The use of kurtosis de-noising for EEG analysis of patients suffering from Alzheimer's disease[1]

Gang Wang[a], Simon J. Shepherd[b], Clive B. Beggs[b], Nini Rao[a] and Ying Zhang[a,*]

[a]*University of Electronic Science and Technology of China, 611731, Chengdu, China*
[b]*Medical Biophysics Group, School of Engineering, Design and Technology University of Bradford, BD7 1DP, Bradford, UK*

Abstract. The use of electroencephalograms (EEGs) to diagnose and analyses Alzheimer's disease (AD) has received much attention in recent years. The sample entropy (SE) has been widely applied to the diagnosis of AD. In our study, nine EEGs from 21 scalp electrodes in 3 AD patients and 9 EEGs from 3 age-matched controls are recorded. The calculations show that the kurtoses of the AD patients' EEG are positive and much higher than that of the controls. This finding encourages us to introduce a kurtosis-based de-noising method. The 21-electrode EEG is first decomposed using independent component analysis (ICA), and second sort them using their kurtoses in ascending order. Finally, the subspace of EEG signal using back projection of only the last five components is reconstructed. SE will be calculated after the above de-noising preprocess. The classifications show that this method can significantly improve the accuracy of SE-based diagnosis. The kurtosis analysis of EEG may contribute to increasing the understanding of brain dysfunction in AD in a statistical way.

Keywords: Electroencephalogram (EEG), Alzheimer's disease (AD), kurtosis, sample entropy (SE)

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia in people over the age of 65 years [1-4]. It is a neurodegenerative disease characterized by progressive cognitive deterioration, decline in the ability to perform daily activities, and behavioral changes. These symptoms are accompanied by histological changes in the brain, characterized by the formation of extracellular amyloid plaques and the intracellular deposition of neurofibrillary tangles [5]. AD can be difficult to diagnose, particularly in the early stages, and definite diagnosis is only possible by necropsy [6]. Consequently, while the person is alive, it is only possible to make a diagnosis of 'possible' or 'probable' AD. It was shown in [1] that clinical diagnosis of possible AD has accuracy rates of about 50%. The low diagnosis rate

makes it urgent to develop methodologies for accurately diagnosing AD in patients, particularly during the early stages when there is the best chance to treat the symptoms of the disease.

Although a definitive diagnosis can be made after death, in recent decades, the electroencephalogram (EEG), neuroimaging diagnostic tools such as PET, SPECT, and fMRI have proved useful in aiding the diagnosis of AD. This paper is focused on EEG. Indeed, several studies have associated changes in quantitative spectral electroencephalogram (q-EEG) with cognitive decline in patients with AD [7]. AD patients are characterized by EEG abnormalities which shift the power spectrum to the lower frequencies and a decrease in coherence amongst the cortical areas [6], although in the early stages of the disease, EEG plots may exhibit normal frequencies [8]. So although useful, EEG diagnosis can be inconclusive and careful interpretation is required in order to identify AD using this method. Indeed, such is the difficulty of accurately diagnosing AD using EEG that it has led researchers to seek superior methods for analyzing EEG data. To this end, many statistical, signal processing and nonlinear dynamical methods have been used with varying degrees of success. Some of these methods work as preprocessing or de-noising, such as blind source separation (BSS) [9, 10]. Some of these methods work as feature extraction, such as windowed signal power (WP) [10], bipolarity [11], coherence or synchronization [12, 13], even-related potentials (ERP)-based feature [14], power frequency (PF) [15, 16] and stochastic synchrony [17, 18]. Other methods work as classifier, such as linear discriminant analysis (LDA) [9], sample entropy (SE) [5, 6, 19-22].

Based on above methods, the authors in [23] conducted different methods of the various combinations as: PF+CSP+LDA, WP+CSP+LDA, SE+CSP+LDA, PF+P4+LDA, WP+P4+LDA, SE+P4+LDA, where P4 denotes one EEG electrode. The performances of these methods are usually influenced by some kind of noise, such as muscle and line noise, and the noise shows different influence in different methods, so one tendency is to develop different de-noising method for different classifier. For example, LDA classifier is utilized in [9], and the AMUSE-BSS is proposed to help de-noise, where the components with the lowest linear predictability are removed from raw data. This paper is focused on how to help sample entropy de-noise.

Sample entropy (i.e. estimation of signal complexity), which was first proposed as a method for investigating the dynamics of heart rate and other time series [24, 25], has been widely applied to the diagnosis of AD [5, 19, 20]. In particular, Abasolo *et al.* [5, 19] demonstrated that SE could be used to discriminate AD patients from normal controls. This approach is based on the hypothesis that EEG plots of AD patients exhibit greater regularity than those of age-matched controls. Thus, the EEG entropy exhibited by AD patients should be different from that exhibited by age-matched controls. More recent research suggests that the SE approach may be limited due to signal noise [23]. Then it can be hypothesized that SE might work better in an AD application if suitable de-noising measures were first undertaken.

While a number of de-noising algorithms have been applied to EEG data from AD patients, such as the most relevant components algorithm based on AMUSE [9, 26] and common spatial patterns (CSP) [23, 27, 28], these appear not to be suitable preprocessing tools of SE. From a nonlinear dynamical viewpoint, the regularity of the EEG data collected from AD patients is greater than that from control patients.

