CHAPTER I
INTRODUCTION

1.1 Background

Nowadays, electroencephalography (EEG) signals with plenty of physiological and pathological information are getting more attention with the rapid development of cognitive neuroscience. Over the last few years, the Brain-Computer Interface (BCI) has become a technology that has caught the attention of researchers. This is because BCI is a technology that can immediately read and translate thoughts in the brain and then can be applied to computer commands. A BCI system provides an alternative pathway between human brain and external devices [1]. This process starts by recording the brain activity and continues through the signal processing part to detect the user's intent. Then an appropriate signal is sent to the external device to control the device according to the detected signal [2]. Among different brain activity monitoring modalities, EEG technique provides an easy and inexpensive solution for BCI systems and is used in many non-invasive BCI studies [3]. But EEG signals belong to non-stationary random signals without ergodic states, and their background noise is very strong, so the analysis and processing of EEG signals have always been a very attractive, but difficult problem.

Motor imagery (MI) based BCI is one of the standard concepts of BCI, systems is based on the recording and classification of circumscribed and transient EEG changes in association with the imagination of different types of movements. Motor imagery can modify the neuronal activity in the primary sensorimotor areas in a very similar way as observable with a real executed movement, so as a result it can serve to generate selfinduced variations of the EEG [4]. In the medical domain, MI EEG classification is integral to neurological rehabilitation, particularly for stroke recovery. By providing real-time feedback, it helps patients retrain their brains and restore motor functions via neuroplasticity. Furthermore, it aids in understanding the neural mechanisms underlying motor planning and execution, contributing to the broader field of neuroscience. In fact, in both imaginary and actual movements (or motor execution, ME), the neural activation of event-related EEG in the Mu rhythm (8−12 Hz) of the motor cortex was observed as being functionally similar. Although there was a difference in the intensity of brain activities during MI and ME, in a previous study that MEG, the event-related synchronization/desynchronization (ERS/ERD) of beta (15−30 Hz) in the contralateral motor cortex and somatosensory cortex has been confirmed commonly [5].

Common spatial pattern (CSP) is a very efficient method and has been mostly applied to MI feature extraction. CSP is a popular method among MI studies for extracting features [2]. CSP has been applied successfully in many MI task recognition studies. Since CSP finds spatial filters to maximize the variance of the projected signal from one class while minimizing it for another class, it provides a natural approach to effectively estimate the discriminant information of MI [6]. The subject-specific band-pass filtering of the EEG prior to spatial filtering influences the effectiveness of the CSP algorithm. Hence the Filter Bank Common Spatial Pattern (FBCSP) algorithm was proposed to effectively select appropriate subject-specific frequency band-pass filtering for the CSP algorithm [7]. So far, numerous research efforts have been dedicated to developing variants of CSP towards further improved EEG pattern decoding accuracy. By exploiting the mutual information between the extracted CSP features, a filter bank CSP (FBCSP) was designed to optimize CSP features from multiple filter bands. By incorporating Fisher's ratio, a discriminant extension of FBCSP (DFBCSP) was further proposed to select the most discriminative filter bands, and hence enhance the separability of CSP features between classes. A sparsityconstrained filter band common spatial pattern (SFBCSP) was also developed to automatically explore a compact combination of multi-band CSP features via a sparse learning strategy for improved pattern separability [8].

For the classification learning, the good learning samples are the key to train the classifier. The irrelevant or redundant information of the samples will increase the complexity of the classification algorithm as well as the time of training and prediction, and directly affect the classification performance [9]. Its purpose is to simplify the data structure, interpret the data information, and improve the robustness, stability and identification of the model.

Therefore, it is necessary to select the optimal features that can best represent the characteristic of object, and eliminate irrelevant and redundant features to improve the efficiency of the classifier [10]. Feature selection is the process of selecting the optimal feature subset from the raw feature set to reduce the dimension of the feature space. the maximal relevance and minimal redundancy (mRMR) criterion is a filter feature metric for pattern recognition [11], it aims to select the feature subset with the maximal dependency, maximal relevance and the minimal redundancy from the raw feature set.

