# SUPERVISED PROJECTIVE LEARNING FOR ELECTROENCEPHALOGRAPHY ANALYSIS

by

Tyler J Grear

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Applied Physics

Charlotte

2021

Approved by:

_____

Dr. Donald Jacobs

_____

Dr. Nathaniel Fried

_____

Dr. Tino Hofmann

# ABSTRACT

TYLER J GREAR. Supervised Projective Learning for Electroencephalography Analysis.
(Under the direction of DR. DONALD JACOBS)

The brain-computer interface has emerged as a promising field of research with the capacity to significantly transform multiple technological sectors. The imagining of limb movement known as motor imagery (MI) serves as an auspicious paradigm for the interface control of such a device. Masked by volume conduction, electroencephalography (EEG) MI signals exhibit a low signal-to-noise ratio. Additionally, the underlying spatio-temporal pattern is manifested spectrally by the suppression and amplification of the respective $\mu$ and $\beta$ frequency bands during sensorimotor cortical activity. These challenges necessitate further advancements in pattern recognition and feature extraction by way of machine learning.

A novel bottom-up approach for the artifact removal and single-channel classification of EEG MI signals known as supervised projective learning for EEG analysis (SPLEEGA) is introduced in this thesis. The underlying engine of SPLEEGA is supervised projective learning with orthogonal completeness (SPLOC). For each electrode, a SPLEEGA eigenchannel provides a characteristic vector space that discriminates MI from resting state signals in both the $\mu$ and $\beta$ frequency bands while simultaneously performing temporal alignment as an emergent property of the developed orthonormal basis.

The 52-subject GigaDB MI dataset from Gwangju Institute of Technology was utilized in this work. At a sparse-sensor configuration, complete separation of MI signals from resting state signals is achieved for 73% of subjects. With a contralateral montage, this is increased to 100% of subjects revealing at least two discriminatory eigenchannels. Utilizing only 7% of data for training during an automatic channel selection procedure, SPLEEGA attains comparable MI classification accuracy to the state-of-the-art using only a single electrode and frequency band. Furthermore, the utility for real-time applications is encouraging due to a rapid classification which takes < 100 ms after an initial calibration is executed.

# ACKNOWLEDGEMENTS

My sincere thanks are extended to Professor Donald Jacobs for his ubiquitous guidance throughout my academic career. This gratitude is also expressed to my thesis committee members Professor Nathaniel Fried and Professor Tino Hofmann. Your insights have been impactful to the quality of this work.

I also wish to show my appreciation to all members of the BioMolecular Physics Group (BMPG). Notably: Mr. Chris Avery, Mr. John Patterson, and Dr. Jenny Farmer. They are each integral components of the BMPG, and make it the think tank that it is.

This work would not have been possible without funding provided by the University of North Carolina at Charlotte (UNCC) Graduate Assistant Support Program (GASP). Lastly, I thank the UNCC University Research Computing (URC) team that facilitated the use of high-performance computing clusters during algorithmic development.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER 1: INTRODUCTION 1

1.1. Spatio-Temporal Challenges 2

1.1.1. Spatial Resolution 2

1.1.2. Temporal Variability 3

1.2. Motor Imagery Classification for Brain-Computer Interfaces 4

1.3. Supervised Projective Learning for Electroencephalography Analysis 7

CHAPTER 2: DATASET AND PREPROCESSING 9

2.1. Giga Science Motor Imagery Dataset 9

2.2. Channel Referencing and Trial Partitioning 10

2.3. Infinite Impulse Response Filter 12

2.4. State Space Representation 14

CHAPTER 3: SPLOC AND THE MODE-FEATURE SPACE PLANE 15

CHAPTER 4: ARTIFACT REMOVAL USING PROJECTION PURSUIT 19

4.1. SPLEEGA Eigenchannel Construction 19

CHAPTER 5: CLASSIFICATION OF EEG SIGNALS 21

5.1. Support Vector Machines 21

CHAPTER 6: AUTOMATIC SINGLE-CHANNEL CLASSIFICATION 24

CHAPTER 7: RESULTS 26

7.1. Artifact Removal Results 26

7.2. Automatic Single-Channel Classification 33

7.3. A Prelude to the Automatic Eigenbrain 38

CHAPTER 8: CONCLUSIONS 40

REFERENCES 41

LIST OF TABLES

TABLE 7.1: Eigenchannel artifact removal results for 52 subjects from GigaDB MI dataset. 29

TABLE 7.2: Automatic contralateral run. Subject; Single-channel classification accuracy; Channel number; Channel name; Optimal frequency band; Artifact removal efficacy. 34

LIST OF FIGURES

FIGURE 1.1: General BCI Paradigm. 1

FIGURE 1.2: Analyzed frequency band vs. Study plot. 5

FIGURE 1.3: SPLEEGA pipeline. 8

FIGURE 2.1: The experimental protocol for MI signal acquisition. 9

FIGURE 2.2: Electrode configurations. 10

FIGURE 2.3: Reference electrode configuration. 11

FIGURE 2.4: Magnitude response plot for $\mu$ IIR filter. 13

FIGURE 2.5: IIR filtered signals. 13

FIGURE 2.6: State space representation. 14

FIGURE 3.1: A diagram of the SPLOC mode-generating algorithm. 17

FIGURE 4.1: SPLEEGA eigenchannel C4 $\mu$ basis spectrum. 19

FIGURE 4.2: SPLEEGA eigenchannel C3. 20

FIGURE 5.1: The foundation for the optimization problem of SVMs. 21

FIGURE 5.2: Application of an SVM to non-linearly separable classes. 23

FIGURE 6.1: Single-channel classification workflow. 25

FIGURE 7.1: The C3 eigenchannel for subject 50. 27

FIGURE 7.2: The C3 eigenchannel for subject 11. 28

FIGURE 7.3: The SPLEEGA brainmap for subject 5. 32

FIGURE 7.4: The SPLEEGA brainmap for subject 11. 33

FIGURE 7.5: The C3-$\mu$ eigenchannel for subject 43. 37

FIGURE 7.6: The SPLEEGA brainmap of s26 for all 64 electrodes. 38

FIGURE 7.7: The SPLEEGA brainmap of select electrodes for subject 26. 39

# CHAPTER 1: INTRODUCTION

The brain-computer interface (BCI) provides a promising avenue for many applications such as neuroprosthetics [1, 2, 3], computer-assisted eye-gaze targeting [4, 5] and controlling exoskeleton apparatus [6, 7]. A systemic workflow for a generalized BCI is shown in Figure 1.1. Effectively utilizing a BCI remains largely untapped due to the open challenge of deciphering electroencephalography (EEG) signals. A novel methodology for the artifact removal and analysis of EEG signals, introduced as supervised projective learning for electroencephalography analysis (SPLEEGA) is presented in this work. Furthermore, SPLEEGA allows for a more-informed manner of automatic channel selection with respect to discrimination from resting state and optimal frequency bands on a per-subject basis.

The challenges facing BCI systems are presented in this chapter, followed by an evaluation of previous work; thereafter, the SPLEEGA algorithm is proposed. Chapters 2 through 5 follow the progression of the SPLEEGA algorithm and motivate the choices for development at each juncture. Chapter 6 focuses on an application of the SPLEEGA algorithm in terms of automatic channel selection and single-channel classification. Chapter 7 will cover results from algorithmic development then conclusions are elucidated in Chapter 8.

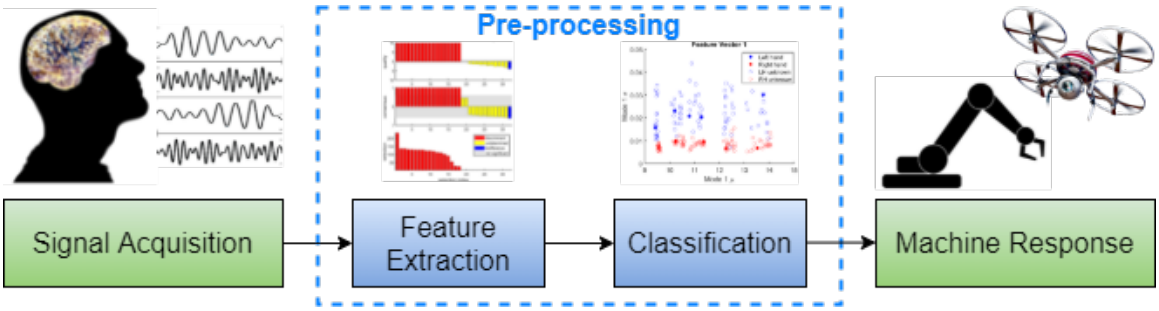


Figure 1.1: General BCI Paradigm. The process begins with experimental acquisition of EEG signals. Next, useful features are constructed for the purpose of classification. Finally, prediction algorithms are developed to identify desired interface control signals.

## 1.1 Spatio-Temporal Challenges

Motor imagery classification serves as a proving ground for the identification of spatio-temporal patterns. These types of patterns are of recent interest in many fields spanning molecular engineering [8, 9] to atmospheric physics [10, 11]. It is not enough to consider just spatial snapshots of a process at a given time, nor just time-series profiles at a given spatial location - the behavior at spatial locations at one time point will affect the behavior at nearby spatial locations at the next time point. Only by considering time and space together can we address how spatially coherent entities change over time or, in some cases, why they change [12]. In this section, the spatio-temporal challenges associated with motor imagery classification are presented in terms of i) spatial resolution; and ii) temporal variability.

### 1.1.1 Spatial Resolution

A favored conduit for communication between a brain and computer is the acquisition of EEG signals through non-invasive techniques. This is largely due to low cost, high temporal resolution sensors, and a minimized risk for the subject [13, 14]. The EEG signals act as the interface control which are procured through experimental trials where subjects imagine limb movement known as motor imagery (MI). The problem of low spatial resolution arises due to the proximity of the electrodes from the signal source occurring somewhere within the brain. Consequently, the cortical current propagates through different resistive mediums prior to measurement. As a result, the recorded potentials at each sensor are a statistical mixture of the underlying brain sources [15].

These factors motivate the use of blind source separation (BSS) which exploits primarily spatial diversity, that is, the fact that different sensors receive different mixtures of the sources. BSS approaches look for structure across the sensors, not across time [16]. A simplified BSS model assumes there are $n$ unknown independent signals $s_1(t), ..., s_n(t)$ that

produce $n$ observed mixtures $x_1(t), ..., x_n(t)$. This is generally expressed as:

$$\mathbf{x}(t) = A\mathbf{s}(t) \tag{1.1}$$

where $A$ represents the $n$ $x$ $n$ mixing matrix that transforms $\mathbf{s}(t)$ to $\mathbf{x}(t)$. Alternatively, Equation 1.1 can be reformulated to find $\mathbf{y}$(t), the source signal approximations:

$$\mathbf{y}(t) = B\mathbf{x}(t) \tag{1.2}$$

The objective of this work can be condensed to constructing the set of basis vectors that comprise $B$ in Equation 1.2 such that the desired motor imagery signals are separated from the resting state at the eigenchannel level. This information can subsequently be used to construct the orthonormal basis $U$ that optimally separates left-hand (LH) from right-hand (RH) motor imagery tasks, known as the SPLEEGA eigenbrain.

### 1.1.2 Temporal Variability

The first temporal issue that presents itself during MI tasks is based on human reaction time (RT). The response to a stimulus can vary from person to person [17]; additionally, RT can also depend on the type of stimuli [18]. The next temporal problem is associated with all signal acquisition techniques by way of electroencephalography. The volume conduction effects that result in spatial distortion also contribute to a temporal smearing across all sensors [19]. Subsequently, spontaneous EEG signals recorded by different electrodes tend to appear more phase locked than they actually are, inducing artificially high between-site coherence [20]. Electroencephalographic coherence and covariance are often used as large-scale measures of the functional interrelations among pairs of neocortical regions. Normalizing this covariance by the product of individual variances results in the correlation function coefficient [20, 21]. This coefficient is a function of the delay time between signals;

furthermore, an evoked potential waveform at one spatial location may be correlated with another spatial location *only after some temporal delay* [22]. These first two considerations (RT and correlative time delay) necessitate an effective methodology for the temporal alignment of motor imagery tasks which is directly addressed in this work.

The nonstationarity inherent in EEG signals is also considered a temporal challenge when conducting analysis. Nonstationarity expresses that the statistical moments such as mean and variance of the signal change over time. This presents a problem in the choice of method to construct the power spectral density. An underlying assumption of the Fourier transform is that a source signal is infinitely periodic [23]. This assumption has utility, a good example being quantum mechanics where spectral decomposition of the Hamiltonian operator is performed. Here, the source wave being decomposed is a probability wave, and thus the stationary states are independent of time. For real-world signals, window-based Fourier approaches have been implemented to address the nonstationarity of EEG signals with success [24, 25]. However, there is a trade-off between the periodicity of a signal and loss of information from approximating its frequency components.

## 1.2 Motor Imagery Classification for Brain-Computer Interfaces

Previous work has shown [26] that an underlying pattern emerges within the power spectral density (PSD) of EEG signals due to expected neuronal population activity during MI tasks [27]. Furthermore, the $\mu$ (7.0-13.0 Hz) and $\beta$ (13.0-30.0 Hz) frequency bands exhibit a structured trend which allows various pattern recognition methodologies to apply. This trend is observed as event-related synchronization (ERS) followed by an event-related desynchronization (ERD) of brain activity over the sensorimotor cortex [28]. An ERS corresponds to an amplification of the evoked $\beta$ waves localized to one region of the brain while simultaneously an attenuation of the $\mu$ frequency band takes place in another region. Thereafter, ERD occurs which returns the system to its high statistically-mixed natural state until an MI task is performed again. Given the known suppression/amplification across

the $\mu/\beta$ frequencies during MI [29], this serves as a pertinent starting point for identifying patterns under the BCI-EEG MI paradigm.

There is not a strict range of frequencies that are guaranteed to be effective for all humans, further adding to the challenge of constructing an efficient BCI. Akin to the biometric uniqueness of a fingerprint, certain brain activity related to MI can vary substantially from subject to subject. This is exemplified in a recent high-quality meta analysis [30] of notable ML methodologies for MI BCIs. The utilized frequency ranges for each study are shown in Figure 1.2. There is no clearly-defined standard for the selection of frequency bands. It has even become a recent trend for deep learning approaches to use a single large frequency range from ~0.1 Hz to 100 Hz, in the spirit of "let the neural network figure it out". This uncertainty in optimal frequency band selection is directly addressed in this thesis; furthermore, the SPLEEGA exploratory procedure pursues where the information required for discrimination exists.

