American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

EEG-Based Movement Imagery Classification Using Machine Learning Techniques and Welch's Power Spectral Density Estimation

Saman Sarraf*

Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada, the Institute of Electrical and Electronics Engineers, IEEE

Email: samansarraf@ieee.org

Abstract

This project implements an EEG-based movement imagery classification using Welch's Power Spectral Density estimation which could be used in Brain Computer Interface systems. This classification which is based on the extracted features from $\boldsymbol{\delta}$, $\boldsymbol{\theta}$, α, β1, β2 EEG components, is performed by K – Nearest Neighbor, Support Vector Machine (Linear, Quadratic and RBF) and Artificial Neural Network. This report covers the background and theory behind this approach, as well as describing the considerations involved in implementing this algorithm. A variety of classifier parameters are changed to demonstrate this implementation's performance.

Keywords: EEG; Machine Learning; Movement Imagery.

1. Introduction

Brain Computer Interface (BCI) systems, which could be developed during these years, create a non-muscular communication channel between the brain and external world [1]. Also, BCI systems would facilitate interactions between disabled people and devices. Functional brain imaging modalities such as fMRI, FNIRS, and EEG image the activities in the brain. The cheapest technique is Electroencephalography, which is non-invasive [2]. Electroencephalography or EEG is a technique for the recording of electrical signals coming from brain activities. An EEG recording is based on voltage fluctuations, resulting from ionic current flows within the neurons in the brain.

----

* Corresponding author.

EEG has a good temporal resolution, and these days it is used in dual modality fMRI-EEG in the functional brain imaging [3]. Although EEG – based BCI systems have the potential to perform correctly, there are still several problems to achieve an ideal accuracy. Brain Computer Interfaces uses processed EEG signals. The processing of EEG signals, which is a vast research area, includes the extraction of convenient and discriminative features from EEG signals, and then classifies these features into several clusters. For example BCI motor-disabled people need to decide to do a task mentally. The BCI will detect the changes in EEG, which have been affected by a patient's mental task, and it will function instead of them [1,2]. Motor-Imagery based BCI (MI-BCI) processes the brain signals from the imaginations of left or right hand movements. It is developed based on the neural activation of movement and movement imaginations. The movement related features of EEG are derived from Movement Related Potential (MRP) or Event-related De-synchronization / Synchronization (ERD/ ERS), which are not studied in this work. The neuronal activation, during the preparation of movement, results in contra-lateral effects of EEG and this is termed as readiness potential (RP) [4]. As artifacts might affect the deterioration of the accuracy of a BCI system, it is mandatory to handle artifacts, or to design BCI systems, whose performance is robust in the presence of artifacts. Many studies show the focus on the noise removal in single-trial EEG, where the average of multi-trials is not available [5]. In this course project real movement stimulations (not a mental task) on the left and right hand were performed by two subjects. EEG signals through four channels were recorded during almost 10 seconds without a baseline. After data collection, several pre-processing tasks were applied on signals such as filtering, removing zeros, and normalization. The important and principal features related to motor cortex stimulation are EEG components such as delta (1.0-3.5 Hz), theta (4.0-7.5 Hz), alpha (8.0-12.0 Hz), beta-l (12.5-20.5 Hz), beta-2 (21.0-30.0 Hz). After feature selection, the window-based feature extraction was performed using the average and standard deviation of mentioned EEG component [6,7,8,9]. In order to compare different classification methods, several pattern recognition approaches such as Support Vector Machine, K-Nearest Neighbor algorithm (K-NN) and Artificial Neural Network were used, and results were compared. Two approaches of feature extraction were considered, called online and offline feature extraction techniques. As mentioned before, BCI systems are often active, and they work in real time. So, an online approach of feature extraction is needed. The online method represents a feature extraction based on dividing EEG signals to several segments and calculating features, while the offline method will measure feature from the whole signals. For segmentation of the EEG signal, it is necessary to use a moving time window, which is small enough to approximate the stationary of the underlying process. In the classification step, some parameters of classifiers were changed in order to achieve a high performance, and then the results of K - NN, SVM, and ANN were compared. The accuracy of this work is lower than one presented in the literature. This was expected, because of lack of data and the imprecise data collection protocol. The most significant conclusion in this work will illustrate that not only a number of data can affect results, but also classifier selection plays an important role.

2. Background

The human brain is formed of several lobes, which are responsible for different functions, such as visual or auditory processing systems and etc. The motor areas are located in both hemispheres of the cortex. They are shaped like a pair of headphones stretching from ear to ear. The motor areas are closely related to the control of voluntary movements, especially fine fragmented movements performed by the hand. The right half of the motor area controls the left side of the body, and vice versa. Two areas of the cortex are commonly referred to as motor: 1) Primary motor cortex, which executes voluntary movements 2) Supplementary motor areas and premotor cortex, which select voluntary movements. In addition, motor functions have been described for: 1) Posterior parietal cortex, which guides voluntary movements in space; 2) Dorsolateral-prefrontal cortex, which decides which voluntary movements to make according to higher-order instructions, rules, and self-generated thoughts [10,11]. In the official 10-20 standard of EEG electrodes placement, C3, C4, P3 and P4 are more associated to motor cortex activities than other electrodes (which have been considered in this project). In a full EEG signal recording, more channels are used to define brain activity. Studies [12,13,14] identify that several EEG components are involved when there is external movement stimulation. Christa Neuper and his colleagues [13] indicate hand movement stimulation returns de-synchronization of alpha (mu) (8-12 Hz) as well as central beta rhythms (13-18 Hz). Furthermore, it could be possible to distinguish between imagination of right and left hand stimulation based on single-trial EEG [13].

Akrami and his colleagues [2] shows an acceptable classification on movement imagery based on logarithmic Power Spectral Density of in the standard bands: delta, theta, alpha, beta, and gamma. Meng Hu and his colleagues. Reference [15] used approximate entropy and the Welch Power Spectral Density [16,17] together for the feature extraction step, in order to classify normal and hypoxia EEG signals. PSD analysis returns the basic information of how power plays the role of the frequency function [15]. Guilherme A. Barreto and his colleagues [18] made a comparison between neural and statistical approaches of EEG signals and concluded that Welch's periodgram allows the classification methods to reach higher accuracy rates (at that time) [18]. Several EEG signal processing techniques such as Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS) [19], Autoregressive Estimation [20], Independent Component and Coherence Analysis and STFT [21], etc. have been used in the online and offline approaches in order to extract suitable features from segmented EEG signals. Statistical approaches such as Student's t-test method and etc. as well as pattern recognition algorithms such as Artificial Neural Network, Bayesian classifier, and Support Vector Machine, etc. would be useful tools for classification. Meng Hu and his colleagues [15] compared the results of classification by Bayesian classifier and SVM in which accuracy was 90.8% and 92.5%, respectively.

Christa Neuper and his colleagues [13] reported 56% and 67% accuracy for the classification of visual-motor imagery and kinesthesis motor imagery by Distinction Sensitive Learning Vector Quantization (DSLVQ) method. Generally, the accuracy reported about motor cortex imagery classification based on different feature selection methods and different pattern recognition algorithms covers from almost 55% to 95%, showing huge differences between features and classifiers selection. Valerie Morash and his colleagues [19] demonstrated that the ERD/ERS preceding movement could be used in order to predict, which of four type movements would occur. Khodayari-Rostamabad and his colleagues [22] at McMaster University indicated that the analysis of pre-treatment EEG signals could forecast the clinical response to clozapine in treatment resistant schizophrenia. In this work, 16–channel EEG was used, and the signals were purified by a band pass filter (0.5 and 80 Hz). 20 discriminative features such as Coherence in a special frequency between two specific electrodes were used and a kernel partial least squares regression (KPLSR) method was performed in the classification step. Using the mentioned method, an overall accuracy rate of almost 84% was reported [22].

