A Novel BCI System Based on Hybrid Features for Classifying Motor Imagery Tasks

Jaidaa Abougharbia
Electronics and Communications Department
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt
jaidaaahany@hotmail.com

Omenya Attallah
Electronics and Communications Department
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt
o.attallah@aast.edu

Mohamed Tamazin
Electronics and Communications Department
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt
tamazin@aast.edu

A.A.A.Nasser
Electronics and Communications Department
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt
menem_1954@yahoo.com

Abstract—Brain-Computer Interface (BCI) is a way to control external devices based on Electroencephalography (EEG) signals. One of the most critical problems facing BCI is realizing high Classification Accuracy (CA) for Motor Imagery (MI) mental tasks. A novel study is proposed which aims to achieve a reliable CA. In this study, three sets of features were extracted in time, time-frequency, and the time and time-frequency domains. Several Support Vector Machine (SVM) classifiers were constructed with different electrode sets to determine the channels which improve the CA. The publicly available dataset BCI competition III datasets Iva was used in this study. The results showed that the proposed method is one among the few, which focuses on achieving higher classification accuracy depending on features from different domains. The highest mean CA of 91.72% was achieved using the hybrid feature set extracted from the time and time-frequency domains. The mean CA of the proposed outperformed other several recent related works. Therefore, the proposed can be used successfully to control wheelchairs and rehabilitation therapies.

Keywords—Brain-Computer Interface, Electroencephalography signal, motor imagery, support vector machines.

I. INTRODUCTION

EEG is a method used to measure the electrical activity of the brain. This activity is generated by billions of neurons and they are measured using scalp electrodes [1]. A BCI is a way of communicating between brain neurons, and external software or hardware such as a computer, robotic limbs or a wheelchair which helped paralyzed and Amyotrophic Lateral Sclerosis patients to move or even to talk [2].

Motor Imagery (MI) is a mental activity where an individual would imagine the movement of his limbs without any actual muscular motion [3]. Efficient feature extraction and classification methods are the two significant steps to translate the brain activities into MI tasks.

Feature extraction is the process of extracting significant features from signals capable of differentiating between different classes of a dataset. The Common Spatial Pattern (CSP), feature extraction method, is commonly used for MI tasks. S. Selim et al. [4] used the CSP approach with a window time interval of (0.5s - 3.5s). The authors also proposed a novel method for channels selection using CSP. In [5], the CSP approach was applied for the training trials before assigning a score to each channel based on L1 norm scores. For feature extraction, electrodes having higher scores are reserved for further CSP processing.

Other feature extraction methods were used as well in MI tasks. J. Kevric et al. [6] extracted High Order Statistics (HOS) features from three different decomposition techniques namely; Empirical Mode Decomposition (EMD), Discrete Wavelet Transform (DWT), and Wavelet Packet Decomposition (WPD). Using WPD with the popular K-nearest neighbor (k-NN) classifier gave the best results. The overall classification accuracy was around 92.8%. On the other hand, in [7], Miao et al. used spatial-frequency-temporal patterns for feature extraction. The highest accuracy reached was 86.38%. Singh et al. [8] applied symmetric positive definite (SPD) to extract features using the prior information of channels. Covariance matrices of EEG trial lying in SPD matrices hold spatial information for EEG trial and can directly be used for classification. Computational time was minimized using this technique. In [9], S. Selim et al. employed the Root Mean Square (RMS) method to extract features. They then filtered the dataset through a bandpass filter from 10 to 30 Hz to capture the bands which contain the MI frequency band. The authors compared the performance of the proposed system with different channels sets. For classification, Linear Discriminant Analysis (LDA) classifier and Least Square-Support Vector Machine (L-SVM) classifier were utilized. Joadder et al. [10] suggested decreasing the number of electrodes to reduce the computational cost and improve the accuracy of the classification models. The authors considered only the three channels C3, Cz and C4. The overall classification accuracy achieved was 76% using an SVM classifier.

Our study introduces a novel hybrid features which combine features in different domains to overcome the limitations of an EEG being a non-linear, non-stationary signal. The WPD is used to decompose signal in time-frequency domains. Afterwards, some statistical and HOS features are extracted from each sub-band. Time features are also extracted. The SVM classifier is used as a classification model. The publicly available BCI competition III dataset IVa is used to construct the proposed BCI system for classifying mental MI tasks.

The rest of the paper is organized as follows: Section II explains the system architecture and the steps of the proposed method. Section III describes the dataset and used parameters. Results
are discussed and compared with recent related works in section IV. Section V concludes our study.

