

COMPARISON OF METHODS FOR EARLY DETECTION OF ALZHEIMER'S DISEASE.

Cindy Goh¹, Emmanuel Ifeakor¹, Tracey Cassar², Charles-Francois V. Latchoumane^{2,3}, Cristin Bigan⁴, Geoffrey Henderson¹, Kenneth Camilleri², Simon Fabri², Jaeseung Jeong³, Nigel Hudson⁵, Paolo Capotosto⁶, Sunil Wimalaratna⁷ and Mircea Besleaga⁴

¹School of Computing, Communications and Electronics, University of Plymouth, Plymouth PL4 8AA, U.K.

²iBERG, Faculty of Engineering, University of Malta, Msida MSD06, Malta

³Department of BioSystems, Korea Advanced Institute of Science and Technology (KAIST), South Korea

⁴ Faculty of Engineering, Ecological University of Bucharest, Romania

⁵ Department of Neurophysiology Derriford Hospital, Plymouth PL1, UK

⁶ Department of Human Physiology and Pharmacology, University of Rome La Sapienza, Italy

⁷ Department of Neurology, Radcliffe Infirmary, Oxford, UK

email: cindy.goh@plymouth.ac.uk

Abstract: In this paper, six methods for early detection of AD - fractal dimension (FD), source localization (SL), Hjorth analysis, cross mutual information (CMI), pdf of zero-crossing intervals (ZCI) and power spectrum (PS) are compared. We selected these methods because they were relatively insensitive to artefacts and gave promising results. The methods were applied to the EEGs of 38 mild AD patients and 45 normal subjects, to extract markers for early detection of AD. The datasets were obtained from three different countries and hence provided a more rigorous comparison. The performances of each method were measured using ROC analysis, sensitivity, specificity, classification accuracy, z-score and Area under ROC curve. Results showed that indices found using time domain methods, such as FD and ZCI outperform those obtained from frequency analysis. In particular, ZCI had the best overall performance for at least 50% of the datasets. We found that, although the PS results tend to agree with SL and ZCI, it is more sensitive to shifts in the alpha-theta rhythms. Results also show that FD and ZCI could potentially be used for early detection as their performances outperformed those used in current clinical AD diagnosis (sensitivity > 55%, specificity > 83%).

Keywords: Alzheimer's disease, Dementia, Early Detection, Fractal dimension, EEG, Zero-crossing intervals, Hjorth index, Cross mutual information, Source localization, Power spectrum, Sensitivity, Specificity, ROC.

1 INTRODUCTION

Dementia is tremendously costly to patients, their families, and society. Alzheimer's disease (AD) is the most common form of dementia and is increasing in prevalence. It has been reported that more than 20 million people worldwide are affected by the disease [1]. A better understanding of the disease, its treatment and diagnostic measures could unravel important findings to advance preventing or even halting the disease and help reduce the number of sufferers. However, the diagnosis of AD is a complex task involving many parameters and the mild symptomatology associated with earlier stages of the disease are difficult to detect. It has been reported in a study

of cortical atrophy [2] that the delay between the actual onset of and clinical diagnosis of AD using current clinical criteria lies between three to five years. By which time, sufferers are at an intermediate to late stage of progression.

At present, several markers are used clinically for the detection and monitoring of treatment progress in AD. Studies in neurofibrillary tangles (NFT) and amyloid plaques [3] are important and could contribute to effective treatment and eventually finding a cure for AD. However, they are conducted post-mortem. Neurochemical alterations in blood, cerebral spinal fluid (CSF) and tissue samples [4] have also been investigated. They involved highly invasive procedures and are used only in advance stages when a diagnosis of probable AD is made. By which time, sufferers have already experienced appreciable losses in cognition and functioning. Significant efforts are also made in neuroimaging techniques such as MRI and PET. Despite its advantages, it is costly and not suitable for routine screening of the large at-risk population. Therefore, there is an urgent need for the development of effective, low cost, non-invasive and reasonably accurate screening markers to detect AD at earlier stages.

Since its first use by Hans Berger in 1924, the electroencephalogram (EEG) have been used to provide additional support in AD diagnosis including correlating results from imaging, biochemical and psychometric tests. A key to the successful use of EEG for early detection of AD lies in the development of robust methods for extraction of markers to accurately detect the disease. Research in this area is active and a large array of methods now exist, including both time and frequency, linear and non-linear analysis techniques [5,6].