Maldonado [12] proposed an embedded feature selection strategy that penalizes feature set cardinality using scaling factors. This approach was applied to two SVM formulations specifically designed for class-imbalanced problems, achieving superior predictive performance compared to established feature selection methods. Then Insik [13] introduced an improved mRMR feature selection method by employing Pearson's correlation coefficient to measure redundancy and the R-value to evaluate relevance. This enhancement demonstrated higher classification accuracy and reduced computational time across various high-dimensional datasets compared to the original mRMR method. However, mRMR only considered the contribution of a single feature. The feature with the maximum contribution from the raw feature set are selected and added to form the optimal feature subset. The existing enhanced methods for mRMR are primarily based on calculating the correlation and redundancy of individual features. But researcher ignore the joint contribution of multiple features, and do not consider the calculation of redundancy and correlation between feature subsets.

This study makes significant contributions to the field of EEG analysis, particularly in motor imagery tasks, by focusing on improving feature selection to enhance classification accuracy and data processing efficiency. The primary contribution of this work is the development of the Equal Grouping Method (EGM), a novel approach that helps reduce feature redundancy by grouping similar features based on their information distribution. This method ensures that only the most relevant features are selected, thus optimizing the feature subset, improving the efficiency of data handling, and speeding up the classification process without compromising the quality of the results. Another key contribution is the application of the Pearson correlation coefficient to assess and manage feature redundancy. By measuring the correlation between features, this study can identify and remove highly correlated features that do not add significant value to the classification task. This helps in simplifying the model, improving its speed, and preventing overfitting, while still maintaining high classification accuracy.

Based on introduction, this study applied an advanced machine learning classification technique (random forest and SVM) to categorize these features after performing feature selection on these datasets by improve mRMR using EGM approach as feature subset and layering with pearson correlation. To ascertain the efficacy of the suggested approach, we carry out in-depth tests and assess using a variety of performance measures (accuracy, dimensionality reduction rate, comprehensive rate, precision, recall, and F-measure). Tests demonstrate that the suggested feature selection approach on random forest and svm classifier performs better than the traditional mRMR.

1.2 Problem Identification

The key challenge is effectively selecting and enhancing discriminative features for EEG classification. Methodologies involve Feature Bank Common Spatial Pattern

(FBCSP) for feature extraction from EEG data across multiple frequency bands. While feature extraction has large dimensional data from each channel sub bands, feature selection can help filter relevant features by dividing and optimizing candidate feature subsets remains complex. Balancing these processes to improve classification accuracy is a significant issue. Also, traditional mRMR still running on single feature calculation and ignore the joint contribution of multiple features.

1.3 Objective

This study proposed a robust and efficient method for feature extraction and feature selection in EEG classification, specifically focusing on enhancing the discriminative power of features using FBCSP and improved feature selection mRMR. The objectives are limited to the following.

1. Increase the discriminative power of extracted features using FBCSP. Capture more detailed frequency-specific information by employing a greater number of sub-bands.
2. Reduce feature dimensionality while keeping only the most relevant and effective features. Then improve the classification performance for motor imagery tasks before and after the model enhancement
3. Comparing and validating the proposed method's performance with results from previous relevant studies.

1.4 Scope of Work

To strengthen the points of concern, the various problems solved in this research are limited to the following:

1. Dataset using from BCI Competition IV IIb with 9 subject train data.
2. The data includes clearly labeled motor imagery tasks. By selected only label left and right class, then ignore another label.
3. Include artifacts such as eye movement (EOG) signals on processing.
4. Apply band-pass filtering to remove noise and isolate relevant frequency bands 8 – 40hz with overlapping 2hz.
5. Measure performance metrics including dimensional reduce rate, accuracy, precision, recall, and F1-score.