When viewed from a statistical signal processing standpoint, the EEG data from controls are so irregular that they can be regarded as near-Gaussian signals, while that from AD patients are so regular that they can be regarded as far-Gaussian signals. Therefore, EEG signals can be evaluated in terms of their level of "Gaussianity", which is quantified by their kurtosis value (which is a measure of non-Gaussianity) [29, 30]. The lower the absolute value of kurtosis, the more near-Gaussian is the signal, and vice versa. If it is applied to EEG data from AD patients, the non-Gaussian components can be

regarded as useful signals while the near-Gaussian component can be regarded as useless noise. Note that the kurtosis is the measure of "Gaussianity". This motivates us to identify AD sufferers from EEG plots using the kurtosis to help denoising.

In this paper, a modified sample entropy (SE) approach was used to evaluate sample EEG plots and found this new methodology to be able to distinguish between patients with and without AD with great accuracy. The 21-electrode EEG data were first decomposed using an independent component analysis (ICA) method [31-33], and then the data were sorted by their kurtosis value in descending order. Finally, the subspace of EEG signal was reconstructed using back projection of only the first several components with far-Gaussian property. After the de-noising preprocess, the SE approach was applied to the EEG data.

Table 1
Kurtosis of EEGs of controls for all electrodes

kurtosis	EEG1	EEG2	EEG3	EEG4	EEG5	EEG6	EEG7	EEG8	EEG9
Fp1	0.6828	0.0985	0.8014	0.0680	0.0846	0.0425	11.4655	11.7081	16.1772
Fp2	4.8635	9.3026	6.1694	0.7516	2.8487	0.7829	0.8904	0.2357	0.9938
F3	4.5403	8.8131	4.4459	0.5435	1.2553	0.5818	2.7173	1.6430	1.3271
F4	1.9128	7.6254	4.2684	0.6340	1.3354	2.2668	4.2738	14.1135	3.3162
C3	2.1537	6.2253	7.1602	0.0848	0.1262	1.4552	7.5946	35.9307	10.9414
C4	2.9638	1.9331	4.2397	0.0494	0.1701	7.6063	3.6163	10.3625	5.3654
P3	1.9374	6.0670	8.4925	0.1752	0.4533	0.7876	1.4478	0.8565	2.4335
P4	2.2931	2.9128	4.6395	0.8932	2.8330	2.3080	0.4756	0.2502	0.6900
O1	1.3404	1.2470	2.4450	1.3907	1.0204	1.2129	2.7040	0.7212	2.8280
O2	21.6754	5.0561	4.0545	0.2709	0.3598	0.2957	5.8991	4.3587	5.7090
F7	1.9853	0.5176	0.7739	0.4236	0.3373	0.4563	2.6497	1.9802	2.1083
F8	0.3838	0.2931	0.1030	0.2490	0.1514	0.2770	0.6528	1.2226	1.0271
T7	2.3961	0.5298	2.1150	1.9071	0.8121	0.6395	0.7568	0.7805	0.6437
T8	4.0216	5.6961	4.0315	0.2212	0.1878	3.4055	0.3872	0.2411	0.4400
P7	0.9496	0.1995	0.3442	0.0805	0.2841	0.0174	0.3416	0.6345	0.6151
P8	3.5042	7.4554	9.2163	0.8296	-0.0656	0.1233	2.1008	17.0041	4.9070
FZ	0.7913	1.9702	1.0507	0.2636	0.3850	0.5184	1.8433	3.0732	2.3913
CZ	0.2943	0.1514	2.3762	0.0041	-0.0321	-0.1079	2.0886	1.5938	4.5597
PZ	12.3836	0.3773	221.0881	0.4186	0.2591	0.6007	0.6468	0.3631	1.1326
FPZ	1.7342	0.7820	15.3238	0.1985	0.1636	0.2011	0.2548	0.3279	0.5295
OZ	1.0717	0.8227	5.2999	-0.3547	-0.2409	-0.1155	1.2644	1.2625	1.2614
Means	3.8163	2.9518	6.2992	0.3969	0.8625	1.6798	1.5715	2.3819	2.2331

2. Method

2.1. EEG data and its kurtosis

In order to test the algorithm, the EEG data were recorded from three times from 3 AD patients and three age-matched controls, respectively. EEGs were collected from the 21 scalp loci of the international 10-20 system (electrode Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8 FZ, CZ, PZ, FPZ and OZ). A specialist physician to check for eye movement and other artifacts inspected all EEGs. Since the EEG data are not stationary, we divided the EEG data from one person into three parts and each part contains 10 000 data points where the sample rate was 100 Hz. Thus, the 9 EEGs data from AD patients and 9 EEGs data from controls were obtained, and each EEG data have the size of 21×10000.

Kurtosis measures the relative flatness of a distribution. Distributions with positive kurtosis are termed "super-Gaussian", while those with negative kurtosis are "sub-Gaussian". If κ_i denotes the kurtosis of the output y_i $(i = 1,2,\ldots,n)$, M_{4i} the *4*th moment of y_i, and M_{2i} the *2*nd moment of y_i, then the kurtosis can be estimated via the following algorithm [34]:

$$
\kappa_i^+ = \frac{M_{4i}^+}{M_{2i}^+} - 3
$$
$$
M_{4i}^+ = \delta M_{4i} + (1 - \delta)|y_i|^4
$$
$$
M_{2i}^+ = \delta M_{2i} + (1 - \delta)|y_i|^2
$$

(1)

where $\delta (0 << \delta < 1)$ is a constant.

