Implementation of a brain computer interface to classify between two motor activities

1 Introduction

This report focuses on EEG data from the EEGMMI DS database [2]. This dataset captures brain activity during different hand-related tasks. To analyze these signals, the study applied the following processing steps.

Signal Preprocessing: The initial step involved loading the EEG data, segmenting it based on hand-related tasks, and applying spatial filtering like Common Spatial Patterns (CSP)

Frequency Isolation: Band-pass filtering targeted specific frequency ranges to isolate neural activity associated with the hand tasks, enhancing relevant signal components.

Feature Extraction: Extracting meaningful information involved computing statistical measures like variance and performing logarithmic transformations in the component space. Also the Auto-Regressive (AR) coefficients were added from the signal to create enhanced feature vectors for classification.

This report analyzes EEG signals from subject S001 during various hand-related tasks, applying signal processing and feature extraction techniques to enable accurate classification of these activities [1].

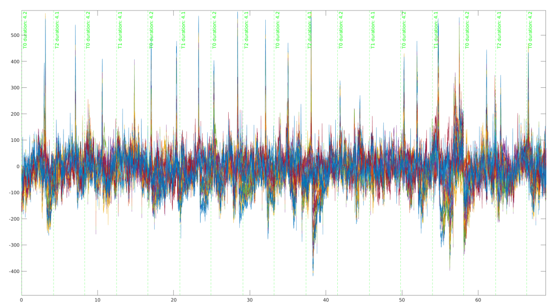


Figure 1: Signal with annotated features

2 Methods

2.1 Preporcessing

The EEG dataset chosen captured brain activity during hand activities, segmented into different events. First, the data was proccessed: loading EEG signals and segmenting them to isolate hand-related tasks [2]. The Common Spatial Patterns (CSP) technique transformed the signals into the spatial component space to enhance discriminative features.

The CSP technique serves as a crucial translator for EEG signals, transforming them into a space where crucial features become clearer and more discernible. This transformation enhances our understanding of EEG signals, simplifying the identification of important elements related to hand activities.

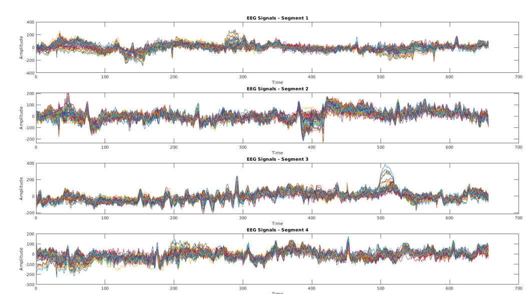


Figure 2: Segments representing distinct hand-related activities in the EEG signal

2.2 Band-Pass filtering

After spatial transformation, the EEG signals were filtered using a band-pass FIR filter, targeting specific frequency ranges associated with motor-related brain activity. This filtering procedure aimed to isolate neural activity relevant to hand tasks.

The bandpass FIR filter can be represented as:

$$H(z) = \sum_{n=0}^{N-1} b(n) z^{-n}$$

These parameters were utilized to generate the filter coefficients $b(n)$ using the 'firls' function.

$$f = [0 \quad 8 \quad 8 \quad 13 \quad 13 \quad \mathrm{fs}/2]/(\mathrm{fs}/2)$$

$$a = [0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0]$$

$$n = 35$$

2.3 Feature Extraction

Feature extraction involved computing statistical measures like variance and logarithmic transformations within the component space. Furthermore, the AR coefficients from the power spectra were included.

The AR coefficients were extracted using a method that considers the immediate and second previous values of the EEG signal. These coefficients, obtained using the arburg function, give insight of how the signal relates to its recent past.

For the obtained results, the code was tested and empirically set $p = 2$ to represent the order of the AR model, implying that the AR coefficients were estimated considering the immediate and second previous values of the signal.

The obtained AR coefficients are then included in the feature vectors alongside other statistical features (such as variance and logarithm of variance).

Variance is a measure that quantifies the spread or dispersion of a set of values around their mean. For a sequence of N values $x_1, x_2, \ldots, x_N$, the variance (σ^2) is calculated as:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

where μ is the mean of the sequence.

The variance of the EEG segments is calculated using the 'var' function.

Moreover, applying the logarithm to the variance ($\log(\sigma^2)$) can significantly enhance the discriminative power of features in classification tasks [3]. This transformation offers improved separability among different classes.

The AR method boosted the EEG signal analysis significantly. By adding AR coefficients to the feature vectors, we saw a huge jump of about 30% in both Classification Accuracy (CA) and Area Under the Receiver Operating Characteristic Curve (AUR) metrics.

2.4 Classifcation

LDA [1] was the main classifier used to tell apart different hand-related tasks based on enhanced EEG signal features, including AR coefficients.

In EEG signal classification, LDA distinguished between different brain activities by learning features from training data and then applying that knowledge to classify new data. It's especially good when dealing with classes that can be separated by straight lines or planes.

