CLASSIFICATION OF EEG SIGNAL USING ENSEMBLE SYNCHRONIZATION VIA WAVELET TRANSFORM

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SIGNAL PROCESSING AND DIGITAL DESIGN

Submitted by:

ANSHUL BHABHA

2K17/SPD/02

Under the supervision of

DR. SUDIPTA MAJUMDAR

DEPARTMENT OF ELECTRONICS AND COMMUNICATION
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042
JUNE, 2019

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Anshul Bhabha, 2K17/SPD/02, of M.Tech, hereby declare that the project Dissertation Titled "Classification of EEG signal using ensemble synchronization via wavelet transform" which is submitted by me to the Department of Electronics and Communication, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **ANSHUL BHABHA**

Date:

ELECTRONICS AND COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Classification of EEG signal using ensemble synchronization via wavelet transform" which is submitted by Anshul Bhabha, Roll No. 2K17/SPD/02, Electronics and Communication Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

DR. SUDIPTA MAJUMDAR

SUPERVISOR

Assistant Professor

Department of Electronics and Communication

Delhi Technological University, Delhi

ACKNOWLEDGEMENTS

My deepest gratitude is to my advisor, Dr. Sudipta Majumdar. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. Her patience and support helped me to finish this dissertation.

I would like to thank my family and friends and the one above all of us, the omnipresent God for the support they provided.

Finally, I would like to thank family of DTU.

ANSHUL BHABHA

ABSTRACT

In this research work classification of the subjects using their respective single trial EEG signals is done based on the ensemble synchronization and the phase based on the wavelet transform. Phase synchronization matrix is formed based on the instantaneous parameters of the wavelet-based phase estimation between the EEG channels. Frobenius norm is used for the normalization of the ensemble synchronization so that it can be compared with the other subjects on the 0 to 1 scale. Lower gamma band intrahemispheric area of the brain was studied as from the other research work this band is mainly responsible for the long-range coordination of the cognitive process and thus is an important factor for the classification of the subject into healthy or the schizophrenic one. It was observed that this process classified the subject into their respective category with an accuracy of 75% as compared to the other models using the Hilbert transform based which gave an accuracy of about 70%. This is due to the time frequency localization property of the wavelet transform.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Significant progresses have been done in the past years for the detection of the early stages of the schizophrenia by the biochemical, genetic, neuroimaging, and neurophysiological biomarkers such as Electroencephalography (EEG). EEG signal tracks the connection between the neurons in the recording sites of the scalp and accordingly produces the electrical signal with milliseconds precision. The brains physiology can be understood from the EEG signal recorded and by identifying the unusual frequency patterns. Different rhythms along with the frequency bands are used for describing the activity of the brain [1]-[3]. The important bands are given as: alpha (8-13 Hz, 30-50 μV amplitude), beta (13-30 Hz, 5-30 μV amplitude), gamma (≥ 30 Hz), delta (0.5-4 Hz), and theta (4-7 Hz, ≥ 20 μV amplitude).

The Brain Computer Interface (BCI) provides direct connection path from the human intentions of the brain signal to the computers. This helps in recording and gathering of the data or what the response of humans are towards a particular input. The various functions of human are related to or involves the processes of thinking and reasoning, involves the combination of different brain areas. Even for performing simple tasks different areas of brains are used and they must be in perfect phase synchronization. Brain synchronization can be referred to as the process which leads to the mental clarification in the decisions and the awareness that occurs when the left and the right hemispheres of the brain are synchronized. When any of the activity occurs, the left and the right hemispheres of the brain simultaneously works together. It can be described as the brain synchrony or the hemispherical synchronization. It is the result of the production of the same brain waves by both of the hemispheres of the brain.

Synchronization occurs in the living nature. Synchronization concept in the living nature has been studied to model the interaction between the different physiological subsystems that represents oscillatory behaviour. One very important is the interaction of the human cardiovascular and respiratory systems. This synchronization term is also related with neuroscience problem. The synchronization is a critical mechanism for the neuronal information processing with brain area and the communication between different brain areas. This synchronization is also responsible for the binding of different visual pattern as a whole. Synchronization also plays an important role in the neurological diseases. So, analysis of synchronization is important for the understanding of the physiological brain function and various neurological disease. Synchronization in living nature is a mechanism by which different brain regions coordinates. Abnormal synchronization results in the neurological disorders. In neural signal processing the synchronization is divided into two classes that is phase synchronization and amplitude synchronization.

Binding problem in neuroscience is the method in which an array of parallel processing occurs in the brain at any given time. The diverse neuronal activities are related to a single stimulus and are bound together or integrated. Binding problem in neuro science refer to method of integration of diverse neuronal activities that relate to a single stimulus due to array of parallel processing occurring in brain at any time. The synchronous activity in specific narrow band frequency ranges are the mechanism involved in distributed network binding and gamma band is important for this mechanism.

Synchronization of these neuronal activity, which are associated with network oscillations [4], provides a means for combining the anatomically distributed processes in the brain. But neuronal processing is associated with the simultaneous oscillations in the frequency bands. Thus, for unified perceptual experience from more than one input or multiple input phase synchronization is important. At the level of neural ensembles, large number of neurons synchronization activity gives rise to macroscopic activity, and these macroscopic activity results in macroscopic oscillations, leading to the development of a signal [5]. Such signals are termed as electroencephalogram (EEG). The pattern changes related to spatial and temporal in the EEG can be used to know the actual movement, imagined movements (thought) or observation of the movements or thoughts.

Phase synchronization is involved in different memory processes, attentive modulation and sensory perception, task execution, active planning, etc [6]. Schizophrenia is a chronic and severe mental or psychotic disorder that manifests itself by changing the patients mental (such as delusions or disturbances in thoughts), perception (which includes hallucinations), and functional activity. It is related with the aberrant brain connectivity. Schizophrenia is a task assigned aberrant synchrony in the scalp of the EEG associated signals. Hence, the phase synchronization scalp EEG across the different regions of brain during the audio–visual integration task performance, in the view of individuating subjects with schizophrenia from the list of healthy subjects using a different approach of ensemble synchronization-based classification metric. Conventionally, the analysis was done by ensembling or average of more than two tasks were taken for extracting the important contrariness. But the signal by noise ratio among the recalled EEG and the instant EEG is much low. Thus, for increasing the transfer rates of information, single-trial recognition is alluring for the brain computer interface system.

For improving the performance against the noise most of the existing electroencephalography analysis are based on or developed by using the averaging techniques over multiple trials. Classical averaging techniques in most of the case allows low ratio of the signal and noise (SNR). But if we average the signal over multiple simulation trials then it is possible to remove noise or its induced properties to some of the extent. It helps in the removal of trial to trial fluctuation in event related potentials (ERPs), thus it also deliberately neglects the transient response of the same type of stimuli. Attaining the single trial is much of a significant task. The reason for the following can be explained as follows: firstly, the response of the human brain to the external incentive is embedded in the EEG signals which have low signal to noise ratio (SNR), due to which the same response is very much difficult to achieve. Second case for the attainment of the single trial is that for a single trial, the correlation between most of the electrodes output is so powerless that the brains absolute response can be incompletely understood [6]. So, the different neuronal activity will not be correctly broken down such as power or the stimuli response latency, etc. Feature extraction is a technique of obtaining the desirable properties from the EEG signals, such that they can define or classify the signals into unique remainders.

1.1 SINGLE TRIAL EEG CLASSIFICATION

The most important researches in the area of the Brain Computer Interface (BCI) is the EEG signal analysis and classification [6]. Newer and more efficient techniques are being devised for analysing the EEG signals and extracting the relevant data from it.

The process of the EEG signal analysis and classification consists of the following three steps:

• SIGNAL PRE-PROCESSING: In this process the filtering of the specific portion is done which is more important from the analysis point of view. For example, the gamma band is much important for the long-range coordination of the signals in the brain. In this process the artifacts are also removed which creeps due to the blinking of the eye, muscular activity, etc.

• FEATURE EXTRACTION: In this step the extraction of those feature takes place of the signal which displays the characteristics properties of the of EEG signal which are unique to the signal and thus are suitable for the purpose of classification. Due to this extraction feature the total amount of data that is fed to the system is drastically reduced and thus it leads to the decrease in the processing time for the BCI system. Most commonly used features for the classification includes the FFTs, the PSDs, the auto regressive coefficients, the multivariate autoregressive coefficients and the time-frequency transforms. Spatial filtering of the multi-channel EEG signal is also performed for taking out the discriminatory information from the signals. For identifying which electrodes provide better discriminatory information different techniques used are the neural network feature selector, fuzzy entropy-based feature ranking and the signal to noise ratio-based technique

• THE CLASSIFICATION PROCESS: In this process the demarcation boundary is to be identified for the different classes of the signals in the feature space. Different approach is used for this classification such as neural networks or the support vector machines or the logistic regression-based classifier.