II. PROPOSED METHOD
The architecture of the proposed BCI system is shown in Fig.1. The first block is the signal acquisition, where the dataset is loaded from each subject. Then, the preprocessing stage to remove the unwanted artifacts from the EEG signals. Afterwards, the feature extraction stage is performed to extract useful features. Finally, the features are classified using a suitable classifier.

A. Signal acquisition
Signal acquisition is the stage of collecting brain signals. There are two types of signal acquisition: invasive and noninvasive.

Although invasive type produces a high-quality signal, the non-invasive type is preferred [11]. The non-invasive signal acquisition can be safely applied to anyone without a lot of effort. Subjects were asked to relax. Signals from 118 channels and of the extended 10/20-system were captured.

B. Preprocessing
Preprocessing, it is a critical step to convert EEG data into a format that is more suitable for analysis and more processing. Preprocessing always refers to artifact's reduction of the raw data. Artifacts removal is necessary for EEG since it is a low voltage signal contaminated with noise. In this study, the preprocessing procedure is divided into three following stages [12]:

- Normalization
The literature showed that the normalization step improves the performance of BCI systems. It is used to get rid of the difference in power levels between the signals due to the position of the electrodes [13]. It produces EEG signals with unit variance and zero means [14].

\[ z_i = \frac{w_i - \bar{w}}{\delta} \]  

where \( z_i \) is the normalized signal, \( w_i \) is the original signal, \( \bar{w} \) and \( \delta \) are the mean and standard deviation, respectively.

- Notch filter
EEG recordings contain interference from the power lines. A Notch filter can be considered as a band-stop filter with a very narrow band to reject the 60 Hz frequency from power lines [15].

- Windowing
Every movement in the dataset was done during the duration of 3.5s; windowing is needed to take only the samples within this period only. As stated in [5], different window sizes were tested, but an optimal window was found to be from 0.5 to 3.5s.

C. Features extraction

- Feature extraction in the time-frequency domain
The EEG signal is non-stationary, so it is required to analyze it in time and frequency domains. Wavelet decomposition techniques are suitable for the non-linear EEG signal; as they decompose signals both in time and frequency domains. They generate a set of sub-bands by translation and dilation of the mother wavelet function. The DWT decomposes each level into approximate and detailed coefficients. Only the approximation components are decomposed using a low pass filter. On the other hand, the WPD decomposes the signal into approximate and detailed coefficients at each level using low and high pass filters as shown in Fig. 2. This gives higher resolution and more information about the EEG dataset [16]. In this study, the WPD is applied to each signal. Afterwards, four statistical features are extracted using equation (2-5). Also, two HOS features are also extracted using equations (6-7).

a. The absolute mean of coefficients in each sub-band

\[ \mu = \frac{1}{N} \sum_{i=1}^{N} |x_i| \]  

Figure 1 BCI system architecture.
b. The average power of coefficients in each sub-band

\[ P_{av} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \]  

(3)

c. The standard deviation of the coefficients in each sub-band

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

(4)

d. The ratio of the absolute mean values of coefficients (signal values) of adjacent sub-bands

\[ \gamma = \frac{1}{N} \frac{\sum_{i=1}^{N} |x_i|}{\sum_{i=1}^{N} |z_i|} \]  

(5)

e. The skewness of the coefficients (signal) in each sub-band

\[ S = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu)^3}{\sigma^3} \]  

(6)

f. The kurtosis of the coefficients (signal) in each sub-band

\[ K = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu)^4}{\sigma^4} \]  

(7)

where \( N \) is the number of samples (length of cue) in each sub-band, and \( X\{x_1, x_2, \ldots, x_N\} \) and \( Z\{z_1, z_2, \ldots, z_N\} \) are two adjacent sub-bands after WPD.

- **Feature extraction in the time domain**

Time domain features are directly extracted from the preprocessed EEG signal, which makes them easy to implement. Time domain features assume data as stationary time signals [17]. In this study, seven features are extracted in the time domain.

RMS is a commonly used time domain feature. It gives information about the amplitude of the signal and the frequency independence [10].

\[ RMS = \sqrt{\frac{1}{M} \sum_{j=1}^{M} y_j^2} \]  

(8)

The entropy is a measure of the disorder in a physical system. Shannon extended this concept to the field of information theory [18]. Afterwards, an improvement of the Shannon [18] entropy is introduced by Renyi [19] by adding a new parameter 'q'.

\[ \text{Renyi} = \log \left( \frac{\sum_{j=1}^{M} y_j^q}{1 - q} \right) \]  

(9)

**Figure 2** WPD for scale level 2.

The Hjorth parameter [1] has three kinds of parameters: activity, mobility, and complexity.

\[ \text{Activity} = \text{var}(y_j) \]  

(10)

\[ \text{Mobility} = \frac{\text{var}(y_j')}{\text{var}(y_j)} \]  

(11)

\[ \text{Complexity} = \frac{\text{mobility}(y_j')}{\text{mobility}(y_j)} \]  

(12)

The waveform length

\[ WL = \sum_{j} |y_j - y_{j-1}| \]  

(13)

The mean absolute value

\[ \mu = \frac{1}{M} \sum_{j=1}^{M} |y_j| \]  

(14)

where \( M \) is the number of samples in each sub-band, and \( Y\{y_1, y_2, \ldots, y_N\} \) are the samples in the time domain.