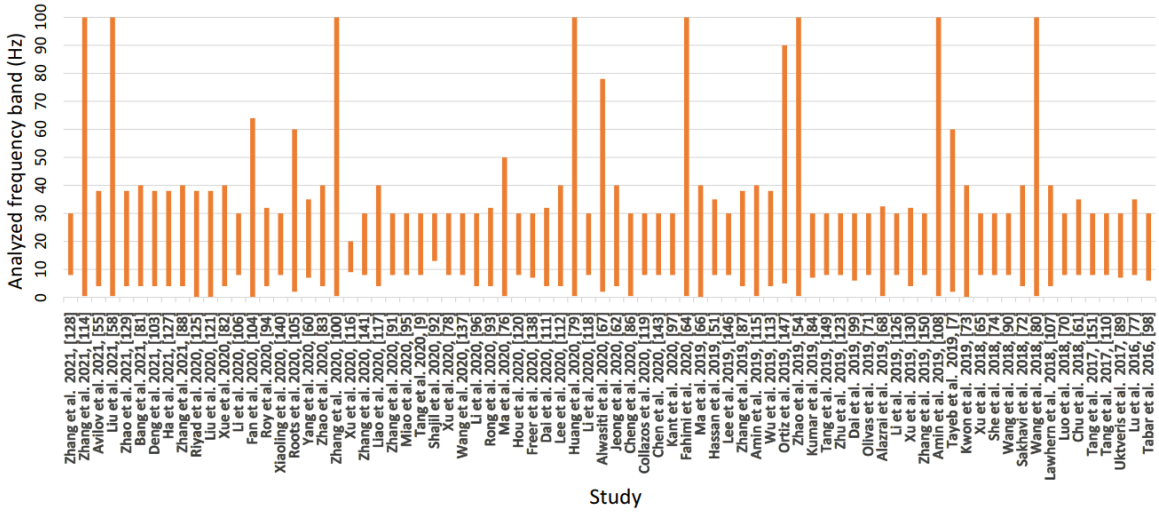


Figure 1.2: Analyzed frequency band vs. Study plot from Altaheri et al. (2021). The vertical orange bars represent the ranges of frequencies used in 78 studies from 2016-2021 indicated by the x-ticks. There are two general trends present here, first being the large single bands from ~0.1 Hz to 100 Hz that are presented to deep learning architectures. Second, the trend of band selection from 7-28 Hz which agrees with the selection for this work. There is a distinction between the high $\beta$ frequency (~18-40 Hz) and low $\beta$ (~14-22 Hz). In this work, the low $\beta$ band was favored over high $\beta$ because the higher end of this band relates directly to stress, anxiety, paranoia, high energy, and high arousal.

Due to recent hardware and algorithmic developments, machine learning (ML) has become a focal point in the field of BCI. This is a result of its pattern recognition and predictive capabilities for complex and nonlinear systems [31]. The ML approaches for BCI-EEG can be decomposed into two processes: i) feature extraction, and ii) classification on both known (training) and unknown (testing) systems. The features extracted, then ultimately selected, weigh heavily on how the problem is posed to the machine. The process of identifying a subset of features to be used for model training and prediction is considered dimension reduction (DR). Both feature extraction and selection represent critical steps for discriminating the information content in EEG signals.

In the context of projection pursuit (PP) [32, 33, 34, 35, 36, 37], customized features are extracted based on a particular objective function to yield a desired outcome. principal component analysis (PCA) is a PP optimization of second statistical moments, given as covariance [38]. Other examples include optimizing covariance with projections to latent structures (PLS) [39], and higher statistical moments such as kurtosis with independent component analysis (ICA) [40].

An effective feature extraction approach for EEG analysis is Common Spatial Pattern (CSP) that aims to acquire optimal components using eigendecomposition with the pooling of class covariance matrices [41, 42]. Frequency bandpass filters with either Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) are used to extract features within time-series data [43, 44]. Alternatively, wavelet-based features with their subsequent BCI-EEG methods [45, 46] provide an informed choice of features that are well-equipped to deal with the nonstationary nature of EEG signals [47]. Classification is executed once the signals are represented by a collection of features.

The first BCI-EEG ML paradigm utilized discriminative filter banks to leverage the variable range of brain activity within the $\mu$ and $\beta$ bands [48, 49, 50]. This approach reveals that the sensorimotor cortical activity is not spatially consistent on a per-subject basis [51], creating a moving target for the classification of ERS/ERD patterns. To overcome

this confounding factor some methods include a more comprehensive set of subbands [52, 53, 54]. Unfortunately, this added computational complexity [55] does not resolve the intrinsic problem of spatial-temporal variation in EEG signals.

Deep learning has become popular in BCI-EEG motor imagery classification. A large body of work using convolutional neural networks [56, 57, 58, 59, 60] are able to handle an extremely low Signal-to-Noise Ratio (SNR) that exists in BCI-EEG systems [53, 58, 60, 61, 62]. The low SNR is thought to be a result of having a conglomerate of biological signals from a multitude of involuntary muscle contraction sources such as eye blinking and heart rhythm. Moreover, EEG sensors are sensitive to interference from nearby electrical power sources [63], further suggesting noise removal is of vital importance. The origin of EEG sensors being burdened by low spatial resolution falls under biological phenomenon associated with volume conductive effects [22]. To better deal with blind source separation, recurrent neural networks equipped with attention mechanisms [64] such as long short-term memory (LSTM) and gated recurrent units to extract sequential relationships in time series has been implemented [65, 66, 67, 68].

A common practice in BCI-EEG studies has been to omit bad trials from subjects or the removal of all data from certain subjects classified as EEG-illiterate [26, 48, 51, 54, 56, 57, 58, 57, 59, 60, 65, 66, 67, 68, 69]. While the removal of statistical outliers is justified in the scientific method given a high SNR, at low SNR this practice in ML creates a minimum failure rate baseline as an initial step. The long term aim would then be to reduce the minimum failure rate in applications through improved methods if possible.

## 1.3 Supervised Projective Learning for Electroencephalography Analysis

The primary purpose of SPLEEGA is the removal of artifacts from EEG signals. Through an exploratory process of multivariate EEG signals, SPLEEGA is capable of identifying underlying spatio-temporal patterns. The SPLEEGA paradigm can be decomposed into 4 generalized procedural steps As illustrated in Figure 1.3: *i*) preprocessing; *ii*) artifact

removal; *iii*) feature extraction; and *iv*) classification.

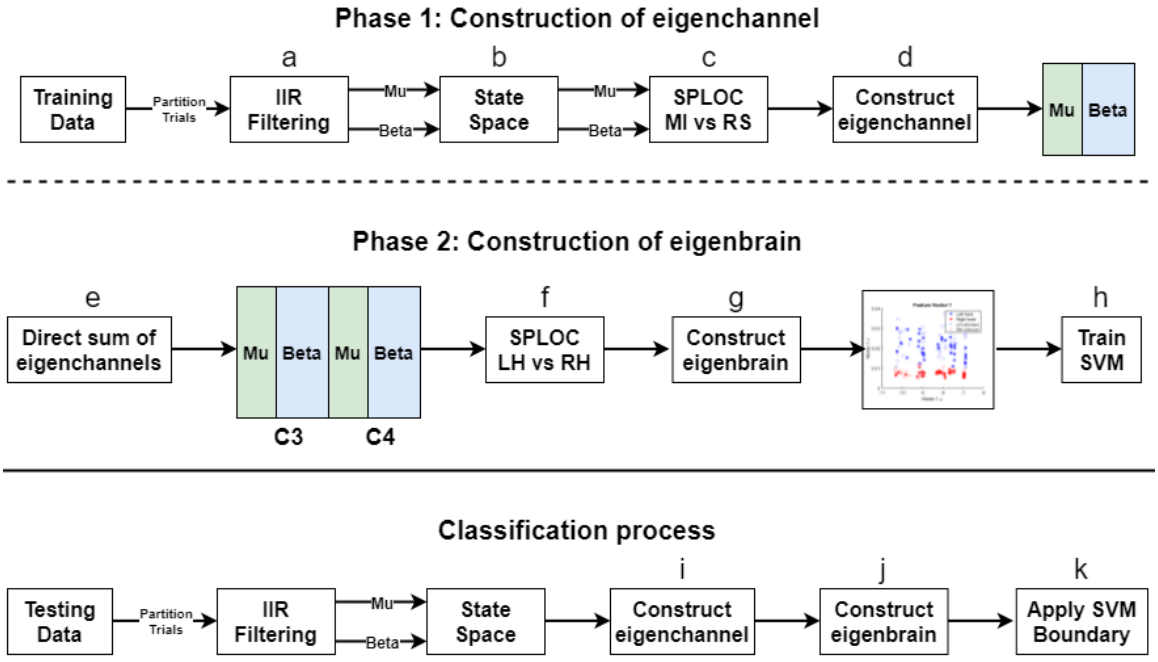


Figure 1.3: SPLEEGA pipeline. (a) An infinite impulse response filter is applied to create the $\mu$ and $\beta$ components of the power spectral density. (b) A time-correlated state space is created per frequency band to increase the dimensionality from 1 to $p$. (c) The subspace that discriminates motor imagery from resting state signals is created by training SPLOC at both frequency bands independently. (d) The SPLEEGA eigenchannel is obtained as the direct sum of the ($\mu$,$\beta$) state spaces projected onto the ($\mu$,$\beta$) discriminant basis vectors from SPLOC. (e) Phase 2 begins by taking a direct sum of the eigenchannels. (f) SPLOC is then used for binary classification of LH and RH MI signals. (g) The discriminant modes from SPLOC are used to construct the eigenbrain. (h) Utilizing the eigenbrain to cluster LH and RH MI data, a support vector machine with a cubic kernel is employed to obtain a decision boundary. On test data, raw EEG signals are manipulated using a sequence of linear operations to (i) construct eigenchannels, (j) construct the eigenbrain, and (k) make predictions based on how the data is represented with respect to the decision boundary.

Dimension reduction is achieved by identifying a discriminant basis set. Importantly, SPLOC constructs an orthonormal basis that isolates the a priori ambiguous metric containing discriminatory information [70, 71]. The application of an SVM then becomes straightforward. An important computational aspect of SPLEEGA is that after an initial calibration phase, brain patterns are revealed by projecting EEG signals onto a small number of basis vectors, making this a rapid calculation [72, 73].

# CHAPTER 2: DATASET AND PREPROCESSING

## 2.1 Giga Science Motor Imagery Dataset

The GigaDB dataset [69] from the Gwangju Institute of Science and Technology is utilized in this study. This data consists of 64 EEG channels and 4 EMG channels. The 64 Ag/AgCl active electrodes were configured according to the international 10-10 standard. All EEG signals were recorded at a sampling rate of 512 Hz during MI tasks using the BioSemi ActiveTwo system with the BCI2000 3.0.2 software. The pool of participants was comprised of 52 subjects. Each recording session was divided into 5 runs where either 20 or 24 MI trials were performed. This resulted in 100 or 120 total left and right-hand trials per subject. As noted by the original authors, 38 of the subjects contributed well-discriminated datasets while the remaining 14 were deemed non-discriminatory. A comprehensive explanation of the subject characteristics and experimental protocols during data acquisition can be found in the original work. The procedure for the acquisition of one motor imagery trial is shown in Figure 2.1.

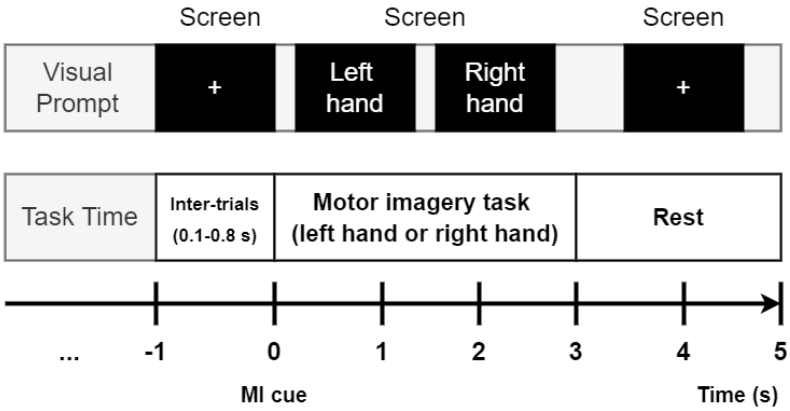


Figure 2.1: The experimental protocol for MI signal acquisition. During EEG recording sessions, subjects were prompted with a visual cue on a screen in order to isolate MI trials.

During experimentation, two criteria were evaluated to identify bad trial indices which are often omitted. The first of these criteria is the removal of a trial if an amplitude greater than $\pm 100\ \mu$V is measured within a 500-2500 ms segment. Second, correlation with EMG (real hand movement) is considered following an approach in [74]. Here, all trials highly correlated with EMG are omitted. In the methodology presented here, the bad trials are *not* omitted as a real-time solution is expected to handle these types of outliers. This study utilizes both sparse (C3, C4) and contralateral (Cz, C1-C6, T7, T8) sensor configurations; subsequently, this reduces the dimensionality of MI trials to 2 and 9 respectively as shown in Figure 2.2.

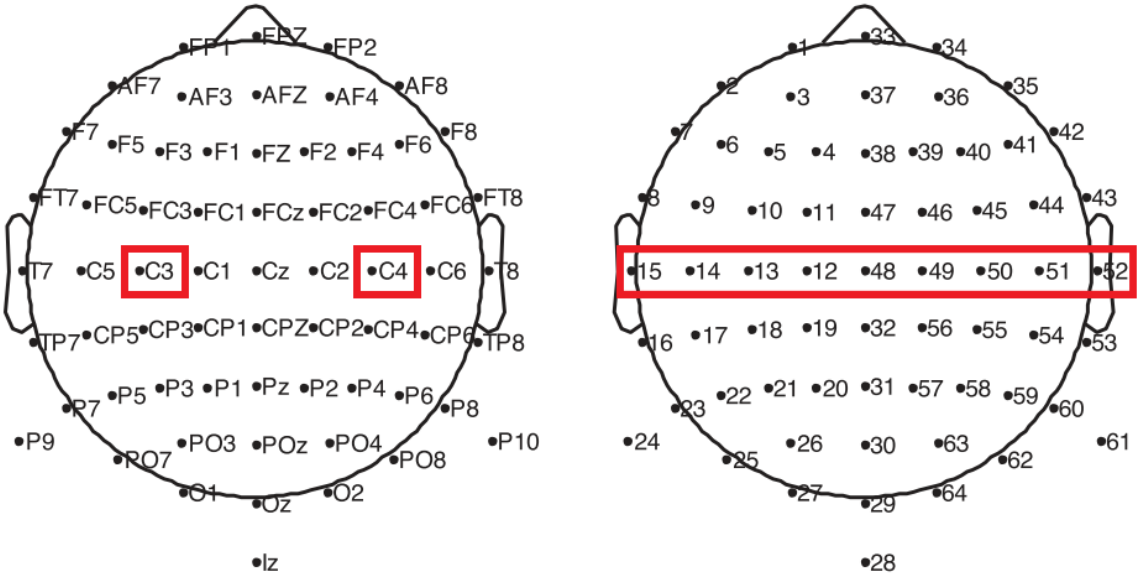


Figure 2.2: Electrode configurations. The 64 electrode 10-10 international standard montage used during signal acquisition. (left) The sparse-sensor selection consists of the C3 and C4 electrodes indicated by red rectangles. (right) The contralateral electrode configuration is comprised of 9 sensors. In the right figure, the corresponding channel numbers to the channel names (left) are used for categorizing the dimensions of the dataset.

## 2.2 Channel Referencing and Trial Partitioning

The LH and RH data matrices of raw EEG signals are of size 358400 x 64 from GigaDB. This is reduced to 358400 x 4 for the sparse sensor array and 358400 x 9 for the contralateral configuration. A reference signal is defined as the average signal of the two mastoid channels

TP7 and TP8 as shown in Figure 2.3. Referencing is performed by subtracting the reference signal from the desired elctrode signals (C3-C4, and contralateral) [75].

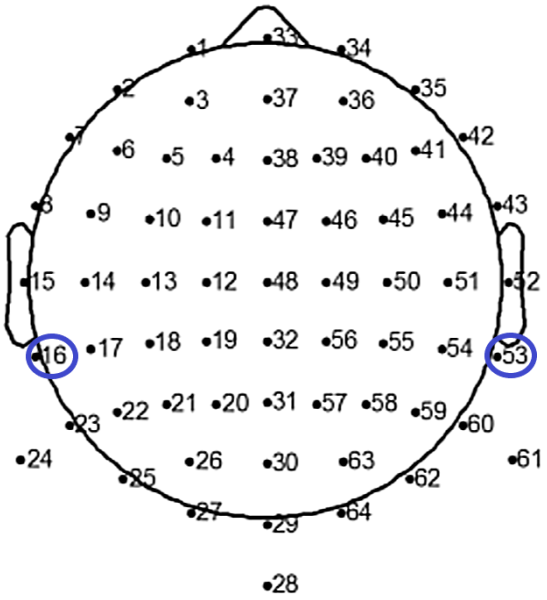


Figure 2.3: Reference electrode configuration. The 64 electrode 10-10 international standard montage used during signal acquisition. Referencing was conducted relative to the mastoid TP7 and TP8 channels shown in blue ellipses.

