T. C.
ALTINBAŞ UNIVERSITY
Graduate School of Science and Engineering
Information Technologies

INVESTIGATION OF SINGLE TRIAL EVENT RELATED POTENTIALS BASED DYNAMIC TIME WARPING USING MACHINE LEARNING TECHNIQUES

Kadhum Kareem Al-Rubaye

M. Sc. Thesis

Supervised by Prof. Dr. Oguz BAYAT

Istanbul, 2019

INVESTIGATION OF SINGLE TRIAL EVENT RELATED POTENTIALS BASED DYNAMIC TIME WARPING USING MACHINE LEARNING TECHNIQUES

by

Kadhum Kareem Al-Rubaye

Information Technologies

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2019

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of ....

Asst. Prof. Dr. Adil Deniz Duru	Assoc. Prof. Dr. oguz BAYAT
Co-Supervisor	Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Academic Title Name SURNAME	Faculty, University	___________________
Academic Title Name SURNAME	Faculty, University	___________________
Academic Title Name SURNAME	Faculty, University	___________________
Academic Title Name SURNAME	Faculty, University	___________________
Assoc. Prof. Name SURNAME	Faculty, University	___________________

I certify that this thesis satisfies all the requirements as a thesis for the degree of ....

Academic Title Name SURNAME
Head of Department

Approval Date of Graduate School of
Science and Engineering: ____/____/____

Academic Title Name SURNAME
Director

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Kadhum Kareem Al-Rubaye

DEDICATION

For My family, thanks to their support, to my teachers, my friends and all those who supported me, i say thank you all for your support and dedicate this work to you.

ABSTRACT

INVESTIGATION OF SINGLE TRIAL EVENT RELATED POTENTIALS BASED DYNAMIC TIME WARPING USING MACHINE LEARNING TECHNIQUES.

Kadhum Kareem Al-Rubaye,

M.Sc., Information Technology, Altınbaş University,

Supervisor: Prof. Dr. Oğuz BAYAT

Co-Supervisor: Asst. Prof. Dr. Adil Deniz DURU

Date: 9/2019

Pages: 60

Human brain electrical responses measured as Electroencephalogram epochs have different characteristics by means of amplitude and frequency content depending on the conditions and stimuli. Event-related potentials are the responses given to the stimuli and can be measured using the EEG. The average of these epochs are computed to remove the background activity and helps to exhibit the response to stimuli solely. In the concept of this study, dynamic time warping based connectivity features are used to classify the single-trial ERP epochs. Color Stroop test was implemented and ERP data are collected from 10 subjects. Support vector machine and K-NN classifiers are used and accurate classification results are achieved with the use of DTW metrics.

Keywords: EEG; ERP; SVM; KNN; DTW; Machine Learning

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

BCI	: Brain-Computer Interface
CV	: CROSS VALIDATION
DTW	: Dynamic Time Warping
ERPs	: Event-Related Potentials
FDTW	: Fast Dynamic Time Warping
KNN	: k-nearest neighbors’ algorithm
SVM	: Support Vector Machine algorithm
WPD	: wavelet packet decomposition

LIST OF SYMBOLS

α	: Alpha Wave
β	: Beta Wave
γ	: Gamma Wave
δ	: Delta Wave
θ	: Theta Wave

1. INTRODUCTION

Electroencephalography (EEG) is a noninvasive method that enables an immediate measure of human brain electrical activity via placement of electrodes on the scalp and collects mental information from this activity like eye movement of hearing music or watching different types of pictures. EEG technology has the ability to record brain electrical activity immediately with high accuracy. Through event-related potentials ERP analysis the EEG provides to researches to investigate psychological mechanisms of much mental behavior .the EEG signal analysis has many uses in various filed and applications, the most important filed that use the EEG signal is the health field which is uses the EEG signals to analyses the human brain activity for Alzheimer patient and epileptic seizure patient, by collecting these signals and analyses it, experts can predict the epileptic seizure .the previous researches aimed to analyses the EEG signal using many kinds of techniques that help to understand these signals and get valuable information from collected signals , these signals are not pure and contain a lot of noise the researches can't get what they need from these signals by a lot of approaches should implemented to clean these signals and process them to put them in suitable form which they can then implement classification techniques on it.

There is a lot of research in the classification of EEG signal to discover some patterns in these signals that help to extract important information about the human brain that helps in the health field. With the help of machine learning techniques and methods now it is possible to classify abnormal signals in the human brain and predicate of occur of this abnormal.

Raja Majid Mehmood and Hyo Jong Lee[64] they proposed a novel method of emotion recognition from brain signals using 14 EEG channels in the experiment aimed to elicit emotional reaction from volunteers while they watching emotional stimulus , in this experiment they defined four emotional states sad, happy, scared and clam they worked on possibility of emotion classification from EEG signal for five male subjects using SVM and KNN machine learning techniques and the highest accuracy for KNN was 61% while the highest accuracy they reach in SVM was 37%.

In[65] paper which uses EEG signals to detect the seizure activity automatically in their experiment, they focus on creating a matrix of multi-feature dimension involving Mean absolute value, Brownian motion and Root Mean Square as key features. Then they fed these features to machine learning to classify and predict it using SVM, KNN and ensemble bagged tree classifiers they succeed to reach accuracy 74.78% for SVM, 89.13% for Weighted KNN and 91.09% for Ensemble Bagged Tree.

The [66] which the authors aimed to study the semantic relationship between pairs of nouns of concrete objects and how this relationship activity is reflected in EEG signals. For feature extraction, they use many techniques, but the main technique was FDTW. They use many classification techniques involved SVM and the average accuracy for all the algorithms was 65.37 %. the SVM accuracy was 64.07%

Furthermore, of EEG classification the feature extraction is an important stage in EEG signal classification these techniques helps to extract valuable patterns from EEG data, in same research[66] Fast dynamic time warping technique used to extract the important features from the data to use this feature then for classification approach.

In [67] authors used Relative Spectral Band Energy, Harmonic Parameters, and Itakura Distance as feature extraction technique to achieve significant data reduction and to determine informative measures for automatic sleep staging .they found out that Itakura distance and central frequency seem to provide promising features for classification of sleep stages.

in their paper[68] the authors used wavelet packet decomposition as feature extraction for classifying EEG signals during the motor imagery tasks is investigated . all the previous research was for general EEG data, for single-trial ERP EEG signal classification there also many types of research have been done.

in [69] the researcher's inventive filters for multi-channel EEG that lead to signals which discriminate optimally between two conditions and they use the single-trial classification data to show the influence of this method and they reach 94, 90 and 84%, respectively.

AIM OF STUDY

The aim of the study is to perform single trial EEG time series classification based on dynamic time warping as a feature extraction technique. Several machine learning classification techniques have been adapted to identify the stimulus type that the participant was presented. For this, Stroop task is performed to collect multichannel EEG time series data. The time series were analyzed in 1 sec length epochs to determine if the stimulus is congruent or incongruent.

2. LITTERATUR REVIEW

The human brain is a great organ of the human body it's responsible of all his function and keeping him alive, the brain contains billions of communicated nerve cells that which are working together to process millions of task that. In last decades studying human brain become widely and there are many technologies invented to make studying of human brain easy and without surgery operation, EEG invented to make studies on the human brain and understand its behavior by connecting the brain to computers by cap wear on the scalp contain many electrodes that recording the brain activity . many types of research are focused on this recorded data by analyzing it and discovering important patterns and information that could help in many fields like neuroscience and cognitive science with of technology and appearing of machine learning we can do a lot of things with it in this filed .

2.1. BRAIN OF HUMAN

The human brain is the most complex organ of human. It controls All kind of physical functions. includes a huge neural network proximate of 10 billion neurons. It is responsible to deal with different man functions like feelings, hunger, thirst, body movements and sleep. Controls almost all important actions are necessary for the human to stay alive. the Brain can communicate with other body parts with exchange the single with them like sending orders to do some functions and receive feeling form outside environment. It is the crux of the central nervous system that involves the brain stem. large brain and spinal cord as shown in Figure 2.1. The spinal cord and the large brain are connected via brain stem. It is divided into three different sections based on anatomy and functions.

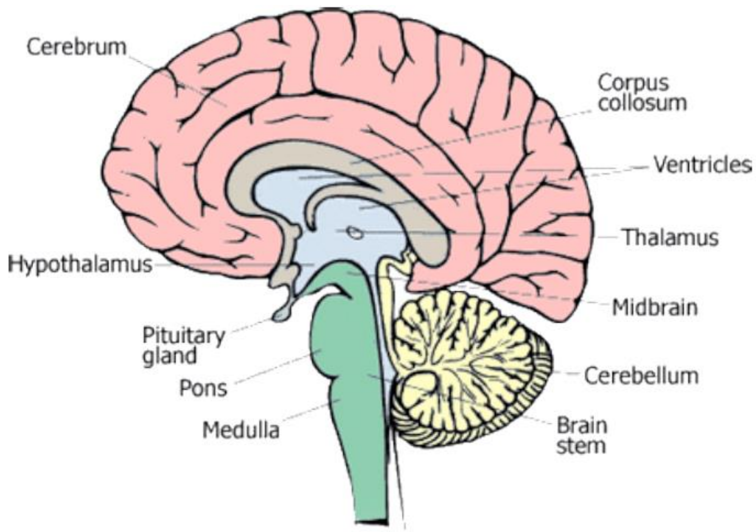

Figure 2.1: Different parts of the Human Brain [3]

2.2. ELECTROENCEPHALOGRAM

Electroencephalogram (EEG) is a Brain-Computer Interface (BCI) communication technique which helps to create nonmacular connection between the brain activities and the computers[1] by interpreting the human brain activity signals to electronic signals can be understood and analysis by computer with helping of many sensors (electrodes) which reflects electrical potentials induced by brain activities these electrodes are distributed around the skull and connected to the EEG device to transforming the signals [2]and arranged in order to electrode placement international system using the standardized electrode placement scheme Figure 2.2 [3]the EEG data are is multivariate time series[2] and measured in microvolts (mV).

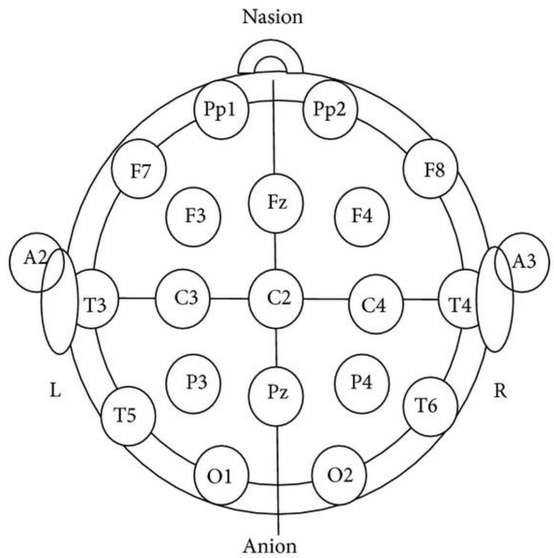

Figure 2.2: Standardized electrode placement scheme [3]

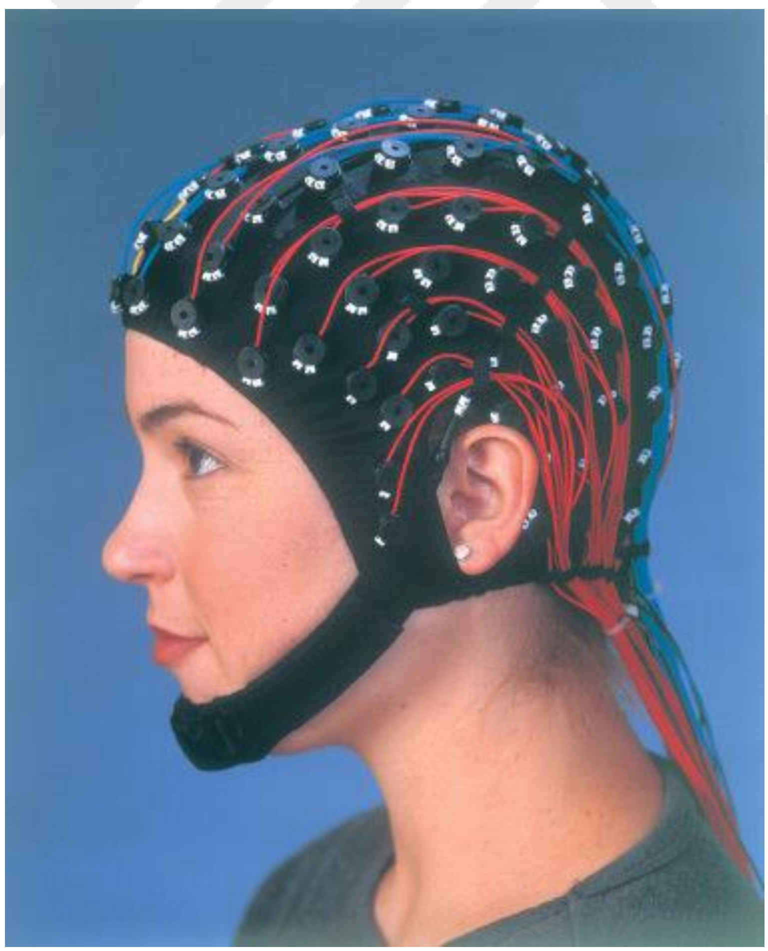

Figure 2.3: EEG device

2.3. EEG SIGNALS TYPE

EEG has many signals frequencies each frequency referring to human activity like relaxing, sleeping, focusing and deep sleeping.

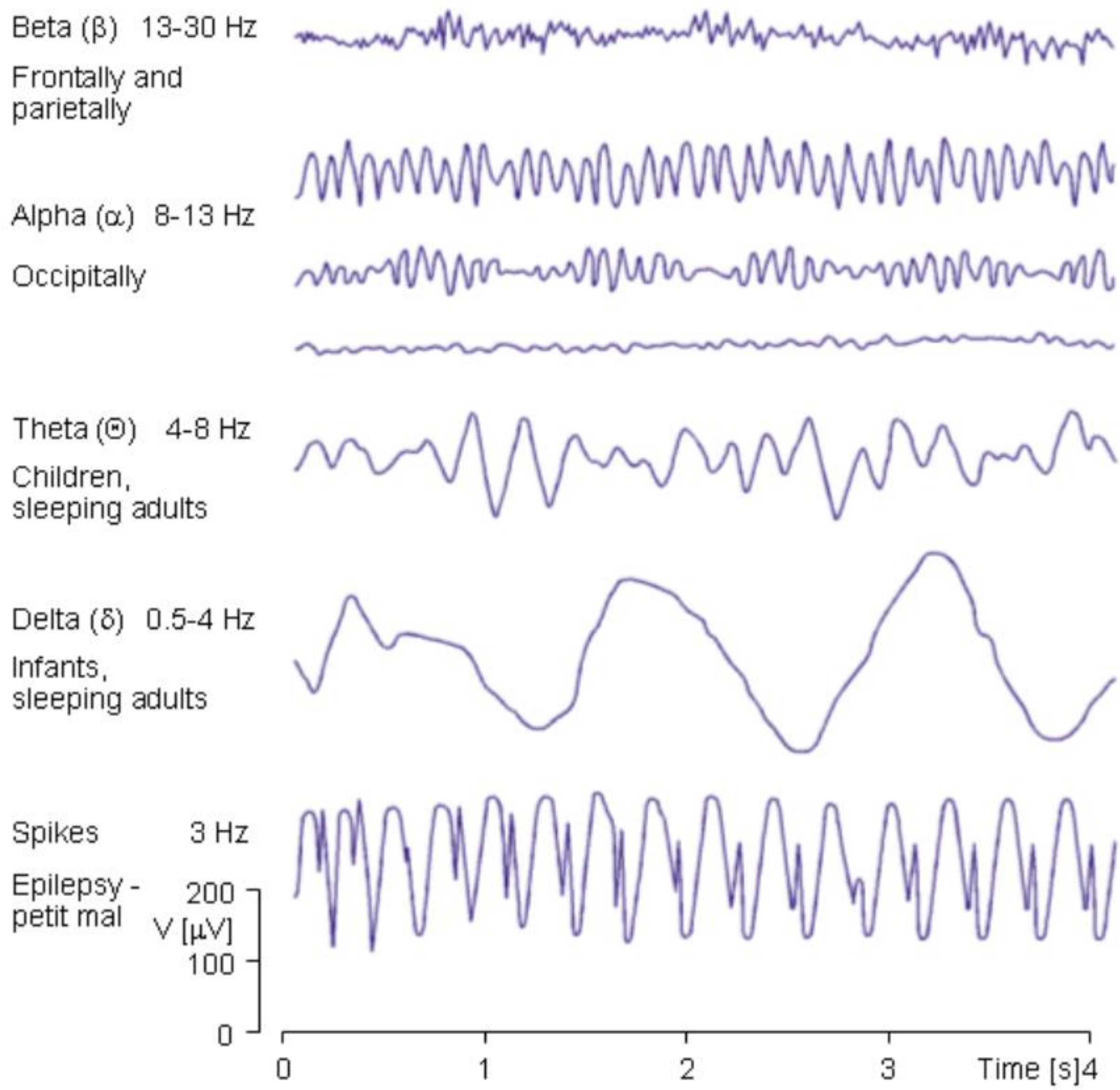

Figure 2.4: EEG Signals [4]

we can recognize five types of waves based on frequency ranges, they are alpha (α), theta (θ), beta, (β), gamma (γ) and delta (δ). Every specific lobe of the cerebral cortex has a specific type of wave however this is not always true. any mental states have its associated type of wave which it is useful to define one’s status at a specific time as shown in Table 2.1:

Table 2.1: Frequency and the Mental States of Waves

Wave	Frequency (Hz)	The Mental State
Delta (δ)	0 to 4	Represent deep Sleep
Theta (θ)	4 to 8	referring to drifting Thoughts, Dreams,
Alpha (α)	8 to 13	Referring to calmness, Relaxation, and abstract Thinking
Beta (β)	13 to 30	Highly Focused, Highly Alertness
Gamma (γ)	> 30	Simultaneous Process, Multi-Tasking

2.4. THE DELTA WAVES (Δ)

It has frequencies ranging between 0 and 4 Hz. The mental states associated with these frequencies are deep sleep, coma or hypnosis and sometimes they are awake. In the case of vigilance, it is always considered a satisfactory phenomenon. The higher the capacity, the more serious the problem. These waves are reduced with age and are usually present in healthy people in wakefulness.