In this research, three experiments were explored using the following selected feature sets: Time domain features (Experiment 1), time-frequency domain features (Experiment 2), and time-frequency domains features (Experiment 3).

**D. Classification**

SVM classifier idea is to separate the data from two classes by finding a weight vector and maximizing the geometric margin. Because it is not always possible to separate two classes correctly, the SVM will search for the hyperplane with the highest margin and lowest training error. The data points closest to the hyperplane are called the support vectors. In our study specifically, the linear SVM classifier is applied, which is often applied to linearly separable problems [20].

**III. EXPERIMENTAL SETUP**

**A. Dataset description**

The Dataset IVa BCI competition III [21] consists of five healthy subjects sitting in a comfy chair with 118 EEG channels.
The number of visual cues subject is 280 executed for 3.5s. These MI cues are (R) right hand, or (F) foot. Short breaks of around 2s are introduced. The EEG signals acquired were then band-pass filtered with cutoff frequencies of 0.05 and 200 Hz. Afterwards, they were digitized at 1000 Hz and downsampled to 100 Hz. In this work, the the downsampled data is used. The subjects have an unequal number of training and testing trials which are represented in (Table-I) some markers have “NaN” value as a target class representing the class for test data.

**B. Database preparation**

After the data is acquired, we removed the mean from the available segments to get a better version of the dataset. After that, the positions of the markers were selected to consider the beginning of the 280 cues and the 3.5s movement duration. The processing procedures were carried out on four different sets. Some studies used three channels, which are C3, Cz and C4 as they contaminated the most discriminative MI information [22]. Other researchers tried the eight centro-parietal channels, namely C3, Cz, C4, CP1, CP2, P3, Pz and P4 [9]. On the other hand, Wang et al. [22] decided to use the 18 channels around the sensorimotor cortex including; C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4 and P6 channels. In this study, these three channels sets and the 25 channels [51-75] suggested in [9] are explored in our proposed experiments.

**C. Parameters**

After preprocessing signals were decomposed using WPD, the number of scale levels chosen was 4, and the wavelet function was “db2” which results in 2^j sub-bands, where j is the number of scale levels. A total of 16 sub-bands after WPD is produced.

**IV. RESULTS AND DISCUSSION**

In this section, we overview the results from the three proposed experiments where classification accuracy is the substantial metric for evaluation purpose. Training and testing trials are combined into one dataset and classified with linear SVM classifier. The ten Cross-Validations (CV) method is applied. Tables II shows a comparison of using feature extraction in different domains with different channel sets; the last row in each experiment represents the average accuracies of all subjects. The mean CA of experiment 3 outperforms the other two experiments for all channel sets. Experiment 1 shows the lowest mean CA, which ranges from (79.78%-86.3%). The combination of the 18 channels presents the highest mean CA of 90.34% and 91.72% for Experiment 2 and 3, respectively. Though the subject "aa" and "ay" got higher accuracy of 86.4% and 95.4% when using the 25 channels compared to other channel sets. On the other hand, the 25 channels combination gave the highest accuracy of 86.3% for experiment 1.

Figure 3 shows a comparison between the mean CA of the three experiments. Experiment 3 results in the highest mean CA (91.72%). On the other hand, the other two experiments mean CA (90.34% and 86.3%) for experiment 2 and 1, respectively. Tables III presents a clear comparison between our proposed system and recent related works. Wang et al. [23] the winners of BCI competition III datasets IVa got the highest accuracy. They used three different feature extraction methods. For al, aw and ay they used CSP algorithm on Event-Related Desynchronization (ERD). While for aa and av they used the hybrid feature set (CSP and Autoregressive (AR)) extracted from 18 channels. They constructed a Linear Discriminant Analysis (LDA) for the classification process and achieved a 94.2% mean CA. Selim et al. [9] extracted features using RMS. The authors built LDA classifier using 18 channels as well. They reach an average CA of 78.77%. Miao et al. [7] extracted features depending on spatial-frequency-temporal patterns. Their mean CA obtained was 86.38%. On the other hand, Singh et al. [8] designed a spatial filter which converts Sample Covariance Matrices into a lower dimension. The authors achieved an average CA of 86.13%. Selim et al. [4] introduced a hybrid feature selection model which reduce the number of CSP features and using the same 18 channels used in [9].