Given the event markers within a response vector, MI trials were analyzed over 4 second intervals (2048 samples) from the onset stimulus time. A signal skip of 0.5 seconds (256 samples) was used to compensate for human reaction time. After the referencing and partitioning steps, there were 100 trials of size 2048 x 2 (sparse sensors) and 2048 x 9 (contralateral) for LH and RH MI. Some subjects contain 120 trials, in these cases 20 trials were randomly omitted. The RS signals were experimentally acquired for 60 continuous seconds prior to MI yielding a 33792 x 64 data matrix. Other than the acquisition time difference, the RS signals are processed in the same way as the LH and RH signals, resulting in 14 trials of size 2048 x 2 and 2048 x 9.

## 2.3 Infinite Impulse Response Filter

Infinite impulse response (IIR) is a property that applies to time-invariant systems where an induced pulse does not converge to zero after some given time [76]. Two IIR filters were applied to the partitioned trials in order to construct the $\mu$ (7.0-13.0 Hz) and $\beta$ (13.0-28.0 Hz) components of all referenced EEG signals. An IIR filter can be expressed by the transfer function:

$$H(z) = \frac{Y(z)}{X(z)} \tag{2.1}$$

Where $Y(z)$ and $X(z)$ represent the z-transformed output and input signals respectively. Note that after the z-transformation, the source input signal $x(t) \rightarrow X(z)$ and the output $y(t) \rightarrow Y(z)$. Expanding $Y(z)$ and $X(z)$ as a ratio of polynomials:

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{l=0}^{N} a_l z^{-l}} = \frac{z^N}{z^M} \frac{b_0 z^M + b_1 z^{M-1} + \ldots + b_M z^M}{z^N - a_1 z^{N-1} - \ldots - a_{N-1} z^{-1} - a_N} \tag{2.2}$$

There are $M$ roots in the numerator of Equation 2.2 known as the zeros, with $N$ roots in the denominator called poles. The goal is then refined to finding the filter coefficients $a$ and $b$ such that the desired frequency response $\mathcal{H}(z)$ is attained:

$$\mathcal{H}(\hat{\omega}) = H(z)\Big|_{z=e^{i\hat{\omega}}} = H(e^{i\hat{\omega}}) \tag{2.3}$$

The real component of Equation 2.3 represents the magnitude response $|\mathcal{H}(\hat{\omega})|$. As a developmental check, the magnitude responses for both filters were computed and an example for $\mu$ is provided in Figure 2.4. Each filter contained a stopband attenuation of 40 Hz, and a passband ripple of roughly 0.1 Hz. The use of bandpass filters was motivated to mitigate undesired frequencies outside the defined ranges due to spectral leakage [77]. A typical LH MI time sequence trial is shown in Figure 2.5 along with its $\mu$ and $\beta$ components. After completion of IIR filtering, each referenced EEG signal becomes bifurcated as two distinct time series, which are subsequently processed independently.

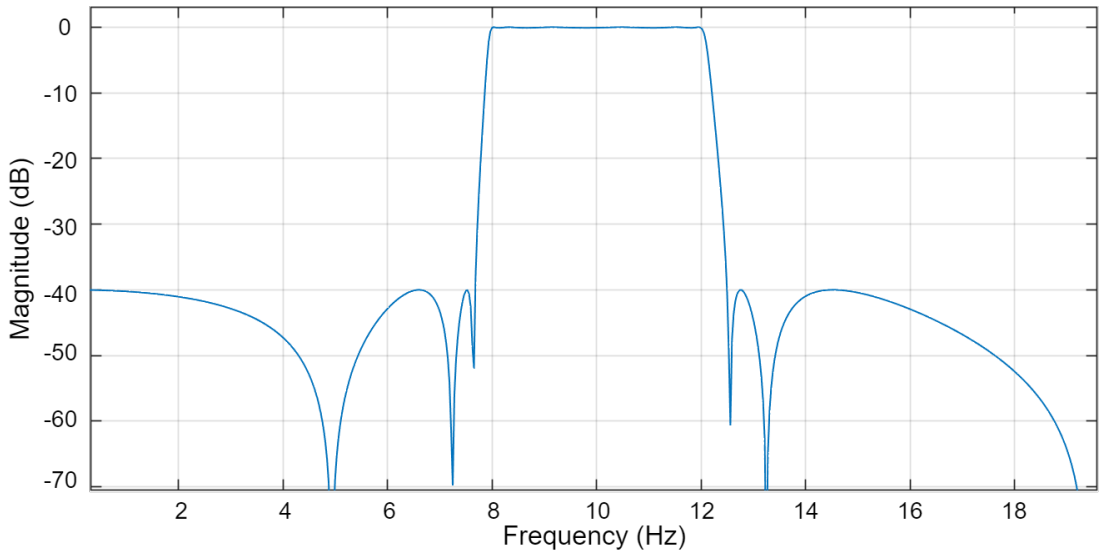


Figure 2.4: Magnitude response plot for $\mu$ IIR filter. The solid blue line represents the transfer function over the desired frequency range. The stopband attenuation of 40 decibels is shown where the transfer function is constrained outside of 8-12 Hz.

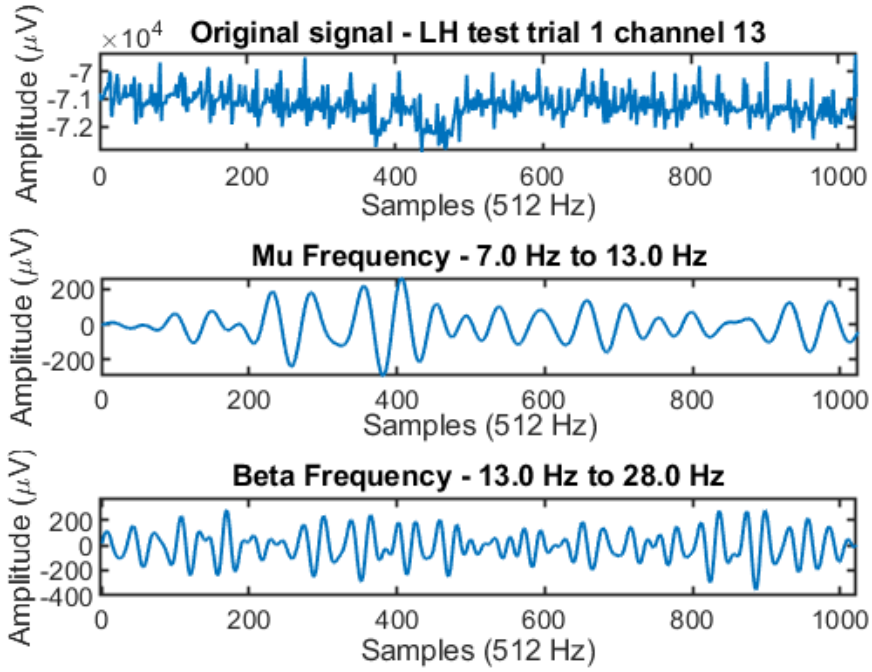


Figure 2.5: IIR filtered signals. (top) A typical referenced EEG signal from the subject executing LH MI. (middle) The time series after a $\mu$ bandpass filter is applied. (bottom) The time series after a $\beta$ bandpass filter is applied.

## 2.4 State Space Representation

Given a signal $x(t)$ for a channel and frequency band, a $p$-dimensional state space is defined by forming the state vector $X(t) = [x(t), x(t+\tau), x(t+2\tau), \ldots, x(t+(p-1)\tau)]$. For the results shown here, $p = 32$ with $\tau = 0.015625$ seconds using a 8 sample step, thus reducing each trial to 2048 samples. As $p$ increases there is less sampling in higher dimensions, although more memory is retained in the state vector for discrimination. Experimentally observed, $p$ ranging from 8 to 128 works well, with $p = 32$ being a good compromise between sampling adequacy and number of discriminant modes comprising the eigenchannels. A general example of expanding one signal at the $\mu$ frequency into its respective state space is presented in Figure 2.6. Within the pipeline, the data matrices per frequency band ($\mu/\beta$) used to construct an eigenchannel are of size 2048 x 32 for LH, RH, and RS which are presented to SPLOC for training in following steps. Worth noting, the expansion into a highly-correlated state space contains mostly redundant information. When the top basis vector from eigenchannel construction is retained, this represents the transformed state vector which contains the most discriminatory information; therefore, this acts as an indirect temporal alignment with respect to optimal discrimination.

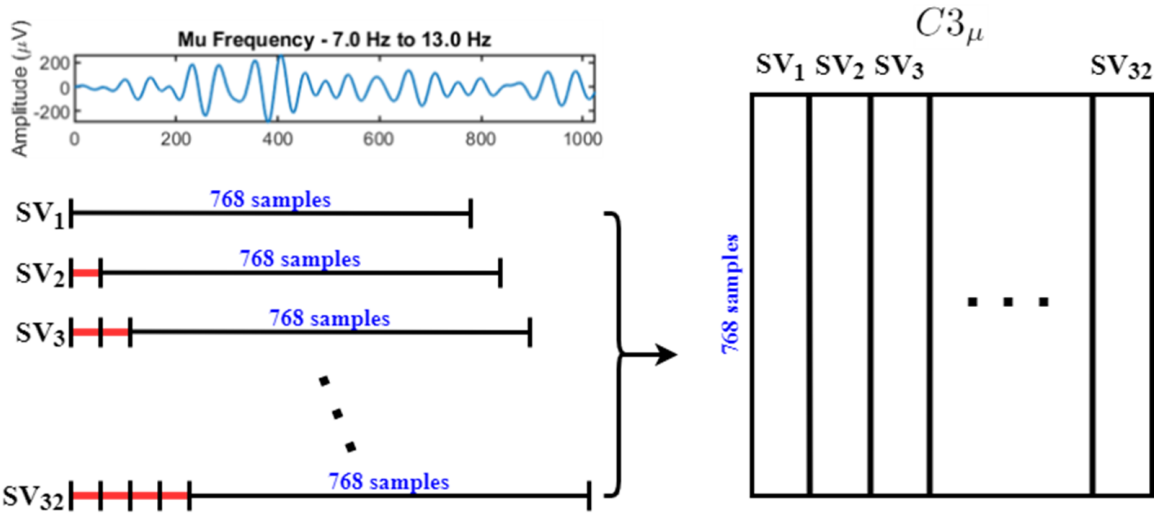


Figure 2.6: State space representation. An example signal from the C3 electrode in the $\mu$ frequency which is expanded into a 32-dimensional highly-correlated state space.

# CHAPTER 3: SPLOC AND THE MODE-FEATURE SPACE PLANE

The underlying engine of SPLEEGA is a new intelligent PP network known as supervised projective learning with orthogonal completeness (SPLOC). SPLOC was created by Professor Donald Jacobs in 2018, then it was further developed by Tyler Grear and Chris Avery from 2019-2020. Here, projection pursuit (PP) is implemented as an iterative process to change the directions of a complete set of basis vectors for the purpose of maximizing a projection index. Using Dirac bra-ket notation, the $i$-th basis vector in column form on the $s$-th step is denoted as $|i, s\rangle$. The molecular conformation of $N$ atoms is described by $2N$ vectors that span a $2N$ dimensional space. This generalizes to higher dimensional structures. Through successive orthogonal rotations, the set of basis vectors remain orthonormal and complete, $\forall s$, such that $\langle j, s|i, s\rangle = \delta_{ji}$ and $\sum_{i=1}^{2N} |i, s\rangle\langle i, s| = I$, where $\delta_{ji}$ is the Kronecker delta and $I$ is the identity matrix.

The initial projection index is comprised of three objective sub-functions that seek to optimize three criterion during rotations: the primary member - signal to noise, with the secondary members - statistical significance and clustering quality. These sub-functions are referred to as the decision triad, and together they decide which basis vectors are discriminate (d-mode), undetermined (u-mode), or indifferent (i-mode). Let $|k_1\rangle$ and $|k_2\rangle$ be two randomly selected vectors defining a plane. An orthogonal rotation of the plane is performed where the step size of the rotation angle $\phi$ defines the resolution of the vector pair sweeps. Upon each iterative rotation, the data is projected onto the basis vectors $|k_1\rangle$ and $|k_2\rangle$, from which two emergent features are extracted: the mean and standard deviation. These are evaluated by the decision triad to determine the final efficacy projection index $E_k$. $E_k$ ranks the procured basis sets according to the best performers from statistical testing.

The primary member of the decision triad contains a scoring function $S_{\mathrm{k}}$ that is applied

to the $k$-th mode. Furthermore, $S_{\mathrm{k}}$ bifurcates the problem by setting the minimum threshold for being a discriminant mode $S_{\mathrm{d}}$ or an indifferent mode $S_{\mathrm{i}}$ independently, with $S_{\mathrm{d}} > S_{\mathrm{i}}$. Let $snr(k|\alpha,\beta) = |\mu_k(\alpha) - \mu_k(\beta)|/\sqrt{v_k(\alpha) + v_k(\beta)}$ define the signal to noise ratio for the $k$-th mode when comparing the $\alpha$-th functional system to the $\beta$-th nonfunctional system, and $sbn(k|\alpha,\beta) = \max(0, snr(k|\alpha,\beta) - 1)$ is signal beyond noise. Let $rex(k|\alpha,\beta) = \max(\sigma_k(\alpha)/\sigma_k(\beta), \sigma_k(\beta)/\sigma_k(\alpha)) - 1$ be the excess ratio of standard deviations from the two systems being compared. Let $S_m = \sqrt{S_i S_d}$ be the geometric mean of the two thresholds. This allows $S_{\mathrm{k}}$ to be expressed as:

$$S_k(\alpha,\beta) = \begin{cases} \sqrt{sbn^2 + rex^2} + 1 & \text{when} > S_d \\ \sqrt{snr^2 + rex^2} + 1 & \text{when} < S_i \\ S_m & \text{otherwise} \end{cases} \tag{3.1}$$

Note, the decision to index a basis vector is determined by $S_{\mathrm{k}}$; however, in order to remain either a i-mode or d-mode, it must pass further testing by the secondary objective functions. Next, the objective function for statistical significance is applied. For this purpose, two Voting Activation Functions (VAFs) are introduced, $f_d$ and $f_i$. Each VAF is presented with a d-mode or i-mode by $S_k$, and a consensus vote accepts or rejects the proposal through the functions:

$$f_d(x) = [1 + \exp(16(x_d - x))]^{-p_d(x)} \tag{3.2}$$

$$f_i(x) = \left[1 - [1 + \exp(16(x_i - x))]^{-1}\right]^{p_i(x)} \tag{3.3}$$

The powers $p_d(x)$ and $p_i(x)$ ensure the functions $f_d(x)$ and $f_i(x)$ change more rapidly than a sigmoid function when respectively $x < x_d$ and $x < x_i$. A consensus vote threshold is set for the scores of $f_d(x)$ and $f_i(x)$ on the proposal that a mode is either discriminant or indifferent, this voting threshold is denoted $V_t$.