Then the kurtosis of the data in each electrode for controls is shown in Table 1 and that of AD patients is shown in Table 2. It can be seen from the two tables that most of the kurtoses of EEGs of AD patients are higher than that of controls. This motivates us to introduce kurtosis-based method to help to de-noise the EEG data.

2.2. ICA for EEG de-noising

Implicit in the concept of de-noising is the expectation that within a noisy signal exists hidden components that impart core information relating to the complex signal under consideration. With respect to EEG data, these components are more sensitive in AD patients than in others, and thus can be considered as useful 'signals'. By comparison, the other components of the EEG data are 'noise' and thus can be considered 'useless signals.' Thus by de-noising, or improving the signal-to-noise ratio (SNR) by filtering off the 'noise,' it should be possible to enhance the performance of subsequent feature extraction and data classification. With regard about this, ICA algorithms [26] can be used for the purpose of filtering. ICA is used to extract independent source signals from their linear mixtures without knowing the mixing parameters. It is a statistical method and has attracted considerable attention amongst the signal processing and neural network community. When applying ICA to EEG analysis, if $x=(x_1, x_2, \ldots, x_m)^T$ denotes an *m*-dimension observation vector, and $s=(s_1, s_2, \ldots, s_n)^T$ an *n*-dimension hidden independent components (ICs) - some of which may be evident in AD patients, then the linear relationship is given by

$$
x = As
$$

(2)

where A is an $m \times n$ unknown matrix, called the mixing matrix. Here it is assumed that m is equal to n, and A is full rank. The basic problem of ICA is then to conduct an $n \times n$ full-rank separating matrix B given just the observation sequence with the restriction that source elements s_i are non-Gaussian independent components. Thus the output vector

$$y = Bx, \quad y = (y_1, y_2, \ldots, y_n)^T \tag{3}$$

provides the estimates of the sources signals. Due to two kinds of inherent indeterminacies in Eq. (1), it is impossible to recover the source signals in exact amplitude or in exact order [26, 30]. Without loss of generality, the source signals are assumed to have unit variance with any scaling factors absorbed into the mixing matrix A. In other words, when a perfect ICA is performed, the global transfer matrix $C=BA$ should be an $n \times n$ generalized permutation matrix. Many algorithms designed for ICA have been proposed, including those employing high order statistics (HOS), such as fast ICA [30] and flexible ICA [34], and those employing second-order statistics (SOS), such as second order blind identifi

Table 2
Kurtosis of EEGs of AD patients for all electrodes

kurtosis	EEG1	EEG2	EEG3	EEG4	EEG5	EEG6	EEG7	EEG8	EEG9
Fp1	25.4143	0.3365	25.3954	0.0950	0.0332	-0.0090	2.6302	0.3716	0.8848
Fp2	9.9425	2.9189	15.3509	20.8243	13.7805	13.2179	6.8940	8.3364	6.2143
F3	11.2298	5.6626	3.5790	6.4451	174.1147	6.9603	24.6065	91.9136	15.6406
F4	2.3398	1.3652	14.6675	0.4203	586.1999	0.0696	5.5400	8.5658	2.8525
C3	15.2431	1.5070	710.9700	5.3769	18.8216	1.8099	5.9622	3.1621	3.9944
C4	1.7673	0.1861	1.0985	1.1406	0.9838	2.4726	1.3227	0.9336	0.6748
P3	11.5065	1.1540	3.0467	0.8154	9.7243	2.3595	3.0037	7.9806	3.5677
P4	4.5861	2.7389	6.3293	4.8936	7.8389	4.2088	2.5446	11.4943	2.8517
O1	5.2264	0.7323	22.2497	1.1920	59.4170	5.0022	0.9179	0.2470	0.4685
O2	28.4244	4.0033	21.8968	0.6144	5.1121	0.4355	1.6995	1.0665	1.6507
F7	11.5029	0.3558	10.9458	0.3430	0.3652	2.7869	3.5040	3.1700	12.3627
F8	0.4262	0.4003	0.9675	0.1333	0.1420	12.2024	1.3489	0.0861	0.8716
T7	1.0529	0.6289	2.0512	6.7394	2.1853	25.8189	1.6907	1.4805	1.6151
T8	2.7040	1.3900	23.7984	8.9052	3.1716	21.9267	2.2837	0.6258	0.7632
P7	4.8916	2.7803	22.9866	27.2410	21.9844	56.3523	0.5131	0.2897	0.3514
P8	27.4883	0.3791	28.5912	1.0208	0.0693	0.0041	23.7412	9.3029	4.6342
FZ	22.5844	1.3900	36.6615	1.3416	-0.0949	-0.0784	3.3059	0.8454	0.5356
CZ	2.7989	1.2735	51.6157	2.2172	1.5684	19.0611	10.6106	1.7113	1.5719
PZ	5.5300	0.1359	11.0895	1.4354	10.4328	6.5352	1.9583	1.1752	0.3318
FPZ	15.2469	1.1504	28.3517	0.2694	1.2227	29.6515	4.7314	1.5520	0.7005
OZ	28.9541	0.1200	15.2411	0.1499	-0.0240	-0.0222	4.6474	2.3591	1.2358
Means	11.0842	2.4350	30.3846	2.3811	136.8885	2.9083	3.2937	3.4279	3.1594

(SOBI) [35], the AMUSE algorithm [26] and mean square cross prediction error (MSCPE) algorithm [8, 31]. Generally speaking, the HOS algorithms are more time-consuming than SOS, therefore in order to save time AMUSE, one of the SOS algorithms, was used in the experiments below.

