II. Review of Literature

Automated Seizure Detection using EEG has been deeply investigated for its technical and commercial feasibility thereby creating a revolution in healthcare sector. “The Indian Epilepsy Association (IEA) was registered in December 1971 as a Public Charity Trust with a mission to raise epilepsy awareness, increase in acceptance of persons with epilepsy and provide relief and rehabilitation to patients and their families. It started with 16 chapters, Bombay is one of them, and today there are about 22 chapters all over India with a total membership of 1597” (http://epilepsysupport.aarogya.com/help-and-information/ indian-epilepsy-association.html). Raising awareness and reducing stigma is necessary to improve the quality of life of people with epilepsy. According to The Hindu a daily newspaper of India “Epilepsy affects about 65 million people worldwide and in India six persons in every 1000 population is affected by epilepsy” (http://www.thehindu.com/news/cities/Tiruchirapalli/nursing-students-step-up-epilepsy-awareness/article4546826.ece). Social prohibition and misconceptions were major impediments in treating epilepsy patients. It is essential that people are sensitized and educated on epilepsy. Recorders & Medicare Systems Pvt Ltd (RMS) is the leading manufacturer of medical equipments in the Indian Subcontinent with the inception of EEG in the year 1977. Amidst the development of assorted instruments for brain research, prevalence of EEG in one of the leading positions for several decades is an amazing reality. The reason for its sustenance is higher temporal resolution, low cost of technology and its feasibility of combining it with advanced tomography techniques. This chapter reviews some EEG based techniques used in detecting seizures.

2.1. EVOLUTION OF EEG: FROM 1929 – 2013

Physiology is the study of the function of the body and its parts (http://quizlet.com/dictionary/physiology/) . Electro-physiology is the study of the electrical properties of biological cells and tissues. The goal of electro physiological recording is to detect the communication signals between neurons in real time. The most important law which forms the basis in Electrophysiology is Ohm's law. The potential difference between two points linked by a current path with a conductance G and current I is indicated as $\Delta V = IR = I/G$ (units: volts). One of the pure electro-physiological equipments are EEG which is used to acquire data pursued by its interpretation. EEG consistently presumes the state of the art technology since 1929. Predominance of several cutting edge technologies still could not rescind the primordial EEG's subsistence. Its development was a groundbreaking work with its existence being pivotal in the field of science and technology. Dr. Hans Berger (1929), was the first scientist to record human EEG by using a 30 mm/s paper speed which henceforth has become the standard.

2.1.1 About EEG

Excavating the roots of EEG as stated by Saeid Sanei and J.A. Chambers (2007) in their book "EEG Signal Processing" some of the information about the reclining history of EEG are discussed. Subsequently David Millet (2001) has explored about EEG in his work "Hans Berger: From Psychic Energy to the EEG" which has also been considered as a basis for this section. The primitive hypothesis that electrical signals were emitted from muscle nerves was corroborated by Carlo Matteucci (1811-1868) and Emil Du Bois-Reymond (1818-1896) who used a galvanometer and thereby establishing a foundation of neurophysiology. The basic use of the galvanometer is to measure direct electric current as it flows to and fro from source. Figure 2.1 represents the first galvanometer. Later a scientist, Richard Caton (1842-1926) from Liverpool further complemented to the above work by using a galvanometer and placing two electrodes on the scalp thereby recording spontaneous electrical rhythms of the mammalian brain 1875.The term Electroencephalogram was derived from the concepts of

Electro- referring to registration of electrical activities of the brain

Encephalo- referring to the emission of signals from the head

Gram (or graphy) - referring to drawing or writing

Thus the three terms were combined to frame Electroencephalogram to denote the electrical and neural activity of the brain.

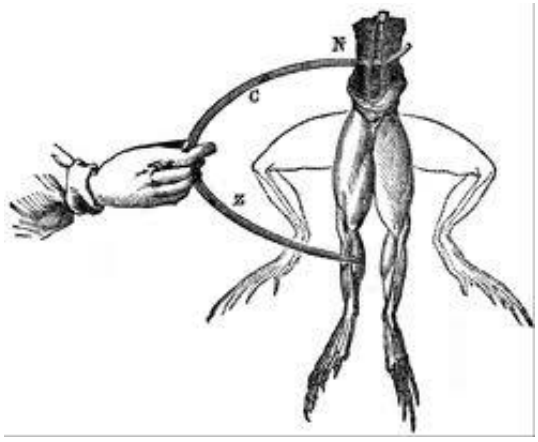

Figure 2.1 First galvanometer

Figure 2.2 Recording brain activity of first patient

Fritsch (1838-1927) along with Hitzig (1828-1907) ascertained that the human cerebrum could be electrically stimulated. Vasili Yakovlevich Danilevsky (1852-1939) investigated the activity of the brain following the electrical stimulation. The evidence of Epileptic seizure in a dog caused by electrical stimulation was investigated by Napoleon Cybulski (1854-1919). Kaufman investigated the association of epileptic attacks with abnormal electrical discharges.

Hans Berger (1873-1941) marked a great record in the history of encephalographers by discovering the existence of human EEGs .His experimentation adopted a string galvanometer (1910) a model of primordial type, and later shifted to a smaller Edelmann model and shortly transiting to a big Edelmann model in 1924.He used a much more powerful galvanometer in 1926 depicted in Figure 2.2 recording brain activity of first patient. He placed annals in 1929 by recording human EEG on a photographic paper for a duration of one to three minutes by using a one-channel bipolar method with fronto-occipital leads. He was the first to discover the alpha rhythm and indicated it to be the main component of EEG. He was engrossed in the study of cerebral localization and associated himself in localization of brain tumors. He explored the correlation between mental activities and alterations in EEG signals. He was the first to investigate on the recording of sleep spindles i.e. epitome of EEG synchronization at sleep in 1930.He was also the one to report on the effect of hypoxia i.e. scanty supply of oxygen to the brain. Hans Berger the inventor of EEG with his icon is shown in figure 2.3.

Figure 2.3 Father of EEG-Hans Berger

Progression in EEG found its path towards the invention of biological amplifier to record brain potentials by Toennies, a group from Berlin. This was followed by the development of differential amplifier by the Rockefeller foundation in 1932. In order to record a broad area of the brain, a mono channel EEG recording was not sufficient and hence the pavement of multichannel recordings was recognized by Kornmuller. Epileptic manifestations and revelation on epileptic spikes were first presented by Fischer and Lowenbach. W. Gray Walter was the pioneer of clinical electroencephalography and he discovered that delta waves were the signals of interest as they attributed to the diagnosis of brain abnormalities. Loomis *et al*. (1937) were the first who mathematically studied the human sleep EEG patterns. Epileptology historically reigned over two periods: before and after the advent of EEG. Berger showed a few examples of paroxysmal EEG discharge in case of presumed petit mal seizures.

The American EEG society was first founded in 1947 and the first International EEG Congress was held in London. Work on EEGs expanded throughout the decade of 1950s and during this era surgical operation for curtailing epileptic convulsion attained its popularity. Microelectrodes prevailed and were used for surgery. Implanted intracerebral electrodes were invented by Mayer and Hayne during the year 1948. A step ahead in technology was the invention of intracellular microelectrode which created a revolution in the era of biomedical technology by implanting the electrode in the spinal cord instead of the brain.

Berger assisted by Dietch applied Fourier analysis to EEG sequences. A scrutiny of sleep disorders with EEG commenced during 1950s .In 1960s work on analysis of neonatal seizures were developed. EEG was the pioneer device to register brain signals .Although numerous neuroimaging devices have arrived, the existence of EEG has become the quintessence, owing to its remarkable advantages. None of the novice technology (fMRI) could afford to reinstate EEG but have the option of integrating them with EEG in order to ensure a higher degree of certainty in predicting certain neurological inconvenience.

2.2 PURSUIT OF AUTOMATED EEG ANALYSIS:

Paper based conventional systems abounded in India, until the emergence of commercial computerized EEG systems a few years ago. The contemporary clinical Electroencephalographers regarded computerization to be a great boon**.**

Historically the prominence in the development of EEG is mainly due to its incorporation of computers. The reason behind the energizing force of computerization is due to the demands of EEG applications, which led to signal acquisition, analysis and display of EEG data. Computation with respect to EEG signals is an alteration in the content or display of electroencephalographic signals.