3. Theory and Approach

The algorithm used in this course project was designed based on literature review and includes four main phases. Figure 1 shows the steps of operational procedures which are:

- Data Collection
- Pre – Processing of EEG Signals
- Feature Selection and Extraction
- Pattern Classification

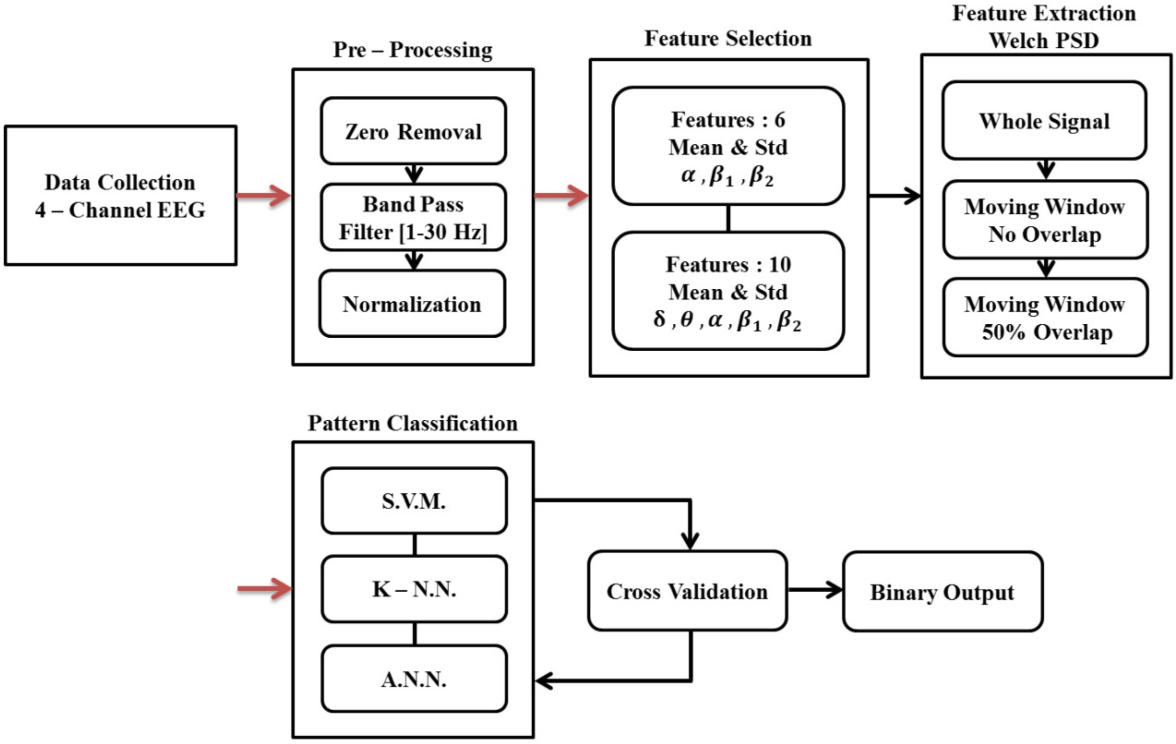

Figure 1: The flow chart above illustrates the steps of the designed algorithm.

This diagram illustrates the steps of movement imagery classification using Welch PSD method. As is shown, there are several steps in each phase such as feature selection and … which will be discussed in next chapter, completely. For instance, it is recognized from Figure 1 that feature extraction was executed six (2*3) times for all EEG signals.

3.1. Welch Power Spectral Density

The various methods of spectrum estimation are categorized as follows: a) Nonparametric methods, b) Parametric methods and c) Subspace methods. Nonparametric methods are those in which the Power Spectral Density is estimated directly from the signal itself. The simplest such method is the periodogram. Other nonparametric techniques such as Welch's method [1], the multitaper method (MTM) reduce the variance of the periodogram. As mentioned before, the Welch method estimates the power spectral density of signal and is an extension of the Short Time Fourier Transform (STFT). This method profits a normalization approach in order to calculate the scaling of the spectrum [1]. The Welch PSD estimation is presented by Equation 1:

$$P_{va}^{welch}[d,k] = \frac{1}{NMU} \sum_{n=0}^{M-1} \left| \sum_{l=(Md+n)N}^{(Md+n)N+2L} w[l-(nN+L)] v_a[l] e^{-\frac{2\pi l d}{N}} \right|^2$$

Equation 1: Welch Power Spectral Density equation

Where the signal is presented by va and k takes values k=0, 1,…N-1, which describes the signal spectrum. W[l] is a symmetric N-point windowing function. This equation shows the averaging of PSD of M consecutive non-overlapping segments of length N each. So NM data points will be used. U is the normalization factor for the window w and it is given by Equation **2**:

$$U = \frac{1}{N} \sum_{l=-L}^{L} w^2[l]$$

Equation 2: Normalization Factor in Welch PSD

The index d, which is an integer, describes the start of the data interval over which the computations are performed. The interval starts at time dNM, e.g. for d = 0 the data points at times 0, 1,⋯, NM -1 are used for spectrum computations [23]. Furthermore, the Welch's method returns a biased estimator of PSD, which is presented by Equation **3**:

$$E\{P_{Welch}(f)\} = \frac{1}{FsLU} \int_{-Fs/2}^{F/2} \left| W(f-\acute{f}) \right|^2 P_{xx}(\acute{f}) d\acute{f}$$

Equation 3: Bias and Normalization in Welch's Method

Here L is the length of the data segments, U is the same normalization constant present in the definition of the modified periodogram, and W[f] is the Fourier transform of the window function. As is the case for all periodograms, Welch's estimator is asymptotically unbiased. For a fixed length data record, the bias of Welch's estimate is larger than the periodogram value, because the length of the segments is less than the length of the entire data sample. The variance of Welch's estimator is difficult to compute because it depends on both windows used and the amount of overlap between segments. Basically, the variance is inversely proportional to the number of segments whose modified periodograms are being averaged [17,23,24,25].

3.2. Pattern Classification Methods

In order to obtain the best-case scenario of data analysis in this course project, several classifiers, such as Support Vector Machine, K–Nearest Neighbour and Artificial Neural Networks with a variety of parameters, were trained and the results of all classifiers were compared together, which will be studied in the next chapter.

3.2.1. K-Nearest Neighbor algorithm (K – NN)

K – NN is a simple and fast algorithm in pattern recognition, which is based on closest training examples in the feature space. The objective of K – NN is to classify an object by a majority vote of its neighbours. K representing the class of nearest neighbours is a positive integer, which is typically small. In binary classification (like this project) K will be an odd number selected by experience or through cross validation [26].

3.2.2. Support Vector Machine (SVM)

Support Vector Machine or briefly "SVM" is a supervised classification method in statistics and computer science, which is a powerful tool for recognizing patterns. It is mostly used in regression analysis and classification problems. This classifier became popular when it had an excellent reputation in manuscript classification projects during the early 2000s. Several kernel functions can be used in SVM such as Linear, Quadratic, Gaussian Radial Basis Function called RBF, Polynomial, etc. [21]. SVM needs a lot of data to be well trained for getting trustable results and high accuracy.

3.2.3. Artificial Neural Network (ANN)

ANN is a mathematical model inspired by biological neural networks. Modern neural networks, which are usually non-linear statistical data modeling, are used to model or simulate complex relationships between inputs and outputs or to find patterns in data. It consists of hidden layers and neurons processing information using connectionist algorithm for computation. ANN could be categorised in Supervised or Unsupervised learning methods depending how it would be trained [21]. In terms of time processing, ANN is a slow algorithm compared to SVM but its sensitivity to number of data is less than SVM.

4. Discussion of Experiments and Results

In this chapter, the whole processing EEG signals will be studied step by step and the results will be discussed. As mentioned in the previous chapter, there are four main phases in the designed algorithm for this classification problem. First, data collection was performed, which was followed by a pre – processing, feature extraction, and finally classification. R2010b MATLAB version 7.11.0.584 (64-bit) was used for programming.

4.1. Data Collection

EEG signals were recorded by CleveLabs system and 4-channel data (C3, C4, P3, and P4) were collected in several trials, including close and open eyes for right and left hand stimulation, as well as several neutral trials, in which the subject was asked to be relaxed and do no task. The sequence of trials was defined randomly by a MATLAB function. Table 1 shows the detail of collected EEG trials. The subject was asked to do the task (left or right hand stimulation or no task) and each trial took almost 10 seconds with the (sampling frequency) Fs = 480 Hz. Figure 2 demonstrates an unprocessed 4-channel EEG signal for the "open eye – left hand stimulation" trial during one epoch of almost one second. As is shown, some zeros happened during the signal recording, which should be removed in the pre-processing step. In practice, when zeros happen during EEG recording, the experiment might be stopped, and the amplifier would be double-checked in order to collect accurate data from patient.