1.2 LOGISTIC REGRESSION

The output variable is known that is denoted by Y and further aim is modelling of the probabilities that is $P(Y=1 \mid X=x)$. All the parameters which are not known have to be estimated by using the maximum likelihood criteria. This estimation can be performed by different ways:

1. First method is to consider p(x) as a linear function of x. The increase or decrease in the value of x causes an increase or decrease in the value of probability. The main

thing to be kept in mind is that the value of the probability must be in the range of 0 and 1. Also the linear functions are bounded functions. In many cases it has been seen that the small change in the value of p needs a big change in the value of x. Linear model doesn't have the capability to pursue this.

2. Second method after method one fails is that, assume log p(x) as a linear function which is linear with respect to x. According to this if there is a change in the input then the probability gets multiplied by a static value. But the difficulty in this is that the log function is inbounded in only one direction but linear functions are unbounded in both of the directions.
3. Third method which is the easiest of all is to consider a log function as $\log \frac{p}{1-p}$ which is also said as the logistic transformation.

This third case is termed as the logistic regression. It can be more specifically defined as

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + x \cdot \beta$$

… (1.1)

Further it can be solved for p as

$$p(x; b, w) = \frac{e^{\beta_0 + x \cdot \beta}}{1 + e^{\beta_0 + x \cdot \beta}} = \frac{1}{1 + e^{-(\beta_0 + x \cdot \beta)}}$$

… (1.2)

For the minimization of this error rate, Y=1 must be considered when the value of $p \geq 0.5$ and the value of Y = 0 when the value of $p < 0.5$. According to the above sentence we can say that the value of Y=1 when the $\beta_0 + x \cdot \beta$ value is non-negative. Thus, this discussed regression that is the logistic regression model gives us the linear classifier. The decision boundary is selected by $\beta_0 + x \cdot \beta = 0$ solution, according to which if x is 1-D then it is a point, if x is 2-D then the decision boundary is a line.

1.3 PHASE SYNCHRONIZATION

Phase relation among the brain regions is an important criterion for the for determining the timing of the action potential in those regions. Phase synchronization is taken into account for the two regions in the oscillation when the phases oscillation in these two regions are taken into account and are found to be correlated.

The synchronization that occurs in the brain's phase is one of the basic mechanisms of the neural mechanism. This is an important criterion for communication between the neurons and is important for multiple cognitive processes [7]. As the data has increased and the accuracy in most of the biological department has increased but the basics of these functions and their underlying fact is missing. Thus, in this topic the main focus was on what is the role of phase synchronization in the various memory process. And this is the most important task that is to be proceeded with. This can help in understanding how the simple phase synchronization can help in the basic understanding of the various neurophsycotic diseases.

Phase synchronization functions with respect to the neural communication and by the experiments involved it backs the inference in the frequency range which is associated mainly with the gamma band that it makes a transient union among the brain regions that reflects a particular property of the stimulus [8]-[9]. For example, when we see the Taj Mahal, information about its color (white), its architecture (historical) and its location is refined in different blocks or regions of the brain. And these data that has been processed must be interlinked to each other such that our brain assigns them to the same object or class. Due to this binding phenomenon it assures us that phase synchronization plays an important role in the neural connectivity.

1.4 HILBERT TRANSFORM

Hilbert transform is a particular linear operator which considers a function, u(t) of a real variable and it gives alternative signal as output that is H[u(t)]. The Hilbert operator is defined as the convolution with the function $1/\pi t$:

$$H[u(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} \, d\tau \qquad \dots (1.3)$$

In terms of the frequency domain this Hilbert transform has a simple representation that is it imparts the phase shift of 90^0 to every components of the Fourier transform. We can take a function cos (ωt) in which ω > 0, then its Hilbert transform will be cos (ωt-

90^0). This Hilbert transform is a much important task in the signal processing as it derives the analytical representation of the real valued signal u(t).

1.5 WAVELET TRANSFORM

Wavelet transform has the resemblance to the Fourier transform (or much more to the windowed Fourier transform) but it has totally different merit function. As the Fourier transform decomposes the signal in the cosines and sines that is the function localised in the Fourier space, the wavelet transform considers those functions that are localized in both the Fourier and the real space. Wavelet transform is given as:

$$F(a,b) = \int_{-\infty}^{\infty} f(x)\, \psi_{(a,b)}^{*}(x)\, dx$$

… (1.4)

where, * is referred to as the complex conjugate and the function ψ is some function. This ψ is chosen such that it follows certain rules.

It can be seen that the wavelet transform consists of the infinite set of various transforms, which further depends on the merit function which is used for its computation. Thus, wavelet transforms can be used in multiple situations and applications. Also, there are many ways in which we can consider the type of the wavelet transform to be used. Such as orthogonal wavelets can be used for the discrete wavelet transform development and the non- orthogonal wavelets for the continuous wavelet transform development. These two transforms have the following properties:

1. The Discrete Wavelet Transform outputs the data vector which have the same length as that of the input. Most of the outputs in this vector are almost zeros. This is because the decomposition is done into the set of functions that are orthogonal to its translations and scaling. Hence, we decompose those signals into lower no. of the wavelet coefficients spectrum that is as many as the input data are present. Redundancy is removed in this process completely. The construction of the wavelets can be done from a scaling function as it describes its scaling properties. The restriction that the scaling functions must be orthogonal to its discrete translations implies some mathematical conditions on them which are mentioned everywhere,

e.g. the dilation equation (1.5). The area between the functions must be normalized and the scaling function must be orthogonal to its integer translations (1.6)

$$\varphi(x) = \sum_{k=-\infty}^{\infty} a_k \, \varphi\,(Sx - k) \qquad \dots (1.5)$$

$$\int_{-\infty}^{\infty} \varphi\,(x)\, \varphi\,(x + l)\, dx = \delta_{0,l} \qquad \dots (1.6)$$

2. The Continuous Wavelet Transform given an output of one dimension greater than the input data. It is the implementation of the wavelet transform by using arbitrary scales and almost arbitrary wavelets. The wavelets used are not orthogonal and the data obtained by this transform are highly correlated. The choice of wavelet that is to be used is of much importance. Using this wavelet choice, we can influence the time and frequency resolution of the result. If we use the Morlet wavelet for example (real part – damped cosine function) we can expect high frequency resolution as such a wavelet is very well localized in frequencies. In contrary, using Derivative of Gaussian (DOG) wavelet will result in good time localization, but poor one in frequencies.

1.6 ORGANIZATION OF THE THESIS

This thesis is organized as follows:

Chapter 1 presents the introduction to the parameters which are used in the work

Chapter 2 presents various quantitative methods used in this work such as logistic regression, ensemble synchronization, likelihood rating, etc. in brief.

Chapter 3 presents brief theory of wavelet transform.

Chapter 4 presents the theory on the logistic regression classifier used in the work.

Chapter 5 presents problem statement and simulation results.

Chapter 6 presents conclusion and future work.

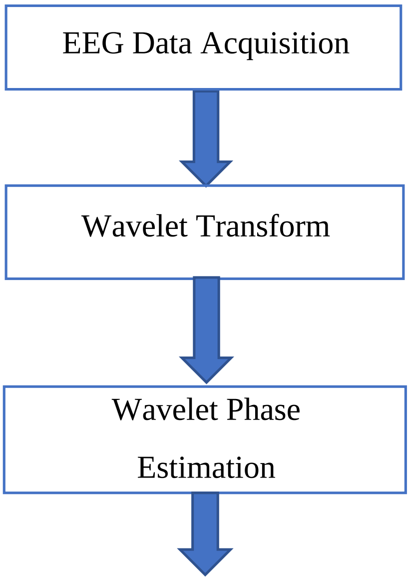
CHAPTER 2

QUANTITATIVE METHODS

2.1 DEVELOPMENT OF THE MODEL

In the development of this model, the main aim was to classify the subjects as either healthy or schizophrenic. Single trial was used so that the system developed can be applied in the real time scenarios too. Ensemble synchronization was used as a feature by combining the data from multiple channels of a subject. Ensemble consisted of all the areas of the brain which was involved in the execution of the cognitive tasks such as judgement, attention, problem solving, thinking or imagination, so that the difference in the subject's presentation among them is anticipated. As we have studied in chapter two that the gamma band is responsible for the extended range of coordination, thus intra-hemispheric synchronization was considered.