AMUSE algorithm is similar to PCA or singular vale decomposition (SVD). AMUSE utilizes PCA/SVD two times. In the first step, PCA/SVD can be applied for whitening. Whitening removes the second-order dependence among the observations x. If the autocorrelation matrix $R_x = E\{xx^T\}$ has the SVD, $R_x = PDP^T$, where D=diag $(d_1, d_2, \ldots, d_n)$ is a diagonal matrix of corresponding eigenvalues, the suitable whitening matrix will be $W = PD^{-1/2}P^T$. Then $\tilde{x} = Wx$ is a standardized vector satisfying $E\{\tilde{x}\tilde{x}^T\} = I$. In the second step, SVD/PCA is utilized as

$$R(x_1) = E\{\tilde{x}(k)\tilde{x}^T(k-1)\} = U\Sigma V^T \tag{4}$$

where Σ is diagonal matrix, and U, V are orthogonal matrices. Then, a global separating matrix is estimated as $B = U^T W$ [26].

$$y = Bx = U^T Wx$$
$$y = (y_1, y_2, \ldots, y_n)^T \tag{5}$$

where the kurtosis of y_i is sort as follows for simplicity:

$$\kappa_1 > \kappa_2 > \cdots > \kappa_n \tag{6}$$

As for the separated signal space, it can be divided into near-Gaussian and far-Gaussian subspaces according to the kurtosis of separated signals. Denote by y_{ICA} the subspace spanned by the far-Gaussian components

$$y_{ICA} = (y_1, y_2, \ldots, y_d)^T \tag{7}$$

where d is the number of the far Gaussian components. Then the near-Gaussian noise can be filtered by back projecting observation x onto the far-Gaussian subspace

$$x_{ICA} = y_{ICA}^T y_{ICA} x \tag{8}$$

In this experiment, the far-Gaussian subspace containing five components with highest positive kurtosis has been selected:

$$m = n = 21$$
$$d = 5 \tag{9}$$

As for the selection of the number of the far Gaussian components (d), when d is less than two or more than eight, worse performance will be achieved than in the case of raw data. When d is ranged from four to six, the results of classification will be better. In this paper, the number of the far Gaussian components is adopt as d=5.

2.3. Calculation of SE

In the detection and diagnosis of AD, the nonlinear characteristics of the EEG time series has been used for some time, such as the correlation dimension (D2) of the EEG time series [36], the first Lyapunov exponent (L1) to distinguish between probable AD and control subjects [20] and SE. However, SE can work well instead of D2 and L1. This point has been proven by [5] and [19]. Moreover, the results of [5] indicate that SE is a useful feature for discriminating between AD subjects and controls.

Formally, given N data points form a time series $x(n) = x(1), x(2), \cdots, x(N)$, the basic idea of SE is to approximate the following statistic [5, 23]:

$$SE(m,r) = -ln \left[\frac{\Pr\left(d\left[x_{m+1}(i), x_{m+1}(j)\right] \leq r\right)}{\Pr\left(d\left[x_m(i), x_m(j)\right] \leq r\right)} \right] \tag{10}$$

where $x_m(i) \in R^m$ is the vector extracted from the time series $x(t)$ as follows:

$$x_m(i) = \left[x(i), x(i+1), \cdots, x(i+m-1)\right] \tag{11}$$

and $d\left[x_m(i), x_m(j)\right]$ is defined as:

$$d\left[x_m(i), x_m(j)\right] = \max_{k=[1,m]} \left(\left|x(i+k-1) - x(j+k-1)\right|\right) \tag{12}$$

In practice, the probability values in expression (10) were approximated by averaging over the entire time series. Because of the nonlinear character of EEG signal, SE can be utilized as a powerful tool to measure the regularity of AD patients' EEG. Furthermore, The Matlab tools for calculating sample entropy provided at http://www.physionet.org/physiotools/sampen/matlab/ are used in the experiments. The calculation of Eq. (10) including the approximation of the probability values are performed through the Matlab codes in the above website.

2.4. Testing for difference between two means

It is frequently in biology that analysis is undertaken of two populations whose distributions overlap, and statistics is used to determine if the populations are significantly different from each other. In situations like this useful technique that can be applied is the t-test, which can be used to evaluate the statistical difference between AD patients and the controls, after the SE values have first been obtained from the EEG signals.

If $Z1$ denotes the set of SE values from one electrode in AD patients and $Z2$ denotes the set of SE values from the controls, the null hypothesis in this case will be: "Z1 and Z2 are equal on average". The t-test therefore allows us to assign a probability level to the likelihood that the null hypothesis is true. The calculation is straightforward, requiring us to calculate both the mean of each set and a measure of the variation of each mean. Here is the formula for calculating t-value

$$t_{value} = \frac{mean(Z1) - mean(Z2)}{z_p \left(1/n_1 + 1/n_2\right)^{0.5}}$$

$$z_p^2 = \frac{\left[(n_1 - 1)Var(Z1) + (n_2 - 1)Var(Z2)\right]}{\left(n_1 + n_2 - 1\right)}$$

(13)

where operator mean() denotes the means of the sets, Var() denotes the sample variances of the two groups, and n_1 and n_2 are the sample sizes of each set.

After getting the *t*-value, the *p*-value can be further calculated. In the experiment, the threshold of *p*-value as [5], i.e. 0.01, was selected. If the *p*-value was lower than 0.01ct the null hypothesis of no difference would be rejected, and infer that the differences between the mean values were considered significant. The calculation of the *p*-value, refer to as the website: http://www.danielsoper.com/statcalc/calc08.aspx, where the degrees of freedom for our experiments are 16 (9 samples for AD + 9 samples for controls - 2).