Table 1: Number of trials in EEG data collection

	Close Eyes	**Open Eyes**
Left hand	**3**	**7**
Right hand	**3**	**7**
Neutral	**2**	**4**

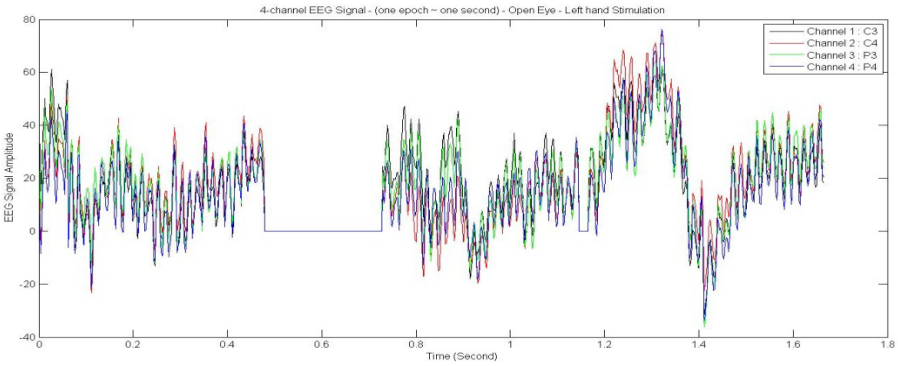

Figure 2: 4-channel EEG Signal - one epoch (one second) - Open Eye - Left hand Stimulation

Since movement stimulation has contralateral effect on EEG, it is expected that during left hand stimulation C4 and P4 electrodes have higher amplitude than C3 and P3. As is demonstrated in Figure 3, because of volume conduction this rule was not respected completely, and amplitude of EEG is spread on all electrodes.

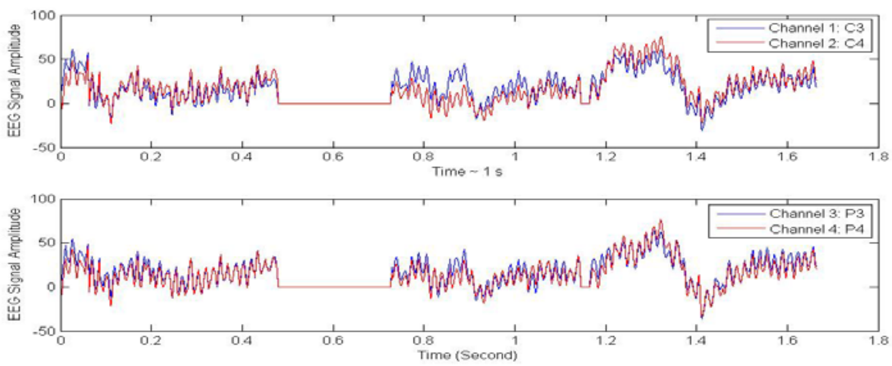

Figure 3: Contralateral Assumption of Motor Cortex Stimulation

4.2. Pre–Processing of EEG Signals

The first phase of EEG signals analysis in the study is pre–processing, which was done in three steps: a) Zero removal b) Filtering and c) Normalization. The clean signals were obtained after the mentioned procedures.

4.2.1. Zero removal

The first EEG signals included several zeros, which happened during the data collection, because of the EEG recording machine. As these unexpected data affect the whole EEG processing, zero removal was performed and consecutive EEG signals were achieved. The Figure 4-3 shows one epoch of EEG before and after zero removal.

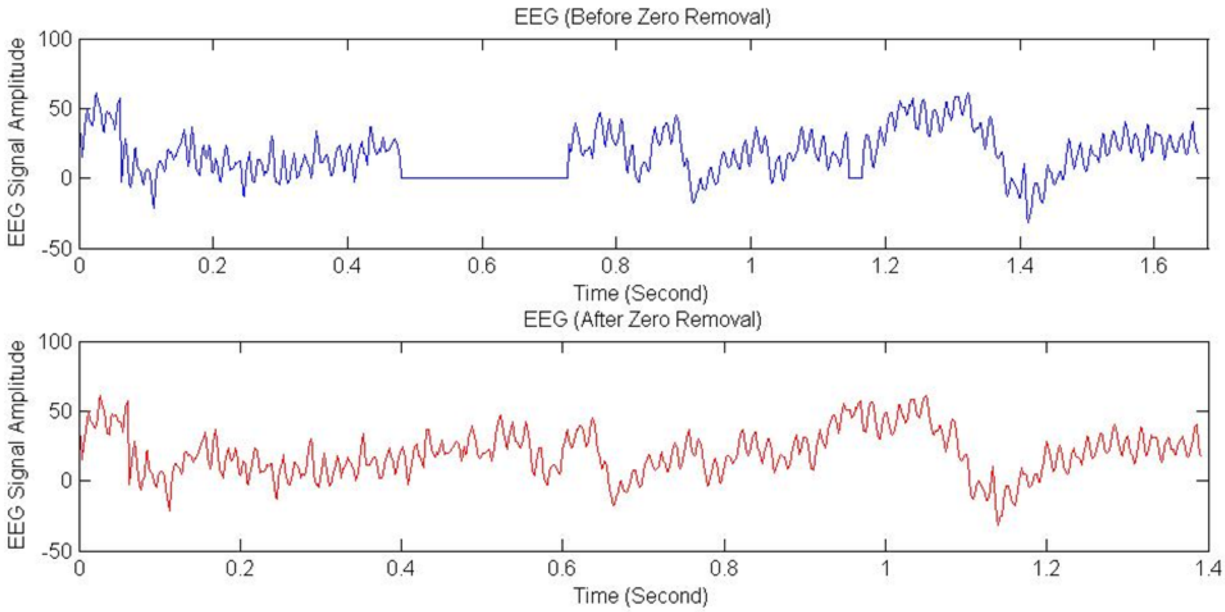

Figure 4: Pre-processing (Zero Removal)

4.2.2. Filtering

It seems the EEG recording machine had used an analog or digital filter. Although the acceptable EEG signals with no major problem were obtained, the Butterworth band pass filter with the order of N = 4 where Fbp = [1 31] was used in order to remove noise and useless frequency bands. As mentioned in background, the EEG components used start from 1 Hz to 30 Hz.

4.2.3. Normalization

In the third step of pre-processing, data normalization was applied on EEG signals to obtain smooth data, since pattern classification algorithms need normalized inputs. Equation 4 and Equation 5 present the data normalization formula of given signal. Figure 5 shows the previous EEG signal after filtering and normalization.

$$Normalized\ X = \frac{X - \bar{X}}{\sigma(X)}$$

Equation 4: Normalization method

$$\sigma(X) = \sqrt{\sum_{i=1}^{N}(x_i - \bar{X})^2} \qquad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Equation 5: Standard deviation of X (left), mean of X (right)

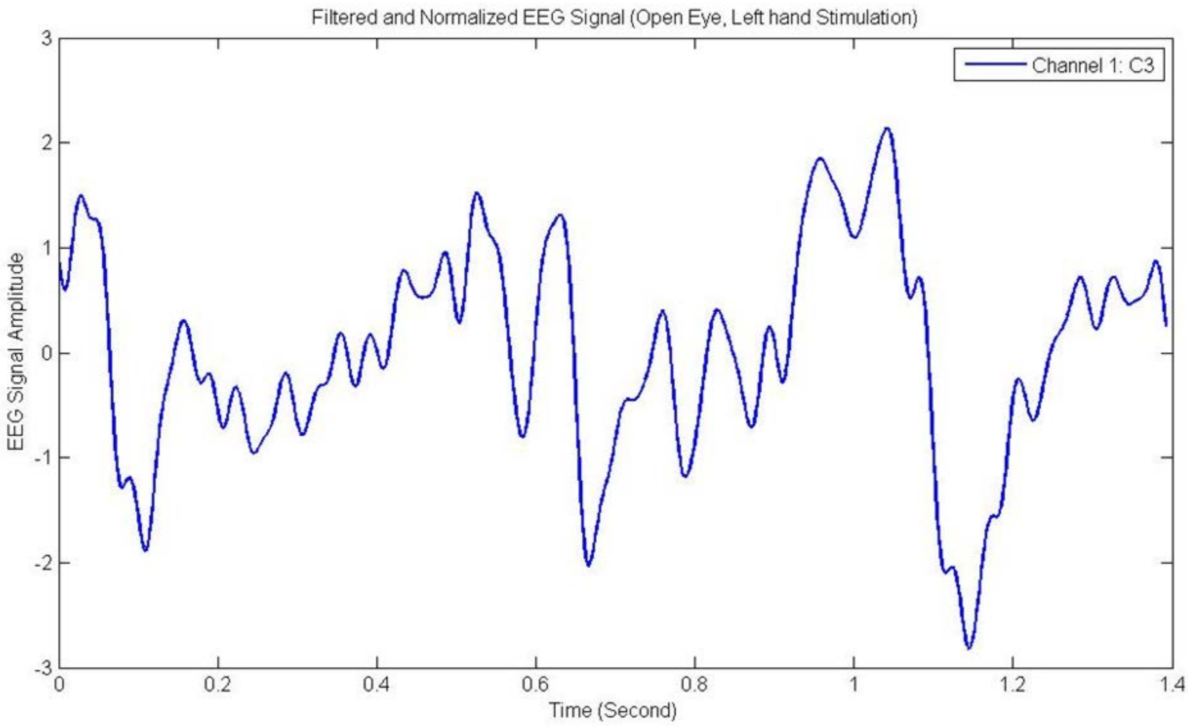

Figure 5: Filtered and Normalized EEG Signal

4.3. Feature Selection and Extraction

Two feature sets were considered, which were based on EEG components. At first, Welch Power Spectral Density of signals (or a segment of signal) was calculated. According to the literature, motor cortex stimulation affects significantly EEG components: alpha (8.0-12.0 Hz), beta-1 (12.5-20.5 Hz), beta-2 (21.0-30.0 Hz), which formed the first feature set.