Thomas F. Collura (1993) in his chronological presentation “History and Evolution of Electroencephalographic Instruments and Techniques” has enlightened about Automated EEG analysis that are discussed in the current section. Carlo Matteucci, secured a place in history in 1942 by making use of the galvanoscopic frog leg which induced a twitch, exploring the concepts of nonlinear transduction, spatial and temporal integration and thresholding. Nevertheless, he introduced the concept of Electrophysiology which provided a gateway to computerization. The first researcher to record EEG in humans was Hans Berger, who along with a Physicist, G. Dietsch worked on two aspects namely,

i) Resistance measurements of skull
ii) Calculating the frequency spectrum of the EEG using Fourier transform

Alfred Loomis designed a sleep monitoring system with several enhancements in EEG waveform processing. He also designed an electronic integrator which calculated and recorded the amount of activity in a particular frequency band by flashing the lights more often, which indicated increased activity in the particular band. The flashing of lights was automatically counted and recorded with drum.

The first frequency analyzer introduced by Grass et al. (1938) was an ingenious combination of mechanical and electrical technology. Using this technology, reports on the frequency spectrum of EEG during epileptic seizures were produced. Subsequent work by other investigators elaborated on the frequency analyzers. Introduction to the frequency band theta by Gary Walter, who later invented another design based wave analyzer called the toposcope, which portrayed 24 channels of EEG representing the frequency, phase and time relationships between each channel. Another frequency analyzer constructed by John Knott comprised of six channels by adopting phase shift oscillators and its output stored using capacitors.

A cross-correlation analyzer was reported by C. W. Goodwin in 1948. Before the advent of an absolute digital system for EEG analysis, a variety of hybrid systems were launched by researchers. Signals were worked out in time, amplitude and the applications were considered in averaging, correlation and frequency analysis. An evoked potential system developed by Dr. George D. Watson in 1947 was an important forerunner in the development of digital EEG processing. He used a mechanical commutator to sample time points and stored EEG sweeps by using capacitors. It was a partially digital device, as time values were discrete but amplitudes were not. Dr. Manfred fabricated the first commercial digital analyzer which could perform Analog to Digital conversion. Computers of average transient systems were acquired and put to use with EEG. The evolution paved path for general purpose processors.

2.2.1 Experimental Trends in Computerized EEG

A substantial amount of effort was putforth in the analysis of automated EEG in 1950 at Massachusetts Institute of Technology (MIT). Every mathematical and statistical technique in existence has been applied to EEG waveforms. The other techniques that conform to their participation were non-linear dynamic analysis, neural networks, fuzzy logic, and machine learning etc., . Erudite presentation by Thomas F. Collura (1995) in his “History and Evolution of computerized Electroencephalography” contributed greatly towards computerized EEG analysis that was helpful for discussion in this section. The automated analyses of human EEG with an analog correlator in 1955 used a magnetic drum to record and play back EEG signals. The Main concern of the system was to produce autocorrelations and cross-correlations of EEG waveforms.

Subsequently, the production of first A/D convertor named, Average Response Computer (ARC) was reported during 1958, which was transistorized to store EEG in digital memory. During the 1960s and 1970s much of the work was targeted at automated detection of abnormalities particularly those encountered by epilepsy. Gotman *et al*. (1976) contributed towards long-term monitoring of abundant data in EEG.

The revolution of computerized EEG with the introduction of microprocessor based system for Electromyography (EMG) in 1978 was instigated by Dr John Cadwell. The primary impetus was to eradicate the complexity of the product. Microprocessor based systems had its own disadvantages i.e. the acquired signals could not be seen at real time. A subsequent product introduced in 1984, was a 16 channel Cadwell 8400 which focused on brain mapping. This invention contained an 8 bit microprocessor, a colour display, printer supporting a set of EEG, EMG protocol, but the major setback was the complexity of the software and the speed at which it operated.

In 1984, a mapping unit based on IBM-compatible PC architecture was introduced by Biologics Corporation, which was a 21 channel system designed to provide topographical maps and analysis of EEG frequency. The system could perform FFT and provided artifact rejection, cursor-based waveform analysis and could simultaneously display up to 12 different maps. Eventually the digital EEG systems approached the real time response and user characteristics of traditional analog based systems.

Fisher *et al*. (1992) were the first to give a promising solution of computerized EEG by recording a low amplitude EEG between 80-120 Hz. The recordings which were revealed in digital EEG printouts and frequency plots exhibited activities which were not apparent using either conventional analog

EEG equipment or digital EEG. Though the clinical value of the activities weren’t firmly established, this technique recorded a significant position in the pace of computerization.

With the advent of digital EEG systems, explorations of new capabilities have come to the forefront. The concern for storage of voluminous data in order to suit long term monitoring and remote access to records was crucial for the enhancement of the technology. Recent developments in computerized monitoring of EEG have addressed to the issues of system design, software, and multi user architectures, which have led to the domination of software, data acquisition unit and computer hardware readily available at an economical price.

The revolution in neuroimaging brought about computer-aided tomography (CAT), magnetic resonance imaging (MRI), functional MRI (fMRI) and positron emission tomography (PET). All these approaches are far-reaching, but ousting EEG is a still a stalemate.

2.3 ELICITATIONS ON ARTIFACTS AND THEIR ATTENUATION

Denoising of Electroencephalographic signals is a demanding preprocessing step former to qualitative EEG analysis. EEG is one of the different modalities of recording signals of cerebral origin, but attempts to trace electrical activities arising from other locations too. These signals which are not of cerebral origin pose a serious impediment in investigating the disorders of the brain and are termed to be artifacts. The recorded activity other than cerebral basis are categorized broadly into physiologic or (intrinsic artifact) and extraphysiologic or (extrinsic Artifact). Eyeball movement, Eye blinks, cardiac signals, muscle noise are certain examples of physiological artifact, as their nativity springs from the patient which is not of cerebral origin. Artifacts which arise from the environment, equipments augment to extraphysiological artifacts e.g line noise, electrode pop-up. Pervasive ocular artifacts encompassing eyeball movement and eye blink, pose serious tribulations in interpreting and analyzing EEG data. Myogenic or Muscular artifacts arising due to frowning, jaw clenching, and spit swallowing also induce perturbations to the EEG signals. These types of artifacts contaminate EEG signals and present serious problems of unacceptable data loss. Denoising of exacerbating artifact signals are the prime concerns of Signal processing. Different methods are carefully reviewed with their affirmative doctrine and shortcomings in order to propose an appropriate methodology. This section discusses some of the work proposed by other authors.

Numerous algorithms and techniques were projected and applied to remove Artifacts present in EEG signals. The signal processing algorithms laid its foundation by adopting Fourier Transform for artifact removal. P.F Prior *et al.* (1973) observed an augmentation of amplitude followed by sudden decline. Babb *et al*. (1974) used intracranial electrodes to design an electronic circuit using seizure detection.

2.3.1 Different methods of handling Artifacts

This section briefly reviews various contemporary artifact removal techniques, which eliminates the artifacts by keeping the neurological phenomenon intact.

2.3.1.1 Regression based Artifact removal

In regression based approaches calibration trials are first conducted to determine regression coefficients and these coefficients are used in the correction phase to estimate the EOG components in the EEG recording which are later removed by subtraction. Regression based approaches for removal of ocular artifacts are discussed below.

Verleger *et al*. (1982) corrected EOG artifacts using a regression based approach which evaluated the transmission rates for eight EEG channels in 67 subjects and proved that EOG artifact correction based method is superior to the conventional rejection technique in terms of reducing the correlation between EOG and EEG.

Gratton *et al*. (1983) used an off-line procedure for dealing with ocular artifacts. The system used here estimates a propagation factor which describes the relationship between the EOG and EEG traces. Different propagation factors were computed for blinks and eye ball movements. Potentials derived from trials and corrected by this procedure seemed to be are more similar to a true potential. The major advantage of the procedure was that it permits retention of all trials in an experiment, irrespective of ocular artifact. Thus, studies of populations characterized by a high degree of artifact, and those requiring eye movements as part of the experimental task are made possible. Furthermore, there is no need to require subjects to restrict eye movement activity. It also has the advantage that separate correction factors are computed for blinks and movements and that these factors are based on data from the experimental session itself rather than from a separate calibration session.

The method implemented by Woestenburg *et al*. (1983) was based on complex regression analysis. The eye-movement activity is reliant on the frequency dependent amplitude and phase characteristics and the regression formula is used in the frequency domain. The method is demonstrated with artificial signal-in-noise EOG and EEG series. A 24 msec time-shift of the EOG was simulated due to the frequency dependent phase characteristics. The complex regression coefficient of the EOG and EEG series was calculated and then common regression was removed from the EEG. The estimated regression coefficient was tested for significance with the F-statistic and significant EOG activity was subtracted in the frequency domain from the EEG. Later each EEG record was inversely transformed to the time domain.