The final test of the decision triad is for quality of clustering. The emergent features $\mu_k$ and $\sigma_k$ characterize how a data packet projects onto the $k$-th mode. Together, $(\mu_k, \sigma_k)$

defines a point on a plane representing a two-dimensional cross-section in the $2p$ (e.g. $4N$) feature space, this plane is called a *mode feature space plane* (MFSP) comprising $N_F$ functional and $N_N$ nonfunctional systems. After the final modes are selected, the complete basis set is ranked-ordered by $E_k$. The MFSP is an important output of SPLOC which has been integrated into SPLEEGA which can be seen in Figure 3.1c. The MFSP acts as a multi-dimensional 'observation window' during multivariate exploration.

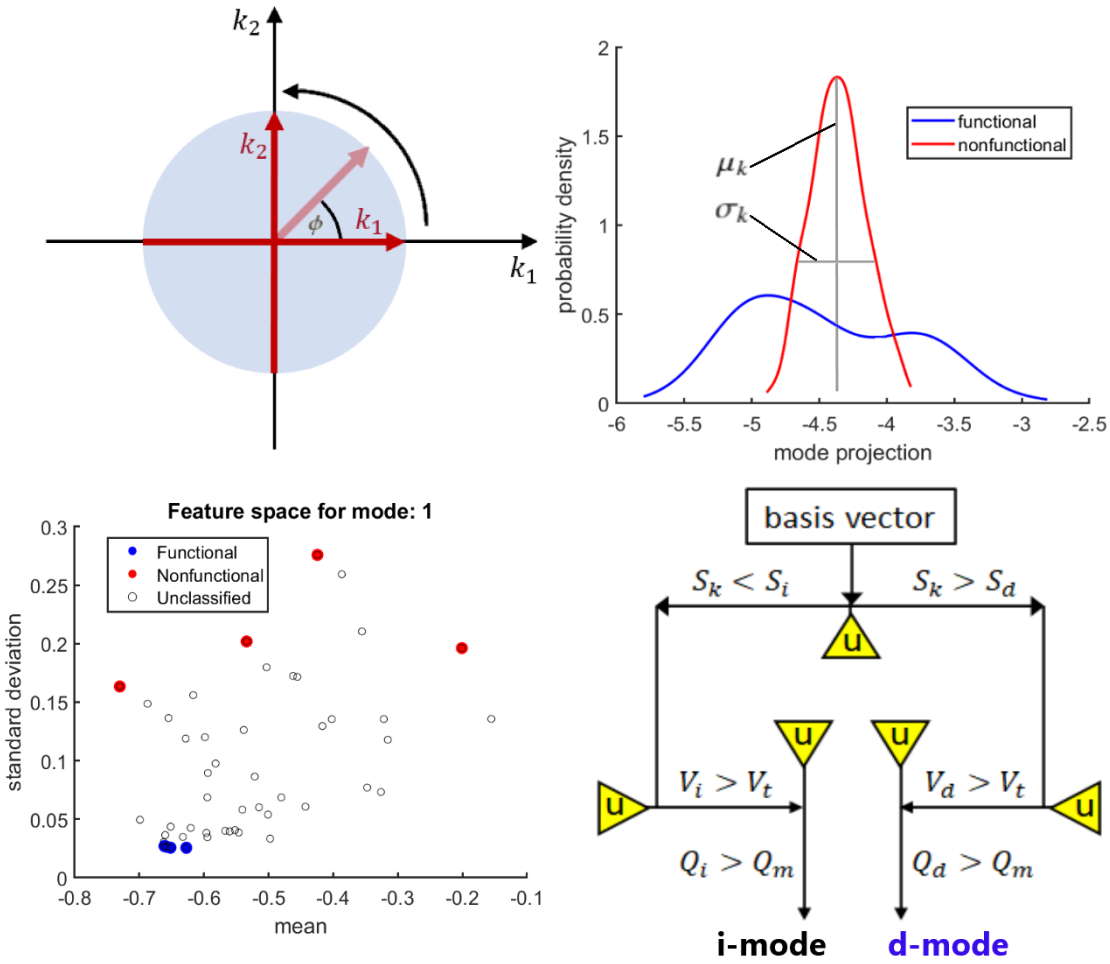


Figure 3.1: A diagram of the SPLOC mode-generating algorithm. (top left) The vector pair $|k_1\rangle$ and $|k_2\rangle$ define a plane which is iteratively rotated from -$\pi$/4 to $\pi$/4. This sweep spans the 2D space of the plane. (top right) With each rotation step $\phi$, the data is projected along $|k_1\rangle$ and $|k_2\rangle$ and the emergent features $\mu_k$ and $\sigma_k$ are extracted for testing of: signal to noise, statistical significance, and quality of clustering. (bottom left) Once a mode passes the first two tests it is evaluated for class separation on an MSFP where each point represents an entire synthetic molecule data packet. (bottom right) A schematic of the decision triad shows how the problem is bifurcated by $S_k$.

The original motivation for SPLOC was to classify molecular dynamics trajectories that represent data packets. For this application, the extended state spaces at different points in the pipeline represent a multivariate time sequence of data (e.g. data packet). Then SPLOC finds the optimal basis set for discrimination as a neural network by presenting to it examples of two distinct classes of data packets. For an eigenchannel, the two classes were MI versus RS. For the eigenbrain, the two classes were LH versus RH.

# CHAPTER 4: ARTIFACT REMOVAL USING PROJECTION PURSUIT

## 4.1 SPLEEGA Eigenchannel Construction

For a state space dimension of 32, SPLOC is presented $N$ LH and $N$ RH data packets as an MI-class, and 14 RS data packets as an RS-class during the eigenchannel construction. This gives $2N$ examples of the MI-class and 14 examples of the RS-class. In this section, channel C4 is used to illustrate the process of constructing an eigenchannel, which is the same for any channel. The example SPLOC basis vector spectrum is shown in Figure 4.1.

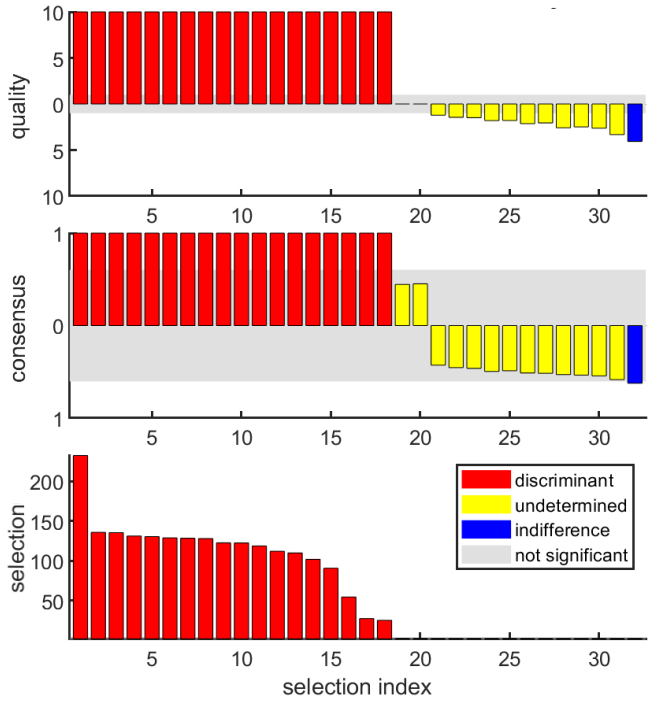


Figure 4.1: SPLEEGA eigenchannel C4 $\mu$ basis spectrum. The primary output of SPLOC is the basis spectrum which represents a complete basis set comprised of discriminant (red), undetermined (yellow), and indifferent (blue) basis vectors. To be considered a discriminant mode, the basis vectors must pass the tests of each triad member: clustering quality (top), consensus (middle), and selection power (bottom). This is a complete basis of 32 dimensions that is the same number of dimensions as the $\mu$ state space which was presented to SPLOC for training.

To construct a single eigenchannel, the ($\mu$,$\beta$) state spaces of C4 are projected along their respective d-modes. This process removes noise, and constitutes artefact removal. The eigenchannel is constructed by the direct sum of the found d-modes for both frequency bands. For example, when 10 d-modes are identified by SPLOC in the $\mu$ band and 5 d-modes from the $\beta$ band, the direct sum of these d-modes results in a 15-dimensional discriminant subspace. In general, each eigenchannel will end up with a discriminant space of different dimensions. An example case for illustration is presented in Figure 4.2.

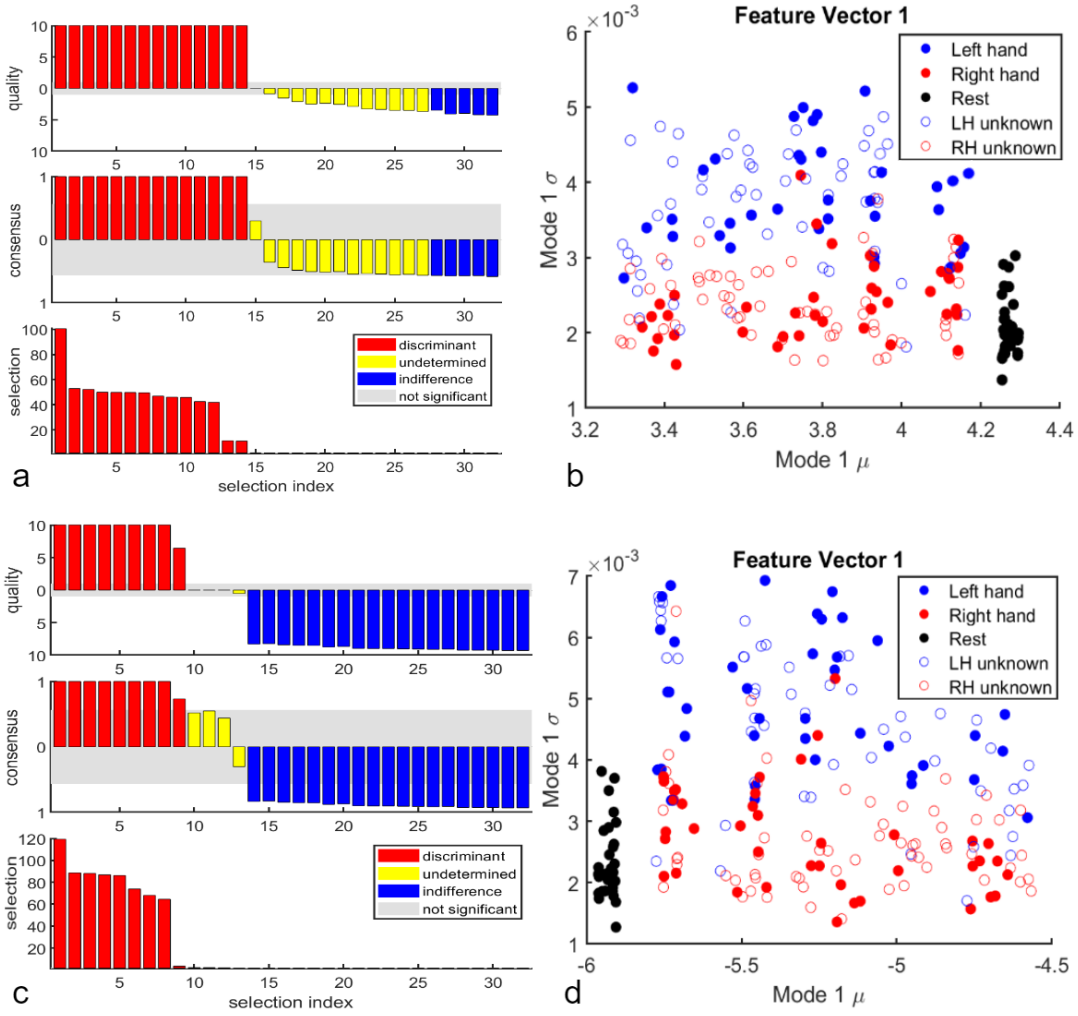


Figure 4.2: SPLEEGA eigenchannel C3. The first SPLOC run during eigenchannel construction separates MI from RS signals. (a) For the $\mu$ frequency, 14 d-modes were identified. (b) The MFSP from the top d-mode for the $\mu$ band exhibits advanced clustering of MI signals (blue and red circles) with respect to resting state (solid black circles). (c) 9 d-modes were captured by SPLOC within the $\beta$ component of eigenchannel C3.

# CHAPTER 5: CLASSIFICATION OF EEG SIGNALS

## 5.1 Support Vector Machines

This section gives a brief review of the basic elements behind a support vector machine (SVM). The SVM was originally developed for the purpose of binary classification; however, research is ongoing in the ML field to generalize for multi-class classification [78]. Here, the foundations for SVMs are presented which were proposed by Vladimir Vapnik in his PhD thesis at Moscow University in 1963 [79], later in 1995 this was generalized for statistical and computational learning [80]. The SVM optimization problem is shown in Figure 5.1.

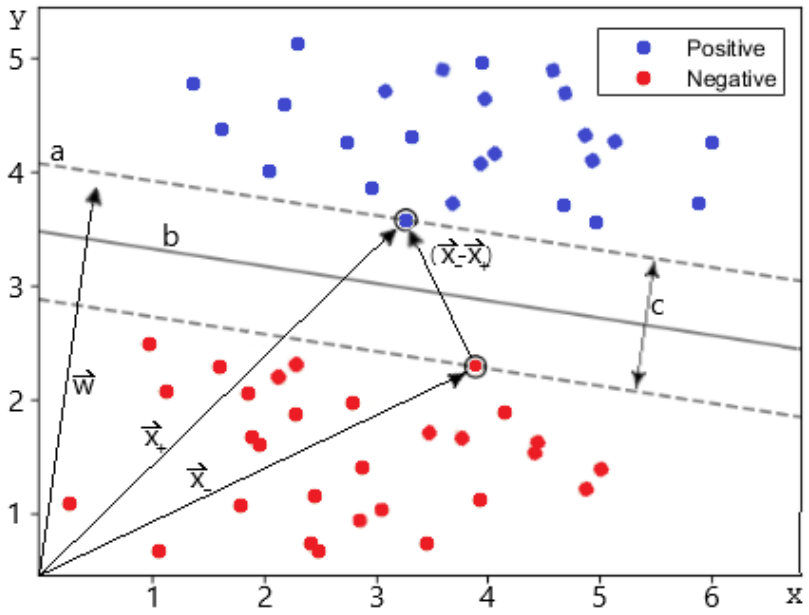


Figure 5.1: The foundation for the optimization problem of SVMs. Two distributions of data from two classes are plotted on an xy-plane where the blue dots indicate positive samples, and the red dots indicate negative samples. (a) Two support vectors are defined by a known data point from each class. (b) A decision boundary is placed such that the margin between support vectors (c) is maximized. A vector of undetermined length $\boldsymbol{w}$ is introduced, constrained to be perpendicular with the support vectors. Two vectors point in the direction of known samples of each class $\mathbf{x}_+$ and $\mathbf{x}_-$. The width of the margin $\boldsymbol{w}'$ is determined by subtracting $\mathbf{x}_+$ from $\mathbf{x}_-$, then projecting along a normalized vector $\mathbf{w}$.

Figure 5.1 shows that the SVM optimization problem takes the form:

$$max\frac{1}{\|\boldsymbol{w}'\|} \quad or \quad min\|\boldsymbol{w}'\| \tag{5.1}$$

It is important to note that with certain data domains linear class separation is not possible due to the overlap in coordinates. For this reason, many SVMs implement a soft boundary that controls the overlap of class samples by the use of slack variables [81]. In order to apply SVMs to non-linearly separable classes a kernel function is introduced that will transform the data from the current space, to a space that is more convenient. This method is often referred to as the kernel trick [82], and is made possible by the total mathematical dependence on the dot product of samples, which can be shown when the problem is formalized as an Euler-Lagrange equation for constrained optimization [83]. The goal is to map the data to a higher dimensional space in order to acquire class separation. To do this, a kernel K is introduced such that:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \psi(\boldsymbol{x}_i) \cdot \psi(\boldsymbol{x}_j) \tag{5.2}$$

where $\psi$ represents the transformation that maps the data to a higher dimensional space. Equation 5.2 implies that the actual transformation into a higher dimensional space does not need to be known, just the kernel function given the constraints of the Euler-Lagrange equation are fulfilled. Not all functions can serve as kernel functions, only those that are considered convex spaces [84]. This is because a convex space does not suffer from local minima or local maxima where some methods such as neural networks exhibit problems dealing with these issues [85]. The mathematical criterion in which a kernel is valid to use in this manner is called Mercer's condition [86]. An example of separating two non-linearly separable classes using a linear hyperplane accompanied by the radial basis function as the kernel is presented in Figure 5.2.

