



Use of complexity based features in diagnosis of mild Alzheimer disease using EEG signals

Nilesh Kulkarni¹

Received: 10 June 2017 / Accepted: 27 October 2017 / Published online: 7 November 2017
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2017

Abstract Alzheimer Disease is one of the neuro-degenerative diseases and most expensive disease in the modern society characterized by cognitive, intellectual as well as behavioral disturbance. Early diagnosis of Alzheimer disease is a major concern due to increasing population and lack of a standard and effective diagnosis. Although EEG is a powerful and relatively cheap tool for diagnosis of Alzheimer disease and dementia; it still not achieves the standards of clinical performance in terms of sensitivity and specificity to accept as a reliable technique for screening of Alzheimer disease. Hence, there is scope for increase in performance of EEG based diagnosis. This paper focuses on Electroencephalography based diagnosis and complexity based features for early diagnosis of Alzheimer disease. From the results, it is observed that the EEG of the Alzheimer patients slows down and it is less complex as that compared to the Normal patients. In this study, different complexity based features such as Spectral Entropy, Spectral Centroid, Spectral Roll-off and Zero Crossing Rate are used. K nearest Neighbor classifier is used for classifying the data between two groups i.e. Normal and Alzheimer disease patients giving accuracy of 96% in present research work.

Keywords Alzheimer disease · Dementia · EEG · Spectral centroid · Spectral entropy · Spectral roll-off · Zero crossing rate · K nearest neighbor classifier

1 Introduction

Alzheimer disease is one of the Neuro-degenerative disease which is found to be complex in the present scenario. It is the common form of dementia and by the time it affects the brains cells [1, 2]. Alzheimer Disease is a chronic Neuro-degenerative disorder that has ranked as third most expensive disease and sixth leading cause of death in United States. It is neurodegenerative disorder characterized by rapid impairment of memory and some other cognitive functions, which are mainly associated with the behavioral disturbances and finally leads to total dependency [3]. An important research is to identify the neuro-anatomical basis of cognitive impairment in Alzheimer disease (AD). The need of research is to understand the changes taking place related to the cognitive impairment and the progression of AD in the brain structure [4–7]. There exist different techniques for diagnosis of Alzheimer Disease and other neurodegenerative diseases such as Epilepsy, Brain Stroke, and Parkinson's Disease, etc. Different neuroimaging methods such as Magnetic Resonance Imaging (MRI), Single Photon Emission Computerized Tomography (SPECT), Positron Emission Tomography (PET) are used today for diagnosis of neural diseases. Imaging has a key role in medical diagnosis, education and noninvasive therapeutics. The new scientific and technological advances boost the complex issues of diseases such as Alzheimer, Epilepsy and many more. Computer Aided Diagnosis (CAD) is a general tool used for a variety of applications such as to diagnose the disease in medical applications. CAD helps the physicians, researchers in diagnosing the disease in less time by identifying the patterns, making fewer efforts. Non-neuroimaging methods such as EEG, Biomarkers are also used today for AD diagnosis. Electroencephalography (EEG) is one of the

✉ Nilesh Kulkarni
nileshkulkarni_svc@sinhgad.edu

¹ Department of E & TC Engg., STES's Sou. Venutai Chavan Polytechnic, Pune 411041, India

tools which can be used for early diagnosis of Alzheimer disease. The Electroencephalogram (EEG) basically reflects the electrical activity of large number of cortical neurons which is mainly associated with the neural information processing of brain regions. In present scenario, there is no significance of objective method based diagnosis of Alzheimer Disease but the use of EEG as a diagnostic tool continues to be challenging part in current studies [8, 9]. Focusing on previous studies obtained in literature, none of the existing systems are not clinically or analytically validated. Due to such reasons, the systems require significant improvements.

