Analysis of epilepsy related abnormalities using EEG

A Dissertation submitted towards the partial fulfilment
of the requirement for the award of degree of

Master of Technology
in
Signal Processing & Digital Design

Submitted by

UPENDRA UPKAR SINDHAV

2K14/SPD/19

Under the supervision of

Sh. Rajesh Birok
(Associate Professor, Department of ECE)

Department of Electronics & Communication Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Delhi-110042

JUNE 2016

DELHI TECHNOLOGICAL UNIVERSITY
Established by Govt. Of Delhi vide Act 6 of 2009
(Formerly Delhi College of Engineering)
SHAHBAD DAULATPUR, BAWANA ROAD, DELHI-110042

CERTIFICATE

This is to certify that the dissertation title "**Analysis of epilepsy related abnormalities using EEG** " submitted by **Mr. UPENDRA UPKAR SINDHAV, Roll. No. 2K14/SPD/19**, in partial fulfilment for the award of degree of Master of Technology in

"**Signal Processing and Digital Design (SPDD)",** run by Department of Electronics & Communication Engineering in Delhi Technological University during the year 2014-2016., is a bonafide record of student's own work carried out by him under my supervision and guidance in the academic session 2015-16. To the best of my belief and knowledge the matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

Mr.Rajesh Birok

Supervisor

Associate Professor (ECE)

Delhi Technological University

Delhi-110042

DECLARATION

I hereby declare that all the information in this document has been obtained and presented in accordance with academic rules and ethical conduct. This report is my own work to the best of my belief and knowledge. I have fully cited all material by others which I have used in my work. It is being submitted for the degree of Master of Technology in Signal Processing & Digital Design at the Delhi Technological University. To the best of my belief and knowledge it has not been submitted before for any degree or examination in any other university.

UPENDRA UPKAR SINDHAV

M. Tech. (SPDD)

2K14/SPD/19

Date: _ JULY, 2016

Place: Delhi Technological University, Delhi

ACKNOWLEDGEMENT

I owe my gratitude to all the people who have helped me in this dissertation work and who have made my postgraduate college experience one of the most special periods of my life.

Firstly, I would like to express my deepest gratitude to my supervisor **Sh. Rajesh Birok**, Associate Professor (ECE) for his invaluable support, guidance, motivation and encouragement throughout the period during which this work was carried out. I am deeply grateful to **Prof. Prem R. Chadha,** H.O.D. (Deptt. Of E.C.E) for their support and encouragement in carrying out this project.

I also wish to express my heart full thanks to all faculty at Department of Electronics & Communication Engineering of Delhi Technological University for their goodwill and support that helped me a lot in successful completion of this project.

Finally, I want to thank my parents, family and friends for always believing in my abilities and showering their invaluable love and support.

UPENDRA UPKAR SINDHAV

M. Tech. (SPDD)

2K14/SPD/19

ABSTRACT

Seizure prediction is an issue in biomedical science which now is conceivable to understand with machine learning techniques. A seizure forecast framework has the ability to help those influenced by epilepsy in better dealing with their pharmaceutical, day by day exercises and enhancing the personal satisfaction. Utilization of machine learning calculations and the accessibility of long haul Intracranial Electroencephalographic (iEEG) recordings have immensely diminished the confusions required in the testing seizure expectation issue. Information, as iEEG was gathered from canines with actually happening epilepsy for the examination and a seizure forecast framework comprising of a machine learning based pipeline was executed to produce seizure notices when potential pre-ictal movement is seen in the iEEG recording. A correlation between the distinctive removed components, dimensionality decrease methods, and machine learning systems was performed to explore the relative viability of the diverse strategies in the use of seizure forecast. The machine learning convention performed essentially superior to a chance expectation calculation in all the broke down subjects. Also, the examination uncovered subject-particular neurophysiological changes in the extricated highlights before lead seizures recommending the presence of an unmistakable, identifiable pre-ictal state.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Motivation For This Work

Epilepsy is a gathering of neurological maladies portrayed by epileptic seizures.[1][2] Epileptic seizures are scenes that can fluctuate from brief and about imperceptible to long stretches of overwhelming shaking.[3] These scenes can bring about physical wounds incorporating sometimes softened bones.[3] epilepsy, seizures have a tendency to repeat, and have no prompt hidden cause.[1] Isolated seizures that are incited by a particular cause, for example, harming are not esteemed to speak to epilepsy.[4] People with epilepsy in a few zones of the world experience disgrace due to the condition.[3]

Around 0.5-1% of the world's populace are affected by epilepsy [1] and World Health Organization (WHO) reports that epilepsy represents almost 1% of the whole worldwide weight of ailments . The instability in the event of seizures is seen as the most noteworthy reason for epilepsy related inability [4, 5, 6]. Predictable uneasiness about when the following seizure will happen has been communicated by even patients with rare seizures [6]. Individualized epilepsy treatment could be made conceivable by the capacity to anticipate seizures in a convenient way so that the patients could be cautioned and take pharmaceuticals just when required. Seizure anticipating has turned into a noteworthy examination enthusiasm because of the potential clinical effects [7].

In 70% cases these epileptic seizures are treatable by prescription .so in the event that we can conjecture the up and coming seizures then by taking pre pharmaceutical one can spare himself/herself from embracement.

Machine learning based methodologies of seizure estimating have the accompanying strides. Estimations from the mind are taken in some structure (distinctive estimations incorporate scalp EEG, iEEG, FMRI, and so forth.). Since crude estimations are generally exceptionally uproarious and less uncovering, they are changed into components which condense the essential changes in the crude signs. A machine learning calculation consolidating these components is then used to make expectations on seizure event. The components to which the crude signs are changed over are normally picked in a manner that the elements are dynamic in seizure related movement. In any case, what these chosen works neglect to consider is the way that the components can be subject ward. A component which is found to add to seizure related action of one subject may not add to that of another. In this postulation, the likelihood of these components being subject ward has been considered and a system to dissect which highlight.

1.2 Problem statement:-

1. My main aim is to forecast epileptic seizure .since pre-ictal and normal EEG patterns are not much distinguishable visibly, so it is good practice to classify them using machine learning.
2. Another problem in methods of classification using machine learning is requirement of high computational power (due to bulk features sets) and also need large time for execution. This problem of computational complexity can be vanished or decreased by imply various feature reduction/dimension reduction

techniques on feature sets before classify them.

1.3. Approach for problem solving:-

In this thesis I have CHB-MIT SCALP EEG database. For solving problems statement's problems, I have extracted various feature from pre-ictal EEG samples and normal EEG samples taken from CHB-MIT SCALP database. For this I have accumulated DWT coefficients of the samples and after it following features are extracted from each DWT of samples:-

a. Mean of absolute values of wavelet coefficients
b. Mean power of wavelet coefficients
c. Standard deviation of wavelet coefficients
d. Ratio of mean values of adjacent sub-bands in a sample

After this I have implemented various dimension reduction algorithms on these feature sets i.e. PCA, LLE, ICA, ISOMAP, MDS and LAPLACIAN EIGEN MAPS.

After this I have done classification using SVM to know the performance of these dimension reduction techniques.

1.4. Contributions/Results

In this thesis I have evaluate performance of various dimension reduction techniques using performance evaluation parameter 'sensitivity' and 'specificity'.
Using these dimension reduction techniques I have reduced computational complexity from 80% to 86.66%, which is excellent performance from the point of

view of real time forecasting. My all approaches gives above 90% of sensitivity.

1.5. Organization

The remaining part of thesis is organized as follows. Chapter 2 provides a back-ground on epilepsy, seizures and EEG. Chapter 3 describes general approach of forecasting of epileptic seizure. Chapter 4 describes the datasets available and the data collection methods. Chapter 5 explains the methods of features detection .Chapter 6 explains the various dimension reduction and features reduction methods so one can forecast epileptic seizures with low computational complexity. Chapter 7 describes the various methods of classification. Chapter8 constituted by related field work and comparative study of previously given methods. Chapter 9 shows my approach for solving this problem. Chapter 10 presents the evaluation of the method, results and discussion. Chapter 11 discusses the conclusion and future directions for this work.

CHAPTER 2: EPILEPSY, SEIZURES AND EEG

2.1. What Is Epilepsy

Epilepsy has been considered as a shallow mental issue for a considerable length of time. In any case, today, epilepsy is distinguished as a neurological issue of the focal sensory system. The basic physiological wonders that cause epilepsy still stay obscure. In any case, it is generally watched that epilepsy is com-mon among the individuals who had experienced cerebrum related wounds or maladies. In kids and youthful grown-ups, hereditary scatters, innate variations from the norm, and birth injury influencing the cerebrum are ordinarily considered as the reasons for epileptic indications. Then again, in full grown-ups and the elderly, strokes, tumors, and cerebrovascular sickness are viewed as the causes.

Despite the fact that epilepsy is an undeniably compounding issue (i.e., every seizure harms the mind), those influenced by epilepsy are equipped for a standard vocation and family lives. Be that as it may, they are not encouraged to participate in exercises, for example, driving, swimming, and so on, which the event of a seizure scene could prompt passing. Aside from these, the reactions of against epileptic drug, repeating scenes of loss of awareness and engine control, and the general misinterpretation about the turmoil make clinical and mental boundaries.

2.2 Seizures and Types of Seizures

A seizure can be portrayed as a mix of inadvertent changes in behavior, development, sensation and cognizance as a consequence of irregular cerebrum movement. Seizures can be epileptic seizures or non-epileptic seizures.

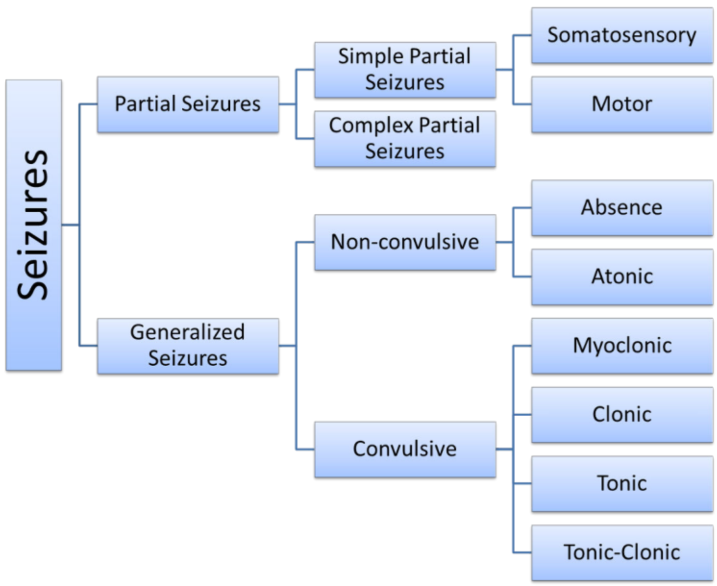


Figure 2.1 – types of seizures

Epileptic seizures happen as an after effect of an irregular cerebrum movement described by synchronized unusual and intemperate electrical action. Then again, non-epileptic seizures happen because of an outer unsettling influence to the focal sensory system, for example, liquor withdrawal, drug misuse, intense sickness, lack of sleep or with regards to mental injury.

Distinctive sorts of seizures are recorded in Figure 2.1. Diverse medicines are required for every sort of seizure and subsequently the capacity to recognize among them is critical. Two noteworthy sorts of seizures are fractional seizures and generalized seizures. While halfway seizures are restricted to a part of the cerebrum, summed up seizures include the whole mind.

2.2.1Partial Seizures

In a halfway seizure, epileptic movement is contained in one a player in the mind. Incomplete seizures that don't influence awareness are named basic halfway seizures, while those that do are delegated complex fractional seizures. A straightforward fractional seizure that begins in the somatosensory zone of the cerebrum is known as a basic halfway tactile seizure, while one that starts from the engine cortex is known as a basic incomplete engine seizure.

2.2.2 Generalized Seizures

In a summed up seizure, epileptic action includes the whole cerebrum from the onset. Summed up seizures which lead to unpredictable solid developments is delegated summed up convulsive seizures, while those that don't, are named summed up non-convulsive seizures. Contingent upon the condition of cognizance after the seizure, we can facilitate arrange convulsive seizures into the myoclonic, clonic, tonic, and tonic-clonic sorts.

Non-definitive seizures that outcome in the loss of cognizance, eye flickering, gazing, and other minor facial developments are called nonattendance seizures. Summed up non-indisputable seizures that don't prompt lost awareness are called atonic seizures.

2.3 What Is Treatment of Epilepsy

Epilepsy influences people with variable degrees of seriousness. Between 70-80% of epilepsy patients experience the ill effects of seizures whose seriousness and recurrence can be constrained with the utilization of antiepileptic medications, each of which basically restricts the limit of neurons to flame at inordinate rates. The right arrangement of these patients' seizures is essential since various seizure sorts require particular medication regiments. Truth be told, the utilization of the wrong antiepileptic medication may exacerbate certain sorts of seizures. The rest of the 20-30% of epilepsy patients experiences the ill effects of seizures that are stubborn to medicine. These patients look for option treatment choices that incorporate surgery, vague nerve incitement, and ketogenic diets.

2.4 Types Of EEG(Electrocorticography (ECoG)/Intracranial Electroencephalography (iEEG)):-

The electro-physiological checking which utilizes cathodes straightforwardly embedded on the uncovered area of the cerebrum to record electrical movement from the cerebral cortex is called Electrocorticography (ECoG), or Intracranial Electroencephalography (iEEG). Customary Electroencephalogram (EEG) on the other hand, screens this action from outside the skull. ECoG can be per-shaped in two routes, (1) in the working room amid surgery or (2) outside of surgery. ECoG is an intrusive methodology as it requires a surgical entry point into the skull. ECoG/iEEG has a reasonable favorable position over neuro imaging strategies due to its high spatial and transient determination. Further, the pollutions because of muscle development, eye squints which consistently hinder the nature of scalp EEG is insignificant in ECoG/iEEG. Be that as it may, the average qualities of EEG and ECoG/iEEG are practically identical, i.e. the run of the mill constituents of an EEG recording can likewise be seen in ECoG/iEEG.

2.5.NORMAL EEG Brain Rhythms:-

An average EEG recording contains the accompanying distinctive rhythms.

•Alpha cadence - EEG action with recurrence between 8-13 Hz that is unmistakable in the occipital locales of typical, loose grown-ups whose eyes are shut.

•Beta cadence - EEG action with recurrence surpassing 13 Hz that is most unmistakably seen in the frontal and focal locales in grown-ups, yet may likewise be summed up.

•Theta musicality - EEG movement with recurrence between 4-7 Hz; this action is strange in conscious grown-ups, yet usually saw in rest and kids underneath the age of 13 years.

•Delta musicality - The delta beat displays a recurrence beneath 3 Hz and amplitudes that surpass those of every other cadence; it is most unmistakable frontally in grown-ups and posteriorly in kids in the third and fourth phases of rest.

•Mu beat - The mu musicality alludes to EEG action with recurrence between 7-11 Hz that is most noticeably saw in the focal area; mu action is stifled by development (clench hand grasping), envisioned movement, or material incitement; conversely, it is improved by stationary nature and uplifted consideration

•Lambda waves - Transient sharp waves going on for a span of approximately 0.25 seconds that happen in the occipital district at whatever point a grown-up sweeps a visual field with level eye development

•Sleep-axles, K-buildings, and vertex Waves - These are interesting waveforms watched just amid the distinctive phases of rest

2.6 Ab-ordinary EEG Brain Rhythms

Strange EEG movement is any action that is common in the EEG of gatherings of individuals with neurological or other ailment dissensions, and missing from that of typical people. Irregular EEG might be a strange waveform and the nonappearance or deviation of typical EEG from very much recorded cutoff points on recurrence, sufficiency, morphology, confinement, and reactivity. The accompanying can be considered as the constituents of a strange EEG.