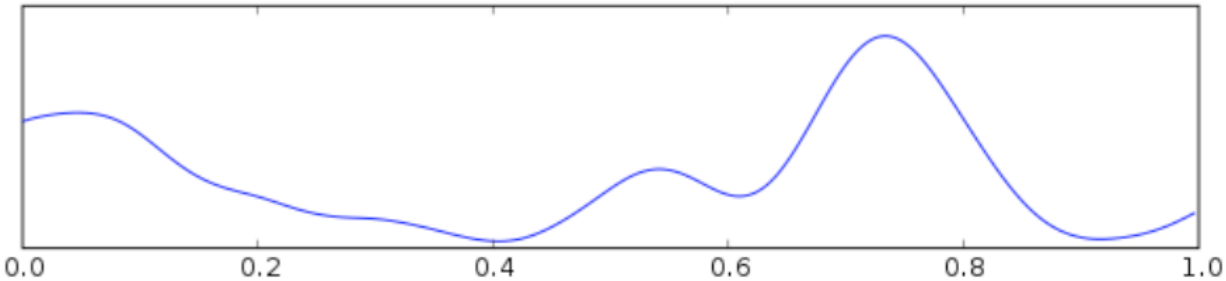

Figure 2.5: Delta Wave [4]

2.5. THE THETA WAVES (Θ)

Their Range between 4 and 8 Hz. The mental states that related to these waves are drift thoughts, creative thinking, and unconscious materials. These waves are founded in central, temporal and parietal parts of the skull. These waves are normal for healthy people as while they are in deep sleep.

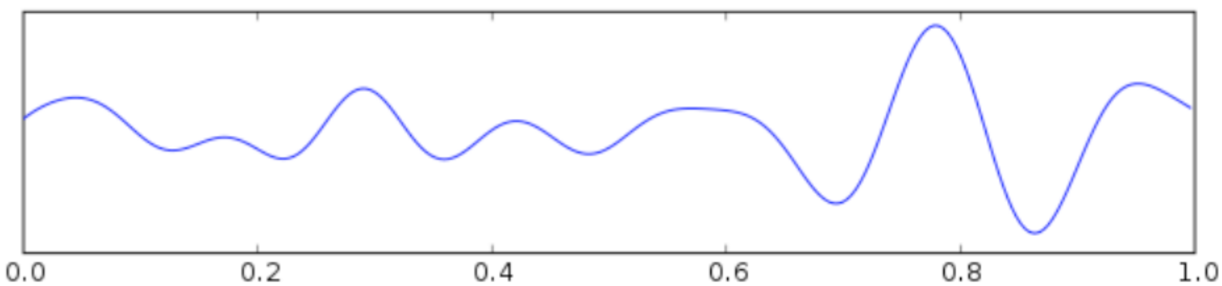

Figure 2.6: Theta Wave [4]

2.6. THE ALPHA WAVES (Α)

Alpha waves are in the range of 8 – 13 Hz. This wave represents the relaxed and calm for the mental states. These waves are present in the back part of the skull and in the occipital part of the brain. These waves have high amplitude compared to other types. This can be detected while the subject is in awake and clam cases. in sometimes conditions, these waves interfere with μ-rhythm. we find this wave is normal in people who are in relax and clam mode.

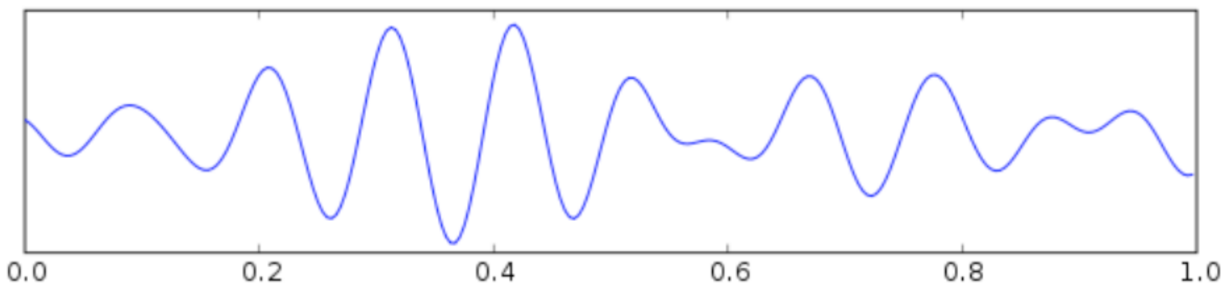

Figure 2.7: Alpha Wave [4]

2.7. THE BETA WAVES (Β)

They have a frequency range is fall between 13 – 30 Hz. We obtain this wave when the mental states of the subject associated with highly focused and alertness, for example, the deep thinking and concentration. these waves are having a large frequency band in compare to other waves. Beta waves appear in the front part of the scalp and the central part of the brain.

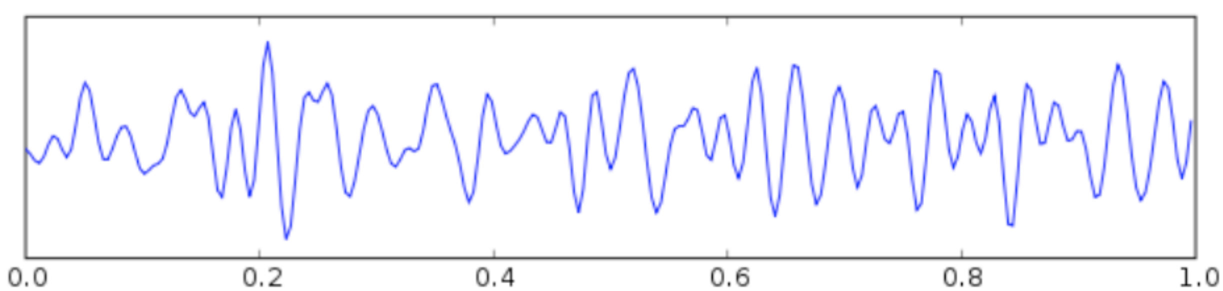

Figure 2.8: Beta Waves [4]

2.8. GAMMA WAVES (Γ)

Frequency band of this wave is falling under the range of 30 Hz. Mental states associated with these are synchronized work and the multi-tasking state. The very low amplitude makes this wave hard to notice. These waves seem in each portion of the brain.

2.9. THE 10-20 SYSTEM OF ELECTRODES PLACEMENT

It is one of the popularly used methods of electrodes placement which is standardized by the American Electroencephalographic Society. This system was founded by jasper [3] which is explains how the electrodes should be arranged on the head according to some condition which can found in [3] . in this system 21 electrodes are organized on the head as shown in Figure 2.9 The system of 10-20 is based on relationships between the position of electrode and the area behind the cerebral cortex. Each position indicates this figure to the left refers to a possible electrode location. Each site has a latter symbol (to identify the lobe) and a numeric number or another addition letter to identify the hemisphere location. The letters P, C, T, F, and O refer to Parietal, Central, Temporal, Frontal and Occipital. The even numbers (2,4,6,8) denote the right skull hemisphere and odd numbers (1,3,5,7) denotes to the left hemisphere. The z letter refers to an electrode located on the midline of the skull. Taking into consideration that the smaller the number, the closer the position is to midline spot. Nasion: the electrodes are arranged at the level of the eyes and top of the nose.Inion: placing the electrodes on the midline at the backside of the head and base of the skull are Inion. Parameters of the skull are measured from the above points. Locations for electrodes are determined by the division of parameters to intervals of 10% and 20%. Three electrodes are placed in the middle of adjacent points as shown in B of Figure 2.9 [5]. Relationship between the location of an electrode and cerebral cortex is the basis of the 10-20 system. Electrode letters are used to determine the placements such as:

A – Ear Lobe

C – Central Lobe

F – Frontal Lobe

Fp = Frontal Polar

O – Occipital Lobe

P – Parietal Lobe

Pg = Nasopharyngeal

T – Temporal Lobe

A collection of letters(s) and an integer number is used to specify the placement.

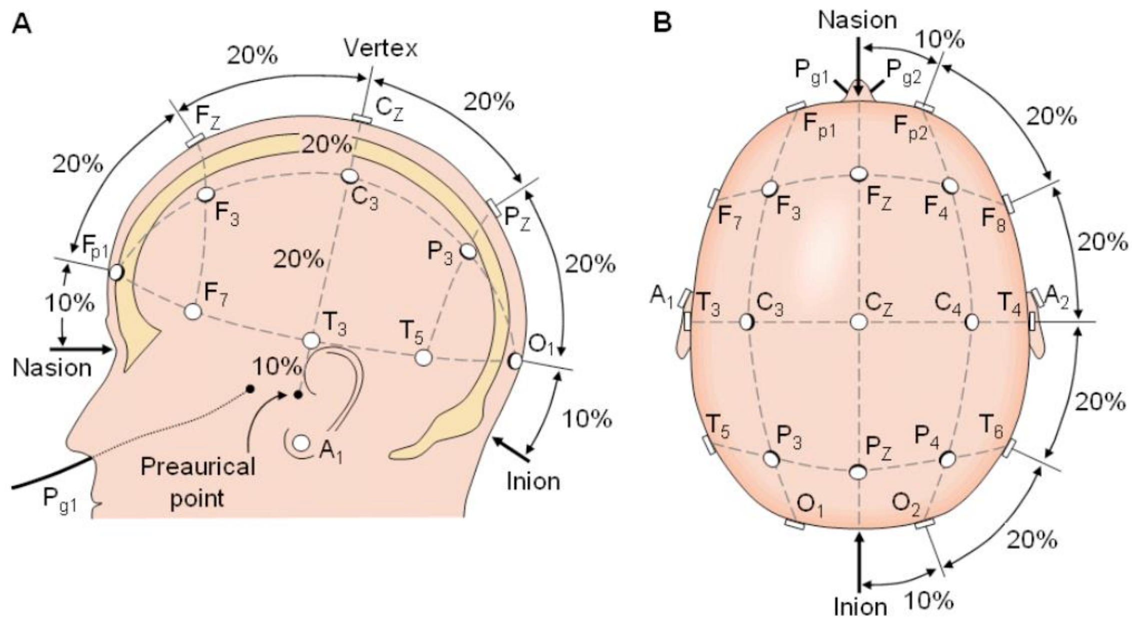

Figure 2.9: 10-20 System [5] Seen from (A) Left and (B) Above the Head

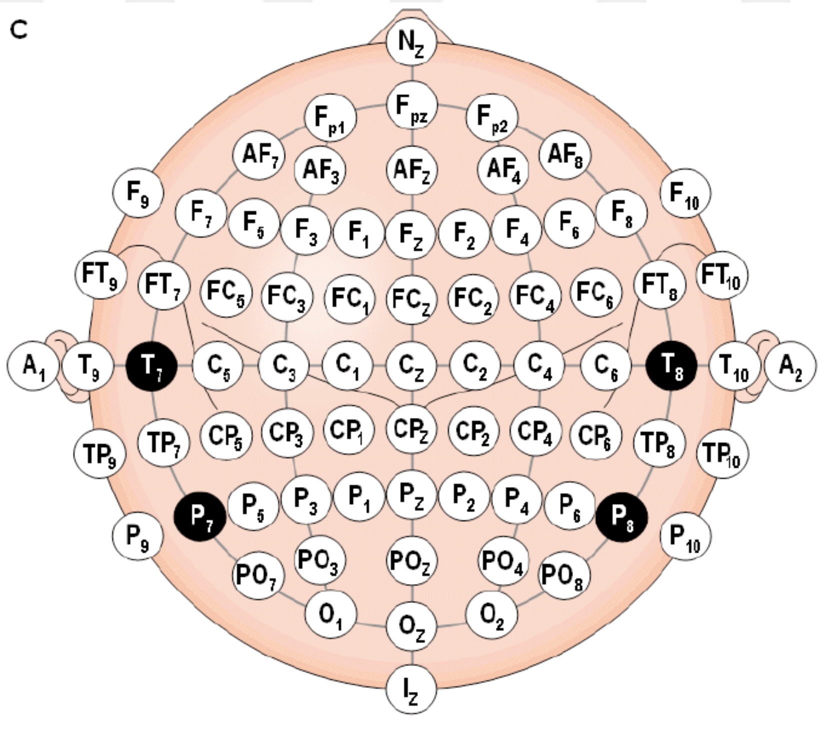

Figure 2.10: Position and Nomenclature of the Intermediate 10% Electrodes [5]

furthermore to 10-20 system, there are many other systems are available to get EEG signals. The Queen Square system was proposed as a standard for recording electric potentials on the skull [5]. EEG measurement can also be implemented using bipolar or unipolar electrodes. in bipolar case, potential differences between pair of electrodes are measured whereas in the unipolar case, the average of all electrodes is compared with the potential of every electrode [5].

2.10. EEG SIGNAL ANALYSIS

Various kinds of mental activities are utilized as input (BCI) systems, such activity kind is based on Event-Related Potentials (ERPs) the ERPs as their names refer are electrical potentials that related to specific events .ERPs consist of an average of single-trial and the single-trial means the first experiment on the subject-stimulus But averaging the single-trials may lead to losing important cognitive information [39].

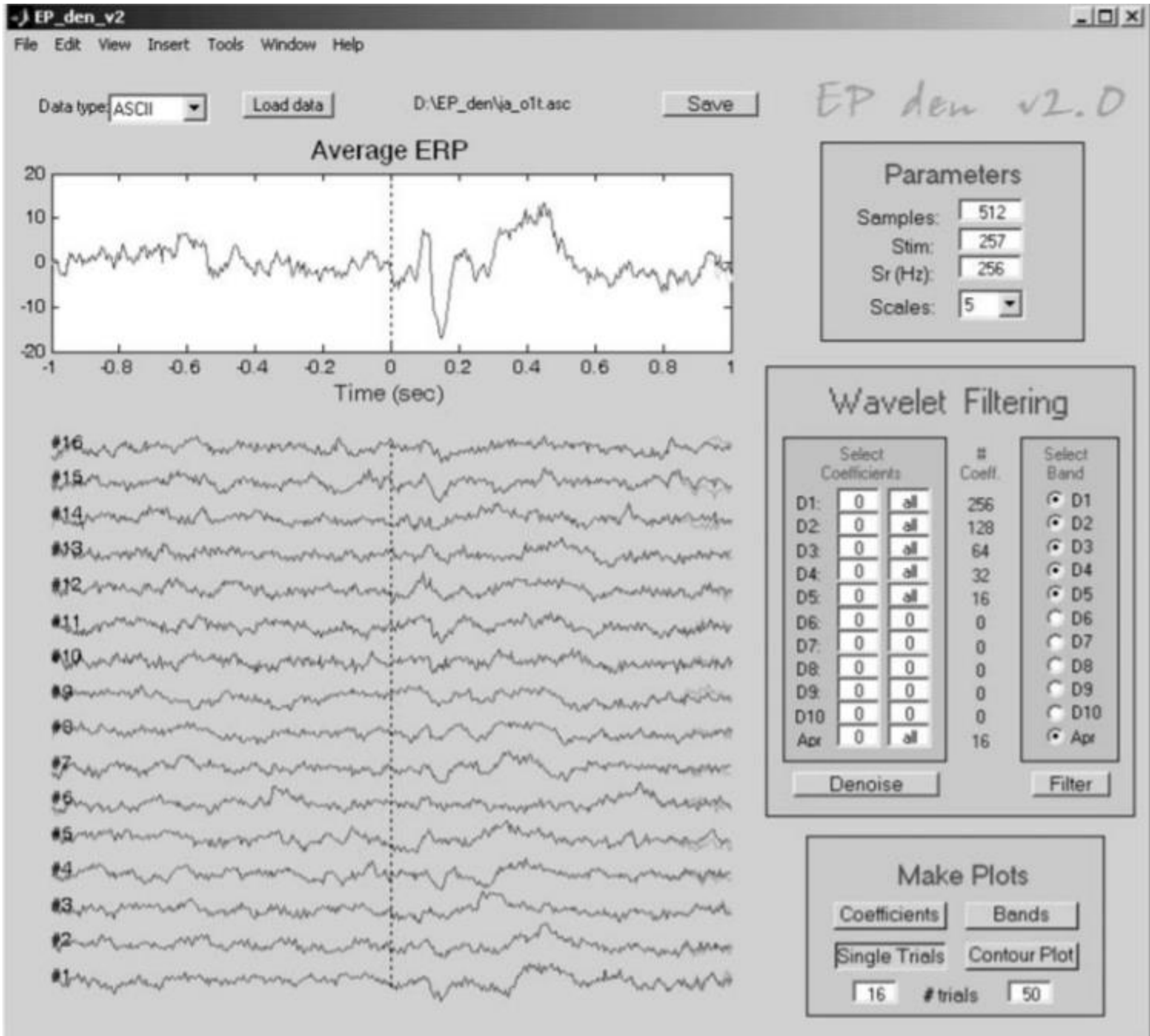

Figure 2.11: In the bottom single-trials signals and at the top the average of them which creates the ERP response[39]