In present study, our aim was to investigate and observe the effects of different complexity based features on EEG signals of both Alzheimer Disease and Normal patients. In literature, we can see that Spectral Based features such as EEG Relative Power, Magnitude Square Coherence, Phase Synchrony and EEG amplitude modulation Energy are widely used which plays a significant role in AD diagnosis giving accuracy of about more than 80%. Our hypothesis in this study is to prove that EEG of AD patient tends to be less complex as compared to the normal subjects due to neuronal loss of cells in brain regions. In further part of the paper we discuss the role of the different non—linear features used for early diagnosis of Alzheimer disease.

2 Methods

The proposed approach consists of four phases, that is, signal acquisition, preprocessing and segmentation, feature extraction, and classification between two groups. The basic flow of work is shown in Fig. 1. The first phase of our proposed approach prepares the input signal for further analysis. Second phase includes the preprocessing of the EEG signals to remove unwanted parts from signals which include the noise content removal such as eye blinking activity, muscle activity etc. Features are extracted in the third phase. The fourth and final phase classifies the signal as belonging to a normal subject group or Alzheimer’s disease group.

Let us discuss each of the block in more detail.

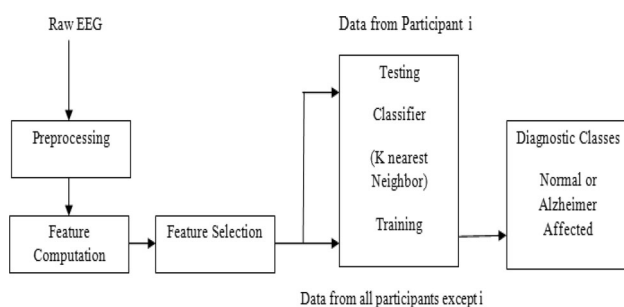


Fig. 1 Block Diagram of the Proposed System

Phase 1—preprocessing Any signal processing algorithm like segmentation and feature extraction relies significantly on the quality of the signal obtained. The quality of EEG signal deteriorates either during its acquisition process or afterwards. The process of EEG acquisition may incorporate certain artifacts. For example noise problem and intensity in homogeneity. These artifacts occur due to the power lines interferences in the machine. Some biological artifacts are also observed in the recordings of the EEG data. They may occur due to the human efforts. Eye blinking of patients, Muscle activity & variations in EEG electrodes are some of the causes for signal contamination in EEG recordings. Hence to remove these and to optimize the signal for further analysis and evaluation, a number of preprocessing steps are required. Signal preprocessing can significantly increase the visual reliability of the signal. It involves a set of techniques which enhances or eradicates certain details of the signals in order to efficiently process it for further analysis. In this study, Bandpass filter is used for filtering the EEG signal from 0.5 Hz to 30 Hz.

Phase 2 & 3 segmentation & feature extraction Biosignals processing serves best in early detection of Alzheimer’s disease as they hold necessary information to distinguish between healthy controls and Alzheimer’s disease patients. But the major concern here is the huge data size of EEG signals. In order to classify these signals by the classifier, the computational time is enormous. Moreover not all information in the signal is required for the classification purposes as most of the information is irrelevant. For this purpose feature extraction is performed to find more relevant and discriminative features in order to classify signals more efficiently. In recent literature a vast variety of features has been extracted from EEG signals for the identification of AD. These features are Time Frequency Based features, Wavelet based features & Complexity based features [8, 9]. These selected features are affected in the earliest stages of AD. Hence the contribution of presented research is towards identification of AD patients in early stages, using a smaller feature set which results in lower computational expense.

Phase 4: classification Different types of classifiers, both supervised as well as unsupervised are efficiently used in machine learning & pattern recognition. In proposed system, supervised approach for classification between two classes, i.e. Support Vector Machine is used. In addition to different classifiers, ensemble of these classifiers can be used to verify the enhanced accuracy rate. The Fig. 1 shows the methodology used for early diagnosis of AD using EEG signals. Along with this, EEG signals of Temporal, Frontal, Central and Parietal lobes of both normal and Alzheimer disease patients were taken into the consideration for study [10, 11].