Combining the kurtosis, AMUSE, SE algorithm and *t*-value test, the new EEG analysis method can be summarized as follows:

Step 1: Each EEG expressed as *x* was decomposed into 21 independent components by AMUSE algorithm using Eq. (5).

Step 2: Calculate the kurtoses of the separated components, and sort them in descending order as Eq. (6).

Table 3
Average kurtosis of EEGs for AD patients and controls for all channels

	AD Patients		Controls		Statistical Analysis	
	Mean	STD	Mean	STD	t_value	P_value
Fp1	3.074	5.1475	2.088	3.0492	0.4945	0.628
Fp2	20.40	22.71	3.196	2.478	2.259	0.0382
F3	72.44	174.7	3.661	2.695	1.181	0.255
F4	149.5	431.9	5.480	4.537	0.9998	0.332
C3	66.56	139.3	4.217	4.093	1.342	0.198
C4	5.949	5.442	5.957	5.127	0.0031	0.998
P3	9.058	17.00	3.039	3.806	1.037	0.315
P4	6.579	8.009	3.575	3.536	1.029	0.319
O1	18.19	43.26	1.136	0.7301	1.182	0.254
O2	6.755	9.679	2.469	1.941	1.302	0.211
F7	4.310	5.629	0.827	0.7421	1.840	0.0843
F8	4.680	9.663	0.4192	0.3787	1.321	0.205
T7	3.531	4.816	0.4702	0.368	1.901	0.0755
T8	7.163	29.25	0.9911	1.151	1.051	0.309
P7	39.41	110.8	0.6218	0.6416	1.050	0.309
P8	7.163	1.516	1.490	11.28	1.469	0.161
FZ	8.009	17.33	0.8348	0.5256	1.241	0.232
CZ	5.578	6.669	10.51	0.5327	2.273	0.0371
PZ	7.240	16.23	9.192	18.24	0.2399	0.813
FPZ	3.578	3.917	1.014	1.760	1.791	0.0923
OZ	4.049	7.718	0.5978	0.5997	1.337	0.200

Step 3: Select the first five components for the far-Gaussian subspace y_{ICA} as Eq. (7), project the EEG x onto this subspace as Eq. (8), and obtain x_{ICA}.

Step 4: Calculate the SE of the 21 electrodes (x_{ICA}).

Step 5: Calculate the p-value between two means

3. Experiments

The kurtoses for the 21 electrodes Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8 FZ, CZ, PZ, FPZ and OZ were first estimated. Secondly, the SE (m=1, r=0.25 times the standard deviation of the data) for all 21 electrode under different conditions (i.e. without de-noising, relevant de-noising and kurtosis-based de-noising) were calculated. Finally, the receiver operating characteristic (ROC) curves for the different conditions were plotted.

3.1. The kurtoses of the EEG from AD patients and controls

Table 3 summarized the average kurtosis of EEGs for the AD patients and controls for all channels. The mean values of almost all electrodes except C4 and PZ are much higher than that of normal controls. However, the standard deviation (STD) of AD EEG's kurtosis is very high, and none of the *p*-values in 21 electrodes are lower than 0.01. Thus, their mean values might be not so significantly dif-

Table 4
Average SE (m=1, r=0.25) of EEGs for AD patients and controls for all channels without denosing

	AD Patients		Controls		Statistical Analysis	
	Mean	STD	Mean	STD	*t*_value	*p*_value
Fp1	1.8583	0.0862	1.891	0.0839	0.8164	0.426
Fp2	1.4919	0.1823	1.6727	0.3242	1.4583	0.164
F3	1.5156	0.1956	1.6663	0.3152	1.2188	0.241
F4	1.7575	0.1398	1.6395	0.2115	1.3961	0.182
C3	1.744	0.2131	1.6741	0.2436	0.6484	0.526
C4	1.8112	0.1299	1.6651	0.2116	1.7641	0.968
P3	1.8373	0.0886	1.8233	0.1371	0.2562	0.801
P4	1.7805	0.0676	1.7914	0.1744	0.1747	0.864
O1	1.8558	0.1652	1.8993	0.0353	0.7719	0.451
O2	1.8157	0.1516	1.8272	0.133	0.1712	0.866
F7	1.8551	0.1367	1.9181	0.0259	1.3582	0.193
F8	1.8989	0.0823	1.9416	0.0243	1.4951	0.154
T7	1.8699	0.0808	1.9388	0.0192	2.4902	0.0241
T8	1.884	0.0661	1.9184	0.0581	1.1724	0.258
P7	1.9057	0.0661	1.9207	0.034	0.6083	0.552
P8	1.9197	0.2339	1.9211	0.0325	0.0172	0.986
FZ	1.7689	0.4489	1.9146	0.0271	0.9719	0.346
CZ	1.7997	0.1751	1.9306	0.0375	2.1927	0.0435
PZ	1.8918	0.092	1.7882	0.3886	0.7783	0.448
FPZ	1.8678	0.1114	1.915	0.0492	1.1625	0.262
OZ	1.8207	0.1606	1.9269	0.0318	1.9466	0.0693

ferent as to be used to a classification. This shows on the viewpoint of statistical signal processing that the regularity of AD seems more like super-Gaussian than normal controls.