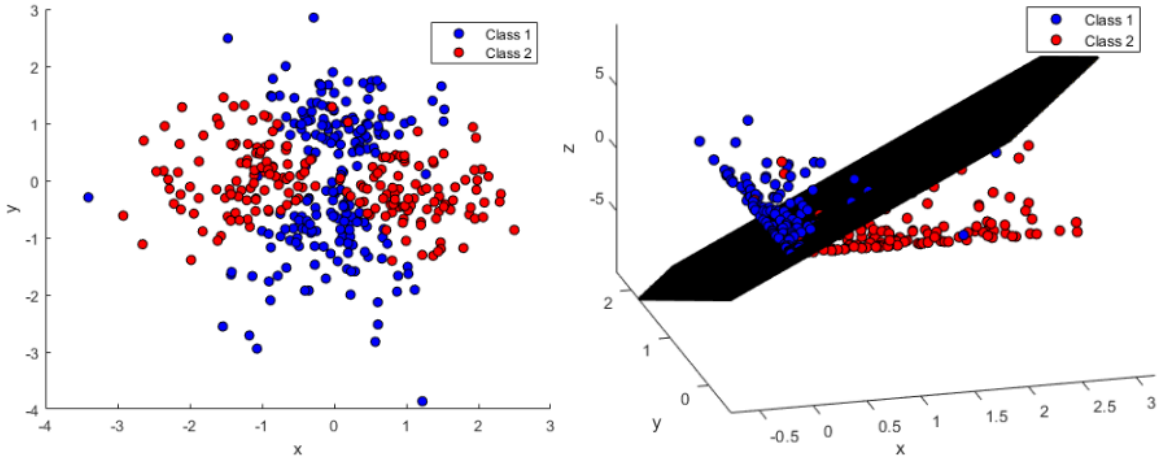


Figure 5.2: Application of an SVM to non-linearly separable distributions of two classes. (left) A 2D plot of the data points indicates that there is no linear boundary that will achieve good separation. (right) The data is projected into a third dimension using a kernel function, class separation is then dictated by a linear hyperplane. (right) Some misclassifications occur which is allowed by utilizing a soft margin when constructing the SVM.

# CHAPTER 6: AUTOMATIC SINGLE-CHANNEL CLASSIFICATION

The development of the automated procedure for single channel classification was motivated largely from the trends observed during SPLEEGA eigenchannel development. In this chapter, SPLEEGA is reformulated as an iterative optimization procedure over the 9 contralteral electrodes. Here, the number of MI trials used for eigenchannel construction were held fixed at 7, resulting in 14 MI trials (7 LH and 7 RH). This was to achieve balance in the class distribution during eigenchannel training since there are 14 available RS trials at 2048 samples. The number of testing trials was held fixed at 50 (100 LH and RH), and the remaining 86 trials (43 LH and RH) were put aside as a validation set. There are 3 iterations of varying frequency ranges for both $\mu$ and $\beta$. For $\mu$, iteration 1 utilized 8-12 Hz, iteration 2 used 7.5-12.5 Hz, and iteration 3 applies the range of 7-13 Hz. For $\beta$, iteration 1 used 14-24 Hz, iteration 2 utilized 16-26 Hz, and iteration 3 contained the frequency range 28-28 Hz. The reasoning for multiple frequency bands during optimization is grounded in the uncertainity of a clearly-defined set of viable bands as discussed in Section 1.2.

The automated SPLEEGA procedure can be expressed in x general steps: First, for channel $N$ an eigenchannel is constructed at the $\mu$ and $\beta$ bands resulting in (discriminant,undetermined,indifferent) subspaces. Only if d-modes are identified at the eigenchannel level they are then projected into the corresponding discriminant subspace and SPLOC is performed for a validation step. For example, with the $\mu$ frequency for channel $N$ 8 d-modes are identified during artifact removal, then the trials are projected into this incomplete discriminant subspace of dimensionality 8. The 8-dimensional trials are then presented to SPLOC for validation where 7 LH trials comprise class 1 with class 2 containing RH trials. Of the 8 dimensions presented to SPLOC during this phase, if d-modes are identified, their corresponding MFSPs are evaluated for prediction accuracy through the

application of an SVM. If no d-modes are available for validation, the algorithm defaults to u-modes. If no u-modes are available for validation, the frequency band of channel *N* is omitted. Continuing the example, only the top mode identifed by SPLOC during validation progresses to an SVM box constraint optimization. Here, the 86 validation trials are used as a check while sweeping over 1250 different box constraints. After box constraint optimization the top mode from each frequency band is retained and a final classification is performed using only the 100 testing trials. This accuracy is then recorded for both bands of each contralateral electrode. The best performing channel and frequency band are recorded and represent the optimal electrode selection over the 9-channel sweep.

Ultimately, the SPLEEGA automated channel selection procedure explores whether a sparse-sensor configuration is sufficient for a large enough population, or more flexibility is required when determining which sensors can contribute most effectively for analysis.

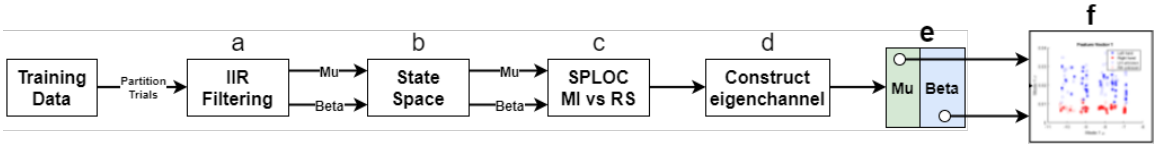


Figure 6.1: (a-d) The preprocessing steps for this procedure remain the same as previous algorithmic protocols. (e) Single-channel classification is performed at each desired frequency band. (f) The SVM model developed using the training data is validated to attain the optimal box constraint before presenting the machine with testing data. The automatic channel selection procedure performs the workflow shown above for *n* iterations, where *n* is the number of desired frequency bands to sweep over during optimization. The small scale of the MFSP in (f) is intentional and used to illustrate that an SVM decision boundary is applied at this point of the pipeline.

# CHAPTER 7: RESULTS

## 7.1 Artifact Removal Results

The purpose of the eigenchannel is to separate MI from RS at both $\mu$ and $\beta$ frequency bands. The results presented in this section are for all 52 subjects from the GigaDB dataset. For this phase, the sparse-sensor configuration was utilized for a rapid sweep over subjects using the two most commonly used electrodes, C3 and C4. To consolidate the artifact removal results, a quantity known as an artifact removal score denoted $\gamma_{AR}$ is introduced. Heuristically, $\gamma_{AR}$ is the sum of 4 components corresponding to 4 (discriminant, undetermined, or indifferent) subspaces from eigenchannel construction for C3 and C4. If a given subspace for a given channel contains at least one d-mode that separates all MI from RS testing trials, this represents a value of 0.25. For C3 $\mu/\beta$ and C4 $\mu/\beta$, a maximum $\gamma_{AR}$ is therefore 1.0. Figure 7.1 and Figure 7.2 present 2 cases (s50 and s11) which respectively correspond to good and bad artifact removal performance.

The separation of MI trials in Figure 7.1 indicates that SPLOC is finding directions within the 32-dimensional space where the characteristics of MI signals differ from those of RS. An observed trend at all training set sizes of the top d-modes identified by SPLOC is the *advanced clustering* between the LH and RH MI signals. This is especially noteworthy given the machine was not informed that the MI-class consisted of two types of systems by way of labeling. This sub-clustering is a general property of SPLOC as noted previously [8]. The significance of advance clustering is that there are consistent patterns inherent in the data. These results suggest that the advanced clustering within the MFSPs during eigenchannel construction may provide a distinctive signature for robust classification of LH and RH systems with the eigenbrain.

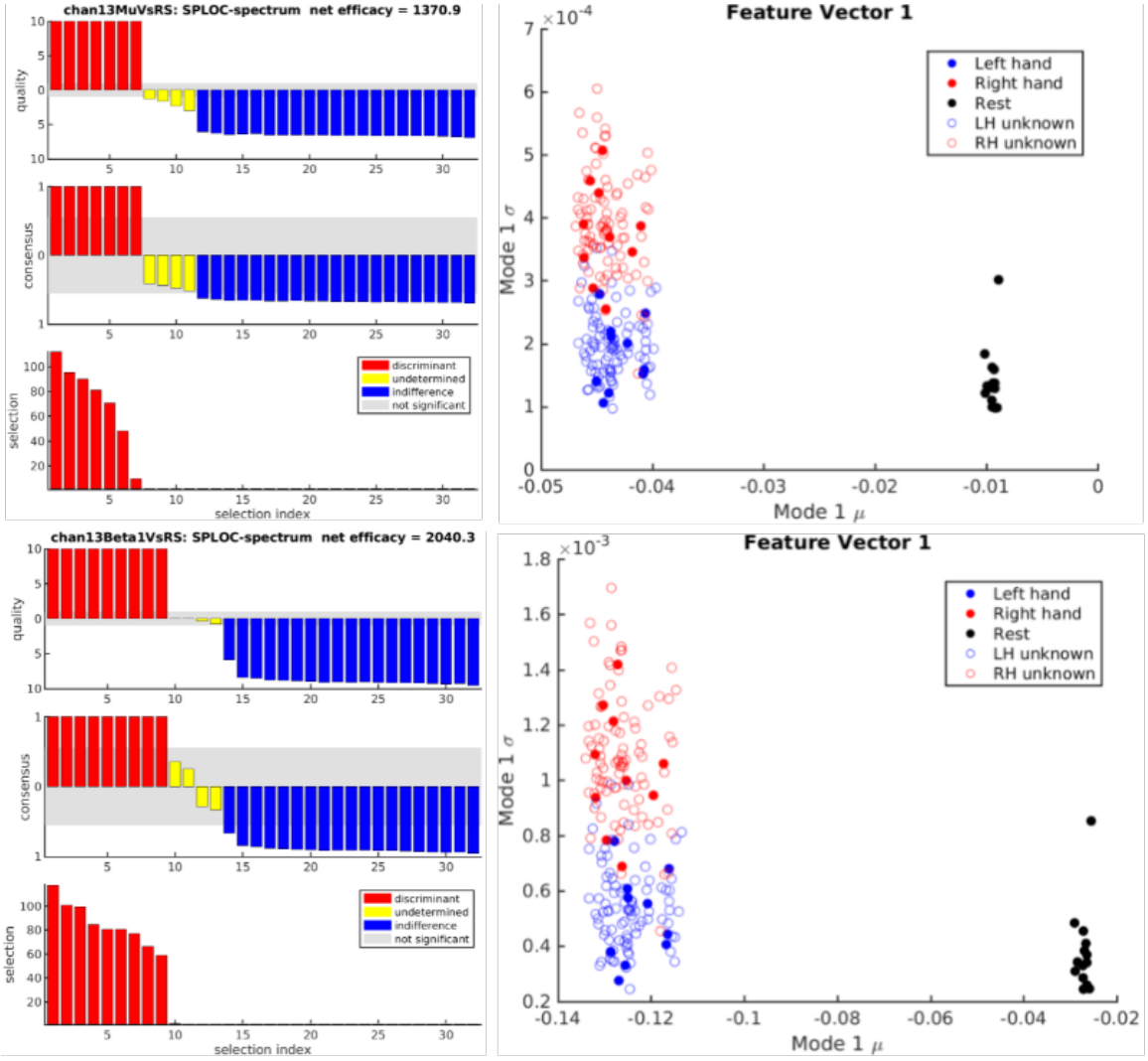


Figure 7.1: The C3 eigenchannel for subject 50. (top left) For the $\mu$ frequency band, SPLOC is trained to distinguish between MI and RS signals, represented by the blue/red and black circles respectively. Solid circles indicate training systems, while empty circles represent MI trials that were unknown to the machine. (top right) Advanced clustering can be observed after projecting along the 1st C3-$\mu$ d-mode and plotting on an MFSP. Similarly the results are favorable for s50 in the (bottom) $\mu$ band discrimination as well. It is readily apparent in both frequency ranges perfect separation from RS systems is achieved.

Moving to the example bad case (s11) as exemplified with the C3 eigenchannel displayed in Figure 7.2. For both the $\mu$ and $\beta$ frequencies, zero d-modes were identified. It can be seen in the BVS for both bands, there are modes that satisfy the 1st (selection power) and 2nd (signal to noise) decision triad evaluations; however, since the quality of clustering metric did not surpass minimum thresholds, a d-mode could not be properly declared. It is

possible that for subject 11 C3 and C4 are not the optimal electrodes for MI classification, this will be discussed further in Section 7.2.

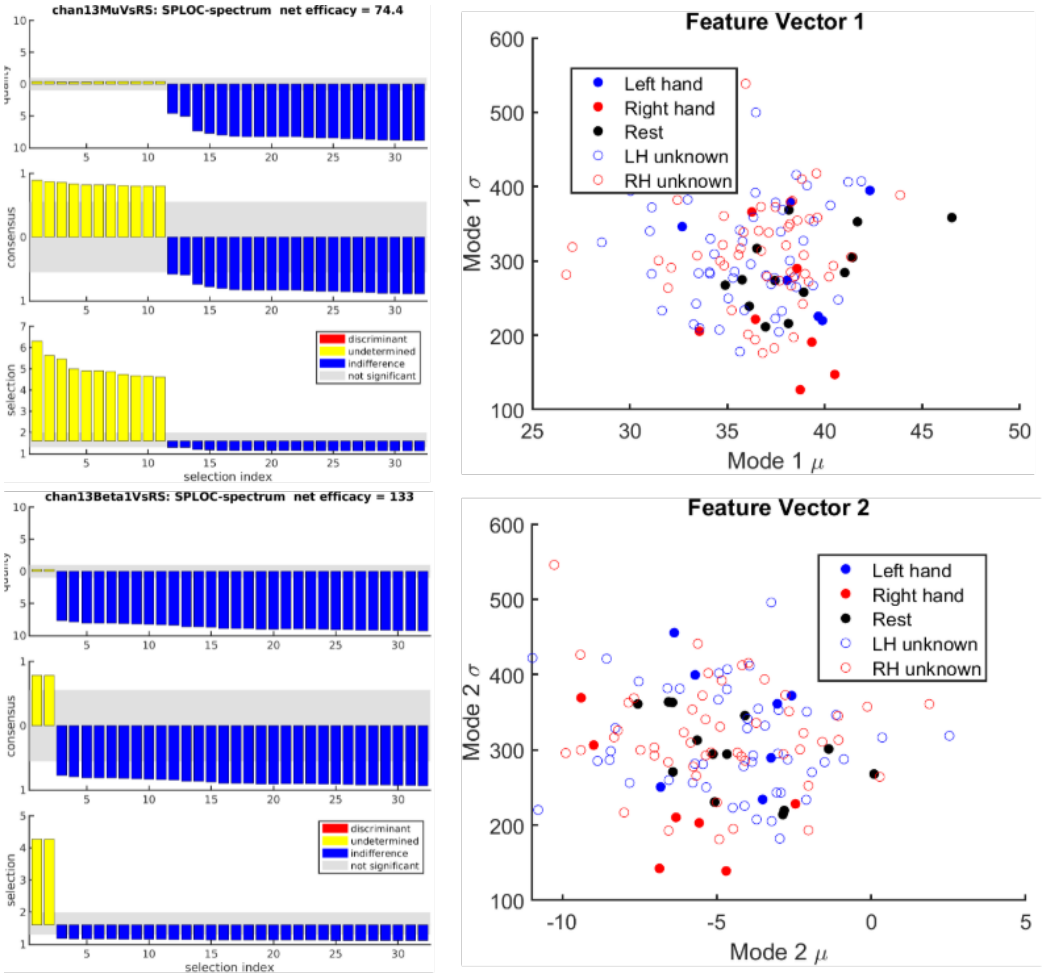


Figure 7.2: The C3 eigenchannel for subject 11. (top left) The basis vector spectrum for the C3-$\mu$ component of the C3 eigenchannel. (bottom left) The basis vector spectrum for the C3-$\beta$ component of the C3 eigenchannel.(right) The trials are projected along the top u-mode direction and the emergent features are plotted in an MFSP. In this case, the separation of MI and RS signals is not attained in either 1-dimensional undetermined subspace for both bands. Although no d-modes were identified, through undirected orthogonal rotations a creative exploration of the undetermined subspace can be performed.