3 Data collection and subjects involved

Data used in the study was obtained from Smt. Kashibai Navale General Hospital and Research Centre, Pune (India) consisting of both Alzheimer patients and control patients. Patients were selected from consecutive, community residing elderly persons 55–80 years of age with the report of decline in cognitive as well as behavioral functioning. Diagnosis of the patients was made by experienced neurologists based on Indian version of *Mini Mental State Examination (MMSE)* and *Clinical Dementia Rating (CDR)*. Resting awake multi channel EEG recordings were obtained from 100 participants separated into 2 groups. The first group was composed of 50 subjects; 30 males and 20 females (mean age: 60 years) giving indication of functional cognitive and behavioral decline. The second group consists of 50 participants of normal subjects consisting of 35 males and 15 females (mean age: 60.5 years), giving no indication of functional cognitive decline. In addition to the AD cohorts, an additional criterion was the presence of functional, behavioral and cognitive decline over the previous 6 months. Patients belonging to the abnormal group were also checked for diabetes, kidney disease, thyroid disease lung and liver disease or vitamin B12 deficiency, as these can also cause cognitive decline. The EEG recordings and the study was approved from Ethical committee of the hospital and the participants.

EEGs were recorded from RMS (Recorders and Medicare systems Private Limited) EEG machine with 12 bits resolutions and sampling rate of 1024 Hz. Impedance was maintained below 10Mohms and the electrodes (Referential Montages) were placed according to the International 10–20 systems. Biauricular referential electrodes were also attached as recommended by American EEG Society. The Power grid interference was eliminated by low pass filtering. As there is evidence of an interhemispheric disconnection in AD and dementia, a virtual hemisphere bipolar montage is also taken into consideration. The obtained signals are also termed as ‘Bipolar signals’. The Bipolar signals recorded in the study and taken into the consideration are Fp1–Fp2, F3–F4, F7–F8, C3–C4, T3–T4, P3–P4, T1–T6, O1–O2,. During EEG examination and recordings, patients were awake and relaxed with eyes closed. The artifacts of EEG signals such as muscle activity and eye blinking were removed manually.

4 Preprocessing

In proposed research work, EEG signal is acquired using EEG electrode cap. But, at the time of acquiring the signals, the signal is contaminated with different noise

artifacts. These artifacts are mainly associated due to the power line interference, muscle activity of the patients and eye blinking. To obtain the noise free signals, preprocessing of the signal is essential. In our present work, we have used the Independent analysis (ICA) and Wavelet based Denoising method for signal Denoising. Till now different algorithms are applied to biomedical signals considering the EEG signal, one of the most popular classes of algorithm is the Independent Component Analysis (ICA). It is the effective tool used for obtaining noise free signal. The important use of the above technique is to perform the dimensionality reduction, and separate the relevant information of the signal. Along with the same, we have also incorporated the use of Wavelet based Denoising technique in order to preprocess the input signal. In our present study, we have used the one dimensional wavelet based Denoising technique for preprocessing the input signal.

5 Feature extraction

There exist different features for diagnosis of Alzheimer disease in literature. Features play a significant role in automated diagnosis of Alzheimer Disease [12–16]. In our study, we have incorporated the use of complexity based for diagnosis. Before discussing the different features used in the study let us see the importance of the Jarque–Bera test.

5.1 Importance of Jarque–Bera test

After the filtering and preprocessing of EEG signals, we have divided the EEG signals into samples of 3–5 s. Each epoch divided into different was test for normality condition using the Jarque–Bera test. Jarque–Bera test of the null hypothesis is such that the sample in vector x (suppose) comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution. The test is basically designed for alternatives in Generating Data Using the Pearson System of distributions. The test returns the value $h = 1$ if it rejects the null hypothesis at the 5% significance level, and $h = 0$ if it cannot. The NaN vales are by default ignored and sometimes are also missing values in x [16, 17].

The Jarque –Bera Test Statistic is given by,

$$JB = \frac{n}{6} \left(s^2 + \frac{(k - 3)^2}{4} \right) \quad (1)$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis. The features of the epochs were computed depending on the values of the h and p ; which denotes the normality conditions. The Statistical

parameters were also test and p –values are also provided in the results. Let us now discuss the different complexity based features used in the feature extraction.