2.2 FLOWCHART OF THE PROPOSED MODEL



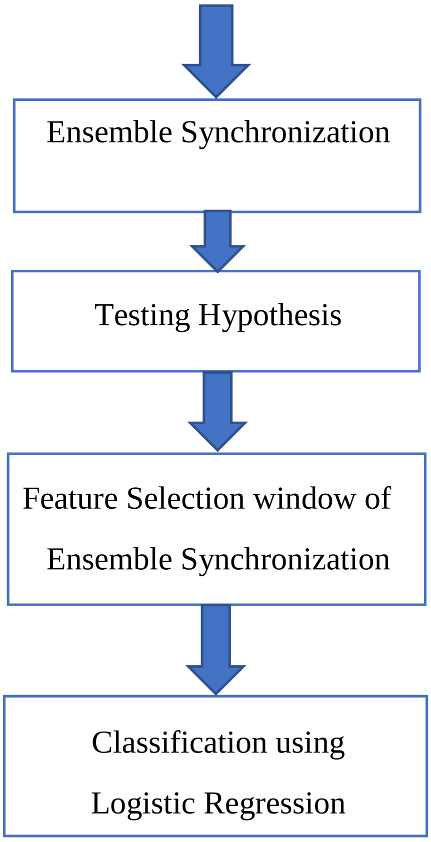


Fig 2.1 Flowchart of the model.

2.3 QUANTITATIVE ANALYSIS

The steps for the classification of the subjects can be divided to seven parts which are named as:

a) Hilbert Phase synchronization,
b) Wavelet phase synchronization,
c) Measurement by ensemble synchronization,
d) Phase synchronization's statistical significance,
e) Logistic Regression,
f) Classification based on logistic regression,
g) Performance of the classifier and
h) Likelihood rating.

2.4 CALCULATION OF PHASE VIA WAVELET

Let g(t) be analytic wavelet function of whose Fourier transform is given by $\hat{g}(w)$ which satisfies the condition for any of the given signal,

$$g(t) \in L^1 (R, dt) \cap L^2 (R, dt)$$

$$\hat{g}(w) \in L^1(R\backslash\{0\}, dw/|w|) \cap L^2(R\backslash\{0\}, dw/|w|)$$

Where, $s(t) \in L^2 (R, dt)$.

For a signal s(t) the wavelet transform that is WT is defined according to the g(t), in which t and b ∈ Real number and a > 0, $\bar{g}$(t) denotes the conjugate of g(t).

$$S(b,a) = \frac{1}{a} \int_{-\infty}^{\infty} s(t)\bar{g}\left(\frac{t-b}{a}\right) dt$$

… (2.1)

2.5 THEOREMS

Theorem 1 [10]: Consider a signal s(t) with finite energy and an analytic equation of wavelet as g(t). We define S (b, a) as the wavelet transform of the above given signal in accordance to g(t). The wavelet transforms of s(t), that is S (b, a) is a complex function in accordance to the real value of the variable "b" and the n the variable "a" is termed as the scaling factor. If we keep the variable "a" as constant then the imaginary part of the above function is the Hilbert transform of the real part. This statement can be proved in [10] and [11].

Theorem 2 [10]: Let g(t) be an analytic wavelet function and its real part be $g_R(t)$, and it is even. Also, we define as in the paper [10]

$$C_g = \int_0^{\infty} \left(\frac{\widehat{g_{R(w)}}}{w}\right) dw$$

… (2.2)

where 0 < $C_g < \infty$. It can be written for any random real number, s(t) ∈ $L^2(R, dt)$, and

$$\frac{1}{C_g}\int_0^{\infty} S(t,a)\frac{da}{a} = s(t) + j\,H[s(t)]$$

… (2.3)

Where, S (t, a) is defined above as the wavelet transform. H[s(t)] is the Hilbert transform of s(t).

2.6 WAVELETS BASED ON THEOREM 2

The most basic and the wavelet which satisfies the theorem 2 is Morlet wavelet. It is given as [19]

$$g(t) = \; e^{jmt}e^{-\frac{t^2}{2}}\,(m > 6).$$

… (2.4)

The Morlet function satisfies the theorem 2 and it can easily be proved by [12]. This Morlet function can further be modified and is expressed as [10],

$$g_{\tau}(t) = \; e^{imt}\; e^{-(1/2)[\sqrt{2}\sigma m/(2\pi\tau)t]^2}$$

$$= cos(mt)\; e^{-(1/2)[\sqrt{2}\sigma m/(2\pi\tau)t]^2}$$

$$+ j\sin(mt)\; e^{-(1/2)[\sqrt{2}\sigma m/(2\pi\tau)t]^2}$$

where,

$C = \sqrt{2}\sigma m/(2\pi\tau)$,

m = Angular frequency,

τ = Frequency in an envelope,

σ = Precision number.

Now when the value of $m^2/(4C^2)$ is much large then it satisfies the Theorem 2. Instantaneous Parameter for the signal s(t) thus can be defined as given below [10],

$$e(t) = \; \sqrt{s^2(t) + \; H^2[s(t)]}$$

$$\theta(t) = \arctan\left(\frac{H[s(t)]}{s(t)}\right)$$

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\arctan\left(\frac{H[s(t)]}{s(t)}\right)\right]$$

Thus, we can formulate for the estimation of instantaneous parameter as,

$$f(t) = \frac{1}{2\pi}\frac{s(t)\frac{dH[\,s(t)\,]}{dt} - H[\,s(t)]\frac{ds(t)}{dt}}{e^2(t) + \epsilon^2 e_{max}^2}$$

Where,

$$e_{max}^2 = \max(\,e^2(t)), \quad \text{and} \quad 0 < \epsilon < 1$$

consider, $c_1 = \frac{\sqrt{2}\sigma m_1}{2\pi\tau_1}$, $c_2 = \frac{\sqrt{2}\sigma m_2}{2\pi\tau_2}$, $\sigma = 5$, $\tau_1 = \tau_2 = 4$, $m_1 = 28.28$, $m_2 = 42.43$, and using all these parameters we can obtain [10]

$$H[s(t)] = e^{-(1/2)(c_1 t)^2}\sin(m_1 t) + e^{-(1/2)(c_2 t)^2}\sin(m_2 t)$$

… (2.5)

which can further be used in place of the Hilbert transform for getting the analytic wavelet signal.

2.7 HILBERT PHASE SYNCHRONIZATION

Phase refers to the instantaneous phase determined by the Hilbert transform. It is termed as the Hilbert phase synchronization. Hilbert phase synchronization between two ECoG (electrocorticoencephalogram) is computed as the phase synchronization between two chaotic oscillators. Phase locking between two periodic oscillators is defined as

$$|\varphi_{n,m}(t)| < c, \qquad \varphi_{n,m}(t) = n\varphi_1(t) - m\varphi_2(t)$$

… (2.6)

where, c is a constant, φ_1 and φ_2 are the phases pf the ECoG from the first and second channel respectively. The n:m phase locking between two signals is given by equation (2.6). Consider a signal s(t) such that it is in time domain. An analytic signal is defined ψ(t) such that the signal can be expressed as:

$$\psi(t) = s(t) + j\,\widehat{s(t)}$$

and,

$$\psi(t) = I(t)\exp\;(j\,\varphi(t))$$

where, $\widehat{s(t)}$ is termed as the Hilbert transform of the signal as defined of s(t), j= $\sqrt{-1}$, I(t) is termed as the instantaneous amplitude or the envelope and $\varphi(t)$ is termed as the phase of the signal which is given by $\psi(t)$.

$$\widehat{s(t)} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}\,d\tau$$

… (2.7)

where only Cauchy principle value is considered in equation (2.7). Physical interpretation of I(t) and $\varphi(t)$ requires the narrow band of signal S(t). So, the lower gamma band (30-40 Hz) is used for synchronisation analysis as it is in narrow band.

Ratio of centre frequency to band width in lower gamma band =35/10=3.5

Ratio of centre frequency to band width in normal gamma range (30-80 Hz) = 55/50=1.1

2.8 WAVELET PHASE SYNCHRONIZATION

Consider a signal s(t) such that it is in time domain. An analytic signal is defined ψ(t) such that the signal can be expressed as

$$\psi(t) = s(t) + j\,W[s(t)]$$

and,

$$\psi(t) = I(t)\exp\;(j\,\varphi(t))$$

where, W{s(t)} is termed as the wavelet transform of the signal as defined of s(t), j= $\sqrt{-1}$, I(t) is termed as the instantaneous amplitude or the envelope and φ(t) is termed as the phase of the signal which is given by ψ(t). We have to take the narrowband signal as due to it the phase will not depend upon the amplitude that is it will be independent of the amplitude. The gamma band within the frequency of 30 to 40 Hz is a narrow band and as a substitute of the total gamma band in the 30 to 80 Hz this lower gamma band is used since it is narrow enough for the amplitude to be independent of the phase. The approach for the Instantaneous phase difference at any time denoted by "t" is given as

$$\varphi_{12}(t) = \varphi_1(t) - \varphi_2(t)$$

where $\varphi_1(t), \varphi_2(t)$ are termed as the first phase and second phase of the EEG signal pair. For two channels the synchronization parameter that is "s" is defined by the formula for a window W_t as given [1]:

$$s_{12}(W_t) = \left| \frac{1}{|W|} \sum_{t=1}^{|W|} \exp(i \,.\, \varphi_{12}(t)) \right|^2 \quad \dots (2.6)$$

where, the dimension of the analysis window "W" is defined by |W|. By the value of "s" we can determine whether the signal is synchronized or not. For hundred percent phase synchronization the value of "s" must be equal to 1 and for no phase synchronization the value of s must be equal to 0. Suppose we have a total of "n" EEG channels taken from each of the patient then the matrix of $I_n(t)$ which is symmetric in nature with dimension n x n is formed. The value of each of the matrix entries are denoted by $a_{ij}(t)$ and in it the i and j denotes the channels of the EEG signal which has been calculated in the above equation.

