THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Deep Learning of Resting-state Electroencephalogram Signals for 3-class Classification of Alzheimer’s Disease, Mild Cognitive Impairment and Healthy Ageing

Citation for published version:
Huggins, CJ, Escudero, J, Parra, MA, Scally, B, Anghinah, R, Lacerda de Araújo, AV, Basile, LF & Abasolo, D 2021, 'Deep Learning of Resting-state Electroencephalogram Signals for 3-class Classification of Alzheimer’s Disease, Mild Cognitive Impairment and Healthy Ageing', *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2552/ac05d8

Digital Object Identifier (DOI):
10.1088/1741-2552/ac05d8

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Neural Engineering

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 26. Aug. 2022

Deep Learning of Resting-state Electroencephalogram Signals for 3-class Classification of Alzheimer's Disease, Mild Cognitive Impairment and Healthy Ageing

Cameron J Huggins[1], Javier Escudero[2], Mario A. Parra[3,4], Brian Scally, Renato Anghinah[5,6], Amanda Vitória Lacerda de Araújo[6], Luis F Basile[7] and Daniel Abasolo[1]

[1] Centre for Biomedical Engineering, Department of Mechanical Engineering Sciences, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK.
[2] School of Engineering, University of Edinburgh, Edinburgh, UK.
[3] School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK.
[4] Universidad Autónoma del Caribe, Programa de Psicología, Barranquilla, Colombia.
[5] Reference Center of Behavioural Disturbances and Dementia, School of Medicine, University of São Paulo, São Paulo, Brazil.
[6] Traumatic Brain Injury Cognitive Rehabilitation Out-Patient Center, University of São Paulo, São Paulo, Brazil.
[7] Division of Neurosurgery, Department of São Paulo Medical School, University of São Paulo, São Paulo, Brazil.

E-mail: cameronjhuggins@gmail.com

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

Objective: This study aimed to produce a novel Deep Learning (DL) model for the classification of subjects with Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI) subjects and Healthy Ageing (HA) subjects using resting-state scalp EEG signals.

Approach: The raw EEG data were pre-processed to remove unwanted artefacts and sources of noise. The data were then processed with the Continuous Wavelet Transform (CWT), using the Morse mother wavelet, to create time-frequency graphs with a wavelet coefficient scale range of 0 to 600. The graphs were combined into tiled topographical maps governed by the 10-20 system orientation for scalp electrodes. The application of this processing pipeline was used on a data set of resting-state EEG samples from age-matched groups of 52 AD subjects (82.3 ± 4.7 years of age), 37 MCI subjects (78.4 ± 5.1 years of age) and 52 HA subjects (79.6 ± 6.0 years of age). This resulted in the formation of a data set of 16,197 topographical images. This image data set was then split into training, validation and test images and used as input to an AlexNet DL model. This model was comprised of 5 hidden convolutional layers and optimised for various parameters such as learning rate, learning rate schedule, optimiser, and batch size.

Main Results: The performance was assessed by a 10-fold cross-validation strategy, which produced an average accuracy result of 98.9% ± 0.4% for the three-class classification of AD vs. MCI vs. HA. The results showed minimal overfitting and bias between classes, further indicating the strength of the model produced.

Significance: These results provide significant improvement for this classification task compared to previous studies in this field and suggest that DL could contribute to the diagnosis of AD from EEG recordings.

Keywords: Alzheimer's Disease, Biomedical Signal Processing, Classification, Deep Learning, Electroencephalogram (EEG), Mild Cognitive Impairment (MCI).

© xxxx IOP Publishing Ltd

1. Introduction

Dementia is a term that describes a collection of diseases that affects approximately 50 million people worldwide (Prince et al., 2015). It is characterised by a measured cognitive decline in two or more domains such as memory, language, behaviour and personality, ultimately leaving the individual unable to perform simple everyday tasks (Weller & Budson, 2018). The global healthcare cost of dementia is upwards of $818 billion, which is increasing year by year (Prince et al., 2015). Alzheimer's Disease (AD) contributes to approximately 60-80% of the global dementia diagnoses and is most prevalent in adults aged 60 and above (Weller & Budson, 2018). The biggest risk for AD is age, with a reported doubling of the disease prevalence every 6.3 years after the age of 60 (Prince et al., 2015). Other factors include health risks such as high Body Mass Index (BMI), high fasting glucose, smoking and increased intake of sugar-sweetened beverages (GBD 2016 Dementia Collaborators, 2019).

AD is related to neurofibrillary tangles and amyloid plaques developing in the cerebral cortex area of the brain, especially in the hippocampus (DeTure & Dickson, 2019). An intermediary stage between Healthy Ageing (HA), sometimes referred to as Healthy Control (HC), and AD has been widely recognised as a stage called Mild Cognitive Impairment (MCI). It is currently unclear who amongst individuals with MCI will develop AD dementia (Kramer et al., 2007). At present, the only definitive way to diagnose AD is via post-mortem examination to find the plaques or tangles within the brain (DeTure & Dickson, 2019). However, with the advent of novel dementia biomarkers which can be gathered in vivo, a new biological definition of AD has been introduced (Jack Jr et al., 2016) (Jack Jr et al., 2018). Biomarkers are expensive, invasive, little specific, and only available in specialised centres (Parra et al., 2019).