5.2 Spectral entropy

Spectral Entropy indicates the amount of unpredictability and disorder in spectrum of EEG. Higher complexity is achieved if higher amount of Spectral is entropy is observed [4]. It is computed in following manner:

1. For the given signal $x(t)$, compute $S(f)$, the power spectral density (PSD), as the Fourier transform of the autocorrelation function of the signal $x(t)$.
2. Depending upon the frequency of interest; extract the Power in the spectral band from 0.5–30 Hz.
3. After calculation of the Spectral band power, normalize the power in the given band of interest.
4. Calculate the Spectral Entropy by using formula,

$$SE = \sum_{f=0.5}^{40} S(f) * \ln \frac{1}{s(f)} \quad (2)$$

5. Compute the spectral entropy as given above.

5.3 Spectral centroid

Spectral Centroid measures the shape of the spectrum of EEG signals. A higher value of SC corresponds to more energy of the signal being concentrated within higher frequencies. Basically, it measures the spectral shape & position of the spectrum [17].

It is computed as follows:

1. Let $x_i(n)$, $n = 0, 1, \dots, N - 1$ be the sample of the i th frame, with $X_i(k)$, $k = 0, 1, \dots, N - 1$, as the Discrete Fourier Transform (DFT) coefficients of the sequence.
2. Compute the Spectral Centroid of the each frame as:

$$C(i) = \frac{\sum_{k=0}^{N-1} k |X_i(k)|}{\sum_{k=0}^{N-1} |X_i(k)|} \quad (3)$$

The mean value of the spectral Centroid across all the frames can be used as the SC feature for each epoch of the frame of the EEG Signal.

5.4 Zero crossing rate

Zero Crossing Rate is the rate at which the signal changes its sign; which can be effectively interpreted as a time domain measure of signal complexity [16].

Zero Crossing Rate is computed as:

1. Let $x(n) = 0, 1, \dots, N-1$ be the samples of the i th frame.
2. Compute the zero-crossing rate for the i th frame as:

$$Z(i) = \frac{1}{2N} \sum_{n=0}^{N-1} |sgn[x_i(n)] - sgn[x_i(n-1)]| \quad (4)$$

where,

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0. \end{cases}$$

5.5 Spectral roll-off

Spectral Roll off represents the frequency below which a certain percentage (usually 80-90%) of the magnitude distribution of the spectrum is concentrated in the spectrum.

It is computed as follows:

1. Let $x_i(n)$, $n = 0, 1, \dots, N - 1$ be the sample of the i th frame, with $X_i(k)$, $k = 0, 1, \dots, N - 1$, as the Discrete Fourier transform (DFT) coefficients of the sequence.
2. Compute the spectral roll-off as the sample that satisfies,

$$\sum_{k=0}^{k(i)} |X_i(k)| = \frac{P}{100} \sum_{k=0}^{N-1} |X_i(k)| \quad (5)$$

where the P parameter is sometimes chosen between 80 and 100.

6 Results

After computations of the above features, the significant results were obtained clearing out our proposed hypothesis. Spectral Entropy, Spectral Centroid, Spectral Roll-off and Zero Crossing rate of the signals in above defined electrodes were calculated and classified by use of classifier. We have involved the use of K nearest neighbor classifier for classifying the data. Based on database available for computation, 50% of the data was trained and remaining 50% data was left out for testing purpose.

The computed values for different features used in different electrodes (averaged values) along with p values are given below (Tables 1, 2, 3, 4).