At the time of cross modal task, the EEG signals was extracted. It was performed by the segmentation of these EEG signals into the trials. Segmentation was done from one second pre and 2 second after the stimulus onset. The equ-ripple FIR filter was used for the band pass filtering of the collected EEG signals so that it can be attained in the lower gamma band within the frequency limit of 30 to 40 Hz. The band pass filter's specification is as follows-

Passband is at 30 and 40 Hz,

Stopband is at 26 and 44 Hz,

Attenuation of pass band = 0.5 dB,

Attenuation of stop band = 40 dB.

The length of the analysis window which was used for the calculation of phase synchronization is 400 milliseconds. It was then moved to one time point in a particular time. For single hemisphere of the brain the dimension of the matrix is 35 x 35 and it is symmetric in nature.

The centre line consists of a total of eight channels and these 8 channels are contained in both of the hemispheres which can be shown as $[(62-8)/2]+8=35$ number of channels, contained for single hemisphere of the brain.

2.9 MEASUREMENT BY ENSEMBLE SYNCHRONIZATION

The matrices that we have obtained for the synchronization between the two channels must be ordered. Frobenius norm is used for this process. Consider that a_{ij} is the entry in matrix of "i" row and "j" column for the matrix named "I". Thus, the norm is defined for each of the entry by the function $||I||_p$, where p is a positive number and n is termed as total number of channels in the EEG signal. Frobenius norm is selected when we consider the value of p to be two [1].

$$||I||_p = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij})^p\right)^{1/p}$$

$$\gamma = \sqrt{\frac{||I||_F^2 - n}{n^2 - n}}$$

… (2.7)

where, "n" defines the number of channels in the EEG of each subject. The value of the function ensemble synchronization lies on the scale 0 – 1. If we have the highest value of synchronization then the value of the ensemble synchronization is exactly equal to 1. For the value of this variable to be 0, it means there is no synchronization. If the value of ensemble synchronization lies between 0 and 1 the it means it is partially synchronized with each other. Using the above values, we can infer which value of the ensemble synchronization shows the maximum synchronization.

2.10 PHASE SYNCHRONIZATION STATISTICAL SIGNIFICANCE

Phase synchronization's statistical significance can be determined by using the null hypothesis H_0 property, that is if we consider two channels then the synchronization which is calculated above can be given $s_{1,2} = 0$ if the two channels selected above are not statistically important. The result which is used for the following hypothesis is to reject H_0 if $s_{1,2} > s_0$. s_0 is dependent on a particular value of p and is given by the term $100 \times (1 - p)$ percentile based on the distribution. Using the phase synchronization that is based upon the wavelet transform in the window of 400 milliseconds, it is observed that $s_0 = 0.51$ in the case of hundred pair of signals. For the value of p to be 0.05 or 95% significance level. As the window is of a small duration so it is observed that the value of s_0 is very high. It is based on out use of selection of the size of the window. If our need is to get a good temporal resolution then we should accordingly select the window length. Same type of the null hypothesis is done on the measure of ensemble synchronization that is γ. The decision rule which is accepted for the testing of null hypothesis that is $H_0$is as follows:

Reject H_0: If γ > γ_0 that is if the observed ensemble synchronization γ is greater than the γ_0, where $\gamma_0 = 100 \times (1 - \text{p})$ percentile.

Accept H_0: If γ < γ_0 that is if the observed ensemble synchronization γ is less than the γ_0, where $\gamma_0 = 100 \times (1 - \text{p})$ percentile.

Thus, by undertaking this ensemble synchronization measure it is observed that the value of $\gamma_0 = 0.51$ which is equal to s_0. By using the above measure, we can say that the value of p is equal to 0.05.

2.11 LOGISTIC REGRESSION

Consider that the number of training data set be m and the number of features extracted to be n. And X is a feature matrix which can be expressed as [1]:

$$X = \left[(x^{(1)})^T \ (x^{(2)})^T \ (x^{(3)})^T \ \dots \ \left(x^{(i)}\right)^T \ \dots \ (x^{(m)})^T \right]^T$$

$$x^{(i)} = \left[1 \ x_1^{(i)} \ x_2^{(i)} \ \dots \ x_j^{(i)} \ \dots \ x_n^{(i)}\right]$$

… (2.8)

The order of $x^{(i)}$ is $1 \times (n + 1)$ and it is termed as a feature vector for the ith sample. The value denoted by the $x_j^{(i)}$ is $x^{(i)}$ jth feature. One line of extra ones is added and is set for the intercept term, mathematically it can be expressed as by the term $x_0^{(i)} = 1$. The term by the braces $(x^{(i)}, y^{(i)})$ represents the ith sample. $y^{(i)}$ is defined as the output class by the vector Y and it has the order $m \times 1$

$$y^{(i)} = \begin{cases} 0, & Healthy\ Subject \\ 1, & Schizophrenia. \end{cases}$$

The hypothesis function here is denoted by h and it maintains the output not to exceed out of the [0,1] range. Weight is used for the proper outcome of the output and is shown by the variable $h_\theta(x^{(i)})$ with the order $1 \times (n \times 1)$ and

$$h_\theta\left(x^{(i)}\right) = \frac{1}{(1 + \exp(-x^{(i)} \cdot \theta^T))}$$

… (2.9)

The classification is done by the single trial as given

$$\left(x^{(i)}, y^{(i)}\right) = \begin{cases} Healthy\ Subject, & if\ h_\theta\left(x^{(i)}\right) < 0.5 \\ Schizophrenia, & if\ h_\theta\left(x^{(i)}\right) \geq 0.5 \end{cases}$$

The main aim what we are left with is the minimization of the cost function such as to search the minimum value of θ. And for avoiding the overfitting we regularize the data which is measured by the variable λ.

2.12 CLASSIFICATION BASED ON LOGISTIC REGRESSION

Total of three features were selected that were the ensemble synchronization measure denoted by γ. This value of γ was an array and a window of 300 milli-second was considered over the three intervals that was from the stimulus onset to 300 milli-second (first feature), from the response time that is from 100 milli-second to 400 milli-second (second feature), and the last feature from stimulus offset to 300 milli-second before it [13]. The trials of who's the values were at maximum of two standard deviations away from mean was considered. The purpose of using this is that if the subject responded very quickly or there was enough time difference then it might have been due to the confusion or due to hate or any personal reasons.

2.13 PERFORMANCE OF THE CLASSIFIER

For each of the stimulus, the logistic regression model was trained for a total of 20 data sets of who's each subject with four EEG channels were considered. Out of the 20, halves were healthy subject and the other half belonged to the schizophrenic patient. Training was done for the 20 subjects and using the above data the accuracy was seen. Of the 10 healthy subjects using the Hilbert based phase analysis and classification 7 outcomes were correct and for the 10 schizophrenic subjects using the Hilbert based phase analysis and classification 7 outcomes were correct, that is it showed an accuracy of 70%. When the wavelet-based method was used for the analysis of the classification of the 10 healthy subjects 7 outcomes were correct and for the 10 schizophrenic subjects 8 outcomes were correct, that is it showed an accuracy of 75%.

2.14 LIKELIHOOD RATING

For each subject the likelihood can be assigned where the likelihood is denoted by $lik\ (s_j)$ and s_j denotes a particular subject. This shows whether the subject belongs to the healthy or the schizophrenic group. The likelihood function is thus given by

$$lik\ (s_j) = \frac{1}{tr\ (s_j)} \sum_{i=1}^{tr\ (s_j)} score\ (i, s_j) \qquad \ldots (2.10)$$

$$s_j \in \begin{cases} healthy\ subject, & lik\ (s_j) < threshold \\ schizophrenic, & lik\ (s_j) \geq threshold \end{cases}$$

CHAPTER 3

WAVELET TRANSFORM

3.1 INTRODUCTION

In the middle of 1980's wavelet transform was introduced as a new kind of mathematical tool. For locally analysing the nonstationary and wide band signal with fast transition wavelet transform shows efficient results. In wavelet transform time signal is mapped to time scale joint representation which is analogous to Wigner distribution, ambiguity function and short-time Fourier transform action over the time signal. In it the signal's temporal aspect is conserved. It is also used for multiresolution analysis of the signal in terms of dilated windows.