The gold standard, or current state of the art, for diagnosing AD uses mental examinations combined with costly and time-consuming neuroimaging scans such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) (Cassani et al., 2018). These imaging tools aim to highlight biomarkers, such as amyloid-β peptides, that indicate the formation of plaques within the brain. This method is highly dependent on trained doctors that interpret and analyse the results to determine the diagnosis. The reported diagnostic accuracy of AD by experts alone is only 77% (Sabbagh et al., 2017).

Early diagnosis of AD, during the MCI stage, would allow healthcare professionals to avoid misdiagnosis, deliver quick and more appropriate treatment options, and provide better overall disease management. To achieve this, the analysis of electroencephalogram (EEG) signals has been suggested by researchers to find features and biomarkers that may aid in AD diagnosis (Dauwels et al., 2010) (Rossini et al., 2020).

Conventionally, EEG analysis has been used clinically to evaluate different conditions such as epilepsy, sleep disorders and strokes (Britton et al., 2016). In addition, these signals have also been analysed in research settings using visual and statistical methods for the diagnosis of AD. The drawback of this is that standalone, EEG analyses have not produced results which can compete with standard practices (Craik et al., 2019). Typically, EEGs are subject to signal processing techniques that provide condensed features for classification. There are three main features of EEGs that differ between AD and HA which are: slowing of the EEG, reduced complexity of the EEG and EEG synchrony (Dauwels et al., 2010). Examples of signal processing methods for the detection of slowing of the EEG include time frequency analysis techniques such as Discrete Fourier Transform (DFT), Power Spectral Density (PSD) and Continuous Wavelet Transform (CWT). Entropy and Lempel-Ziv complexity are examples of algorithms used to explore EEG complexity and the Pearson Correlation Coefficient and Coherence Function have been used to identify perturbations in EEG synchrony (Dauwels et al., 2010). The problematic poor performance of EEG analysis for AD diagnosis has drawn interest from new developments in Artificial Intelligence (AI), with the ability to analyse signals with increased complexity and depth.

Deep Learning (DL) is a subsection of AI that has become popular in recent years due to advancements of Graphics Processing Units (GPUs) in computing. It aims to mimic the learning of the human brain by using complex algorithms to obtain features of data that cannot be seen using conventional statistical analysis methods. Thus, millions of learnable parameters, particularly useful in image classification problems, are built to detect perturbation features. These algorithms have proven results that exceed human performance in classification problems such as ImageNet (Dodge & Karam, 2017). Many biomedical engineering applications, including EEG analysis, can produce graphical image outputs that can be used as inputs to these DL networks.

It is hypothesised that improved accuracies for the diagnosis of AD can be achieved using DL to correctly classify the EEG recordings from AD, MCI or HA subjects. This study employed a novel signal processing technique to convert complex EEG recordings into usable input images for a DL network. A pre-trained DL model was optimised for this study's focus, trained using the ground truth values associated with each image. It was then cross validated to evaluate the reported classification accuracy result.

The outline of this paper is as follows. Section 2 details the related literature and studies in this field. Section 3 describes the materials and methods used. Section 4 displays the results of the study and Section 5 contains the discussion. Finally, section 6 presents the conclusions and further work.

2. Related Studies

This section details seven studies that relate to the research topic of improving diagnosis of AD using DL of EEG signals. It is split into two sections detailing three-class and two-class classification.

The first reported paper to complete three-class (tertiary) classification of AD vs. MCI vs. HA using DL techniques was by Morabito et al. (2016). The authors propose a 2-layers deep Convolutional Neural Network (CNN) that uses extracted features from time-frequency maps as the input to the network. The signal processing method used was CWT with the Mexican Hat mother wavelet. The reported accuracy was relatively low at 82% and used a reasonably small data set consisting of 23 AD, 23 MCI and 23 HA subjects.

A discriminative deep probabilistic model was proposed by Bi & Wang (2019) which produced an accuracy result of 95.04% for three-class classification of AD, MCI and HA. The employed signal processing technique created spectral topography maps from raw signals, showing the relative power per frequency band across each electrode in one image. One shortfall of this paper is that, although a combined 12,000 images were used across training and validation, the data originated from a very small sample size of four subjects per class. This could indicate that the results are more susceptible to error and less accurate than what is stated within the paper.

A paper published by Ieracitano et al. (2019) used 2D greyscale Power Spectral Density (PSD) images as an input to a CNN for the three-class classification of AD, MCI and HA. The authors report an accuracy of 83.33% for the CNN, which was superior compared to shallow machine learning techniques such as SVM, MLP and Linear Discriminant Analysis (LDA). This result showed clear accuracy benefits for using CNNs within this field; however, the authors correctly conclude that without future work, this method would not be sufficient alone for clinical diagnosis of AD. In addition to the large data set of 63 HA, 63, MCI and 63 AD subjects, a clear explanation of the importance of avoiding bias by balancing and age-matching the data set was included. Despite the large number of subjects, there were only 2,340 images used within the model due to the sampling size which is below the general guidelines of 1,200 images per class (Stanford Vision Lab, 2010).