Amongst other conventional techniques, the most established is frequency spectral analysis and many studies show that by looking at the EEG power density at selected frequency bands, a "slowing of the EEG" is observed in the alpha rhythms of AD patients. The slowing of EEG rhythms have been considered as an indirect affirmation of deficits in neurotransmitters, particularly acetylcholine, which occurs at early stage AD [7]. Several nonlinear dynamical analysis (NDA), typically those involving time embedding techniques, have been

reported to reveal information about the EEG and its underlying system, which were unavailable from spectral analysis [6, 8]. Previous studies have found that correlation dimension (D_2) and the first Lyapunov exponent (L_1) exhibit lower values when applied to the EEGs of AD patients [9, 10]. This indicates a reduced complexity in the underlying systems generating the EEGs and can be seen as a consequence of neuronal death in AD [6]. Despite promising results, these techniques are sensitive to parameters such as initial conditions, the embedding dimension and time delay. In addition, they required large amounts of data to produce appreciable outcomes, which are not always feasible.

Fractal analysis is an alternative approach to time embedding methods. It works directly on the EEG to quantify the waveform shape complexity. In this context, the FD is a measure of the structural details of the waveform itself [11]. The probability density function (PDF) of the zero-crossing intervals (ZCI) is another method which derives information directly from the EEG. It shares some features as FD and has been successfully applied to the EEG before [12]. It was shown to be robust in the characterization of sleep spindles. In a previous work, we have also applied the ZCIs to the EEG of AD patients and found that this measure provides good separation between AD and normal subjects [11].

Recently, several entropy measures have been applied to characterize the EEG in AD. In [13], sample entropy was used applied to the EEG of AD patients and results show increase regularities in the waveform, indicating abnormalities in its dynamics. However, it is sensitive to the run length and tolerance window used in the algorithm. The concept of mutual information (MI) is another NDA method applied to the EEG for AD diagnosis. MI analysis have shown promise when used to quantify functional segregation and integration in the brain as well as complexity [14]. Recently, it has also been used to measure the amount of information transmission in different cortical areas of the AD brain and results indicate a reduction in functional connectivity [15]. Another set of concept for detecting changes in the EEG of AD patients is the Hjorth descriptors. Based on spectral moments computed from the EEG, three descriptors, namely the activity, mobility and complexity are derived [16]. When applied to the EEG of AD patients, results show significant increase in complexity and a reduction in mobility.

Another potential method for EEG analysis is source localization (SL), which involves the use of inverse techniques to locate the sources generating the EEG. There are several approaches for finding the sources, including the minimum current approach and the low resolution electromagnetic tomography (LORETA) [17]. These methods have shown promising results when applied to the EEG of AD patients [18]. It is evident that many promising methods exist. However, despite promising results and claims about their usefulness, no comparison of the methods has been conducted. It is therefore important to compare key methods to verify these claims and thereby provide more insights into how methods should

be use in the analysis of EEG for AD diagnosis.

The objective of this paper is to rigorously compare the performances of a number of key methods for detecting AD using markers obtained from the EEG.

We have chosen to compare six candidate methods, namely, FD, Hjorth descriptors, SL, ZCIs, cross mutual information (CMI) and power spectrum (PS) analysis of the EEG. They are compared based on their performance to accurately extract markers from the EEG for early detection of AD. The methods were partly chosen due to their insensitivity to artefacts, making them more suitable in practice. Furthermore, we have used a relatively large dataset obtained from different countries; allowing for more rigorous assessment of the methods as compared to using only a single dataset.

The remainder of the paper is organized as follows. Section 2 provides an overview of the methods. In Section 3, we provide a brief description of the data. Section 4 describes the performance measures used for assessing the methods. In Section 5, we present and discuss the results and we conclude the paper in Section 6.

2 METHODS

2.1 Fractal Dimension

In this paper, the dimension of zero-set approach is used to estimate the FD [11]. Given an EEG channel, the FD is obtained by finding the set of instances where the EEG intersects with the zero volt reference. The set is then covered with $N(\Delta t)$ line segments of a chosen length, Δt . The total length of all the line segments ($30\text{ms} \leq \Delta t \leq 500\text{ms}$) is:

$$L(\Delta t) = N(\Delta t) \cdot \Delta t \approx L_0 \Delta t^{D-1} \quad (1)$$

D is estimated using linear regression by taking the log of equation 1, such that $\log(L(\Delta t)) \approx \log(L_0) + (1 - D) \cdot \log(\Delta t)$

To compute the FD for all channels of an EEG recording, the raw EEG of every channel in the set was first divided into one second segments. The estimated FD for each segment was then estimated using equation 1 and plotted on a histogram (128 bins). The mode of the histogram was then used as the composite measure of the FD. This is expressed in equation 2:

$$FD_{index} = \sum_{i=1}^N \frac{D_n^i FD_i}{\sum_{i=1}^N D_i^n} \quad (2)$$

Where D_i is the histogram height (density) at FD estimate FD_i and n is a control constant set to 4.

