19] and a multi-segment majority voting (MSMV) decision fusion post-processing. Each EEG trial was segmented and the final classification label for a trial was obtained through a majority vote of the labels assigned to each segment. The impact of the segmentation approach on classification accuracy was investigated. The impact of the MSMV post-processing step was investigated within the context of six classifiers, namely LDA, NB, RF classifiers, and SVM classifiers with linear, polynomial, and radial basis function (RBF) kernels. Furthermore, the impact of MSMV decision fusion was investigated using five different static EEG

channel subsets. When using a window size of 2s and a window increment size of 0.5s, the MSMV decision fusion technique led to a statistically significant improvement in classification accuracy for five of the six classifiers. The majority voting decision fusion step was also shown to improve the sensitivity of the classifiers. Peak classification performance was obtained using the LDA classifier and the SVM classifier with RBF kernel. In both instances, MSMV decision fusion led to an improvement of approximately 5%.

This study also showed that using channel subsets comprising of the central, central-parietal and/or central-frontal channels significantly improved classification accuracy and sensitivity when compared to using the full cohort of EEG channels. Many studies use the full cohort of EEG channels available [14, 21] however, from the cross-classifier analysis in this study, subsets using the central, central-parietal, and central-frontal electrode groups should be considered for static channel subsets in these studies. To our best knowledge no other study has investigated the effect of designing a static channel subset based on scalp regions in this way.

Methods

This section begins with a description of the dataset used, and then presents the proposed classification approach. It then discusses the EEG channel subsets used, hyperparameter tuning, and the evaluation approach.

Dataset Description

Open-access data from the BCI Competition III, dataset IVa was used [26]. The dataset consists of EEG data with 118 channels collected from 5 subjects, labelled as aa, al, av, aw and ay. Subjects were asked to perform two classes of MI activities: imagining movement of the right hand (class 1) and movement of the right foot (class 2). 140 trials were recorded for each class and trials were 3.5s long. Data was recorded at a sampling frequency of 1000Hz and was downsampled to 100Hz. No artefact removal was carried out.

The Proposed MSMV Decision Fusion Approach

The MSMV decision-level fusion approach involved dividing the data for a trial into segments, extracting a feature vector from each segment, acquiring a classification label for each feature vector, and then obtaining the final label for the trial using majority voting, where the most popular label is assigned to the trial. Figure 1 shows a graphical abstract of the methodology.

Each trial was first pre-processed. In the pre-processing step, the selected subset of channels for a given trial was passed through an elliptic bandpass filter with passband from 8Hz to 32Hz that extracts the *mu* and *beta* frequency bands [2]. The data was then mean-centered. A standard CSP feature vector was extracted from each

segment [21]. CSP features were used because they are widely adopted in the literature [3, 17- 19].

Windowing was used to divide each trial into segments. A window of size (length) x_s was moved over the data in increments of y_s , and for each increment of the window, a feature vector was obtained. Thus, the first position of the window had a start point at $0s$ and endpoint at x_s , and the second position of the window had a start point at $(0s+y_s)$, and endpoint at (x_s+y_s) , and so on until the endpoint of the window reached the end of the trial. Table 1 shows different windowing schemes tested in terms of window length and increment size pairings. Segmentation schemes are denoted as (window size in s , window increment size in s) in this paper.

The effect of MSMV decision fusion was investigated using six classifiers, namely: SVM classifiers with linear, polynomial, and RBF kernels, denoted as SVM-linear, SVM-poly and SVM-RBF, respectively, and LDA, NB and RF classifiers.

Channel Subsets

The impact of MSMV decision fusion on classification performance was studied within the context of five different EEG channel subsets. Figure 2 shows the standard electrode placement for EEG recording and the following five electrode montages were used in this study:

- 1) Central channels (C): C5, C3, C1, Cz, C2, C4, C6, a total of 7 channels;
- 2) Central and Central-Parietal channels (C+CP): C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, a total of 14 channels;
- 3) Central and Central-Frontal channels (C+CF): C5, C3, C1, Cz, C2, C4, C6, FC5, FC3, FC1, FCz, FC2, FC4, FC6, a total of 14 channels;
- 4) Central, Central-Parietal and Central-Frontal (C+CP+CF): C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, FC5, FC3, FC1, FCz, FC2, FC4, FC6, a total of 21 channels;
- 5) All available 118 channels in the dataset.

Tuning of Classifier Hyper Parameters

For each EEG subset, the hyperparameters of each classifier were tuned using a grid search approach. During a grid search, a hyperparameter space was constructed using different combinations of hyperparameters for the classifier. Afterwards, each set of hyperparameters was used to evaluate the 10-fold cross-validation classification error of every subject in the dataset and the average classification error across the five subjects was calculated.

The hyperparameter set with the lowest error was selected for further testing and analysis. The following values were used in the grid searches:

- **LDA classifier:** Values of linear coefficient threshold (Δ) from 2^{-10} to 2^{10} , and including the value of zero, and values of regularization coefficient (γ) from 2^{-10} to 2^{-1} as well as the values 0, $2^{-0.5}$ and $2^{-0.25}$. Values were spaced in a geometric progression with common ratio 2.
- **SVM classifiers:** For linear, RBF and polynomial SVMs, the search for the regularization parameter (C) and the kernel scale (g) spanned values from 2^{-10} to 2^{10} , spaced in a geometric progression with common ratio 2. For the SVM-poly classifier, the search was carried out for polynomial orders of 2 and 3.
- **NB:** The kernel functions Gaussian, Epanechnikov, box and triangular were included in the search. Box widths in $[10^{-1}, 10^1]$ with the power increasing in steps of $\frac{1}{50}$ were considered.
- **RF:** The number of trees was increased from 20 to 1000 in steps of 20, node-based predictions were increased from 1 to 8 in steps of 2, and the observations per leaf were varied from 4 to 20 in steps of 4.

All classifiers were implemented in MATLAB and the parameters were derived without segmenting the EEG data.