The most recent published paper by Ieracitano et al. (2020) used a “multi-modal machine learning” approach to classify EEG recordings in dementia. The authors make use of a novel combination of features extracted from Bispectrum analysis and CWT time-frequency analysis to classify AD vs. MCI vs. HA. The results indicate a maximum of 89.22% accuracy for this classification, which is a significant increase in using either signal processing method individually. The results published are supported with statistical evidence and used the largest number of features seen within this field equalling 207,900. However, the authors do not innovate within machine learning, using a simple Multi-Layered Perceptron (MLP) as the classifier for this problem. It successfully identifies areas of future research, such as linking this work with a feasibility study using a novel spiking neural network architecture called NeuCube (Capecci et al., 2014).

Three of the seven papers analysed only reported two-class classification which severely reduced the complexity requirement of the models proposed.

Zhao & He (2015) used time-domain EEG signals as the input to a Restricted Boltzmann Machine (RBM) network to obtain features from the data, which are classified using an SVM. The unique feature of this paper is that the authors conducted a detailed optimisation experiment on the number of hidden layers and number of nodes within each layer for this network. They conclude that a low number of layers (L=3) and high number of nodes (n=2000) was preferential, resulting in a two-class (binary) classification accuracy of 92%. The authors reduced computing time by only using 10 out of the 30 available subjects (15 AD, 15 HA) but showed that the sample size was small in comparison to a relatively large 19,200 input images.

Kim & Kim (2018) investigated early diagnosis of AD by producing a model that detects the difference between MCI and HA subjects. The model comprises of a Deep Neural Network (DNN) with feature-based inputs relating to the Relative Power (RP) of different frequency bands within EEG signals. The model’s performance was poor in comparison to other papers analysed, with just 75% accuracy over a small data set of 10 HA and 10 MCI subjects. In comparison to the paper by Zhao & He (2015), the authors also investigate the effect of the number of hidden layers within the network but report the highest accuracy with the largest number of hidden layers tested (L=4). This difference could be explained by various factors such as the difference in NN type, data set and type of two-class classification.

Fan et al. (2018) made use of a complexity measure, Multiscale Entropy (MSE), to extract features of the EEG signals, in numerical form, for the input to a DL Linear Regression (LR) model. A calculated 46,470 features were used which produced a maximum accuracy result of 82%, low compared to the other papers in this section. An interesting data set classification was used which compared HA subjects to three different severities of AD based on increasing Clinical Dementia Rating (CDR) scores (AD1, AD2 and AD3). The specificity of each AD diagnosis has advantages and disadvantages; it could lead to better diagnosis accuracy results but limits the model to this data set as the CDR scale is not readily available for many other databases. Although the data set was large with 123 subjects, it was also highly unbalanced with only 15 HA subjects. The authors were able to suggest future improvements including altering the MSE method, collecting more data and assessing the interactions between electrodes.

A summary of the key information presented in each discussed paper is detailed in Table 1.

Cassani et al. (2018) produced a review paper on EEG for AD diagnosis and includes recommendations for the future of this research area. The author's recommendations focus on the EEG databases, which should be balanced around demographics such as age, gender, number of subjects and education level as well as being as large as possible. The paper also suggests that clearer and more detailed information should be provided about the process of machine or deep learning. This is apparent after reviewing the above seven papers, as they all provide different levels of detail surrounding the experimentation and design process and make it difficult to reproduce or build upon their results. The reviewed papers show the possibility of high accuracy classification using a combination of signal processing and DL of EEGs, and there are lots of different methods that are yet to be explored.

3. Materials and Methods

3.1 Model Design

The overall design of the proposed model within this study is outlined in the experimental flow diagram in Figure 1. First, the raw EEG signals were pre-processed and split into epochs of 5 seconds. Each sample was then subject to signal processing and converted into Red, Green and Blue (RGB) colour images that were suitable as an input to an optimised DL Neural Network (NN). The resulting image data set was then randomly split into 10 folds, from which a discrete portion of the folds were used in the training, validation, and test splits of the DL NN model. The model was then validated using k-fold cross validation and assessed using confusion metrics.

Table 1: A summary of the previous directly related papers in the field of Deep Learning of EEG signals for the diagnosis of Alzheimer's Disease, showing the database size & age matching, signal processing methods employed, machine learning architectures used and final classification results.

Legend: **AD** = Alzheimer's Disease, **CNN** = Convolutional Neural Network, **CWT** = Continuous Wavelet Transform, **DCssCDBM** = Discriminative Contractive Slab and Spike Convolutional Deep Boltzmann Machine, **DNN** = Deep Neural Network, **HA** = Healthy Aging, **LASSO** = Least Absolute Shrinkage and Selection Operator, **MCI** = Mild Cognitive Impairment, **MLP** = Multi-layer Perceptron, **MSE** = Multiscale Entropy, **RBM** = Restricted Boltzmann Machine, **RGB** = Red, Green & Blue, **RP** = Relative Power.