Elbert *et al*. (1985) described the propagation of ocular potentials across the scalp on a biophysical basis. He incorporated 3 EOG derivations that are generally necessary to account for ocular disturbances in the EEG. The inadequacy of using 3 derivations suggested that just one EOG derivation provides enough information to remove ocular potentials from any EEG recording. He also demonstrated that the frequency dependence of the ocular influence cannot be neglected, if fast and slow EOG activities have to be removed. Lins *et al*. (1993b) suggested a low pass filter to EOG artifacts and used this as a basis to compute regression co-efficient. DiMatteo *et al*. (2001); Wallstrom *et al*. (2002) conceptualized the use of nonlinear filter to remove high frequency activity when the amplitude fluctuations are small and retain high frequency activity when the amplitude fluctuations are large. Implementing such an adaptively filtered EOG essentially segregates cerebral activity associated with ocular artifacts.

2.3.1.2 Artifact removal using adaptive filters

Filters in general can be classified broadly into two categories, fixed and adaptive. The optimum choice of a fixed filter is based on the criterion that knowledge of both the signal as well as the noise should be known in advance. On the other hand, adaptive filters have the capability to adjust their own parameters automatically and do not require any priori information. Noise cancellation makes use of reference input, which is subtracted from the primary input signal resulting in an attenuation of noise signals.

He *et al*. (2004) described a method for removing ocular artifact specially from frontal channels based on adaptive filtering by using vertical EOG and horizontal EOG as two reference inputs. Each of the reference input was first processed by a Finite Impulse Response (FIR) filter of length M and subtracted from the original EEG signal. In order to track the non-stationary portion of EOG signals, a Recursive Least Square (RLS) algorithm is used with a forgetting factor λ=. 9999.

Garces Correa *et al.* (2007) adopted a method to remove artifacts by implementing a cascade of three adaptive filters. The rationale behind the usage of three adaptive filters in cascade was to eliminate line interference, ECG artifacts and EOG artifacts respectively. In each stage a Finite Impulse Response (FIR) filter was implemented which adjusted it’s coefficients to produce an output similar to the artifacts present in the EEG and the working was based on a

Least Mean Squares (LMS) algorithm. In all cases, line-frequency, ECG and EOG artifacts were attenuated.

Senthil Kumar et. al (2009) implemented a similar method to remove ocular artifacts based on Recursive Least Square (RLS). Ashok Babu et al.(2011) designed specific filters to remove the artifacts in EEG ,which adopted an adaptive filtering method that uses RLS (Recursive Least Square) algorithm and FRLS (Fast Recursive Least Squares) to remove ocular artifacts from EEG recordings through wavelet transform. RLS & FRLS algorithms were compared with wavelet transforms. Elapsed time was decreased by using the FRLS algorithm compared to other techniques.

2.3.1.3 Blind Source Separation (BSS)

BSS is a method of recovering stochastic independent signals from a linear mixture of signals. The term "blind" implies that neither the mixing co-efficient nor the probability distributions of the original source signals are known. This problem is more prominent in the field of biological signal processing i.e. EEG, MEG, ECG. Jutten and Heralt (1991) were the pioneers to propose a heuristic learning algorithm to solve this problem . They observed a poor performance when the number of sources exceeded two. This open problem was resolved by many researches which included the proposal of ICA (Independent Component Analysis) proposed by Comon (1994), PCA (Principal Component Analysis) by Oja *et al*. (1995), entropy maximization by Bell and Sejnowski (1995) and natural gradient approach by Amari *et al*. (1995).

2.3.1.3.1 Independent Component Analysis (ICA)

ICA is solely a method of signal and data analysis which tries to recover independent source signals s_i (t) (i=1.. N) that are statistically independent at each time step t and linearly mixed with an unknown matrix A, and their mixtures x_i (t) (i=1.. N) =As, are also not known. Information about the sources or about the mixing process is unknown except for the information that there are N recorded mixtures. A tantamount version is to recover u=Wx, of original sources s, by finding a square matrix W.

Joyce *et al*. (2004) proposed an automated ICA based method for artifact removal. This algorithm can isolate correlated ocular components with a higher degree of accuracy. The approach was also suitable to other sources of contamination like cardiac, line interference, myogenic interference etc., thus extending the scalability.

Aurthur Flexer *et al*. (2005) proposed a method of recording EEG from blind subjects, which posed special problems since it showed a higher quantity of eye movements which were all the more irregular .It was further verified that ICA stages an astonishing performance by comparing results of four blind subjects with results from one subject without eye bulbs which did not show eye movement artifacts at all.

Krishnaveni *et al*. (2005) proposed a method by applying Independent Component Analysis (ICA) to multichannel EEG recordings which could remove a wide range of artifacts from EEG recordings by eliminating the perturbations arising from the scalp. The estimated sources should be as independent as possible, for better removal of artifacts from EEG. Various ICA algorithms like OGWE, MS-ICA, SHIBBS, and Kernel-ICA, JADE and RADICAL are assessed and individual components obtained are compared to Mutual Information (MI) estimator based on k-neighbor statistics without using the probability density functions. The recommendation is for RADICAL as it performs best at separating the source signals from the observed (mixed) EEG signals.

Le van *et al*. (2006) suggested a system for automatic artifact removal in scalp EEG based on independent component analysis which demonstrated the potential not only to de-correlate but also to work with higher-order dependencies. A Bayesian classifier was trained using numerous statistical, spectral, and spatial features.

Arnaud *et al*. (2007) presented a robust ICA technique based on higher order statistics rather than circumventing to degree two, to detect ocular as well as muscular artifacts. The three popular ICA algorithms used for comparison were Infomax, SOBI and FastICA . The scheme can have useful applications in the area of medical imaging. Invariant features were used as a watermark and the extracted features were selected for embedding purpose. Root Mean Square Error (RMSE) was used as a similarity measure and this technique had the ability to resist geometric distortions.

2.3.1.3.2 Principle Component Analysis (PCA)

PCA is a mathematical phenomenon that applies an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linear uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. The goal of PCA is to obtain principal components which are uncorrelated. Moreover, PCA gives projections of the data in the direction of the maximum variance. The Principal Components (PCs) are ordered in terms of their variances: the first PC defines the direction that captures the maximum variance possible; the second PC defines the variance (in the remaining orthogonal subspace). PCA algorithms use only second order statistical information. PCA cannot separate all the ocular artifacts from the EEG signals, especially when they have comparable amplitudes.

Silvia *et al*. (2004) implemented a method of recording EEG from 10 locations using the 10 – 20 system. Diagonal EOG from the right eye was also recorded. PCA was applied in order to reduce ocular artifacts and the logic used was that the first or the second PC was subtracted when the correlation coefficient between the component and EOG was greater or equal to 0.9 or 0.95, respectively. The performance of the method was tested on simulated and real data, varying EOG amplitude and artifact transmission characteristics.

Lagerlund *et al*. (1997) explored the performance of PCA by Singular Value Decomposition (SVD) in order to analyze an epoch of a multichannel

Electroencephalogram (EEG) which resulted in multiple linearly independent components with each of the components temporally and spatially de- correlated. A new EEG can be reconstructed, by a linear combination of component waveforms .The variation developed with this technique is that the factors that reconstruct the modified EEG from the original are stored as a matrix. This matrix is applied successively to multichannel EEG to create a new EEG continuously in real time, without performing the time-consuming SVD. This matrix acts as a spatial filter and is applied in this method to remove artifacts, including ocular movement and electrocardiographic artifacts. Both the ocular and myogenic artifacts were considered but certain limitations were observed as the system was unable to completely separate the artifacts from the cerebral activities when the amplitude was analogous.

2.3.1.4 Discrete Wavelet Transform

The Wavelet Transform (WT) has gained widespread acceptance in the field of signal processing. A transformation of a signal is just it's representation in another form, while preserving its information intact. Wavelet transformation provides the time frequency representation of the signal. The main purpose of it's advent was to surpass the shortcomings of Short Time Fourier Transform (STFT). The conception of multi-resolution has been adopted, by which different frequencies are analyzed with different resolution as opposed to STFT, which utilizes constant frequency at all the frequencies. Wavelets have their energy concentrated in time or space and are apt for transient signals.

A denoising technique proposed by Tatjana Zikov *et al*. (2002) provides a solution for the removal of ocular artifacts in EEG. This technique does not rely upon reference EOG or visual inspection, but is based on the threshold limit estimated from uncontaminated baseline EEG which is recorded from the same subject.

Krishnaveni *et al*. (2006) projected a work on eliminating ocular artifact by automatically identifying the artifact zones. She proposed a method to automatically identify Ocular Artifact (OA) zones by applying thresholding to OA zones, thereby preserving the significant information intact.