Evaluation of MSMV Decision-Level Fusion

10-fold cross-validation was used to evaluate performance. In the training phase, the trials were segmented and used to train the classifier. In the testing phase, the majority voting step was carried out to obtain the final classification labels. To evaluate the efficacy of the majority voting step, the same hyperparameters found through the grid search, which was carried out on the non-segmented EEG data, were used.

Performance Evaluation

Classification performance was measured using accuracy and sensitivity, calculated with the same approach as O'Reilly et al. [27]. The ideal value for these measures is 100%. Common statistical tests, namely the Wilcoxon signed rank test and the Kruskal-Wallis test were used to analyse results. The Pearson correlation coefficient was also used in analysis. A 0.05 level of significance was used for all statistical tests.

Results

The impact of MSMV decision fusion on the accuracy of each classifier is recorded in Figure 3. Plots of median accuracy against window size are shown, with separate plots for each increment size being displayed. The increment size is indicated by the marker used, with circular, square and diamond markers denoting the increment sizes 0.5s, 0.25s and 0.1s, respectively. The value for each data point was found by calculating the average of the channel subset results across subjects, and then finding the median result across the five channel subsets (C, C+CP, C+CF, C+CP+CF, 118). The horizontal blue lines represent the control cases for each classifier, when no multi-segment decision fusion is used. Data points representing statistically significant changes from the control case are highlighted in red. A Wilcoxon signed rank test was used for the statistical analysis since the data was found to be non-normal following an Anderson-Darling test. The median results were plotted since the Wilcoxon signed rank test uses the distribution about the median value to determine statistical significance.

Figure 4 shows the results in greater detail, illustrating how average classification accuracy varies with window size for all classifier and subset pairings. To create the plots, the classification results for each windowing scheme were averaged across subjects, then averaged across the window increment sizes (0.5s, 0.25s and 0.1s). The horizontal line in each plot represents the control case, which is the average classification accuracy obtained when the whole trial is classified without MSMV decision fusion.

Table 2 records the peak classification accuracy obtained for each classifier and the corresponding sensitivities to classes 1 and 2. It also includes the details of the configuration which resulted in peak performance, where ‘configuration’ refers to the channel subset, window size and window increment size. The ‘Control’ results were those obtained for the configuration channel subset but without using MSMV decision fusion.

A two-part correlation analysis was carried out: the first part (‘part 1’) investigated if there was a significant correlation between window size and classification accuracy, and the second part (‘part 2’) investigated if there was a significant correlation between window increment size and classification accuracy. In both cases, the correlation was evaluated for each classifier separately. To obtain the results used for analysis, the accuracies for each segmentation approach were averaged across subjects, and then the median values across the channel subsets were used. In part 1 the correlations between window size and accuracy were evaluated separately for each channel increment size. In part 2, the correlations between window increment size and accuracy were evaluated separately for each window size.

The results for parts 1 and 2 are recorded in Tables 3 and 4, respectively. Table 3 records the Pearson correlation coefficient, ρ , and corresponding p-value obtained for each classifier and window increment size pairing, and Table 4 records the value of ρ and corresponding p-values for each classifier and window size pairing. Values of ρ that were statistically significant ($p\text{-value}<0.05$) are highlighted in italics.

In Table 3, statistically significant values of ρ are always positive, indicating that larger window sizes resulted in higher classification accuracies. For the LDA, SVM-linear and SVM-poly classifiers, there is a significant positive correlation between classification accuracy and window size regardless of increment size. There was no significant correlational relationship between the accuracy of the SVM-RBF classifier and window size for any window increment size. In the case of the NB classifier, there was a significant relationship between window size and classification accuracy when a window increment of 0.5s was used, whereas for the RF classifier there was a significant relationship when an increment of size 0.1s was used. The results in Table 4 indicate there was no significant correlation between classification accuracy and window increment size at any of the window sizes.

A channel subset analysis was then carried out. Figure 5 is a box plot showing how average classification accuracy, sensitivity to class 1 and sensitivity to class 2 varied for different channel subset groups. The plot was constructed from the average results obtained from all six classifiers for the segmentation scheme (2s,0.5s) which, as illustrated previously in Figure 3, resulted in statistically significant improvements in classification accuracy for all classifiers except the SVM-poly classifier. Since the data was found to be non-normal following an Anderson-Darling test, a Kruskal-Wallis test was used to determine whether there were any significant differences in accuracy, sensitivity to class 1 and sensitivity to class 2 across the channel subsets. There was a significant difference in classification accuracy ($p=0.022$) and sensitivity to class 2 ($p=0.0155$), however there was no significant difference in the sensitivity to class 1 ($p=0.4144$).

Table 5 [14, 17, 20, 21, 28, 29] compares the peak results obtained in this study to results in the literature from other conventional and deep learning systems.

Discussion

This Section first discusses the effectiveness of the MSMV decision fusion approach for MI EEG classification. It goes on to examine peak classification performance, and then discusses the impact of window size and window

increment size on classification accuracy. The effectiveness of the different channel subsets used was then analysed. Finally, the Section closes with a comparison to the literature.

Analysing the Impact of MSMV Decision Fusion on Performance

The results in Figure 3 illustrate that multi-segment decision fusion can have a significant effect on classification accuracy for all classifiers except the SVM-poly classifier. In fact, for all classifiers except the SVM-poly, using a window size of 2s and increment sizes of 0.5s or 0.25s resulted in a significant improvement in classification accuracy. The LDA, NB and RF classifiers were most likely to experience a significant improvement in classification accuracy with MSMV decision fusion; the LDA and NB classifiers experienced a significant improvement in classification accuracy for 73% of the (window size, window increment size) pairings, and the RF classifier experienced an improvement with 52% of the pairings. The SVM-linear and SVM-RBF classifiers experienced a significant improvement in classification accuracy for only 5% of the (window size, window increment) pairings. When a window size of less than 1s was used, MSMV decision fusion could lead to a significant decrease in performance. For example, the windowing scheme (0.5s,0.5s) led to a significant decrease in performance for the LDA, SVM-linear and SVM-RBF classifiers, and the LDA and SVM-linear classifiers also experienced a significant decrease in performance for the windowing schemes (0.5s,0.1s) and (0.75s,0.1s). The results in Figure 3 indicate that the effectiveness of MSMV decision fusion depends strongly on the classifier used and on the size of segmentation window. Notwithstanding this, since the segmentation schemes (2s,0.5s) and (2s,0.25s) gave a significant improvement in classification accuracy for five of the six classifiers MSMV with these schemes would be recommended for further analysis.