•Spike and sharp waves - Spike waves are drifters with pointed crests displaying spans between 20-70 milliseconds; sharp waves are similar to spike waves, however show longer lengths ordinarily between 70-200 milliseconds

•Periodic releases - Periodic releases allude to time-constrained blasts that are rehashed at a specific rate; blasts may show an assortment of lengths, frequencies, amplitudes, morphologies, and confinements

•Rhythmic hyper synchrony - Rhythmic hyper synchrony alludes to rhythmic movement rising up out of a tranquil foundation and showing un-normal recurrence, abundancy, morphology and restriction of any degree; musical action may either be constant or irregular

•Electro cerebral dormancy - Electro cerebral inertia alludes to a vari-capable length period not brought about by instrumental or physiological antiques that show amazing weakening of the EEG in respect to a patient-particular benchmark.

CHAPTER 3: GENRAL APPROACH FOR EPILEPTIC SEIZURE FORECASTING

With the goal of dissecting the iEEG information in a successive and organized design, we built up a pipeline comprising of various functionalities in each of its stages. The pipeline comprises of four confined segments: feature extraction, dimensionality decrease, machine learning classifier and forecasting and evaluation. The four segments are separately disengaged from each other so that diverse methods can be utilized on every segment without adjusting alternate segments. Figure 3.1 demonstrates the diverse components in the pipeline.

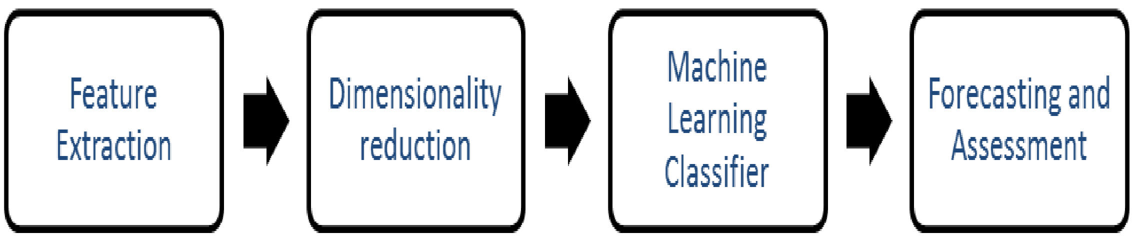


Figure 3.1: diverse components in the pipeline

All these blocks are described in subsequent chapters in great details.

CHAPTER 4 DATASETS

To test my methodology I used CHB_MIT-Scalp database[8].This SCALP EEG data, collected at the Children's Hospital boston, consists of graph recordings from medical specialty subjects with intractable seizures. Subjects were monitored for up to many days following withdrawal of anti-seizure medication so as to characterize their seizures and assess their movement for surgical intervention.

Recordings, sorted into twenty three cases, were collected from twenty two subjects (5 males, ages 3–22; and seventeen females, ages 1.5–19). (Case chb21 was obtained one.5 years when case chb01, from constant feminine subject.) The file SUBJECT-INFO contains the gender and age of every subject. (Case chb24 was side to the current assortment in Dec 2010, and isn't presently enclosed in SUBJECT-INFO.)

Each case (chb01, chb02, etc.) contains between nine and forty two continuous .edf files from one subject. Hardware limitations resulted in gaps between consecutively-numbered .edf files, throughout that the signals weren't recorded; in most cases, the gaps are ten seconds or less, however sometimes there are for much longer gaps. so as to safeguard the privacy of the topics, all protected health info (PHI) within the original .edf files has been replaced with surrogate info within the files provided here. Dates within the original .edf files are replaced by surrogate dates, however the time relationships between the individual files happiness to every case are preserved. In most cases, the .edf files contain precisely one hour of digitized graph signals, though those happiness to case chb10 are 2 hours long, and people happiness to cases chb04, chb06, chb07, chb09, and chb23 are four hours long; sometimes, files during which seizures are recorded are shorter.

All signals were sampled at 256 samples per second with 16-bit resolution. Most files contain twenty three graph signals (24 or twenty six during a few cases). The International 10-20 system of graph conductor positions and language was used for these recordings. during a few records, different signals are recorded, equivalent to AN EKG signal within the last thirty six files happiness to case chb04 and a cranial nerve stimulant (VNS) signal within the last eighteen files happiness to case chb09. In some cases, up to five “dummy” signals (named "-") were interspersed among the graph signals to get AN easy-to-read show format; these dummy signals is neglected.

The file RECORDS contains a listing of all 664 .edf files enclosed during this assortment, and therefore the file RECORDS-WITH-SEIZURES lists the 129 of these files that contain one or a lot of seizures. In all, these records embody 198 seizures (182 within the original set of twenty three cases); the start ([) and finish (]) of every seizure is annotated within the .seizure annotation files that accompany every of the files listed in RECORDS-WITH-SEIZURES. additionally, the files named chbnn-summary.txt contain info concerning the icon used for every recording, and therefore the time period in seconds from the start of every .edf file to the start and finish of every seizure contained in it.

CHAPTER5: VARIOUS FEATURE EXTRACTION METHODS FOR EPILEPTIC EEG DATA

5.1. Feature Extraction

From numerous studies we all know that characterizing the changes within the raw iEEG signals alone isn't enough to spot the pre-ictal signatures. Therefore, a change of the raw iEEG into some options is important. Researchers have used an outsized variety of features to explain the changes within the encephalogram recordings. However, with the increasing usage of machine learning techniques for deciding, the hunt for distinguishing complicated feature descriptors became less significant. Machine learning techniques are capable of expressing the complicated relationships between simple features which may be extracted very simply.

5.2Classification of feature extraction methods-

On the basis of various categories of features are used, methods of feature extraction may lie in various domains. These domains are as follows[9]-

Time Domain

Frequency domain

Wavelet domain

PCA and IDA domain

EMD domain

Singular value decomposition(SVD)

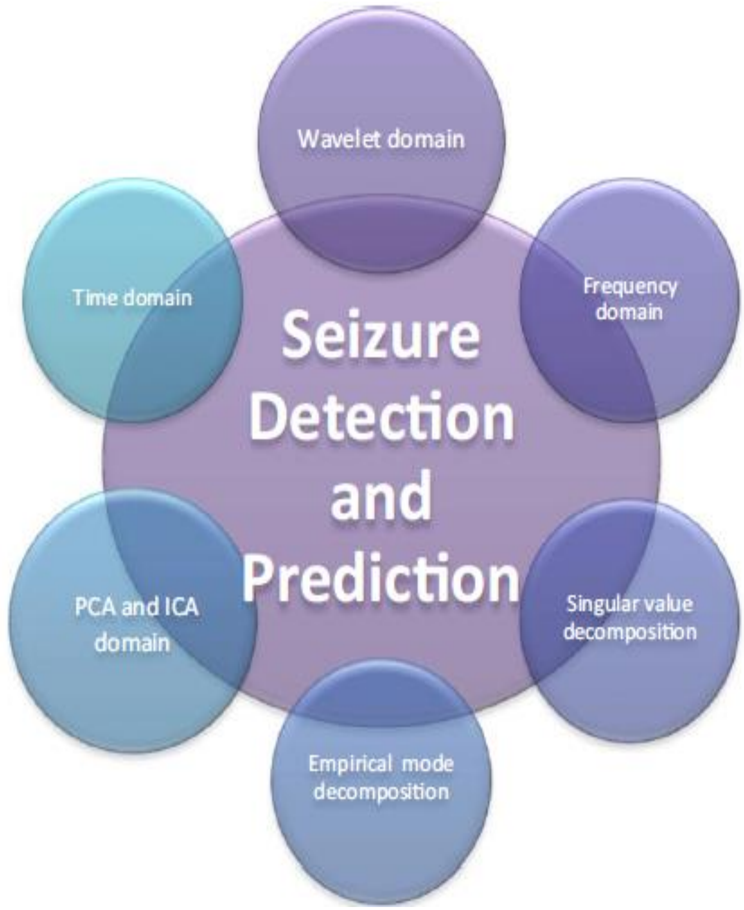


Figure 5.1: classification of feature extraction methods

5.2.1 Time-domain methods

Various parameters can be used as features-set of data .if these features have incorporated over time information ,then they are characterised in time domain.

Various features lie under time domain are as follows-

1. Mean
2. Variance,
3. zero-crossing rate
4. Entropy,

5. auto-correlation with template signals-

NOTE:-For estimation of auto-correlation, a dynamic time warping (DTW) Approach is used for best alignment between the signal segments (Which is to be tested) and the template signal.

5.2.2 Frequency-domain methods

In many research Frequency-domain technique has been used for EEG Seizure features detection. Fourier transform magnitude and phase is used for frequency domains features. For discrete domain DFT approach is used. As another approach for frequency domain features estimation is entropy estimation. since entropy is calculated in various bands of frequency and then set of these entropies is called as feature-set, Entropy has frequency domain information

Different entropy features are as follows-

1. phase entropy
2. approximate entropy
3. sample entropy

The phase entropies are calculable from the higher-order spectra of electroencephalogram signal epochs as discriminating features for ictal, pre-ictal (epochs before seizure occurring period), and inter-ictal activities. The approximate and sample entropies are logarithmic metrics that confirm the closeness and matching between the incoming electroencephalogram signal pattern and also the recorded emulates.

5.2.3 Wavelet-domain methods:-

Wavelets are widely utilized in the sector of electroencephalogram signal analysis, particularly for seizure detection and prediction. The wavelet transform in itself will be considered some form of sub-band decomposition, however with down sampling. The wave remodel will be enforced on analog yet as digital signals. we are a lot of interested in the DWT. The DWT will be implemented with low-pass (LP) and high-pass (HP) filtering additionally to a decimation method, and it should be invertible as shown in Figure 6a . The DWT will be enforced with a single level or multi levels as shown in Figure

6b, c. For the multi-level wave decomposition, more decomposition up to the desired level are performed on the low-pass branch solely. Another implementation of wavelet analysis is that the wavelet packet transform, which performs more decomposition on the low-pass and high-pass branches.

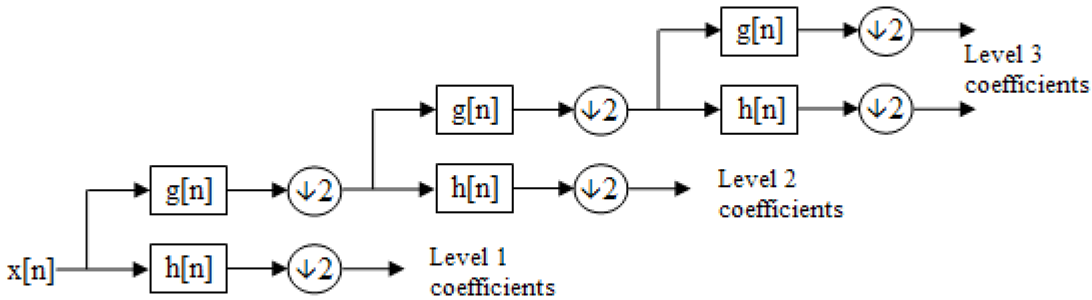


Figure 5.2 multilevel wavelet decomposition

The wavelet decomposition and reconstruction filters H0, H1, G0, and G1 should satisfy the proper reconstruction condition. These filters will be obtained by solving the subsequent equation .

$$
\begin{aligned}
Y(z) &= \frac{1}{2}\{X_0(z) + X_0(-z)\}G_0(z) + \frac{1}{2}\{X_1(z) + X_1(-z)\}G_1(z) \\
&= \frac{1}{2}X(z)\{H_0(z)G_0(z) + H_1(z)G_1(z)\} \\
&\quad + \frac{1}{2}X(-z)\{H_0(-z)G_0(z) + H_1(-z)G_1(z)\}
\end{aligned}
\tag{5.1}
$$

5.2.4 Empirical mode decomposition:

The EMD is a signal decomposition, which transforms a signal into a group of intrinsic mode functions (IMFs).For EEG seizure detection, these IMFs show different behaviour with normal and abnormal activities in the signals.EMD could be a methodology of breaking down a signal while not leaving the time domain. It may be compared to alternative analysis strategies like Fourier Transforms and wave decomposition. The method is helpful for analyzing natural signals that are most frequently non-linear and non-stationary. This elements from the assumptions of the strategies we've so far learned (namely that the systems in question be LTI, a minimum of in approximation). EMD filters out functions that kind a whole and nearly orthogonal basis for the first signal. Completeness is predicated on the tactic of the EMD; the approach it's decomposed implies completeness. The functions, referred to as Intrinsic Mode Functions (IMFs), are so adequate to explain the signal, even if they're not essentially orthogonal. The explanations are represented in Huang et al., printed within the honorary society Proceedings on science, Physical, and Engineering Sciences: "...the real that means here applies solely regionally. for a few special information, the near elements may definitely have sections of knowledge carrying identical frequency at totally different time durations. However regionally any two elements ought to be orthogonal for all sensible purposes".The fact that the functions into that a symptom is decomposed are all in the time-domain and of identical length because the original signal permits for variable frequency in time to be preserved. Getting IMFs from global signals is very important as a result of natural processes typically has multiple causes, and every of those causes could happen at specific time intervals. This kind of knowledge is clear in associate degree EMD analysis, however quite hidden within the Fourier domain or in wave coefficients. Process• The EMD can break down a symptom into its part IMFs.

An imf could be a function that:

1.Has only 1 extreme between zero crossings, and

2.Includes a mean of zero.

In order to explain the method, we tend to borrow from our poster the subsequent section:

The sifting method:-

sifting method is what EMD uses to decomposes the signal into IMFs.

The sifting method is as follows:

For a signal X(t), let m1 be the mean of its higher and lower envelopes as determined from a cubic-spline interpolation of native maxima and minima. The locality is set by associate degree impulsive parameter; the calculation time and therefore the effectiveness of the EMD depends greatly on such a parameter.

-The 1st part h1 is computed: $$h1=X(t)-m1 \quad (5.2)$$

-In the second sifting method, h1 is treated because the information, and m11 is that the mean of h1's higher and lower envelopes:

$$h11=h1-m11 \quad (5.3)$$

-This sifting procedure is continual k times, till h1k is associate degree imf, that is: $$h1(k-1)-m1k=h1k \quad (5.4)$$

Then it's selected as c1=h1k, the first imf element from the info, that contains the shortest period element of the signal. we tend to separate it from the remainder of the data: X(t)-c1 = r1 The procedure is continual on rj: r1-c2 = r2,....,rn-1 - cn = rn

• The result's a collection of functions; the amount of functions within the set depends on the first signal.

CHAPTER 6: VARIOUS METHODS FOR FETURES REDUCTION AND DIMENSION REDUCTION

One of the foremost vital tasks in any pattern recognition system is to seek out an informative, yet small, set of options with increased discriminatory power1. Techniques that may introduce low-dimensional feature illustration with enhance discriminatory power are of dominant importance, due to the thus known as curse of spatial property. Feature extraction consists of finding a group of measurements or a block of data with the target of describing during a clear way the data or an occurrence present during a signal. These measurements or options are the elemental basis for detection, classification or regression tasks in medicine signal process and is one in all the key steps within the data analysis method. Features represent a brand new type of expressing the info, and may be binary, categorical or continuous: they represent attributes or direct measurements of the signal. Let's say, features could also be the age, health standing of patient, case history, electrodes.

since analyzing all the features within the predictive pipeline appeared unfeasible because of the size of the dataset, applying spatial property reduction on these features was thought of. To account for the variations within the spatial property reduction techniques, we tend to used 2 spatial property reduction techniques, one unsupervised and one supervised technique. Principal element associate degree lysis (PCA) is an unsupervised spatial property reduction technique used extensively in several applications . On the opposite hand, partial least squares (PLS) regression may be a supervised spatial property reduction technique, that finds those constituents within the options that mostly contribute to the discriminability of the various categories within the dataset .PLSR and PCA are each ways to model a response variable once there are an oversized variety of predictor variables, and those predictors are extremely cor- related also collinear. each ways construct new predictor variables, referred to as parts, as linear combos of the first predictor variables, however they construct those parts in numerous ways that. PCA creates parts to clarify the determined variability within the predictor variables, without considering the response variable in any respect. On the opposite hand, PLSR will take the response variable under consideration, and so typically results in models that are able to match the response variable with fewer components. if or not that ultimately interprets into a lot of ungenerous model, in terms of its sensible use, depends on the context. Position or graph signal descriptors (amplitude, voltage, phase, frequency, etc). The aim of extracting features is to spot “patterns” of brain activity: features are often used as input to a classifier2. The performance of a pattern recognition system depends on each the options and therefore the classification algorithmic program utilized. Additional formally, feature extraction assumes there are N samples and D features, for a N X D information matrix. It’s conjointly attainable to get a feature vector at the sample n type the feature matrix, that is, x may be a unidimensional vector x=[x1, x2,…, xD] known as pattern vector. Our representation of the globe is made by process massive numbers of sensory inputs – as well as, let’s says, the pixel intensities of pictures, the ability spectra of sound, and therefore the joint angles of articulated bodies. Whereas complicated stimuli of this way are often described by points during a high-dimensional vector area, they usually have a way additional compact description. Coherent structure within the world ends up in sturdy correlations between inputs, generating observations that lie on or on the brink of a sleek low-dimensional manifold. to match and classify such observations- in impact, to reason concerning the world- rely crucially on modelling the nonlinear pure mathematics of those low-dimensional manifolds. Suppose the info lie on or close to a sleek non-liner manifold of lower spatial property d<<D. To a decent approximation then, there exists a linear mapping- consisting of a translation, rotation, and rescaling-that maps the high-dimensional coordinates of every neighborhood to global internal coordinates on the manifold3. By design, the reconstruction weights Wij reflect intrinsic geometric properties of the info that are invariant to precisely such transformations. Thus expect their characterization of native geometry within the original information area is anticipated to be equally valid for native patches on the manifold. Above all, a similar weights Wij that reconstruct the ith datum in D dimensions ought to conjointly reconstruct its embedded manifold coordinates in d dimensions[10].