Reference	**Database Size & Age Matching**	**Signal Processing Method**	**Machine Learning Architecture**	**Results**
(Ieracitano et al., 2020)	63 HA 63 MCI 63 AD Balanced	Features from Mexican Hat CWT & Bispectrum Estimation.	MLP	AD-MCI-HA = 89%
(Bi & Wang, 2019)	4 HA 4 MCI 4 AD Balanced	2D RGB images by combining spectral topographical maps	DCssCDBM 2 hidden layers	AD-MCI-HA = 95%

(Ieracitano et al., 2019)	63 HA
	63 MCI
	63 AD
	Balanced
		2D greyscale Periodogram images	CNN
			1 hidden layer
				AD-MCI-HA = 80%
				MCI-HA = 92%
				AD-HA = 91%
				MCI-AD = 84%

(Kim & Kim, 2018)	10 NC 10 MCI Age-matched	Features from RP	DNN 4 hidden layers	MCI-HA = 75%

(Fan et al., 2018)	15 HA
	15 AD1
	69 AD2
	24 AD3
	Unknown
		Features from MSE Analysis	LASSO Model	HA-AD1 = 42%
				HA-AD2 = 69%
				HA-AD3 = 79%
				AD1-AD3 = 82%
				AD2-AD3 = 72%
				AD1-AD2 = 71%

(Morabito et al., 2016a)	23 HA
	23 MCI
	23 AD
	Balanced
		2D RGB images from Mexican Hat CWT	CNN
			2 hidden layers
				AD-MCI-HA = 82%
				MCI-HA = 85%
				AD-HA = 85%
				MCI-AD = 78%

(Zhao & He, 2015)	15 HA 15 AD Unknown	Raw Data	RBM 3 hidden layers	AD-HA = 92%

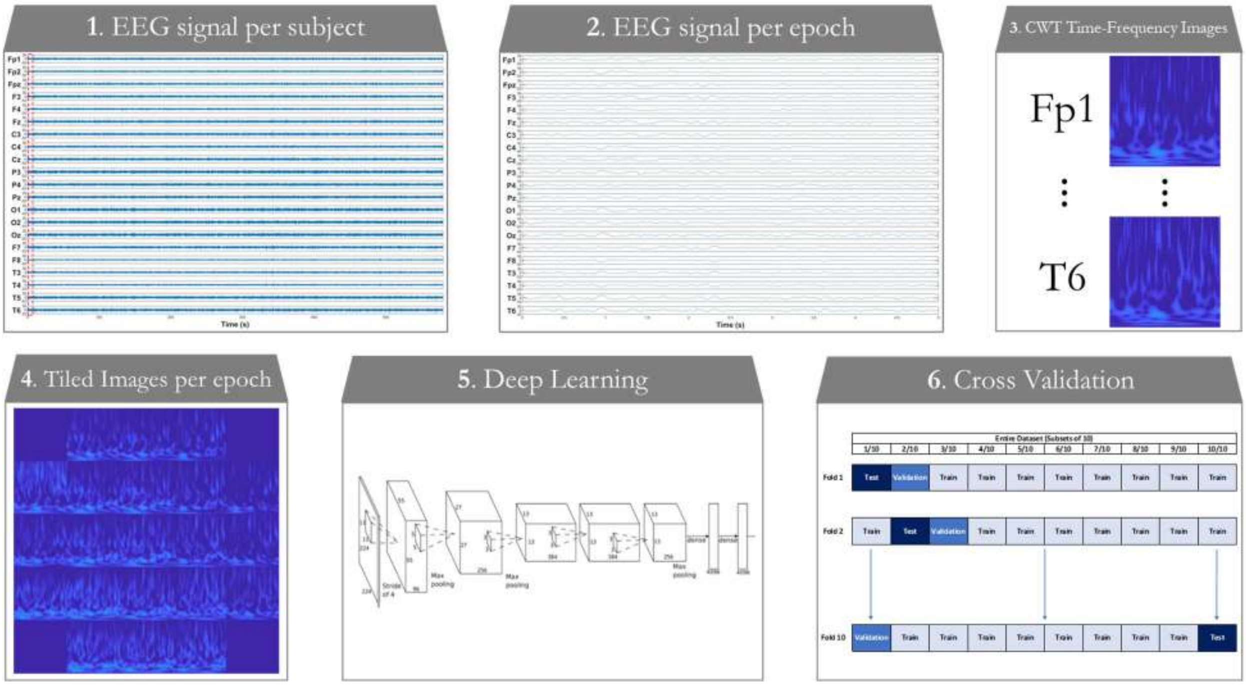


Figure 1: Experimental Flow Diagram showing a simplified version of the proposed model. Image for '5. Deep Learning' taken from (Krizhevsky et al., 2012). The training folds/test folds describe a proportion of the data assigned to train or test the model respectively.