Figure 4 shows the average accuracy results for all classifier and channel subset pairings. As was observed in Figure 3, larger window sizes tend to result in greater accuracy. For example, when a window size of 2s is used, 93.33% of classifier-channel subset pairings have an improved classification accuracy when compared to the control case, but when a window size of 0.5s is used, only 33% of classifier-channel subset pairings had an improved performance. The correlation analysis later in this section further explores the relationship between window size and performance.

Peak Performance

Considering the results in Table 2, the overall peak average accuracy was 84.50%, and this was obtained for two different classifiers: the LDA classifier and the SVM-RBF classifier. In both cases, the C+CP channel subset and the windowing scheme (2s,0.1s) gave peak performance. Without MSMV decision fusion, the LDA classifier

obtained an accuracy of 80.93%, and the RBF classifier obtained an accuracy of 80.76%, indicating that in both cases the MSMV decision fusion approach led to an improvement in the classification accuracy of approximately 5%. The averaged results in Table 2 indicate that MSMV decision fusion had the potential to boost the accuracy by an average of 3.70% across all classifiers. MSMV decision fusion also tended to improve the sensitivities of the classifiers to both classes in all cases except for the naïve Bayes classifier's sensitivity to class 2, which decreased by 1.1%.

Correlation Analysis

Part 1 of the correlation analysis was carried out to identify the relationship between window size and performance. The results in Table 3 confirm that a larger window size was generally correlated with greater classification accuracy. This result is in agreement with the observations made in relation to Figures 3 and 4, and the statistical analysis associated with Figure 3, which concluded that for a window size of 2s there was a statistically significant improvement in classification accuracy for all classifiers except the SVM-poly classifier, which experienced no significant change. These results also concur with the overall peak performance results obtained by the LDA and SVM-RBF classifiers in Table 2, which were linked to a window size of 2s.

However, not all the classifiers in Table 2 had a peak performance that coincided with a window size of 2s. The RF and SVM-linear classifiers had a peak performance associated with window sizes of 1.5s and 1.75s, respectively. This could be because for window sizes below 2s the EEG data can tend to approach approximate stationarity. For example, a window of 1.25s has been used under the assumption of approximate stationarity [30,31], and it is known that CSP feature extraction suffers when data is non-stationary [32]. However, windows which are smaller than 1s often lead to lower accuracies, as shown in Figure 3, possibly because there was not enough data in the window for the CSP features to be strongly discriminative. These results may indicate that choice of window size for some classifiers is a trade-off between two opposing factors: 1) smaller windows leading to data which is closer to stationarity, and 2) larger windows capturing enough data for more discriminative classification.

Part 2 of the correlation analysis, presented in Table 4, confirmed that there was no significant correlation between accuracy and window increment size. However, larger increments result in fewer segments and therefore less computational processing. Previously, from Figure 3, it was established that window increments of 0.5s and 0.25s were both linked to significantly improved accuracy. Based on this correlation analysis, a window increment size of 0.5s would be recommended.

Channel Subset Analysis

Considering the accuracy results in Figure 5, channel subset C+CP+CF had the highest median accuracy, outperforming even the case when all 118 EEG channels were used. Channel subsets C+CP and C+CF also had higher median accuracies than the full set of 118 channels. However, using just the central channels resulted in a substantial drop in median accuracy. Consider the plot of sensitivity to class 2: using channel subsets C+CP+CF, C+CP and C+CF resulted in an increase in the median sensitivity when compared to using 118 EEG channels and using just the central channels resulted in a notable decrease in median accuracy. Therefore, the trends in the accuracy results are mirrored in the results for sensitivity to class 2, and the differences across channel subsets were statistically significant for both performance measures. In the case of sensitivity to class 1, there was no statistical difference in results across the channel subsets. These results suggest that the improvement in classification accuracy obtained using channel subsets C+CP+CF, C+CP and C+CF relative to using 118 channels resulted from an improvement in sensitivity to class 2 obtained using the channel subsets. This improvement may have been due to a reduction in noisy or redundant channels. There is still an open debate [22, 23] in the literature about whether reducing the number of EEG channels can improve classification accuracy, however from the results of this cross-classifier study, the channel subsets C+CP+CF and C+CP gave the highest accuracy and would be recommended.

Comparison to the Literature

Although the core aim of this study was to investigate whether MSMV decision fusion could be used to improve the classification accuracy of an MI EEG classification system, the results were compared to the literature for completeness. Considering the results from classical implementations as highlighted in italics in Table 5, the MSMV decision fusion approach offered better performance. The key differences between the classical systems in [14, 20] and the approach in this paper were the MSMV decision fusion step and the use of EEG channel subsets.

The results obtained from the MSMV decision fusion approach outperformed several conventional processing techniques [20, 21, 28]. Yang et al. [20] used algorithmic channel selection through time-series analysis, obtaining 9 subject specific channels and an accuracy of 78%. Like Yang et al. [20], our study used CSP features and an LDA classifier, but obtained an average classification accuracy of 84.50%. Our implementation had the added benefit of using a universal, static, channel subset which did not involve algorithmic channel selection pre-processing. Olias et al. [21], used normalized CSP and tangent space logistic regression and obtained

a classification accuracy of 82%, which was lower than the proposed approach. The MSMV decision fusion approach also outperformed the CSP feature extraction approach reported by He et al. [17] which used the Rayleigh coefficient to increase discriminative performance.