Suitable spatial property Reduction Techniques :-

6.1. Principal component Analysis:-

It is a second order technique and it's the most effective within the mean-square error sense. in numerous domains, it's known as Singular worth Decomposition(SVD), empirical orthogonal function (EOF), Hotellingre model and therefore the Karhunen-Loeve remodel. By finding many orthogonal and linear mixtures of the first variables, PCA finds to cut back the info dimension. Within the field of engineering arithmetic and statistics, PCA has been with success applied during a sort of domains. for top dimensional information, the computation of the chemist vectors can be unworkable since the variance matrix is proportional to the info point's spatial property.

6.2. Kernel PCA:-

In high dimensional information, PCA is employed for the look and modeling of the linear variabilities. However the characteristic nature of the high dimensional information set is that it's a non-linear nature. In such cases PCA cannot verify the variability of knowledge accurately. to deal with the matter of non-linear spatial property reduction kernel PCA are often used. With the usage of kernels, some nonlinear mapping is completed in order that the principal elements are often computed expeditiously in high – dimensional feature areas. Kernel PCA finds the principal elements wherever the newest structure is low dimensional and is simpler to find.

6.3. Sammon Map:-

It is a typical example of an ideal non-linear feature extraction technique. it's wide used for the 2-D visualization of knowledge that has quite high dimension. so as to attenuate the error operate, Sammon's algorithmic program uses gradient descent technique.

6.4. Multidimensional Scaling (MDS):-

For the straightforward analysis of proximity information on a moderate set of stimuli, multidimensional scaling are often used. it's wont to specific or reveal the hidden structure that underlies within the information. Here the points are largely placed in a very low dimensional area and it's wide thought of as an exploratory information analysis technique. The visualization of the structures within the information is often simply learnt by the researchers if the illustration of the patterns of proximity is completed in 2 or 3 dimensions. Multidimensional scaling is generally classified into 2 varieties particularly MDS analysis of interval of ratio information and MDS analysis of ordinal information.

6.5. Locally Linear Embedding:-

One of the used approaches for the issues addressing to the non-linear spatial property reduction is locally Liner Embedding. It’s done by the computation of low-dimensional information that preserves the neighborhood embedding of high dimensional data. Our aim here is to map the high-dimensionality regional information point’s toone coordinate system that is allotted to a global position.

6.6. ISOMAP:-

It is a non-linear spatial property reduction technique supported the Multi-Dimensional technique. The classical MDS are often described in a generalized non-linear manner which is named as an Isomap. Within the input area, MDS isn't performed; however it's performed at the nonlinear information manifold’s geodesic area. Isomap doesn't apply the classic MDS to the line distances. It invariably applies to the geodesic distances so as to seek out a low-dimensional mapping that preserves those combine wise distances.

6.7. Factor Analysis:-

It is a applied mathematics technique that is closely involving the Principal component Analysis and spatial property reduction. It’s essentially a generative model. From a group of latent, the discovered information has been created and through the analysis of mathematical equations the unobserved variables has been created. The factors invariably follow a variable Gaussian distribution and it's assume dto be additional uncorrelated to noise.

6.8. Independent component Analysis (ICA):-

ICA may be a variant of principal component analysis (PCA) where the elements are assumed to be reciprocally statistically independent rather than ,simply uncorrelated. It’s a procedure technique for separating a variable signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian supply signals. Independent component Analysis describes a model for variable information describing massive info of samples. The variables within the model are assumed non-Gaussian and reciprocally independent and that they are known as the independent elements of the discovered information.

6.9. Generalized SVD:-

It is a really versatile tool as a result of it are often particularized to correspondence analysis underneath the acceptable decisions of the constrain matrices. It can even be particularized to discriminant analysis and canonical correlation analysis. Generalized SVD performs an analogous decomposition and it modifies or relaxes the conditions of orthogonality according the 2 constrain matrices.

6.10. Maximum Variance unfolding:-

It is additional like Isomap in a respect when upon the information, a neighborhood graph is outlined and therefore the pairwise distances within the resulting graph is preserved. Here the Euclidian distances are maximized between the info points underneath the condition that the corresponding distances within the neighborhood graph is left unchanged. Then as results of it, the optimization drawback comes into existence sand so it is often solved expeditiously using semi definite programming.

6.11. Diffusion Maps:-

It is originating from the sector of system dynamics. On the graph of selected information, a markov random walk is outlined upon it. By playacting the stochastic process for variety of times the info points proximity is obtained. Using this sort of means that solely diffusion maps are outlined. If the info is described during a low dimension then the pair wise diffusion distances are preserved there itself. Unlike the PCA and fa, it's a whole linear technique and it will simply incorporate over second order info. Projection Pursuit is employed largely for the analysis of non-Gaussian datasets. Over second order strategies it’s additional computationally intensive. Any side of non-Gaussianity is typically measured through projection pursuits solely. A projection index invariably defines the interest of a direction and appears for the directions that optimize that index.

6.12. Multi-unit Objective functions:-

To specify the target functions there are many alternative ways in which. In spite of their completely different formulations all of them are closely connected and that they even become equivalent in some conditions. Underneath bound conditions the non-linear PCA technique is applicable wherever the mutual info technique is most similar to maximum likelihood principle. a number of the foremost important Multi-unit Objective functions are maximum likelihood and network entropy, Mutual info and Kullback-Leibler divergence, Non-linear cross-correlations, Non-linear PCA and Higher-order cumulant tensors.

6.13. Multi-layer autoencoders:-

They are feed-forward neural networks with hidden layers sometimes having an odd variety. The network is trained well for the minimization method of the mean square error between the input and therefore the output. If the activation functions are linear and it's employed in the neural network, then the performance of autoencoder are going to be additional similar to PCA. to find out a non-linear mapping between the high dimensional and low dimensional information representation, the machine encoders use sigmoid activation functions. They have been used with success for the appliance of issues involving the missing information and HIV analysis.

6.14. Laplacian Eigen maps:-

It is additional like LLE technique wherever the preservation of the native properties of the manifold is given a previous importance and so the Eigen maps will simply realize the low dimensional information illustration.Between near neighbors.the native properties are supported the pair wise distances. They invariably compute a low dimensional illustration of selected information in which the distances between a data point and its nearest neighbors are reduced.

CHAPTER 7: METHODS FOR EPILEPTIC DATA CLASSIFICATION

Here we need a classification of EEG data in two classes; one class is Epileptic and another one is non-epileptic. Our main aim is to distinguish between normal EEG data and epileptic pre-ictal EEG data, for this purpose we can use following methods-

Support Vector Machine
Gaussian Mixture Model (GMM)
Artificial neural network
K-Nearest Neighbor

7.1 Support Vector Machine:-

In the figure 3, there are several linear classifiers (hyper planes) that separate the info. But only 1 of those achieves most separation. the rationale we'd like it's as a result of if we have a tendency to use a hyper plane to classify, it'd find yourself nearer to at least one set of datasets compared to others and that we don't wish this to happen and so we see that the thought of most margin classifier or hyper plane as a plain answer. ensuing illustration provides the utmost margin classifier example that provides an answer to the on top of mentioned drawback [8].

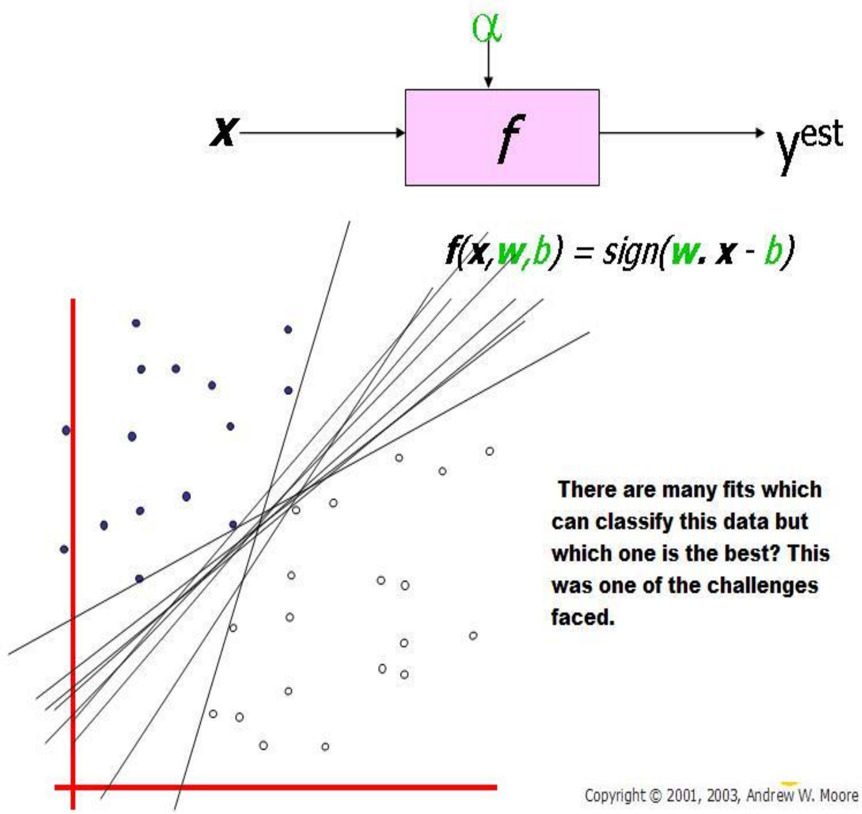


Figure 7.1: Here we have a tendency to see that there are several hyper planes which may be slot in to classify the info however that one is that the best is that the right or correct answer. the requirement for SVM arises. (Takenandrew W. Moore 2003)[11]

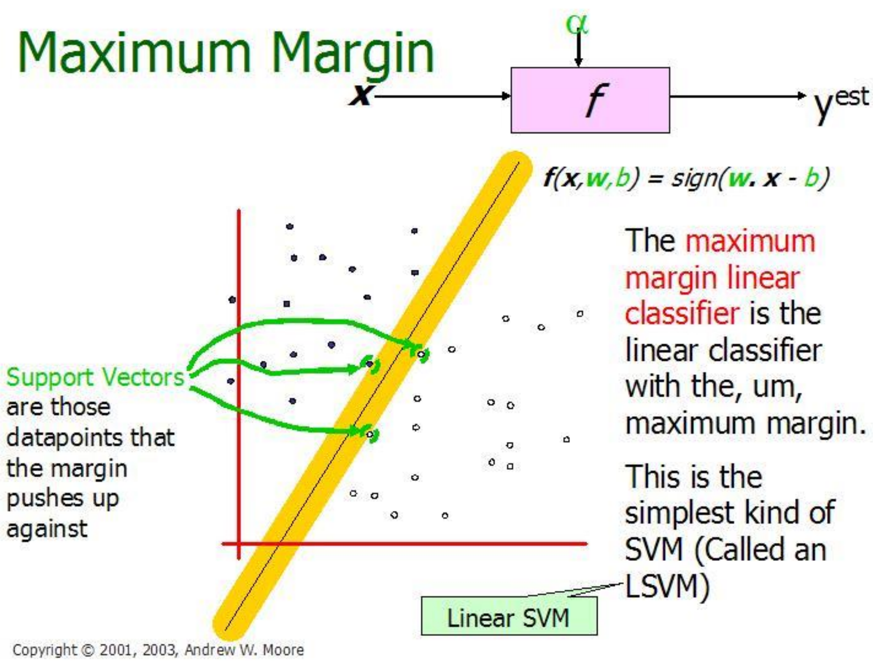


Figure 7.2: Illustration of Linear SVM. (Taken from Andrew W. Moore slides 2003) [12]. Note the legend is not described as they are sample plotting to make understand the concepts involved.

Expression for max margin is given as:

$$\text{margin} \equiv \arg\min_{\mathbf{x}\in D} d(\mathbf{x}) = \arg\min_{\mathbf{x}\in D} \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}} \qquad (7.1)$$

The on top of illustration is that the most linear classifier with the maximum range. During this context it's associate example of a straightforward linear SVM classifier. Another fascinating question is why most margin? There are some smart explanations that embrace higher empirical performance. One more reason is that even if we've created a little error within the location of the boundary this offers us least probability of inflicting misclassification. The opposite advantage would be avoiding native minima and higher classification. Currently we have a tendency to try and categorical the SVM mathematically and for this tutorial we have a tendency to try and present a linear SVM. The goals of SVM are separating the info with hyper plane and extend this to non-linear boundaries exploitation kernel trick . For scheming the SVM we have a tendency to see that the goal is to properly classify all the info. For mathematical calculations we've -

[a] If Y_i= +1 ; $wx_i + b \geq 1$
[b] If Y_i= -1; $wx_i + b \leq 1$
[c] For all i; $y_i \ (w_i + b) \geq 1$ (7.2)

In this equation x may be a vector point and w is weight and is additionally a vector. Thus to separate the info [a]should be larger than zero. Among all doable hyper planes, SVM selects the one wherever the gap of hyper plane is as giant as possible. If the training information is nice and each check vector is found in radius r from training vector. Currently if the chosen hyper plane is found at the farthest doable from the info [12]. This desired hyper plane that maximizes the margin conjointly bisects the lines between nearest points on convex hull of the two datasets. Therefore we've [a], [b] & [c].