3.2 Subjects

This study consisted of 141 subjects originating from the Behavioural and Cognitive Neurology Unit of the Department of Neurology and the Reference Centre for Cognitive Disorders at the Hospital das Clinicas in São Paulo, Brazil (henceforth referred to as the 'Brazil' study) (Cassani et al., 2017). The Brazil study diagnosed AD and HA subjects according to the National Institute of Neurological Disorders and Stroke and Alzheimer's Disease and Related Disorders (NINCDS-ADRA) criteria. Typically, within this study AD subjects presented a Mini Mental State Examination (MMSE) score of ≤24 and a CDR of 0.5-2, and HA subjects presented an MMSE score of ≥25 and a CDR of 0. The AD subjects were required to have shown functional and cognitive decline over the previous 12 months. The MCI subjects are presented with an MMSE score of ≥24 and a CDR of 0-0.5, with objective evidence of impairment in one or more cognitive domains but retaining independence in functional abilities (Albert et al., 2011) (Petersen & Knopman, 2006) (Petersen & Negash, 2008). In addition, all subjects were required to display absence of conditions that cause cognitive decline specified as diabetes mellitus, kidney disease thyroid disease, alcoholism, liver disease, lung disease or vitamin B12 deficiency (Cassani et al., 2017).

The EEG recordings from the Brazil study were collected using a resting-awake eyes-closed method using the Braintech 3.0 EEG device (EMSA Equipa-mentos Médicos Inc., Brazil) with a sampling frequency of 200Hz. The electrode positioning consisted of 21 electrode, positioned on the scalp according to the 10-20 layout system (Fp1, Fp2, Fpz, F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, Oz, F7, F8, T3, T4, T5, T6), in which each point corresponds to a brain region. A time series of 587 seconds (~10 minutes) was captured for each subject (excluding 2 AD, 1 MCI and 3 HA subjects which had shorter recordings).

3.3 EEG Pre-processing

The data were pre-processed by utilising a sequence of different methods. They were first filtered using a 1-60Hz band-pass, FIR filter with an order of 330 and de-noised using Independent Component Analysis (ICA) and notch filters at 21 and 42 Hz to remove significant oscillatory noise artefacts and its harmonic. They were processed further using the Multiple Artefact Rejection Algorithm (MARA) (EEGLAB plugin for MATLAB®) to automatically classify and remove other noise related ICA components. The de-noised data set was then split into age-matched groups of 52 AD subjects (82.3 ± 4.7 years of age, MMSE score of 21.0 ± 4.8), 37 MCI subjects (78.4 ± 5.1 years of age, MMSE score of 25.4 ± 2.7) and 52 HA subjects (79.6 ± 6.0 years of age, MMSE score of 27.5 ± 1.6). It is noted that 22 AD subjects, 19 MCI subjects and 24 HA subjects did not have recorded MMSE scores but were still used for this analysis. Misclassification of older HA subjects and younger AD subjects can occur when comparing subjects of a large age-range due to the natural progressive neurological degradation during human ageing. Therefore, it is important to use data sets with age-matched subject groups to ensure reliable results (Dukart et al., 2011).

The time series was then split into 5 second epochs, removing 3.5 seconds from the start and end of the signal to account for any discrepancies or noise. Using all the raw data available from the 141 subjects over the 21 electrode positions, a total of 340,137 samples of 1,000 data points were made available for analysis. The number of artefact-free epochs for each subject differed, however all available data were included to provide as many data as possible for analysis.

3.4 Signal Processing

For each sample, the data were processed using the Continuous Wavelet Transform (CWT), shown by Equation (1).

$$\boldsymbol{W(a,b)} = \int_{-\infty}^{\infty} \boldsymbol{x(t)}\frac{\boldsymbol{1}}{\sqrt{|\boldsymbol{a}|}}\boldsymbol{\psi}^{*}\left(\frac{\boldsymbol{t-b}}{\boldsymbol{a}}\right)\boldsymbol{dt} \qquad (1)$$

Where: t = time; ψ = mother wavelet; $x(t)$ = time domain signal; a = range of scales; b = translations; $W(a,b)$ = wavelet coefficients (amplitudes of a series of wavelets), * = complex conjugate.

The CWT function is an analysis method that transforms the raw, one-dimensional input signal into the time-frequency (often called the time-scale) domain. The transform is controlled by a mother wavelet with a zero average which is subject to adaptations over a range of scales (*a*) and time translations (*b*), resulting in the output wavelet coefficients (*W*). The signal is fit to match the wavelet, so it is important that an appropriate shaped mother wavelet is chosen for each individual application. There is a plethora of families of mother wavelet choices such as Haar, Daubechies, Coiflets, and Morlet, with the Mexican Hat wavelet proving most popular within this specific field as shown in Table 1 (Mallat, 2009). Despite this, a family of wavelets called the Morse wavelets, introduced by Olhede & Walden (2002), was chosen for this study (Greco et al., 2003). This mother wavelet is a form of analytic wavelet, which are complex-valued wavelets designed for the analysis of modulated oscillations that have useful information in both magnitude and phase (Lilly & Olhede, 2010). Due to this unique feature, the Morse wavelet is directly applicable to the complex and non-stationary nature of EEG signals.