Senthil Kumar *et al*. (2008) proposed a statistical method for removing ocular artifacts from EEG recordings through wavelet transform without using an EOG reference channel. Inorder to de-noise the corrupted EEG signals, SWT along with Donoho's method has been employed in their work. A statistical empirical formula was used for calculating the thresholding limits.

Salwani *et al*. (2005) projected a technique to remove ocular artifacts using Lifting Wavelet Transform (LWT) a technique to remove ocular artifact. A comparative study was performed to choose the best of the wavelets; and the wavelets under consideration were haar, reverse biorthogonal, daubechies 4 and cdf4 (cohen, daubechies and feauveau). Lifting operations were performed on these wavelets and a level dependent threshold was chosen to remove the ocular artifacts.

2.4. Feature Extraction

Tremendous progression in the extraction of quantitative features of an EEG signal was introduced more than 70 years ago. Feature Extraction is a process which transforms the raw signal to a relevant data structure, called feature vector. It can also imply a dimensionality reduction, which eliminates redundant data from the feature vector. Three categories of features are predominant in EEG system. They are time domain, frequency domain and time frequency domain.

2.4.1 Time domain features

It is the analysis of mathematical functions or signals with respect to time. Time domain EEG features can be categorized as depicted below:

Statistical features

Hjorth's descriptors

Non –linear features

2.4.1.1 Statistical features

Non-stationary nature of the signals can be quantified by measuring the statistics of the signals. Some of the statistical features under the contemplation of EEG signals are mean, Standard deviation, maximum peak value, skewness, kurtosis etc. However, the signals could be considered to be stationary if there is no substantial variation in statistics. A study based on analysis of seizure prediction performed by Mormann *et al*. (2007) elucidated that the statistical analysis were carried out by Lehnertz *et al*. (1998), Litt *et al.* (2002) and Le van Quyen *et al*. (2001), Drury *et al*. (2003) and Harrison *et al*. (2005).

Mcsharry *et al*. (2002) adopted linear and nonlinear methods for automatic seizure detection in EEG recordings. A nonlinear statistical technique Multidimensional Probability Evolution (MDPE) based on the time evolution of the probability density function within a multi-dimensional state space, and a linear statistical technique variance were used to detect seizures.

D'Alessandro *et al*. (2003) has developed a system for epileptic seizure prediction by using a hybrid feature selection by placing multiple Intracranial EEG electrodes. The approach applied an intelligent genetic process to EEG signals and multiple first level quantitative features derived from those signals. The second and third level features used are minimum, maximum, mean, variance, standard deviation, skewness, kurtosis, slope, integral, derivative and sum. It was predominant that all potential second-level features were considered in the research for developing the objective feature vector.

2.4.1.2 Hjorth's descriptors

It is one of the preferred features by neurologists in assessing EEG activity and characterizes the EEG signal in terms of amplitude, time scale and complexity. It's parameters are activity, mobility and complexity which depend on the variance, 1st order derivative and 2nd order derivative of standard deviation of the signal respectively.

Forrest Sheng bao (2008) adopted a new approach to automated Epileptic Diagnosis of EEG using Probabilistic Neural Network (PNN). The feature extraction inherited was Hjorth's descriptor with mobility and complexity. As Hjorth complexity appears very inconsistent among classes, since PNN uses normalized features, normalized Hjorth complexity was calculated and it was confirmed that a consistent distribution exists within each class. An interesting finding from his work was that a set with low Hjorth mobility had a high normalized Hjorth complexity. He further extended his paper in (2010) by developing an open source Python module for EEG Feature Extraction.

Melvin Ayala (2009) in her research proposed a platform for applications of artificial neural networks in pediatric Epilepsy. The activity parameter was extracted out of 1 second window for each electrode. The study capitulates that the Hjorth's parameter activity was adequate to relate EEG to epileptic and non-epileptic subjects. It outperformed the other time domain and frequency domain features.

Temko *et al*. (2011) offered an SVM based multichannel, patient independent neonatal seizure detection system. Hjorth parameter was one among the Time domain features presented in his study.

2.4.1.3 Non-Linear features

A non linear system is one whose output is not directly proportional to its input. These systems are of great interest to the engineers, physicists and mathematicians as most of these systems are inherently nonlinear in nature. They are utmost difficult to solve but give rise to an interesting phenomenon such as chaos. These systems are not random in nature. Some of the important non – linear parameters are discussed subsequently.

2.4.1.3.1 Correlation dimension (CD)

Correlation dimension is a measure of dimensionality of space occupied by a set of random points is referred to be a type of fractal dimension. It can be calculated using the distances between each pair of points in the set of N number of points. CD was used in different studies relented by ((Frank *et al*., 1990), (Iasemidis and Sackellares 1996) and (Guevara 1997))

2.4.1.3.2 Largest Lyapunov exponent (LLE)

It is another significant non-linear parameter which measures how fast signal trajectories separate from each other**.** They characterize the average rate of divergence of these neighboring trajectories. A negative exponent implies that the orbits approach a common fixed point. A zero exponent means the orbits maintain their relative positions; they are on a stable attractor. Finally, a positive exponent implies that the orbits are on a chaotic attractor. This analysis yielded by researchers in EEG analysis such as (Iasemidis and Sackellares 1991, Pradhan and Sadasivan 1996, Osowski *et al*. 2007) and Seizure detection (Moser *et al*. 1999, Iasemidis and Sackellares 2001)

2.4.1.3.3 Entropy (En)

Entropy is a thermodynamic quantity describing the amount of disorder in the system. From an information theory perspective, the above concept of entropy is generalized as the amount of information stored in a more general probability distribution. Different entropy estimators exist which have been applied to EEG data to quantify the complexity of the EEG signal. These techniques do not describe how the EEG signals changes with time. Intuitively it seems that, if an EEG signal is irregular, the position of a particular point will not be easily predicted, whereas in a regular signal the position of the point will be more reliably predicted.

2.3.1.3.4 Hurst Exponent (HE)

The Hurst Exponent is a measure that has been widely used to evaluate the self-similarity and correlation properties of the time series produced by a fractional Gaussian process. Hurst exponent is used to evaluate the presence or absence of long-range dependence and its degree in a time series.

Natrajan *et al*. (2004) demonstrated his work by extracting nonlinear parameters like Correlation Dimension (CD), Largest Lyapunov Exponent (LLE), Hurst Exponent (HE) and Approximate Entropy (ApEn) which are further used in evaluating the EEG signal

Kannathal *et al*. (2005) proposed a system to detect epileptic seizures in EEG signals which appeared to be random. Chaotic measures like CD, LLE, HE and Entropy were used to characterize the EEG signals. These non linear measures were good discriminators of normal and epileptic signals. She extended her work using entropy measures for feature extraction along with Adaptive Neuro Fuzzy Inference System (ANFIS) . Different entropy estimators were used to discriminate normal and epileptic data. The classification ability of entropy was investigated using ANFIS classifier.

Guler *et al*. (2005) used Lyapunov Exponents (LE) in an amalgamation with Recurrent Neural Network (RNN) . Ubeyli unmitigated the above work, to evaluate the diagnostic accuracy of RNN by employing LE trained with the Levenberg Marquardt algorithm. He observed that the proposed RNN employing LE could be very much useful in the long term analysis of EEG. Further Guler and Ubeyli (2007) performed Wavelet transform and Chaos analysis using Support Vector Machines (SVM). The dynamics of the EEG signals are quantified using LLE representing system chaocity and CD representing system complexity. Significant differences were observed when both these parameters were used in conjunction within specific EEG subbands. CD differentiated the higher frequency sub bands i.e. beta and gamma and LLE discriminated the lower frequency sub band.

Srinivasan *et al*. (2007) proposed a neural network based epileptic EEG detection system that uses Approximate Entropy (ApEn) as an input feature. ApEn, being a statistical parameter indicates that predictability of current amplitude values highly depend upon previous amplitude values. It could be noted that the value of ApEn drops sharply during an epileptic seizure. ApEn was proposed for the first time in detecting seizures and was proved to achieve higher accuracy rates.

Balli *et al*. (2010) investigated the characterization ability of linear and non linear features. The characterization ability of non linear features presented in his proposal were CD, LLE, ApEn, Hurst Exponent, non linear prediction error , higher order auto covariance and asymmetry which are further compared to Auto Regressive (AR) reflection coefficients and AR model coefficients. Overall results suggested that a combination of linear and non linear features proved to be a better approach for characterization and classification.