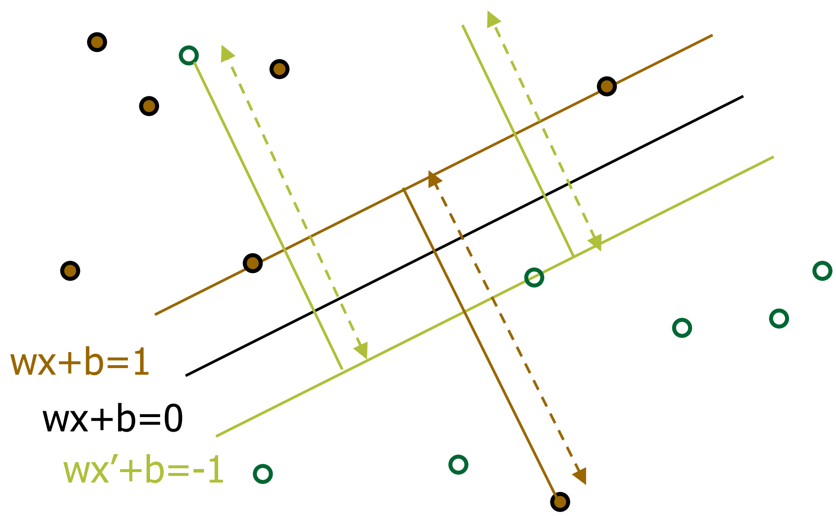


Figure 7.3: presentation of Hyper planes

Distance of nearest purpose on hyperplane to origin will be found by increasing the x as x is on the hyper plane. Equally for the opposite aspect points we've an identical state of affairs.
Therefore finding and subtracting the 2distances we have a tendency to get the summed distance from the separating hyperplane to nearest points. Most Margin = M = 2 / ||w||
Now increasing the margin is same as minimum . currently we've a quadratic improvement drawback and that we ought to solve for w and b. to resolve this we'd like to optimize the quadratic function with linear constraints. The answer involves constructing a dual drawback and wherever a Langlier's multiplier factor αi is associated. we'd like to seek out w and b such $\Phi(w) = \frac{1}{2}|w'||w|$ is minimized

And for all $\{(x_i, y_i)\}$: $y_i (w * x_i + b) \geq 1.$

$$(7.3)$$

Now solving: we get that $w = \Sigma\alpha_i * x_i$; $b = y_k - w * x_k$ for any x_k such that $\alpha k \neq 0$
Now the classifying function will have the following form: $f(x) = \Sigma\alpha_i y_i x_i * x + b$

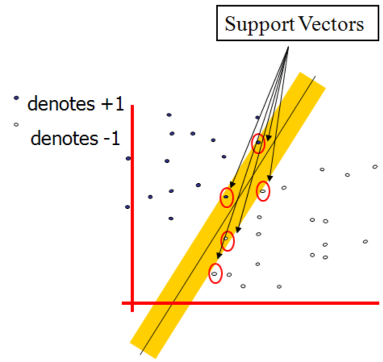


Figure 7.4: presentation of Support Vectors (Copyright © 2003, Andrew W. Moore)[2]

SVM illustration:-

In this we have a tendency to present the QP formulation for SVM classifications giant. Variety of ways for quick SVM training are planned .

SVM classification:

$$\min_{f,\xi_i} \|f\|_K^2 + C\sum_{i=1}^{l}\xi_i \quad y_i f(\mathbf{x}_i) \geq 1 - \xi_i, \text{ for all i } \xi_i \geq 0 \qquad (7.4)$$

SVM classification, Dual formulation:-

$$\min_{\alpha_i} \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad 0 \leq \alpha_i \leq C, \text{ for all i; } \sum_{i=1}^{l}\alpha_i y_i = 0 \qquad (7.5)$$

This is often a straightforward illustration solely. Variables ξi referred as slack variables and that they measure the error created at point (xi,yi). Training SVM becomes quite difficult once the amounts of training points are large.

7.2.Gaussian Mixture Model :-

We can think about building a Gaussian Mixture Model as a sort of bunch rule. using associate degree reiterative technique called Expectation Maximization, the method and results are almost like k-means bunch. The distinction is that the clusters are assumed to every have an independent normal distribution, every with their own mean and variance matrix.

Comparison To K-Means clustering

When activity k-means clustering, you assign points to clusters using the straight geometer distance. The geometer distance may be a poor metric, however, once the cluster contains vital variance. Within the below example,

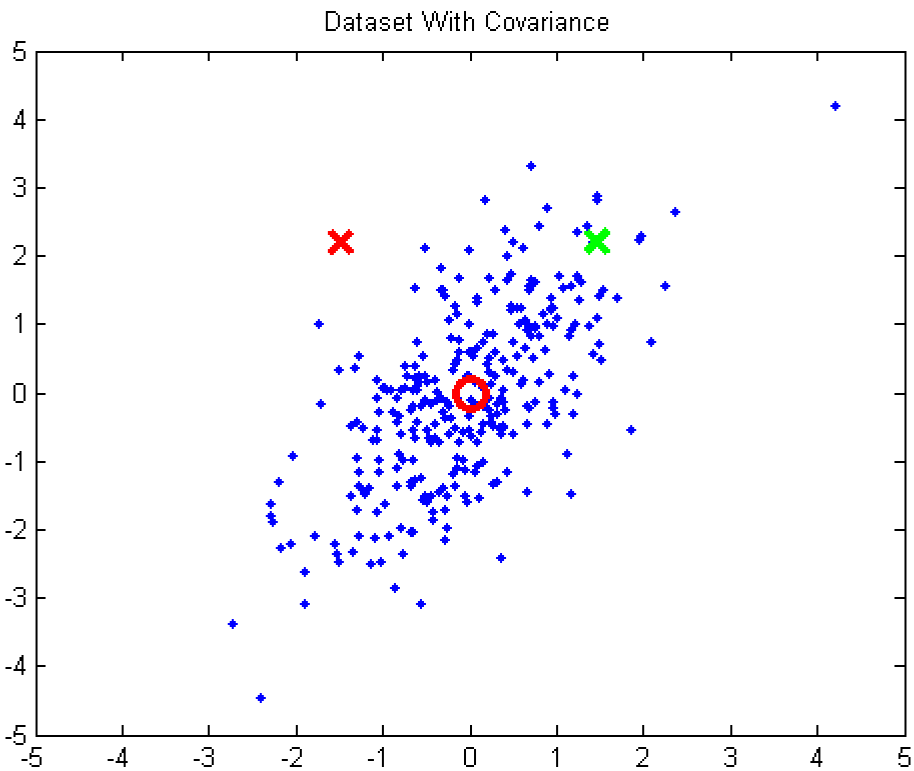


Figure 7.5 dataset with covarience

We've a bunch of points exhibiting some correlation. The red and inexperienced x's area unit equal from the cluster mean using the geometer distance, however we will see intuitively that the red X doesn't match the statistics of this cluster close to as well because the green X. If we were to have these points and normalize them to get rid of the variance (using a method referred to as whitening), the green X becomes a lot of nearer to the mean than the red X.

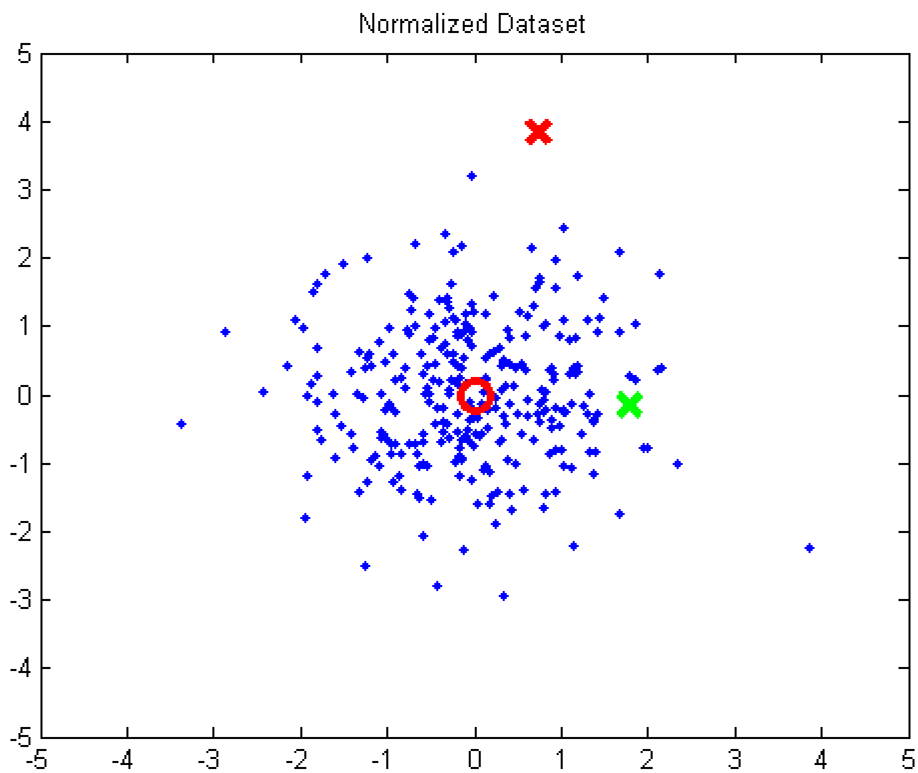


Figure 7.6 normalized dataset

The Gaussian Mixture Models approach can take cluster variance under consideration once forming the clusters. Another vital distinction with k-means is that standard k-means performs a tough assignment of knowledge points to clusters–each point is assigned to the nearest cluster. With Gaussian Mixture Models, what we'll end up is a assortment of independent Gaussian distributions, and then for every information, we'll have a likelihood that it belongs to every of those distributions / clusters.

Expectation Maximization:-

For GMMs, we'll realize the clusters employing a technique referred to as “Expectation Maximization”. this can bean reiterative technique that feels a great deal just like the reiterative approach employed in k-means clustering. In the “Expectation” step, we'll calculate the likelihood that every information belongs to every cluster (using our current calculable mean vectors and variance matrices). This looks analogous to the cluster assignment step in k-means. In the “Maximization” step, we’ll re-calculate the cluster suggests that and covariance supported by the possibilities calculated within the expectation step. This looks analogous to the cluster movement step in k-means. Initialization To kick-start the EM rule, we’ll at random choose data points to use because the initial suggests that, and we’ll set the variance matrix for every cluster to be equal to the variance of the complete training set. Also, we’ll offer every cluster equal “prior probability”. A cluster’s “prior probability” is simply the fraction of the dataset that belongs to every cluster. We’ll begin by assumptive the dataset is equally divided between the clusters. Expectation In the “Expectation” step, we tend to calculate the likelihood that every data point belongs to every cluster. We’ll would like the equation for the likelihood density function of a variable Gaussian.A variable Gaussian(“multivariate” simply suggests that multiple input variables) is a lot of advanced as a result of there's the likelihood for completely different variables to possess different variances, and even for there to be correlation between the variables. These properties area unit captured by the variance matrix.

$$g_j(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)}$$

Symbol	Meaning
$g_j(x)$	The PDF of the multivariate Gaussian for cluster j; the probability of this Gaussian producing the input x
j	Cluster number
x	The input vector (a column vector)
n	The input vector length
Σ_j	The n x n covariance matrix for cluster j
$\|\Sigma_j\|$	The determinant of the covariance matrix
Σ_j^{-1}	The inverse of the covariance matrix

((7.6)

The likelihood that example point i belongs to cluster j will be calculated using the following:

$$w_j^{(i)} = \frac{g_j(x)\phi_j}{\sum_{l=1}^{k} g_l(x)\phi_l}$$

Symbol	Meaning
$w_j^{(i)}$	The probability that example i belongs to cluster j
$g_j(x)$	The multivariate Gaussian for cluster j
ϕ_j	The "prior probability" of cluster j (the fraction of the dataset belonging to cluster j)
k	The number of clusters

(7.7)

We'll apply this equation to each example and each cluster, giving us a matrix with one row per example and one column per cluster.

Maximization:-

You can gain some helpful intuition concerning the maximization equations if you're aware of the equation for taking a weighted average. to search out the common value of a set of m values, wherever you have got a weight sw_defined for every of the values, you'll be able to use the subsequent equation:

$$\bar{y} = \frac{\sum_{i=1}^{m}(w_i y_i)}{\sum_{i=1}^{m} w_i} \quad (7.8)$$

With this in mind, the update rules for the maximization step are below-

$$\phi_j := \frac{1}{m}\sum_{i=1}^{m} w_j^{(i)},$$

$$\mu_j := \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}},$$

$$\Sigma_j := \frac{\sum_{i=1}^{m} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}}$$

..(7.9)

The equation for mean (mu) of cluster j is simply the common of all data points within the training set, with every example weighted by its likelihood of belonging to cluster j. Similarly, the equation for the variance matrix is that the same because the equation you'd use to estimate the variance of a dataset, except that the contribution of every example is once more weighted by the likelihood that it belongs to cluster j. The previous likelihood of cluster j, denoted as letter, is calculated because the average likelihood that a data point belongs to cluster j.

7.3.Artificial neural network:-

An artificial neural network may be a system supported by the operation of biological neural networks, in alternative words, is an emulation of biological neural system. Why would be necessary the implementation of artificial neural networks? Though computing recently is actually advanced,there are bound tasks that a program created for a typical microchip is unable to perform; nevertheless a software system implementation of a neural network is created with their advantages and drawbacks.

Advantages:

• A neural network will perform tasks that a linear program can't.
• When part of the neural network fails, it will continue with none drawback by their parallel nature.
• A neural network learns and ought not to be reprogrammed. • It is enforced in any application.
• It is enforced with none drawback.

Disadvantages:

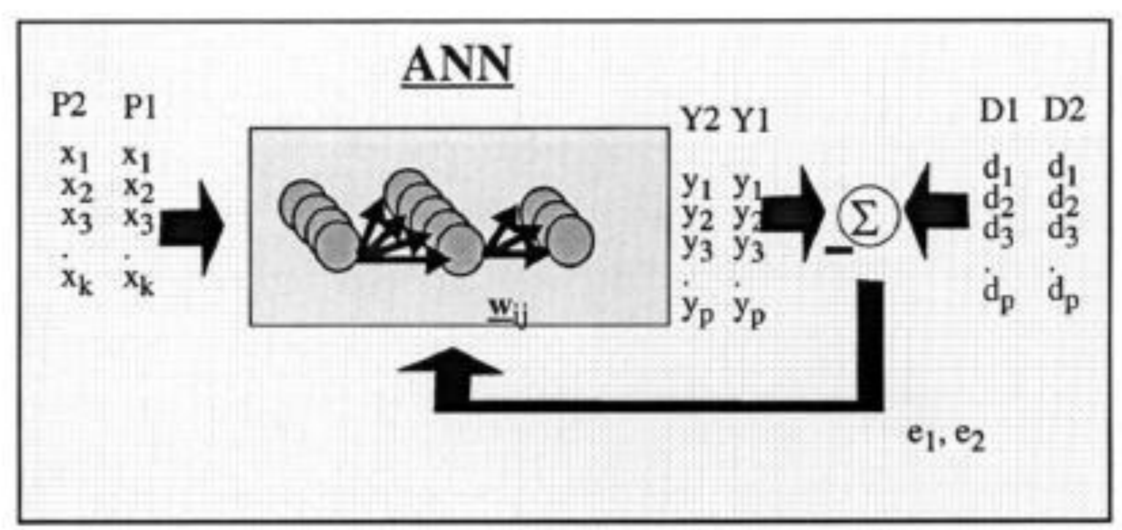
• The neural network desires training to work.
• The design of a neural network is totally different from the design of microprocessors so has to be emulated.
• Requires high time interval for giant neural networks.

Another side of the artificial neural networks is that there are totally different architectures, that consequently need differing kinds of algorithms, however despite to be AN apparently advanced system, a neural network is comparatively straightforward.

Artificial neural networks (ANN) are among the most recent signal-processing technologies within the engineer ‘stool case. the sphere is extremely knowledge domain, however our approach can limit the read to the engineering perspective. In engineering, neural networks serve two necessary functions: as pattern classifiers and as nonlinear adjective filters. we'll offer a quick summary of the idea, learning rules, and applications of the foremost necessary neural network models. Definitions and elegance of Computation a man-made Neural Network is AN adjective, most frequently scheme that learns to perform a function (an input/output map) from information. Adaptive implies that the system parameters are modified throughout operation, ordinarily referred to as the training part. Once the training parts the artificial Neural Network parameters are fastened and therefore the system is deployed to resolve the matter at hand (the testing part). The artificial Neural Network is constructed with a scientific stepwise procedure to optimize a performance criterion or to follow some implicit internal constraint, that is usually noted because the learning rule. The input/output training information are basic in neural network technology, as a result of they convey the required data to "discover" the optimum operational purpose. The nonlinear nature of the neural network process components (PEs) provides the system with a lot of flexibility to realize much any desired input/output map, i.e., some Artificial

Neural Networks are universal mappers . There’s a method in neural computation that's value describing.

An input is given to the neural network and a corresponding desired or target response set at the output (when this can be the case the training is termed supervised). a mistake consists from the distinction between the specified response and therefore the system output. This error data is fed back to the system and adjusts the system parameters in a very systematic fashion (the learning rule). the method is recurrent till the performance is appropriate. it's clear from this description that the performance hinges heavily on the info. If one doesn't have information that cowl a major portion of the operational conditions or if they're noisy , then neural network technology is perhaps not the correct answer. On the opposite hand, if there's lots of information and therefore the downside is poorly understood to derive an approximate model, then neural network technology may be a good selection.



The style of neural computation.