However, the MSMV decision fusion approach was outperformed by some implementations that use algorithmic channel or feature selection techniques [14, 17]. For example, Baig et al. [14] used differential evolution to select a subset of 13-18 features from a total of 236 features extracted from the 118 EEG channels. He et al. [17] used a genetic algorithm to obtain subject-specific subsets of EEG channels that were then used for classification. Differential evolution [14] and genetic algorithms [17] are wrapper selection techniques which introduce significant additional computational expense when compared to using static EEG channel subsets, as in our study. Our study has confirmed that MSMV can improve classification performance across several EEG channel subsets. A future step could involve applying the MSMV decision fusion approach to systems using algorithmic channel selection.

Consider the results in Table 5 from deep learning-based systems. Our results were on a par with those obtained by She et al. [28], however they were inferior to those obtained by Kumar et al. [29]. Notwithstanding this, deep learning approaches, such as CNN-based systems [29], require significant tuning and training times which could impact their practical use for the time being.

The results for subject *aw* were poor when compared to some of the other methods. This may have been due to two factors: i) the grid search not covering the optimal classifier parameters for *aw*, and ii) the choice of channel subset may not have been optimal for *aw*. Carrying out a new grid search for the classical LDA and SVM-poly classifiers, for subsets C+CP+CF and C+CP respectively, the accuracy was increased to 83.93% and 83.21%. Optimal parameters were $\Delta = 0.0001$ and $\Upsilon = 0.88889$ for the LDA classifier, and $C=1000$, $g=46.416$ and order of 3 for the SVM-poly classifier. Performing a similar analysis when 118 EEG channels were used for *aw*, the accuracy increased to 90.71% for the LDA classifier and 90.36% for the SVM-polynomial classifier (parameters: $\Delta = 0.01$ and $\Upsilon = 0.1111$ for the LDA classifier, and $C=1000$, $g=251.44$, order of three for the SVM-poly classifier), which was on a par with some best results in the literature. In this work an effort was made to pick the best classifier parameters and channel subsets based on the results obtained across the entire population of subjects, however this result illustrates the importance of fine tuning the classifier to the subject under study.

Conclusion

Accurate and robust MI EEG classification is important for future BCI neurorehabilitation technologies.

In this paper, we have investigated whether decision fusion through majority voting post-processing can improve the performance of a conventional MI EEG classification system. The effect of this method was studied in the context of six different MI EEG processing pipelines using CSP features and the following classifiers: SVM-linear, SVM-poly, SVM-RBF, LDA, NB and RF. The impact of MSMV decision fusion was studied across five different EEG channel subsets. Some useful conclusions are summarized as follows:

- MSMV decision fusion with the segmentation schemes (2s,0.5s) and (2s,0.25s) significantly improved the classification accuracy in five of the six classifiers.
- When using MSMV decision fusion, larger windows were correlated with better accuracy. The size of the window increment was not correlated with classification accuracy.
- An overall peak classification performance of 84.50% was obtained for the LDA and SVM-RBF classifiers for channel subset C+CP and windowing scheme (2s,0.1s). In this case the MSMV decision fusion approach led to an improvement of approximately 5% when compared to the control case.
- The C+CP+CF and C+CP channel subsets had the best performance across classifiers and led to higher accuracies when compared to using all 118 electrodes.
- The cross-classifier analysis carried out in this paper indicated that introducing central-parietal and/or central-frontal channels can improve the classification results when compared to using just the central EEG channels and these scalp groups would be recommended for consideration when constructing static EEG subsets.

Future work could focus on applying MSMV decision-fusion in to classification pipelines using other MI EEG features or applying MSMV to systems using novel channel selection techniques [33, 34]. Although the grid search was found to be an effective way of tuning a range of classifiers for MI EEG classification, Bayesian optimization could be further investigated to improve hyperparameter tuning for conventional classifiers due to its verified effectiveness. In addition, we will also work on exploring extraction of the most useful signals within the EEG data as well as applying advanced machine learning tools such as deep learning for more effective denoising, feature extraction and data classification.

Declarations

Funding

This work was supported in part by the Dazhi Scholarship of the Guangdong Polytechnic Normal University (GPNU), the Education Department of Guangdong Province (2019KSYS0091), the National Natural Science

Foundation of China (62072122), the Scientific and Technological Planning Projects of Guangdong Province (2021A0505030074) the Scientific and Technological Planning Projects of Guangdong Province (2021A0505030074) and the University of Strathclyde Scholarship.

Conflict of Interest

The authors have no conflicts of interest to declare in relation to this article.