Based on above values computed, we have used K nearest Neighbor classifier for classifying the EEG data between two group's i.e. Normal v/s Alzheimer patients. k-NN is a simple intuitive and efficient method of classification used by researchers and scientists for classifying signals. This classifier makes a decision on comparing a newly labeled sample (testing data) with the baseline data (training data). For the given set of input values, it finds the

Table 1 Results obtained for SE parameters

Electrode	<i>P</i> value	Mean (CN)	Mean (AD)
F3	0.012	5.7615	3.9189
F4	0.001	5.2619	2.8722
C3	0.00021	5.3981	3.644
P4	0.0033	5.0386	3.9801

Table 2 Results obtained for SC parameter

Electrode	<i>P</i> value	Mean (CN)	Mean (AD)
F3	0.001	0.1104	0.074
F4	0.001	0.1102	0.0863
C3	0.00021	1.992	0.987
P4	0.00311	0.1942	0.0744

Table 3 Results obtained for SR parameter

Electrode	<i>P</i> value	Mean (CN)	Mean (AD)
F3	0.012	1.3433	0.9807
F4	0.001	1.9987	0.9921
C3	0.0021	1.99	0.987
P4	0.0033	1.9342	0.2311

Table 4 Results obtained for ZCR parameter

Electrode	<i>P</i> value	Mean (CN)	Mean (AD)
F3	0.01	0.11	0.0143
F4	0.001	0.0863	0.0018
C3	0.0076	0.622	0.0916
P4	0.001	0.0744	0.0344

k (closest neighborhood) in training dataset and assigns a class which appears frequently in its neighborhood. In similar manner, the algorithm for k-NN can be given as:

- 1 The k-nearest neighbor classification is performed by using a training data set which contains both the input and the testing variables which are to be classified.
- 2 Then test data which only contains input variables is compared with reference set of values.
- 3 K-NN classifier works with k patterns, the distance of unknown ‘k’ determines its class, by considering nearest neighbor points. The value of K can be varied.
- 4 Majority voting scheme where class gets one vote for each instance in neighborhood samples is classified accordingly.
- 5 The given target data is then said to be classified.

The Classification output of the K Nearest Neighbor classifier is shown below.

In K Nearest Neighbor classifier, we have to specify the value of K for classification. Default value of K is equal to 1. But, to obtain more accuracy we can vary the value of K from 1 to 10. We have used default value K = 1 in our study. Based on the features calculated and classifier used, we have calculated the accuracy of classification based on following terminology [18].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{7}$$

$$Specificity = \frac{TN}{(FP + TN)} \tag{8}$$

where, TP stands for True Positive (AD individuals correctly classified), TN stands for True Negative (NC individuals correctly classified), FP stands for False Positives (NC individuals misclassified), FN stands for False Negative (AD individuals misclassified).

In our study, we have trained 50 EEG signals from Temporal, Frontal, Parietal and Central electrodes randomly. Out of which remaining 50 EEG signals were left out for testing comprising of both Normal and Alzheimer Affected persons. The following results were obtained after testing,

1. Total number of Correctly Identified AD individuals (TP) = 24.
2. Total number of Correctly Identified Normal individuals (TN) = 23.
3. Total number of misclassified AD individuals (FN) = 01.
4. Total number of misclassified Normal individuals (FP) = 02.

Correspondingly, we have obtained the following results after calculating the values (Table 5),

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) = 24 + 23/(24 + 22 + 01 + 03) = 47/50 = 94\%$$

$$Sensitivity = TP/(TP + FN) = 24/(24 + 01) = 24/25 = 96\%.$$

$$Specificity = TN/(FP + TN) = 23 / (22 + 03) = 23/25 = 92\%.$$

Table 5 Results indicating the Accuracy obtained in the research work using KNN classifier

Accuracy	Sensitivity	Specificity
94%	92%	96%

7 Conclusion and research challenges

On the basis of the above results and features used, we have evaluated the different EEG based complexity measures to observe and study them if they carry any diagnostic useful information for diagnosis of Alzheimer disease. In medical concept, it is signified that AD affects the neuronal activity of the patients. In this study, we have stated the hypothesis that EEG signals of AD patients has less signal complexity as that compared to the CN subjects. The above used features show decreased complexity values for AD patients, which practically confirms our hypothesis. The difference in the complexity based feature values among the cohort are small, but indicates its significance on the electrodes of EEG. The AD group features consists of lower values, suggesting that AD subjects tends to be less complex. The features used carry relevant information in the central, parietal, temporal and frontal lobe. This reduced complexity occurs due to the appearance of the neurofibrillary plaques and tangles. Spectral Entropy, Spectral Roll-off, Spectral Centroid and Zero Crossing Rate values were also lower for Alzheimer patients in the frontal and temporal lobes. It is observed that there exists a higher amount of spectral content in higher frequencies for CN group. This is predicted as the high level of complexity in CN subjects. It is to highlight that when we combine these features with one another they can provide more diagnostic information. In this way, we can conclude that complexity based features can be effectively used for AD diagnosis using EEG signals.