Narrow windows are used for the analysis at high frequency and for the analysis at lower frequency wide windows are used. The wavelet transform can also be called as constant-Q analysis. With the help of wavelet functions like translation and dilation the base of wavelet transform is formed. For reconstructing the original signal by the inverse wavelet transform such kind of wavelet functions provide the suitable required conditions. The regularity condition is also fulfilled by the wavelets such that decrement in wavelet coefficient occur along with the decrement in the scale. Locality is shown in both the time and frequency domain by the wavelet transform.

The discrete wavelet transforms with discrete translation and dilation of continuous wavelets [14] can be used for reducing the product of time and bandwidth. Scaling based function i.e. orthonormal wavelet transform is used for the multiresolution analysis of the signal. At every level of resolution, the discrete translation of scaling function forms an orthonormal basis. The scaling function basis forms the wavelet basis. At every level of resolution these two basis function are mutually orthogonal. Average approximation is made by projecting function orthogonally over the scaling function. The difference between the two approximations at two adjacent resolution level is represented by the orthogonal projection onto wavelet basis [15]. Both the regularity condition as well as the orthonormality condition is satisfied by the scaling functions and the wavelets.

The two discrete high pass and low pass band filters those are Para unitary perfect quadrature mirror filters with additional regularity development in sub-band coding theory are used in multiresolution analysis framework for the computation of the decomposition and reconstruction in discrete orthonormal wavelet. O(L) operations where length is L of the data vector is required for discrete wavelet transform in tree algorithm. Here the product of bandwidth and time of the output of the wavelet transform is increased slightly according to the signal. It is a powerful tool for image processing, pattern recognition and multiresolution analysis of local spectrum of nonstationary signal like seismic, radar, sound, sonar and ECG signals.

3.2 CONTINUOUS WAVELET TRANSFORM

Consider **L** as the vector space of the measurable, square-integrable functions. The continuous wavelet transform of a function f(t) ϵ **Z** is merely the decomposition of the function $h_{s,\tau}(t)$ called the wavelets:

$$W_f(s,\tau) = \int f(t)\, \psi^*_{s,\tau}(t)\, dt$$

… (3.1)

Here * denotes the conjugate function but most of the wavelets are real valued functions. The generation of the wavelets occurs from the single basic wavelet (mother wavelet) by the method of scaling and translating that is given as:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-\tau}{s}\right)$$

… (3.2)

Here τ is the translation factor and s is the scaling factor. Usually scaling factor is taken as positive i.e. s >0. When s >1 then wavelets are dilated. And when s <1 then wavelets are contracted. $h_{s,\tau}(t)$ is generated from the basic wavelets having different locations τ and scales s having identical shapes. $\sqrt{s}$ is a constant which denotes the energy normalisation of the wavelets.

The normalisation of the wavelet is done as following:

$$\int \left|\psi_{s,\tau}(t)\right|^2 dt = \int |\psi(t)|^2 dt = 1$$

… (3.3)

It is normalised such that the wavelets have the same energy scaled by the factor s. The normalisation can also be done with respect to the amplitude as following:

$$\int \left|\psi_{s,\tau}(t)\right| dt = 1$$

… (3.4)

In above case the normalisation constant is s^{-1} instead of $s^{-1/2}$ and with the help of basic wavelets, the wavelets can be generated as:

$$\psi_{s,\tau}(t) = \frac{1}{s}\,\psi\,(\frac{t-\tau}{s})$$

… (3.5)

On substituting equation (3.2) into equation (3.5) the wavelet transform can be written as correlation between the signal and the scaled wavelets h(t/s):

$$W_f(s,\tau) = \frac{1}{\sqrt{s}} \int f(t)\,\psi^*\left(\frac{t-\tau}{s}\right) dt$$

… (3.6)

MORLET WAVELET

The most commonly used complex wavelet is the Morlet wavelet. The mother wavelet is given as,

$$\psi(t) = \pi^{-\frac{1}{4}}\left(e^{j\omega t} - e^{-\frac{\omega^2}{2}}\right) e^{-\frac{t^2}{2}}$$

… (3.7)

where, ω is termed as the central frequency of the mother wavelet. For correcting the non-zero mean of the complex sinusoid $e^{-\frac{\omega^2}{2}}$ term is used. This term will be negligible when $\omega > 5$. Hence in most of the work the mother wavelet of the Morlet wavelet is:

$$\psi(t) = \pi^{-\frac{1}{4}} \left(e^{j\omega t}\right) e^{-\frac{t^2}{2}}$$

… (3.8)

The Morlet wavelet transform provides a bridge between the frequency and time information [16] which can further be used in the biomedical signal processes for the neurological aspects as well as the head related trauma aspects.

3.3 DISCRETE WAVELET TRANSFORM

In continuous wavelet transform one-dimensional time signal is mapped to a two-dimensional time-scale joint representation. The time bandwidth product of the continuous wavelet transform is given as the square of the signal. But the main aim of the signal processing is to represent the signal as efficiently as possible with less parameters. The use of discrete wavelet transform can thus reduce the time bandwidth product of the outputs.

CHAPTER 4

LOGISTIC REGRESSION AS A CLASSIFIER

This chapter deals with the estimation of the probability function that is $P\ (y_q \mid S_p\ , x_q)$ or we can say that if the system is S_p and is fixed then for the input x_q what is the probability of getting the output that is denoted by y_p. If the output is categorically defined then we can approximate the value of $P\ (y_q \mid S_p\ , x_q)$ and is termed as the weighted logistic regression and its based on the memory.

Consider a system such that its denoted by S_p, where the input space is taken to be two dimensional and the output to be Boolean. Consider we have the data (x_q, y_q) which is not depicted in the graph, then to label it we should know what type of the system it is. In this proposed weighted logistic regression, it is considered that the information of the system comes through the earlier observation of the system.

4.1 CLASSIFICATION METHODS

As the approximation of the function which is denoted by $P\ (y_q \mid S_p\ , x_q)$ is a question for classification among themselves because the output y is categorically defined. Many classification techniques have been defined that includes:

a) Nearest neighbour,
b) Kernel regression,
c) Bayes classifier,
d) Decision tree classification,
e) Support vector machine,
f) Perceptron
g) Artificial neural networks / Deep learning

The most basic classifier is the nearest neighbour. k-nearest neighbour is more popular and it is the derivative of the nearest classifier. If we consider kernel regression method then it is more vital. The methods which is named above are termed as the memory-based methods, and if we consider the non-memory-based method, these are neural network, decision-tree, hierarchical mixtures of experts (HME), etc. But as compared to each other both of the classifiers take the understanding of the system, which is gathered from the training data sets. One of the most important features of the memory-based classification is that only when a particular query is processed then they comply to all the data which has been processed. Thus, the memory-based classifiers are necessary for the processing of the real time data as it is desirable for the continuous stream of the data. One more advantage of the memory-based classifiers is that they can self-tune based on the distribution and the noises of the training data set. In the case of the non-memory-based classifiers even when the query is not made it tries to learn the systems properties. Brief of some mostly used classifiers are discussed below:

1. NEAREST NEIGHBOUR: This classifier doesn't work accurately in most of the cases as its much sensitive. The sensitiveness is mostly for the noise in the case of single nearest data set. To overcome the following reasons k- nearest neighbour is performed and it works well as compared to the single nearest neighbour classifier. The disadvantage for the k nearest is that it is not able to detect the boundary of changing outlines.
2. KERNEL REGRESSION: It is a better method as compared to the k-nearest neighbour but we cannot extrapolate it. If there is a query away from the centroid of another set then this classifier is not sufficient to classify such data sets. Rather than classifying, it gives 50% probability to the aforementioned case.
3. BAYE'S CLASSIFIER: It sets much stronger approximation on how the data are distributed. The classifier classifies into two class if the output generated is Boolean. Dual clusters formation takes place, that is one when the output is generated as zero and the another one when the data produces an output equal to one. The formation of the clusters that occur is of the shape of ellipse. This is because the distribution in the input space is gaussian in nature.
4. DECISION TREE: In this classifier the input data set is divided or bifurcated into small sections. This small section which are partitioned are labelled as the output of each of the categories. The decision tree which has been used mostly divides the outcome into two categories only. And if we try to extend this tree then it leads to the overfitting problem.