References

- [1] Padfield N, Zabalza J, Zhao H, Masero V, and Ren J. EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. *Sensors (Switzerland)* 2019; 19(6): 1–34; doi: 10.3390/s19061423
- [2] Graimann B, Allison B, and Pfurtscheller G. Brain–Computer Interfaces: A Gentle Introduction. In: *Brain-Computer Interfaces*. Springer: Heidelberg, Germany; 2009; doi: 10.1007/978-3-642-02091-9_1
- [3] Z. Tang, C. Li, and S. Sun. Single-trial EEG classification of motor imagery using deep convolutional neural networks. *Optik (Stuttg)*. 2017; 130: 11–18.
- [4] Choi J and Kim H. Real-time Decoding of EEG Gait Intention for Controlling a Lower-limb Exoskeleton System. In: 7th International Winter Conference on Brain-Computer Interface (BCI), Feb. 2019. doi: 10.1109/IWW-BCI.2019.8737311
- [5] Armin J, Kahani D, Lakany H and Conway B A. Heterogeneous real-time multi-channel time-domain feature extraction using parallel sum reduction on GPU. In: 8th Graz Brain Comp. Interface Conf. 2019, Bridging Science and Application, Sept. 2019; doi: 10.3217/978-3-85125-682-6-33
- [6] Rodríguez-Bérmudez G and García-Laencina P J. Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces. *J. Med. Syst.* 2012; 36 :51-63; doi: 10.1007/s10916-012-9893-4
- [7] Asensio-Cubero J, Gan JQ, Palaniappan R. A study on temporal segmentation strategies for extracting common spatial patterns for brain computer interfacing. *UKCI 2011 - Proceedings of the 11th UK Workshop on Computational Intelligence*. 2011.
- [8] Asensio-Cubero J, Gan J Q and Palaniappan. Extracting optimal tempo-spatial features using local discriminant bases and common spatial patterns for brain computer interfacing. *Biomed. Signal Process. Control*. 2013; 8(6): 772–778; doi: 10.1016/j.bspc.2013.07.004
- [9] Samuel O W, Asogbon M G, Geng Y, Pirbhulal S, Mzurikwao D, Chen S, Fang P et al. Determining the optimal window parameters for accurate and reliable decoding of multiple classes of upper limb motor imagery tasks. In: 2018 IEEE Int. Conf. Cyborg Bionic Syst. CBS, Oct. 2018; doi: 10.1109/CBS.2018.8612159
- [10] Kevric J and Subasi A. Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomed. Signal Process. Control*. 2017; 31: 398–406; doi: 10.1016/j.bspc.2016.09.007
- [11] Ko L W, Lu Y and Chang Y. Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface. *IEEE Comput. Intell. Mag.* 2019; 14(1): 96–106; doi: 10.1109/MCI.2018.2881647
- [12] Oikonomou V P, Georgiadis K, Liaros G, Nikolopoulos S and Kompatsiaris I. A comparison study on EEG signal processing techniques using motor imagery EEG data. In: *IEEE Symp. Comput. Med. Syst.*, Jun. 2017; doi: 10.1109/CBMS.2017.113
- [13] Wahid MF, Tafreshi R, Langari R. A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2019 Dec 23;28(2):427-36; doi: 10.1109/TNSRE.2019.2961706

- [14] Baig M Z, Aslam N, Shum H P H, Zhang L. Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Syst. Appl.* 2017; 90: 184–195; doi: 10.1016/j.eswa.2017.07.033
- [15] Rodríguez-Bérmudez G and García-Laencina P J. Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces. *J. Med. Syst.* 2012; 36 :51-63; doi: 10.1007/s10916-012-9893-4
- [16] Kumar S, Sharma A, and Tsunoda T. An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information. *BMC Bioinformatics.* 2017; 18; doi: 10.1186/s12859-017-1964-6
- [17] He Li, Hu Y, Li Y, and Li D. Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *Neurocomputing* 2013; 121: 423–433; doi: 10.1016/j.neucom.2013.05.005
- [18] Shenoy H V and Vinod A P. An iterative optimization technique for robust channel selection in motor imagery based brain computer interface. In: 2014 IEEE Int. Conf. Syst. Man Cybern., Jan. 2014; doi: 10.1109/SMC.2014.6974191
- [19] Li M, Ma J, and Jia S. Optimal combination of channels selection based on common spatial pattern algorithm. In: *IEEE Int. Conf. Mechatronics Autom. ICMA 2011*, Aug. 2011; doi: 10.1109/ICMA.2011.5985673
- [20] Yang Y, Bloch I, Chevallier S, Wiart J. Subject-specific channel selection using time information for motor imagery brain–computer interfaces. *Cogn. Comput.* 2016 Jun 1;8(3):505-18; doi: 10.1007/s12559-015-9379-z
- [21] Olias J, Martin-Clemente R and Sarmiento-Vega MA. EEG signal processing in MI-BCI applications with improved covariance matrix estimators. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2019; 27(5): 895–904; doi: 10.1109/TNSRE.2019.2905894
- [22] Yang B, Hu C, Wang J, Li B, Wang W. Research on motor imagery EEG modeling based on window optimization and a few channels PSD. In: *Int. Conf. Digit. Signal Process.*, Nov 2018. doi: 10.1109/ICDSP.2018.8631698
- [23] Siuly S, Li Y, and Zhang Y. Comparative study: motor area EEG and all-channels EEG. In: *EEG Signal Analysis and Classification*. Cham, Switzerland: Springer; 2016. pp. 211–225; doi: 10.1007/978-3-319-47653-7_11
- [24] Armstrong S, Sale MV, Cunnington R. Neural oscillations and the initiation of voluntary movement. *Front. Psychol.* 2018 Dec 18; 9:2509.
- [25] Schutter DJ, Putman P, Hermans E, van Honk J. Parietal electroencephalogram beta asymmetry and selective attention to angry facial expressions in healthy human subjects. *Neurosci. Lett.* 2001 Nov 13;314(1-2):13-6.
- [26] Dornhege G, Blankertz B, Curio G, and Müller K R. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans. Biomed. Eng.* 2004; 51(6): 993-1002; doi: 10.1109/TBME.2004.827088
- [27] O'Reilly C and Nielsen T. Automatic sleep spindle detection: Benchmarking with fine temporal resolution using open science tools. *Front. Hum. Neurosci.* 2015; 9(June): 1–19; doi: 10.3389/fnhum.2015.00353
- [28] She Q, Chen K, Ma Y, Nguyen T, and Zhang Y. Sparse representation-based extreme learning machine for motor imagery EEG classification. *Comput. Intell. Neurosci.* 2018; Article ID: 9593682; doi: 10.1155/2018/9593682
- [29] Kumar S, Sharma A, Mamun K and Tsunoda T. (2016, Dec). A deep learning approach for motor imagery EEG signal classification. In: 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Dec. 2016; doi: 10.1109/APWC-on-CSE.2016.017
- [30] Cranstoun S D, Ombao H C, von Sachs R, Guo W and Litt B. Time-frequency spectral estimation of multichannel EEG using the auto-SLEX method. *IEEE Trans Biomed Eng.* 2002; 49(9): 988-996; doi: 10.1109/TBME.2002.802015
- [31] Qin D, Houchins B, Parten M E, Parker N, Homan R and Petrosian A. A comparison of techniques for the prediction of epileptic seizures. In: 8th IEEE Symposium on Computer-Based Medical Systems, Jun. 1995; doi: 10.1109/CBMS.1995.465436

- [32] Wojcikiewicz W, Vidaurre C and Kawanabe M. Stationary common spatial patterns for non-stationary EEG data. *Front. Comput. Neurosci. Conference Abstract: Bernstein Conf. on Computational Neuroscience*, Sept. 2010; doi: 10.3389/conf.fncom.2010.51.00051
- [33] Wang Q, Lin J, Yuan Y. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Trans Neural Netw Learn Syst.* 2016 Mar 18;27(6):1279-89; doi: 10.1109/TNNLS.2015.2477537
- [34] Lyu D, Liu B, Geist M, Dong W, Biaz S, Wang Q. Stable and efficient policy evaluation. *IEEE Trans Neural Netw Learn Syst.* 2018 Oct 30;30(6):1831-40; doi: 10.1109/TNNLS.2018.2871361

Tables

Table 1: A table of the different window sizes and increment sizes used for EEG trial segmentation. Each window size and window increment pairing were investigated.