The results from a 52-subject sparse-sensor configuration run are presented in Table 7.1. Here, the number of d-modes acquired during eigenchannel construction for C3-$\mu$, C3-$\beta$, C4-$\mu$, C4-$\beta$ were recorded and provided in columns 2-5. The $\gamma_{AR}$ was recorded for all 52 subjects which is provided in column 6, with column 1 listing the subject nomenclatures.

Table 7.1: Eigenchannel artifact removal results for 52 subjects from GigaDB MI dataset.

| **Subject** | **C3 $\mu$ d-modes** | **C3 $\beta$ d-modes** | **C4 $\mu$ d-modes** | **C4 $\beta$ d-modes** | $\gamma_{AR}$ |
|---|---|---|---|---|---|
| s01 | 11 | 10 | 23 | 8 | 1.00 |
| s02 | 16 | 8 | 14 | 10 | 1.00 |
| s03 | 11 | 10 | 14 | 6 | 1.00 |
| s04 | 9 | 8 | 11 | 10 | 1.00 |
| **s05** | **16** | **8** | **17** | **9** | **1.00** |
| s06 | 8 | 10 | 9 | 13 | 1.00 |
| s07 | 8 | 10 | 12 | 10 | 1.00 |
| s08 | 7 | 9 | 0 | 0 | 0.50 |
| s09 | 8 | 9 | 13 | 5 | 1.00 |
| s10 | 0 | 0 | 6 | 5 | 0.50 |
| **s11** | **0** | **0** | **0** | **0** | **0.00** |
| s12 | 13 | 9 | 14 | 8 | 1.00 |
| s13 | 17 | 9 | 15 | 8 | 1.00 |
| s14 | 0 | 0 | 12 | 7 | 0.50 |
| **s15** | **0** | **0** | **0** | **0** | **0.00** |
| s16 | 17 | 10 | 15 | 10 | 1.00 |
| s17 | 7 | 8 | 0 | 0 | 0.50 |
| s18 | 9 | 13 | 14 | 7 | 1.00 |
| s19 | 7 | 6 | 17 | 9 | 1.00 |
| s20 | 17 | 9 | 0 | 0 | 0.50 |
| s21 | 10 | 6 | 10 | 6 | 1.00 |
| s22 | 0 | 0 | 12 | 11 | 0.50 |
| s23 | 6 | 8 | 16 | 6 | 1.00 |
| s24 | 8 | 7 | 5 | 5 | 1.00 |

Table 7.1 continued from previous page

| **Subject** | **C3 $\mu$ d-modes** | **C3 $\beta$ d-modes** | **C4 $\mu$ d-modes** | **C4 $\beta$ d-modes** | $\Gamma_{AR}$ |
|---|---|---|---|---|---|
| s25 | 0 | 0 | 16 | 7 | 0.50 |
| s26 | 10 | 10 | 7 | 7 | 1.00 |
| s27 | 6 | 4 | 0 | 0 | 0.50 |
| s28 | 10 | 11 | 0 | 0 | 0.50 |
| s29 | 8 | 11 | 15 | 12 | 1.00 |
| s30 | 0 | 0 | 7 | 8 | 0.50 |
| s31 | 9 | 10 | 16 | 7 | 1.00 |
| s32 | 15 | 11 | 0 | 0 | 0.50 |
| s33 | 15 | 7 | 13 | 7 | 1.00 |
| **s34** | **0** | **0** | **0** | **0** | **0.00** |
| s35 | 17 | 11 | 17 | 8 | 1.00 |
| s36 | 11 | 11 | 17 | 12 | 1.00 |
| s37 | 14 | 5 | 11 | 6 | 1.00 |
| s38 | 10 | 4 | 15 | 7 | 1.00 |
| s39 | 10 | 10 | 12 | 11 | 1.00 |
| s40 | 11 | 10 | 14 | 6 | 1.00 |
| s41 | 10 | 10 | 17 | 10 | 1.00 |
| s42 | 16 | 6 | 8 | 9 | 1.00 |
| **s43** | **0** | **0** | **0** | **0** | **0.00** |
| s44 | 15 | 9 | 10 | 7 | 1.00 |
| **s45** | **0** | **0** | **0** | **0** | **0.00** |
| s46 | 0 | 0 | 12 | 11 | 1.00 |
| s47 | 13 | 9 | 13 | 10 | 1.00 |
| s48 | 16 | 10 | 13 | 4 | 1.00 |

Table 7.1 continued from previous page

| **Subject** | **C3 $\mu$ d-modes** | **C3 $\beta$ d-modes** | **C4 $\mu$ d-modes** | **C4 $\beta$ d-modes** | $\Gamma_{AR}$ |
|---|---|---|---|---|---|
| s49 | 4 | 7 | 7 | 5 | 1.00 |
| s50 | 7 | 9 | 9 | 8 | 1.00 |
| s51 | 15 | 12 | 17 | 9 | 1.00 |
| s52 | 12 | 10 | 12 | 5 | 1.00 |

The average $\gamma_{AR}$ for all 52 subjects was 0.80. Perfect separation of MI signals from RS signals was achieved for 38 (73.08%) subjects, each having a $\gamma_{AR}$ of 1. Interestingly, there were 38 subjects having a $\gamma_{AR}$ = 1, and 14 subjects that scored a $\gamma_{AR}$ < 1. This corresponds to the original authors who stated that within the GigaDB dataset, there are 38 discriminatory subjects and 14 they identified as non-discriminatory [69].

The average number of d-modes identified per subject was 8.9 ± 5.7 for C3-$\mu$, 7.0 ± 4.1 for C3-$\beta$, 10.1 ± 6.2 for C4-$\mu$, and 6.3 ± 3.9 for C4-$\beta$. In Table 7.1 it is shown that for any given channel if the $\mu$ component of the eigenchannel was successful, so was the $\beta$ component. Conversely, it never occurred that within one electrode either the $\mu$ or $\beta$ eigenchannel component succeeded independently of the other. This would imply a correlation in the performance of an eigenchannel regardless of selected frequency bands.

Selecting one good ($\gamma_{AR}$ = 1) and one semi-bad case ($\gamma_{AR}$ = 0.5) which are respectively s05 and s20 as shown in Figure 7.3. These example subjects are represented using a showcase output of SPLEEGA, the brainmap. Here, a top-down comprehensive view of the artifact removal process can be visualized. Subject 5 is one of the advanced clustering cases, whereas subject 20 was not. For subject 20, the C3 eigenchannel was successful at both bands; however, for the C4 channel, zero d-modes were procured, suggesting an alternative electrode would be optimal.

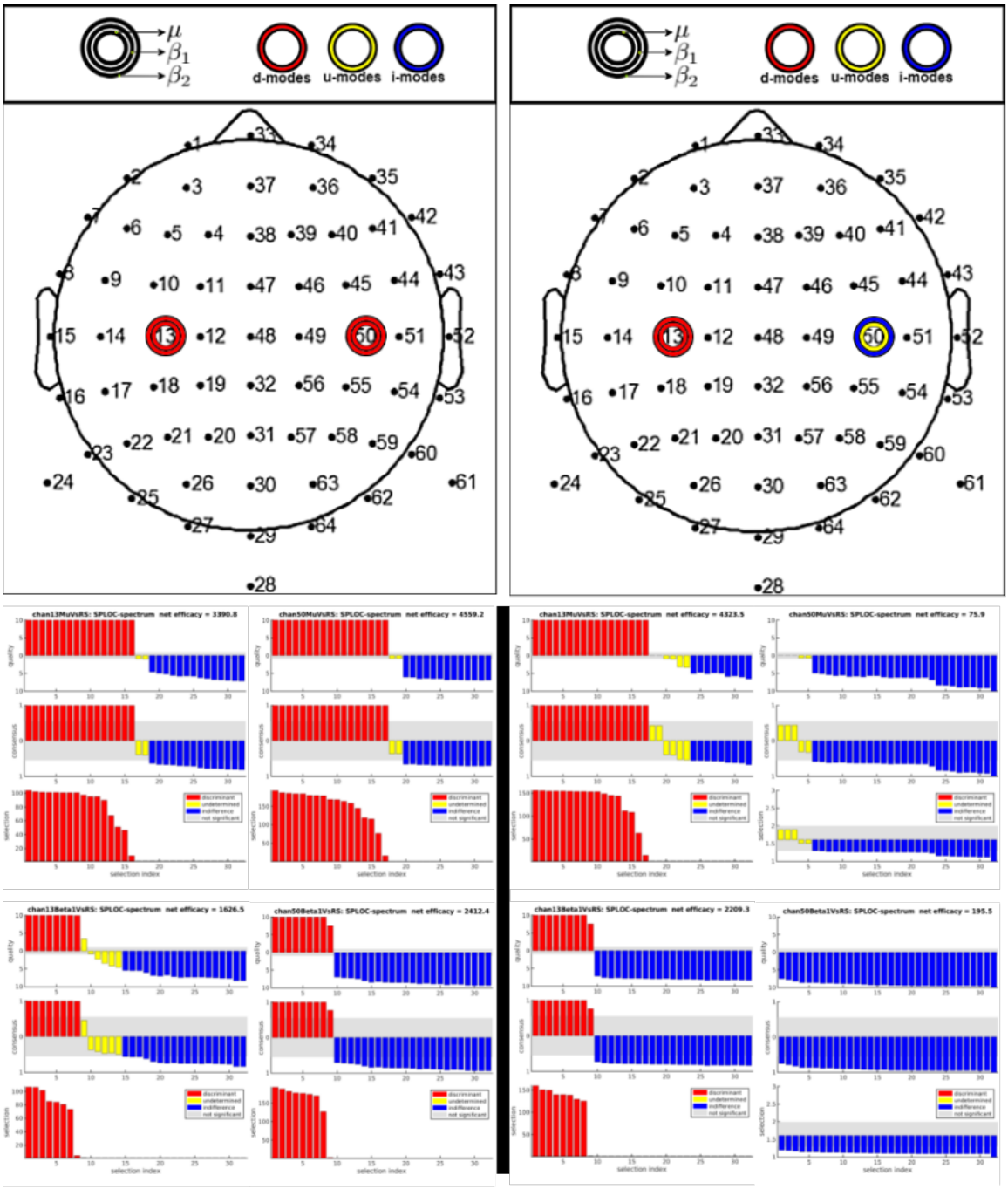


Figure 7.3: (left) The s05 SPLEEGA brainmap for the sparse-sensor configuration experiment. At least 1 d-mode that separates MI from RS is indicated by a red ring for the $\mu$ (inner ring) and $\beta$ (outer ring). The coloring scheme follows suit with the SPLOC basis vector spectra. When only u-modes and no d-modes are available, this is expressed as a yellow ring. (right) For s20, there were no d-modes identified for either band of C4 (channel number 14); furthermore, C4-$\beta$ was not able to develop an undetermined subspace spanned by u-modes. The vantage provided by the brainmap allows for quick interpretation of the artifact removal process. The basis vector spectra along the bottom are intentionally displayed at small scale for the explicit connection of the brainmap colored rings and the developed SPLOC subspaces.

## 7.2 Automatic Single-Channel Classification

As a preemptive measure to contextualize classification results, at the time of writing this manuscript the average accuracy over all subjects for state-of-the-art methods falls within 66-69% [30, 67, 69, 87]. Here, the number of MI trials used for eigenchannel construction was held fixed at 7, resulting in 14 MI trials. This was to achieve balance in the class distribution during training since there are 14 available RS trials at 2048 samples. Carrying over observations from the results in Section 7.1, subjects 11, 15, 34, and 45 contained zero discriminant C3 or C4 eigenchannels; therefore, each had a $\gamma_{AR}$ corresponding to 0. Additionally, 10 subjects (s08, s10, s14, s20, s22, s27, s28, s30, s32, and s46) only separated MI from the resting state in one of the two electrodes ($\gamma$ = 0.5) during the sparse-sensor experiment. Considering the per-subject variability of brain activity during MI tasks, the utility of sweeping over multiple frequency bands and channels is evident in Figure 7.4.

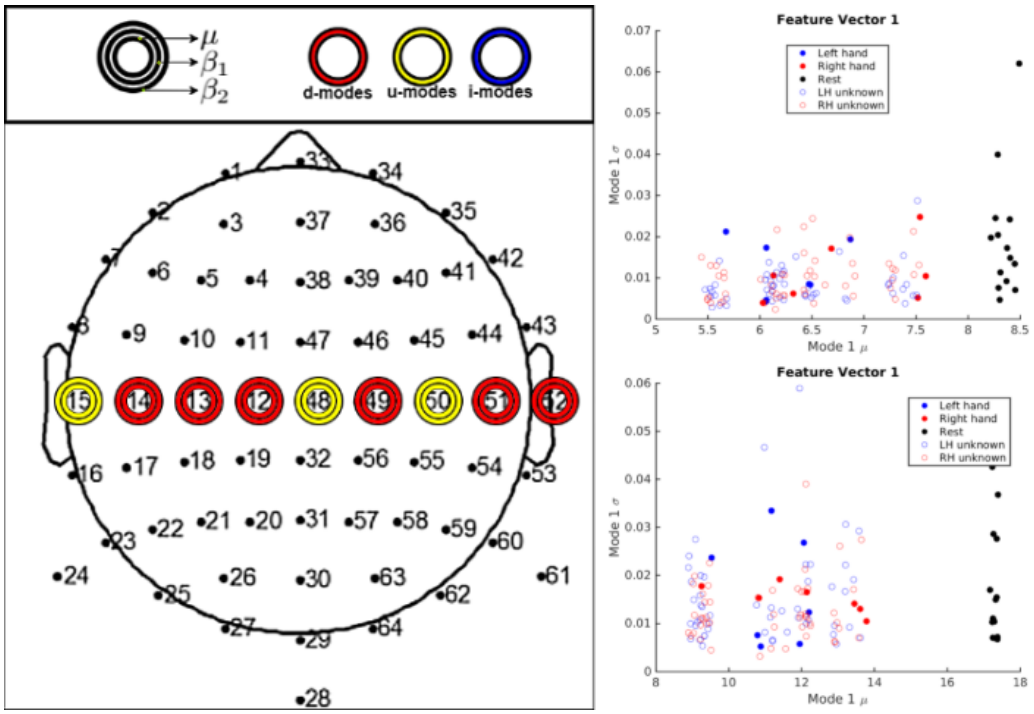


Figure 7.4: (left) The SPLEEGA brainmap for s11 using contralateral electrodes. As with the results from Section 7.1, channel C4-50 still does not attain discriminatory eigenchannels. Considering optimal frequency bands and the C5-14 (top right) and C6-51 (bottom right) electrodes, perfect separation of MI and RS is achieved.