The second feature set included: delta (1.0-3.5 Hz), theta (4.0-7.5 Hz), (8.0-12.0 Hz), beta-1 (12.5-20.5 Hz), beta-2 (21.0-30.0 Hz).

Since data collected belonged to one subject and had the same distribution, the average and the standard deviation of the Welch PSD on the mentioned components were calculated as features. The detail of feature selection is mentioned in Table 2 and Table 3 .

Table 2: The first set of features

The first feature set (6)	
Operand	Component
Mean of Welch PSD	alpha
	beta-1
	beta-2
Standard deviation of Welch PSD	alpha
	beta-1
	beta-2

Table 3: The second set of features

The second feature set (10)	
Operand	Component
Mean of Welch PSD	delta
	theta
	alpha
	beta-1
	beta-2
Standard deviation of Welch PSD	delta
	theta
	alpha
	beta-1
	beta-2

Feature extraction from pre-processed EEG signals was performed by three approaches. First of all, the whole signal was considered and two sets of features applied, so two feature vectors, including six and ten values were obtained, respectively. Although the feature space could be small by this approach, the minimum of noise and high SNR will be achieved. The mentioned method called off-line feature extraction might be more useful, when lots of data (EEG signals) are used to train classifier.

Secondly, a moving window scanned all of a signal, and extracted the features based on two mentioned feature sets. The size of the moving window was equal to one epoch, which divided each trial into 10 segments. Hence, each epoch has almost one second length. Two feature vectors of 60 and 100 values were obtained for each signal (signal means each channel of any trial).

Finally, a moving window, having 50% overlap, scanned the signals as in the second step. Two feature vectors including 114 and 190 values were received, respectively. As not many dataset were used in this work, it would be expected that the third method of producing feature vectors could get the better results. In other words, applying a moving window with overlap could extract more than details from signal resulting in a better continuous classification. Although this approach seems to be more accurate, it could decrease SNR, and it might extract information from noise.

Lastly, in order to form a unique feature vector, all vectors were added together. Firstly, all feature vectors of a trial (having 4 channels) were joined together, and then all matrices coming from all trials were added together. Equation 6 and Equation 7 show mathematically the final feature vector (fv) formation.

$$FV_{trial(i)} = \{fv_{trial(i)(channel\ 1)}, \dots, fv_{trial(i)(channel\ 4)}\}$$

Equation 6: Feature vector of each trail (including 4 channels)

$$FV_{final} = \{FV_{trial(1)}, FV_{trial(2)}, \dots\}$$

Equation 7: Final feature vector of all trials

4.4. Classification

The aim of the classification step is to identify the right or left stimulations, based on these processed EEG signals. Thus, in the output of classifiers there will be a matrix of labels related to right hand or left hand stimulation. 20 trials of data (except neutral trials) were used. 70% of trials (14 trials) trained classifiers and 30% (six trials) tested them. The sets of train and test data were randomly defined by a function in MATLAB. In order to compare the results, the same set of train and test data were experienced for all classifiers. The classification step was repeated 5 times with randomly selected different sets of train and test data for all classifiers (K – NN, SVM and ANN) including six different feature vectors to achieve reliable results, and to demonstrate the repeatability and reproducibility of the designed algorithm. At the end of each classification step, the performance of the system including precision and accuracy was calculated by Equation 8:

$$Precision = TP/(TP + FP) \qquad Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

Equation 8: Precision (left) and Accuracy (right) formula. TP: True Positive, TN: True Negative, FP: False Positive and FN: False Negative. For example: TP in right hand stimulation refers an output value which was classified correctly (right hand as right hand) and FP in the same experiment refers an output value which was classified improperly (right hand detected as left hand) and etc.

4.4.1. K – Nearest Neighbor (K – NN)

K – NN classifier applied on 14 trials data and tested by six trials data. In terms of simplicity and speed, K – NN is the first choice. In order to obtain a reliable result, the cross validation had to be performed on the distance factor K. As mentioned before, all classification steps were repeated five times on each feature vector, but for avoiding the text complexity, only more trustable result will be discussed in Figure 6.

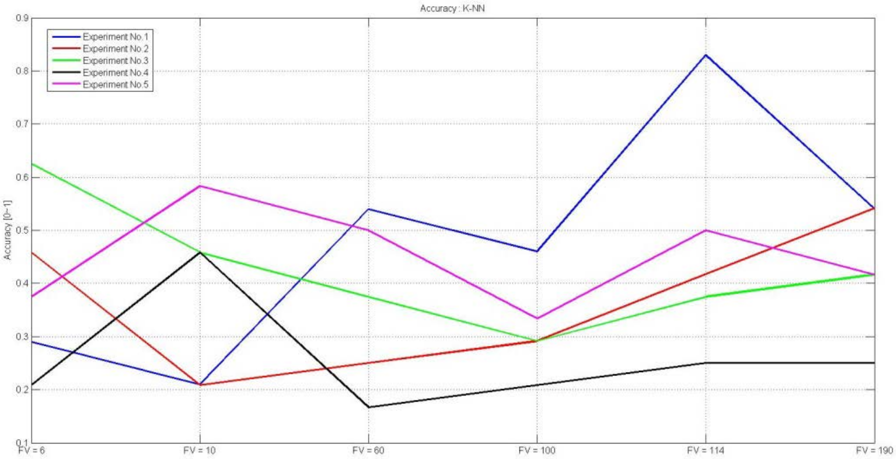

Figure 6: K-NN classification was repeated 5 times for 6 feature vectors. All related accuracies are shown above. The discussion on results will be kept on the first experiment. (The first train and test data set)

4.4.1.1. Cross Validation

For each feature vector classified by K – NN, five odd values for K (K = 1, 3, 5, 7, 9) were applied and finally, the K value minimizing Mean Squared Error between classifier output values and test label ones was considered as the optimum K. Next, the classification was done with this K value. The Figure 4 6 demonstrates a sample of K – NN cross validation.

4.4.1.2. K – NN: Results

After applying the optimum K value for classifying each feature vector, the results below were obtained (Figure 8). This figure shows the best results happened when there were 114 and 190 features indicating to feature extraction by the moving window with 50% overlap. Although the accuracy of FV = 114 is better than the others, it seems the results of both FV = 114 and 190 could be more reliable.