Here, the logistic regression is capable of being used as a memory-based classifier and it also includes the properties of other memory-based classifiers [17].

4.2 SETTING UP CONDITIONAL PROBABILITY

Until this time our study is mainly involved in the approximation of the conditional probability for the continuous variables and estimation of the function or variable distribution. But in most of the cases of classification such as in the case of regression we are mostly interested for the relationship among the input and the output. It should be noted that the output is mainly discrete whether it may be a binary outcome or multiple discrete outcome as compared to the continuous input. We also have a case where the input can be discrete. The classification of such data can be done by following a particular rule according to which we can guess the binary output as compared to the input variable. This is termed as classification which is used in most of the cases when statistical data is present and the machine learning. Merely, by observing the data and guessing the output either as "yes" or "no" is no way for the classification, mainly when no rule is followed. Summarizing the above condition, we can say that we need probabilities so that the given set of data can be fit for a stochastic model.

The probability that will be considered is the distribution of the output Y, if we know the input data, that is P(Y|X). It would describe how accurate the predictions are and will move the system towards more reliable one. If the system estimates that there is a 53% probability that it will rain and it doesn't rain then our system will be better than as compared to the system according to which there is a 97% probability of rain and this 97% is also not reliable. The non-parametrical probability can be approximated and it can be done by the use of kernels for the discrete variables. Consider two classes and assign it with either 0 and 1. The Y is the outcome variable and we can say that P (Y = 1) = E [Y]. Using the same we can say that P (Y = 1|X = x) = E [Y|X = x]. Now we will get the most favourable estimate of the regression function for the output variable, and it will serve for the approximation of the conditional probability function. Two properties that can defer to the condition discussed above are:

a. The probability value always should lie in the range of 0 and 1, while the smoothers which is defined above will not maintain it. If the value of y_i we are getting is either 0 or 1 then also the smoothening will not maintain the probability in the range of 0 and 1.

b. The second property that can defer to the idea is that rather than approximating the output we can model the system.

Consider, P (Y = 1 | X = x) = p(x ; θ), where p is any function parameterized by θ. Assume that the outcomes are not dependent to each other. Then the conditional likelihood function can be defined as [18]-[19],

$$\prod_{i=1}^{n} P\,(\,Y = y_i |\, X = x_i\,) = \prod_{i=1}^{n} p(x_i;\theta)^{y_i}(1 - p(x_i;\theta))^{1-y_i}$$

… (4.1)

In the Bernoulli trial, if the success probability is p the likelihood is given by

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

… (4.2)

And, the maximization of this likelihood function is given by p when its value is $p = \hat{p} = n^{-1}\sum_{i=1}^{n} y_i$. Suppose each of the trial has its own success probability that is p_i, then this likelihood comes out to be

$$\prod_{i=1}^{n} {p_i}^{y_i}(1-p_i)^{1-y_i}$$

… (4.3)

Now, we will have to apply some constraints for the estimation of the Bernoulli system using the maximum likelihood. Consider that p_i is not any random number but it is related to each other, such type of the approximation provides with the non-trivial estimation of parameter. The assumption that $p_i = p(\,x_i\,;\,\theta\,)$ gives us the idea that p_i should always be the same if the value of x_i is same. Also, if the value of x_i is same then p is continuous in nature which tells us that there is similarity between the values of x_i and p_i.

4.3 LOGISTIC REGRESSION

The output variable is known that is denoted by Y and further aim is modelling of the probabilities that is P (Y=1 | X=x). All the parameters which are not known have to be estimated by using the maximum likelihood criteria. This estimation can be performed by different ways:

1. First method is to consider p(x) as a linear function of x. If there is an increase or decrease in the value of x then it would also lead to an increase or decrease in the value of probability. The main thing to be kept in mind is that the value of the probability must be in the range of 0 and 1. Also the linear functions are bounded functions. In many cases it has been seen that the small change in the value of p needs a big change in the value of x. Linear model doesn't have the capability to pursue this.
2. Second method after method one fails is that, assume log p(x) as a linear function which is linear with respect to x. According to this if there is a change in the input then the probability gets multiplied by a static value. But the difficulty in this is that the log function is inbounded in only one direction but linear functions are unbounded in both of the directions.
3. Third method which is the easiest of all is to consider a log function as $\log \frac{p}{1-p}$ which is also said as the logistic transformation.

This third case is termed as the logistic regression. It can be more specifically defined as [13]

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + x \cdot \beta$$

… (4.4)

Further it can be solved for p as

$$p(x; b, w) = \frac{e^{\beta_0 + x \cdot \beta}}{1 + e^{\beta_0 + x \cdot \beta}} = \frac{1}{1 + e^{-(\beta_0 + x \cdot \beta)}}$$

… (4.5)

For the minimization of this error rate, Y=1 must be considered when the value of $p \geq 0.5$ and the value of Y = 0 when the value of $p < 0.5$. According to the above sentence we can say that the value of Y=1 when the $\beta_0 + x \cdot \beta$ value is non-negative. Thus, this discussed regression that is the logistic regression model gives us the linear classifier. The decision boundary is selected by $\beta_0 + x \cdot \beta = 0$ solution, according to which if x is 1-D then it is a point, if x is 2-D then the decision boundary is a line.

Logistic regression is mostly used and is used in the analysis of the discrete data and statistics. Four main reasons are:

a. Tradition.
b. For the analysis of the contingency table the term log $\frac{p}{1-p}$ has an important role. Classification can be understood as having a contingency table with binary rows and infinite number of rows. So, the logistic regression is termed as the linear interpolation of the contingency table.
c. Logistic regression is distribution based on the exponential function.
d. It has good accuracy as compared to most of the other classifiers.

CHAPTER 5

SIMULATION RESULT

Single trial EEG classification which is based on the task by using the patterns generated through the impulses of the brain is one of the challenging problems. All the mental activity that includes the thought process, senses and perception is extremely fast and thus it must be investigated in a defined time window. In the previous works the classification of healthy patient and the schizophrenic has been attempted with an accuracy of 71%.

Solution: Logistic regression along with the ensemble synchronization for the classification of single trial EEG signal was performed. For the synchronization among the different channels in a window chosen, wavelet transform was used. The wavelet considered was a Morlet wavelet. Using this synchronization values obtained between the different pairs of channels, normalization was done and the average was taken and a new measurable quantity that is the ensemble synchronization was obtained. As the ensemble synchronization values were normalized in the range of 0-1 so it was compared between each channel to understand the synchrony. The features were obtained from the ensemble synchronization array. A window of 300 interval was taken in three distinct regions of the graph and was used through the logistic regression model for the classification.

The dataset for the EEG signal has been obtained form the Laboratory of Neurophysiology and Neuro-Computer Interfaces (formerly the Human Brain Study Group - HRCM).

EEG data archive for two classes that is the healthy and the schizophrenic subjects were given. Two classes that is the healthy comprised of total 39 patients out of which only the 10 subjects were considered in this work, and the schizophrenic subjects of total 45 patients out of which only the 10 subjects were considered in this work. The txt file contained the column with 16 channels of data. The amplitude of the EEG signal was in terms of mkV. The first 7680 samples belonged to the first channel after that the second channel from 7681^{st} sample to $15,360^{th}$ sample and so on.

In this EEG data the sampling rate considered was 128 Hz so the 7680 samples referred to the 60 seconds of the EEG recording. The location of each of the electrodes where the signals were taken are shown in the given figure. A total of 16 channels were considered for the extraction of data that is named from channel 1 to channel 16 as F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2 respectively.

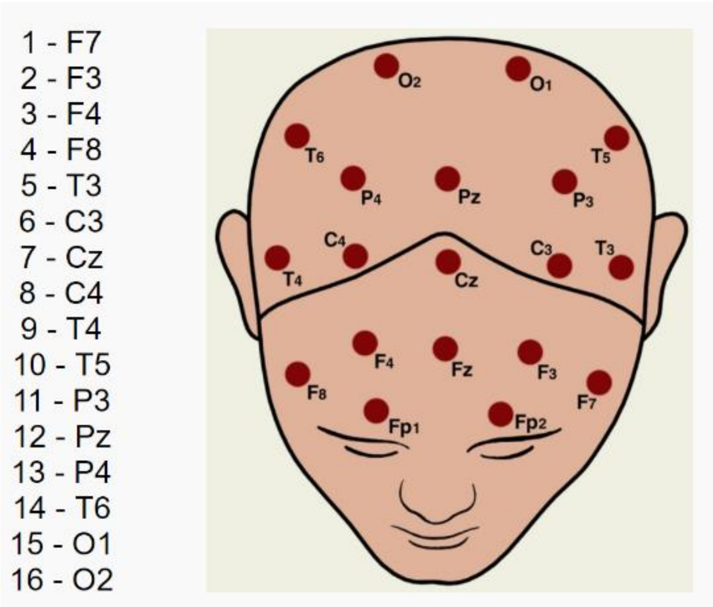


Fig 5.1. Topographical position of the channels.