Figure 7.7 style of neural computation

The Mathematical Model:-

An input is given to the neural network and a corresponding desired or target response set at the output (when this can be the case the training is termed supervised). a mistake consists from the distinction between the specified response and therefore the system output. This error data is fed back to the system and adjusts the system parameters in a very systematic fashion (the learning rule). The method is recurrent till the performance is appropriate. It’s clear from this description that the performance hinges heavily on the info. If one doesn't have information that cowls a major portion of the operational conditions or if they're noisy, then neural network technology is perhaps not the correct answer. On the opposite hand, if there's lots of information and therefore the downside is poorly understood to derive an approximate model, then neural network technology may be a good selection. This process ought to be contrasted with the standard engineering style, manufactured from exhaustive scheme specifications and communication protocols. In artificial neural networks, the designer chooses the constellation, the performance operate, the training rule, and therefore the criterion to prevent the coaching part, however the system mechanically adjusts the parameters. So, it's troublesome to bring a priori data into the planning, and once the system doesn't work properly it's conjointly onerous to incrementally refine the answer. However ANN-based solutions are very economical in terms of development time and resources, and in several troublesome issues artificial neural networks offer performance that's troublesome to match with alternative technologies. Denker ten years ago state that "artificial neural networks are the competition thanks to implement a solution" actuated by the simplicity of their style and since of their catholicity, solely umbrageous by the standard style obtained by learning the physics of the matter. At present, artificial neural networks are rising because the technology of alternative for several applications, admire pattern recognition, prediction, system identification, and management.

When making a useful model of the biological neuron, there are 3 basic parts of importance. First, the synapses of the neuron are modelled as weights. The strength of the affiliation between AN input and a nerve cell is noted byte value of the weight Negative weight values reflect restrictive connections, whereas positive values designate

excitatory connections[Haykin]. Consecutive 2 parts model the particular activity at intervals the nerve cell cell.AN adder sums up all the inputs changed by their several weights. This activity is noted as linear combination. Finally, AN activation operate controls the amplitude of the output of the nerve cell. an appropriate vary of output is sometimes between zero and one, or -1 and one. Mathematically, this method is represented within the figure-

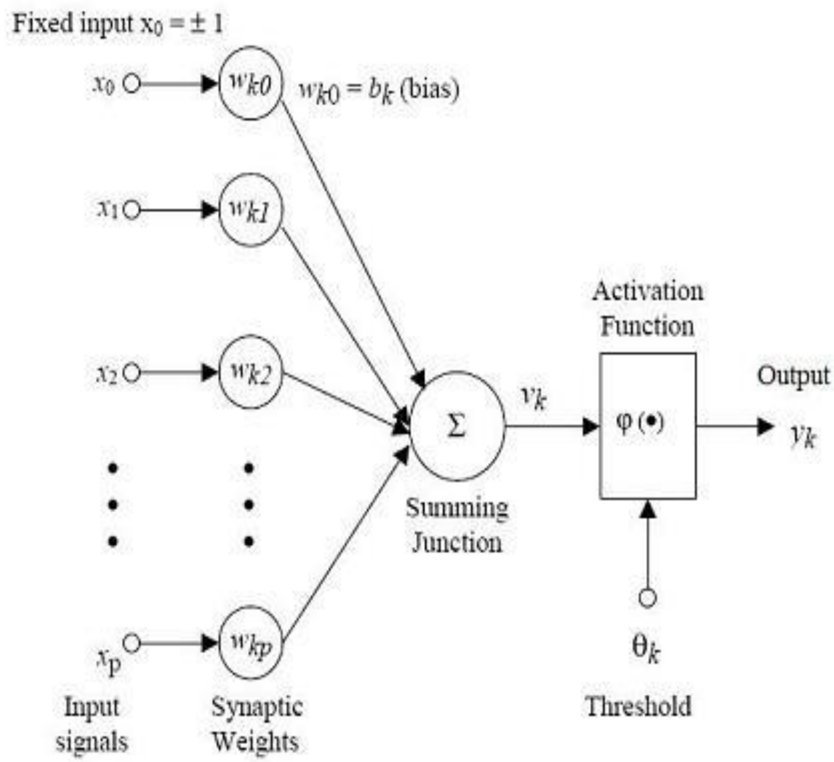


Figure 7.8 ANN MODEL

by this model the interval activity of the nerve cell is shown to be:

$$v_k = \sum_{j=1}^{p} w_{kj} x_j \tag{7.10}$$

The output of the neuron is yk, so there will be the outcome of some activation function on the value of vk.

Activation functions:-

As mentioned previously, the activation operate acts as a squashing function, specified the output of a nerve cell in a neural network is between sure values (usually zero and one, or -1 and 1). In general, there are 3 varieties of activation functions, denoted by Φ(.) . First, there's the threshold function that takes on a value of zero if the summed input is a smaller amount than a definite threshold value (v), and therefore the value one if the summed input is bigger than or up to the threshold value.

$$\varphi(v) = \begin{Bmatrix} 1\ if\ v \geq 0 \\ 0\ if\ v < 0 \end{Bmatrix} \tag{7.11}$$

Secondly, there's the Piecewise-Linear function. This function once more will get the values of zero or one, however also can fight values between that looking on the amplification factor in a definite region of linear operation.

$$\varphi(v) = \begin{Bmatrix} 1\ if\ v \geq \frac{1}{2} \\ v\ if - \frac{1}{2} > v > \frac{1}{2} \\ 0\ if\ v < 0 \end{Bmatrix} \tag{7.12}$$

Thirdly, there's the sigmoid function.
This function will vary between zero and one, however its conjointly typically helpful to use the -1 to one vary. AN example of the sigmoid function is that the hyperbolic tangent functions.

$$\varphi\left(v\right) = \tanh\left(\frac{v}{2}\right) = \frac{1-\exp(-1)}{1+\exp(-1)}) \tag{7.13}$$

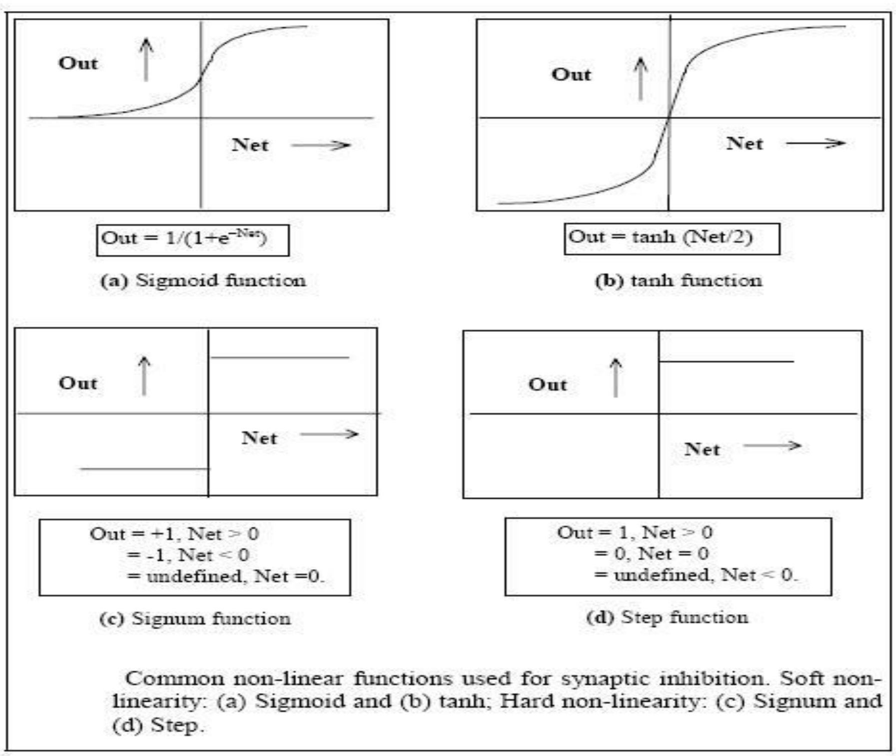
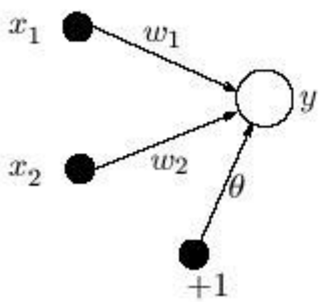


Figure 7.9 common linear functions used for synaptic inhibition

The artificial neural networks that we tend to describe are all variations on the parallel distributed process (PDP) plan. The design of every neural network is predicated on terribly similar building blocks that perform the process. During this chapter we tend to 1st discuss these process units and discuss totally different neural network topologies. Learning methods as a basis for an adjustive system.

Networks with threshold activation functions:-

A single layer feed-forward network consists of 1 or a lot of output neurons o, every of that is connected with aweigh factor wio to all or any of the inputs i. within the simplest case the network has solely 2 inputs and one output, as sketched in figure-



(We leave the output index o out). The input of the nerve cell is that the weighted add of the inputs and the bias term. The output of the network is created by the activation of the output nerve cell, that is a few operate of the input:

$$y=F\left(\sum_{i=1}^{2} w_i\, x_i + \theta\right) \tag{7.14}$$

The activation function F is linear so we've got a linear network, or nonlinear. In this section we tend to take into account the threshold (or or sgn) function:

$$f(s) = \begin{Bmatrix} 1\ if\ s > 0 \\ -1\ if\ s < 0 \end{Bmatrix} \tag{7.15}$$

The output of the network therefore is either +1 or -1 looking on the input. The network will currently be used for a classification task: it will decide whether or not AN input pattern belongs to at least one of 2 categories. If the entire input is positive, the pattern are assigned to category +1, if the entire input is negative, the sample are assigned to category -1.The separation between the 2 classes during this case may be a line, given by the equation: input:

$$w_1x_1 + w_2x_2 + \theta = 0 \qquad (7.16)$$

There are two learning methods for these types of net-works: 1. 'perceptron' learning rule and 2.the 'delta' or 'LMS' rule. Both methods are iterative procedures which adjust the weights. For each weight the new value is computed by doing a correction to the old value. The threshold is updated as:

$$w_i(t+1) = w(t)_i + \Delta w_i(\mathrm{t}) \qquad (7.17)$$

$$\theta(t+1) = \theta(t) + \Delta\theta(t) \qquad (7.18)$$

7.4.K Nearest Neighbors – Classification:-

K nearest neighbors may be a easy algorithmic program that stores all obtainable cases and classifies new cases supported by a similarity measure (e.g., distance functions). KNN has been utilized in statistical estimation and pattern recognition already within the starting of 1970's as a non-parametric technique.

Algorithm:-

A case is classed by a majority vote of its neighbors, with the case being allotted to the category most typical amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is solely appointed to the category of its nearest neighbor.

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

(7.15)

It ought to also be noted that every 3 distance measures are solely valid for continuous variables. Within the instance of categorical variables the performing distance should be used. It additionally brings up the problem of standardization of the numerical variables between zero and one once there's a mix of numerical and categorical variables within the dataset.

Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

(7.19)

Choosing the optimum value for K is best done by initial inspecting the info. In general, an oversized K value is a lot of precise because it reduces the noise however there's no guarantee. Cross-validation is differently to retrospectively verify a decent K price by victimization Associate in Nursing freelance dataset to validate the K price. Traditionally, the optimum K for many datasets has been between 3-10. That produces far better results than 1NN.

CHAPTER 8: Related work and comparative study of previously given methods

With the supply of well-established signal process and statistical methods, techniques used on seizure prediction vary generally. Intelligibly, features that characterize the changes in electroencephalogram are being given the foremost importance from the starting of seizure prediction analysis. These feature set could be linear, non-linear, univariate (single channel) or variable (multiple channels) and also the changes in these features may be characterized using machine learning or thresholding

. Early work on seizure prediction geological dating back to the Seventies was performed based on surface electroencephalogram recordings within the absence of seizures using linear approaches to extract seizure precursors [18]. Rogowski et al. [19] and Salant et al. [20] have used autoregressive modeling to spot pre-ictal changes within 6 s before seizure onset. Siegel et al. [21] identied characteristic changes between the one-minute periods before a seizure and similar baseline periods for individual patients. during this study, statistical condense of the endings was assessed and also the influence of different vigilance states was discussed.

Spike incidence rate within the electroencephalogram was evaluated as potential pre-ictalchange during a few studies. Whereas lange et al. [22] identified a reduced focal spiking rate and an enhanced rate of bilateral spikes before seizures; different studies showed no visible changes in spike rates before seizures [23, 24, 25]. Advances within the mathematical theories of non-linear systems in the opened up a great deal of latest approaches in modeling dynamically changing advanced systems. Statistic analytics became applicable in seizure prediction with the availability of long run electroencephalogram recordings. within the early Nineteen Nineties, Iasemidis et al. [26] found that the most important Lyapunov exponent may be a seizure precursor as its behavior was ascertained to vary throughout the pre-ictal stages of intracranial electroencephalogram recordings. Martinerie et al. [27] reported decrease within the correlation density before the seizures pre-ictally. They additionally developed a measure referred to as dynamic similarity index that quantized changes in dynamics relative to a relentless pre-ictal reference window. Their works identified reduced dynamic similarity in pre-ictal regions of intracranial and scalp electroencephalogram recordings [28, 29, 30]. However, these studies centered solely on the pre-ictal regions of the recordings and neglected the baseline characteristics on the identified measures. so the analysis of the investigated measures' applicability in seizure prediction was incomplete because of the unknown specificities. Later works analyzed the specificities by comparison the measures in each inter-ictal and pre-ictal regions. Navarro et al. [31] ascertained that in chosen examples of their subjects, the similarity measures showed additional frequent drops before seizures than throughout inter-ictal regions. phase synchronization between different areas in brain was evaluated as a seizure precursor by Mormannet al. [32].These results were corned by 2 works of le Van Quyenet al. [33, 34] on cerebral cortex brain disease. Chavez et al. [35] reported that preictal changes in phase synchronization occur preponderantly in beta band based on their analysis once bandpass altering of electroencephalogram. Several measures together with the correlation dimension [36, 37] (as a measure for dynamic complexity), dynamic entrainment [38] (denied because the convergence of largest Lyapunov exponents in sure chosen channels), accumulated signal energy [39, 40], simulated neural cell models or part synchronization were shown to be appropriate for differentiating inter-ictal from pre-ictal knowledge.

A number of studies revealed ranging from 2003 raised skepticism in seizure prediction because the earlier optimistic studies couldn't be reproduced. De Clercq et al. and Winterhalder et al. questioned the optimistic results obtained by similarity index .Correlation dimension was reevaluated in and also the previous work on that was challenged. Similarly, the work on accumulated energy [39] couldn't be reproduced in . Studies by Lai et al. raised doubts concerning the quality of the Lyapunov exponent for seizure prediction.

McSharry et al. questioned the performances of nonlinear features like correlation density [27]. Upon evaluation, the studies showed that this measure was additional or less a reaction of the variance in electroencephalogram signals. As a suggestion for more studies on non-linear measures, the authors pointed out that usage of nonlinear or difficult features shouldn't be taken into account unless it will be shown that these measures so outmatch easy linear measures.

Starting from 2004, machine learning algorithms were getting used for generation of seizure warnings rather than manual labeling. Machine learning methods are able to map the advanced relationships between the features extracted from the electroencephalogram recordings to seizure annotations. This ability of machine learning techniques remarkably exaggerated the prediction capability when used with the features that showed pre-ictal changes in earlier studies. from 2004,features were utilized in every study. Feature engineering has been explored multiple times from the starting of seizure prediction analysis. Several linear, non-linear features and their combinations are tried as bio-markers for brain disease with no specific try providing extraordinary results. Therefore, we have a tendency to believe that the essential assumption of one or combination of features being the bio-markers for all people is wrong. Also, usage of machine learning for warning generation is very necessary to realize the most effective trade-o between sensitivity and specificity. Therefore, during this work, we've used a machine learning primarily based approach to spot patient specific bio-markers and used them for seizure prediction to realize best results. In doing that, we have a tendency to additionally performed a pre-ictal analysis to say that the bio-markers identified from the machine learning based analysis do so show pre-ictal changes resulting in seizures.