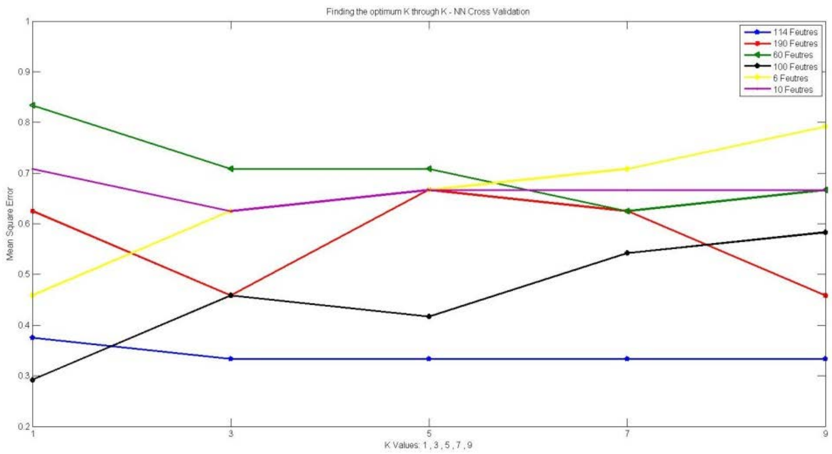

Figure 7: Finding the optimum K through K – NN Cross Validation

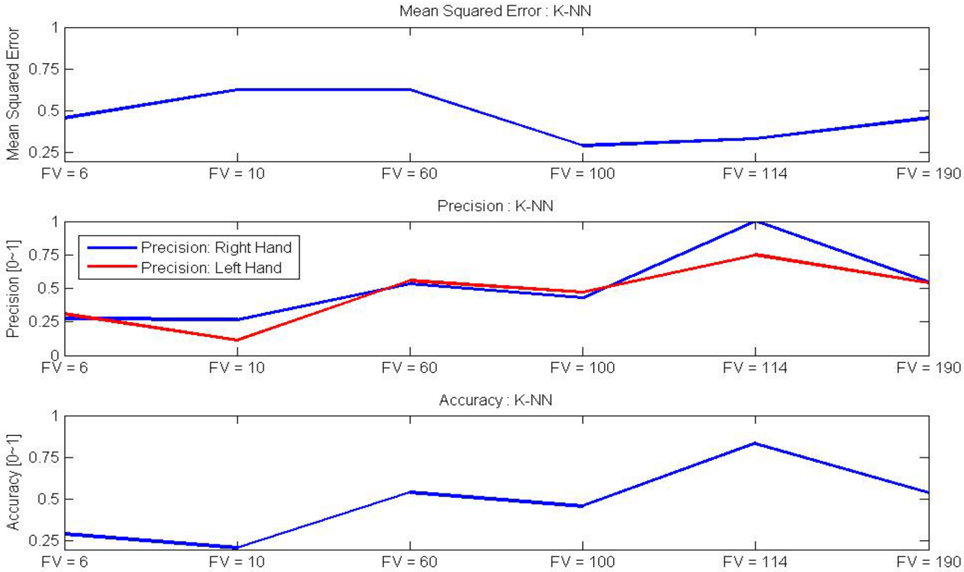

Figure 8: MSE (Top), Precision (Middle) and Accuracy (Bottom) for six experiments including 6, 10, 60, 100, 114 and 190 features. The best result belongs to 114 and 190 features.

The results of K – NN classification (Table 4) reveals that the accuracy depends on the number of features, generally. It might conclude that the simple feature vectors (FV = 6, 60, 114) could be classified better by a simple classifier such as K – NN. More complex feature vectors should be classified by a more complicated network.

Table 4: K – NN Classification values (MSE, Precision Right and Left, Accuracy) for six Feature Vectors

	FV1 = 6	**FV2 = 10**	**FV3 = 60**	**FV4 = 100**	**FV5 = 114**	**FV6 = 190**
MSE	0.46	0.63	0.63	0.29	0.33	0.46
Precision R	0.27	0.27	0.53	0.43	1.00	0.55
Precision L	0.31	0.11	0.56	0.47	0.75	0.54
Accuracy	0.29	0.21	0.54	0.46	0.83	0.54

4.4.2. Support Vector Machine (SVM)

The SVM classifier includes several kernel functions such as linear, quadratic, Radial Basis Function (RBF), etc. Three mentioned functions were used in the SVM classification. The SVM was performed for the same data sequence used in the K–NN. Therefore, the classifier was trained by 14 trials data and tested by six ones. Five times, all procedures were repeated for six different feature vectors, and Figure 4 8 shows a comparison of them. As mentioned before, to avoid making complex this report, the discussion will be done on the results showing more reliability.

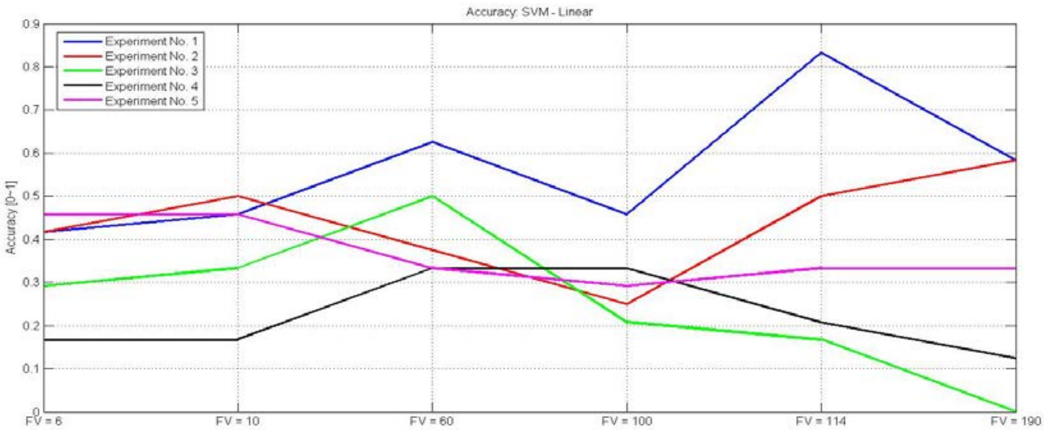

Figure 9: SVM classification was repeated 5 times for 6 feature vectors. All related accuracies are shown above. The discussion on results will be kept on the first experiment. (The first train and test data set)

4.4.2.1. SVM – Linear: Results

First kernel function performed was “Linear” function. Figure 10 illustrates MSE, Precision and Accuracy in this step. As is seen, the best accuracy happened for FV = 114 and then FV = 190. (FV = 114 means the feature extraction based on a moving window 50% overlap and 6 features). The SVM linear classification values are indicated on Table 5.

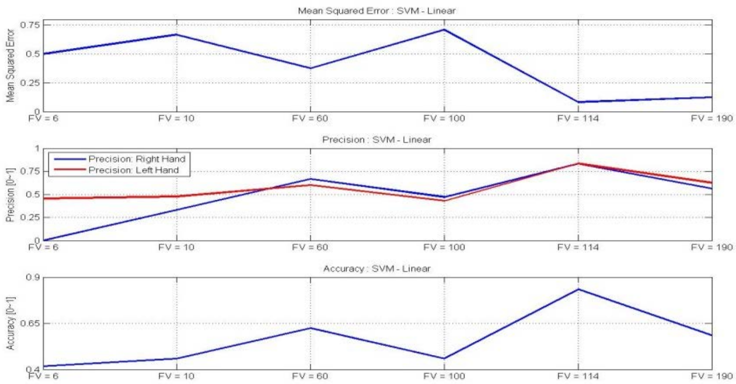

Figure 10: MSE (Top), Precision (Middle) and Accuracy (Bottom) for six experiments including 6, 10, 60, 100, 114 and 190 features. The best result belongs to 114 and 190 features.

Table 5: SVM Linear classification values (MSE, Precision Right and Left, Accuracy) for six Feature Vectors

	FV 1 = 6	FV 2 = 10	FV 3 = 60	FV 4 = 100	FV 5 = 114	FV 6 = 190
MSE	0.50	0.67	0.38	0.71	0.08	0.13
Precision R	0.00	0.33	0.67	0.47	0.83	0.56
Precision L	0.45	0.48	0.60	0.43	0.83	0.63
Accuracy	0.42	0.46	0.63	0.46	0.83	0.58

4.4.2.2. SVM Quadratic: Results

The same procedure was implemented using "Quadratic" kernel function for SVM. Figure 11 and Table 6 expose the related results. In this experiment, the best accuracy belongs to FV = 190. Also, FV = 114 returned acceptable results.

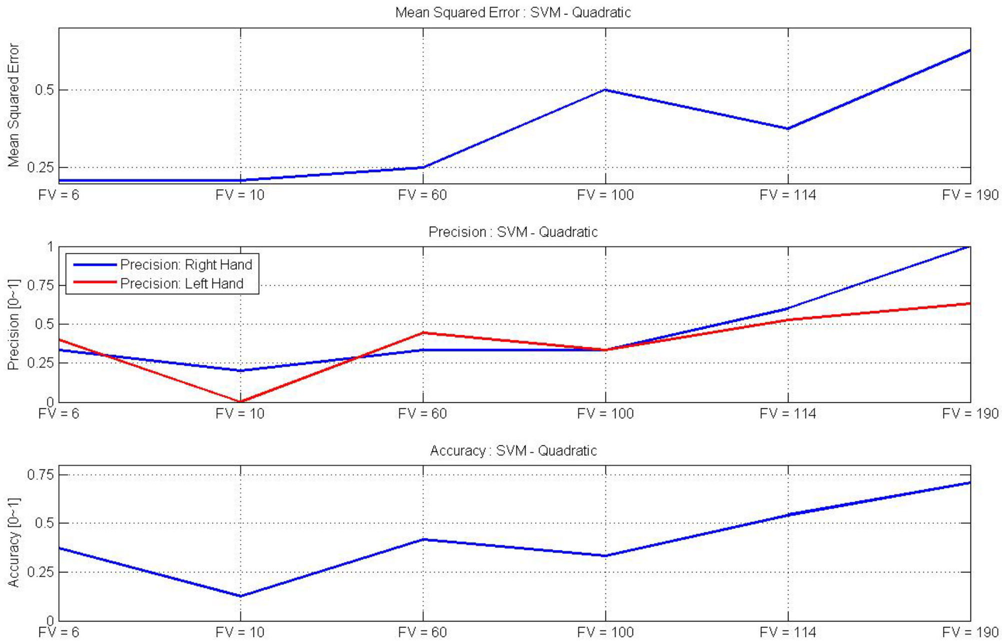

Figure 11: MSE (Top), Precision (Middle) and Accuracy (Bottom) for six experiments including 6, 10, 60, 100, 114 and 190 features. The best result belongs to 114 and 190 features.