A time series complexity analysis of Electroencephalographic signals was introduced using Sample Entropy (SampEn). This statistical parameter quantifies the regularity in the signals which are purely deterministic or stochastic in nature. Yuedong Song *et al*. (2010) proposed a new method for epileptic seizure detection using SampEn, the only input feature and Extreme Learning Machine (ELM) for classification purposes. The results demonstrated that an accuracy as high as 95.5% was achieved through a combination of SampEn and ELM at a faster rate. Yuedong Song *et al*. (2012) further extended his work by presenting another novel method for Seizure detection using Optimized Sample Entropy (O-SampEn) in combination with ELM. The results illustrated that not only a high accuracy rate was achievable but also fast computation was also possible

Bai D (2007) used sample entropy for EEG based epilepsy detection. He analyzed the method to have higher precision rate than ApEn. Data analysis results show that when there is a burst in epilepsy the values of both ApEn and SampEn significantly decrease. Furthermore, the SampEn is more sensitive to EEG changes caused by the epilepsy, about 15%-20% higher than the results of the ApEn.

2.4.2 Frequency domain features

Frequency domain is the analysis of mathematical functions or signals with respect to frequency. A digitized time series signal can be represented as a superimposition of different frequencies in other words a signal is broken down into different spectral components. These features are characterized by the power of the brain in several frequency bands. Fourier transforms are the cornerstone for the analysis of signals in the frequency domain.

2.4.2.1 Fast Fourier Transforms (FFT)

FFT is the most commonly used transformation technique for analyzing the spectrum content of any deterministic biosignal.

Polat and Gunes (2007) aimed to detect seizures in EEG signals using a system based on Fast Fourier Transform (FFT) and Decision tree. The implementation was carried out by performing feature extraction using FFT and decision making with decision classifier. The progression of the proposed system was accomplished using k-fold cross-validation, classification accuracy, sensitivity and specificity. Experimental results showed that the proposed method was successful in the design of a new intelligent diagnostic system.

2.4.2.2 Walsh Hadamard Transform

A transformation similar to FT is the implementation of Walsh-Hadamard transform, which follows a divide and conquer algorithm to simplify the computations.

Azadeh Bastani *et al*. (2011) in their study had implemented Walsh-Hadamard Transform for feature extraction. Examining that most of the study carried out with respect to EEG signals were based on Fourier transform, he compared the efficiency of WHT with FT and proved that WHT was better than FT. Walsh Hadamard transform was ideal compared to Fourier spectral features due to its simplicity and swiftness in calculation.

2.4.2.3 Power Spectral Density (PSD)

A power spectrum describes the energy distribution of time series with respect to frequency domain. PSD estimation methods are broadly classified as follows

Parametric methods: These methods are model based methods and require model parameters of time series for estimating the PSD. The most popular parametric based models are Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), Multiple signal classification (MUSIC) models. An appropriate model should be built first so that it correctly reflects the behaviour of the system that generates the time series, so as to enable a reliable PSD estimation.

Nonparametric methods: These methods are based on a computation that uses the concept of data windowing. Examples of these methods include periodogram, Welch's method, Capon method which are all based on calculations of DFT. Distortions can emanate due to windowing and these methods are more robust than parametric methods . But if the model appropriately fits the signal than parametric methods are the best.

2.4.2.3.1 Parametric Methods

One of the most common parametric methods for power spectral density estimation is performed using Auto Regressive Coefficients. A number of techniques exist for computing AR coefficients. The main two categories are Least Squares and Burg method.

Palaniappan *et al*. (2000) considered the advantages of autoregressive (AR) modeling over the classical Fourier Transform methods and proposed a method for spectral analysis of EEG signals. Using the spectral values, the neural network is trained and classification is done accordingly.

Guler *et al*. (2001) analyzed the EEG signals using AR method. Parameters in AR method were realized by using Maximum Likelihood Estimation (MLE). Results were compared with Fast Fourier Transform (FFT) method and were observed that the AR method performed well in the analysis of EEG signals. They proved that AR method would be successful in the diagnosis of diseases.

Abdulhamit Subasi (2005) dedicated his major contributions to the research of Epilepsy. He incorporated Wavelet Neural Network (WNN) with an AR model for detecting epileptic signals. Logistic Regression (LR) as well as Feed forward Error Back Propagation Artificial Neural Networks (FEBANN) and Wavelet Neural Networks (WNN) based classifiers were developed and compared for accuracy in classification of EEG signals. FFT and AR model by using MLE were input to the classification system. A reliable classifier was obtained by applying AR with MLE in connection with WNN architecture. The comparative studies between the classifiers were primarily based on the analysis of the Receiver Operating Characteristic (ROC) curves. The WNN-based classifier outperformed the FEBANN and LR based counterpart.

An extension of the above work was carried out by Abdulhamit Subasi (2007). EEG signals were processed using AR method and PSD was acquired. The parameters of autoregressive method were estimated by different methods such as Yule-Walker, covariance, modified covariance, Burg, least squares, and MLE. The Cramer-Rao Bounds (CRB) were derived from the estimated AR parameters of the EEG signals and the performance evaluation using these values were carried out. Finally, the EEG signals were selected according to the computed CRB values. According to the computed CRB values, the optimal AR spectral estimation method outperformed the other methods in the EEG epileptic signal analysis.

2.4.2.3.2 Non-Parametric methods

The power spectral density, PSD, expresses how the power or variance of a time series is distributed with frequency. Mathematically, it is defined as the Fourier Transform of the autocorrelation sequence of the time series.

Alkan *et al.* (2006) performed a comparative study on detecting epileptic seizures by using Auto Regressive and Welch's method. Parameters of AR methods were determined using Yule Walker, covariance and modified covariance methods. The performance of these methods was evaluated using

PSDs. Covariance methods performed the best when compared to the other methods.

John Musson *et al*. (2010) performed a survey on the comparative analysis of PSD estimation of EEG signals. Accurate estimation of PSD requires interpolation and smoothing techniques, which are profound in the periodogram of the non-parametric techniques. Non-parametric methods are computationally efficient when they are coupled with FFT, by which they tend to preserve fine-frequency fluctuations. The survey recommended Welch's method for generic applications.

2.4.3 Time Frequency Domain features

In signal processing, time–frequency analysis comprises those techniques that localize a signal in both the time and frequency domains simultaneously.

Ubeyli (2009) illustrated the use of the combined neural network by employing Wavelet coefficients for EEG. The EEG signals were decomposed into time–frequency representations using discrete wavelet transform and statistical features were calculated to depict their distribution. The first-level networks were implemented using the statistical features as inputs. The second level networks were trained using the outputs of the first-level networks as input data.. The combined neural network model achieved better accuracy rates than that of the stand-alone neural network model.

Tzallas *et al*. (2007) presented a method by analyzing EEG signals based on time-frequency features. Initially, selected segments of the EEG signals were analyzed using time-frequency methods and several features were extracted for each segment, representing the energy distribution in the time-frequency plane. These features are used as an input to ANN, which provided the final classification of the EEG segments concerning the existence of seizures or not. The evaluation results were very promising with an overall accuracy from 97.72% to 100%.

Marcus Musselman *et al*. (2012) proposed a novel work for discrimination of epilepsy based EEG signals. Features were extracted from the bilinear Time–Frequency Distributions (TFD) of the EEG signal. A one-against-one decomposition was used to break the multi-class problem into binary sub problems solvable with a Support Vector Machine (SVM). The decomposition permitted binary sub problem dependent feature libraries to be constructed from EEG TFD. The newly introduced algorithm was able to outperform the best reported accuracy in literature for the problem considered in their paper.

2.5 Classification

Classification in its broadest sense is a type of forecast or a decision made on the basis of the currently available information, thereby employing a procedure for repeatedly making judgments in novice state of affairs. The main historical strands that clasp the perspective of classification can be identified as statistical, machine learning and neural network.

2.5.1 Statistical Classifier

Statistics is one of the oldest disciplines to study data and make inferences based on the information contained in the data. A statistical classifier is required whenever information is to be gleaned from the data. They are characterized by having an explicit underlying probability model in each class rather than simply a classification. Some of statistical classifiers used for Epilepsy detection were Fischer linear Discriminant analysis, Linear Discriminant Analysis, Naives Bayes Classifier.

2.5.1.1 Fisher's Discriminant Analysis

It is one of the most commonly implemented and matured classification procedures .The basic idea is to divide the sample space into two dimensions or hyper- planes in many dimensions. Fathima *et al*. (2011) in his paper, proposed a method to classify EEG signals into normal and seizure classes based on the statistical distributions. After ranking the features using Fisher's discriminant analysis, variance, skewness and Coefficient of Variation (CoV) were found to form the best set of features. The accuracy attained for classification in their paper was 96.9%.