During the automatic channel selection procedure, the 14 subjects that previously had not attained an $\gamma_{AR}$ of 1.0 have now reached successful artifact removal. Resulting in all 52 subjects containing at least two discriminatory electrodes having a $\gamma$ of 1.

Originally, the SPLEEGA eigenbrain was the conduit for final classification [88]; however, during experimentation it was observed that advanced clustering can be identified both within a single channel and within a single frequency band as well. This observation was explored in the automatic channel selection procedure, where classification was performed using these individual optimized channels. According to the neuroscience, the ERS/ERD spatio-temporal amplification/suppression in the $\mu$ and $\beta$ bands occurs between two spatially distinct regions of the brain. It was expected a single channel approach would not result in sufficient information for the identification of the standard MI underlying pattern. Through a 52-subject contralateral run, it is presented that much discriminatory information between LH and RH MI tasks exists within a subject-specific ideal electrode. The results from this contralateral experiment are presented in Table 7.2.

Table 7.2: Automatic contralateral run. Subject; Single-channel classification accuracy; Channel number; Channel name; Optimal frequency band; Artifact removal efficacy.

| **Subject** | **Accuracy** | **Channel #** | **Chan Name** | **Frequency** | **Efficacy** |
|---|---|---|---|---|---|
| s01 | 62% | 13 | C3 | 7.5-12.5 Hz | 5194.8 |
| s02 | 66% | 52 | T8 | 16-26 Hz | 3128.5 |
| s03 | 58% | 50 | C4 | 8-12 Hz | 4279.2 |
| s04 | 62% | 52 | T8 | 8-12 Hz | 6649.2 |
| **s05** | **88%** | **50** | **C4** | **8-12 Hz** | **3265.6** |
| **s06** | **76%** | **15** | **T7** | **16-26 Hz** | **1093.4** |
| s07 | 62% | 52 | T8 | 14-24 Hz | 3528.7 |
| s08 | 68% | 13 | C3 | 16-26 Hz | 2639.3 |
| s09 | 64% | 12 | C1 | 8-12 Hz | 1815.5 |
| s10 | 54% | 13 | C3 | 8-12 Hz | 1284.2 |

**Table 7.2 continued from previous page**

| Subject | Accuracy | Channel # | Chan Name | Frequency | Efficacy |
|---|---|---|---|---|---|
| s11 | 58% | 14 | C5 | 16-26 Hz | 2176.1 |
| s12 | 58% | 50 | C4 | 7.5-12.5 Hz | 5474.5 |
| **s13** | **74%** | **50** | **C4** | **7.5-12.5 Hz** | **2443.3** |
| s14 | 54% | 14 | C5 | 8-12 Hz | 2483.3 |
| s15 | 56% | 14 | C5 | 16-26 Hz | 2589.7 |
| s16 | 60% | 50 | C4 | 18-28 Hz | 1928.8 |
| s17 | 64% | 51 | C6 | 18-28 Hz | 3960.1 |
| s18 | 60% | 49 | C2 | 16-26 Hz | 3586.8 |
| s19 | 58% | 49 | C2 | 8-12 Hz | 3229.9 |
| s20 | 56% | 12 | C1 | 8-12 Hz | 3556.8 |
| s21 | 64% | 49 | C2 | 8-12 Hz | 5089.7 |
| s22 | 54% | 12 | C1 | 16-26 Hz | 3905.2 |
| s23 | 60% | 14 | C5 | 7.5-12.5 Hz | 1944.7 |
| s24 | 58% | 48 | Cz | 7.5-12.5 Hz | 4414.3 |
| s25 | 66% | 15 | T7 | 7.5-12.5 Hz | 4056.3 |
| **s26** | **90%** | **50** | **C4** | **8-12 Hz** | **3245.5** |
| s27 | 56% | 12 | C1 | 7.5-12.5 Hz | 4580.2 |
| s28 | 62% | 50 | C4 | 16-26 Hz | 1639.9 |
| s29 | 64% | 14 | C5 | 8-12 Hz | 5452.9 |
| s30 | 58% | 48 | Cz | 8-12 Hz | 3389.1 |
| **s31** | **70%** | **49** | **C2** | **14-24 Hz** | **3477.5** |
| s32 | 62% | 52 | T8 | 14-24 Hz | 4277.6 |
| s33 | 64% | 50 | C4 | 7.5-12.5 Hz | 3142.5 |
| s34 | 66% | 49 | Cz | 18-28 Hz | 5040.2 |

**Table 7.2 continued from previous page**

| Subject | Accuracy | Channel # | Chan Name | Frequency | Efficacy |
|---|---|---|---|---|---|
| s35 | 68% | 52 | T8 | 18-28 Hz | 4649.0 |
| **s36** | **72%** | **13** | **C3** | **16-26 Hz** | **1181.6** |
| s37 | 58% | 15 | **T7** | 8-12 Hz | 1294.5 |
| s38 | 62% | 48 | Cz | 7.5-12.5 Hz | 6524.4 |
| s39 | 64% | 52 | T8 | 14-24 Hz | 5282.5 |
| s40 | 62% | 52 | T8 | 14-24 Hz | 5282.5 |
| s41 | 60% | 50 | C4 | 8-12 Hz | 3835.8 |
| s42 | 62% | 12 | C1 | 14-24 Hz | 3060.5 |
| **s43** | **86%** | **14** | **C5** | **7.5-12.5 Hz** | **2065.4** |
| s44 | 66% | 52 | T8 | 16-26 Hz | 1455.5 |
| **s45** | **58%** | **12** | **C1** | **7.5-12.5 Hz** | **2534.3** |
| s46 | 66% | 14 | C5 | 16-26 Hz | 5793.7 |
| s47 | 60% | 14 | C5 | 7.5-12.5 Hz | 2891.1 |
| s48 | 66% | 13 | C3 | 14-24 Hz | 3217.2 |
| **s49** | **86%** | **51** | **C6** | **8-12 Hz** | **1879.9** |
| **s50** | **98%** | **50** | **C4** | **8-12 Hz** | **3385.4** |
| s51 | 56% | 15 | T7 | 7.5-12.5 Hz | 2563.2 |
| s52 | 60% | 13 | C3 | 8-12 Hz | 2310.8 |

The average classification accuracy was calculated as 64.65%. This directly falls near the current state-of-the-art methods for the same dataset. For any channel and frequency bands during optimization, if there was a tie in classification accuracy over multiple bands or channels, the selection was based on efficacy. Efficacy in the context of LH vs. RH MI classification is an indirect qualitative measure of how well RS was separated from MI signals. A notable trend observed during the automated contralateral procedure was the $\mu$

band defined as 7-13 Hz was never an optimal frequency range over all subjects.

Throughout algorithmic development, the most common trend with respect to clustering at the eigenchannel level was a conspicuous grouping of five (GRP-5). The GRP-5 pattern was common enough to necessitate further consideration of this underlying structure in the 2-dimensional slices of higher-dimensional space. Revealed in the original work was a possible explanation, the number of separate runs during recording sessions. The number of runs was exactly 5 within a 90 minute EEG recording session. An example case containing the GRP-5 pattern was reproduced with subject 43 as presented in Figure 7.5. First tracking the MI trials prior to random shuffling allows for the identification of recording runs separated by increments of 20 (in the case there are 100 MI trials available). Projecting along the top C3-$\mu$ eigenchannel d-mode and constructing the MFSP, an unexpected temporal trend emerges. From left to right in the example provided, there is a temporal progression respectively from run 1 to run 5. This is yet to be determined as a spatio-temporal pattern; however, the temporal structure is unavoidable. Verification of this GRP-5 pattern is required to determine whether this was a systematic error introduced during preprocessing of the SPLEEGA algorithm.

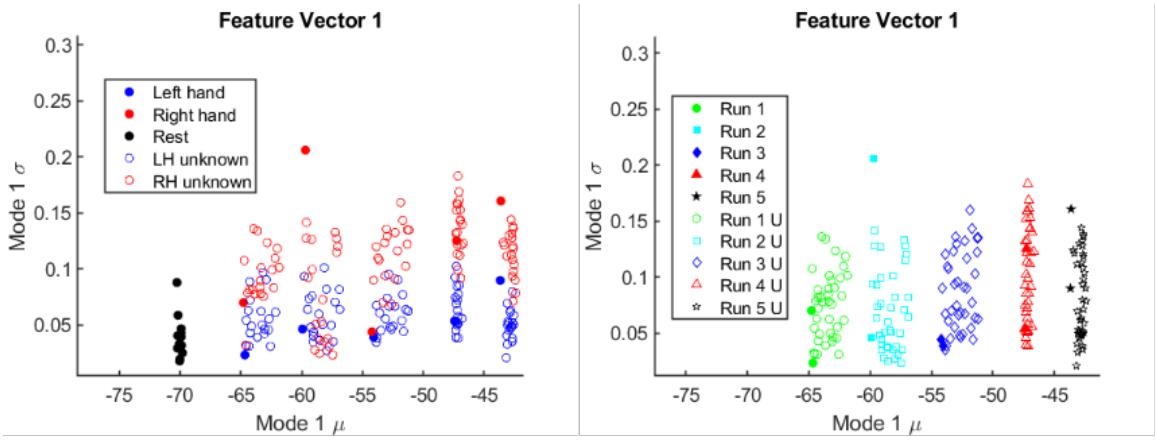


Figure 7.5: (left) The C3-$\mu$ eigenchannel for subject 43. (right) The GRP-5 pattern is uncovered to be a temporal pattern exhibiting a shift in the characteristics of signals correlating to the chronological order of the inter-session runs.

## 7.3 A Prelude to the Automatic Eigenbrain

In this section, a brief glimpse in the direction of future work is provided. The results from both the sparse-sensor and contralateral experiments would suggest that increasing the automatic procedure to include all 64 electrodes is a natural next step in development. As a proof-of-concept, a complete 64 electrode eigenbrain for subject 26 is introduced.

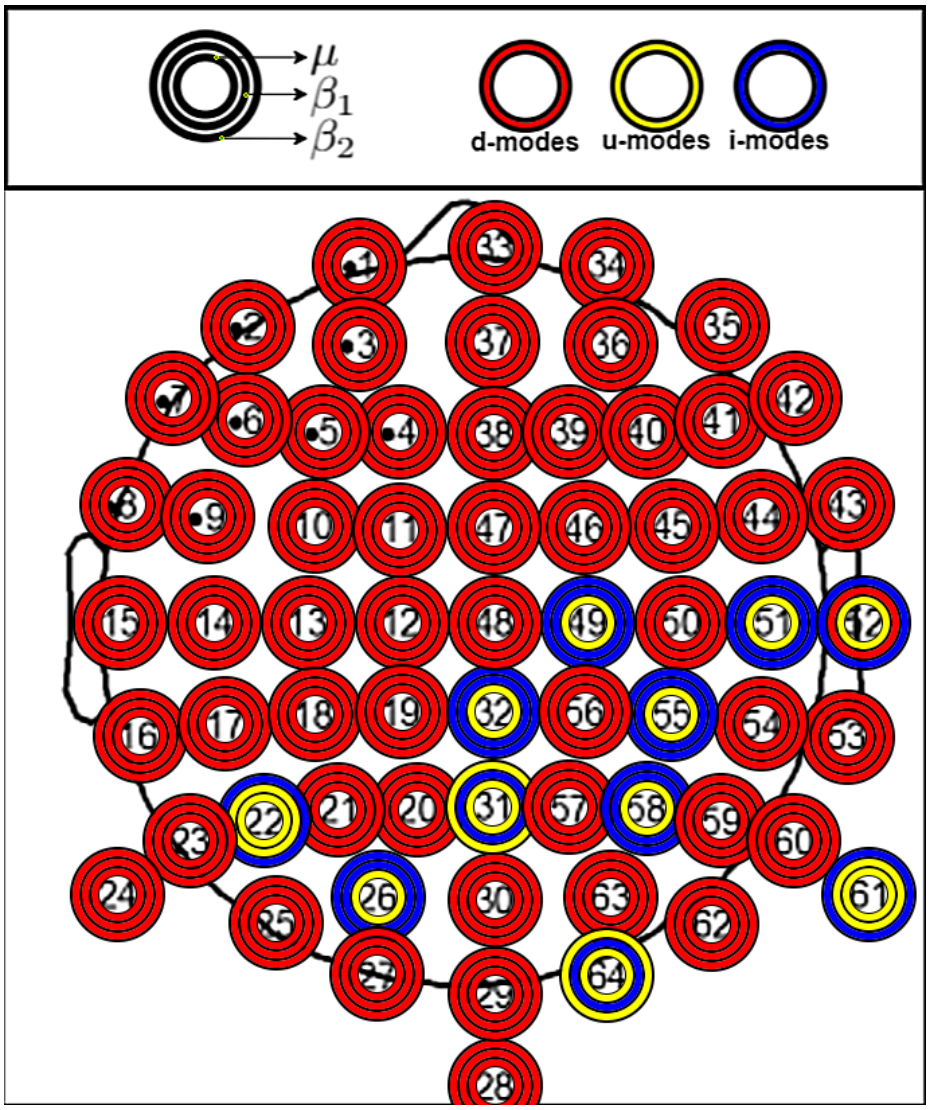


Figure 7.6: The SPLEEGA brainmap of s26 for all 64 electrodes. The automated channel selection procedure at this range of electrodes shows encouraging results. Allowing for the biometric uniqueness of each human's brain, this methodology could provide a more-informed manner of automatic channel selection within the context of discrimination.

Automatically scanning the 64-electrode montage allows for a more precise identification of advanced clustering. This expands the search space for advanced clustering in subjects where it was previously unseen. For the example provided in Figure 7.7, 5 channels were automatically selected as candidate electrodes. Only 1/5 were considered in the contralateral case (C4-50), and it could be well-argued it is the worst in terms of advanced clustering of the 5 identified. It is still quite the case that for all brains, there may exist an eigenbrain vector space inherent to each person that can distinguish control signals of interest from all other interfering signals.