2.11. CLASSIFICATION TECHNIQUES

In general, Classification is the simplification of a set of data by dividing it into the classes and assign each element into a class of its type, in the term of language classification means "systematic arrangement in groups or categories according to established criteria[20] Whereas in technique term precisely in datamining it means classifying every item in a dataset to one of predefined set of groups or classes .the classification in the data analysis is where a classifier or model is built for forecasting categorical labels [21]. Classification represents data extraction function that appoints elements in a group to categories or target categories. The objective of the classification is to accurately forecast the target category for every case in the dataset. For example, building classifier based on EEG data collected from many epilepsy patients that can predict Epileptic seizures date for the patient[40] There are a lot of classification technique that helps to create batter classifier depends on the type of the data like support vector machine, KNearestneighbor...etc. at first, we should prepare the data and this happened at the preprocessing approach, in this approach, we ensure that our data is clean and have no noise as possible as we can to help the classifier to predict the accurate result. The second approach trains the classification algorithm with the pretreated data to build intelligence model that can predicate the desire class label, in this approach we divide the data into training part which is training the algorithm to build the classifier and the test part that tests our model, after model building the evolution of model comes next to improve our model we have many accuracy metrics like, Holdout, cross-validation and bootstrap

2.12. MACHINE LEARNING

is a subfield of artificial intelligence (AI) that uses data mining technique to automatically discover patterns in data and create predictive model can learn from previous data, machine learning deal with algorithms that give the computers the ability to learn [27]In the context of data mining field, it is very important to take the correct approach for handling the task appropriately. It is often accomplished by a machine learning technique. A basic difference between human beings and machines has been for a long time that humans tend to handle with their problems and solve them automatically. The human being can learn from their previous experiences and they were trying to solve their problems by looking for new approaches by locating the problem carefully. The underlying machine programs can’t look at the output result of their task so they can’t identify their problems and they can’t make their behavior get batter. machine-learning locating this problem and trying to solve it by involving the computers programs so they can learn from previous data and improve their performances. So, the main goal of machine leering is too great intelligence computer programs that can learn by them self without human help. Since the 1990’s machine learning techniques are used in data mining fields, adaptive software systems like text and language learning fields. As a popular example: A computer application collection personal data about customers of e-commerce shop using machine learning model to improve the personal advertisements By learning from that data furthermore to improve the user experience to make the shopping approach simple and clear any ambiguously. Machine learning can be supervised and unsupervised, in supervised learning the machine learning model trained with labeled data[62], in another hand the training dataset in the unsupervised machine is not labeled and its depend on the model to discover the pattern in the dataset[62]. In supervised learning, the training set consists of pairs of input and desired output, and the goal is that of learning a mapping between input and output spaces. As an illustration, in figure 2.12. the inputs are points in the two-dimensional plane, the outputs are the labels as assigned to each input (circles or crosses), and the goal is to learn a binary classifier. Unsupervised learning: In unsupervised learning, the training set consists of unlabeled inputs, that is, of inputs without any assigned desired output. For instance, in Fig. (1.13), the inputs are again points in the two-dimensional plane, but no indication is provided by the data about the corresponding desired output. Unsupervised learning generally aims at discovering properties of the mechanism generating the data. In the example of Fig. (1.13), the goal of unsupervised learning is to cluster together input points that are close to each other, hence assigning a label – the clustered index – to each input point (clusters are delimited by dashed lines) [31].

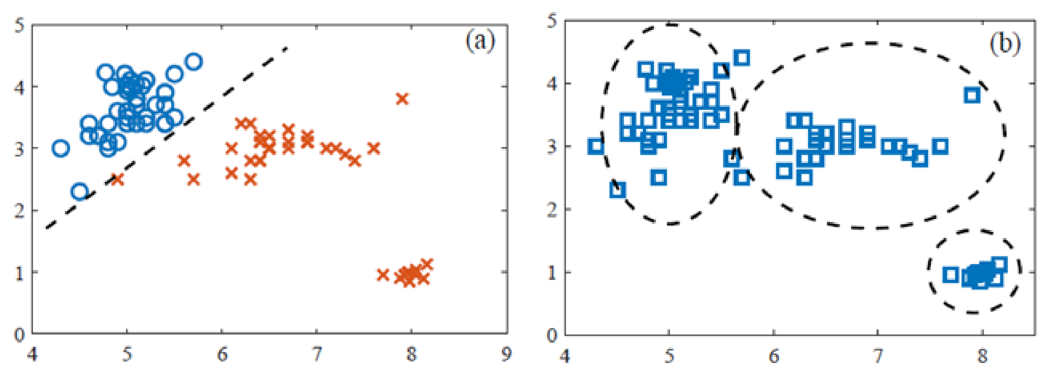

Figure 2.12: Representation of (a) supervised machine learning, (b)unsupervised machine learning

2.13. SUPPORT VECTOR MACHINES

Are supervised classification technique based on statistical learning theory.it was evaluated by Vapnik [25], SVM is supervised discriminative machine learning classifier formally defined by a separating hyperplane, it separating the class using flat plan within predictors space figure 2.13.

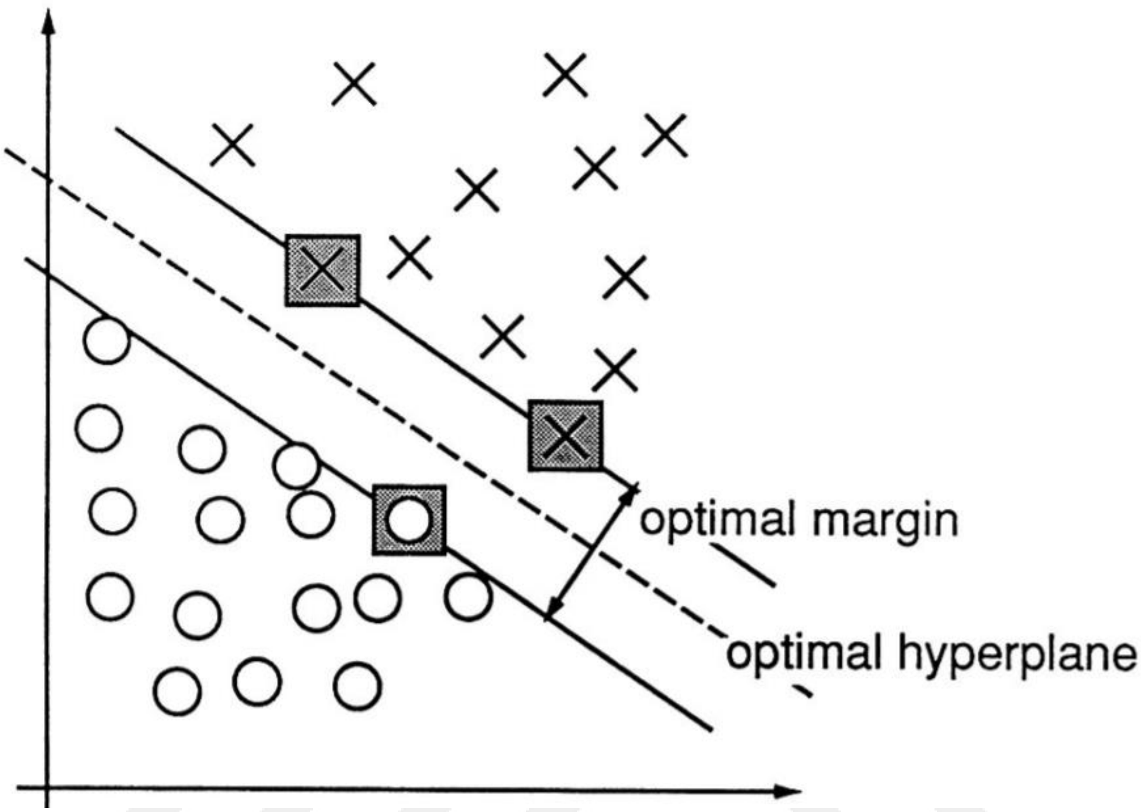

Figure 2.13: Illustration of a separable problem in a 2-dimensional space. grey squares refer to support vectors, define of largest separation margin between the two classes [33].

SVM can deal with classification and regression. the main idea of SVM is to find alignment hyperplane that can dive dataset into two classes. The SVM hyper-plane has maximum margin which represents the distance between support vectors and the hyper-plane. According to Vapnik theory, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified , SVM has many parameters that can inflict on classification result, by manipulating this parameter we can change the hyperplane shape and get satisfied result, the algorithm's name comes from support vectors, which are lists of the predictor values taken from cases that located closest to the decision boundary separating the classes . SVM has many parameters helps to find the best margin of hyperplane to sperate the class in best way on sides of the hyperplane and help the algorithm to get batter accuracy, regularization C parameter controls the trade-off between maximizing the margin and allowing for incorrectly classified training part[41] , gamma parameter is controlling the width of the Gaussian kernel. It specifies the scale of what it means for points to be close together and kernels, the function of the kernel is to take data as input and convert it into the desired form These functions could be different kinds. For example, linear, nonlinear, radial basis function (RBF), polynomial, and sigmoid. manipulate the values of the parameters can give different classification accuracy figure 2.14 .

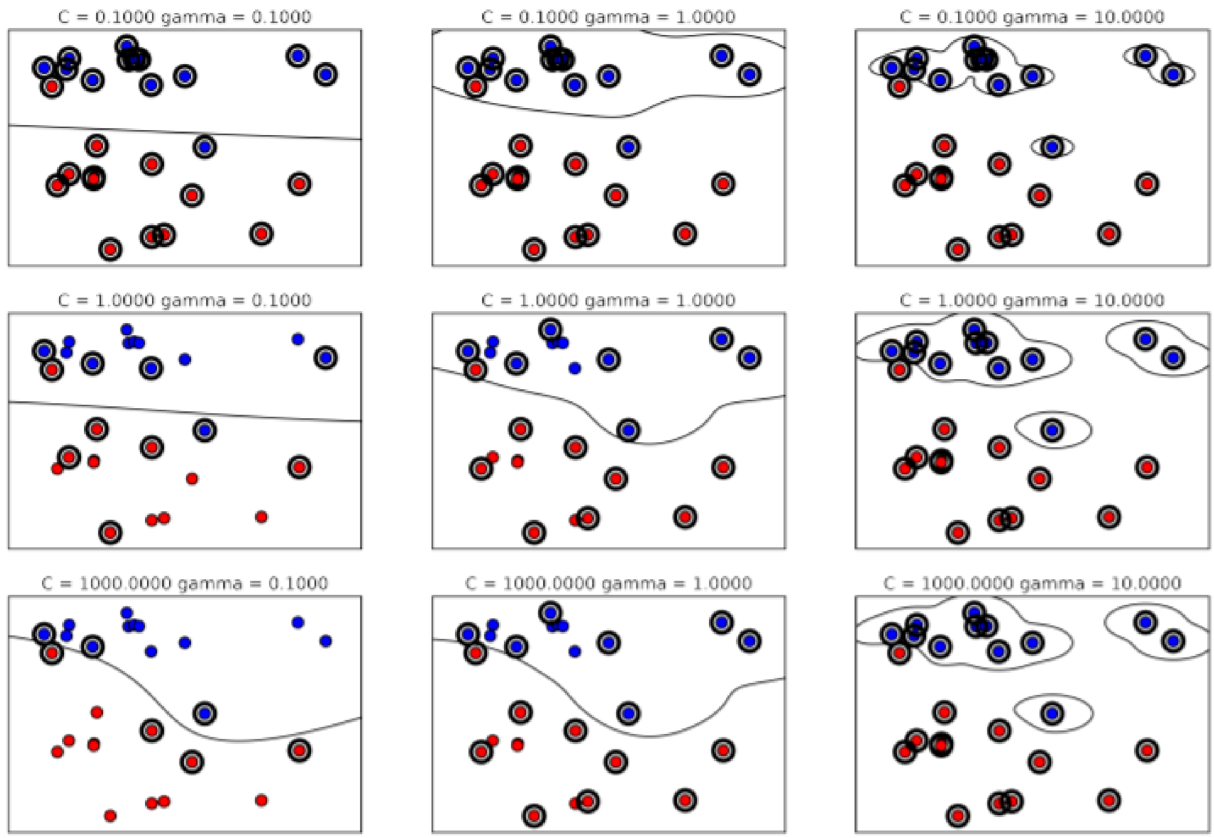

Figure 2.14: Supper Vector Machine parameters influence

2.14. K-NEAREST NEIGHBORS ALGORITHM (KNN)

K-nearest neighbor algorithm belongs to supervised learning algorithms group where the result of new instance query is classified based on the majority of K-nearest neighbor category, it is taking the minim distance between the input and k of nearest neighbor (k is any integer odd number). The purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory. Given a query point, we find K number of objects or (training points) closest to the query point. The classification is using majority vote among the classification of the K objects. Any ties can be broken at random. K Nearest neighbor algorithm used neighborhood classification as the prediction value of the new query instance. K-Nearest Neighbor (KNN) is one of the most basic and simple classification techniques

[28]. KNN is usually considered when there is no or very little knowledge available for the distribution of data. It is a potential non-parametric classification technique which completely bypasses the problem of probability densities [29].

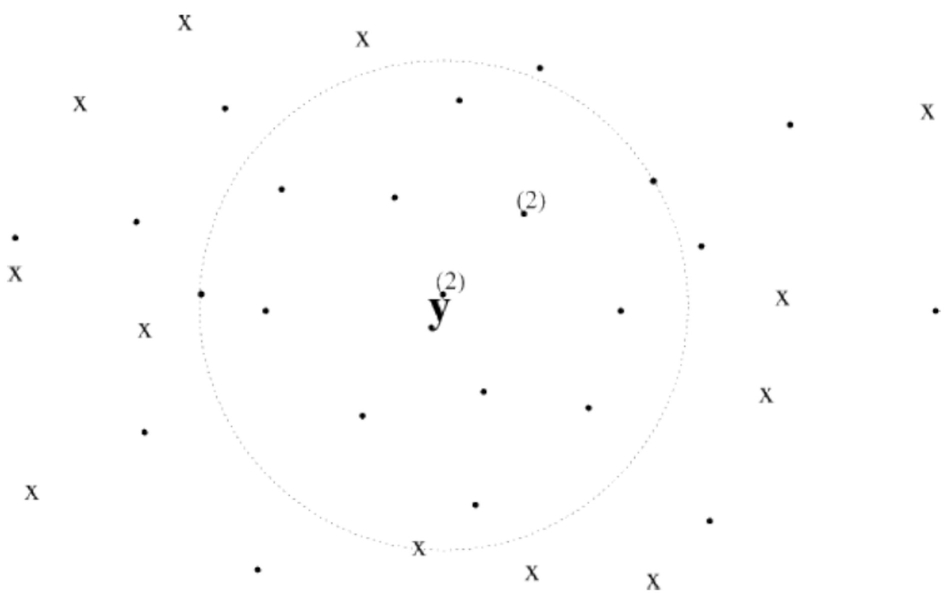

Figure 2.15: Classification using KNN [29]

KNN concludes two steps mainly. First, find the k-nearest neighbors, second calculating the distance by Euclidean distance given in equation 1.

$$S(A,B) = \sqrt{(A2 - A1)^2 + (B2 - B1)^2} \tag{2.1}$$

2.15. DATA PREPROCESSING

Data preprocessing is an important step in machine learning before making the classification approach, the real-world data often obtained clean and arranged and in most time it comes with noise, missed values and variance range of values , the standard steps of preprocessing approach involve dataset creation, data integration, data cleaning, feature selection and normalization [62].

2.16. NORMALIZATION

Normalization is a preprocessing method applied on data to prepare it for machine learning. The main goal of normalization is to change the values of numerical feature in unified scale without destruction variations It is aimed to give all the attribute equal weight [62].