CHAPTER 9: PROPOSED METHODOLOGY

9.1. Methodology

as shown in Figure 9.1 method was created for characteristic the foremost appropriate combination of spatiality reduction technique paired with SVM that gave the very best sensitivity and specificity in classifying epileptic and non-epileptic information. Every 1minute data section within the training matrices and test data was normalized and decomposed into wavelets from that features were extracted. The steps for decomposing the signal and extracting features are represented in detailed in sections four.4 and 4.5 severally. Next, the features were reduced using CA or ICA or LLE or ISOMAP or MDS or LAPLACIAN Eigen MAPS. Finally, the SVM formula was enforced with the training matrix created using the reduced set of features via PCA or ICA or LLE or ISOMAP or MDS or LAPLACIAN Eigen MAPS. The SVM classifier training using PCA and ICA were just like the paper by authors Subasi and Gursoy. Another models using LLE or ISOMAP or MDS or LAPLACIAN eigen MAPS are employed in this thesis to make a training set was introduced to look at the accuracy of the classifier. Principal elements represent the variation in information and also the independent elements establish the independent sources contributive to the variance within the information.

Once the SVM was trained and also the classifiers obtained for every model, the test information set was classified. The sensitivity and specificity of every classification technique or model was computed using equations 9.1 and 9.2. The sensitivity (Equation 9.1) of the model is that the live of its ability to correctly identify the epileptic channels, and also the specificity (Equation 9.2) of the model is the measure of its ability to properly establish non-epileptic channels.

$$\text{SENSITIVITY} = \frac{true\ positive}{true\ positive + false\ negative} * 100 \qquad (9.1)$$

$$\text{SPECIFICITY} = \frac{true\ negative}{true\ negative + false\ positive} * 100 \qquad (9.2)$$

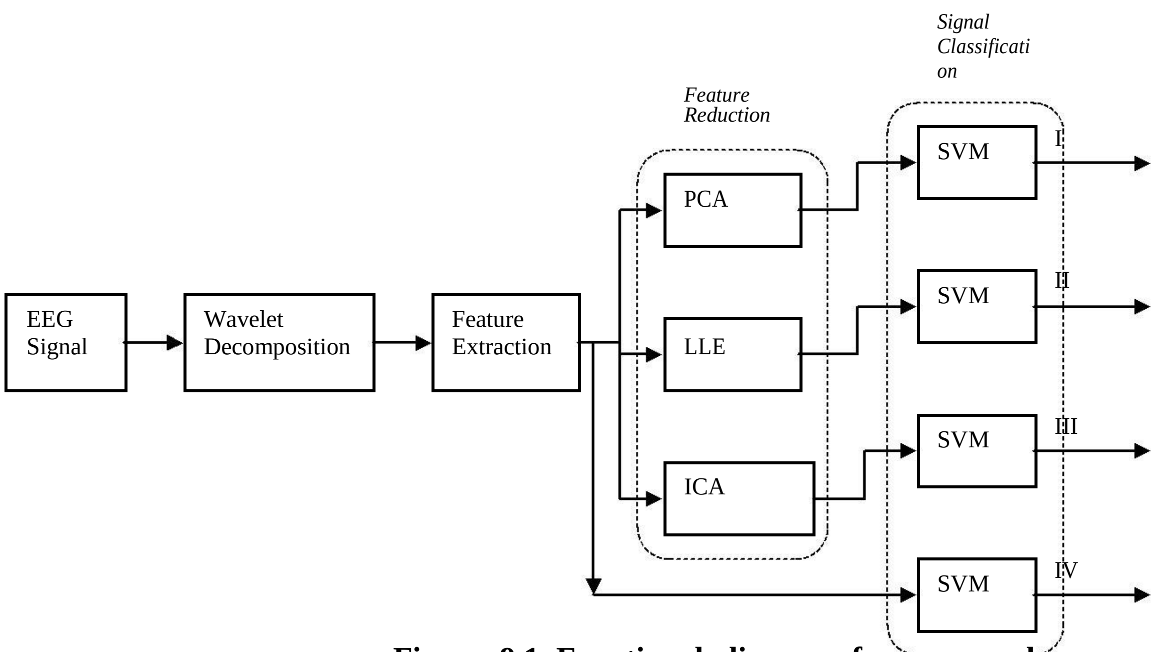


Figure 9.1 Functional diagram for proposed methodology

9.2. Overview of software system used:-

All software system development was done in MATLAB R2013a (Mathworks, Natick, MA) for performing arts the subsequent steps:

1.Reading the CHB-MIT_SCALP encephalogram data for analysis

2.Performing discrete wavelet decomposition (DWT) on the samples.

3.Reducing features using PCA or ICA or LLE or ISOMAP or MDS or LAPLACIAN Eigen MAPS algorithms

4.Implementing machine learning using SVM algorithmic rule on the training matrix

5. Getting performance measures of sensitivity and specificity for test data set.

EEGLAB, an open supply MATLAB compatible package, was used to import the patient information for analysis in step one. EEGLAB is an open source setting for electrophysiological signal process and was developed by the Swartz Center for machine neuroscience (SCCN) and is distributed under the gnu General Public License22. The wavelet toolbox is a component of the MATLAB 2013a that permits analysis and synthesis of signals and pictures using wavelet techniques and was used for performing step 2.For dimension reduction I used Matlab toolbox for dimensionality Reduction (v0.8.1b) by laurens van der Maaten[13]. SVM is additionally a function obtainable within the Statistics toolbox of MATLAB 2013 and was used for classifying information into 2 categories-

9.3. Analysis using discrete wavelet transforms:-

the wavelet technique applied to the encephalogram signals reveals features involving the transient nature of the signal in each the time and frequency content. The DWT analyzes the signal at numerous frequency bands by moldering the signal into coarse and fine data. This decomposition into frequency bands is achieved by using filter banks that divide a symptom into various spectral parts known as sub-bands.

The signals are divided into high pass (Hipass_D) and low pass (Lopass_D) spectral characteristics wherever the high pass filter is like applying a wavelet to the initial signal and therefore the low pass filter is like applying a scaling or smoothing function. Figure 4.2 shows the sub-band decomposition of the signal using DWT during which every stage is formed from high pass and low

pass filters. the primary filter Hipass_D is that the discrete mother wavelet, high pass in nature and therefore the second filter Lopass_D is its mirror version, low pass in nature. The down sampled outputs of the filters are the detailed decomposition known as D1 and approximate decomposition known as A1. The approximation signal A1 is any decomposed using the high pass and low pass filter combine.

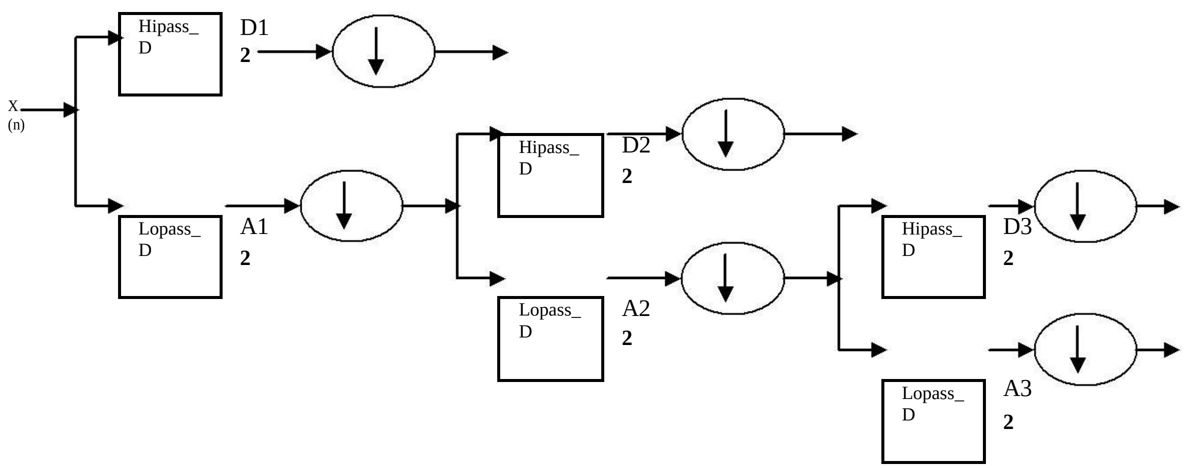


Figure 9.2 Sub band decomposition of DWT

Table 9.1 Ranges of frequency bands for the scalp EEG signal with a sampling frequency of 250 Hz

Decomposed Signal	Frequency-Range (Hz)
D1	33.75-67.5
D2	16.87-33.75
D3	8.43 -16.87
D4	4.21-8.43
A4	0-4.21

Table 9.1 shows the amount of decomposition levels chosen for the Scalp encephalogram signal based on the dominant frequency elements of the signal. The frequency of the scalp encephalogram signals used for analysis was 250Hz. the degree shown in Table 9.1 were chosen specified every frequency sub-band retains frequencies necessary for classification of the signal. The scalp encephalogram signals were decomposed into D1 D4 detailed coefficients and final approximation coefficient A4.

The signal was decomposed using the functions from the wavelet toolbox in MATLAB. Daubechies filter ("db6") was used for getting the detailed and approximation coefficients.

9.4. Feature Matrix:-

After the wavelet coefficients were reconstructed, the detailed coefficients D1, D2, D3,D4 and approximate coefficient A4 were used to generate the “Feature matrix”. Supported Table 9.1, these coefficients represent the range of frequency bands from 0 – 67.5Hz. Feature matrix is the illustration of the signal using statistics over the set of wavelet coefficients D1-D4 and A4. The fifteen statistical options representing the time-frequency distribution of the Scalp encephalogram signal for every sub band (as represented in Table 9.1) are:

1. Mean of the absolute values of the wavelet coefficients: Four features representing the statistical distribution of the signal.
2. Average power of the wavelet coefficients: Four features representing the statistical distribution of the signal.
3. Variance of the wavelet coefficients: Four features representing the quantity of change within the statistical distribution .
4. Ratio of the mean values between adjacent sub bands: 3 features representing the relative modification within the statistical distribution.

These statistical features were chosen during this thesis to emulate the feature extraction methodology explained within the paper written by Subasi and Gursoy[2]. The authors selected these statistical measures as a result overpromising results were ascertained using these statistical measures with classification of respiratory organ sounds. every 1minute data phase, sampled at 250, was more divided into sub segments of 1second or 250 samples and an overlap of 0.5 seconds or 125 samples between adjacent sub segments was accustomed calculate the statistical features for the feature matrix. therefore every 60 second data phase resulted in an exceedingly feature matrix with 119 rows, one for every one second sub segment and fifteen columns for every of the fifteen statistical features.

9.5. Model for dimension reduction:-

For dimension(feature) reduction I used MATLAB toolbox .using this toolbox by calling numerous feature reduction algorithms, I got new reduced feature set .For every sub-sample's feature set of 15 features there was either 2 or 3new features. thus I got new reduced feature set of dimensions of N X 2 or N X 3 instead of previous original feature set of dimensions N X 15.

9.6. Obtaining SVM Classifiers:-

SVM formula classifies the 2 categories (epileptic and non-epileptic) by finding an best hyperplane with the biggest margin that separates the information points of the categories. The support vectors are the information points nearest to the current hyperplane and on the margins of the border separating the categories. For non-separable data, a softer margin is known that separates several if not all points. The basic premise of SVM is to provide a classifier (based on the training matrix) that predicts the classification of the test data set given solely the test data attributes. Training an SVM classifier was achieved in 3 distinct steps; opening was to train the machine classifier, the second step was to classify the data using the classifier and also the third step was to tune the classifier for best classification. fitcsvm() could be a MATLAB operate that's obtainable from the statistical tool chest for training SVM. The training matrix containing the epileptic and non epileptic knowledge beside the category matrix containing the classification (class = -1 for epileptic and class = +1 for non-epileptic) was used to train the SVM. For all the said models, the information matrix modified based on the amount of columns used for making the training matrix.

The radial basis function (RBF) kernel that could be a Gaussian operate was used for training SVM classifiers. This operate non-linearly maps the information points onto a better dimensional house wherever it becomes near to linearly severable under the change of variables. The ensuing SVM trained classifier contains the optimized parameters that helped classify the take a look at data set. predict() operate from MATLAB beside the classifier was accustomed classify the test knowledge set, wherever every row corresponds to a brand new time purpose. The take a look at knowledge set contains all 119 time points for the sixty second knowledge phase for every channel enclosed within the take a look at knowledge. SVM classifier then classifies each time point for the channel as epileptic or non-epileptic. so as to spot whether or not the whole channel will be classified as epileptic or non-epileptic, the preponderance of classification was used. For the channel to be classified as epileptic, fifty one of the 119 time points are needed to own a classification as epileptic (class = -1) and for the channel to be classified as non-epileptic, fifty one of the 119 time points are needed to own a classification as non-epileptic or (class = +1).

CHAPTER10 RESULT, DISCUSSION AND EVALUTION OF METHODOLOGY

10.1. Results

Figure 10.1 shows the scalp EEG data of ictal activity in epileptic data.