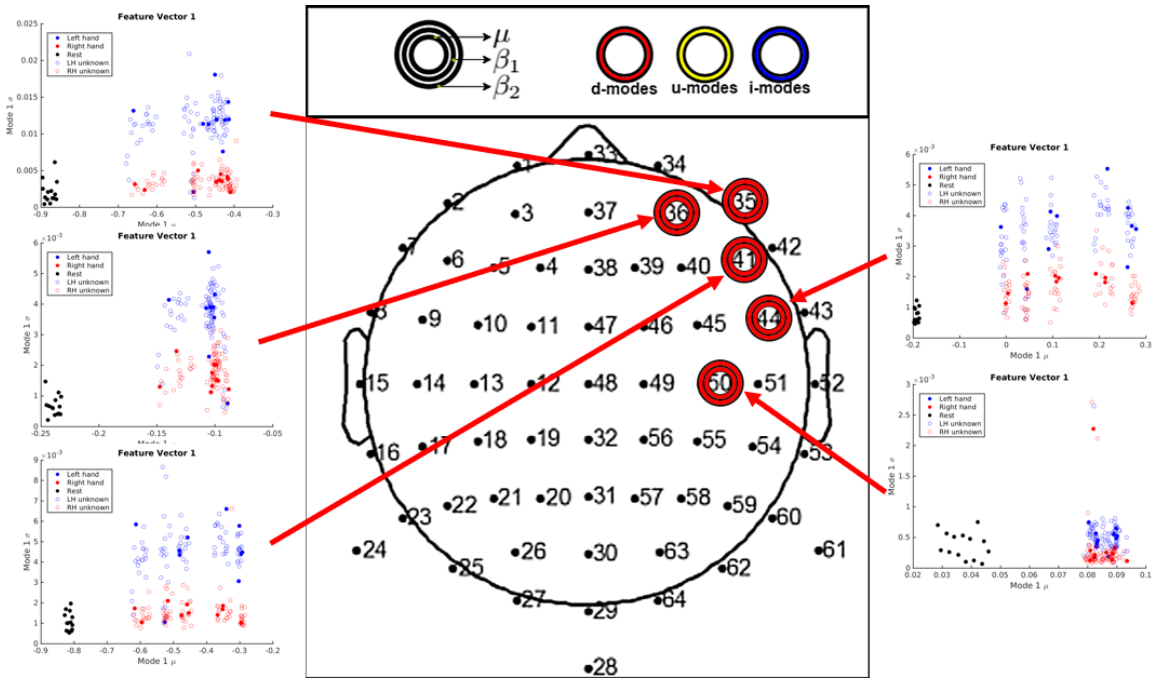


Figure 7.7: (center) The SPLEEGA brainmap of select electrodes for subject 26. Isolating the 5 best electrodes with strong advanced clustering allows for a more-informed manner of automatic channel selection. MFSPs are shown for channels: AF8-35-$\mu$ (top), AF4-36-$\beta$ (left middle), F6-41-$\mu$ (middle right), FC6-44-$\beta$ (bottom right), and C4-50-$\mu$ (bottom left).

# CHAPTER 8: CONCLUSIONS

An intelligent projection pursuit network was employed to construct a SPLEEGA eigenchannel for the blind source separation of MI signals. Across all 52 subjects, it was shown that complete separation of MI and resting state signals is achieved through an automatic channel-selection process. This was performed absent the removal of bad trials given an effective real-time BCI should be able to handle these types of statistical outliers. Furthermore, it was shown that a single electrode can provide sufficient discriminatory information for EEG MI classification.

The results from the contralateral experiment suggest this protocol is a promising route that automatically obtains the optimal electrode for each subject. Competitive classification accuracy was achieved during automatic single-channel classification with regards to the current state-of-the-art methods. A key component to this work has been the use of SPLOC which was recently developed as a general-purpose projection pursuit neural network. Subject-specific patterns of electrode discrimination have been successfully visualized through a SPLEEGA brainmap which acts to increase the discriminatory resolution of analysis as more electrodes are considered.

Multiple underlying patterns were recognized through the use of SPLEEGA. Notably, advanced clustering demands further analysis; in particular, when advanced clustering is observed it is seemingly independent of the $\mu$ and $\beta$ frequency bands. It will be explored whether advanced clustering can be induced for more subjects or perhaps this spatio-temporal pattern is inherent in the subjects themselves. Additionally, the temporal trend corresponding to the chronological progression of recording sessions will be further evaluated. This may provide explanatory power regarding unusually low performance metrics of recent approaches using the GigaDB dataset.

# REFERENCES


[1] A. A. Eshraghi, R. Nazarian, F. F. Telischi, S. M. Rajguru, E. Truy, and C. Gupta, "The cochlear implant: historical aspects and future prospects," *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, vol. 295, no. 11, pp. 1967–1980, 2012.

[2] C. T. Moritz, S. I. Perlmutter, and E. E. Fetz, "Direct control of paralysed muscles by cortical neurons," *Nature*, vol. 456, no. 7222, pp. 639–642, 2008.

[3] E. C. Leuthardt, G. Schalk, D. Moran, and J. G. Ojemann, "The emerging world of motor neuroprosthetics: a neurosurgical perspective," *NEUROSURGERY-BALTIMORE THEN HAGERSTOWN MD-*, vol. 59, no. 1, p. 1, 2006.

[4] P. Biswas and J. DV, "Eye gaze controlled mfd for military aviation," in *23rd International Conference on Intelligent User Interfaces*, pp. 79–89, 2018.

[5] L. Murthy, A. Mukhopadhyay, V. Yellheti, S. Arjun, P. Thomas, M. D. Babu, K. P. S. Saluja, D. JeevithaShree, and P. Biswas, "Evaluating accuracy of eye gaze controlled interface in military aviation environment," in *2020 IEEE Aerospace Conference*, pp. 1–12, IEEE, 2020.

[6] C. J. Walsh, K. Endo, and H. Herr, "A quasi-passive leg exoskeleton for load-carrying augmentation," *International Journal of Humanoid Robotics*, vol. 4, no. 03, pp. 487–506, 2007.

[7] J. K. Proud, D. T. Lai, K. L. Mudie, G. L. Carstairs, D. C. Billing, A. Garofolini, and R. K. Begg, "Exoskeleton application to military manual handling tasks," *Human Factors*, p. 0018720820957467, 2020.

[8] T. Grear, C. Avery, J. Patterson, and D. J. Jacobs, "Molecular function recognition by supervised projection pursuit machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–15, 2021.

[9] C. Avery, *Elucidating Dynamic Mechanisms for Extended Spectrum Antibiotic Resistance in Class a Beta-Lactamase Through Machine Learning on Molecular Dynamics Simulations*. PhD thesis, The University of North Carolina at Charlotte, 2019.

[10] F. Amato, F. Guignard, S. Robert, and M. Kanevski, "A novel framework for spatio-temporal prediction of environmental data using deep learning," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[11] I. Veselovskii, Q. Hu, P. Goloub, T. Podvin, M. Korenskiy, Y. Derimian, M. Legrand, and P. Castellanos, "Variability in lidar-derived particle properties over west africa due to changes in absorption: towards an understanding," *Atmospheric Chemistry and Physics*, vol. 20, no. 11, pp. 6563–6581, 2020.

[12] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. Chapman and Hall/CRC, 2019.

[13] A. Gevins, H. Leong, M. E. Smith, J. Le, and R. Du, "Mapping cognitive brain function with modern high-resolution electroencephalography," *Trends in neurosciences*, vol. 18, no. 10, pp. 429–436, 1995.

[14] G. Pfurtscheller and C. Andrew, "Event-related changes of band power and coherence: methodology and interpretation," *Journal of clinical neurophysiology*, vol. 16, no. 6, p. 512, 1999.

[15] T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems 8: Proceedings of the 1995 Conference*, vol. 8, p. 145, MIT press, 1996.

[16] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[17] B. J. Kemp, "Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off conditions.," *Developmental Psychology*, vol. 8, no. 2, p. 268, 1973.

[18] R. J. Kosinski, "A literature review on reaction time," *Clemson University*, vol. 10, no. 1, 2008.

[19] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, "Spatial and temporal resolutions of eeg: Is it really black and white? a scalp current density view," *International Journal of Psychophysiology*, vol. 97, no. 3, pp. 210–220, 2015.

[20] J. S. Bendat and A. G. Piersol, *Random data: analysis and measurement procedures*, vol. 729. John Wiley & Sons, 2011.

[21] J. Shaw, "An introduction to the coherence function and its use in eeg signal analysis," *Journal of medical engineering & technology*, vol. 5, no. 6, pp. 279–288, 1981.

[22] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, "Eeg coherency: I: statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalography and clinical neurophysiology*, vol. 103, no. 5, pp. 499–515, 1997.

[23] J. B. J. baron Fourier, *Théorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822.

[24] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning method for epileptic seizure detection using short-time fourier transform," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 213–222, 2017.

[25] S. K. Bashar and M. I. H. Bhuiyan, "Classification of motor imagery movements using multivariate empirical mode decomposition and short time fourier transform based hybrid method," *Engineering science and technology, an international journal*, vol. 19, no. 3, pp. 1457–1464, 2016.

[26] V. Morash, O. Bai, S. Furlani, P. Lin, and M. Hallett, "Classifying eeg signals preceding right hand, left hand, tongue, and right foot movements and motor imageries," *Clinical neurophysiology*, vol. 119, no. 11, pp. 2570–2578, 2008.

[27] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. Rao, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *Proceedings of the National Academy of Sciences*, vol. 107, no. 9, pp. 4430–4435, 2010.

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

[29] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[30] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: a review," *Neural Computing and Applications*, pp. 1–42, 2021.

[31] Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," *Nature Communications*, vol. 11, no. 1, pp. 1–11, 2020.

[32] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on computers*, vol. 100, no. 9, pp. 881–890, 1974.

[33] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American statistical Association*, vol. 76, no. 376, pp. 817–823, 1981.

[34] P. J. Huber, "Projection pursuit," *The annals of Statistics*, pp. 435–475, 1985.

[35] D. Cook, A. Buja, J. Cabrera, and C. Hurley, "Grand tour and projection pursuit," *Journal of Computational and Graphical Statistics*, vol. 4, no. 3, pp. 155–172, 1995.

[36] P. J. Bickel, G. Kur, and B. Nadler, "Projection pursuit in high dimensions," *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, pp. 9151–9156, 2018.

[37] A. Barcaru, "Supervised projection pursuit–a dimensionality reduction technique optimized for probabilistic classification," *Chemometrics and Intelligent Laboratory Systems*, vol. 194, p. 103867, 2019.

[38] C. Croux, P. Filzmoser, and M. R. Oliveira, "Algorithms for projection–pursuit robust principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 87, no. 2, pp. 218–225, 2007.

[39] M. Hubert and K. V. Branden, "Robust methods for partial least squares regression," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 10, pp. 537–549, 2003.

[40] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[41] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background eeg," *Brain topography*, vol. 2, no. 4, pp. 275–284, 1990.

[42] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[43] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[44] V. Singh, K. Veer, R. Sharma, and S. Kumar, "Comparative study of fir and iir filters for the removal of 50 hz noise from eeg signal," *International Journal of Biomedical Engineering and Technology*, vol. 22, no. 3, pp. 250–257, 2016.

[45] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[46] P. Jahankhani, V. Kodogiannis, and K. Revett, "Eeg signal classification using wavelet feature extraction and neural networks," in *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, pp. 120–124, IEEE, 2006.

[47] A. Y. Kaplan, A. A. Fingelkurts, A. A. Fingelkurts, S. V. Borisov, and B. S. Darkhovsky, "Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges," *Signal processing*, vol. 85, no. 11, pp. 2190–2212, 2005.

[48] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, IEEE, 2008.

[49] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.

[50] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in neuroscience*, vol. 6, p. 39, 2012.

[51] S. Liang, K.-S. Choi, J. Qin, W.-M. Pang, Q. Wang, and P.-A. Heng, "Improving the discrimination of hand motor imagery via virtual reality based visual guidance," *Computer methods and programs in biomedicine*, vol. 132, pp. 63–74, 2016.

[52] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *Journal of neuroscience methods*, vol. 255, pp. 85–91, 2015.

[53] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse bayesian classification of eeg for brain–computer interface," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2256–2267, 2015.

[54] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery bci systems," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 15–29, 2015.

[55] S. Kumar and A. Sharma, "A new parameter tuning approach for enhanced motor imagery eeg signal classification," *Medical & biological engineering & computing*, vol. 56, no. 10, pp. 1861–1874, 2018.

[56] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented csp features for multi-class motor imagery of eeg signals classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2620–2623, IEEE, 2015.

[57] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of eeg motor imagery signals," *Journal of neural engineering*, vol. 14, no. 1, p. 016003, 2016.

[58] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[59] Z. Tang, C. Li, and S. Sun, "Single-trial eeg classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, 2017.

[60] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[61] L. Liu, "Recognition and analysis of motor imagery eeg signal based on improved bp neural network," *IEEE Access*, vol. 7, pp. 47794–47803, 2019.

[62] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan, and J. R. Wolpaw, "Brain-computer interface (bci) operation: signal and noise during early training sessions," *Clinical Neurophysiology*, vol. 116, no. 1, pp. 56–62, 2005.

[63] S. Leske and S. S. Dalal, "Reducing power line noise in eeg and meg data via spectrum interpolation," *NeuroImage*, vol. 189, pp. 763–776, 2019.

[64] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[65] T.-j. Luo, F. Chao, *et al.*, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC bioinformatics*, vol. 19, no. 1, p. 344, 2018.

[66] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent eeg signal analysis," *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 715–719, 2019.

[67] S. Kumar, A. Sharma, and T. Tsunoda, "Brain wave classification using long short-term memory network based optical predictor," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[68] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from eeg using a deep attention-based lstm network," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3113–3122, 2019.

[69] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "Eeg datasets for motor imagery brain–computer interface," *GigaScience*, vol. 6, no. 7, p. gix034, 2017.

[70] Y.-B. Wang, Z.-H. You, L.-P. Li, Y.-A. Huang, and H.-C. Yi, "Detection of interactions between proteins by using legendre moments descriptor to extract discriminatory information embedded in pssm," *Molecules*, vol. 22, no. 8, p. 1366, 2017.

[71] A. Dehzangi, K. Paliwal, J. Lyons, A. Sharma, and A. Sattar, "Exploring potential discriminatory information embedded in pssm to enhance protein structural class prediction accuracy," in *IAPR International Conference on Pattern Recognition in Bioinformatics*, pp. 208–219, Springer, 2013.

[72] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from eeg signals," *Sensors*, vol. 19, no. 1, p. 210, 2019.

[73] M. Mahmood, D. Mzurikwao, Y.-S. Kim, Y. Lee, S. Mishra, R. Herbert, A. Duarte, C. S. Ang, and W.-H. Yeo, "Fully portable and wireless universal brain–machine interfaces enabled by flexible scalp electronics and deep learning algorithm," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 412–422, 2019.

[74] T. M. Vaughan, L. A. Miner, D. J. McFarland, and J. R. Wolpaw, "Eeg-based communication: analysis of concurrent emg activity," *Electroencephalography and clinical Neurophysiology*, vol. 107, no. 6, pp. 428–433, 1998.

[75] Y. Qin, P. Xu, and D. Yao, "A comparative study of different references for eeg default mode network: the use of the infinity reference," *Clinical neurophysiology*, vol. 121, no. 12, pp. 1981–1991, 2010.

[76] S. Sarkar and K. L. Boyer, "On optimal infinite impulse response edge detection filters," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 11, pp. 1154–1171, 1991.

[77] D. A. Lyon, "The discrete fourier transform, part 4: spectral leakage," *Journal of object technology*, vol. 8, no. 7, 2009.

[78] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, 1999.

[79] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[80] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.

[81] T. Fletcher, "Support vector machines explained," *Tutorial paper*, pp. 1–19, 2009.

[82] M. Hofmann, "Support vector machines-kernels and the kernel trick," *Notes*, vol. 26, no. 3, pp. 1–16, 2006.

[83] O. P. Agrawal, "Formulation of euler–lagrange equations for fractional variational problems," *Journal of Mathematical Analysis and Applications*, vol. 272, no. 1, pp. 368–379, 2002.

[84] J. A. Clarkson, "Uniformly convex spaces," *Transactions of the American Mathematical Society*, vol. 40, no. 3, pp. 396–414, 1936.

[85] K. Kawaguchi, "Deep learning without poor local minima," *arXiv preprint arXiv:1605.07110*, 2016.

[86] I. Steinwart and C. Scovel, "Mercerâs theorem on general domains: On the interaction between measures, kernels, and rkhss," *Constructive Approximation*, vol. 35, no. 3, pp. 363–417, 2012.

[87] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in eeg decoding with deep learning," *Frontiers in human neuroscience*, vol. 14, p. 103, 2020.

[88] T. Grear and D. Jacobs, "Classifying eeg motor imagery signals using supervised projection pursuit for artefact removal," *Proceedings of IEEE Systems Man and Cybernetics Conference (in print)*, 2021.