2.5.1.2 Decision tree

It is a method of classification based on recursive partitioning of the sample space by finding explicit and establishing rules-like relationships among the input and output variables using search heuristics. Decision trees are built of nodes, branches, and leaves that indicate the variables, conditions, and outcomes, and the most predictive variable is placed at the top node of the tree. Tzallas *et al*. (2009) proposed a work based on epileptic seizures using time frequency analysis. Short-time Fourier transforms and several time frequency distributions were used to calculate the Power Spectrum Density (PSD) of each segment. The analysis was performed in three stages: 1) time frequency analysis and calculation of the PSD of each EEG segment; 2) feature extraction, measuring the signal segment fractional energy on specific time frequency windows 3) Existence of seizure was detected using Decision trees.

2.5.1.3 Linear Discriminant Analysis

Kunjan Patel *et al*. (2009) devised an algorithm to monitor epilepsy in patients using Ambulatory Electroencephalograph (AEEG) technology. The main intention was to circumvent interruptions to the patients routine life. A low power real time seizure detection algorithm suitable for AEEG devices is proposed herein. The performance of various classifiers was tested and the most effective was found to be the Linear Discriminant Analysis (LDA) classifier. The algorithm provided 87.7% accuracy, with 94.2% sensitivity and 77.9% specificity.

2.5.1.4 Naïve Bayes Classifier

The Naïve Bayesian Classifier (NBC) is a simple and effective algorithm based on clustering of data and applying Bayes theorem on the clustered data.

This model estimates the class probability by assuming that the attributes of one class are conditionally independent from the other class. Poornendu Prakash Tiwari *et al* (2012) investigated the use of the Naïve Bayes classifier and Back propagation neural network for the given signals. Naïves Bayes classifier performed comparatively better.

2. 5.2 Machine Learning Algorithm

Machine Learning encompasses automatic computing procedures based on logical or binary operations, which learns a task from a series of examples.

2.5.2.1 Support Vector Machine (SVM)

SVM is a binary classifier, where the data can be transformed into a higher-dimensional space in which elements belonging to two different classes can be linearly separated, where the dimension of the high-dimensional space is substantially larger than the input space.

Shoeb *et al*. (2004) presented a machine learning approach in constructing classifiers that detect the onset of an epileptic seizure through EEG. This problem is challenging because the brain's electrical activity is composed of numerous classes with overlapping characteristics. The key steps involved in realizing a high performance algorithm included shaping the problem into an appropriate machine learning framework, and identifying the features critical to separating seizure from other types of brain activity. The algorithm detected 96% of 173 test seizures with a median detection delay of 3 seconds and a median false detection rate of 2 false detections per 24 hour period.

Nicoletta Nicolaou *et al*. (2012) investigated the detection of epileptic seizures based on Permutation Entropy (PE) and SVM. PE was the first of its kind to be used for feature extraction and based on the PE values SVM is used to classify normal and epileptic EEG. EEG is characterized by lower PE in case of epilepsy. It is shown that average sensitivity of 94.38% and average specificity of 93.23% was obtained in the proposed system.

Gonzalez *et al.* (2010) developed a robust technique for automatic detection of the epileptic seizures using SVM. The three features namely, energy, decay (damping) of the dominant bequency and cyclo- stationarity of the signals are used for detecting seizures. The use of SVM achieves high sensitivity and at the same time shows an improvement in terms of computational speed in comparison with other traditional systems.

2.5.3 Neural Network

The mathematical model inspired by biological neural networks aims at modeling complex relationships between the inputs and the outputs as an outcome of finding patterns in data.

2.5.3.1 Supervised Learning

A technique used for approximating the input output behaviour of the complex systems by establishing a rule for classification of new observation into one of the existing classes.

2.5.3.1.1 Back Propagation Network (BPN)

A feed forward network with at least one hidden layer is called a BPN or Back Propagation Neural Network(BPNN) . It uses an iterative gradient descent algorithm to minimize the mean-squared error between the desired output and the actual network output.

Tzallas *et al.* (2006) presented a method of time frequency analysis of EEG signals. Features were extracted and used to represent the energy distribution in the time-frequency plane. PCA is employed for dimensionality reduction. N number of features are used as an input to ANN resulting in M classes. The hidden layers are sigmoid units with hyperbolic tangent as activation function, while the outputs are linear. Half of the pattern of the data set was randomly selected to be used for training, while the rest was used for testing. The network is trained using a standard back propagation algorithm.

Sivashankari *et al*. (2009) has proposed a novel automated approach for detecting epileptic seizures in EEG signals by obtaining statistically independent components and subsequently training the components using Back Propagation Neural Networks (BPNN). The experimental results reveal that the accuracy in detecting seizures is quite high.

2.5.3.1.2 Multi Layer Perceptron Network (MLPN)

A Multilayer Perceptron (MLP) is a feed forward Neural Network model that maps sets of input data onto a set of appropriate outputs. Subasi *et al*. (2005) in his paper dedicated to the functioning of Lifting Based Discrete Wavelet Transform (LBDWT), MLPNN and LR for classification of Epileptic seizures. A novel and reliable classifier was obtained by applying LBDWT in connection with MLPNN. The comparison between the classifiers was performed based on receiver operating characteristics. It was noted from the results that MLPNN outperformed LR.

Ubeyli *et al*. (2009) proposed a paper using statistics representing the features of the EEG. MLPNN architectures were formulated and used as a basis for detection of electroencephalographic changes. Lyapunov exponents, wavelet coefficients and the power levels of PSD values obtained by eigenvector methods of the EEG signals were used as inputs of the MLPNN trained with Levenberg Marquardt algorithm. The classification results confirmed that the proposed MLPNN had the potential in detecting the electroencephalographic alterations.

Mahmut Hekim (2012) proposed an MLPN based classification of EEG signals, using the average power based on rectangle approximation window average power was extracted from the power spectral densities of frequency sub-bands of the signals using ANFIS and MLPNN. The experiments showed that both classifiers with the proposed approach resulted in satisfactory classification accuracy rates, although MLPNN performed little better than ANFIS.

2.5.3.1.3 Radial Basis Function Network (RBFN)

In RBFN the learning is performed in two stages. One which takes place in hidden layer usually occurs in the unsupervised bottom up self organizing method. The other is learning in the output layer which is carried out by top down supervised approach such as least square estimation.

Kesban Aslan *et al*. (2008) proposed a model for classification of epilepsy using EEG signals using RBFN. The correct classification of this data was performed by two expert neurologists before they were executed by neural networks. A comparative study was performed using RBFNN and MLPNN. Experimental results showed the predictions of the RBFN model with a total classification accuracy of 95.2% when compared with MLPNN with a total classification accuracy of 89.2%. These results indicate that RBFNN model may be used in clinical studies for classification of epilepsy groups.

Harikumar *et al*. (2011) performed an analysis of Singular Performance Value Decomposition (SVD) and Radial Basis Function Neural Networks (RBFNN) for classifying Epilepsy Risk Levels in EEG signals. The fuzzy pre classifier is used to classify the risk levels of epilepsy. Further the SVD and RBF neural network is employed on the results obtained from the first step to identify the optimized risk level. The efficiency of the results of the above methods is compared based on the benchmark parameters such as Performance Index (PI), and Quality Value (QV).

2.5.3.1.4 Probabilistic Neural Network (PNN)

PNN is a kind of distance-based ANN using bell shaped activation function comprising of input Layer, Radial Basis layer and Competitive layer.

Srinivasan *et al*. (2007) proposed a neural network based automated epileptic EEG detection system that used statistical parameter ApEn as the input feature. The value of the ApEn drops sharply during an epileptic seizure and this fact was used in their proposed system. Two different types of neural networks, namely, Elman and probabilistic neural networks were employed in their work for achieving an overall accuracy .

Forrest Sheng Bao *et al*. (2008) developed a new approach towards the development of automated epileptic Seizures using PNN. The diagnostic system uses an interictal EEG system to detect epilepsy in a patient. The system was developed by extracting four features from EEG data to build a PNN. Leave -one -out cross validation was used which reflects an impressive accuracy in the system. It uses a Bayesian strategy and hence is considered to be more appropriate for medical application. It has the advantage that it is not necessary to train the network over the entire data set as more and more patient's data become available as it has the property that the decision boundaries can be modified in real time.

2.5.3.1.5 Recurrent Neural Network (RNN)

RNN is a particular class of multilayer network which in addition to the feed forward connections have self-connections or connections to units in the previous layers. This recurrence acts as a short-term memory which aids the network in remembering the past.

Srinivasan *et al*. (2005) discussed an automated epileptic seizure detection system by extracting time and frequency domain features using a special type of recurrent neural network known as Elman network. The experimental results show that the accuracy rates were high when used with a single feature alone.