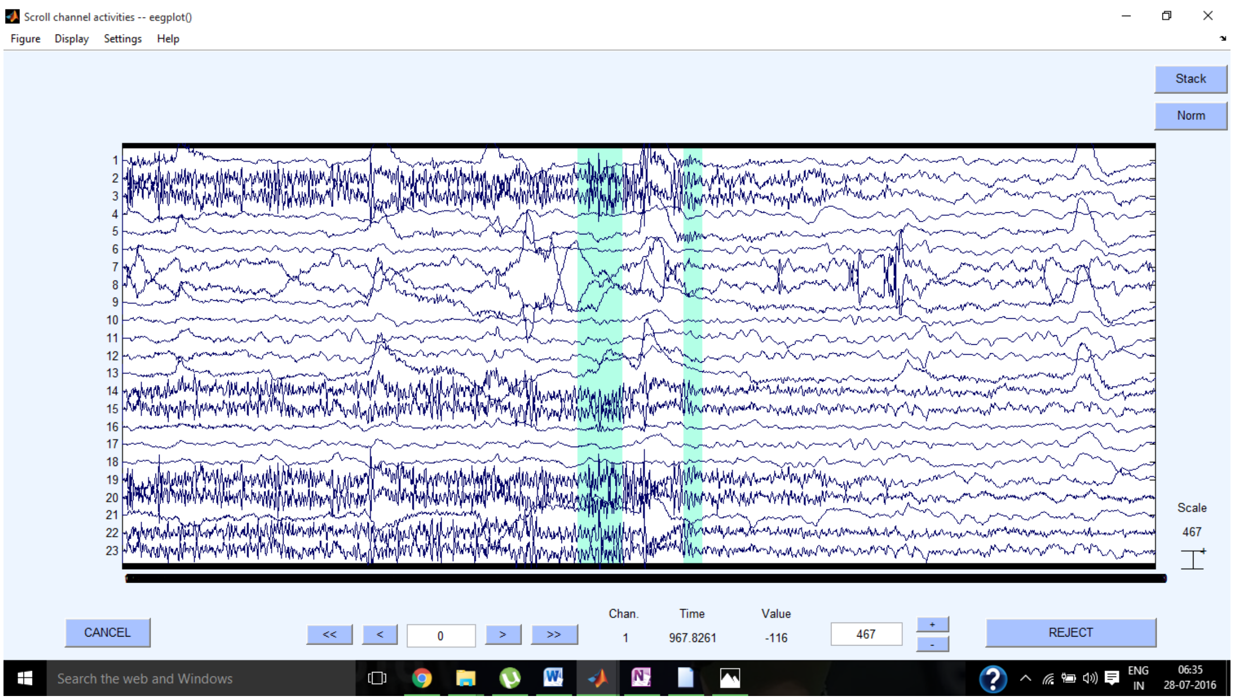


Figure 10.1 ictal Epileptic scalp EEG **data**

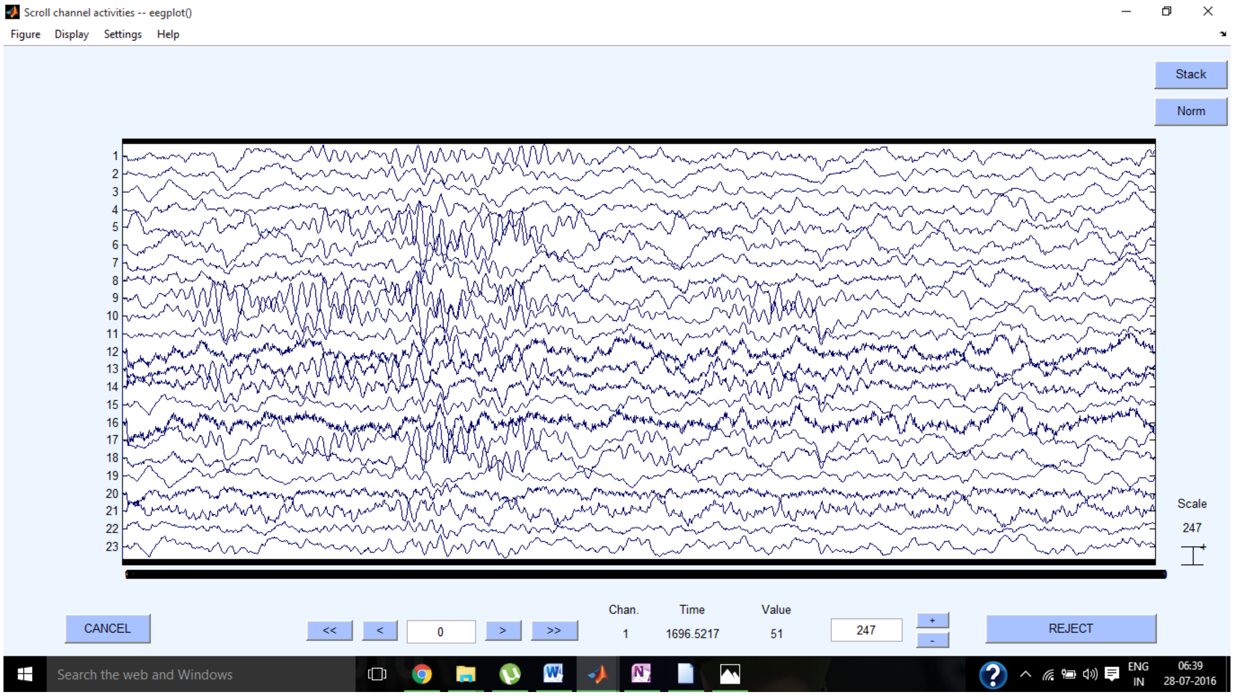


Figure 10.2 pre-ictal Epileptic scalp EEG **data**

Figure 10.2 shows the pre-ictal EEG waveform

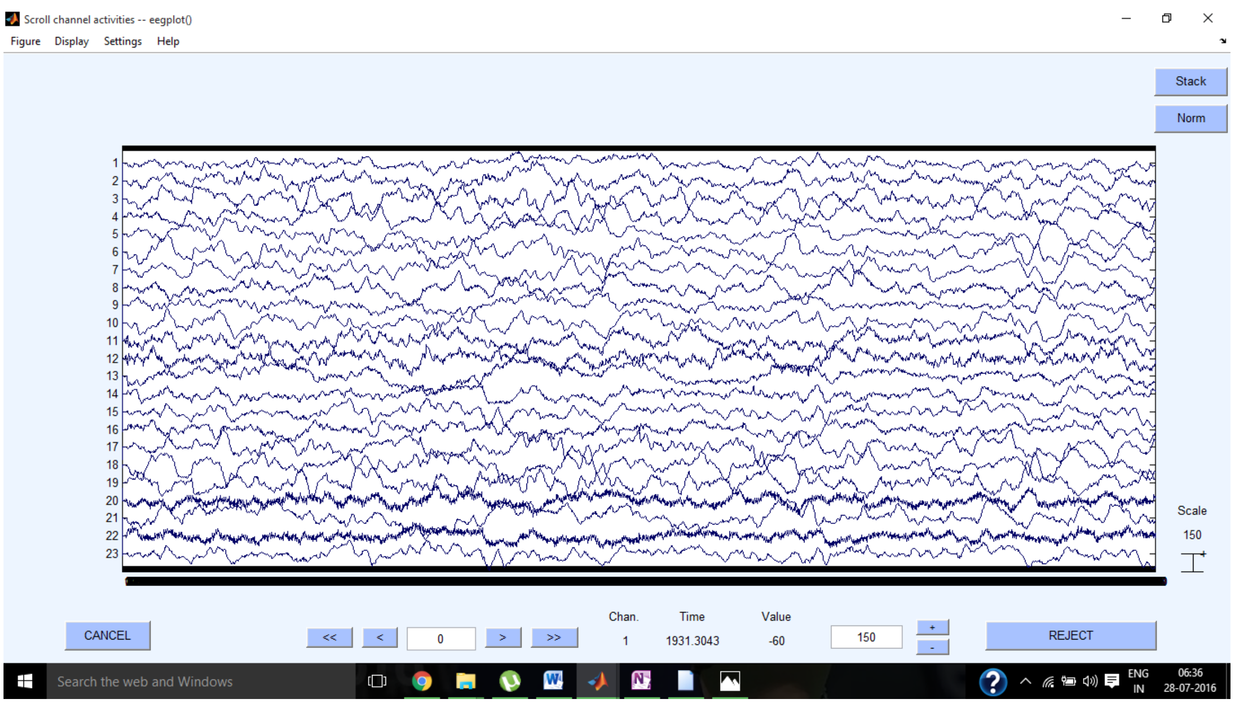


Figure 10.3 non- Epileptic scalp EEG

Here now I want to extract DWT coefficients of total119 subsamples of 1 sec each .here below in figures for same channel amplitude waveform and DWT waveforms are generated.

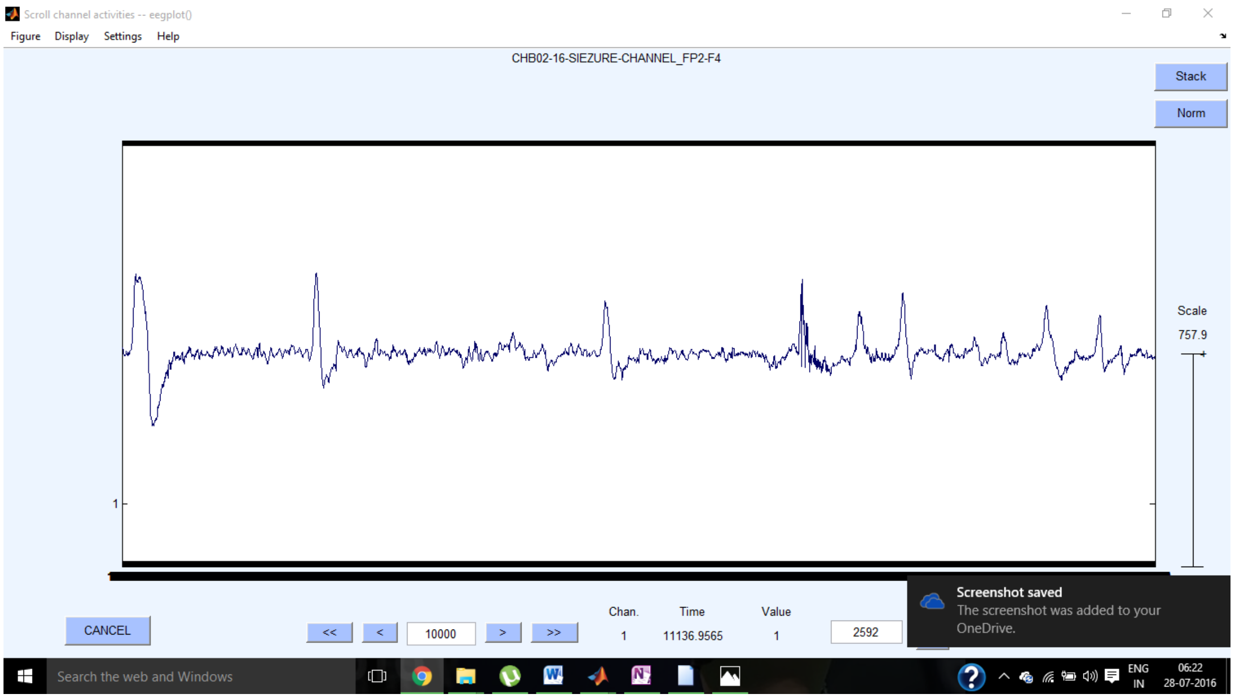


Figure 10.4 seizures for chb02-16-channel FP2-F4

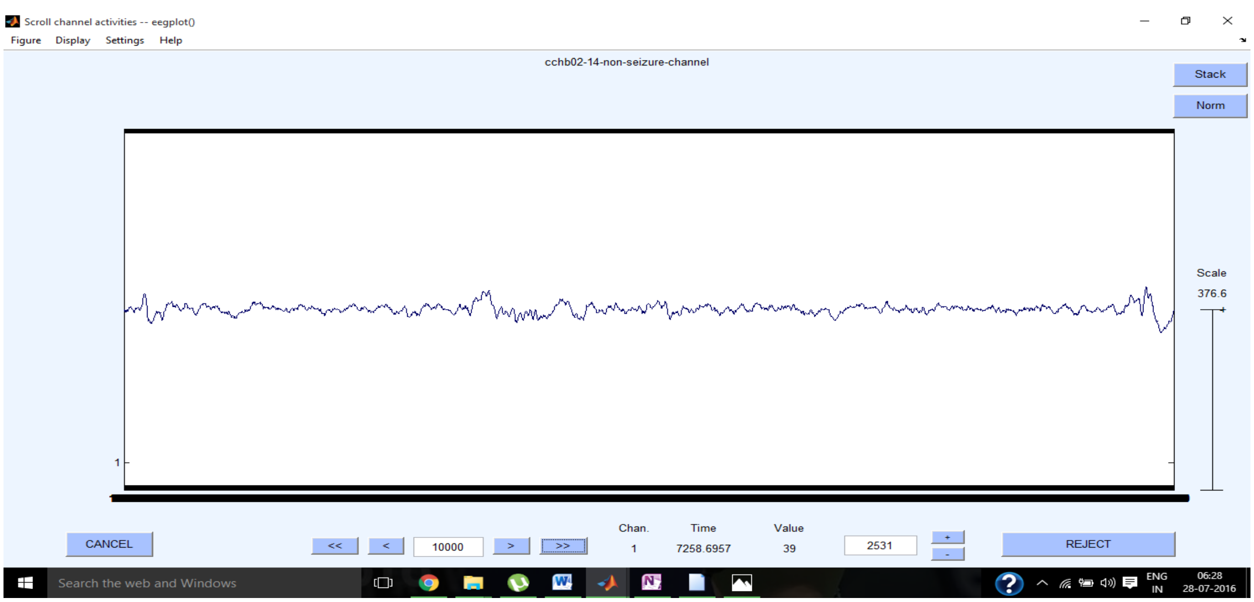


Figure 10.5 non- seizure for chb02-14-channel FP2-F4

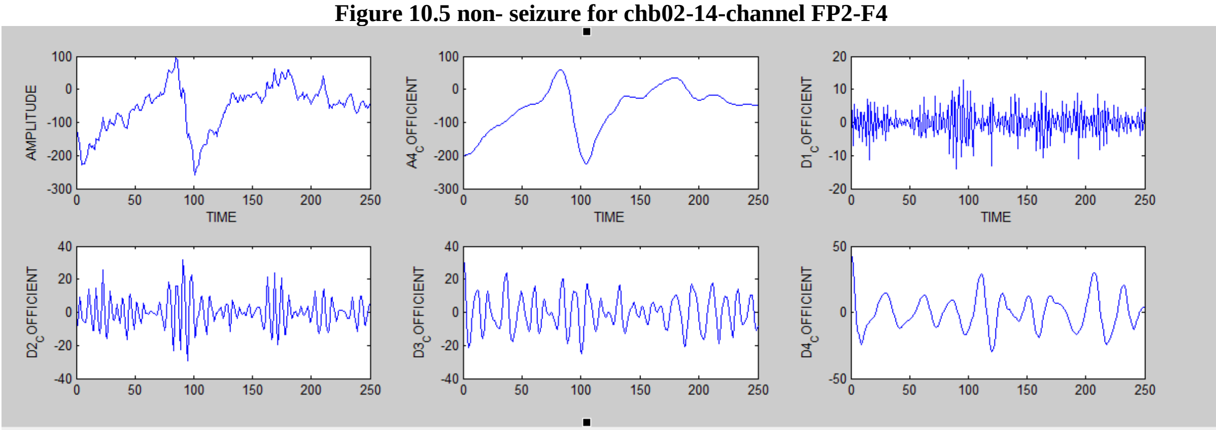


Figure 10.6 Epileptic data

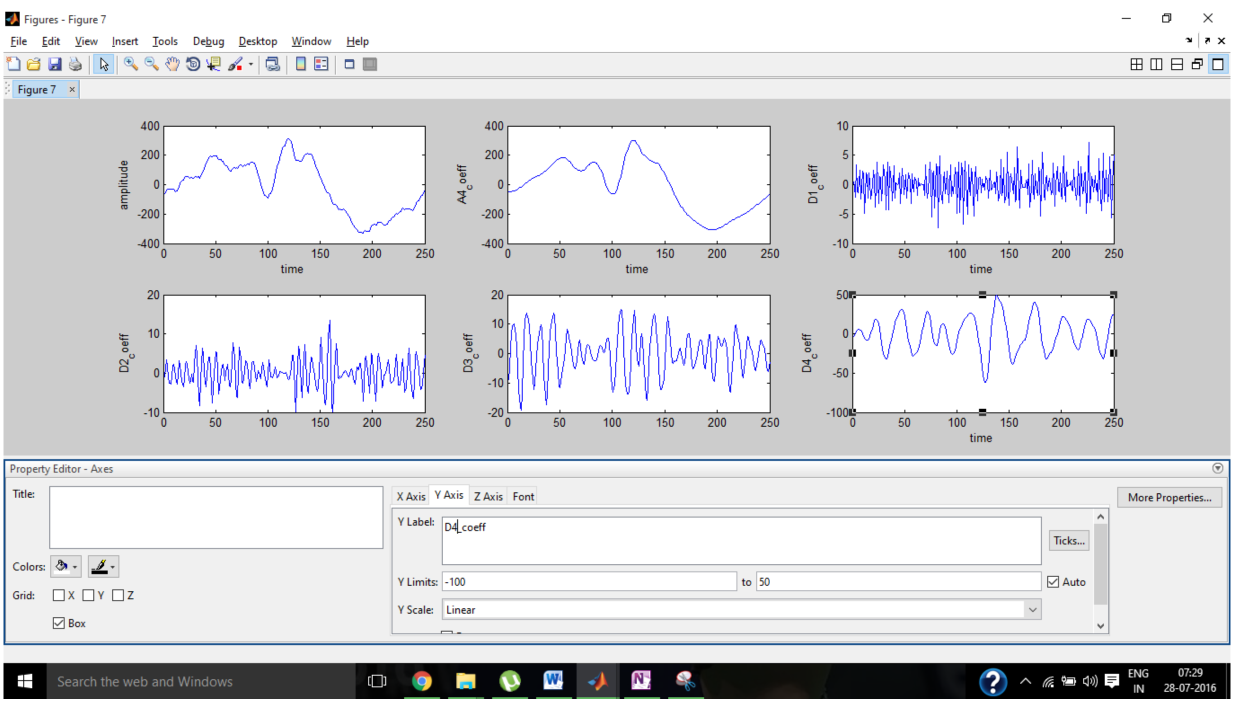


Figure 10.7 NON-Epileptic data

Extracted coefficients provided a illustration of energy distribution of the scalp EEG signal in time and frequency. Every coefficient was reconstructed using wrcoef() function from the wavelet tool case. The feature matrix obtained from these coefficients had 119 time points and fifteen features as explained in section 9.5. If preponderance (greater than 51%) shows that the channel is epileptic because the SVM classifier classifies epileptic data as -1 and non-epileptic data as +1. Figure 10.8 shows the channel as being non-epileptic. once more the preponderance of your time points is shown as non-epileptic that is +1.