Table 6: SVM Quadratic classification values (MSE, Precision Right and Left, Accuracy) for six Feature Vectors

	FV 1 = 6	FV 2 = 10	FV 3 = 60	FV 4 = 100	FV 5 = 114	FV 6 = 190
MSE	0.21	0.21	0.25	0.50	0.38	0.63
Precision R	0.33	0.20	0.33	0.33	0.60	1.00
Precision L	0.40	0.00	0.44	0.33	0.53	0.63
Accuracy	0.38	0.13	0.42	0.33	0.54	0.71

4.4.2.3. SVM Radial Basis Function (RBF): Results

The same method of classification was done, but there was an additional step. As is known, there is an important factor in SVM RBF called "sigma", which has to be optimized in order to return the optimal result. Consequently, the "Grid–Searching (Cross Validation)" was performed to find the optimum "sigma" which minimizes the MSE in the related classification.

4.4.2.4. Grid - Searching (Cross Validation)

The SVM RBF classifier was trained and tested each time with different feature vectors, while the sigma factor was changing from σ = 0.5 to 50 by the step of 0.1. Then the optimum "sigma" minimizing MSE was found, and the SVM RBF classifier was trained with the optimal sigma, again. Figure 12 illustrates the Grid – Searching procedure.

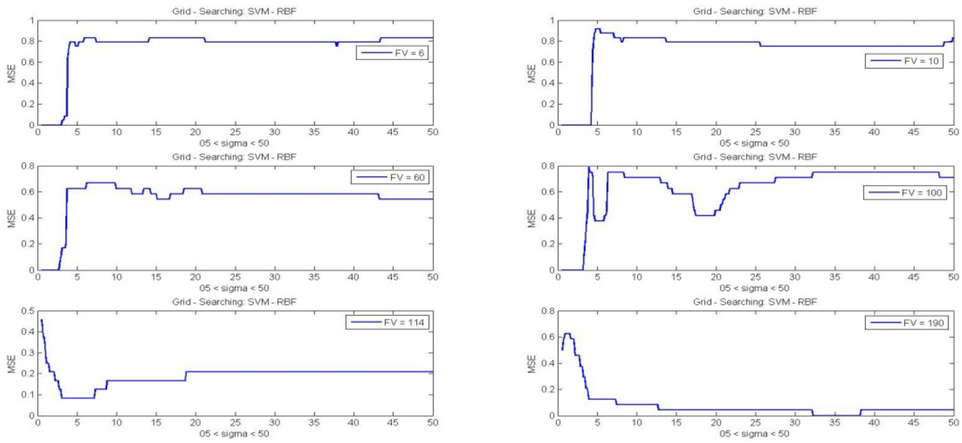

Figure 12: Grid - Searching (Cross Validation) SVM - RBF

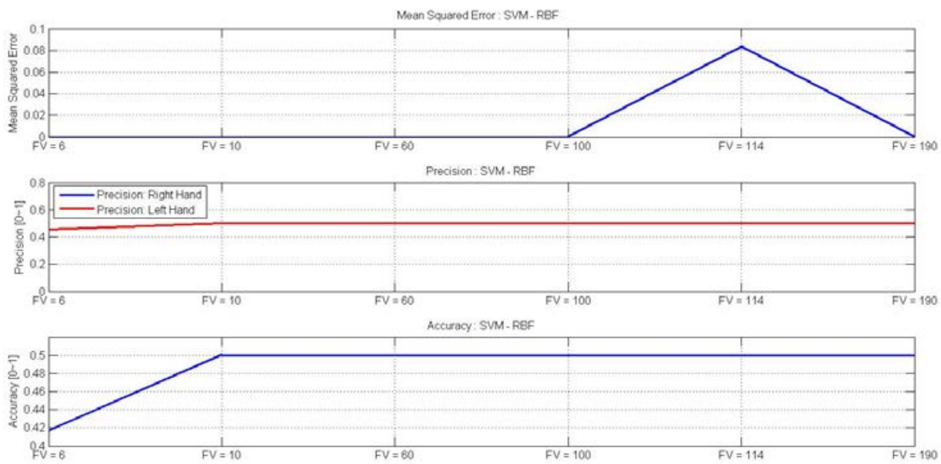

Figure 13: MSE (Top), Precision (Middle) and Accuracy (Bottom) for six experiments including 6, 10, 60, 100, 114 and 190 features. The best result belongs to 114 and 190 features.

Table 7:SVM RBF classification values (MSE, Precision Right and Left, Accuracy) for six Feature Vectors

	FV1 = 6	**FV2 = 10**	**FV3 = 60**	**FV4 = 100**	**FV5 = 114**	**FV6 = 190**
MSE	0.00	0.00	0.00	0.00	0.08	0.00
Precision R	0.00	NaN	NaN	NaN	NaN	NaN
Precision L	0.45	0.50	0.50	0.50	0.50	0.50
Accuracy	0.42	0.50	0.50	0.50	0.50	0.50

As is imaged in Figure 13 and Table 7, not reliable and trustable results were obtained from SVM RBF. It might be justified that, although the standard and complete approach was performed for SVM RBF (i.e. grid – searching), the data used for training were not sufficient and could not converge the designed system to a stable point. SVM RBF is one the most complex kernel functions in machine learning, which is highly sensitive to a number of data (big feature space). Generally, one of the characteristics of SVM is to have a high sensitivity to number of data and feature space. Thus, it is recommended to use the SVM, when there are lots of data for training and testing.

4.4.3. Artificial Neural Network (ANN): Results

Artificial Neural Network can act as a predictive model, as well as a classifier method. One of the advantages of ANN to SVM is lower sensitivity to number of data. But the whole processing steps takes a long time. In ANN classification, it is mandatory to define a margin based on a probability for the output of network. The output of ANN is not exactly similar to the label of test data, and a small fluctuation is often seen on them. For instance, ANN output based on label of test data, which is equal to “1”, might be “0.85” or “1.1”.

The label of test data (and train data) was zero for left and one for right stimulation. Thus, a margin of 0.505 was considered to distinguish the labels from ANN output. Equation 9 defines the mentioned probability:

$$If\ Y \begin{cases} \geq\ 0.505 \ \rightarrow\ Y = 1 \\ <\ 0.505 \ \rightarrow\ Y = 0 \end{cases}$$

Equation 9: Probability for finding labels from ANN output

Where “Y” presents one value of ANN output. This probability converts one value in ANN output to one label. The margin value was selected according to experience. Like K – NN and SVM classification, the whole process was repeated five times for all feature vectors by a feed-forward back-propagation network, but to avoid report complexity the discussion will be done on the most acceptable result (Figure 14).

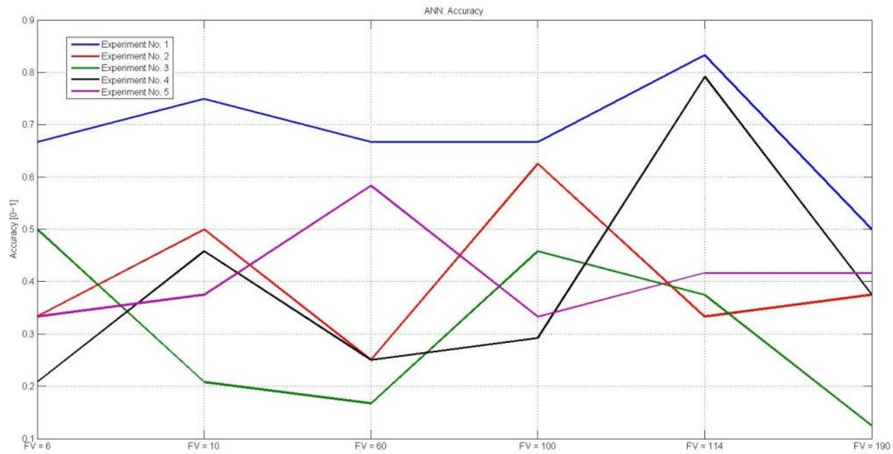

Figure 14: SVM classification was repeated 5 times for 6 feature vectors. All related accuracies are shown above. The discussion on results will be kept on the first experiment. (The first train and test data set)

4.4.3.1. Cross Validation

In order to find the best structure for ANN in this work, the cross validation on number of layers and neurons was performed. According to literature, the number of hidden layers changed from one to three, while the number of neurons changed from three to six. It means that for each feature vector, 12 tests were implemented to find the best network structure showed in figures below.