Nihal Fatma Guler *et al*. (2005) aimed in his study to classify EEG signals with high accuracy by employing a RNN with Lyapunov exponents as features and trained with the LM algorithm. This approach is under the consideration that the EEG signals are chaotic in nature and hence was a need for a reliable classification method. The RNN achieved higher accuracy rates when compared to FFNN models and could be useful for analyzing long term EEG signals.

2.5.3.1.6 Wavelet Neural Network (WNN)

WNN is a neural network based on wavelet transform, in which discrete wavelet function is used as the node activation function and the wavelet space is used is of significant importance in pattern recognition. The feature extraction is realized by a weighted sum of the inner product of wavelet base and signal vector.

In 2006, Jahankhani *et al*. proposed a feature extraction using WT and a learning based classifier was used for classification. The results of the model were evaluated in terms of training performance and classification accuracies which confirmed that the proposed scheme had potential in classifying the EEG signals. Ales Prochazka *et al*. (2009) devoted in his paper the use of DWT an alternative to the commonly used Discrete Fourier Transform (DFT). Classification was done by neural network. Problems of multichannel segmentation were also mentioned in his paper.

All these techniques focus on improvement in terms of various aspects like integrity, reliability and robustness. Notwithstanding the application potential of such methodologies, automated detection of EEG is still an open problem, where improvements are still being investigated.

2.5.3.2 Un-Supervised Learning

A technique where the learning is based only upon the input data and is independent of the desired output data; Such a type of learning is called unsupervised learning where the error is not calculated to train a network.

2.5.3.2.1 Self Organizing Feature Maps (SOFM)

A Self Organizing Map (SOM) or referred to as SOFM is a neural network which classifies patterns according to their intrinsic similarity and maps the result onto a spatial structure. It has two layers, the input layer, and the competitive (or Kohonen) layer. These connections are randomized from the start and then modified by the training procedure.

Van Meijjel *et al*. (1997) has demonstrated the use of SOM in order to overcome the shortcomings in detection of EEG spikes, using error back propagation. These types of neural networks require examples of spike wave complexes and non spike wave complexes. Finding examples of spike wave complexes were more difficult than non spike wave complexes and also had an added disadvantage of occupying larger input space than spike wave complexes.

James *et al*. (1999) proposed a system to detect the presence of epiletiform in EEG using Kohonen's Self organizing feature maps. Features were extracted using mimetic approach. The proposed system incorporates the outputs of sixteen modules into a single channel module which is trained through Kohonen's Learning vector quantization technique. Results obtained showed that the trained SOFM has a sensitivity of 63% and specificity of 79%.

Most of the researchers prefer feed forward networks trained by back propagation algorithm for EEG spike detection. Kurth *et al*. (2000) examined the off line spikes using Kohonen's feature map. Three different sized networks were examined and trained using Kohonen's Feature Map (KFM) and the results from these networks were appraised with the findings of certified electroencephalographer . This was done in order to investigate the quality of results obtained from KFM, were promising or not.

Reza Khosrowabadi *et al*. (2010) employed a system using SOM for boundary detection in EEG based emotion detection. EEG features were extracted using magnitude squared coherence of EEG signals. The boundaries of EEG features were extracted using SOM and a 5 fold cross validation using K-NN classifier was performed .The proposed method proved with accuracies up to 84.5%.

2.5.3.2.2 Adaptive Resonance Theory (ART)

ART is a self organizing neural network architecture that clusters the input pattern space and produces weight vector templates appropriately. It learns new patterns and doesn't alter the patterns that have been learnt previously. ART 2 is a streamlined format of ART networks, which clusters data into different classes by learning when the classes are labeled by an expert and are saved in a look-up table for use by the system. A new input data is recognized by finding the class assigned by ART 2 by searching through the table.

Ozdamar *et al.* (1992) in his paper has performed an unsupervised learning for detecting spikes in epilepsy using ART 2. In their study, 20 inputs were trained using EEG data containing spike and non-spike waveforms an ART 2 neural network with. For the purpose of comparison a MLP was constructed and evaluated similarly. Network performance of ART 2 was close to that of MLP. It was revealed that ART 2 could be trained with just one or a few iterations when compared to BPNN which in turn requires around thousands of iterations.

2.6 Other Classifiers

2.6.1 Fuzzy Classifiers

Sadati *et al.* (2006) exercised Fuzzy classifier for detecting epilepsy using EEG. An Adaptive Neural Fuzzy Network (ANFN) was used for classification. The results were compared with other classifiers such as SVM, ANFIS and FFBNN. It is shown that a classification accuracy of about 85.9% can be achieved using ANFN.

Subasi (2007) implemented a new approach based on an ANFIS for detecting epileptic seizures. The proposed ANFIS model was an amalgamation of neural network and the fuzzy logic. Feature extraction was performed using the WT and the ANFIS was trained with the gradient descent and least squares method. The results were highly promising, and a comparative analysis suggested that the proposed modeling approach outperforms ANN model in terms of training performances and classification accuracies.

2.6.2 ELM Classifier

Yuedong Song *et al*. (2010) implemented an automatic seizure detection system using sample entropy and ELM. BPNN and ELM were implemented in his study. SampEn was the only feature used in his study and when assimilated together with ELM achieved high classification accuracy of 95.67%. In 2012, a novel method was proposed by them which adopted an optimized sample entropy combination of ELM and demonstrated that the proposed amalgamation had a great potential for real time detection of EEG signals.

Yuan *et al*. (2011) performed classification of EEG signals based on ELM extreme learning machine. Non linear dynamic features are extracted from EEG signals using ApEn and HE. ELM algorithm was employed to train a SLFN network with the non linear features. A Detrended Fluctuation Analysis (DFA) was performed to characterize EEG signals. A comparative analysis was performed using Back propagation algorithm, SVM and ELM and the ultimate performance was achieved by ELM.

A discussion on the summary of each sub-section with relevance to the proposed work is envisaged subsequently. The decisive factor coagulating the pre-processing section are as follows

Regression based artifact removal needs calibration trials to determine regression coefficients for EOG, EMG and electrical artifacts and this becomes a very tedious and extensive process. Hence regression based artifact removal method is not considered in this study.

Artifact removal using adaptive filters have the capability to adjust their parameters effortlessly, but still requires determining reference inputs, which is a wearisome task. This method was implemented for a comparative study.

Principle Component Analysis converts correlated variables into a set of uncorrelated variables called Principal Components. It uses only second order statistical information as opposed to **ICA**, which makes use of Higher Order

Statistics .HOS are less encountered by Gaussian noise than 2nd order measures. Hence delineating this aspect ICA was considered in this study.

Discrete Wavelet Transform implements the concept of trapping the neural information in the approximation part and ensnares the artifactual information in the detail part effortlessly was an enticing factor in the selection of DWT in this study.

The factors influential for choosing methods for feature extraction are discussed as follows

Statistics play a dominant role in extorting the features of the signal.The non-stationarity of the signals are quantified by measuring the statistics of the signal and the features under contemplation are mean and standard deviation. Higher order statistics are not adopted here because the signals at the preprocessing level engross HOS.

Hjorth's Descriptors mainly depend upon the calculation of variance and standard deviation of the signal. But since the calculation of standard deviation is attempted in the statistical feature approach, hence calculating Hjorth's parameters could be just a replication.

Some of the Non-linear features attempted for discussion are discussed subsequently

In real time systems like EEG it is very difficult to prove or exclude the chaos. Predominantly, a complicated system like the brain is a manifestation of a mixture of noise, some cyclic processes and random fractal signals. Even though noises or artifacts are removed, such compositions are frequently reported to deceive the algorithms used to detect chaotic dynamics. Hence the **Correlation Dimension** estimates should be interpreted with extreme caution and hence neglected in this part of the study.

Largest Lyapunov Exponent, which is the quantitative estimation of the sensitive dependence on initial condition (SDIC). When the differential equations are known, deriving the exponents are much easier, but extracting exponents from a time series is a complex problem and requires utmost care in its application and the interpretation of its results.

Entropies detect the irregularity of a signal, and there do exist many flavors of entropy. The skeptics underlie as which entropy would be the best for EEG study.

The **Hurst exponent** is the asymptotic behaviour of the rescaled range as a function of the time span of a time series. If the same time series is considered and the number of observations are increased, then the rescaled range will also tend to increase. In a complex time series signal like EEG, one is intended to take utmost care in choosing the rescaled range.

FFT considers signals which are sinusoidal in nature and tend to induce complexity while processing the signal. In **FWHT**, the complexity of the signal is reduced as it involves coefficient addition and subtraction instead of multiplication and division.

Parametric methods are highly successful when a model reflecting the behavior of the system is appropriately built and would have a cautioning effect, if it is failed to be built.