3.2. SE under different conditions

Tables 4-6 summarized the average SE values of EEGs for AD patients and controls for all channels under no de-noising, relevant de-noising and kurtosis de-noising conditions, respectively. From the data in these tables, it can be seen that for most of the electrodes p-value>0.05, with the lowest p-value in Table 4 being only 0.0241 (i.e. for electrode T7), and 0.0246 for electrode F3 in Table 5. This indicates that the relevant denoising method made no improvement, and there are no significantly different mean values in all the electrodes for the two methods. If the accuracy for classification using P4 electrode as [5, 23] under this circumstance is calculated, the accuracy will be about 67%, which although consistent with that reported in [23] is not a good result. However, as can be seen in Table 7, if T7 from Table 4, or F3 from Table 5 are selected, then the accuracy of the diagnosis improves to around 80%.

Comparison of the data in Table 6 with that in Table 4 and Table 5 reveals that the kurtosis-based de-noising method resulted in a significant improvement in performance. The lowest p-value in Table 6 reached 0.000074, which indicates significant differences between the mean SE values of AD patients and normal controls. Table 7 summarizes the accuracies of different methods, and shows that the accuracy of kurtosis de-noising method is improved from about 80% to almost 95%.

Table 5
Average SE (m=1, r=0.25) of EEGs for AD patients and controls for all channels after relevant denosing [9]

	AD Patients		Controls		Statistical Analysis	
	Mean	STD	Mean	STD	t_value	p_value
Fp1	1.8652	0.1143	1.9118	0.0408	1.1536	0.266
Fp2	1.5331	0.2352	1.7254	0.2059	1.8458	0.0835
F3	1.5028	0.1801	1.7221	0.1948	2.4803	0.0246
F4	1.7668	0.1503	1.6690	0.1693	1.2961	0.213
C3	1.7199	0.1942	1.6915	0.2564	0.2654	0.794
C4	1.7585	0.2253	1.6952	0.2298	0.5904	0.563
P3	1.7667	0.1202	1.8294	0.1133	1.1396	0.271
P4	1.7317	0.1117	1.7979	0.1376	1.1203	0.279
O1	1.8591	0.1791	1.8691	0.0915	0.1490	0.883
O2	1.8751	0.1336	1.8255	0.1170	0.8374	0.414
F7	1.8839	0.1501	1.9094	0.0567	0.4764	0.640
F8	1.8866	0.1489	1.9193	0.0435	0.6326	0.536
T7	1.8281	0.1905	1.9200	0.0323	1.4273	0.173
T8	1.8864	0.0818	1.9044	0.0472	0.5706	0.576
P7	1.9016	0.0908	1.8981	0.0472	0.1018	0.920
P8	1.9142	0.0735	1.9289	0.0348	0.5442	0.594
FZ	1.7463	0.4002	1.9246	0.0324	1.3319	0.202
CZ	1.8077	0.1723	1.9305	0.0441	2.0709	0.549
PZ	1.8819	0.0947	1.7842	0.3953	0.7218	0.481
FPZ	1.8790	0.1431	1.9149	0.0645	0.6867	0.502
OZ	1.8603	0.2092	1.9141	0.0652	0.7368	0.472

Furthermore, the ROC curves were plot for the no de-noising, relevant de-noising and kurtosis de-noising methods in Figure 1. ROC curves were obtained by plotting the sensitivity values on the *y*-axis against their equivalent {1-specificity} values for all the available cut-off points (in this case, the kurtoses or SE values) on the *x*-axis. The sensitivity value represents the percentage of patients with a diagnosis of AD who are tested as positive, or the true positive rate. The specificity means the proportion of controls correctly recognized, or the true negative rate [5]. Figure 1 reveals that there are no significant differences between the no de-noising method and relevant de-noising method, and that the area under ROC curve of our proposed method (kurtosis de-noising) is much larger than those of the other two methods. Table 7 summarized the corresponding values under ROC curves, where the accuracy of the proposed method is more than 90%.

Generally speaking, the larger the area under ROC curve, the better the performance [5]. With values between 0.90 and 1, the precision of the diagnostic test is considered to be excellent, good for values between 0.80 and 0.89, fair for values between 0.70 and 0.79, poor for values between 0.60 and 0.69, and bad for values between 0.50 and 0.59. As a result, the kurtosis-based de-noising method can be considered excellent for electrode F3, the relevant de-noising method can be good for F3, the SE method without de-noising can be good for T7 and poor for P4.

As for the classification, the simple threshold method was adopted. The classifications were depicted in Figure 2, where 'o' denotes normal controls, '*' denotes AD patients, and '-' denotes threshold. The corresponding accuracies were shown in Table 7. Both visual detections in Figure 2 and accuracies in Table 7 show that the proposed method performs better than the others.