From each of the .txt file the 7680 samples were extracted in the .mat file and was named as norm_1_ch1 that is for the first normal patient's channel one data similarly schzph_1_ch1 for the first schizophrenic patient's channel one data.

A total of 20 patients were considered out of which 10 belonged to the healthy group and the other 10 belonged to the schizophrenic group. For each of the subject four channels of the EEG signals were considered and these four signals were ensembled to a single array by the ensemble measure. The synchronization matrix for the four EEG channel of each subject were obtained and was used to obtain the ensemble synchronization measure in the range of 0 and 1 by normalization using the Frobenius norm.

The synchronization matrix for the first normal patient is as shown:

	Channel 1	Channel 2	Channel 3	Channel 4
Channel 1	1	0.4005	0.0039	0.0156
Channel 2	0.4005	1	0.0103	0.0381
Channel 3	0.0039	0.0103	1	0.2822
Channel 4	0.0156	0.0381	0.2822	1

Fig 5.2. Phase synchronization matrix for first normal patient.

The ensemble measure calculation from the above phase synchronization matrix was done for each of the patient and as a sample for the first normal patient, it plotted for 7680 samples as shown in the next page.

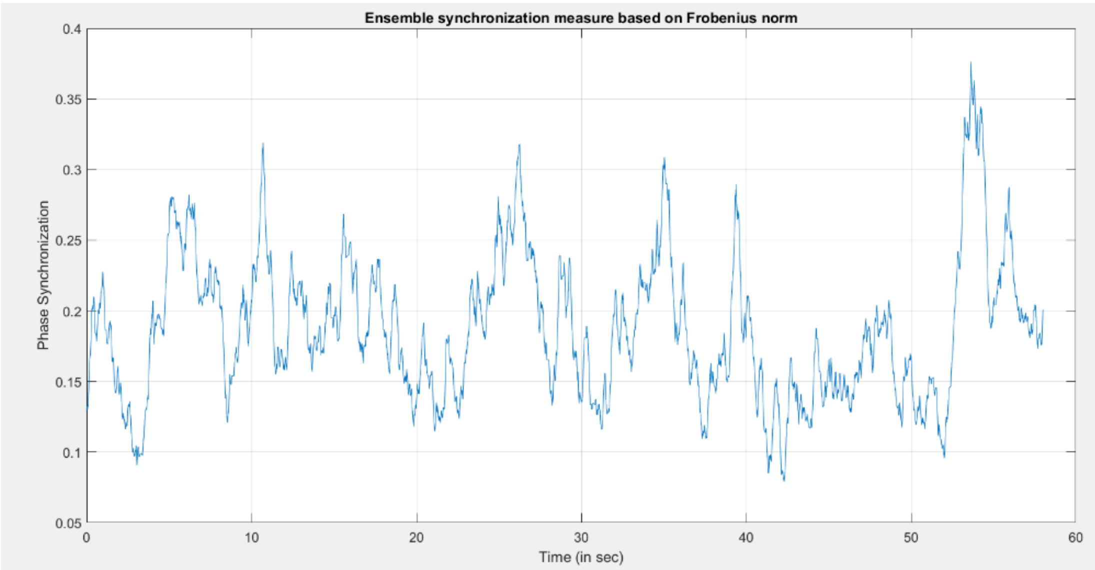


Fig 5.3. Ensemble synchronization measure for first normal patient (Hilbert based).

Serial No.	Stimulus Onset	Stimulus Offset	Response Time	Classifier Value
1	0.3742	0.3313	0.3704	0
2	0.406	0.3699	0.3085	0
3	0.3705	0.3769	0.4284	0
4	0.3405	0.301	0.3013	0
5	0.5432	0.4115	0.4855	0
6	0.4639	0.4639	0.5177	0
7	0.3664	0.3678	0.3574	0
8	0.2028	0.2261	0.2213	0
9	0.2908	0.2629	0.2991	0
10	0.4112	0.3528	0.3369	0
11	0.4699	0.4878	0.4068	1
12	0.3138	0.3009	0.2940	1
13	0.2595	0.2204	0.2244	1
14	0.4219	0.4229	0.4905	1
15	0.3133	0.2913	0.3235	1
16	0.3392	0.3372	0.342	1
17	0.2827	0.2518	0.2829	1
18	0.3162	0.3020	0.3328	1
19	0.2784	0.3111	0.2928	1
20	0.3650	0.4115	0.2828	1

Fig 5.4. Features extracted for the Hilbert based classification.

Using the Hilbert based classification out of the 10 normal subjects it classified 7 as correct and out of the 10 schizophrenic subjects it classified 7 as correct. Thus, out of overall 20 subjects 14 were correctly classified. Thus, the classification accuracy was 70%.

Using the Wavelet based classification out of the 10 normal subjects it classified 7 as correct and out of the 10 schizophrenic subjects it classified 8 as correct. Thus, out of overall 20 subjects 15 were correctly classified. Thus, the classification accuracy was 75%.

Serial No.	Stimulus Onset	Stimulus Offset	Response Time	Classifier Value
1	0.2707	0.2548	0.2779	0
2	0.299	0.2585	0.2251	0
3	0.2993	0.2993	0.3510	0
4	0.2451	0.2176	0.2132	0
5	0.4368	0.3356	0.3280	0
6	0.3610	0.3522	0.3977	0
7	0.2927	0.3020	0.2064	0
8	0.1775	0.1792	0.1521	0
9	0.2202	0.2117	0.2395	0
10	0.3158	0.2671	0.2633	0
11	0.3322	0.3554	0.2717	1
12	0.2153	0.2023	0.2113	1
13	0.1962	0.1428	0.1679	1
14	0.3334	0.3665	0.3469	1
15	0.2177	0.2152	0.2031	1
16	0.2360	0.2397	0.2564	1
17	0.2080	0.1788	0.2102	1
18	0.2533	0.2379	0.2518	1
19	0.2027	0.2263	0.2352	1
20	0.2834	0.3217	0.2352	1

Fig 5.5. Features extracted for the Wavelet based classification.

ROC: A receiver operating characteristics curve or a ROC curve illustrates graphically the diagnostic ability of the classification that can be performed by the binary classifier by varying the discrimination threshold. It is drawn by taking true positive rate (or sensitivity or recall) in the y axis against the false positive rate (or fall-out or 1-specificity). ROC is a tool which is used for selecting the optimal model and discarding the sub-optimal models.

The ROC Curve of the Classification based on the Hilbert phase and the proposed wavelet based is shown in Figure 5.6 and 5.7.

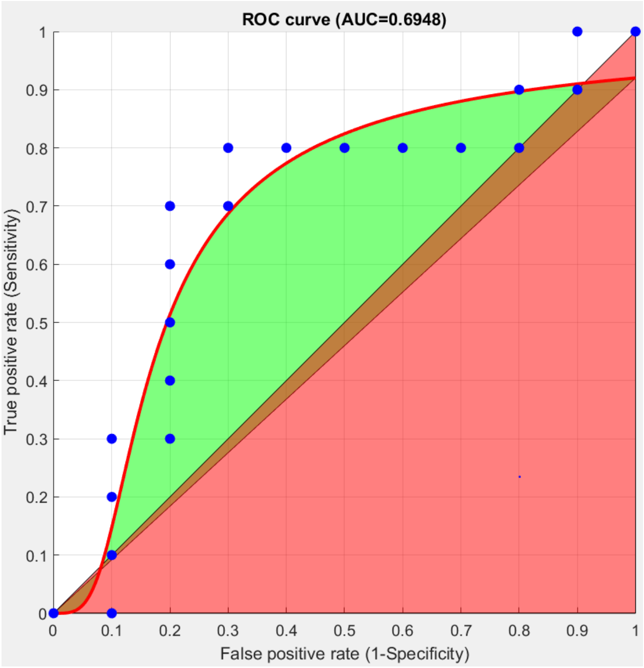


Fig 5.6. ROC for the Hilbert based Classification.

Confusion Matrix: A confusion matrix is the most popular table which is used for describing the performance of the classification model that is a classifier. It is used on the set of the testing data for which we already have the true values. Different parameters are named as TP, TN, FP, FN, precision, accuracy, kappa, F1 score, FPR and TPR.

The basic terms involved in the confusion parameters are as follows:

1. True Positive: It refers to the case where we have predicted YES and they do have the disease.
2. True Negative: The case where the prediction is NO and they do not have that disease too.
3. False Positive: In this case we have predicted YES but they don't actually have the disease. It is termed as the TYPE I error.
4. False Negative: Here we have predicted NO but they have the disease. It is termed as TYPE II error.