2.17. Z-NORMALIZATION

or zero-mean normalization it uses the mean and standard of values for normalization.

$$v_i^` = \frac{v_{i-\bar{A}}}{\sigma_A} \tag{2.2}$$

2.18. CROSS VALIDATION

There are many model selection techniques that split the data to train-test [54]. Because we have a large amount of data it's preferred to apply k-fold cross-validation technique [54] to reach batter accuracy of the classification model. Cross-validation (CV) is a popular technique for tuning hyperparameters and producing robust measurements of model performance. Two of the most common types of cross-validation are k-fold cross-validation and hold-out cross-validation .in hold-out method Before classification approach we split our dataset into two portions : test portion and train portion , every portion taking some percentage of the dataset amount ,In k-fold This technique involves randomly splitting the dataset into k partitions or folds of approximately equal size. The first fold is kept for testing and the model is trained on k-1 folds the classification algorithm is evaluated by the average of the k accuracies resulting.

2.19. GRIDSEARCHCV

GridSearchCV scikit-learn's[57] algorithm finds the optimal hyperparameters to train a model based on a scoring function that it is given to assess its performance. The usefulness of GridSearchCV lies in the fact that the algorithm runs many iterations of a model while testing a different combination of hyperparameters with each iteration[58] This negates the need to run the model manually to test hyperparameters. GridSearchCV then runs an exhaustive search for the best combination of hyperparameters by iteratively running through all possible combinations that were specified to the algorithm, picking the best hyperparameters.

2.20. PERFORMANCE MEASURES AND EVALUATION

We can measure the and performance of the classifier and classification approaches using many statistical methods every method give us a report of a specific function in the classification process.

2.21. F1-SCORE

F1 Score is the weighted average of Precision and Recall. Thus, this score takes both false positives and false negatives into account.

$$\frac{2 * p * r}{p + r} \tag{2.3}$$

2.22. PRECISION

Also named positive prediction value it is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$\frac{tp}{tp + fp} \tag{2.4}$$

2.23. RECALL

It is the value of True Positives divided by the value of True Positives and the value of False Negatives.

$$\frac{tp}{tp + tn} \tag{2.5}$$

2.24. CONFUSION METRIX

A confusion matrix is a table that is used to explain the performance of the classifier on a set of test data for which the true values are known.

		Predicted class		
		yes	*no*	Total
Actual class	*yes*	*TP*	*FN*	*P*
	no	*FP*	*TN*	*N*
	Total	P'	N'	$P + N$

Figure 2.16: confusion matrix

2.25. ACCURACY

Accuracy is the most obvious performance metric and it represents the total number of predicted values divided by the total number of test data

$$\frac{tp + tn}{tp + fp + tn + fn} \quad (2.6)$$

2.26. DYNAMIC TIME WARPING

DTW is a well-known technique to get an optimal alignment between the two-time series.

2.27. TIME SERIES

is a series of indexed data points in chronological order. Most commonly, the time series is a sequence that is captured at equally spaced points in time. Thus, they are a series of data at a separate time.

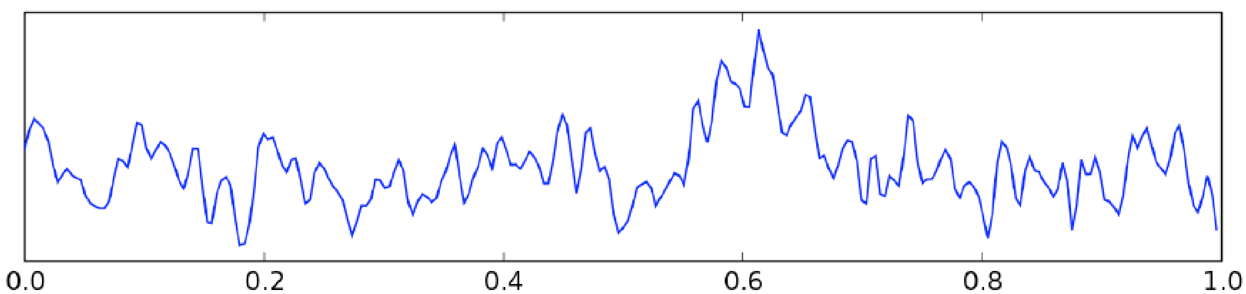

Figure 2.17: EEG time-frequency signals

2.28. SIMILARITY MEASUREMENT

suppose two time series A = {a1, a2, a3, …, am} and B = {b1, b2, b3, …., bn}. A distance/ similarity measured (A, B) gives us the 'distance/ similarity' between the two-time series A and B .there are more than one distant/similarity metrics and algorithms (Euclidean distance, Manhattan, … etc.). but the use of these metrics can be impropriated as well it is sensitive to offset, amplitude, noise, scaling, phase shifts and temporal distortion [41]The first two issues can be reduced by normalization of individual time series, but the other problems contains. The traditional methods of measuring the matching between two-time series are by comparing two corresponding points gives intuitive result because of it my comparing two points that are not corresponding well [63]. But Dynamic Time Warping produce the solution for this problem by creating optimal alignments between the two-time series Figure 2.18 .

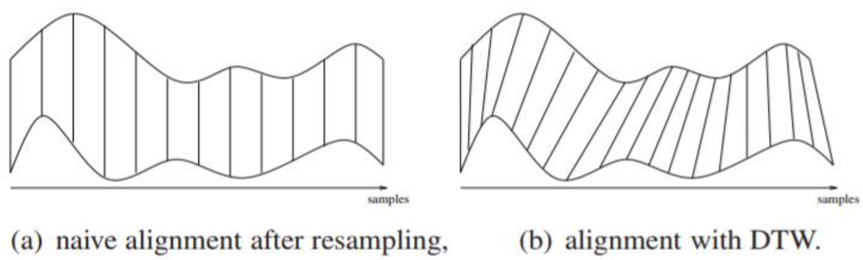

Figure 2.18: Two-time similar series have different alignments[63].

2.29. DTW

DTW[15] algorithm was released for the first time in the area of speech recognition was familiar [16]. Intuitively, the time series are warped in a non-linearly form to match one other. dynamic time warping technique used widely in many since fields, manufacturing, data mining, applications in DNA matching, speech recognition, and gesture recognition. DTW has been successfully applied to automatically handle with time distortion and different speeds associated with time-dependent data. The DTW technique has also been found functional in many other area [17]. DTW was successfully used to Retrieve information widely [18]. let’s suppose that we have two time series A={a1,a2,a3,….,am} and B = {b1,b2,b3,…,bn} the two time series can be arranged on the sides of n-by-m grid called (cumulative cost matrix), with one of time series on the top and the other one up the left side. Both time series start on the bottom from the left of the grid figure (1.18).

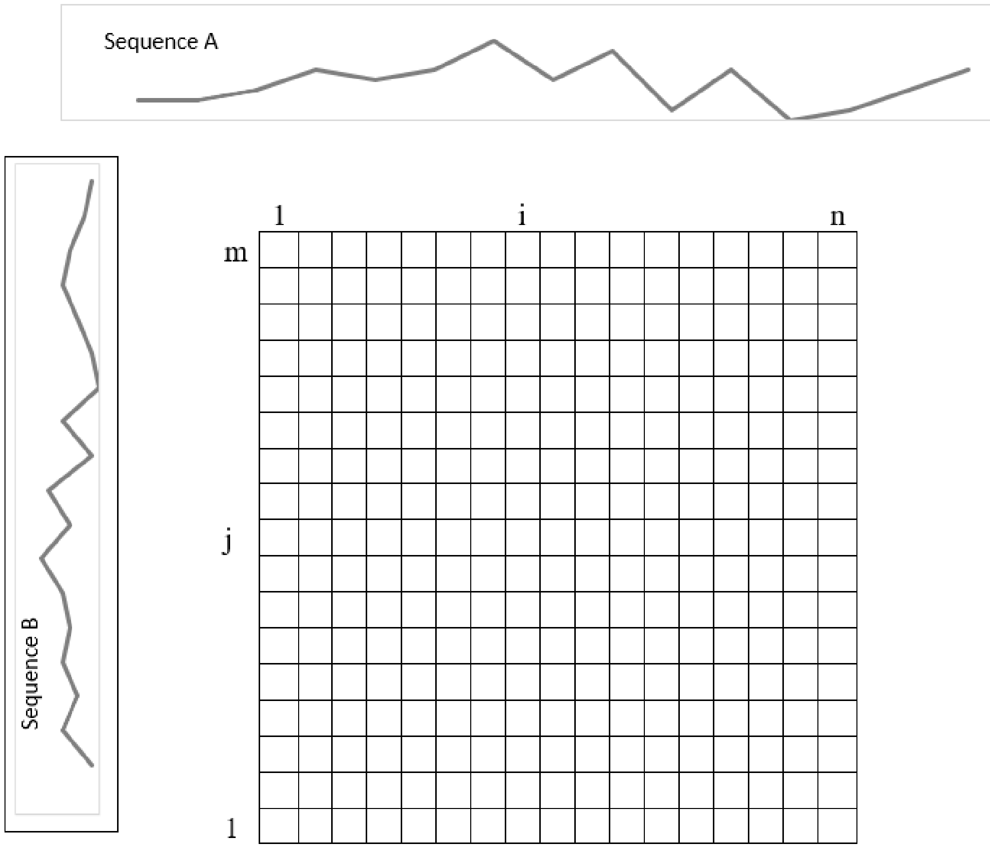

Figure 2.19: Arrangement of time series on the sides of the grid.

Each cell (i, j) in the accumulated cost matrix, (where i represents the index of the time series A, j represents the index of the time series B) contains the alignment distance result d(i, j) between the two time series A_i an B_i points which is can be calculated by Euclidian distance(1.1).

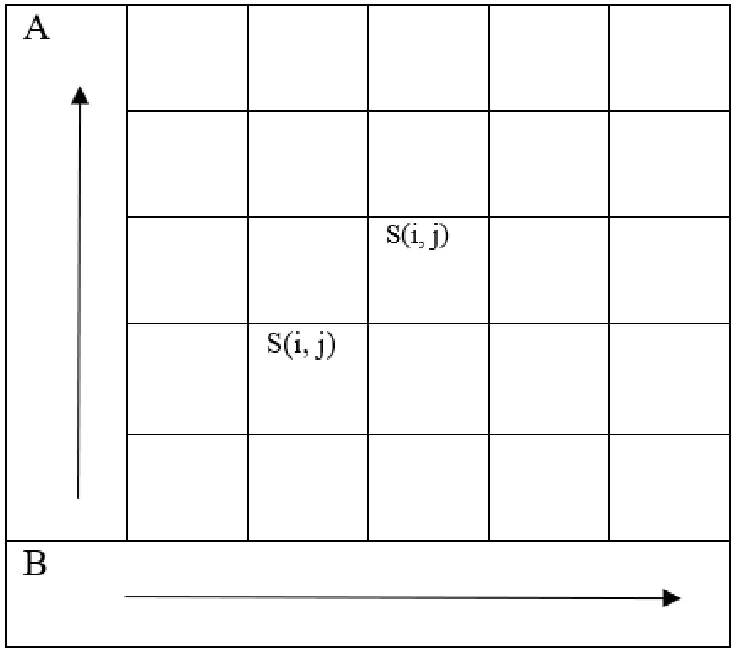

Figure 2.20: represent the matrix that contains the result of the distance between two-time series pints.

After obtaining the accumulated cost matrix, the optimal alignment path W is found by using the DTW equation (Equ 1.7).

$$D(i,j) = \min \begin{Bmatrix} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{Bmatrix} + d(\, x_i, y_i) \qquad (2.7)$$

A warping path between A and B is the a minimal among all the possible warping paths. A warping path W figure 2.20 , is a adjacent set of matrix elements that assigns the elements of A and B to each other. $W_{k.}$ = (i, j) k is defined as the K^{th} element of W and W = (w1, w2, .., $w_{k.}$), max(N, M) ≤ K

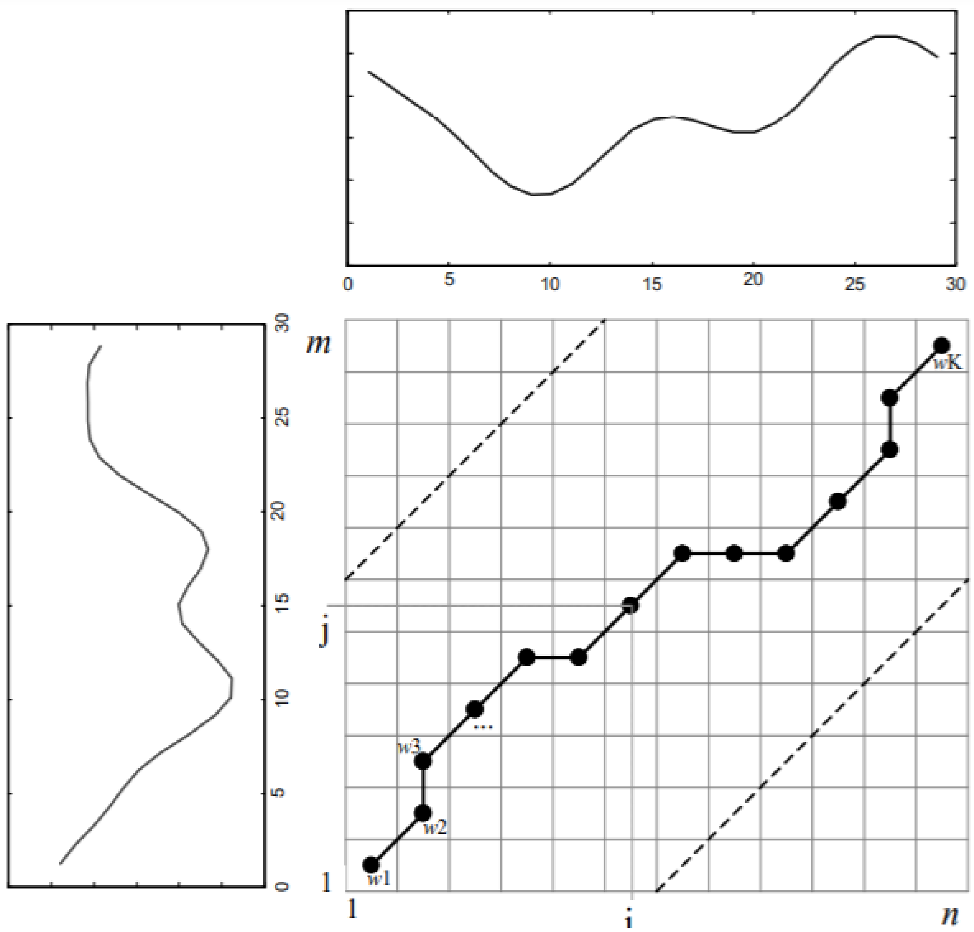

Figure 2.21: The warped path between two-time series[15]

Where d(.,.) is Euclidian distance(1.1) the warping path typically subjected to a bunch of restrictions control its implementation and improve its work and these restrictions are:

1.monotonicity

This condition constraints the warping path from long jumps while matching the series. The following condition imposes the points in W to be monotonically spaced in time. Given that wk = (i, j), then wk−1 = (i', j'), where i − i > 0 and j − j > 0.

2. continuity

This constrains making the steps in the grid are Limited to neighboring points only, ik-ik-1 ≤1 and jk-jk-1≤ 1.

3. warping window

This condition makes allowed points can be limited to fell inside a given warping window $|ik-jk| \leq w$ where w is a positive integer window width.

4. slope constraint

Allowable warping paths can be constrained by restricting the slope, thereby avoiding excessively large movements in a single direction.

5. boundary conditions: The first and last elements of X and Y are matched. The warping path starts and ends in the diagonally opposite corner of the matrix D, namely, w1 = (1, 1) and wK = (N, M).

The warped path represents the best alignment path between the two-time series.

3. METHODOLOGY

In cognitive neuroscience application it's important to know the mental status of subjects like whether its congruent or incongruent to and that can be discovered by helping of BCI techniques like EEG which is used in the experiment to recording the brain signals and extract the single-trial data then extract features for the classifier using DTW technique , and using of machine learning techniques by Applying the classification algorithms on our collected data to build classifier can predict the congruent and incongruent status of subject who are subjected to the Stroop's task, and with using the high modern performance computers to deal with huge numbers of data and programming techniques it's become easy to make such experiment easily and get satisfied results.

3.1. PARTICIPANTS

The mean age of the volunteers was 25 and all neurological, psychological, chronic, status for the volunteers was healthy . 10 of the volunteers were elite and the control group. Participants, Marmara University Sports Academy This study carried out in the Department of Marmara the ethics committee of the University Faculty of Medicine 2013-0191 with consent and voluntary consent They participated.

3.2. EXPERIMENTAL DESIGN AND STIMULI

One of the gold standards of attention measurement is Stroop the experiment has been used in the literature for a long time. Stroop test using visual stimuli, color names used as stimulus. Color names, with its meaning Available incompatible colors or in distracting colors. If there was color matching with writing the right arrow key of the keyboard is pressed, if there is no color matching the left arrow key pressed. While they flashing on screen. For measuring Stroop performance as well as electrophysiological responses, the duration of their reactions is calculated. An example of the experimental design is presented in Figure (3.1). The experiment design was prepared by the MATLAB program within the body of Marmara University School of Physical Education and Sports. The stimulating presentation software is coded to send the moments of the stimuli momentarily to the device that collects the EEG data. The experiments were performed on subjects who were interested in karate sport on the elite level and in non-elite control group volunteers in any sport.