Future work in this study includes the automated diagnosis and classification of EEG data using various classifiers such as Neural Networks, Support Vector Machines, etc. to increase the diagnostic accuracy for distinguishing between AD and CN group.

Acknowledgement Authors gratefully acknowledges the Principal of the institute Dr. P. B. Mane, Prof. D. K. Shedje, Head of Electronics Engg, Dept. and entire staff of Post Graduate section of the Institute for their support and guidance in the present work.

References

1. Mattson M (2004) Pathways towards and away from Alzheimer's disease. *Nature* 430:631–639
2. Meek PD, McKeithan K, Shumock GT (1998) Economics considerations of Alzheimer's disease. *Pharmacotherapy* 18:68–73

3. World Health Organization and Alzheimer's disease International (2012) Dementia: a public health priority. Technical report
4. De Carli C, Kaye JA, Horwitz B, Rapoport SI (1990) Critical analysis of the use of computer-assisted transverse axial tomography to study human brain in aging and dementia of the Alzheimer type. *Neurology* 40:872–883
5. Förstl DH, Kurz A (1999) Clinical features of Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* 249:288–290
6. Bianchetti A, Trabucchi M (2001) Clinical aspects of Alzheimer's disease. *Aging Clin Exp Res* 13:221–230
7. Duthey B (2013) Alzheimer's disease and other dementias. In: *A public health approach to innovation*, pp 1–74
8. Kulkarni N, Bairagi VK (2014) Diagnosis of Alzheimer disease using EEG signals. *Int J Eng Res Technol (IJERT)* 3(4):1835–1838
9. Dauwels Justin, Vialatte Francois, Cichocki Andrzej (2010) Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr Alzheimer Res* 7(6):487–505
10. Cichocki A, Shishkin SL, Musha T, Leonowicz Z, Asada T, Kurachi T (2005) EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clin Neurophysiol* 116(3):729–737
11. Melissant C, Ypma A, Frietman EE, Stam CJ (2005) A method for detection of Alzheimer's disease using ICA enhanced EEG measurements. *Artif Intell Med* 33(3):209–222
12. Jeong J (2005) EEG dynamics in patients with Alzheimer's disease. *Artif Intell Med* 33(1):202–222
13. Poil SS, De Haan W, Flier WM, Mansvelder HD, Scheltens P, Linkenkaer-Hansen K (2013) Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Front Aging Neurosci* 5:1–6 (**Article ID 58**)
14. Cassani R, Falk TH, Fraga FJ, Kanda PAM, Anghinah R (2014) The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis. *Front Aging Neurosci* 6(55):1–13
15. Traimboli LR, Falk TH, Fraga FJ, Anghinah R, Lorena AC (2011) EEG sepectro-temporal modulation energy: a new feature for automated diagnosis of Alzheimer disease. In: *Proc Intl Conf IEEE-EMBC, Boston, USA*, pp 3828–3831
16. Staudinger T, Polikar R (2011) Analysis of complexity based EEG features for diagnosis of Alzheimer disease. In: *Proc Intl Conf IEEE-EMBC, Boston, USA*, pp 2033–2036
17. Jarque C, Bera A (1987) A test for normality of observations and regression residuals. *Int Statist Rev* 55(2):163–172
18. Rueda Andrea, Gonzalez Fabio A (2014) Extracting salient brain patterns for imaging based classification of neurodegenerative diseases. *IEEE Trans Med Imaging* 33(6):1262–1274