2.2 Cross Mutual Information

CMI measures the linear and nonlinear dependence of two variables. In this context, it is used to measure the functional connectivity between two time series X and Y [7 - 9]. The computation of the CMI (between every

channel pair in the EEG recording) is given in steps 1 - 10 below:

1. Normalize each time series in the following way:
 $X = X - X_{min}$, $X = X/X_{max}$.
2. Normalize the time series into N bins. The same number of bins was used for both time series. Numerical experiments verify that 16 bins and epochs of more than 3000 sampling points provide stable estimates of the MI.
3. Compute the joint probability density for the measurements of X and Y , $P_{XY}(x_i, y_j)$ using a histogram estimation based on the partitioned signals (see step 2).
4. From the joint probability density we deduce the probability density, P_X and P_Y of each signal, X and Y .
5. Given the probability distribution of X , Y and the joint probability density P_{XY} , we compute the mutual information defined as:

$$I_{XY} = - \sum_{x_i, y_j} P_{XY}(x_i, y_j) \log_2 \frac{P_{XY}(x_i, y_j)}{P_X(x_i)P_Y(y_j)} \quad (3)$$

6. Calculate I_{XY} for several delay values τ ranging from 0 to 500 ms (at any sampling frequency):

$$I_{XY_\tau} = - \sum_{x(t), y(t+\tau)} P_{XY_\tau}(x(t), y(t+\tau)) \log_2 \left(\frac{P_{XY_\tau}(x(t), y(t+\tau))}{P_X(x(t))P_Y(y(t+\tau))} \right) \quad (4)$$

7. Compute the power of each time series X and Y as follows:

$$Power(X) = \sum_{i=1}^N (X_i^2) \quad (5)$$

$$Power(Y) = \sum_{i=1}^N (Y_i^2) \quad (6)$$

8. Normalize I_{XY} (at each time delay) by the mean power of X and Y . Repeat this for all pairs of channel.
9. For each pair of channel (X, Y) we consider the following value of CMI:

$$I_{XY} = \text{mean}(I_{XY}(1 : \text{delay}) + I_{XY}(1 : \text{delay})) \quad (7)$$

10. Obtain a mean value of CMI across all delays and in both directions X - Y and Y - X .

2.3 Hjorth Index

A set of concepts applicable to EEG analysis are the so-called Hjorth descriptors [16]. These are derived from the moments of the spectral density $S_{xx}(f)$. The n th spectral moment is defined as follows:

$$a_0 = \int_{-\infty}^{\infty} (2\pi f)^n S_{xx}(f) df \quad (8)$$

The zero order moment is:

$$a_0 = \int_{-\infty}^{\infty} S_{xx}(f) df \quad (9)$$

So $a_0 = \sigma^2$ (the variance).

The second order moment is defined by:

$$\begin{aligned} a_2 &= \int_{-\infty}^{\infty} (2\pi f)^2 S_{xx}(f) df \\ &= \left. -\frac{d^2 R_{xx}(\tau)}{d\tau^2} \right|_{\tau=0} \\ &= E \left[\frac{dx(t)}{dt^2} \right]^2 \end{aligned} \quad (10)$$

as in [38], where R_{xx} is the autocorrelation function, and the fourth order moment is:

$$\begin{aligned} a_4 &= \int_{-\infty}^{\infty} (2\pi f)^4 S_{xx}(f) df \\ &= \left. -\frac{d^4 R_{xx}(\tau)}{d\tau^4} \right|_{\tau=0} \\ &= E \left[\frac{d^2 x(t)}{dt^2} \right]^2 \end{aligned} \quad (11)$$

Based on these quantities, Hjorth derived the following parameters, called descriptors:

$$\text{activity}, A = a_0 \quad (12)$$

$$\text{mobility}, M = \left[\frac{a_2}{a_0} \right]^{1/2} \quad (13)$$

$$\text{complexity}, C = \left[\frac{a_4}{a_2} - \frac{a_2}{a_0} \right]^{1/2} \quad (14)$$

When these descriptors were applied to the data, we found that lower average complexity was obtained for normal subjects and a higher average mobility was obtained for the AD patients.

In order to track the changes in EEG as a measure of normality or closeness to AD, the following index based on Hjorth descriptors has been defined:

$$HjorthIndex = \frac{\overline{C_{Left}} + \overline{C_{Right}}}{\overline{M_{Left}} + \overline{M_{Right}}} + \frac{100}{\overline{M_{Left}} + \overline{M_{Right}}} \quad (15)$$

Where $\overline{C_{Left}}$ and $\overline{C_{Right}}$ are the average complexity on the left and right of the brain respectively and $\overline{M_{Left}}$ and $\overline{M_{Right}}$ are the average mobility on the left and right of the brain respectively. H_0 is a constant set at 100.

2.4 Source Localization

Source localization aims to estimate the sources giving rise to the EEG and provides information on the generation and propagation of these sources [17]. The current analysis aims at identifying which brain areas are the most or least affected in AD patients as compared to the normal subjects, so as to provide better means of early detection. In this paper, we investigate the spontaneous rhythm generators through a distributed source model where sources are modelled as current dipoles and a dense set of dipoles located at fixed points on a grid covering the brain volume is assumed. The extraction of the SL index or marker from the EEG is summarised as follow:

1. The spatio-temporal EEG is filtered into four frequency bands: Delta (2 to 4Hz), Theta (4 to 8Hz), Alpha (8 to 13Hz), Beta (13 to 21.3Hz).
2. For each frequency band, source localisation is performed at every 3rd sample. A 3-layer spherical head model with 755 voxels is assumed. The LORETA inverse solution was used to estimate the current density at each individual voxel.
3. By taking the square of the estimated current density, the energy at each voxel is found.
4. The total energy for a particular brain region is then calculated by summing the energy over all time period, for all voxels falling within that brain area:

$$E_{A1} = \sum_{t=1}^T \sum_{v=1}^{N_{A1}} m_{vt}^2 \quad (16)$$

Where E_{A1} is the total energy over time in brain region $A1$, m_{vt}^2 is the energy at voxel v at time instant t and N_{A1} is the total number of voxels in region $A1$.

5. The alpha-theta ratio based on the energy is then used as a marker to discriminate between AD patients and normal subjects:

$$SL_{A1} = \frac{E_{A1,\alpha}}{E_{A1,\alpha} + E_{A1,\theta}} \quad (17)$$

Where SL_{A1} is the alpha-theta ratio of the total energy, $E_{A1,\alpha}$ and $E_{A1,\theta}$ are the total energy over time in the alpha and theta band for region $A1$ respectively.

2.5 PDF of Zero-crossing Intervals

The alpha-theta ratio distribution of the ZCIs is applied to the EEG to differentiate between AD patients and normal subjects as it was demonstrated in our previous work that this method provides a good measure for AD detection [11]. In this context, it is defined as the time interval between a positive to negative voltage transition to the

next positive to negative voltage transition. If one examines an EEG record from a single channel, a list of ZCIs may be extracted and the PDF of these data may be plotted. This was computed for EEGs in the alpha (8 to 13Hz) and theta band (4 to 8Hz) and the alpha-theta ratio is obtained as follow:

$$ZCI_{\alpha,\theta} = \frac{PDF(ZCI_{\alpha,\theta})}{PDF(ZCI_{\alpha}) + PDF(ZCI_{\theta})} \quad (18)$$

Where $ZCI_{\alpha,\theta}$ is the alpha-theta ratio based on the PDF of the ZCIs, $PDF(ZCI_{\alpha})$ and $PDF(ZCI_{\theta})$ is the PDF of the ZCIs in the alpha and theta band respectively.

2.6 Power Spectrum Analysis

For completeness, we also computed the alpha-theta ratio based on the power spectrum of the EEG. The EEG data was filtered in four frequency bands namely, Beta (13 - 21.3Hz), Alpha (8 - 13Hz), Theta (4 - 8Hz) and Delta (2 - 4Hz). The power in each frequency band was obtained by taking the sum of the squared magnitude of each signal. The alpha-theta ratio is computed as follow:

$$Power_{\alpha,\theta} = \frac{Power_{\alpha}}{Power_{\alpha} + Power_{\theta}} \quad (19)$$

3 DATA

Four datasets (A, B, C and D) obtained from three clinical centres in three different centres (UK, Italy and Romania) were pooled together to create a large dataset, giving a total of 45 normal and 38 AD subjects. Each dataset consisted of EEGs of healthy normal, age-matched subjects and mild AD patients obtained using strict protocols. Each AD subject was diagnosed using a range of clinical and psychometric test batteries. Dataset A was used to develop and optimise the algorithms and datasets B to D, one from each centre, were used to assess the performance of the methods. Dataset A exhibits the classic spectral characteristics in AD, with elevated power in the theta band. Dataset B has similar characteristics, but for AD patients, the ratio of the alpha-theta power in the frontal region was closer to the normal subjects and contained more artefacts. For Dataset C, although AD patients have higher power in the alpha than theta band, the ratio of alpha-theta power is lower than the normal group. Dataset D has similar spectral features with Dataset C, however the ratio of alpha-theta power for AD patients were higher than normal subjects. In comparison to the other datasets, Dataset D contained the most artefacts. No pre-processing was performed and whole records including artefacts were analyzed. This approach leads to a prediction of the usefulness of the methods in clinical practice. Table 1 provides a summary of the statistics of the datasets.

| Dataset | | A | B | C | D |
|---------------------|------------------|------|------|------|---------|
| No. of Subjects | | 11 | 41 | 20 | 11 |
| Sampling Freq. (Hz) | | 128 | 128 | 128 | 512 |
| No. of Channels | | 21 | 21 | 19 | 22 |
| Duration (s) | | 240 | 240 | 60 | 300 |
| Normal | No. of Subjects | 8 | 24 | 10 | 3 |
| | M | – | 10 | 5 | 3 |
| | F | – | 14 | 5 | 0 |
| | Mean Age (years) | >65 | 69.4 | 78.3 | 73.5 |
| | MMSE | – | – | 22.8 | 20 - 28 |
| AD | No. of Subjects | 3 | 17 | 10 | 8 |
| | M | – | 10 | 5 | 6 |
| | F | – | 14 | 5 | 2 |
| | Age (years) | > 65 | 77.6 | 78 | 75 |
| | MMSE | – | – | 29.2 | 20 - 28 |

Table 1: Summary of dataset statistics

4 MEASURES OF PERFORMANCES

The methods were assessed based on their ability to correctly detect the AD patients from the cohorts. The output of each method is a single index and can be viewed as a marker. Thus, for every subject in the cohort, there would be one index per method, e.g. ZCI index, Hjorth index, etc. Figure 1 shows the FD indices for all subjects in Dataset A. We can see that AD patients have lower FD compared to normal subjects.

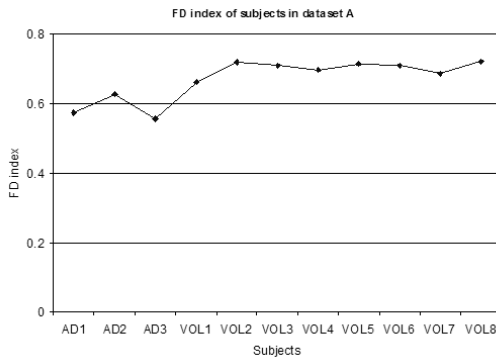


Fig. 1: FD indices of subjects in dataset A.

The following measures were used to quantify the performance of each method:

- **Classification accuracy, sensitivity and specificity.** The classification accuracy is a well-known measure of a method's performance. It is the ratio of correctly classified cases to the total number of cases. Sensitivity and specificity measure the number of true positive and negative cases respectively. Given that A is the number of true positive cases, B is the number of false positive cases, C is the number of false negative cases and D is the number of true negative cases, these measures can be expressed as:

$$Sensitivity = \frac{A}{A + C} \quad (20)$$

$$Specificity = \frac{D}{B + D} \quad (21)$$

$$Accuracy = \frac{A + D}{A + B + C + D} \quad (22)$$

Although the classification accuracy is derived from the sensitivity and specificity, it is important to consider all three measures when evaluating performances. This is because two methods with rather different performances could have the same classification accuracy. For example, incorrect diagnosis from one could be all false negatives while those from the other method might be nearly all false positives, however, based on equation 22, both methods would report the same classification accuracies. At present, there are no benchmarks for these measures. In this paper, we will use the minimum levels of 83%, 55% and 75% for specificity and sensitivity and classification accuracy respectively [19] as these are accepted in clinical diagnosis of AD.

- **Area under the ROC curve (AUC).** The AUC provides a measure of the probability that a system can correctly detect each case when presented with every possible pair of one healthy normal and one AD case from the sample [19]. The closer the AUC is to 1, the better the method.
- **z score.** The z score measures how well a method can differentiate between the AD and normal groups (see equation 23). A higher z indicates better distinction between normal subjects and AD patients. An advantage of this measure is that it is not affected by the units or magnitude of the measured quantity. The z score is given by:

$$z = \frac{\mu_{norm} - \mu_{AD}}{\sqrt{\frac{(n_{norm}-1) \cdot \sigma_{norm}^2 + (n_{AD}-1) \cdot \sigma_{AD}^2}{n_{norm} + n_{AD} - 2}}} \quad (23)$$

where μ_{norm} and μ_{AD} are the mean of the normal and AD group respectively. n_{norm} is the number of normal subjects and n_{AD} is the number of AD subjects in the normal and AD groups. σ_{norm} and σ_{AD} are the associated standard deviations. For ease of comparison, the z score have been normalized across all regions for every method. This was done by taking each method in turn and dividing the z score of each region by the maximum z scores for all regions for that method.

- **Pareto Ranking.** In order to assess the overall performance of each method, we have chosen to use Pareto ranking, which is based on the concept of Pareto dominance [20]. To do this, each method

is assigned a rank based on its performance, relative to other methods. For example, if we compare the results from two methods (A and B) in terms of two criteria: (i) the number of correctly detected true positives i.e. high sensitivity and (ii) the number of correctly classified cases i.e. high classification accuracy. In this case, the method which gives the best performance for both criteria is superior and is given a rank of 1, while the other method is ranked 2. However, in the case where method A has higher sensitivity and method B has higher classification accuracies, both methods would have the same rank since neither outperforms the other. As compared to weighting where assessment criteria are prioritize by pre-assigned weights, Pareto ranking offers the advantage of an unbiased comparison between methods.

5 RESULTS AND DISCUSSIONS

The results presented in this section is based on the performance measures defined in Section IV. The methods are compared in terms of their ability to correctly detect AD. Figure 2 presents the performance of the methods for each dataset for all performance measures. In Figure 3, a comparison summary is provided using Pareto ranking to indicate the overall performance of each method for each dataset.

For dataset A, the performances presented in Figure 2(a) show that FD analysis of the EEG provides a better detection than the other methods. It was able to detect the AD patients from the cohort with 100% classification accuracy (100% for both sensitivity and specificity) and the normalized z scores also indicate a relatively good degree of separation between AD patients and normal subjects. In comparison to FD, the markers based on the ZCIs performed relatively well. However, it was observed that the mean indices between normal and AD groups were closer as compared to those from FD analysis. Therefore, in comparison to the latter, this method could provide a less distinct separation between the two groups. This explains the lower z score obtained for this method. The Pareto rankings in Figure 3 show that Hjorth analysis and SL gave the same overall performances. From Figure 2(a), the two methods had the same performances in terms of specificity, classification accuracies and AUCs. However, we see that the SL indices gave higher false positives while the Hjorth indices had lower z scores. Hence, based on the Pareto concept, the methods do not outperform each other. On the whole, the results for this dataset suggest that all the indices could potentially be used for early detection of AD as they were better than the levels used in current clinical AD diagnosis. However, as the methods were developed using this dataset, there is a need to verify their performances with other datasets.

For dataset B¹, the overall performances of the methods were comparatively better than the minimum levels

¹Results from CMI were insignificant $p > 0.05$, hence disregarded.

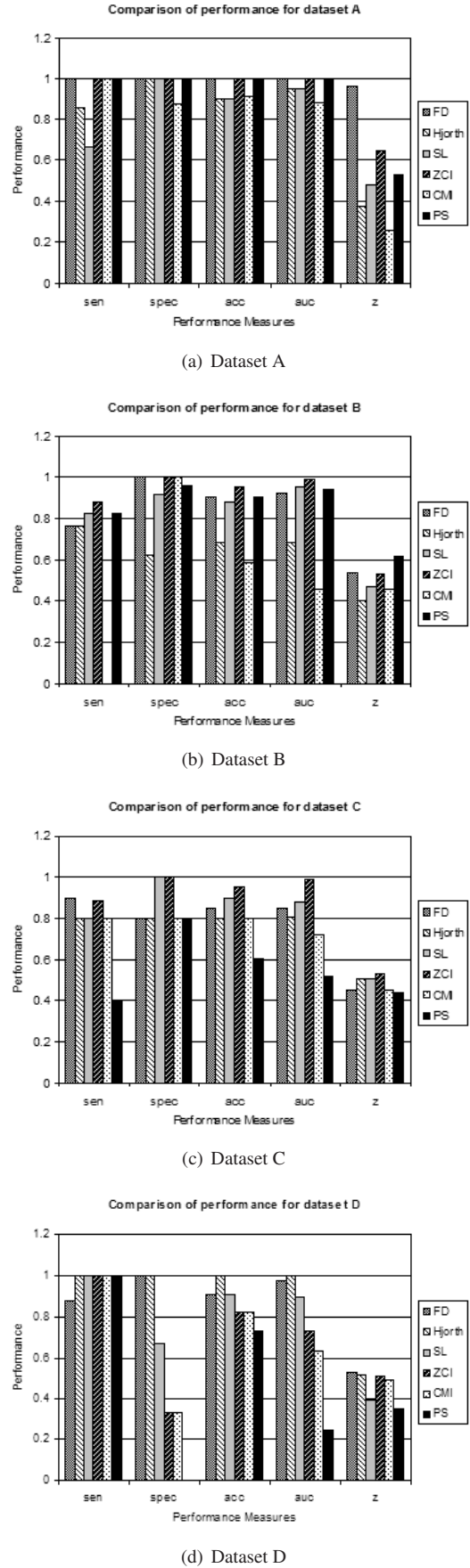


Fig. 2: Comparison of performances for datasets A to D.

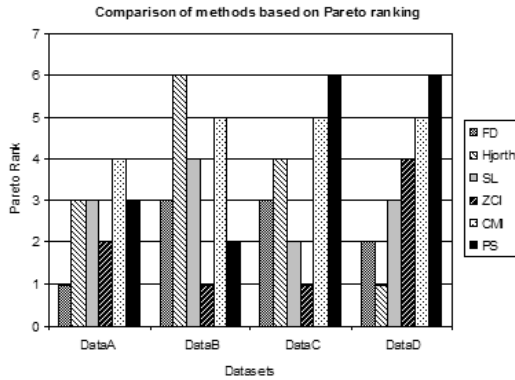


Fig. 3: Comparison of method performances using on Pareto ranking.

we aim to achieve. If the methods were to be used for screening purposes and a more stringent criteria is required, the results in Figure 2(b) suggest that FD and ZCI indices are potentially the most efficient. However, if one is more concern with the overall accuracy of the diagnostic system, then the PS indices would be better since it had less false positives and the normalized z score and AUC indicate better separation and detection within the cohorts. For this dataset, we found that the Hjorth indices for the two groups were very close. Specifically, many normal subjects had indices in the AD range, which resulted in the poor specificity and lower classification accuracies. In comparison to the results for dataset A, we see that not all methods could be use for AD diagnosis. Only FD, ZCI, SL and PS indices could potentially be used as they were better than the minimum specificity, sensitivity and classification accuracy levels used in current clinical diagnosis of AD.

The performances in Figure 2(c) shows that the indices from FD, ZCI, SL and Hjorth analysis have better sensitivity, specificity and classification accuracies than CMI and PS. In the case of PS analysis, the performances were inferior to the other methods. Using this index, a large number of normal subjects were misdiagnosed as having AD (sensitivity $\sim 40\%$). For this dataset, we observed that the distributions of the spectral power in the alpha and theta bands for both AD patients and normal subjects were fairly similar, i.e. higher alpha, lower theta. One of the reasons for the poor performance could be that the method is more susceptible to shifts in the powers of the alpha and theta bands and given the similarities in the spectral characteristics, it was not able to differentiate between the two groups efficiently. Although PS and ZCI analysis share similar characteristics in that both methods are based on the variations in the alpha-theta rhythms of the EEGs of AD patients and normal subjects, the latter gave better performances. One reason could be that in the case of PS analysis, FFT is first performed on the EEGs to obtain the frequency components whereas ZCI works directly in the time domain and hence is not affected by the non-stationarity of the dataset. This allows for greater differentiation between similar, but different,

lower frequency components.

For dataset D, results in Figure 2(d) show that the Hjorth indices had the best performances across all measures. The results show that the indices from PS gave the worst performances and were insignificant ($p > 0.05$). It was observed that AD patients in this dataset had similar spectral characteristics to dataset C. In fact, some AD patients had higher ratios of alpha-theta power than normal subjects. This dataset also contained more artefacts. The effects of the alpha-theta rhythms in the AD group were also evident in the poor performances of ZCI analysis. The results also show that it is important to consider the sensitivity and specificity levels in addition to the classification accuracy. For example, although the classification accuracy of the PS index was better than the minimum level we aim to achieve (75%) and had 100% sensitivity, all normal subjects were misdiagnosed as having AD (0% specificity).

From the overall performances shown in Figure 3, indices from ZCI are the most efficient in detecting AD as they gave the best ranks for two of the four datasets. Figure 3 also show that the methods' performances vary between datasets. For example, for Dataset A, FD had the best performances while for Dataset D, Hjorth indices were superior. This highlights the importance of having larger and more diversified datasets to compare the methods as opposed to when a single dataset is used.

6 CONCLUSION

In this paper, we compared the performance of six candidate methods for the early detection of AD using four different datasets from three different countries. On the whole, we found that time domain methods such as FD and ZCI had better overall performances in comparison to frequency domain methods. For PS analysis, although the results tend to agree with ZCI and SL, it is more susceptible to frequency shifts in the alpha-theta bands. We found that the results from CMI were inferior to the other methods for all datasets. However, this method has shown promising results when applied to EEGs of severe AD patients [14, 21]. The datasets considered in this paper contained only mild AD patients and disruptions in terms of functional connectivity at this stage of the disease may not be apparent in the EEG. Therefore, its is important to investigate this method further. We have also shown that by using an extensive dataset and well-known performance measures, more thorough comparisons of the methods were achieved.

7 ACKNOWLEDGMENT

This work is supported by the BIOPATTERN project, EU Network of Excellence (FP6-2002-IST-1 No 508803).

8 *

REFERENCES

- [1] C. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. Menezes, E. Rimmer, and M. Scazufca, "Global prevalence of dementia: a Delphi consensus study," *The Lancet*, vol. 366, pp. 2112–2117, 2005.
- [2] N. C. Fox, W. R. Crum, R. I. Scahill, J. M. Stevens, J. C. Janssen, and M. N. Rossor, "Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images," *Lancet*, vol. 358, pp. 201–205, 2001.
- [3] C. Hulette, "Neuropathological and neuropsychological changes in "normal" aging: evidence for pre-clinical Alzheimer's disease in cognitively normal individuals," *Journal Neuropathology Exp Neurology*, vol. 57, pp. 1168–1174, 1998.
- [4] V. Jelic, "Early Diagnosis of Alzheimer's Disease - Focus on Quantitative EEG in Relation to Genetic, Biochemical and Neuroimaging markers," Ph.D. Thesis, Karolinska Institute, Stockholm, Stockholm, Sweden, 1999.
- [5] N. Thakor and S. Tong, "Advances in Quantitative Electroencephalogram Analysis methods," *Annual Review Biomedical Engineering*, vol. 6, pp. 453–495, 2004.
- [6] J. Jeong, "Nonlinear dynamics of EEG in Alzheimer's Disease," *Drug Development Research*, vol. 56, pp. 57–66, 2002.
- [7] G. Alder, "The EEG as an indicator of cholinergic deficit in Alzheimer's disease," *Fortschr Neuro.Psychiatric*, vol. 68, no. 8, pp. 352–356, 2000.
- [8] W. S. Pritchard, D. W. Duke, K. L. Coburn, N. C. Moore, K. A. Tucker, M. W. Jann, and R. M. Hostetler, "EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures," *Electroencephalography and Clinical Neurophysiology*, vol. 91, pp. 118–130, 1994.
- [9] J. Jeong, S. Y. Kim, and S. H. Han, "Non-linear dynamical analysis of the EEG in Alzheimers disease with optimal embedding dimension," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 3, pp. 220–228, March 1998.
- [10] C. Besthorn, R. Zerfass, C. Geiger-Kabisch, H. Sattel, S. Daniel, U. Schreiter-Gasser, and H. Förstl, "Discrimination of Alzheimer's disease and normal aging by EEG data," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 2, pp. 241–248, August 1997.
- [11] G. Henderson, E. C. Ifeachor, N. R. Hudson, C. Goh, N. Outram, H. S. K. Wimalaratna, C. D. Percio, and F. Vecchio, "Development and Assessment of a Novel Method for Detecting Dementia using the Human electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1557–1568, 2005.
- [12] J. R. Smith, I. Karacan, and M. Yang, "Automated EEG Analysis with Microcomputers," *Sleep*, vol. 1, no. 4, pp. 435–443, 1979.
- [13] D. Abasolo, R. Hornero, P. Espino, D. Álvarez, and J. Poza, "Entropy analysis of the EEG background activity in Alzheimer's disease patients," *Physiological Measurement*, vol. 27, pp. 241–253, 2006.
- [14] H. Qi, B. Wan, and L. Zhao, "Mutual information entropy research on dementia EEG signals," *The Fourth International Conference on Computer and Information Technology*, pp. 885–889, 2004.
- [15] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 115, no. 7, pp. 1490–1505, July 2004.
- [16] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, pp. 306–310, 1970.
- [17] R. D. Pascual-Marqui, "Review of methods for solving the eeg inverse problem," *International Journal of Bioelectromagnetism*, vol. 1, pp. 75–86, 1999.
- [18] C. Babiloni, G. Binetti, E. Cassetta, D. Cerboneschi, G. Dal Forno, C. Del Percio, F. Ferreri, R. Ferri, B. Lanuzza, C. Miniussi, D. V. Moretti, F. Nobili, R. D. Pascual-Marqui, G. Rodriguez, G. L. Romani, S. Salinari, F. Tecchiok, P. Vitali, O. Zanetti, F. Zappasodi, and P. M. Rossini, "Mapping distributed sources of cortical rhythms in mild Alzheimer's disease. A multicentric EEG study," *NeuroImage*, vol. 22, no. 1, pp. 57–67, May 2004.
- [19] J. Tilbury, P. Van-Eetvelt, J. Garibaldi, J. Curnow, and E. Ifeachor, "Receiver Operator Characteristic Analysis for Intelligent Medical Systems - A New Approach for Finding Confidence Intervals," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 952–963, 2000.
- [20] K. Deb, *Multi-objective Optimization using Evolutionary Algorithms*, 1st ed. John Wiley and Sons Ltd, 2001, ISBN 0-471-87339-X.
- [21] J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the EEG in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 112, no. 5, pp. 827–835, May 2001.