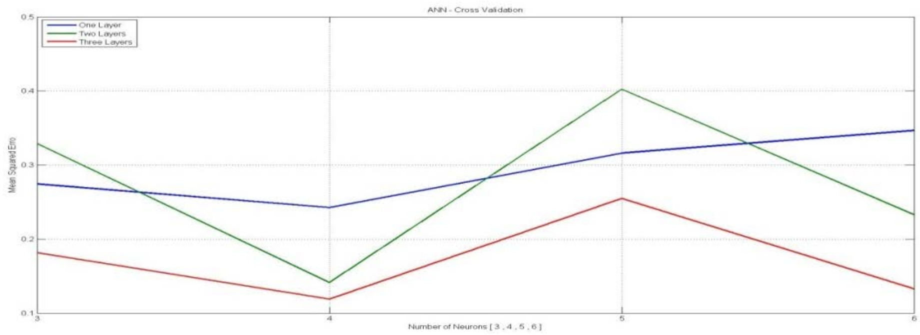

Figure 15: Cross Validation – one, two and three layers over three to six neurons (vertical axis: MSE, horizontal axis: number of neurons) blue: 1 layer, green: 2 layers and red: 3 layers. The min of MSE happens to 3 layers with 4 neurons.

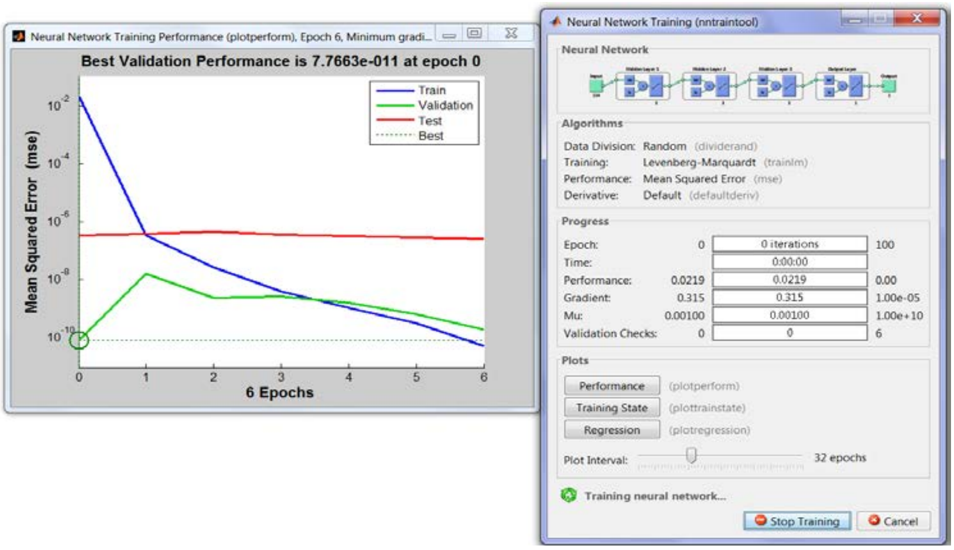

Figure 16: Neural Network Training

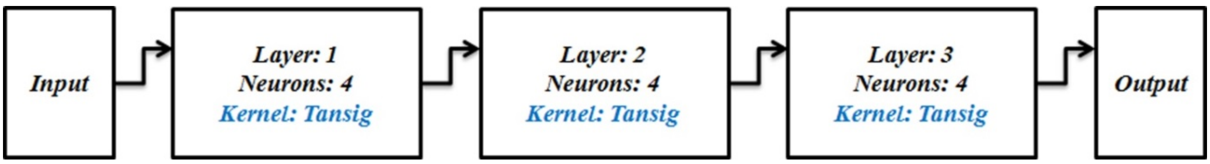

Figure 17: The structure above (feed-forward back-propagation 3 layers 4 neurons) was found in most cross validation cases as the best one minimizing the MSE of test data.

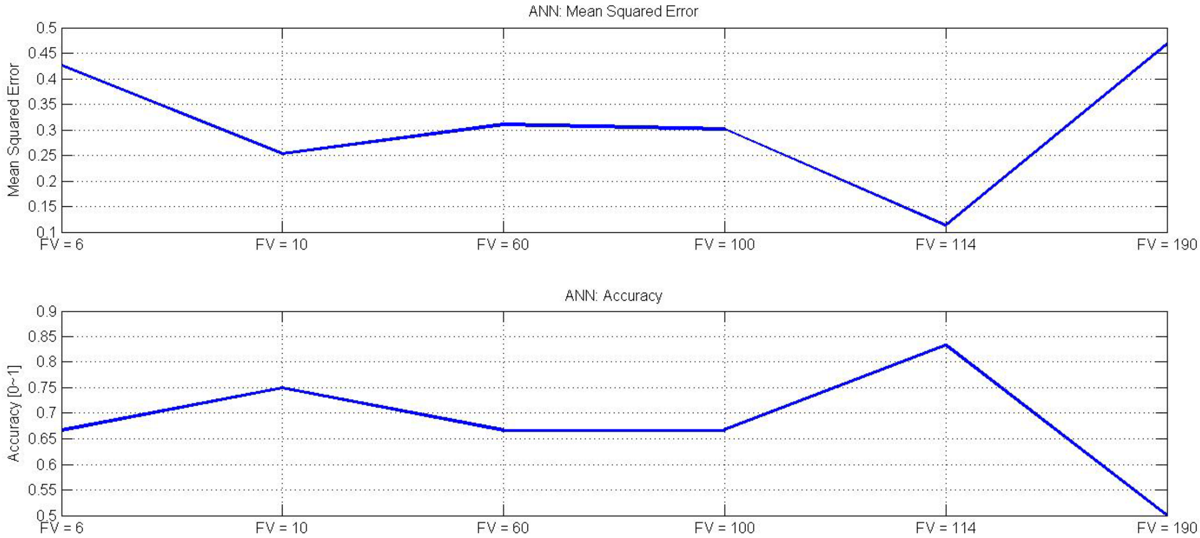

Figure 18: MSE (Top) and Accuracy (Bottom) for six experiments including 6, 10, 60, 100, 114 and 190 features. The best result belongs to 114 and 190 features.

Table 8: Neural Network classification values (MSE, Precision Right and Left, Accuracy) for six Feature Vectors

	FV1 = 6	**FV2 = 10**	**FV3 = 60**	**FV4 = 100**	**FV5 = 114**	**FV6 = 190**
MSE	0.43	0.25	0.31	0.30	0.11	0.47
Accuracy	0.667	0.750	0.667	0.667	0.833	0.500

The results of ANN classification (Table 8) reveal that the best accuracy happens when FV = 114 features. As mentioned before, FV = 114 is related to feature extraction with a moving window (50% overlap) while the mean and standard deviation of Welch PSD of alpha, beta-1 and beta-2 were considered as features.

4.5. Classifiers Comparison and Discussion

The results below revealed that the 5th feature vector which was performed by calculating the mean and standard deviation of α, β1 and β2 based on a moving window with 50% overlap, achieved the highest accuracy and the lowest MSE in all classification methods except SVM – RBF. Thus, based on these EEG signals, motor cortex stimulation had significant effect on α, β1 and β2, as expected. Among classifiers the most reliable results were obtained by Artificial Neural Network, since it showed a lower sensitivity to number of data and having an enough complex network to find the pattern from data. In some outputs, feature vectors based on six features showed a better relationship with linear networks such as SVM – Linear, while 10 features exposed a better performance with more complex classifier such as SVM – Quadratic. On the other hand, based on these data, there was no significant effect of delta and theta component on this classification. As mentioned in the theory chapter, the complex kernel functions, such as RBF, are sensitive to a large number of data. SVM – RBF could not be trained well, and did not show a reasonable performance. So, its results could not be trusted in this work. The general trend would be towards larger feature space. It means that online feature extraction could be a better option. The K – NN classification was the fastest approach as expected. SVM and ANN were in the second and third place, respectively. The cross validation of ANN took long time while grid – searching of SVM was done quicker.