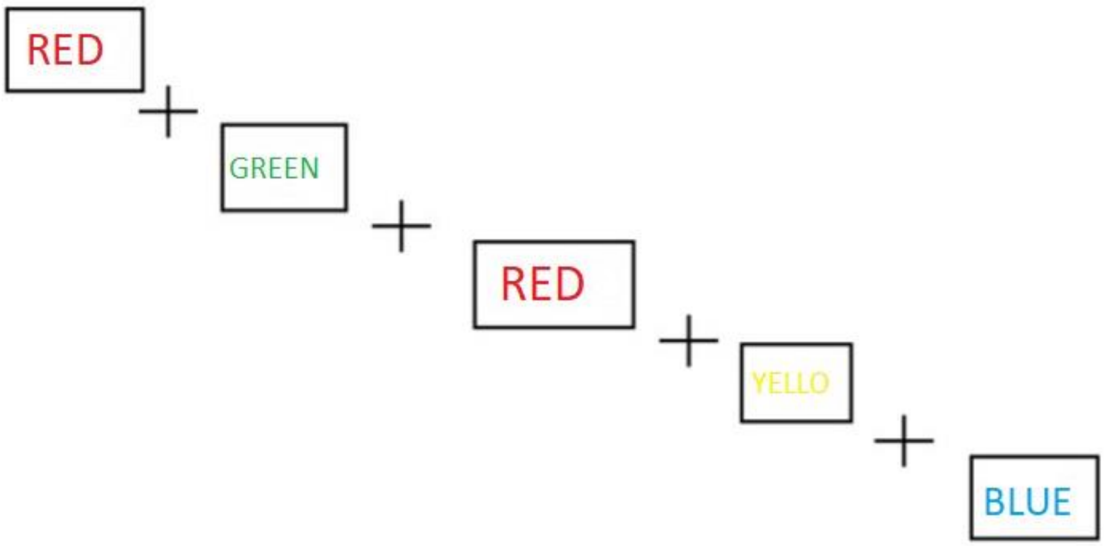

Figure 3.1: Stroop sample design from experimental design

3.3. EEG RECORDING

EEG recording of V-Amp with BrainAmp software 1000 Fp1, Fp2, F3, Fz, F4, FCz, T3, Cz, Ag / AgCl from electrodes T4, CPz, Pz, P7, P8, C3, C4, Oz Left ear nozzle reference through electrodes and right the earlobe is considered as grounding It was performed. Contains EEG artifacts from eye movements regions were selected and not taken into account in further analysis.

3.4. TOOLS EMPLOYED

in this experiment we use python programming language and its data since libraries to prepare data and feature extraction , for dynamic time warping algorithm we using (dtaidistance) library[13] c version (which enable much faster DTW alignment) , (networkx) library[14] for figuring DTW similarity , (scikit-learn) for classification and data preprocessing and other data since libraries for data manipulating In our experiment we collected EEG data of single-trial ERPs for two status of subject (congruent and incongruent) .

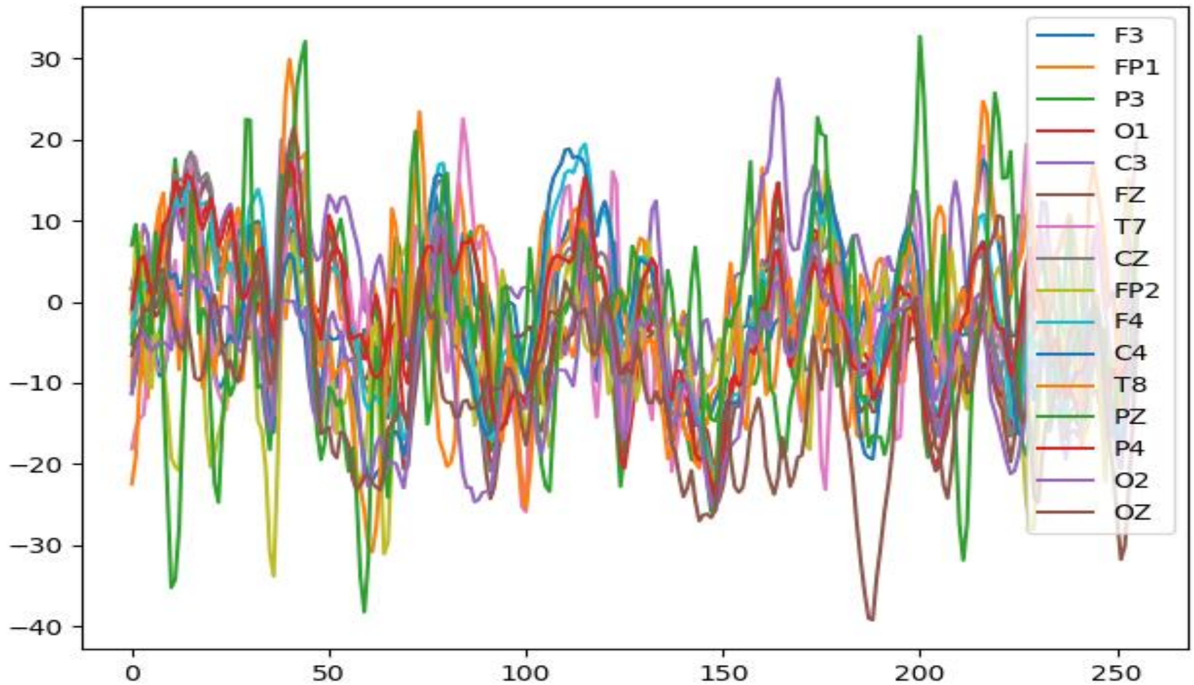

Figure 3.2: Congruent subject single-trial EEG signal frequency for one second

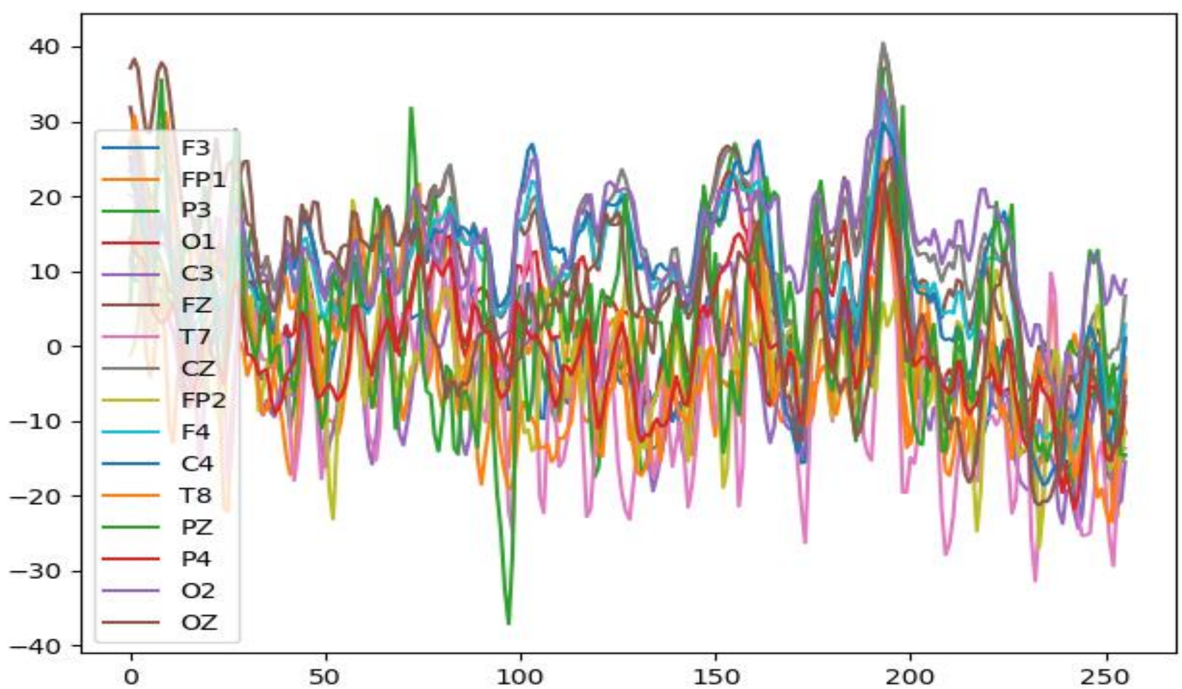

Figure 3.3: Incongruent subject single-trial EEG signal frequency for one second

Every channel has a different frequency for every status as shown in the figures (3.4) and (3.5).

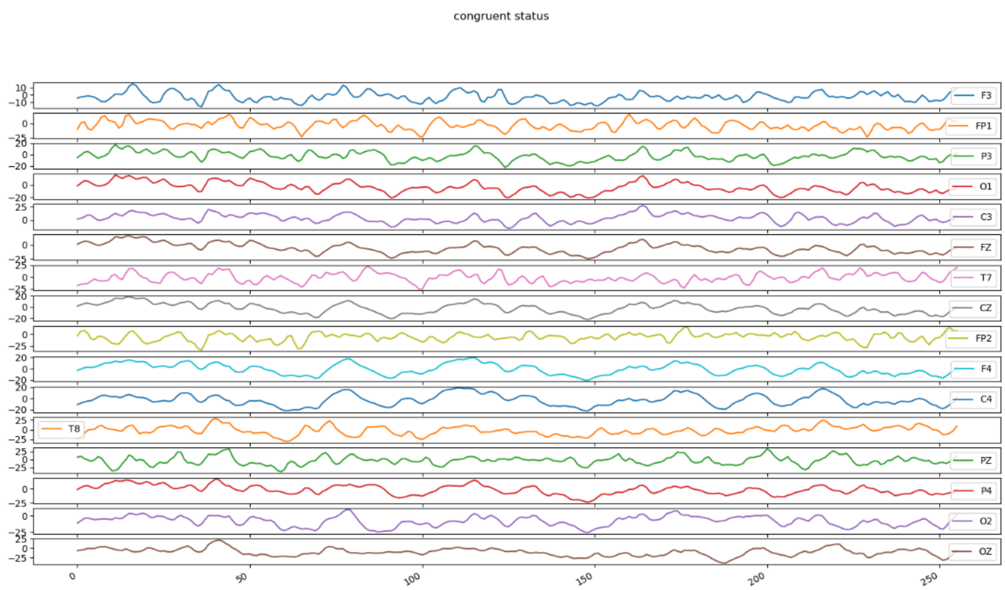

Figure 3.4: EEG single-trial ERP signals incongruent status

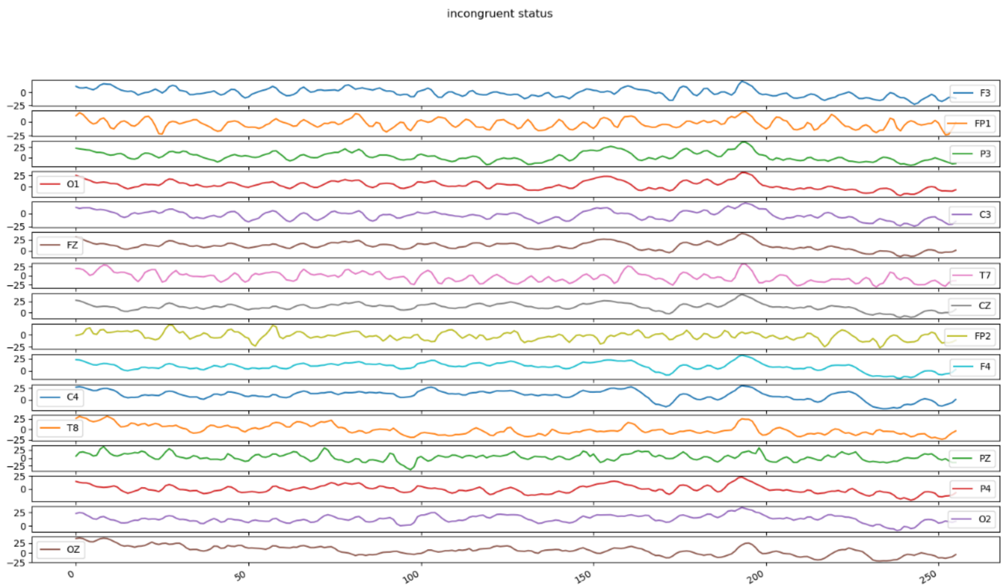

Figure 3.5: EEG single-trial ERP signals in incongruent status

3.5. FEATURE EXTRACTION FROM EEG DATA

EEG signals are complex, it's very hard to extract information from it by traditional way. Nowadays, by helping of computers, we can apply complex automatic processing algorithms that allow us to extract 'hidden' information from EEG signals. the feature extraction basically helps to increase classification accuracy [43]. There are many techniques such as time-domain features (mean, standard deviation, entropy, …)[44], frequency domain features (Fourier transform, wavelets, …)[45] and finally synchronicity features, which try to looks to the relationship between 2 or more EEG channels (coherence, correlation, mutual information, ….).

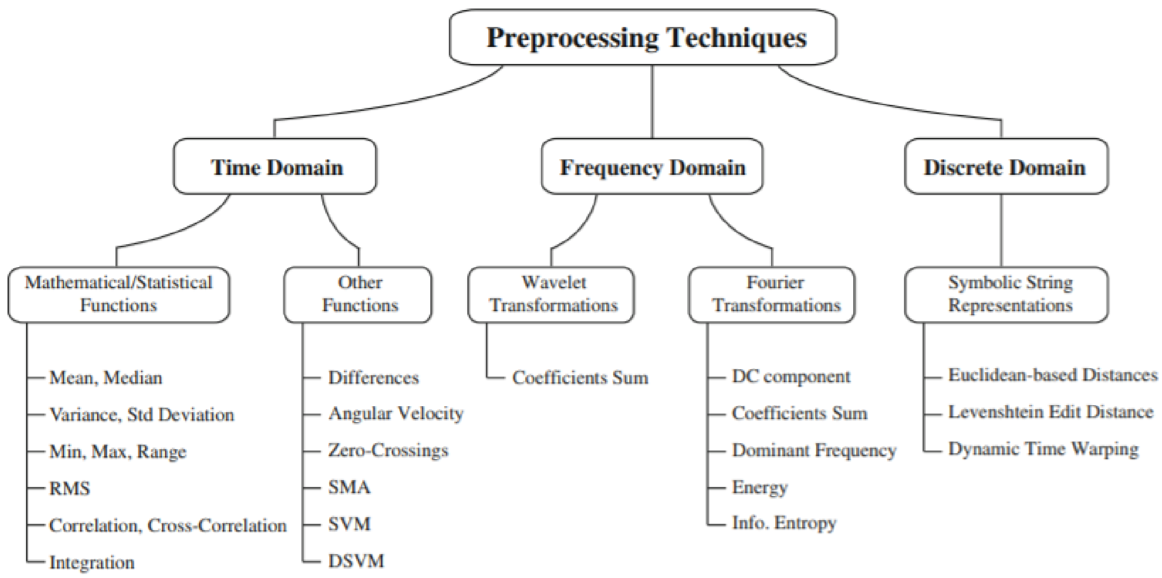

Figure 3.6: Classification of techniques applied to sensor signals for feature extraction [60]

most of the research of extracting synchronicity features are made on EEG data like feature extraction for detecting error-related potentials [46] and emotion detection EEG data feature extraction[47], this research will use single-trial EEG data to extract synchronized feature using DTW algorithm technique.

3.6. DYNAMIC TIME WARPING

DTW is similarity measure technique that helps to find the best alignment between two-time series DTW compresses and extends the time axes of pairs of digitized waveforms to reduce the effects of minor differences in shape due to noise and normal, random shape fluctuations[37]].To find the synchronized channels of single-trial EEG signal there are many techniques was applied[48] The DTW algorithm helped us to extract features that prove the relationship between EEG channels to detect the synchronized parts of the brain that works when the subject is incongruent or incongruent status

3.7. DATA PREPARING

We divided collected data to two groups each group contain 200 dataset each dataset has 16 feature representing 10-20 EEG system electrodes and every simple is represent one millisecond of single-trial data for one subject recorded from EEG , group 1 contain congruent dataset collection and group 2 contain incongruent dataset collection, both groups of dataset belongs for the same subject.

3.8. DATA PREPROCESSING

To make useful comparisons between two time series, both series must be normalized[61] Before we applying dynamic time warping method on data we normalized each series using z-normalizing method, it is not sufficient to normalize the whole dataset[61], the normalizing help us to simplify the data and avoid the high variance in data without making big changes in data figure(3.6).

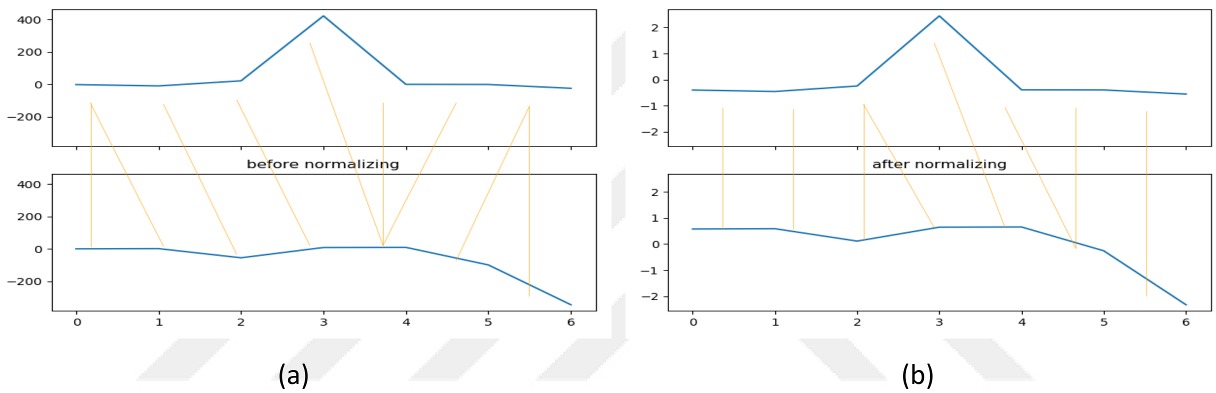

Figure 3.7: illustration of normalization influence on time series data

3.9. APPLYING DTW

by applying the dynamic time warping algorithm on each group, we detect the similarity between every channel.