3.5 Image Data Set Generation

The output of the CWT function can be displayed visually as a time-frequency map. In this situation, a plot called a scalogram is created which plots frequency against time, with the energy of the CWT coefficients indicated by the colour of the plot.

When computing the CWT over many EEG samples, it is important to determine a standard for the colour bar scale. The colours presented on the plot are governed by the scale choice, including the maximum and minimum values and the scale base (i.e. logarithmic or linear). The scalogram plot, by design, plots absolute values of the wavelet coefficients, so a minimum value of 0 was chosen. To obtain the maximum value on the scale and the type of scale, the maximum value of the coefficients for each sample were plotted on a linear scale and a logarithmic scale. These plots indicated that a logarithmic scale would be more appropriate for this task as the data contained a high quantity of low maximum coefficient values in contrast to a low quantity of high maximum coefficient values, as shown in Figure 2. By rounding the highest maximum coefficient to the nearest whole number, a maximum scale value of 6x10^{2} was chosen that could encompass all the signals.

The colour map used for the time-scale representation of the coefficients was ‘parula’, with the default number of unique colours set at 256. This value was kept relatively low to reduce image complexity and therefore computational expense. A dark blue colour represents an intensity value of 0 (absolute coefficient value = 0), and a light-yellow colour represents an intensity value of 1 (absolute coefficient value = 6x10^{2}), changing colour logarithmically between these values (Mathworks, 2020).

The generated scalogram plots were then saved and combined into tiled images based on the 10-20 system. The images relating to each electrode in a 5 second epoch were

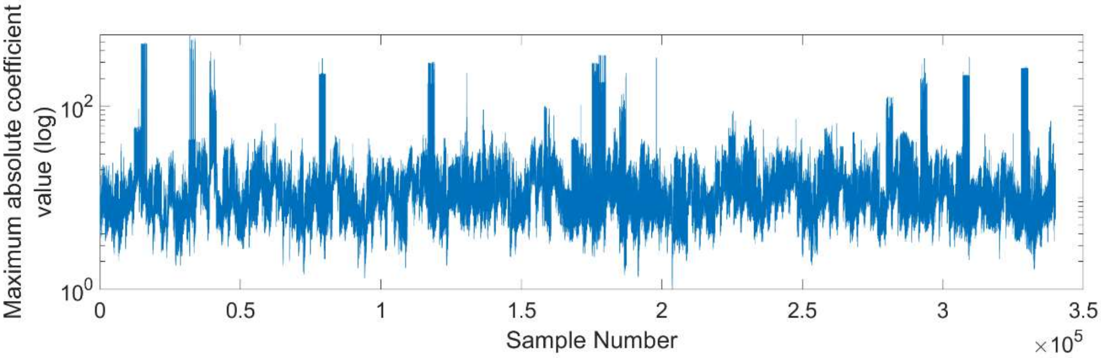


Figure 2: A line graph of the maximum absolute coefficient value on a logarithmic scale against sample number (n=340,137).

combined using the orientation in Figure 3 to produce topographical images based on the 10-20 system.

The resulting image data set consisted of 16,197 images, of which there were 6,020 AD, 2,389 MCI and 2,888 HA images.

-	Fp1	Fpz	Fp2	-
F7	F3	Fz	F4	F8
T3	C3	Cz	C4	T4
T5	P3	Pz	P4	T6
-	O1	Oz	O2	-

Figure 3: A 5x5 tiled image created from individual electrode position images.

3.6 Deep Learning Model

The DL model used was a modified AlexNet architecture that was optimised for three-class classification, shown in Figure 4. AlexNet, is a deep Convolutional Neural Network (CNN) that consists of eight main layers, five of which are convolutional layers and the remaining three as fully connected layers (Krizhevsky et al., 2012).

The architecture had to be altered by reducing the weights on the final fully connected layer from the standard 1,000 classes to 3. A variety of parameters and hyperparameters were optimised throughout the training of this model and were altered one-by-one over a range of values to provide a model which produced the best output accuracy performance.

3.6.1 Deep Learning Parameters

This section will detail the DL parameter and hyperparameter choices that were optimised to tune the model. The tuned model provided the best results in terms of accuracy, loss and generalisation performance. These were manually adjusted and tested sequentially to obtain the final model. The order of optimisation was chosen by starting with the most impactful parameters first and ending with fine tuning parameters, relating to the performance determined by the final classification accuracy.

The optimiser choice was tuned first, comparing three common optimisers: Stochastic Gradient Descent (SGD), Moving Average of Squared Gradients (RMSProp) and Adaptive Moment Estimation (ADAM). Optimiser choice is important as it is used to update the weights of the model during training. SDG is a gradient descent method whereas RMSProp and ADAM are both adaptive techniques. In short, RMSProp is an extension of SGD and ADAM is a further extension of RMSProp (Ruder, 2017). As expected, the ADAM optimiser produced the best accuracy performance (Kingma & Ba, 2015).