Time frequency Domain features

The concept of windowing and Fast Walsh Hadamard Transform is a combination of time, frequency domain features adopted in this study.

In order to choose one of the best classifiers, some of the significant reasons under consideration are discussed below

Fischer Discriminant Analysis requires that, all the data points in each cloud needs to be collapsed onto one point and projected on some line and then making the two points as far as possible, by which it would be very easy to predict as to which class a data point belongs to, but this is not feasible in the practicality. Also the optimality could be achieved only if the underlying distribution of groups is multivariate normal otherwise the procedure becomes more robust.

Decision trees can help to classify information very quickly, but a lot of time is taken to construct and analyze it. Another inherent challenge with decision trees is that the tree is only as accurate as the information incorporated. If inaccurate information is introduced at any node on the tree, the error is propagated throughout the rest of the tree which could lead to multiple wrong decisions, rather than one wrong decision. Decision trees have a hard time adapting to different variations of the same or similar situations.

LDA requires parameters to be estimated and more number of estimations will have a degrading impact over the robustness of the classification algorithm.

Naïve Bayes classifier outperforms sophisticated classifiers, even when the underlying assumptions of independent predictors are far from true. It requires a large number of records to obtain good results, and the advantages are especially pronounced when there are a large number of predictors.

The main complication of **SVM** for practical usage is that a complex quadratic programming problem must be solved and this demands a high computational complexity and computing time cost. For larger data sets this shortcoming is more obvious. Furthermore, proper kernel and its parameter selection are always an open problem.

The most common problem encountered in **BPN**, is the problem of “Local Minima”, and the time to converge to an optimum solution.

A **MLFFNN** requires many repeated presentations of the input patterns and the weights required to be adjusted before the network is able to settle down into an optimal solution. Additionally, lack of efficient and fast learning algorithm for large scale problems. Fixing the number of layers and neurons in each layer is also one of the challenges.

RBFs are more sensitive to the curse of dimensionality and have a greater difficulty, if the number of input units is large, but in a complicated system like brain studies the dataset is enormous even after dimension reduction.

A novice classifying pattern, makes **PNN** slower than MLPNN. An added disadvantage is the amount of memory space required to store the model is large.

RNN facilitates neurons send feedback signals to each other adopting a flexibility in the architecture of RNN therefore, a very pensive decision has to be made in deciding the architecture.

WNN involving wavelets are rapidly vanishing features and it is required that the parameters are needed to be c 81d ned in order to prevent degenerated wavelet shapes.

SOFM, an unsupervised learning where the training data set contains only input variables. SOFM is attempted to be useful where an exploratory data analysis is necessary, such as data mining.

ART is not a self generated means of analyzing its own activity.Though its knowledge is easier to understand and extract, it remains inherent and can't be transferred to other tasks.

Description of problems in terms of linguistics is a major advantage in using **Fuzzy Classifiers**, as it enables the creation of inference system based on human thinking, which is very easy to use and modify the knowledge base.

The alluring fact behind **ELM Classifier** is that, the input weights and hidden layer biases are randomly chosen and the output weights are analytically determined instead of tuning them and hence circumventing the problem of local minima.

2.7 Review related to the current research work

Muhammad *et al*. (2011) in their work on artifact removal has employed spatial constraints on ICA. Spatially constrained ICAs were proposed to extract artifact-only independent components, which was subsequently followed by Wavelet Denoising (WD) to remove cerebral activity from extracted ICs. Ultimately, the extracted artifact-only signals are projected back and subtracted from EEG data to get a clean EEG for further analysis and processing.

The phase I of this study has emanated from the work done by Muhammad *et al*. Emphasis is carried out in extracting artifact-only signals which arrive in a very extensive procedure and needs to haul the artifact information till it is projected back aı 81e racted from EEG signals. As a result it has an impact on time complexity and memory utilization of the processor. In order to avoid a protracted procedure, certain modifications have been considered in the preprocessing stage of our study.

Cerebral information is only extracted as opposed to artifactual information. Wavelet Denoising is further performed to remove any artifact information. Three divergent spatially constrained algorithms are achieved by adopting FastICA, InfomaxICA and ExtendedInfomaxICA resulting into Spatially Constrained FastICA, Spatially Constrained InfomaxICA and Spatially Constrained Extended Infomax ICA. Wavelet Transformation was carried out by Daubechies followed by Otsu and fuzzy shrink thresholding.

The fuzzy shrink thresholding was adopted in this research work after probing through the work done by Jamal Saeedi *et al*. (2010). The work adopted by them comprises of adopting a wavelet-based fuzzy denoising technique for single and multi-channel images denoising. Fuzzy feature was modeled using intra-scale dependency of the wavelet co-efficient. This fuzzy feature was used to distinguish between noisy and noiseless coefficients and to enhance wavelet coefficients information in the shrinkage step. A fuzzy membership function was used to shrink the wavelet coefficients based on the fuzzy feature. The same concept which was used for image denoising was carried out in EEG signal denoising.

Azadeh bastani *et al*. (2011) has delineated in his study a comparison between Walsh Hadamard transform and Fourier analysis of EEG signals. Their work proved the performance of Walsh Hadamard transform to be more effective in extracting the features of the signals for sleep stages of EEG. Owing to lesser number of operations, simplicity of implementation by digital circuits and the swiftness of the WHT, it has been recommended that WHT, is preferred to Fourier transform (FT), as FT requires a greater volume of operations for real time processing of signals. Based on the evaluation on comparative analysis between WHT and FT, it was heuristically stated that Fast Walsh Hadamard Transform (FWHT) could be an attractive alternative to the FFT. Phase II of this study implements FWHT for extraction of EEG signals in the bands. Ravish *et al*. (2012) performed an analysis based upon the extraction of parameter values such as power distribution in various phases of EEG time series, and spectral power in various frequency bands of EEG. The power distribution in alpha band and delta band was computed to detect seizures. Based on this doctrine, the PSD values are calculated in each of the bands extracted from FWHT. The statistical variations (i.e. minimum, maximum, mean and standard deviation) of PSD are the features fed as input for the next phase classification.

The third phase of study i.e. classification was carried out by using Hybrid Extreme Learning Machine (HELM) with Modified Levenberg-Marquardt (MLM) algorithm uses Analytical Hierarchy Process (AHP) method with minor modifications in the algorithm was proposed by Amal Mohamed Aqlan *et al*. (2008). Yuedong Song *et al*. (2010) implemented an automatic seizure detection system using sample entropy and ELM. An implementation of PSD and HELM was a projection to Songs research work. Amal Mohamed *et al*. in their novel algorithm had concentrated much on the learning speed of FNN. The main reason for the adoption of his algorithm was to replace the slow gradient based learning algorithm and tediousness involved in tuned iteration of parameters in these learning algorithms. This method uses the AHP method to select the input weights and hidden biases, the ELM algorithm to analytically determine the output weights and the LM algorithm to learn the network. His model was implemented in our Epilepsy study of EEG with a focus on overcoming the speed by implementing yet another modification to the LM algorithm. In order to perform a comparative study another classification based on Neuro Fuzzy model, ANFIS was adopted and a minor modification was made by implementing MLM in backward pass instead of Gradient Descent (GD) resulting into FANFIS (Fast ANFIS).

2.8 CHAPTER SUMMARY

The review of literature divulges that, though tremendous progression is apparent in automated detection of Epilepsy using EEG, it is conspicuous that there lacks an existence of a broad framework of ideas, which could assimilate findings from too different scales and levels of observations. Artifact removal process could have progressed without using spatial constraints, but implementing such a constrained environment creates the possibility of feature reduction. Further artifacts could have been removed using either ICA or DWT, but assimilating such a combination paves a further way to still remove the finer impediments. Feature extraction could have implemented using Frequency domain or Time domain algorithms, but working with frequency-time domain feature extraction satiates to choose the lesser number of features. Classification could have been implemented using any one of the classifiers. Probing through the merits and shortcomings of all the classifiers, one of the best classifier was chosen to support voluminous amount of data.

No single focused level of analysis suffices the detection of Epilepsy. Much effort has been made by scientists in detecting epileptic seizures, but providing an integrating solution using the techniques of signal processing and soft computing is still sparse. Sifting through manifold techniques and their issues were the prime stride in the progress of this research which aims to provide insight in choosing an appropriate conceptual and empirical framework of algorithms. This research confronts to propose an amalgamation of techniques for detecting seizures by ascribing the removal of the daunting artifacts, applying automatic processing algorithms to extract hidden features and using an appropriate classification algorithm. The techniques and methods used are elucidated in the subsequent chapter, Methodology.