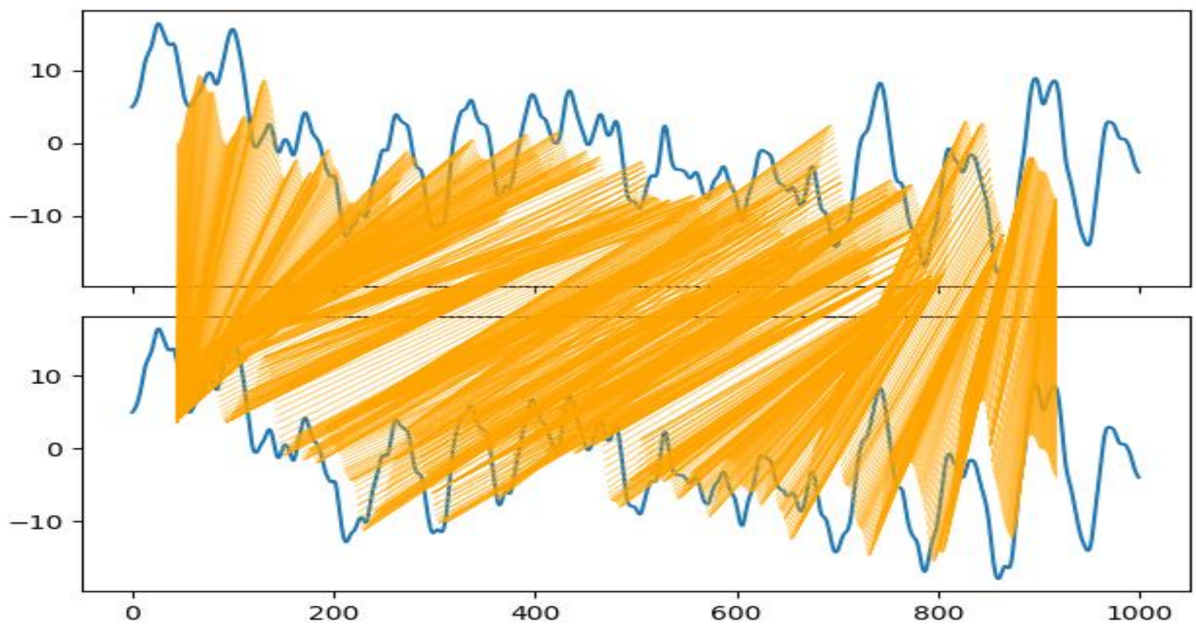

Figure 3.8: DTW between two channels

and other channels to extract appropriate futures for classification, the result of DTW will be the features of new dataset , the new dataset is consist of 120 features (this number comes from the comparing of 10-20 system electrodes with each other which are 16 electrodes) , every feature represents a result of DTW between two channels and every row represent all the similarity result for one second of single-trial EEG signal of one subject in congruent or incongruent status as illustrated in figure (3.8)

DTW (Electrode A, Electrode B)										
D1	DTW(F3,Fp)	DTW(F3,P3)	DTW(Fp,O1)	…..	DTW(Fp,P3)	DTW(Fp,O1)	…	DTW(P3,O1)	…..	DTW(A,B)
D1										
Dn										

Figure 3.9: The shape of a new dataset that created by the result of Applying the DTW on every status group, Dn refers to dataset-n which represents one second of congruent or incongruent status for one subject, DTW (Electrode A, Electrode B) refer to Applying

DTW algorithm was applied by using the python code and DTW python library (dtaidistance.dtw). to represent the synchronized area of brain(that have strong similarity) for each situations (congruent and incongruent)figure we used graph theory techniques and its python library (networkx) ,we represent the nodes as EEG channels(electrodes) and the similarity value between channels as edges (for short we plot only edges that are great than the mean value of total values).

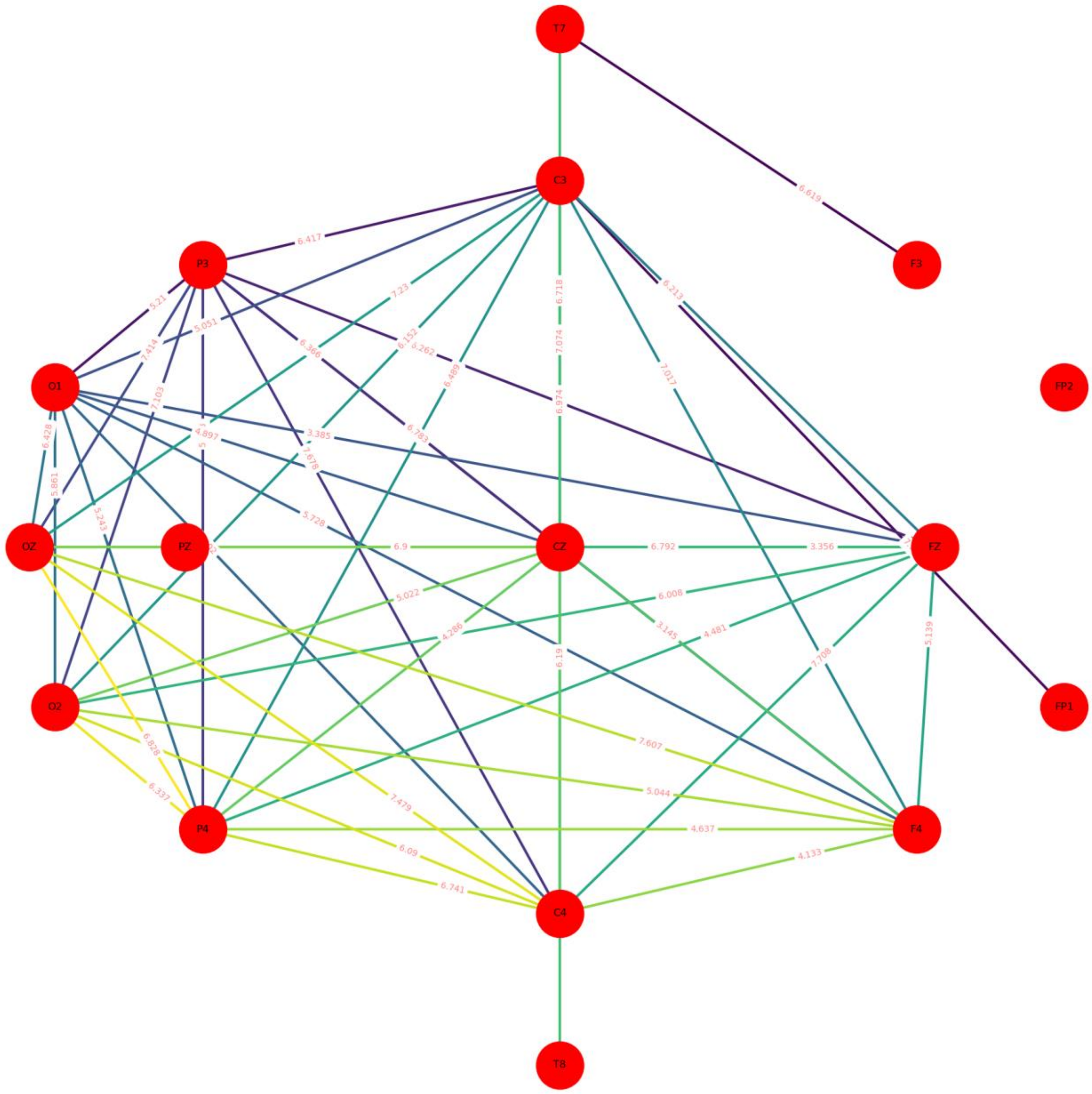

Figure 3.10: representation of applying of DTW on one subject of the one-second incongruent status collected from single-trial ERP EEG data

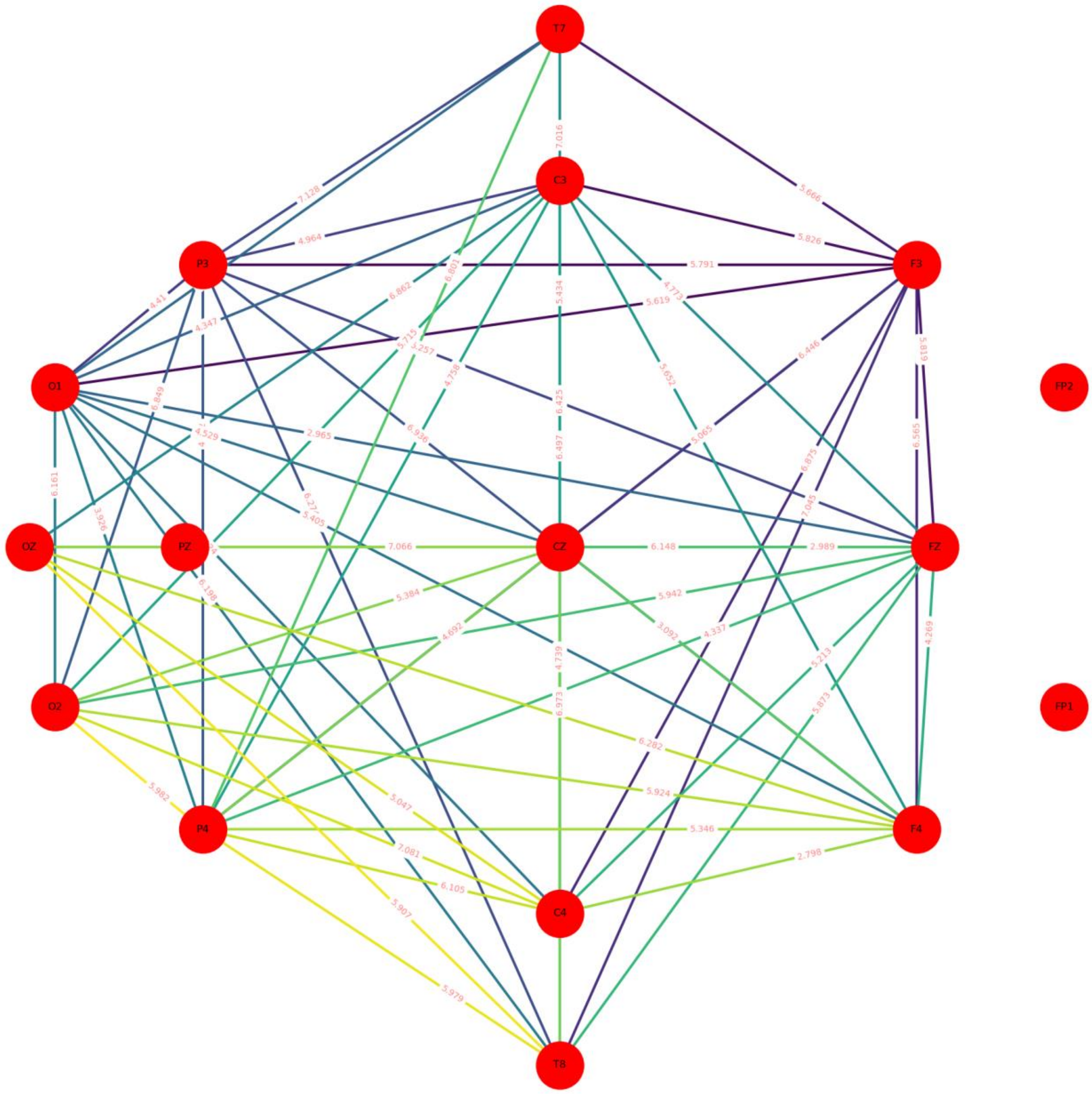

Figure 3.11: representation of applying of DTW on one subject of one second in incongruent status collected from single-trial ERP EEG data

3.10. DATA CLASSIFICATION

Each dataset having EEG data in .csv file format was classified using machine learning techniques that involved in python library. During the classification phase, the model is trained to classify negative or positive congruent/incongruent. Our data is ready now to be classified and tested to have good classifier, In this study we tested two classification algorithms SVM and KNN in order to find the appropriate classification algorithm that has the best performance for our data to build our classifier that can predicate efficiently , the data have binary label congruent(1) and incongruent (2) every label have different mount of instances as illustrated in figure (3.11).

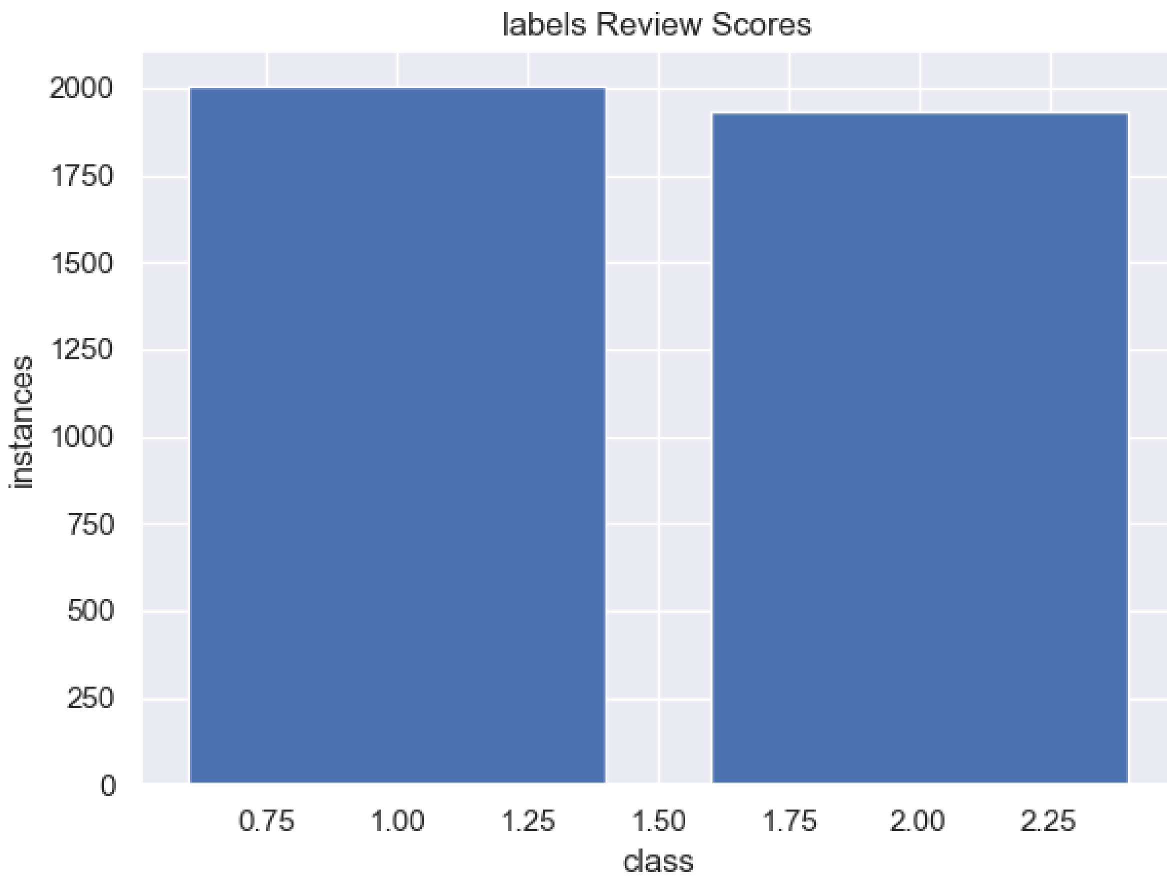

Figure 3.12: class label distribution in data

based on data that we collect from the EEG and extracted by dynamic time warping algorithm and using these data to train the classifier

3.11. IMPLEMENTATION SUPPORT VECTOR MACHINE

To classify the prepared data the collected from the EEG we build a SVM classifier in python using the sklearn library which specified for data science and machine learning algorithms and tools , we split the dataset using k-fold cross-validation ,because we have binary classification which means only two class labels (1,2) and for the SVM classifier kernel we set it to Linear because our data is Hight dimensional and this type of data makes a linear model more flexible . whereas the gamma parameter is set to which controls the width of the Gaussian kernel. the kernel type gives different result so when the kernel changes the accuracy of classifier changes too. Then, to the overfitting problem on the training dataset and tuning the accuracy of classifier we maintain the parameters of the SVM classifier, the c values set to auto, gamma parameter set to auto, in data testing and training we chose the k-fold (10- fold)cross-validation technique

3.12. KNN IMPLEMENTATION

the KNN as one of the most classification technique used for classifying EEG data associated with specific affective/emotional states. typically for classifying the signals and images. This classifier makes the final decision on comparing the testing data with the training data [55]. KNN because of its nature of dealing with the distribution of data with very little knowledge or no knowledge works better with EEG data. KNN helps to reach higher accuracy for classifying EEG data. In this experiment we use KNN classification technique to build a classifier that can predict if the signal is congruent or incongruent, KNN classification algorithm has two main parameters which are k and distance measure parameters, we set the value of k to 3 and we use the Euclidean method for distance parameter

4. RESULT

In this chapter, we will show the results of our work and the techniques we used and discuss them to explain how did we reach these results and we will compare between these techniques to show their influence on our data and which on is batter than other.