The learning rate is one of the most impactful parameters as it describes how often the weights of the model are updated and is therefore important to be paired with the optimiser. Four learning rates were tested, 1x10^{-3}, 1x10^{-4}, 5x10^{-5} and 1x10^{-5}, with a learning rate of 1x10^{-4} producing the best accuracy performance.

The learning rate can either be kept constant or reduced over time with a policy. This policy can allow greater control over the weights at the later stages of training for fine tuning and avoidance of overfitting. The reduction is split into the Period and the Factor. The period describes the rate of change and the factor describes the magnitude of the reduction. During this step, Factors of 1/2, 1/3 and 1/5 were combined with Periods of 10 and 5 epochs and were compared against a stationary learning rate with no policy. This resulted in a chosen policy of Factor = 1/3 and Period = 10 epochs.

The batch size describes how many sample images are used in each iteration of the training. Smaller batch sizes result in noisy updates and are computationally expensive but they offer a good regularising effect. In contract, larger batch sizes decrease processing time but may have poorer generalisation to unseen data. Batch sizes of 50, 100 and 150 were chosen as an initial starting point, moving on to batch sizes of 16, 32, 64 and 128 due to MATLAB's computational efficiency when working with powers of 2. Although a smaller batch size produced superior accuracy results, a batch size of 64 was used as it produced results that had a good compromise between the advantages of both small and large batch sizes, running each fold in approximately 30 minutes using a GTX 1050Ti GPU.

Validation patience is a method of early stopping and is governed by the number epochs used to train the model. Patience is beneficial as it reduces the chance of overfitting and gives the model enough time to generalise. It does this by stopping the model a certain number of epochs after it has reached its lowest validation loss. The validation patience of 10 allowed the model to reach its minimum validation loss whilst avoiding overfitting. One epoch is comprised of using each sample in the whole data set once.

Weight Learn Rate Factor (WLRF) and Bias Lean Rate Factor (BLRF) were the last parameters that were optimised and relate directly to the final fully connected layers of the network. They determine the relationship between the weight and bias learning rates and the global learning rate, used to fine-tune the result. Values of 10, 20, 30 and 40 for both WLRF and BLRF were tested which varied the final accuracy by ~0.1%. From this, values of 20 for both WLRF and BLRF was chosen as it produced the best combined validation accuracy result.

A summary of each parameter, its value and the justification for the choice can be seen in Table 2.

3.7 Reporting Metrics

To assess the performance of the proposed model, Confusion Matrices and Receiver Operating Characteristic (ROC) curves were generated, in addition to the calculation of four different confusion metrics (Hossin & Sulaiman, 2015).

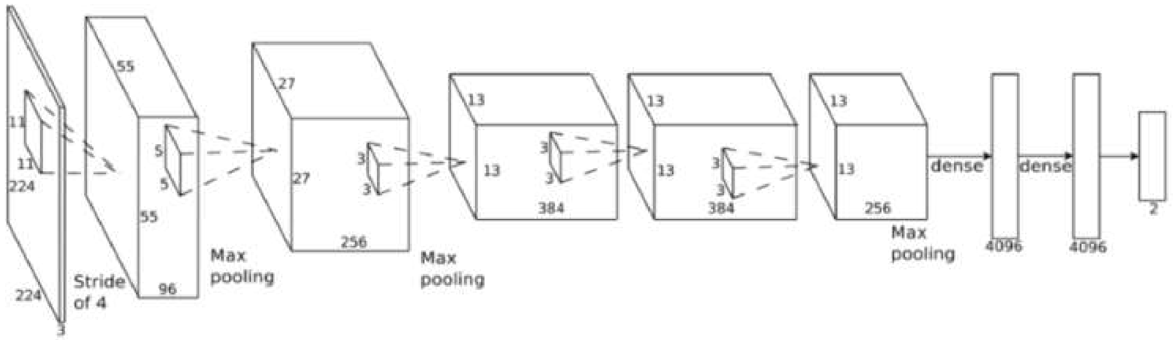


Figure 4: A schematic of the AlexNet deep learning architecture, showing the tensor size as cuboids, kernel size as dashed pyramids and informative descriptions of areas with max pooling and stride lengths, culminating in an example binary classifier (Krizhevsky et al., 2012).

Confusion matrices show the relationship between the predicted classes of each image in the test data set (Predicted Output) in relation to the true class (ground truth) of each image in the test data set (True Output), given by the predefined label. In general terms, the results along the diagonal show correct predictions and off-diagonal values show the incorrect predictions. Within this, there are four bins each prediction can belong to:

- **True Positive (TP)** – Predicting the sample as positive when it is actually positive (i.e. AD classified as AD).
- **True Negative (TN)** – Predicting the sample as negative when it is actually (i.e. MCI classified as MCI and HA classified as HA).
- **False Positive (FP)** – Predicting the sample as positive when it is actually negative (i.e. MCI/HA classified as AD). Also known as a Type I error.
- **False Negative (FN)** – Predicting the same as negative when it is actually positive (i.e. AD classified as MCI/HA). Also known as a Type II error.