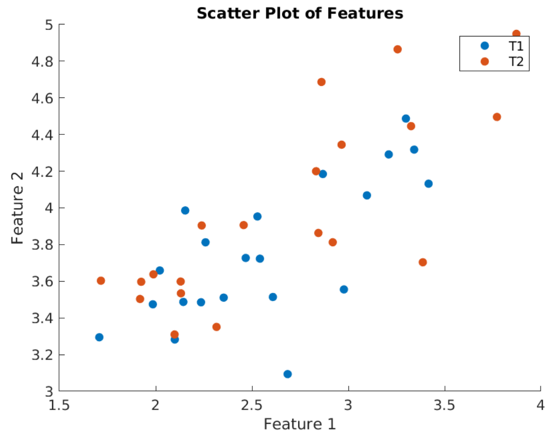


Figure 3: Scatter plot of $\log(\sigma^2)$ transformed features

3 Results

The table [1] provides AUC (Area Under the Curve) and CA (Classification Accuracy) metrics for different datasets without utilizing AR coefficients. It appears that the AUC and CA values are consistently lower across various datasets when AR coefficients are not employed.

The table [2] presents AUC and CA metrics for datasets where AR coefficients were used. Notably, there is a considerable increase in AUC and CA values when AR coefficients are included, especially in the "All records" dataset.

Dataset	AUC	CA
All records - Test	56.06	50.00
All records - Learn	56.06	50.00
20% holdout - Test	48.61	47.06
20% holdout - Learn	57.31	47.06
20% holdout - Test	47.94	47.06
20% holdout - Learn	56.80	47.06
Cross-valid - Test	46.75	45.35
Cross-valid - Learn	56.13	45.35
Cross-valid - Test	46.30	45.35
Cross-valid - Learn	56.24	45.35
Leave one o - Test	40.26	41.86
Leave one o - Learn	55.83	41.86

Table 1: Summary of AUC and CA witout using AR coefficients

4 Discussion

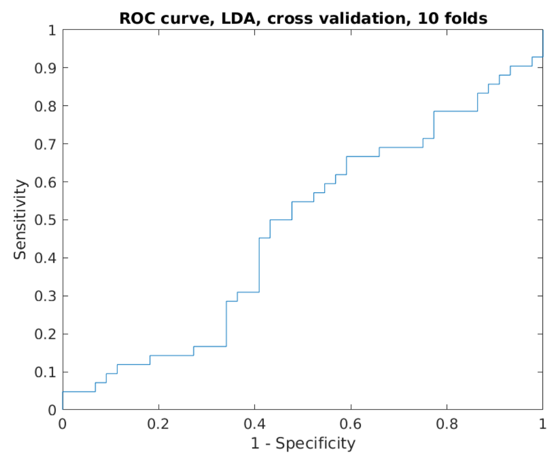
The ROC curves and associated performance metrics, particularly when employing AR coefficients in classifiers, show substantial differences compared to those without AR coefficients [4b] [1]. AR coefficients, likely indicating more features or a different representation, result in notably improved performance with higher AUC values, indicating enhanced discrimination between classes [4a].

These higher AUC values suggest better classifier performance in distinguishing between classes [2]. Integrating AR coefficients significantly boosts classifiers' accuracy in classifying instances. Overall, incorporating AR coefficients ap-

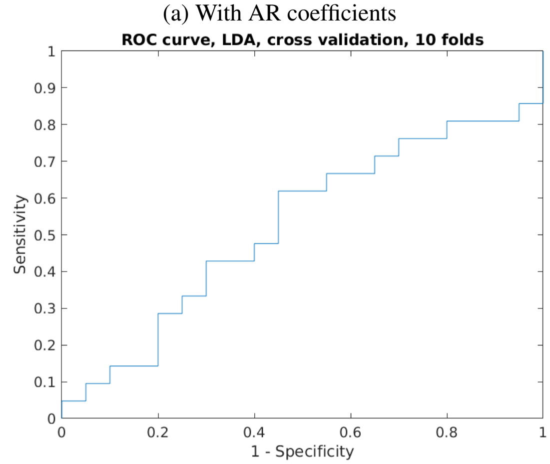
Dataset	AUC	CA
All records - Test	97.40	96.51
All records - Learn	97.40	96.51
20% holdout - Test	50.00	47.06
20% holdout - Learn	97.14	47.06
20% holdout - Test	48.56	50.00
20% holdout - Learn	96.48	50.00
Cross-valid - Test	45.51	43.02
Cross-valid - Learn	96.72	43.02
Cross-valid - Test	47.65	48.28
Cross-valid - Learn	96.78	48.28
Leave one o - Test	46.48	47.67
Leave one o - Learn	97.29	47.67

Table 2: Summary of AUC and CA using AR coefficients

pears to elevate the classifiers' ability to accurately differentiate and classify, possibly by capturing patterns for better predictions.



(a) With AR coefficients



(b) Without AR coefficients

Figure 4: ROC curves with and without AR coefficients.

References

[1] Alan Julian Izenman. "Linear Discriminant Analysis". In: *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York, NY: Springer New York, 2008, pp. 237–280. ISBN: 978-0-387-78189-1. DOI: 10.1007/978-0-387-78189-1_8. URL: https://doi.org/10.1007/978-0-387-78189-1_8.

[2] Gerwin Schalk et al. "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System". In: *IEEE Transactions on Biomedical Engineering* 51.6 (2004), pp. 1034–1043. DOI: 10.1109/TBME.2004.827072.