4.1. SUPPORT VECTOR MACHINE CLASSIFIER EVOLUTION RESULTS

The classifier is trained and evaluated with 10-fold cross-validation. The classifier parameters have been selected depending on GridSearchCV technique and we get the best params to use it in classification the data and this parameter was (C=3, kernel='rbf',gamma=10,degree=5). the classification reports have been illustrated in table 4.1.

Table 4. 1: Support vector machine classification report

Support vector machine classification report				
Class label	**precision**	**Recall**	**F1-score**	**support**
1 (congruent)	0.99	1.00	0.99	1932
2 (incongruent)	1.00	0.99	0.99	2005
avg	0.99	0.99	0.99	3937

Table 4. 2: SVM confusion matrix

Confusion matrix		
	1(congruent)	**2 (incongruent)**
1 (congruent)	1926	6
2 (incongruent)	19	1986

4.2. KNN CLASSIFIER EVOLUTION RESULTS

The classifier is trained and evaluated with 10-fold cross-validation, The classifier parameters have been selected depending on Grid Search technique and we get the best params to use it in classification the data and this parameter was (algorithm='auto', leaf_size=30, metric_params=None, n_jobs=-1, n_neighbors=3, p=2, weights='uniform').

Table 4. 3: KNN classification report

KNN classification report				
Class label	**precision**	**Recall**	**F1-score**	**support**
1 (congruent)	1.00	1.00	1.00	1932
2 (incongruent)	1.00	1.00	1.00	2005
avg	1.00	1.00	1.00	3937

Table 4. 4: KNN confusion matrix

Confusion matrix		
	1	2
1 (congruent)	1932	0
2 (incongruent)	0	2005

4.3. COMPARISON BETWEEN ALL CLASSIFICATION TECHNIQUES

here we show all techniques used to classify data and the accuracy reached and illustrate the comparison among them.

Table 4. 5: Comparison Between SVM and KNN accuracy

Classification technique	accuracy
SVM	0.99
KNN	1.0

4.4. TECHNIQUE USED IN CLASSIFICATION AND THEIR ACCURACY

From this illustration, We can notice that the KNN accuracy batter than SVM that mean the KNN classification algorithm performs batter on our data.

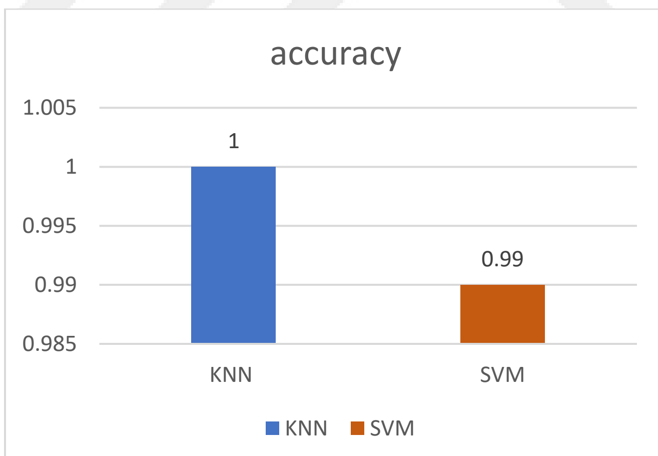

Figure 4.1: Technique used in classification and their accuracy

4.5. TIME-SERIES NORMALIZATION INFLUENCE ON DATA

The normalization is highly required before classification and machine learning process but in feature extraction, it is a little bit tricky. when we applied DTW algorithm on datasets without normalizing them the new dataset the obtained which suppose it contains the extracted features have no meaningful data, vary high variance datapoint figure (4.2), unrelated features figure (4.2), bad classification accuracy for both SVM and KNN algorithms Table(3.6).

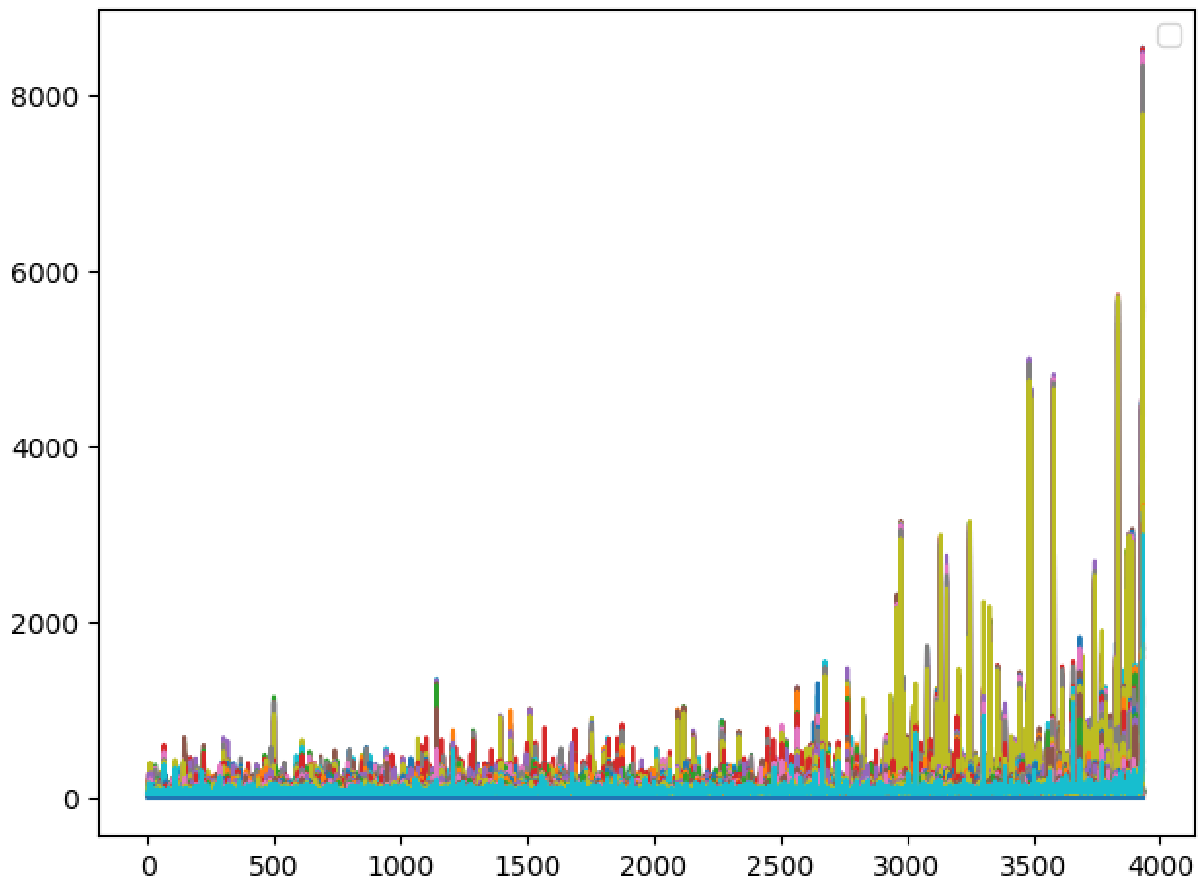

Figure 4.2: Extracted features before normalizing time series

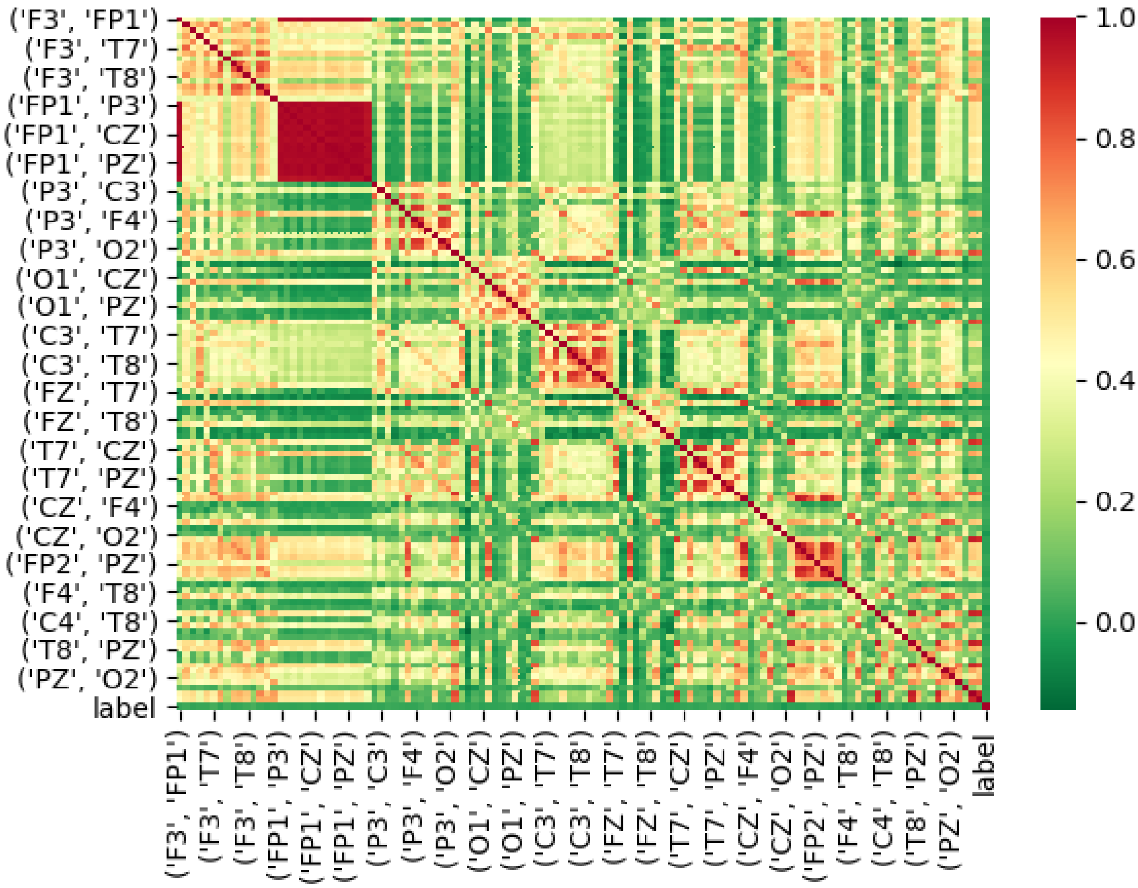

Figure 4.3: Correlation matrix for extracted features before normalizing time series

Table 4. 6: Comparison Between SVM and KNN accuracy before normalizing time series

Classification technique	accuracy
SVM	0.53
KNN	0.56

In applying DTW similarity detection it's highly recommended to normalize each time series before comparison we mentioned in section (2.5.6), after applying z-score normalization on each time series before making comparison in DTW algorithm the result was meaningful, less variance in data figure(4.4) and there was related feature in the new dataset (3.5). The classification became very Hight accuracy table(3.6).

Table 4. 7: Comparison Between SVM and KNN accuracy after normalizing time series

Classification technique	accuracy
SVM	0.99
KNN	1.0

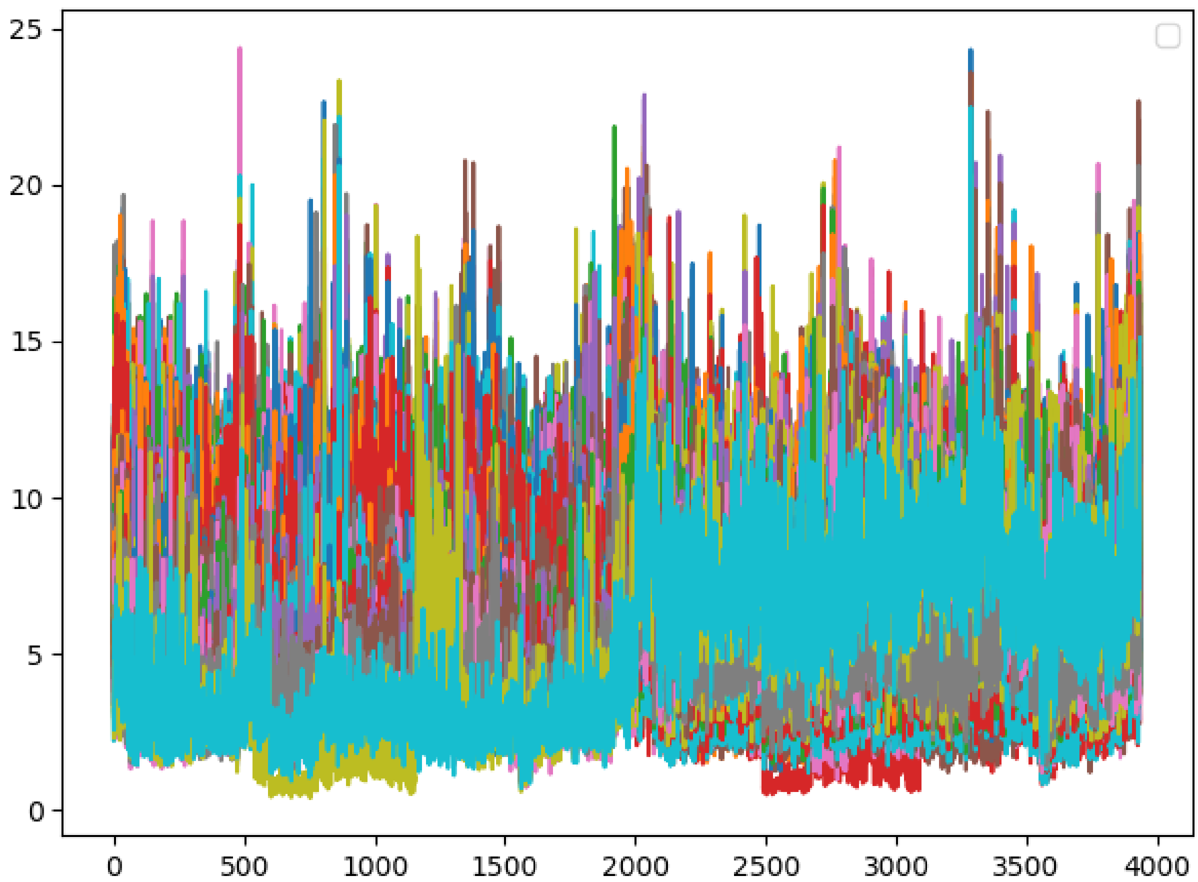

Figure 4.4: Extracted features after normalizing time series

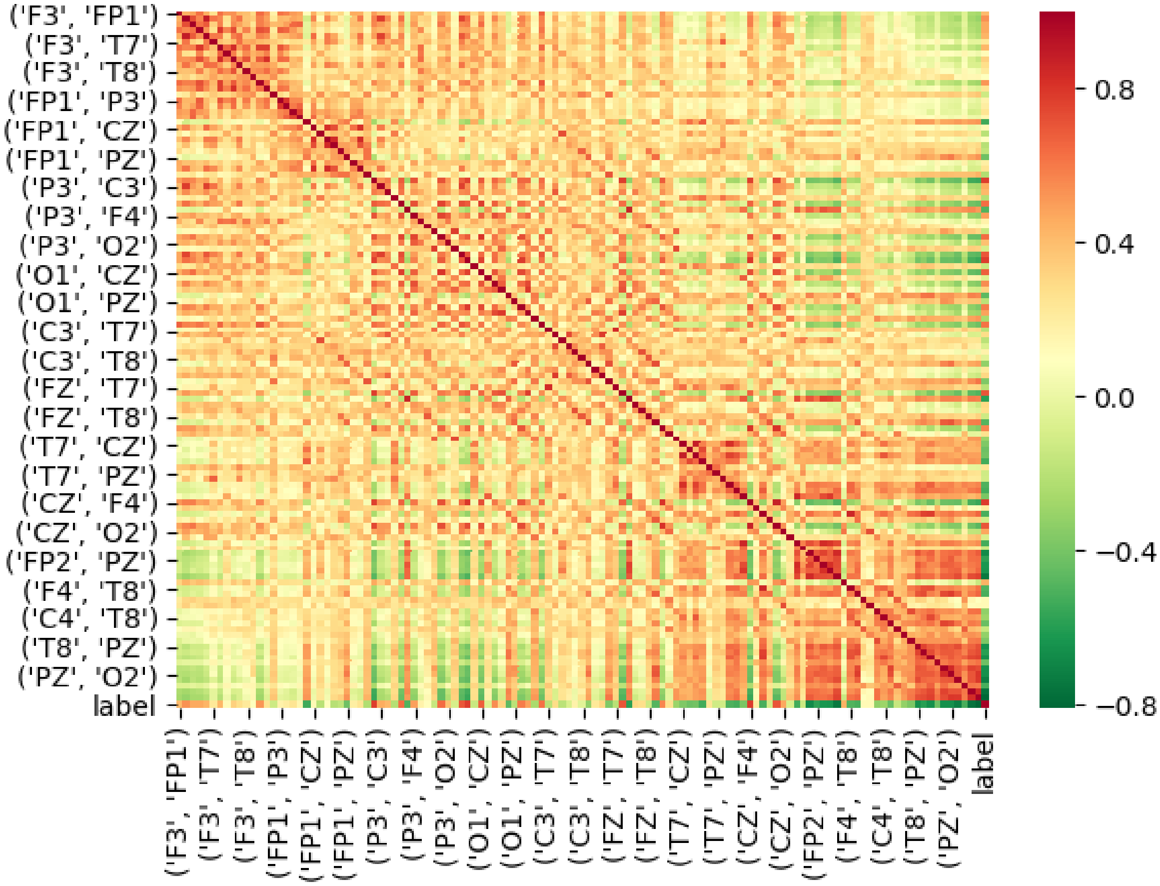

Figure 4.5: Correlation matrix for extracted features after normalizing time series

5. DISCUSSION

Classification and feature extraction are common techniques applied on single-trial ERP of EEG data to extract specific patterns in data and classify them in order to obtain valuable information that can help in many research fields. In this study we could to prove to detect congruent and incongruent states from only one second of single-trail ERP of EEG data that collected from 25 volunteers by using DTW as feature extraction and SVM, KNN as classification techniques and get 98% accuracy for SVM and 100% accuracy for KNN. Furthermore, we prove DTW can work efficiently and give perfect result when we normalize the time series before feeding them to DTW algorithm and that can be obvious in 4.5 section. whereas in [68] research the researcher used WPD as feature extraction technique to obtain features from EEG data that downloaded from ‘’BCI competition 2003 ‘’ website and obtain 90.8% accuracy using PNN classification technique. In another feature extraction from EEG data FDTW [66] technique was used as a feature extraction method for 18 sets of EEG record and used set of classification techniques SVM techniques was among them and its accuracy was 65.37 %.Another research[64] for classification EEG data using SVM and KNN was made on emotion recognition and Hjorth parameters were employed as a feature extraction method of all EEG channels at every epoch. Five male subjects selected in their experiment and they get the highest accuracy with KNN 62% with k=3 and highest accuracy SVM accuracy was 38.9.

6. CONCLUSION

There are many feature extraction techniques used to extract useful features from single-trial ERP data In this work we have subjected Stroop's task experiments on many subjects then we record the   EEG data to process the single-trial signal to investigate about congruent and incongruent status for one second data of subject by using DTW as feature extraction method, to identify congruent and incongruent status we used two classification algorithms SVM and KNN ,we successfully reached significant accuracy the result of each classifier has been illustrated in table [3.5]. From the result of each classifier, we can recognize that both classifiers have approximate accuracy of classification on our data. For data preprocessing the normalizing of time series before feeding it to DTW algorithm was a very effective stage on classification accuracy.

REFERENCES

[1] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang and A. Cichocki, "Sparse Bayesian Classification of EEG for Brain–Computer Interface," in IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 11, pp.2256-2267,Nov.2016.

[2] Thulasidas, Manoj, Cuntai Guan, and Jiankang Wu. "Robust classification of EEG signal for brain-computer interface." IEEE Transactions on Neural Systems and Rehabilitation Engineering 14.1 (2006): 24-29.

[3] Jasper, Herbert H. "The ten-twenty electrode system of the International Federation." Electroencephalogr. Clin. Neurophysiol. 10 (1958): 370-375.

[4] K. S. Sri and J. C. Rajapakse, “Extracting EEG rhythms using ICA-R,” in IEEE

[5] J. Malmivuo and R. Plonsey, Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press, 1995.

[6] P. Zhang and D. F. Galletta, Human-Computer Interaction and Management Information Systems: Foundations. M.E. Sharpe, 2006

[7] A. C. Ramirez-Hernandez, J. A. Rivera-Bautista, A. Marin-Hernandez, and V. A. Garcia-Vega, “Detection and Interpretation of Human Walking Gestures for Human- Robot Interaction,” in Eighth Mexican International Conference on Artificial Intelligence, 2009. MICAI 2009, 2009, pp. 41 –46.

[8] Y.-P. Lin, C.-H. Wang, T.-L. Wu, S.-K. Jeng, and J.-H. Chen, “ EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine,” in Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, 2009, pp. 489 –492.

[9] M. Li, Q. Chai, T. Kaixiang, A. Wahab, and H. Abut, “ EEG Emotion Recognition System,” in In-Vehicle Corpus and Signal Processing for Driver Behavior, K. Takeda, H. Erdogan, J. H. L. Hansen, and H. Abut, Eds. Springer US, 2009, pp. 125–135.

[10] T. Nykopp, “Statistical modelling issues for the adaptive brain interface,” Helsinki University of, 2001.

[11] S. Kaneta, I. Wakabayashi and T. Kawahara, "Feasibility of BMI improvement applying a Stroop effect," 2016 18th International Conference on Advanced Communication Technology (ICACT),Pyeongchang,2016,pp.681-684.

[12] Folstein, J. R., & Van Petten, C. “Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology”, 45(1), 152-170, 2008.

[13] https://github.com/wannesm/dtaidistance

[14] https://networkx.github.io/documentation/stable/install.html

[15] Berndt, Donald J., and James Clifford. "Using dynamic time warping to find patterns in time series." KDD workshop. Vol. 10. No. 16. 1994.

[16] Keogh, E. & M. Pazzani. Derivative Dynamic Time Warping. In Proc. of the First Intl. SIAM Intl. Conf. on Data Mining, Chicago, Illinois, 2001.

[17] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," IEEE Transactions on Affective Computing, vol. 5, pp. 327-339, 2014.

[18] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, et al., "Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping," in ACM SIGKDD international conference on Knowledge discovery and data mining, KDD, 2013, pp. 262-270.

[19] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, pp. 43-49, 1978.

[20]https://www.merriamwebster.com/dictionary/classifin

[21] A Study On Classification Techniques in Data Mining ,

[22] Data Mining: Concepts and Techniques

[23] V. N. Vapnik. The Nature of Statistical Leaming Theory. Springer, New,York, second edition; 2000

[24] Y. Udovychenko et al., "Current Density Distribution Maps Threshold Processing", 2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO), pp. 313-315, Apr. 2014.

[25] V. N. Vapnik, “An overview of statistical learning theory,” Neural Networks, IEEE Transactions on, vol. 10, no. 5, pp. 988 –999, Sep. 1999.rding to established criteria

[26] C. Sammut and G. I. Webb, eds., Encyclopedia of Machine Learning. Springer, 2010

[27] Programming Collective Intelligence by Toby Segaran

[28] H. Parvin, H. Alizadeh, and B. Minaei-Bidgoli, “MKNN : Modified k-nearest neighbor,” in Proc. World Congress on Engineering and Computer Science (WCECS) San Francisco, USA, 2008.

[29] B. V. Dasarathy, “Nearest neighbor (NN) norms: NN pattern classification techniques,” Los Alamitos: IEEE Computer Society Press, 1990, vol. -1, 1990

[30] Y. Wu, K. Ianakiev, and V. Govindaraju, “Improved k-nearest

[31] O. Simeone, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," in IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 4, pp. 648-664, Dec. 2018.

[32] J. Huang, Y. Wei, J. Yi and M. Liu, "An Improved KNN Based on Class Contribution and Feature Weighting," 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, 2018, pp. 313-316.

[33] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[34] R. Salazar-Varas and D. Gutiérrez, "Feature extraction for multi-class BCI using EEG coherence," 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER),Montpellier,2015,pp.94-97.

[35] Al-Fahoum, Amjed S., and Ausilah A. Al-Fraihat. "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains." ISRN neuroscience 2014 (2014).

[36] P. Geethanjali, Y. K. Mohan and J. Sen, "Time domain Feature extraction and classification of EEG data for Brain Computer Interface," 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, Sichuan, 2012, pp. 1136-1139.

[37] Huang, H-C., and B. H. Jansen. " EEG waveform analysis by means of dynamic time-warping." International journal of bio-medical computing 17.2 (1985): 135-144.

[38] Bablani, Annushree, Damodar Reddy Edla, and Shubham Dodia. "Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test." Procedia computer science 143 (2018): 242-249.

[39] Quiroga, R. Quian, et al. "What can we learn from single-trial event-related potentials?." Chaos Complexity Lett 2 (2007): 345-363.

[40] Husain, Shaik Jakeer, and K. Srinivasa Rao. "Epileptic seizures classification from EEG signals using neural networks." 2012 International Conference on Information and Network Technology (ICINT 2012). Vol. 37. 2012.

[41] Gudmundsson, Steinn, Thomas Philip Runarsson, and Sven Sigurdsson. "Support vector machines and dynamic time warping for time series." 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, 2008

[42] Bablani, Annushree, Damodar Reddy Edla, and Shubham Dodia. "Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test." Procedia computer science 143 (2018): 242-249.

[43] Wei, Qingguo, Xiaorong Gao, and Shangkai Gao. "Feature extraction and subset selection for classifying single-trial ECoG during motor imagery." 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2006.

[44] Geethanjali, P., Y. Krishna Mohan, and Jinisha Sen. "Time domain feature extraction and classification of EEG data for brain computer interface." 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2012.

[45] Al-Fahoum , Amjed S., and Ausilah A. Al-Fraihat. "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains." ISRN neuroscience 2014 (2014).

[46] Pang, Zilong, et al. "A new approach for EEG feature extraction for detecting error-related potentials." 2016 Progress in Electromagnetic Research Symposium (PIERS). IEEE, 2016.

[47] Mangalagowri, S. G., and P. Cyril Prasanna Raj. "EEG feature extraction and classification using feed forward backpropogation algorithm for emotion detection." 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT). IEEE, 2016.

[48] Salazar-Varas, Rocio, and David Gutiérrez. "Feature extraction for multi-class BCI using EEG coherence." 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, 2015.

[49] Outrata, Jan. "Boolean factor analysis for data preprocessing in machine learning." 2010 Ninth International Conference on Machine Learning and Applications. IEEE, 2010.

[50] Quinlan, J. Ross. C4. 5: programs for machine learning. Elsevier, 2014.

[51] Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. "Cluster analysis: basic concepts and algorithms." Introduction to data mining 8 (2006): 487-568.

[52] Kotsiantis, S. B., Dimitris Kanellopoulos, and P. E. Pintelas. "Data preprocessing for supervised leaning." International Journal of Computer Science 1.2 (2006): 111-117.

[53] Mohamad, Ismail Bin, and Dauda Usman. "Standardization and its effects on K-means clustering algorithm." Research Journal of Applied Sciences, Engineering and Technology 6.17 (2013): 3299-3303.

[54] Rodríguez-Pérez, Raquel, Luis Fernández, and Santiago Marco. "Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study." Analytical and bioanalytical chemistry 410.23 (2018): 5981-5992.

[55] Murugappan, Murugappan, Nagarajan Ramachandran, and Yaacob Sazali. "Classification of human emotion from EEG using discrete wavelet transform." Journal of biomedical science and engineering 3.04 (2010): 390.

[56] Hossin, Mohammad, and M. N. Sulaiman. "A review on evaluation metrics for data classification evaluations." International Journal of Data Mining & Knowledge Management Process 5.2 (2015): 1.

[57] scikit-learn.orgstablemodulesgeneratedsklearn.model_selection.GridSearchCV.html

[58] Bain, H. M., P. Brea, and E. T. Adamson. "Using Machine Learning Techniques to Forecast Solar Energetic Particles." AGU Fall Meeting Abstracts. 2018.

[59] Shuai, Yong, Yujie Zheng, and Hao Huang. "Hybrid Software Obsolescence Evaluation Model Based on PCA-SVM-GridSearchCV." 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2018

[60] Figo, Davide, et al. "Preprocessing techniques for context recognition from accelerometer data." Personal and Ubiquitous Computing 14.7 (2010): 645-662.

[61] Rakthanmanon, Thanawin, et al. "Searching and mining trillions of time series subsequences under dynamic time warping." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[62] Han, Jiawei, Micheline Kamber, and Jian Pei. "Data mining concepts and techniques third edition." The Morgan Kaufmann Series in Data Management Systems (2011): 83-124.

[62] Davis, Jonathan J., and Andrew J. Clark. "Data preprocessing for anomaly based network intrusion detection: A review." computers & security 30.6-7 (2011): 353-375.

[63] Rath, Toni M., and Raghavan Manmatha. "Word image matching using dynamic time warping." 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.. Vol. 2. IEEE, 2003.

[64] Mehmood, Raja Majid, and Hyo Jong Lee. "Emotion classification of EEG brain signal using SVM and KNN." 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2015.

[65] Das, Tarak, et al. "Classification of eeg signals for prediction of seizure using multi-feature extraction." 2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech). IEEE, 2017.

[66] Calvo, Hiram, Jose Luis Paredes, and Jesús Figueroa-Nazuno. "Fast Dynamic Time Warping Feature Extraction for EEG Signal Classification." 2016 Fifteenth Mexican International Conference on Artificial Intelligence (MICAI). IEEE, 2016.

[67] Estrada, E., et al. "EEG feature extraction for classification of sleep stages." The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 1. IEEE, 2004.

[68] Ting, Wu, et al. "EEG feature extraction based on wavelet packet decomposition for brain computer interface." Measurement 41.6 (2008): 618-625.

[69] Müller-Gerking, Johannes, Gert Pfurtscheller, and Henrik Flyvbjerg. "Designing optimal spatial filters for single-trial EEG classification in a movement task." Clinical neurophysiology 110.5 (1999): 787-798.

APPENDIX A

Normalize time series and applying DTW code:

```
for i in range(len(data)):
    for j in range(i,len(data)):
        if j == i:
            continue

DTW_result.append(dis.dtw.distance_fast(np.array(stats.zscore(data[i]),dtype='d') ,
np.array(stats.zscore(data[j]),dtype='d')))
```

Aplaying DTW on conguret and inconguretn data and visualize it with Graph Theory.

```
for i in data:
    for j in data:
        if j == i:
            continue

DTW_result[np.around(dis.dtw.distance_fast(np.array(stats.zscore(data[i]),dtype='d') ,
np.array(stats.zscore(data[j]),dtype='d')),decimals=3)] = (i, j)
dataInfo=pd.DataFrame(DTW_result)
meanValue=sts.mean(DTW_result.keys())


new={i:DTW_result[i] for i in DTW_result if i<meanValue}
print(new)

edgeslabel = dict([[v, k] for k, v in new.items()])
electrodes = []
# print(data_matrix)
for i in data:
    electrodes.append(i)
# --------------------------------------------------plot---
g = nt.Graph()
edgeColor = [i for i in range(len(edgeslabel.values()))]
edges = edgeslabel.keys()
plt.figure(figsize=(20, 20))
# edge_width=[ i/10.100 for i in edgs.keys()]
edge_width = 3
g.add_nodes_from(electrodes)

pos = {'F3': (0.67303, 0.54501), 'FP1': (0.95106, -0.30902), 'FP2': (0.95106,
0.30902), 'P3': (-0.67303, 0.54501),
       'O1': (-0.95106, 0.30902), 'C3': (4.3298e-017, 0.70711), 'FZ': (0.70711, 0),
'T7': (6.1232e-017, 1),
       'CZ': (6.1232e-017, 0), 'F4': (0.67303, -0.54501), 'C4': (4.3298e-017, -
0.70711), 'T8': (6.1232e-017, -1),
       'PZ': (-0.70711, -8.6596e-017), 'P4': (-0.67303, -0.54501), 'O2': (-0.95106, -
0.30902), 'OZ': (-1, -1.2246e-017)}
# g.add_edges_from(edgs.values())
nt.draw(g, pos,with_labels=1,node_size=3000,node_color='r')
nt.draw_networkx_edges(g,pos,edgelist=edges,width=3,edge_color=edgeColor,nodelist=g.no
des)
nt.draw_networkx_edge_labels(g, pos, edge_labels=edgeslabel,
```

```
font_color='red',alpha=0.5,label_pos=0.3)
#
nt.draw_networkx_edges(g,pos,edgs.values(),width=edge_width,edge_color=edgeColor,cmap=
'hot')
fig = plt.gcf()
fig.set_size_inches(18.5, 18.5)
fig.savefig('draw_only_similarity_electrodes_for_congruent_subject.png', dpi=100)
plt.show()
```

Using GridSearchCV of sklearn python library to get best parametres for modeles .

```
models = []

knn_paras={'n_neighbors':[1,100]}
knn_GS=GridSearchCV(KNeighborsClassifier(),param_grid=knn_paras)

svm_paras={'C':[0.1,0.100],'gamma':[0,100],'kernel':['rbf','linear','poly']}
svm_GS=GridSearchCV(SVC(),param_grid=svm_paras)

models.append(('svm', svm_GS))
models.append(('KNN', knn_GS))


results = []
names = []

kfold = model_selection.KFold(n_splits=10)
for name, model in models:
   names.append(name)
   print(name,model.best_params_)
```