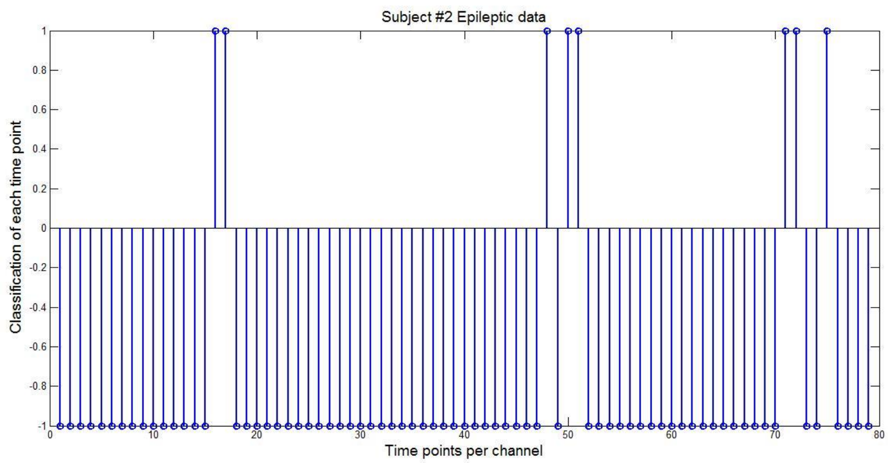


Figure 10.8 Preponderance of time points showing the channel as epileptic

n

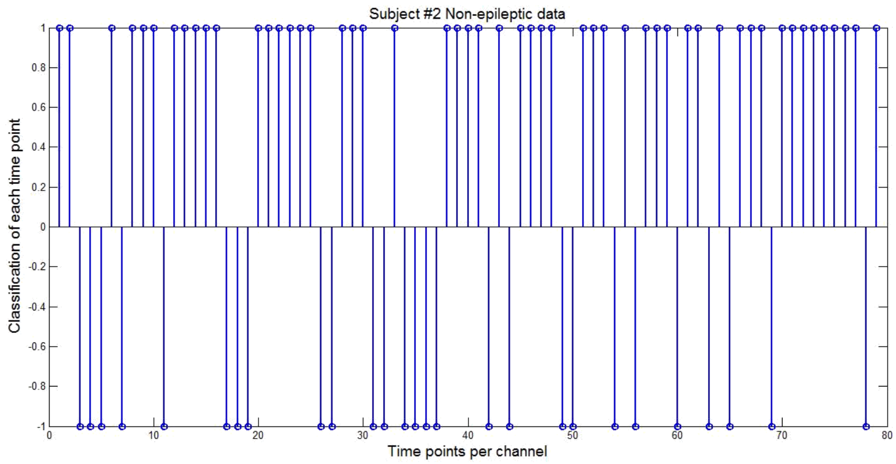


Figure 10.9 Preponderance of time points showing the channel as non-epileptic

Using this rule, a training matrix for every model was created using information from twelve seizures tough by Subject #2. The training matrix contained epileptic information from a complete of 253 channels with 20000 time points and non-epileptic information from a hundred ninety channels with 15000 time points. The feature matrix for every of the channels was computed on an individual basis using the rule and every one the feature matrices were combined to form the training matrix. It ought to be noted that the training matrix contains a lot of epileptic channels than non-epileptic channels; the reason was to extend the sensitivity of the SVM classifiers in classifying the epileptic from non-epileptic information.

The test information set was created using the epileptic and non-epileptic information of Subject CHB02-14 not utilized in the training matrix. This test information set was used with every model to spot the model's sensitivity and specificity.

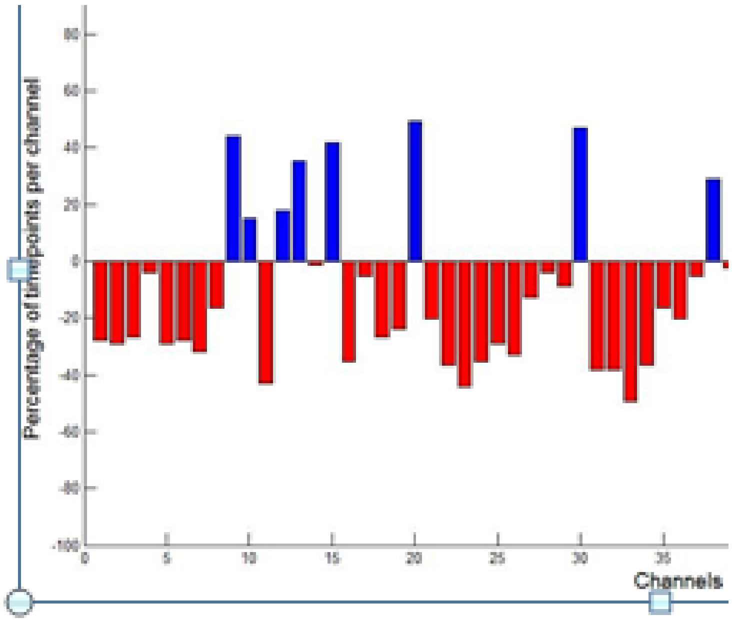


Figure 10.10 Channel classification performed by classifier when ICA is used
NOTE: here I used the sample of 20 sec with 1 second window size and .5 sec overlap

10.2. Evaluation of proposed methodology:-

Table 10.2 Sensitivity and specificity of the four models

DIM_REDUCTION TECHNIQUE	**Sensitivity**	**Specificity**	**Comment**
PCA	100%	60%	
ICA	100%	50%	
LLE	97%	90%	Best performance model
ISOMAP	98%	40%	
LAPLACIAN EIGEN MAPS	91%	79%	
MDS	93%	84%	

10.3. DISCUSSION

The test data set was ready to give the sensitivity and specificity of each model studied during this thesis. the utilization of varied spatiality reduction strategies to scale back spatiality of the feature matrix used to train the SVM classifier to differentiate between the pre-ictal and non-epileptic signals were explored. The target of this thesis was the classification of data alongside minimal computational complexity and good sensitivity.

After choosing the data length of extraction for analysis as 60 seconds, the separate wavelet remodel was used to realize the wavelet coefficients of the signal segment. Wavelet coefficients D1, D2, D3, D4 and A4 were used for features generation. These features were mean of the absolute values, average power, variance, and ratio of the mean so as to get the statistical distribution alongside modification within the statistical distribution in every sub band. For making the feature matrix, the signal was divided into 1 second sub segments with associate overlap of 0.5 seconds. Every channel of data was drawn within the feature matrix as a complete of 119 time points with a complete of fifteen features.

The feature matrix was then used with dimension reduction methods. Then a replacement feature set is acquired with low spatiality.in our case for all of cases new dimensions were 2 or 3.then using SVM, signal was classified as elliptic or non-elliptic. My technique using LLE got sensible results, which were evaluated using sensitivity and specificity parameter.

CHAPTER 11: FUTURE WORK AND CONCLUSION

11.1. Future Work

Since, proposed methodology performed well, but in future we need to try different feature extraction methods or different statistical features for improving the accuracy of classification of SVM.

In future this methodology should also be tested using another classifiers line k-nearest neighbor and ANN.

Since this proposed method classify the channel not the whole signal as epileptic or non-epileptic, so in future it is need to develop such methods which can classify multi-channel signal as epileptic or non-epileptic by incorporating whole channel information(features).

Although this method used dimension reduction techniques for reducing computational complexity, but also did separate analysis of each channel, so here a bulk feature set will be created.so if we can develop a method which can select some of channels of our interest then there will be small feature set(less than 3 or 4 channel features set), so one will need less computational capability. This approach will be ideal for real time systems.

11.2. Conclusion

This thesis contributes to the detection of epilepsy by providing an automatic classification technique that enables the info to be sorted as epileptic or non-epileptic. Whereas many commercially available software packages exist for helping medical professionals in creating this distinction, recording equipment that may intuitively classify the available data and learn to enhance its performance, is a lot of helpful. This recorder or support vector machine may be used to classify the info as either epileptic or non-epileptic based on presently available data. With clinical SCALP EEG single patient data for training and testing, using LLE model the sensitivity and specificity of the SVM classifier were accordingly 97% and 90%.

REFERENCES

[1] S. Shorvon and P. Farmer, \Epilepsy in developing countries: A re-view of epidemiological, sociocultural, and. .treatment aspects," Epilep-sia, vol. 29, no. s1, pp. S36{S54, 1988..
.
[2] .C. J. Murray, A. D. Lopez et al., \Global comparative .assessments in the health sector: Disease burden, expenditures and intervention packages," World Health Organization, Geneva, 1994.

[3] P. Kwan and M.. J. Brodie, \Early identi cation of .refractory epilepsy," New England Journal of Medicine, vol. 342, no. 5, pp. 314{319, 2000.

[4] A. Schulze-Bonhage and A. K•uhn, \Unpredictability of seizures and the burden of ep.ilepsy," Seizure Prediction. in Epilepsy: From Basic Mechanisms to Clinical Applications, 2008..

[5] A. Schulze-Bonhage, F. Sales, K. Wagner, R. Teotonio, A. Carius,
A. Schelle, and M. Ihle, \Views of patients .with epilepsy on seizure prediction devices," Epilepsy & Behavior, vol. 18, no. 4, pp. 388{396, 2010.

[6] R. S. Fisher, \Epilepsy from. the patient's perspective: Review of results of a community-based survey," Epilepsy & Behavior, vol. 1, no. 4, pp. S9{S14, 2000.

[7] R. S. Fisher, \Therapeutic devices .for. epilepsy," Annals of Neurology, vol. 71, no. 2, pp. 157{168, 2012.

[8].CHB-MIT aScalp EEG database *“Apache/2.4.10 (Debian) OpenSSL/1.0.1t Server at physionet.org Port 443”*

[9].Alotaiby et al. EURASIP Journal on Advances in Signal Processing 2014, 2014:183
http://asp.eurasipjournals.com/content/2014/1/183

[10]. Dimensionality Reduction Techniques for Processing
Epileptic Encephalographic Signals
R.HARIKUMAR* and P.SUNIL KUMAR
DOI: http://dx.doi.org/10.13005/bpj/587
(Received: February 03, 2015; accepted: March 19, 2015)

[11]. Takenandrew W. Moore 2003

[12]. Andrew W. Moore slides 2003

[13] Matlab toolbox for dimensionality Reduction (v0.8.1b) by laurens van der Maaten

[14] Subasi, A. & Ismail Gursoy, M. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* 37, 8659–8666 (2010).

[15] I. Leppik, Contemporary Diagnosis and Management of the Patient with Epilepsy. Handbooks in Health Care, 1997. [.Online]. Available: https://books.google.com/books?id=WMxrAAAAMAAJ

[16] J. S. Ebersole and T. A. Pedley, Current Practice of Clinical Electroen-. cephalography. Lippincott Williams & Wilkins, 2003.

[17] F. S. Tyner, J. R. Knott, and W. B. Mayer, Fundamentals of EEG Technology.: Clinical Correlates. Lippincott Williams & Wilkins, 1989, vol. 2.

[18] S. Viglione .and G. Walsh, \Proceedings: Epileptic seizure prediction," Electroencephalography and Clinical Neurophysiology, vol. 39, no. 4, pp. 435{436, 1975.

[19] Z. Rogowski,. I. Gath, and E. Bental, \On the prediction of epileptic seizures," Biological Cybernetics, vol. 42, no. 1, pp. 9{15, 1981.

[20] Y. Salant, I. Gath, and O. Henriksen, \Prediction of epileptic seizures from two-channel .EEG," Medical and Biological Engineering and Computing, vol. 36, no. 5, pp. 549{556, 1998.

[21] A. Siegel, C. L.. Grady, and A. F. Mirsky, \Prediction of spike-wave bursts in absence epilepsy by EEG power-spectrum signals," Epilepsia, vol. 23, no. 1, pp. 47{60, 1982.

[22] H. H. Lange, J.. P. Lieb, J. Engel, and P. H. Crandall, \Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy," Electroencephalogr.aphy and Clinical Neurophysiology, vol. 56, no. 6, pp.

[23].J. Gotman and M. Marciani, \Electroencephalographic spiking activ-ity, drug levels, and seizure occurence in epileptic patients," Annals of Neurology, vol. 17, no. 6, pp. 597{603, 1985.

[24]J. Gotman and D. Ko er, \Interictal spiking increases after seizures but does not after decrease in medication," Electroencephalography and Clinical Neurophysiology, vol. 72, no. 1, pp. 7{15, 1989.

[25]A. Katz, D. A. .Marks, G. McCarthy, and S. S. Spencer, \Does interictal spiking change prior to seizures?" Electroencephalography and Clinical Neurophysiology, vol. 79, no. 2, pp. 153{156, 1991.

[26]L. D. Iasemidis, .\Epileptic seizure prediction and control," Biomedical Engineering, IEEE Transactions on, vol. 50, no. 5, pp. 549{558, 2003.

[27]J. Martinerie, C. A.dam, M. Le Van Quyen, M. Baulac, S. Clemenceau, B. Renault, and F. J. Varela, \Epileptic seizures can be anticipated by non-linear analysis," Nature Medicine, vol. 4, no. 10, pp. 1173{1176, 1998.

[28]M. Le Van Quy.en, J. Martinerie, M. Baulac, and F. Varela, \Anticipat-ing epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," Neuroreport, vol. 10, no. 10, pp. 2149{2155, 1999.

[29]M. Le Van Quyen, C. Adam, J. Martinerie, M. Baulac, S. Clemenceau, and F. Varela, \Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures," Euro-pean Journal of Neuroscience, vol. 12, no. 6, pp. 2124{2134, 2000.

[30]M. Le Van Quyen, .J. Martinerie, V. Navarro, P. Boon, M. D'Have, C. Adam, B. Renault, F. Varela, and M. Baulac, \Anticipation of epilep-tic seizures from standard EEG recordings," The Lancet, vol. 357, no. 9251, pp. 183{188, 2001.

[31]V. Navarro, J. Martinerie, M. Le Van Quyen, S. Clemenceau, C. Adam, M. Baulac, and F. Varela, \Seizure anticipation in human neocortical partial epilepsy," Brain, vol. 125, no. 3, pp. 640{655, 2002.

[32].F. Mormann, K. Lehnertz, P. David, and C. E. Elger, \Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," Physica D: Nonlinear Phenomena, vol. 144, no. 3, pp. 358{369, 2000.

[33]M. Le Van Quyen, J. Martinerie, V. Navarro, M. Baulac, and F. J. Varela, \Characterizing neurodynamic changes before seizures," Journal of Cl.inical Neurophysiology, vol. 18, no. 3, pp. 191{208, 2001.

.

[34] M. L. V. Quyen, V. Navarro, J. Martinerie, M. Baulac, and F. J. Varela, \Toward a neurodynamical understanding of ictogenesis," Epilepsia, vol. 44, no. s12, pp. 30{43, 2003.

[35] M. Chavez, M. L. Van Quyen, V. Navarro, M. Baulac, and J. Martinerie, \Spatio-temporal .dynamics prior to neocortical seizures: Amplitude ver-sus phase couplings," IEEE Transactions on Biomedical Engineering, vol. 50, no. 5, pp. 571{583, 2003.

[36] K. Lehnertz and C. E. Elger, \Can epileptic seizures be predicted? Ev-idence from nonlinear. time series analysis of brain electrical activity," Physical Review Letters, vol. 80, no. 22, p. 5019, 1998.

[37] C. E. Elger and K. Lehnertz, \Seizure prediction by non-linear time series analysis of brain electrical activity," European Journal of Neuro-science, vol. 10, no. 2, pp. 78.6{789, 1998.

[38] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D.-S. Shiau, \Quadratic binary program.ming and dynamical system approach to de-termine the predictability of epileptic seizures," Journal of Combinato-rial Optimization, vol. 5, no. 1, pp. 9{26, 2001.

[39] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry,
P. Pennell, C. Epstein, R. Bakay, M. Dichter et al., \Epileptic seizures may begin hours in advance of clinical onset: A report of ve patients," Neuron, vol. 30, no. 1, pp. 51{64, 2001.

[40] S. Gigola, F. Ortiz, C. Dattellis, W. Silva, and S. Kochen, \Prediction of .epileptic seizures using accumulated energy in a multiresolution frame-work," Journal of Neuroscience Methods, vol. 138, no. 1, pp. 107{111, 2004.