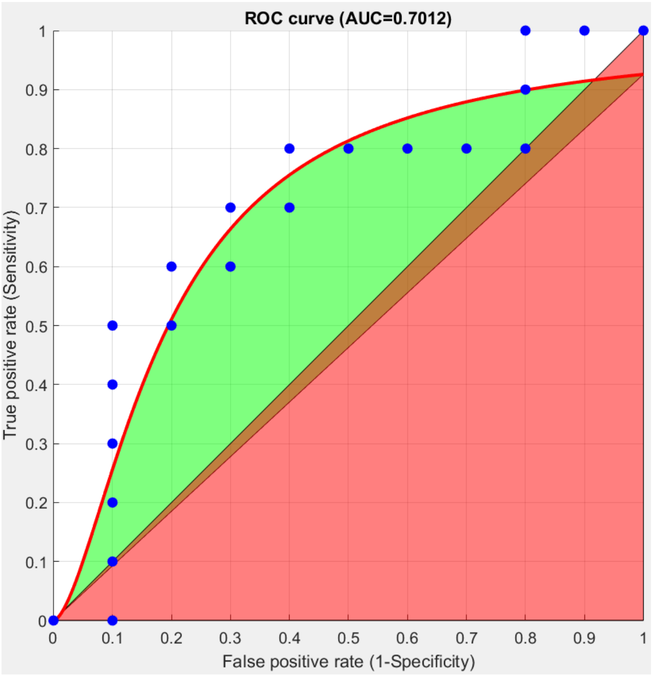


Fig 5.7. ROC for the Wavelet based Classification.

	PREDICTED: NO	PREDICTED: YES	
ACTUAL: NO	TN = 7	FP = 3	10
ACTUAL: YES	FN = 3	TP = 7	10
	10	10	

Fig 5.8. Hilbert based Confusion Matrix.

	PREDICTED: NO	PREDICTED: YES	
ACTUAL: NO	TN = 7	FP = 3	10
ACTUAL: YES	FN = 2	TP = 8	10
	9	11	

Fig 5.9. Wavelet based Confusion Matrix.

Different list of rates which are often extracted from a confusion matrix for the binary classifier are as follows:

1. Accuracy: It tell us how often is our classifier correct.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision: When the prediction is YES, how often is it correct.

$$Precision = \frac{TP}{TP + FP}$$

3. True Positive Rate: When the answer is actually YES, how often does it predict YES. It can be also termed as Recall.

$$TPR = \frac{TP}{TP + FN}$$

4. False Positive Rate: When the answer is actually NO, how often does it predict YES.

$$FPR = \frac{FP}{FP + TN}$$

5. Kappa: It gives us an idea of how well the classifier has performed as compared to how well it had performed by chance. It considers the possibility that the outcome is classified by chance.

6. F1 Score: It can be defined as the measure of the recall and precision considered at a particular threshold.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Further different parameters which are often extracted from a confusion matrix for the binary classifier can be illustrated by Figure 5.10.

TABLE I

COMPARISON BETWEEN THE PRESENT HILBERT BASED METHOD AND THE PROPOSED WAVELET BASED METHOD.

S.No	Stimulus Parameters	Hilbert Method (HM)	Wavelet Method (WM)
1	Training Set Accuracy (%)	70	75
2	True Positive Rate (TPR)	0.70	0.77
3	False Positive Rate (FPR)	0.30	0.20
4	Area Under Curve	0.694	0.701
5	Kappa	0.4	0.5
6	Standard Error	0.12	0.11
7	Precision	0.70	0.78
8	Overall Error (%)	30	25
9	Sensitivity	0.70	0.70
10	Specificity	0.70	0.80
11	F1_Score	0.70	0.74

The performance of the logistic classifier for the detection of healthy or the schizophrenic subject is trained and the testing is done on the 80 channels of the 20 subjects. The ROC is shown for the Hilbert-based and the wavelet-based classifier in Figure 7.5 and 7.6 respectively. The performance is measured by the area under the curve of the ROC which is the most essential parameter. For the reliable performance of the classifier the area under the curve must be high. It can be seen that the area under curve for the wavelet-based classification is better than as compared to the area under the curve for the Hilbert-based classification.

The single-trial average accuracy of the classifier is observed to be 70% and area under curve as 0.69 for Hilbert-based classification but to be 75% and 0.70 for the wavelet-based classification.

		True condition			
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$ / F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$	

Fig 5.10. Confusion Matrix and parameters.

CHAPTER 6

CONCLUSION AND FUTURE WORKS

Electroencephalography and Magnetoencephalography is one of the best options for obtaining high temporal resolution of the brain waves. It gives us a detail as to how the cognitive ability works. Most important tool for the measurement of the ensemble action is the phase synchronization which it has been performed using the wavelet-based method thus, increasing the accuracy of the classification.

Since, the single trial EEG classification is one of the most difficult tasks then the task-based classification is much more difficult. Here, the classification of the subjects based on the EEG data form the patients using the ensemble synchronization is performed. The future aim is to increase the number of subjects while limiting the channels with the inclusion of other bands.

REFERENCES

[1] Pineda, J. A. et al. “The effects of self-movement, observation, and imagination on μ rhythms and readiness potentials (RP’s): toward a brain-computer Interface (BCI)”, IEEE Trans. Rehabil. Eng., 2000, 8(4): 219.

[2] Juergen Fell and Nikolai Axmacher “The role of phase synchronization in memory processes”, PMID: 21248789, DOI: 10.1038/nrn2979, 2011 Feb;12(2):105-18.

[3] Buzsáki, G. Rhythms of the Brain (Oxford University Press, New York, 2006).

[4] Steriade, “Impact of network activities on neuronal properties in corticothalamic systems”, J. Neurophysiol. 86, 1–39 (2001).

[5] Lakatos, P. et al. “An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex”, J. Neurophysiol. 94, 1904–1911 (2005).

[6] Yong Li, Guoya Dong, Xiaorong Gao, Shangkai Gao, Manling Ge, Weili Yan “Single trial EEG classification during finger movement task by using hidden Markov models”, Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering Arlington, Virginia · March 16 - 19, 2005.

[7] Buzsáki, G. & Draguhn, “A Neuronal oscillations in cortical networks”. Science 304, 1926–1929 (2004).

[8] S. C. Gandevia and J. C. Rothwell, “Knowledge of motor commands and the recruitment of human motoneurons," Brain, vol. 110, no. 5, pp. 1117-1130, 1987.

[9] G. Pfurtscheller, Ch. Neuper, D. Flotzinger, and M. Pregenzer, “EEG-based discrimination between imagination of right- and left-hand movement," Electroenc. clin. Neurophys., vol. 103, no. 5, pp. 1-10, 1997.

[10] Jinghuai Gao, Xiaolong Dong, Wen-Bing Wang, Youming Li, and Cunhuan Pan “Instantaneous Parameters Extraction via Wavelet Transform” IEEE transactions on geoscience and remote sensing, vol. 37, no. 2, march 1999.

[11] M. Taner, F. Koehler, and R. E. Sheriff, “Complex seismic trace analysis,” Geophysics, vol. 44, pp. 1041–1063, 1979.

[12] M. Taner and R. Sheriff, “Application of amplitude, frequency and other attributes to strati-graphic and hydrocarbon determination,” in Applications to Hydrocarbon Exploration: Memoir 26, C. Payton. Ed. Tulsa, OK: Amer. Assoc. Petroleum Geologists, 1977, pp. 389–416.

[13] Pradeep D. Prasad, Harsha N. Halahalli, John P. John, and Kaushik K. Majumdar “Single-Trial EEG Classification Using Logistic Regression Based on Ensemble Synchronization” IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 3, May 2014,1074-1080.

[14] S. Mallat “A theory for multiresolution signal decomposition: the wavelet representation”, IEEE Tram Pattern Anal. &Mach. Intell., 1989, PAMI-11, pp.674-693.

[15] I. Daubechies “Orthonormal bases of compactly supported wavelets”, Communications on Pure and Applied Mathematics, 1988,41, pp.909-996.

[16] Rabeb Touati, Imen Massaoudi, Afef Elloumi Oueslati, Zied Lachiri and Noureddine Ellouze "Nucleosome location method based on Morlet wavelet analysis scalograms investigation", 2nd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016 March 21-24, 2016, Monastir, Tunisia, pp.307-312.

[17] T. Haifley “Linear logistic regression: An Introduction”, 2002 IRW final report, pp.184-187.

[18] D. W. Homer and S. Lemeshow, “Applied Logistic Regression”, Second Edition, Wiley Interscience (2000).

[19] Logistic Regression – CMU Carnegie Mellon University Statistics https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf