A novel deep learning approach using AlexNet for the classification of electroencephalograms in Alzheimer's Disease and Mild Cognitive Impairment

Rachel Drage, Javier Escudero, *Senior Member, IEEE*, Mario A. Parra, Brian Scally, Renato Anghinah, Amanda Vitória Lacerda de Araújo, Luis F Basile and Daniel Abasolo, *Member, IEEE*

***Abstract*—Alzheimer's Disease (AD) is the most common form of dementia. Mild Cognitive Impairment (MCI) is the term given to the stage describing prodromal AD and represents a 'risk factor' in early-stage AD diagnosis from normal cognitive decline due to ageing. The electroencephalogram (EEG) has been studied extensively for AD characterization, but reliable early-stage diagnosis continues to present a challenge. The aim of this study was to introduce a novel way of classifying between AD patients, MCI subjects, and age-matched healthy control (HC) subjects using EEG-derived feature images and deep learning techniques. The EEG recordings of 141 age-matched subjects (52 AD, 37 MCI, 52 HC) were converted into 2D greyscale images representing the Pearson correlation coefficients and the distance Lempel-Ziv Complexity (dLZC) between the 21 EEG channels. Each feature type was computed from EEG epochs of 1s, 2s, 5s and 10s segmented from the original recording. The CNN architecture AlexNet was modified and employed for this three-way classification task and a 70/30 split was used for training and validation with each of the different epoch lengths and EEG-derived images. Whilst a maximum classification accuracy of 73.49% was obtained using dLZC-derived images from 10s epochs as input to the model, the classification accuracy reached 98.13% using the images obtained from Pearson correlation coefficients and 5s epochs.**

***Clinical Relevance*— The preliminary findings from this study show that deep learning applied to the analysis of the EEG can classify subjects with accuracies close to 100%.**

I. Introduction

Alzheimer's Disease (AD) is the most common form of dementia worldwide. It accounts for nearly 70% of all dementia diagnoses [1] and has an annual estimated societal healthcare cost of $818 billion worldwide. Mild Cognitive Impairment (MCI) is a term used to represent a decline in cognitive function. Confusion with healthy subjects is particularly problematic for early-stage diagnosis because age-matched healthy control (HC) subjects also present natural neurological degradation due to ageing and significant heterogeneity exists within the human ageing population. Although MCI is likely to result in dementia, approximately 30% of MCI individuals may also remain in stable MCI state or revert to that of normal ageing [1]. Current methods for clinical AD diagnosis in symptomatic individuals involve cognitive and functional assessments, the patient's medical history, and expensive, time-consuming neuroimaging scans [2]. Clinical diagnostic accuracy is reported to be approximately 79% on average and is highly dependent on the study of test results by specialists [3]. Definitive diagnosis is only possible by necropsy [4].

The electroencephalogram (EEG) signal has been used extensively in the characterization of AD: AD is a cortical dementia and the EEG reflects the electrical activity of neurons in the cortex of the brain. Furthermore, acquiring the EEG non-invasively is an easy, portable, and cost-efficient procedure compared to other neuroimaging modalities [1]. The analysis of EEG signals with suitable signal processing techniques could provide relevant information for AD diagnosis.

The use of artificial intelligence techniques has grown exponentially in recent years, with increased interest in applying machine learning (ML) and deep learning (DL) methods to diagnostic biomedical applications. DL is a subset of the ML algorithms which emulate the structure of biological neuronal tissue. DL models consist of multiple layers of artificial neurons, known collectively as a neural network (NN), able to tune millions of learnable parameters. DL has the distinct advantage over traditional ML techniques because deep NNs can process and learn latent discriminative features from raw data, in what is known as end-to-end learning. The volume and scope of research employing DL methods for EEG classification tasks have increased exponentially in the past decade, with the most significant domains being sleep staging, epilepsy seizure detection and prediction, and brain–computer interfaces [5].

This study uses DL in a novel way to explore the classification problem presented by early-stage AD diagnosis. It is hypothesized that the combination of DL with EEG-derived features using linear and non-linear methods would improve the classification accuracy of AD and MCI.

Rachel Drage and Daniel Abasolo are with the Centre for Biomedical Engineering, School of Mechanical Engineering Sciences, University of Surrey, Guildford, UK (+44(0)1483682971; e-mails: rach_drage@hotmail.co.uk, d.abasolo@surrey.ac.uk).

Javier Escudero is with the School of Engineering, Institute for Digital Communications, University of Edinburgh, Edinburgh, UK (e-mail: javier.escudero@ed.ac.uk).

Mario A. Parra is with School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK (e-mail: mario.parra-rodriguez@strath.ac.uk).

Brian Scally is with the Institute of Psychological Sciences, University of Leeds, Leeds, UK (e-mail: scallybrian@gmail.com).

Renato Anghinah is with the Reference Center of Behavioural Disturbances and Dementia, School of Medicine, University of São Paulo, São Paulo, Brazil (e-mail: anghinah@terra.com.br).

Amanda Vitória Lacerda de Araújo is with the Traumatic Brain Injury Cognitive Rehabilitation Out-Patient Center, University of São Paulo, São Paulo, Brazil (e-mail: amandavitooria@gmail.com).

Luis F Basile is with the Division of Neurosurgery, University of São Paulo, São Paulo, Brazil (e-mail: lbasile@gmail.com).

II. MATERIALS AND METHODS

A. EEG database

The EEG data used in this study were recorded from 141 subjects at the Neurology Department of the Samaritan Hospital in São Paulo, Brazil, between 2010 and 2016. The EEG database consists of age-matched groups of 52 AD patients (age of 82.3 ± 4.7, mean ± standard deviation), 37 MCI subjects (age of 78.4 ± 5.1), and 52 healthy controls (HC, age of 79.6 ± 6.0). All recordings were 587s in length except for 2 AD, 1 MCI and 2 HC subjects (with recording lengths between 119s and 572s), The scalp electrode placement followed the International 10-20 system using 21 electrodes: Fp1, Fp2, Fpz, F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, Oz, F7, F8, T3, T4, T5 and T6. The sampling frequency was 200 Hz.

The raw EEG recordings were filtered using a 1-60Hz band-pass, FIR filter with an order of 330 and de-noised using independent component analysis and notch filters at 21 and 42 Hz. EEGs were processed further using the Multiple Artefact Rejection Algorithm (EEGLAB plugin for MATLAB©) to remove other noise sources. Furthermore, prior to splitting the data into non-overlapping epochs, 3.5s from the start and end of each signal were removed to account for discrepancies or settling noise.

B. Pearson Correlation Coefficients

Pearson correlation coefficients between the different electrodes were computed from each EEG data epoch and arranged into matrices with 21x21 pixels. Greyscale images were produced from the correlation matrices by normalizing the absolute values to a greyscale pixel interval and resizing the image to fit the model's input dimensions. Fig. 1 shows examples of the resulting images using 5s epochs.

C. Distance Lempel-Ziv Complexity

Distance Lempel-Ziv Complexity (dLZC) is a recently introduced method to compute the LZC of pairs of signals [6]. The dLZC of channel pairs with few subsequences in common would be higher than in signal pairs with a large proportion of subsequences in common. Therefore, the dLZC measures how similarly complex two signals are [6]. If a signal $x(n)$ is coarse-grained to form a binary sequence P and signal $y(n)$ to form a binary sequence Q, the dLZC can be computed as follows [6]:

$$dLZC(x,y) = \frac{c(PQ)-c(PP)+c(QP)-c(QQ)}{b(2n)}, \quad (1)$$

where $c(PQ)$ denotes the complexity for the concatenation of P and Q, $c(QP)$ denotes the complexity for the concatenation of Q and P, $c(PP)$ is the complexity for the concatenation of P and P, $c(QQ)$ is the complexity for the concatenation of Q and Q and $b(2n)$ normalizes the dLZC considering the length of the sequences. A detailed description of the dLZC algorithm can be found in [6].

The resulting dLZC matrix was of dimension 21x21, with each value representing the dLZC values computed for every electrode channel pair combination. These matrices were converted into 21x21-pixel greyscale images. Fig. 2 shows an example of these images for one subject from each of the three groups.

Figure 1. Greyscale 21x21-pixel images (not to scale) representing the correlation matrix computed across all channels from 5s data epochs for one subject from each class.

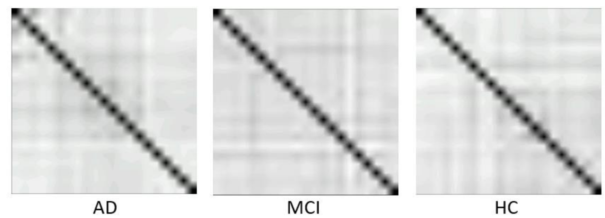

Figure 2. Greyscale 21x21-pixel images (not to scale) representing the dLZC matrix computed across all channels from 5s data epochs for one subject from each class.

D. Deep Learning Architecture

Transfer Learning with AlexNet was employed in this investigation for the three-way classification task of AD, MCI, and HC. AlexNet is the groundbreaking convolutional NN (CNN) that won the 2012 ImageNet Large Scale Visual Recognition Challenge [7]. In AlexNet, a series of five convolutional layers, employing max pooling and rectified linear unit (ReLU), reduce the image size from 224x224 to 13x13 while increasing the depth of the filter response from 96 to 256. The final fully connected layer of AlexNet was replaced with a new final classification layer with three nodes, as this is equal to the number of classes (AD, MCI, and HC) in this classification problem.

A set of standard hyperparameter values based on previous work [8] were used in the DL: ADAM optimizer, mini-batch size of 64, piecewise learn rate schedule, initial learn rate of 10^{-4}, learn rate drop factor of 0.33, learn rate drop period equal to 10, and a validation patience equal to 10.

Epoch lengths of 1s, 2s, 5s and 10s were used to obtain the model input types (images from correlation coefficient and dLZC matrices). Before input to the model, the images were rescaled from size 21x21 to 224x224 using bilinear interpolation. Table I gives the number of training and validation image samples resulting from the different epoch lengths for a training-validation split of 70-30.

TABLE I. NUMBER OF IMAGE SAMPLES AVAILABLE FOR TRAINING AND VALIDATION (70-30 SPLIT) FOR THE DIFFERENT EPOCH LENGTHS USED

Epoch Length		**Number of Image Samples**	
Seconds/epoch	*Data points/epoch*	*Training*	*Validation*
1	200	56400	24171
2	400	28248	12107
5	1000	11338	4860
10	2000	5669	2429

Classification results are presented using confusion matrices, tables summarizing the success of the model at predicting samples belonging to various classes.

III. RESULTS

The epoch-based classification accuracies obtained using the different images described in the previous section as inputs to the model are summarized in Table II. When using images created with the correlation coefficients between the different EEG channels, it can be observed that the accuracy remained stable for all epoch lengths tested, with a maximum value of 98.13% when using 5s epochs. Fig. 3 shows the corresponding confusion matrix, where the high accuracy of the model classifying the 3 classes becomes evident.

On the other hand, the classification accuracies obtained using the images created with dLZC varied significantly, with values ranging from 56.96% when using 1s epochs to 73.49% with images obtained using 10s epochs. The confusion matrix corresponding to the highest classification accuracy is included in Fig. 4, where it is possible to see that the model performed best at classifying the AD patients and worst when identifying patients with MCI.

TABLE II. EPOCH-BASED CLASSIFICATION ACCURACIES ACHIEVED BY ALEXNET TRAINED ON IMAGES DERIVED FROM CORRELATION COEFFICIENTS AND DLZC USING EPOCH LENGTHS OF 1S, 2S, 5S AND 10S

Epoch length (s)	Classification Accuracies (%)	
	Model input: Images derived from correlation coefficients	*Model input: Images derived from dLZC*
1	97.69	56.96
2	98.08	61.61
5	98.13	68.50
10	97.82	73.49

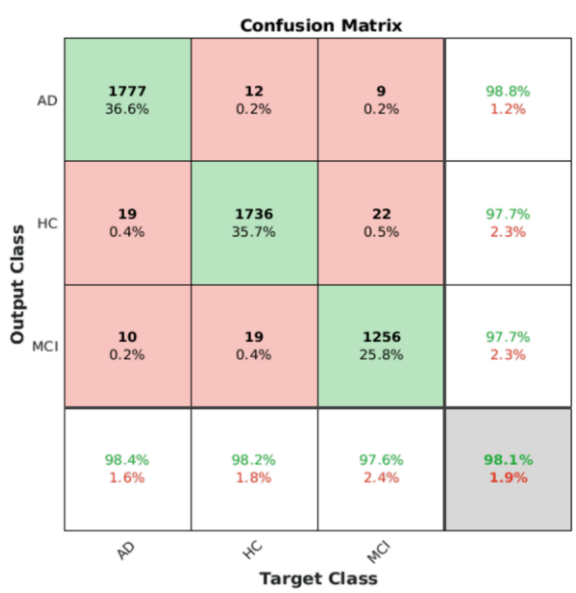

Figure 3. Confusion matrix for the best performing epoch length trials (5s) when images derived from Pearson correlation coefficients between electrodes were used as the model input to AlexNet.

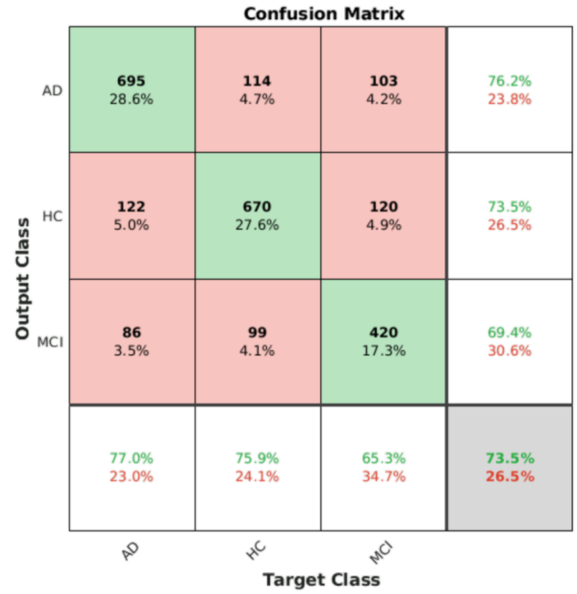

Figure 4. Confusion matrix for the best performing epoch length trials (10s) when images derived from dLZC between electrodes were used as the model input to AlexNet.

IV. DISCUSSION

This study presents a novel application of DL to the challenging problem of differentiating AD from its prodromal diagnostic stage MCI and normal healthy ageing. To this aim, EEG recordings from patients with AD, patients with MCI, and age-matched HC were segmented into epochs of different lengths. To capture linear and non-linear dependencies in the signals from across the brain regions, matrices of Pearson correlation coefficients and dLZC between all pairs of electrodes were computed and subsequently transformed into 2D greyscale images that were fed as inputs to the DL framework, with 70% of the data used for training and 30% for validation. The pre-trained CNN architecture AlexNet was used [7], with the final classification layer replaced with one containing three classes. The epoch-based accuracy was very high when the correlation coefficients were used as inputs to the model (maximum accuracy of 98.13% using 5s epochs), whilst the accuracies with dLZC were significantly lower (maximum value of 73.49% using 10s epochs).

This study investigated four different epoch lengths (1s, 2s, 5s, and 10s). The most frequently used epoch length in similar studies using DL and EEG signals for the classification of AD is 5s with no overlap between adjacent epochs [8-10]. However, other studies have used shorter segments (e.g., 1s in [11] and 2s with 50% overlap in [12]). In the current study, the classification accuracy of the CNN using the matrices of Pearson correlation coefficients is very robust to varying epoch length. In contrast, there is significant variability in the classification accuracy obtained using dLZC, which increased with epoch length. It has been suggested that relatively short 10s epochs may not be able to provide long-term complexity information [13]. Thus, longer epochs might be needed to capture discriminative patterns from EEGs with dLZC.

Pearson correlation coefficients have not been previously used in this characterization problem despite the ability of the method to capture the strength of the linear relationship between pairs of signals. The non-linear nature of EEG signals motivated the use of dLZC to generate inputs reflecting non-linear dependencies between the different EEG electrodes. This method has shown promise in classification of AD, albeit with a much smaller EEG database than the one used in the current study [6]. Despite the inherent non-linear nature of the EEG, the performance of the model with dLZC-derived images was significantly worse than when a much simpler linear method was used. Nevertheless, the current study has only used one CNN framework; a deeper network capable of processing and learning higher order abstract features present within the images due to a greater number of convolution layers may be able to extract discriminative features from the dLZC matrices.

Most of the existing studies using DL to classify AD patients from EEG signals have reported lower epoch-based accuracies than the current study. Some studies have focused on binary classification problems. A classification accuracy of 59% between MCI and control subjects was reported in [12] using a small database with 10 subjects from each class. A higher accuracy was obtained in [14], where AD and HC were classified with 92% accuracy from signals from 15 patients and 15 control subjects. Another study found an accuracy of 98% using 17 AD patients and 35 subjects with MCI [11]. Nevertheless, the classification problem becomes more challenging when there are three different classes present (AD, MCI, and HC). Bi and Wang found a classification accuracy of 95% using a very small database with 4 subjects in each class [15]. A bigger database was used in [9], where a classification accuracy of 82% was achieved with 23 AD patients, 23 controls, and 23 subjects with MCI, although substantial feature extraction was required before the application of a deep CNN. A very similar accuracy of 83% was reported in [10], where the database used contained 63 AD patients, 63 patients with MCI, and 63 controls and features related to the power spectral density of the EEGs were used as inputs to the CNN. Finally, a classification accuracy of 98.9% was obtained with the same database of the current study using images representing the time-frequency information from the EEGs and AlexNet [8].

Several limitations of the study should be discussed. The training and validation data sets need to be labelled for the supervised classification. Given that definite AD diagnosis is only possible by necropsy and that even the most experienced specialists may fail in the diagnosis of about 10–15% of the cases for early-stage AD [16], there may be an inconsistency in the labelling that could lead to biased results. Nevertheless, it was for the AD class that the algorithm yielded the best classification accuracy. Furthermore, the database used in the study is unbalanced, with 52 subjects in each of the AD and HC classes, but a lower number of 37 subjects in the MCI class. This could result in a bias against the MCI class in the confusion matrices. The interpolation process used to rescale the images from size 21x21 to 224x224 creates new virtual features and the impact of this in the results should be explored in the future. Moreover, cross-validation could be used to limit the possible impact of data selection on the classifier results. Another limitation is the possibility of intra-class variability, especially in the case of MCI, as not all patients would go on to develop AD. The heterogeneity that characterizes MCI and normal ageing might, therefore, be driving some of the outcomes. Future research should focus on enhancing specificity of DL-based classification algorithms.

Despite these shortcomings, the current study presents some very promising results in the field of automatic classification of AD patients, showing that DL could potentially help in the early diagnosis of this form of dementia.

REFERENCES

[1] R. Cassani, M. Estarellas, R. San-Martin, F. J. Fraga, and T. H. Falk, “Systematic review on resting-state EEG for Alzheimer’s disease diagnosis and progression assessment,” *Dis. Markers*, vol. 2018, 5174815, 2018.

[2] J. Dauwels, F. Vialatte, and A. Cichocki, “Diagnosis of Alzheimer’s Disease from EEG Signals: Where Are We Standing?” *Curr. Alzheimer Res.*, vol. 7(6), pp. 487–505, 2010.

[3] T. G. Beach, S. E. Monsell, L. E. Phillips, and W. Kukull, “Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010,” *J. Neuropath. Exp. Neur.*, vol. 71 (4), pp. 266–273, 2012.

[4] M. W. Bondi and G. E. Smith, *Mild Cognitive Impairment and Dementia: Definitions, Diagnosis, and Treatment*. Oxford University Press, 2013

[5] Y. Roy, H. Banville, I. M. Carneiro de Albuquerque, A. Gramfort, T. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: A systematic review,” *J. Neural Eng.*, vol. 16(5), 051001, 2019.

[6] S. Simons and D. Abásolo, “Distance-based Lempel-Ziv complexity for the analysis of electroencephalograms in patients with Alzheimer’s disease,” *Entropy*, vol. 19(3), 129, 2017.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Adv. Neur. In.*, vol. 25(2), pp. 1097–1105, 2012.

[8] C. J. Huggins, J. Escudero, M. A. Parra, B. Scally, R. Anghinah, A. V. L. de Araújo, L. F. Basile, and D. Abasolo, “Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing”, *J. Neural Eng.*, vol. 18(4), 046087, 2021.

[9] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, and P. Bramanti, “Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings,” in *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, 2016, pp. 1–6.

[10] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, “A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings,” *Neurocomputing*, 323, pp. 96–107, 2019.

[11] Z. You, R. Zeng, X. Lan, H. Ren, Z. You, X. Shi, S. Zhao, Y. Guo, X. Jiang, and X. Hu, “Alzheimer’s disease classification with a cascade neural network,” *Front. Public Health*, vol. 8, 2020.

[12] D. Kim and K. Kim, “Detection of early stage Alzheimer’s disease using EEG relative power with deep neural network,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA, 2018, pp. 352-355.

[13] M. Fan, A. C. Yang, J. L. Fuh, and C. A. Chou, “Topological pattern recognition of severe Alzheimer’s disease via regularized supervised learning of EEG complexity,” *Front. Neurosci.*, vol. 12, 2018.

[14] Y. Zhao and L. He, “Deep Learning in the EEG diagnosis of Alzheimer’s disease,” in *Computer Vision - ACCV 2014 Workshops. ACCV 2014. Lecture Notes in Computer Science*, vol. 9008, C. Jawahar, S. Shan, Eds., Springer, 2015, pp. 340–353.

[15] X. Bi and H. Wang, “Early Alzheimer’s disease diagnosis based on EEG spectral images using deep learning,” *Neural Netw.*, vol. 114, pp. 119–135, 2019.

[16] A. Alberdi, A. Aztiria, and A. Basarab, “On the early diagnosis of Alzheimer’s disease from multimodal signals: A survey,” *Artif. Intell. Med.*, vol. 71, pp.1–29, 2016.