Our study shows a high average CA of 91.72% by generating hybrid features from both domains and using the set of the 18 channels. The performance of our study is comparable to the results reported in the literature. This proves the effectiveness and strength of the proposed method.

**V. CONCLUSION**

This paper proposed a novel BCI system for classifying mental motor imagery tasks. The proposed method extracted features in the time, frequency domain for different sets of electrodes. It then fuses these features to construct several SVM models. The results showed that the proposed method could successfully classify mental motor imagery tasks. The hybrid fused features using 18 channels has reached a reliable mean CA of 91.72, which is higher than most of the recent related studies. For this reason, the proposed system can be used as a useful system for controlling a wheelchair. Moreover, it can improve the present rehabilitation treatments where a suitable response is carried once the patient performs the correct movement. This will correspondingly enhance rehabilitation outcomes over time.

**TABLE I Number of training and evaluation trials for each subject**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Training Trials</th>
<th>Test Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;aa&quot;</td>
<td>168</td>
<td>112</td>
</tr>
<tr>
<td>&quot;al&quot;</td>
<td>224</td>
<td>56</td>
</tr>
<tr>
<td>&quot;av&quot;</td>
<td>84</td>
<td>196</td>
</tr>
<tr>
<td>&quot;aw&quot;</td>
<td>56</td>
<td>224</td>
</tr>
<tr>
<td>&quot;ay&quot;</td>
<td>28</td>
<td>252</td>
</tr>
</tbody>
</table>

[Image 315x91 to 566x235]
<table>
<thead>
<tr>
<th>Experiments</th>
<th>Channels Subjects</th>
<th>C3 Cz C4</th>
<th>C3 Cz C4 CP1 CP2 P3 Pz P4</th>
<th>C5 C3 C1 C2 C4 C6 C5 CP3 CP1 CP2 C4 CP6 P5 P3 P1 P2 P4 P6</th>
<th>[51:75]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>Time domain features</td>
<td>aa</td>
<td>70.0</td>
<td>74.3</td>
<td>76.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>al</td>
<td>86.4</td>
<td>90.7</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>av</td>
<td>68.2</td>
<td>70.0</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aw</td>
<td>86.4</td>
<td>92.5</td>
<td>93.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ay</td>
<td>87.9</td>
<td>86.4</td>
<td>86.8</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td>79.78</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Time-frequency features</td>
<td>aa</td>
<td>70.4</td>
<td>77.5</td>
<td>84.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>al</td>
<td>91.4</td>
<td>95.7</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>av</td>
<td>68.2</td>
<td>71.4</td>
<td>77.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aw</td>
<td>90.7</td>
<td>96.1</td>
<td>97.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ay</td>
<td>88.6</td>
<td>90.4</td>
<td>93.2</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td>81.86</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Hybrid features (time and time-frequency feature domains)</td>
<td>aa</td>
<td>72.1</td>
<td>81.1</td>
<td>89.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>al</td>
<td>93.6</td>
<td>95.7</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>av</td>
<td>69.6</td>
<td>73.6</td>
<td>77.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aw</td>
<td>90.4</td>
<td>97.9</td>
<td>97.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ay</td>
<td>90.4</td>
<td>92.5</td>
<td>94.3</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td>83.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Accuracy (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. [23] (winners)</td>
<td>CSP/AR/LDA</td>
<td>96 100 81 100 98</td>
<td>94.2</td>
</tr>
<tr>
<td>Meng et al. [5]</td>
<td>CSP + channel selection + SVM</td>
<td>82.4 98.6 76.8 94 96.6</td>
<td>89.68</td>
</tr>
<tr>
<td>Selim et al. [9]</td>
<td>RMS+LDA</td>
<td>74.11 96.43 60.71 71.88 84.2</td>
<td>78.77</td>
</tr>
<tr>
<td>Miao et al. [7]</td>
<td>spatial-frequency-temporal patterns</td>
<td>81.25 100 65.31 93.30 92.06</td>
<td>86.38</td>
</tr>
<tr>
<td>Singh et al. [8]</td>
<td>SR-MDRM</td>
<td>79.46 100 73.46 89.28 88.49</td>
<td>86.13</td>
</tr>
<tr>
<td>Selim et al. [4]</td>
<td>CSP\AM-BA-SVM</td>
<td>86.61 100.00 66.84 90.63 80.95</td>
<td>85.01</td>
</tr>
<tr>
<td>Joadder et al. [10]</td>
<td>Channel selection +SVM</td>
<td>59.2 91.0 58.5 83.5 87.8</td>
<td>76.0</td>
</tr>
</tbody>
</table>
VI. REFERENCES