Window Size	Window Increment
2s, 1.75s, 1.5s, 1.25s, 1s, 0.75s, 0.5s	0.5, 0.25, 0.1

Table 2: Comparing the peak accuracy results obtained using MSMV decision fusion to the control case when no decision fusion is used. The results also include the configuration that resulted in peak classification accuracy, and a comparison of the sensitivities to classes 1 and 2 obtained for the configuration, and the for the control case.

Classifier	Accuracy		Sensitivity (Class1/Class2)		Configuration Channel subset – (Win Size, Win Inc Size)
	Multi-Segment Fusion	Control	Multi-Segment Fusion	Control	
LDA	84.50%	80.93%	83.57%/84.29%	80.57%/81.57%	C+CP- (2s,0.1s)
NB	82.64%	80.00%	86.29%/79.00%	80.01%/79.86%	C+CP+CF – (2s,0.25s)
RF	83.86%	80.93%	84.02%/85.14%	81.08%/81.20%	118 – (1.5s,0.25s)
SVM-Linear	83.93%	81.07%	85.00%/84.00%	82.57%/79.29%	C+CP+CF - (1.75,0.1s)
SVM-Poly	82.36%	80.43%	82.00%/82.71%	79.71%/81.14%	C+CF – (2s, 0.1s)
SVM-RBF	84.50%	80.46%	84.71%/84.29%	82.14%/78.57%	C+CP – (2s, 0.1s)
Average	83.63%	80.67%	84.27%/83.24%	81.01%/80.27%	-

Table 3: Results from a Pearson correlation analysis for each classifier and window size. The correlation coefficients, ρ , and corresponding p-values obtained for each window increment size are recorded. Statistically significant values of ρ are in italics.

Classifier	0.5s increment		0.25s increment		0.10s increment	
	p-value	ρ	p-value	ρ	p-value	ρ
LDA	0.0010	<i>0.9508</i>	0.0035	<i>0.9183</i>	0.0098	<i>0.8756</i>
NB	0.0261	<i>0.8133</i>	0.5323	0.2872	0.3229	0.4402
RF	0.1152	0.6484	0.1087	0.6572	0.0236	<i>0.8208</i>
SVM-Linear	0.0343	<i>0.7906</i>	0.0009	<i>0.9523</i>	0.0066	<i>0.8940</i>
SVM-Poly	0.0370	<i>0.7838</i>	0.0242	<i>0.8192</i>	0.034	<i>0.7912</i>
SVM-RBF	0.0526	0.7491	0.2415	0.5107	0.0527	0.7490

Table 4: Results from a Pearson correlation analysis for each classifier and window increment size. The correlation coefficients, ρ , and corresponding p-values for each window size are recorded. Statistically significant values of ρ are in italics.

Window Size	2s		1.75s		1.5s		1.25s		1.00s		0.75s		0.5s	
	p-val	ρ	p-val	ρ	p-val	ρ	p-val	ρ	p-val	ρ	p-val	ρ	p-val	ρ
LDA	0.24	-0.93	0.81	-0.29	0.45	-0.76	0.38	-0.83	0.36	-0.84	0.28	-0.91	0.25	-0.92
NB	0.30	-0.89	0.68	-0.48	0.24	-0.93	0.08	-0.99	0.06	-0.99	0.01	-0.99	0.08	-0.99

RF	0.60	0.58	0.64	0.53	0.62	0.56	0.65	0.52	0.70	0.45	0.60	0.59	0.58	0.61
SVM-Linear	0.76	0.37	0.49	0.72	0.40	0.81	0.40	0.81	0.37	0.84	0.34	0.86	0.35	0.86
SVM-Poly	0.91	-0.14	0.60	0.60	0.65	0.52	0.77	0.37	0.73	0.41	0.82	0.28	0.86	0.22
SVM-RBF	0.42	0.79	0.37	0.83	0.42	0.79	0.37	0.83	0.35	0.85	0.30	0.89	0.39	0.82

Table 5: Comparing the classification accuracy obtained with MSMV decision fusion to results obtained with other conventional approaches ([14, 17, 20, 21]) and deep learning-based approaches ([28, 29]). Results highlighted in italics were obtained for classical conventional systems.

Papers	Features & Classifiers used	Channels	Classification Accuracy (%)					Average
			aa	al	av	aw	ay	
Olias et al. [21]	CSP+LDA (classical)	118	67.68	96.81	62.12	85.12	87.68	79.88
	Normalised CSP+tangent space logistic regression		70.31	96.31	67.87	87.75	91.62	82.77
Baig et al. [14]	CSP+SVM (classical)	118	82.00	94.00	70.00	87.00	87.00	84.00
	CSP with differential evolution feature selection +SVM		95.80	98.80	89.80	99.20	96.50	96.02
He et al. [17]	CSP with Rayleigh coefficient+LDA	118	67.90	88.30	59.20	87.60	80.40	76.68
	CSP with Rayleigh coefficient+LDA with GA optimisation to select channel subset	Subject-specific (13-18)*	86.40	98.50	75.10	93.90	87.10	88.20
Yang et al. [20]	CSP features with LDA classifier and time series analysis-based channel selection	Subject-specific (9 channels)	67.00	89.00	61.00	81.00	92.00	78.00
She et al. [28]	Hierarchical extreme learning machine with deep architecture	118	61.70	100	73.88	88.17	79.64	79.33
Kumar et al. [29]	CSP+deep neural network	118	90.00	97.50	73.00	97.00	96.00	90.70
This paper	CSP+SVM-RBF + MSMV (Scheme: 2.0s,0.1s)	C+CP (14)	79.64	94.64	75.00	78.57	94.64	84.51
	CSP+LDA + MSMV (Scheme: 2.0s,0.1s)	C+CP (14)	79.64	95.71	73.21	78.93	95.00	84.51

* aa- 39 channels, al- 18 channels, av- 47 channels, aw – 26 channels, ay – 16 channels

Figures

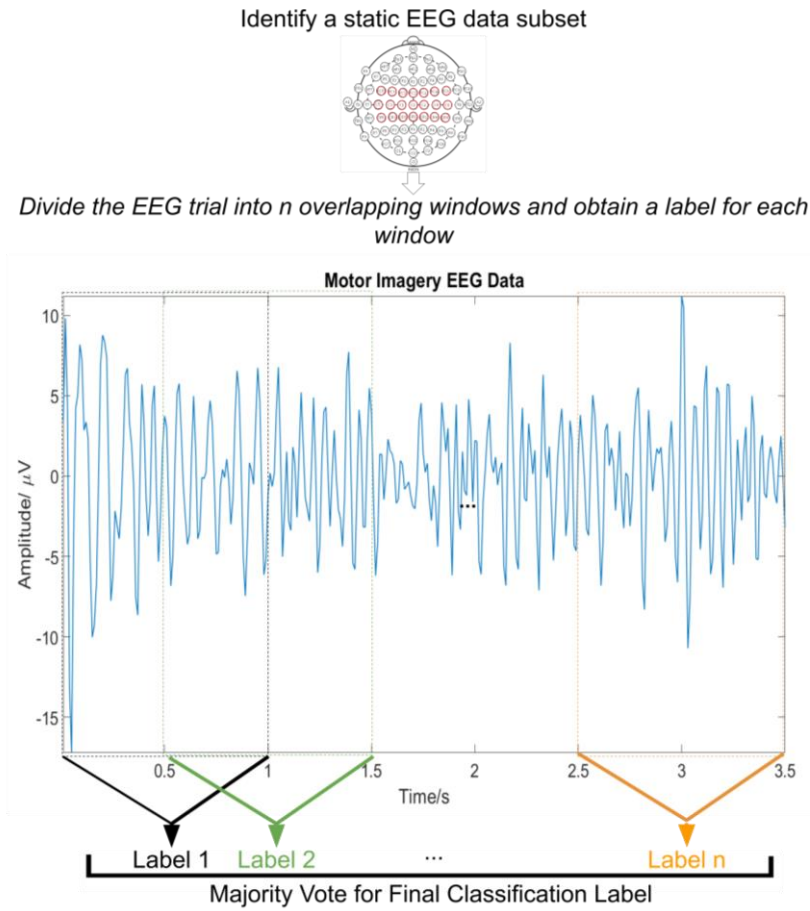


Figure 1: An illustration of the proposed methodology for multi-segment majority voting decision fusion. The image shows the windowing and majority voting scheme for a window of size 1s and with an increment of 0.5s.

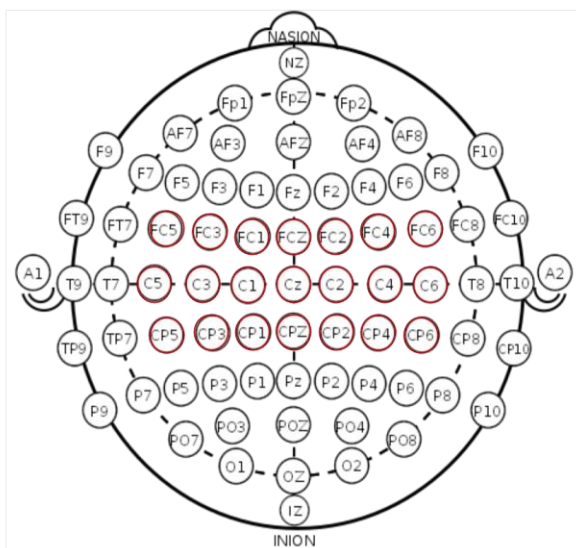


Figure 2: The standard 10-20 electrode placement for EEG recording, with the 21 electrodes used to construct subsets highlighted in red.