Table 6

Average SE (m=1, r=0.25) of EEGs for AD patients and controls for all channels after kurtosis denosing

	AD Patients		Controls		Statistical Analysis	
	Mean	STD	Mean	STD	*t*_value	*p*_value
Fp1	1.6623	0.1788	1.8403	0.064	2.8113	0.0125
Fp2	1.6371	0.1269	1.7537	0.172	1.6375	0.121
F3	1.5376	0.1414	1.8096	0.0616	5.2874	***0.000074***
F4	1.7002	0.1198	1.6677	0.2032	0.4138	0.685
C3	1.7048	0.1758	1.6915	0.249	0.1307	0.898
C4	1.6363	0.1991	1.6664	0.2183	0.3061	0.763
P3	1.7359	0.1191	1.8161	0.1333	1.3459	0.197
P4	1.7037	0.1123	1.7672	0.1386	1.0669	0.302
O1	1.6983	0.1412	1.8551	0.0755	2.9385	***0.00964***
O2	1.7346	0.145	1.7494	0.1522	0.2122	0.835
F7	1.7548	0.1159	1.7695	0.1443	0.2382	0.915
F8	1.6894	0.1555	1.8018	0.1432	1.5957	0.13
T7	1.7833	0.1166	1.7565	0.1973	0.3505	0.731
T8	1.7327	0.1563	1.8602	0.1062	2.0242	0.06
P7	1.6797	0.1836	1.8269	0.094	2.14	0.0481
P8	1.7158	0.1305	1.8048	0.101	1.6172	0.125
FZ	1.668	0.3549	1.8112	0.0911	1.1731	0.258
CZ	1.7156	0.1597	1.8157	0.1683	1.294	0.214
PZ	1.7274	0.1247	1.6477	0.4114	0.556	0.586
FPZ	1.7128	0.1597	1.8566	0.0649	2.5023	0.0236
OZ	1.719	0.1909	1.8201	0.1257	1.3259	0.203

Table 7

Accuracy results under three different conditions

Method	Electrode	Threshold	Accuracy	Area under ROC curve
No de-noising	P4	1.800	66.7%	63.6%
No de-noising	T7	1.904	83.3%	81.5%
Relevant de-noising	F3	1.700	77.8%	80.2%
Kurtosis de-noising	F3	1.700	94.4%	97.5%

Furthermore, the comparisons were done with other de-noising algorithms, ASUME [9] and common spatial patterns (CSP) [27], in Table 8, which shows that the proposed algorithm (Kurtosis de-noising) works better than the other two methods.

4. Conclusions

In this study, the Gaussianity of the EEG signal from a number of AD patients and normal controls using kurtosis methods were analyzed. The average kurtosis of the EEG signals from the AD patients was found to be much higher than that of the controls. This implies that the EEG signals from individuals with AD contain different information, which if identified could prove useful in diagnosing the condition. The SE approach is based on the hypothesis that the regularity of EEG signals form AD patients is higher than that of age-matched controls, with the result that the entropy of this group's EEG signals is lower than that of age-matched controls - indicating an abnormal dynamic within the AD group. However, given that the SE approach appears to be limited by signal noise [23], a kurtosis-based de-noising method was used as a preprocessing step of SE for AD diagnosis. The regularity achieved appeared to be super-Gaussian. The statistical analysis was consistent with the nonlinear dynamic study, suggesting that both embodied the property found in the AD EEG signals. Classification of the results verified the efficacy of the approach and it may be that the new method has great potential in the early diagnosis and prediction of AD.
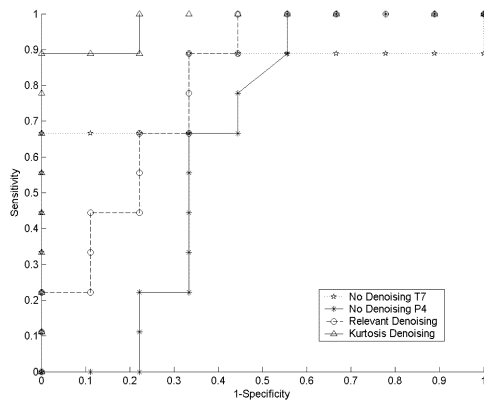

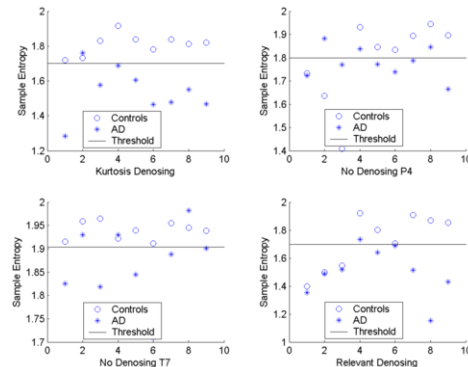Fig. 1. Receiver operating characteristic (ROC) curves different methods.

Fig. 2. Threshold classification for the SE values with for the SE values for different methods in Table 7.

Table 8
Comparison results with ASUME [9] and CSP [27]

Method	Electrode	Accuracy
AMUSE [9]	F3	78.7%
CSP [27]	F3	75.5%
Kurtosis de-noising	F3	94.4%

References

[1]   R. Turner, Biomarkers of Alzheimer's disease and mild cognitive impairment: Are we there yet? Experimental Neurology **183** (2003), 7–10.

[2]   C. Chiang, J.H. Steen and C. Chan, A resident's behavior simulation model for nursing home healthcare services, Bio-Medical Materials and Engineering **24** (2014), 69–75.

[3]   F. Fraga, T. Falk, P. Kanda and R. Anghinah, Characterizing Alzheimer's disease severity via resting-awake EEG amplitude modulation analysis, PLoS One, **8** (2013), e72240, doi:10.1371/journal.pone.0072240.

[4]   R.M. Chapman, et al., Brain event-related potentials: Diagnosing early-stage Alzheimer's disease, Neurobiology of Aging **28** (2007), 194–201.

[5]   D. Ab´asolo, et al., Entropy analysis of the EEG background activity in Alzheimer's disease patients, Physiological Measurement **27** (2006), 241–253.

[6]   J. Jeong, EEG dynamics in patients with Alzheimer's disease, Clinical Neurophysiology **115** (2004), 1490–1505.