1.5 Research Method

Below is the comprehensive, step-by-step methodology, beginning with an extensive review of related literature.

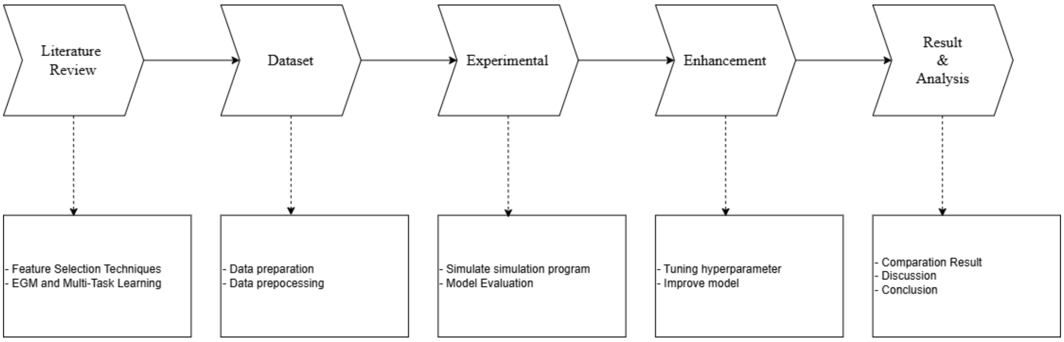

Figure 1. 1 Research framework

Figure 1.1 describe the research process with a thorough literature review to understand the existing methods and techniques related to feature selection, machine learning classification, and the application of mRMR and EGM-based approaches. Then next step relevant datasets are collected, potentially from publicly available sources. The data should be diverse and cover a broad range of instances to ensure a representative dataset for analysis. The collected datasets consist of numerous features that may include raw signal data, derived statistical features, or other relevant indicators. Ensuring data diversity and quality at this stage is crucial for robust model development. The raw data collected is often noisy and may contain missing or irrelevant information. Experiment process starts from scenario 1, the focus is on selecting the most relevant features from the preprocessed dataset. The proposed method uses mRMR combined with an EGM-based adding pearson correlation. All experimentals will running on related simulation task as we need. Then enhanced feature selection step improves upon traditional methods by combining initial mRMR with EGM approches and pearson correlation based on calculation multiple features. The synergy of these techniques addresses the challenges of high-dimensional, noisy datasets, resulting in a more efficient and effective process that selects a compact set of highly relevant features for classification. This enhancement is crucial for achieving better overall model performance. After collecting the result, then evaluating the improvement in accuracy, precision, recall, and F-measure based on experimental and analyzing the strengths and potential weaknesses of the proposed approach, as well as its impact on the overall model performance, drawing conclusions based on the results, highlighting the contribution of the research, and suggesting possible directions for future studies to further enhance feature selection and classification techniques.

1.6 Hypothesis

Based on the challenges identified in the study, we hypothesize that the proposed feature selection method, utilizing mRMR combined with the EGM approach for multi-task learning, will significantly enhance the classification accuracy of EEG datasets compared to traditional mRMR methods. We believe that applying Random Forest and SVM classifiers following this feature selection process will yield higher precision, recall, and F-measure values in EEG classification, as opposed to using these classifiers without feature selection. Furthermore, we hypothesize that the multi-task learning approach in feature extraction will improve the model's ability to generalize spatial patterns across tasks, surpassing the performance of conventional CSP-based methods. Additionally, the proposed feature extraction and selection technique is expected to achieve a higher dimensionality reduction rate without compromising the comprehensive classification rate, thus providing a more efficient analysis of EEG data. Ultimately, we anticipate that integrating FBCSP with mRMR and feature selection will improve by adding the EGM method and pearson correlation algorithm will strike a better balance between enhancing discriminative power and maintaining the generalization of features, leading to a significant improvement in the overall performance of EEG-based classification systems.