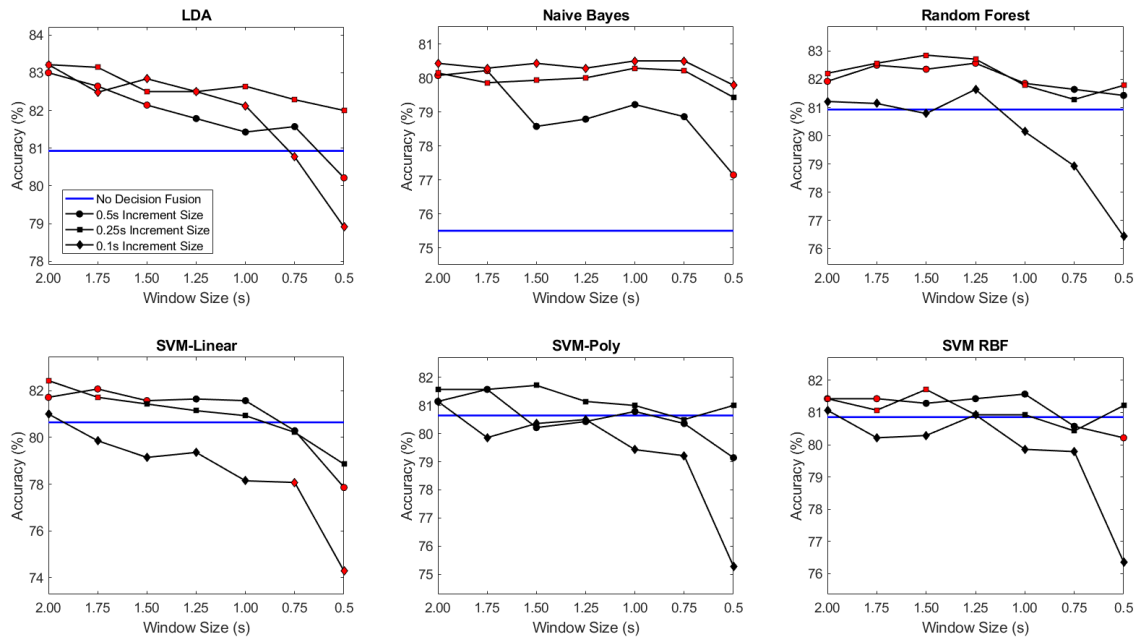


Figure 3: Plots of accuracy against window size for each classifier. The blue horizontal line denotes the control case when no decision fusion was used. The data marker type represents the window increment size. A Wilcoxon signed-rank test was used to compare the MSMV decision fusion results to the control case. Red data points indicate a statistically significant change in performance and black data points indicate no statistically significant change.

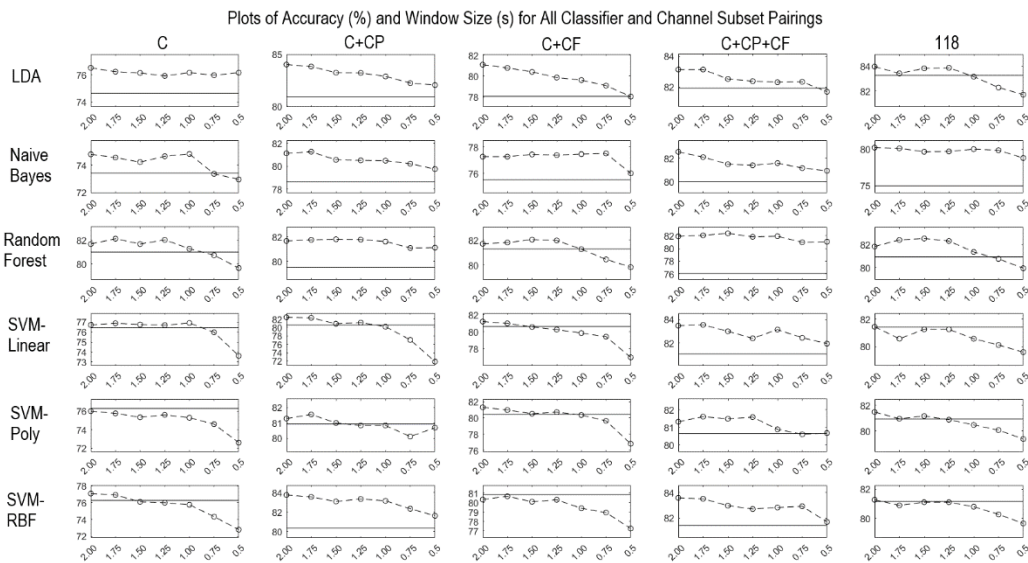


Figure 4: Plots of mean classification accuracy against window size for all the classifier and channel subset pairings. Results were averaged across subjects and then the mean classification accuracy obtained across the different window increments were plotted. The horizontal lines are the mean classification accuracy obtained with that classifier and channel subset when MSMV decision fusion was not used, which was the control case.

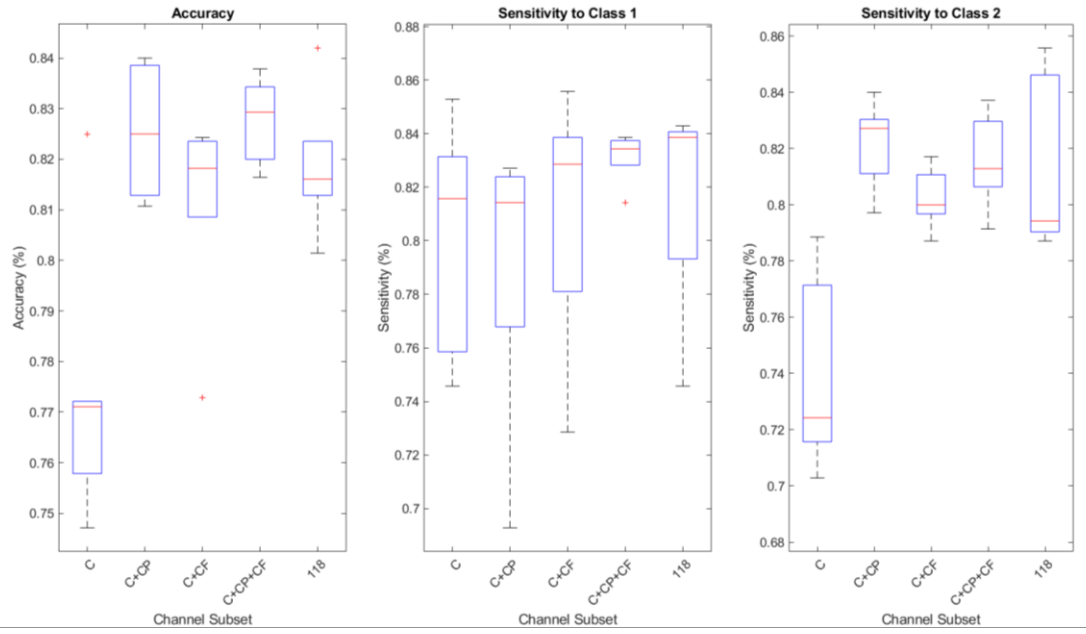


Figure 5: Box plots comparing the accuracy, sensitivity to class 1 and sensitivity to class 2 for different channel subsets. In each box, the horizontal red line represents the median value, whilst the top and bottom edges of the box represent the 75th and 25th percentiles of the data.