[7]   K. Bennys, G. Rondouin, C. Vergnes and J. Touchon, Diagnostic value of quantitative EEG in Alzheimer's disease, Neurophysiologie Clinique **31** (2001), 153–160.

[8]   O.N. Markand, Organic brain syndromes and dementias, in: Current Practice of Clinical Electroencephalography, D.D. Daley and T.A. Pedley, Eds., Raven Press, New York, 1990, pp. 4001–4423.

[9]   A. Cichocki, et al., EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease, Clinical Neurophysiology **116** (2005), 729–737.

[10] F.B. Vialatte, J. Solé-Casals, M. Maurice, C. Latchoumane, N. Hudson, S. Wimalaratna, J. Jeong and A. Cichocki, Improving the quality of EEG data in patients with Alzheimer's disease using ICA, Lecture Notes in Computer Science **5506** (2009), 979–986.

[11] T. Musha, et al., A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease, Clinical Neurophysiology **113** (2002), 1052–1058.

[12] T. Koenig, et al., Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment, Neurobiology of Aging **26** (2005), 165–171.

[13] F. Vecchio, et al., Sources and coherence of cortical EEG rhythms could predict progression from mild cognitive impairment to Alzheimer disease, Clinical Neurophysiology **117** (2006), (Suppl. 1) 1–2.

[14] R. Chapman, et al., Brain event-related potentials: diagnosing early-stage Alzheimer's disease, Neurobiology of Aging **28** (2007), 194–201.

[15] K. Van der Hiele, et al., EEG and MRI correlates of mild cognitive impairment and Alzheimer's disease, Neurobiology of Aging **6** (2006), doi:10.1016/j.neurobiolaging.2006.06.006.

[16] F. Vialatte, et al., Blind source separation and sparse bump modeling of time frequency representation of EEG signals: New tools for early detection of Alzheimer's disease, IEEE Workshop on Machine Learning for Signal Processing, 2005, 27–32

[17] S.T. DeKosky and K. Mare, Looking backward to move forward: Early detection of neurodegenerative disorders, Science **302** (2003), 830–834.

[18] J. Dauwels, F. Vialatte, T.M. Rutkowski and A. Cichocki, Measuring neural synchrony by message passing, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, 2007, NIPS2007.

[19] D. Ab´asolo, et al., Analysis of regularity in the EEG background activity of Alzheimer's disease patients with approximate entropy, Clinical Neurophysiology **116** (2005), 1826–1834.

[20] J. Jeong, et al., Nonlinear dynamic analysis of the EEG in patients with Alzheimer's disease and vascular dementia, Clinical Neurophysiology **18** (2001), 58–67.

[21] W. Pritchard, et al., EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures, Electroencephalography and Clinical Neurophysiology **92** (1994), 118–130.

[22] J-H Park, S. Kim, C-H Kim, A. Cichocki and K. Kim, Multiscale entropy analysis of EEG from patients under different pathological conditions, Fractals **15** (2007), 399–404.

[23] W.L. Woon, A. Cichocki, F. Vialatte and T. Musha, Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings, Physiological Measurements **28** (2007), 335–347.

[24] D.E. Lake, J.S. Richman, M.P. Griffin and J.R. Moorman, Sample entropy analysis of neonatal heart rate variability, American Journal of physiology, Regulatory, Integrative and Comparative Physiology **283** (2002), 789–797.

[25] J.S. Richman and J.R. Moorman, Physiological time series analysis using approximate entropy and sample entropy, American Journal of Physiology, Regulatory, Integrative and Comparative Physiology **278** (2000), 2039–2049.

[26] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, Wiley, New York, 2003.

[27] H. Ramoser, et al., Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Transaction on Rehabilitation Engineering **8** (1998), 441–446.

[28] G. Blanchard and B. Blankertz, BCI Competition 2003: Data set: IIa. Spatial patterns of self-controlled brain rhythm modulations, IEEE Transactions on Biomedical Engineering **51** (2004), 1062–1066.

[29] A. Delorme, T. Sejnowski and S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis, Neuroimage **34** (2007), 1443–1449.

[30] A. Hyvärinen and E. Oja, Independent component analysis: Algorithm and applications, Neural Networks **13** (2000), 411–430.

[31] G. Wang, C. Li and L. Dong, Noise estimation using mean square cross prediction error for speech enhancement, IEEE Transactions on Circuits and Systems-I **57** (2010), 1489–1499.

[32] G. Wang, N. Rao, S. Shepherd and C. Beggs, Extraction of desired signal based on AR model with its application to atrial activity estimation in atrial fibrillation, EURASIP Journal on Advances in Signal Processing **2008** (2008), Article ID 728409.

[33] G. Wang, N. Rao and Y. Zhang, Atrial fibrillatory signal estimation using blind source extraction algorithm based on high-order statistics, Science in China, Series F: Information Sciences **51** (2008), 1572–1584.

[34] S. Choi, A. Cichocki and S. Amari, Flexible independent component analysis, Journal of VLSI Signal Processing **26** (2000), 25–38.

[35] J.F. Cardoso and A. Souloumiac, Blind beamforming for non-Gaussian signals, IEE Proceedings F, Radar and Signal Processing **140** (1993), 362–270.

[36] F.D. Wong, L.S. Chao and X.D. Zeng, A supportive attribute-assisted discretization model for medical classification, Bio-Medical Materials and Engineering **24** (2014), 289–295.