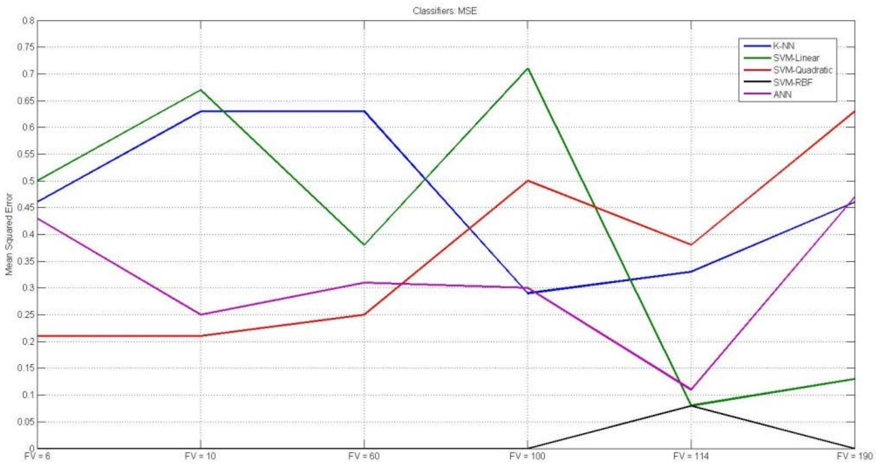

Figure 19: The Mean Squared Error Comparison between K-NN, SVM-Linear, SVM-Quadratic, SVM-RBF and ANN classification

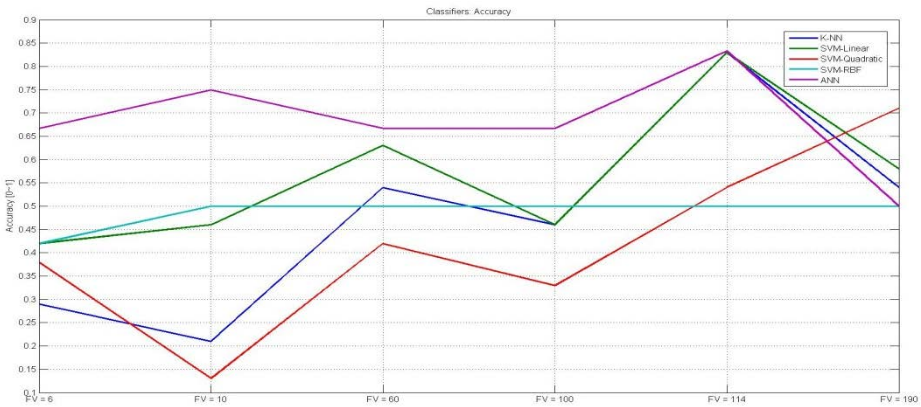

Figure 20: The Accuracy Comparison between K-NN, SVM-Linear, SVM-Quadratic, SVM-RBF and ANN classification

5. Conclusion

EEG-based movement imagery classification is one of the basic concepts used in the Brain Computer Interface systems. Several signal processing methods such as ERD – ERS, Welch PSD, Independent Component Analysis, etc. are used in the BCI. This course project was performed by Welch PSD based on Mean and Standard Deviation of EEG components, which involved motor cortex stimulation (delta – theta – alpha – beta 1 – beta 2). K – NN, three types of SVM, and ANN were trained by feature vectors of the mentioned EEG components. Because of the data set used in this work, the overall accuracy 70.5% (which was calculated by taking average of all classifiers values for FV = 144 [more reliable result]) was lower than literature (which was up to 90%), as expected. For data collection, it is highly recommended that subjects follow a standard instruction. Collecting data of several subjects will lead the classification to more accurate and reliable results. Typically, number of data and required speed of classifier and several other parameters would indicate which pattern recognition methods has to be chosen, and a compromise between the mentioned factors should be reached. Hence, the classifier must be selected based on number and type of data, and not only based on the classifier reputation in the literature. One classifier might have a high reputation in a high feature space, while it does not classify well with few training data. If there is no time restriction, cross validation and testing different classifiers can be helpful to define the most convenient method. To summarize, this report not only tried to implement a classification on EEG signals, but also made a comparison between several machine – learning methods.

References

[1] P. D. Welch, "The use of the fast Fourier transform for the estimation of power spectra," IEEE Transactions on Audio Electroacoustics vol. 15, pp. 70-73, 1967.

[2] A. P. V. Neethu Robinson, "A Wavelet-CSP Method to Classify Hand Movement Directions in EEG based BCI System," IEEE - ICICS, pp. 978 - 982, 2011.

[3] Athena Akrami, Soroosh Solhjoo, Ali Motie-Nasrabadi, Mohammad-Reza Hashemi-Golpayegani, "EEG-Based Mental Task Classification: Linear and Nonlinear Classification of Movement Imagery," Proceedings of the 2005 IEEE, pp. 4626-4629, 2005.

[4] Ernest Niedermever et al., Fernando Lopes da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Lippincott Williams & Wilkins, 2004.

[5] Marcel van Gerven, Jason Farquhar, Rebecca Schaefer, Rutger Vlek, Jeroen Geuze, Anton Nijholt, Nick Ramsey, Pim Haselager, Louis Vuurpijl, Stan Gielen and Peter Desain, "The brain–computer interface cycle (TOPICAL REVIEW)," JOURNAL OF NEURAL ENGINEERING, pp. 1741-1751, 2009).

[6] WU Jin, ZHANG Jiacai, YAO Li, "An Automated Detection and Correction Method of EOG Artifacts in EEG-Based BCI," IEEE, pp. 978-982, 2009.

[7] Saman Sarraf, Jian Sun, "ADVANCES IN FUNCTIONAL BRAIN IMAGING: A COMPREHENSIVE SURVEY FOR ENGINEERS AND PHYSICAL SCIENTISTS.," International Journal of Advanced Research, vol. 4, no. 8, pp. 640-660, 2016.

[8] Saman Sarraf, Cristina Saverino, Halleh Ghaderi, John Anderson, "Brain network extraction from probabilistic ICA using functional Magnetic Resonance Images and advanced template matching techniques," in Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference, Toronto, 2014.

[9] Saman Sarraf, Ghassem Tofighi, "Deep Learning-based Pipeline to Recognize Alzheimer's Disease using fMRI Data," in Future Technologies Conference (FTC), 2016.

[10] K. Saladin, Anatomy & Physiology: The Unity of Form and Function, New York: McGraw Hill, 2004.

[11] Dorland, Dorland's Illustrated Medical Dictionary, A Saunders Title, 2007.

[12] Clemens Brunner , Muhammad Naeem, Robert Leeb, Bernhard Graimann, Gert Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," Pattern Recognition Letters 28, p. 957–964, 2007.

[13] R. S. ,. M. R. G. P. Christa Neuper, "Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial EEG," Cognitive Brain Research 25, p. 668 – 677, 2005.

[14] J. D. B. H. Tao Wang, "Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns," Clinical Neurophysiology 115, p. 2744–2753, 2004.

[15] Meng Hu, Jiaojie Li, Guang Li, Xiaowei Tang and Qiuping Ding, "Classification of Normal and Hypoxia EEG Based on Approximate Entropy and Welch Power-Spectral-Density," in International Joint Conference on Neural Networks , Vancouver, 2006.

[16] Therrien C. W., Discrete Random Signals ans Statistical Signal Processing, New Jersey: Prentice-Hall, 1992.

[17] Guilherme A. Barreto, Rewbenio A. Frota and Fatima N. S. de Medeiros, "ON THE CLASSIFICATION OF MENTAL TASKS: A PERFORMANCE COMPARISON OF NEURAL AND STATISTICAL APPROACHES," Department of Teleinformatics Engineering, Federal University of Ceara, Ceara, Brazil, 2004.

[18] Valerie Morash , Ou Bai, Stephen Furlani , Peter Lin , Mark Hallet, "Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries," Clinical Neurophysiology 119 , p. 2570–2578, 2008.

[19] T. Lipping, R. Ferenets, P. Puumala, K. Suominen, E. Karvonen, "EEG Independent Component and Coherence Analysis from Scalp and Depth Electrodes During Propofol Anesthesia," in Proceedings of the 25th Annual Intemational Conference of the IEEE EMBS , Cancun, Mexico, 2003.

[20] Ahmad Khodayari-Rostamabad , Gary M. Hasey , Duncan J. MacCrimmon , James P. Reilly, Hubert de Bruin, "A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy," Clinical Neurophysiology 121, p. 1998–2006, 2010.

[21] BEN H. JANSEN, JOHN R. BOURNE, SENIOR AND JAMES W. WARD, "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis," Biomedical Engineering, IEEE, pp. 630 - 638, 1981.

[22] Saman Sarraf , Cristina Saverino, Ali Mohammad Golestani, "A Robust and Adaptive Decision-Making Algorithm for Detecting Brain Networks Using Functional MRI within the Spatial and Frequency Domain," in The IEEE International Conference on Biomedical and Health Informatics (BHI) , Las Vegas, 2016