These definitions relate directly to the calculation of the per class accuracies shown in Equation (2).

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{FN} + \boldsymbol{TN}} \qquad (2)$$

Where: TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

3.8 Model Validation

The proposed model was validated using k-fold cross-validation (*k*=10) using the 10 splits of the image data set. This method is a statistically based evaluation method which is applied to a DL model with a limited data set to show its expected performance on new data. The data set is shuffled and split into folds randomly, reducing the chance of the order of the data affecting the model results.

For each fold, 8/10 folds were used for training, 1/10 fold was used for validation and 1/10 fold was used for testing. Using the results from them 10 folds, and average (mean) and standard deviation (std) of the reporting metrics were calculated.

4. Results

For each subject in the database, the EEG signals were split into sample epochs of 5 seconds for each electrode. For each sample, time-frequency maps were created using the CWT for each electrode (number of images produced from data analysis = 340,137), then the electrode images were combined using the 10-20 system to create one image per epoch (number of images = 16,197). The images, an example of which is shown in Figure 5, were randomly split into 10 folds, comprising of 8/10 folds training data, 1/10 fold validation data and 1/10 fold test data, ensuring that there was the correct proportion of classes within each split. These were then fed into an optimised AlexNet DL model to predict the classes associated with each image. The model was split into 10 folds and assessed using k-fold cross validation for robustness, producing results in approximately 30 minutes per test.

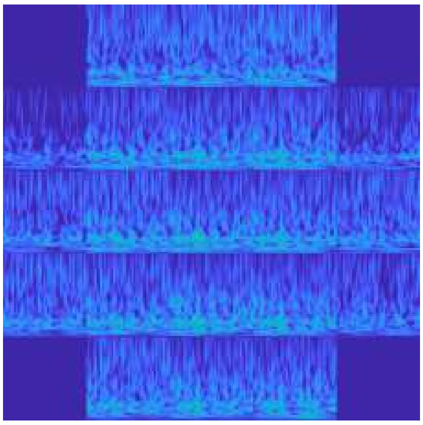


Figure 5: An example output image, showing a 5x5 tiled image of CWT scalograms as detailed in Figure 3.

Table 2: A summary of the optimised parameters, detailing the parameter name, chosen value and justification of the choice.

Parameter	Value	Description
Architecture	AlexNet	Shallowest investigated network but produced the fastest and most accurate results.
Optimiser	ADAM	Proven in experiment and literature to be a high performing optimiser for image classification learning.
Weight Learn Rate Factor	20	Optimised to provide the highest accuracy values.
Bias Learn Rate Factor	20	Optimised to provide the highest accuracy values.
Mini Batch Size	64	Maintains a good relationship between the usage of GPU memory and model fine detail.
Validation Frequency	177	Provides a validation point at the end of every epoch.
Validation Patience	10	Provides enough time for the validation accuracy results to stabilise.
Initial Learning Rate	1e^{-4}	Set to allow for a learning rate curve that asymptotes to zero roughly following a smooth curved trajectory.
Learn Rate Schedule	Piecewise	Allows the learning rate to be altered as the model progresses.
Learn Rate Drop Period	10	Reduces learning rate by 1/3 every 10 epochs which were optimised for this application to avoid overfitting.
Learn Rate Drop Factor	0.33	

The results of the 10 fold cross validation tests are shown in Table 3 and present an overall test accuracy of 98.9 ± 0.4%. The confusion matrices of fold 1-10 are shown in Figure 6a-j respectively. To show additional information about the model, the validation loss and accuracy values were used as a comparrison to the test accuracy. The average validation loss value, 0.07 ± 0.02, is very low, which showed that the model was fit to the data well. Across the folds, the average validation accuracy was 98.8 ± 0.3% which is similar to the overall average test accuracy, indicating that the model has been trained appropriately to stop over or underfitting.

In relation to class specific classification, there was an accuracy for HA of 99.3% ± 0.06%, MCI of 98.3% ± 0.06% and AD of 98.8% ± 0.05%. This shows a very small bias towards the HA and AD classes, indicative of the bias within the original data set. The overall range of 1.0% shows that this bias is minimal, and the skewed data set has been accounted for in the model.

The low standard deviation values across all the accuracy values also shows robustness and consistency between the ten validation tests.

5. Discussion

This paper has presented a classification method with an overall 3-class classification accuracy for AD vs. MCI vs. HA of 98.9% ± 0.04%. The resulting classification accuracy showed minimal bias between class predictions over a large database of 52 AD, 37 MCI and 52 HA subjects. The proposed method extracted artefact free, 5-second-long EEG signals and analysed them using the CWT signal processing method with a Morse mother wavelet. Image maps created from the output scalogram graphs were then used as the input to an optimised DL model based on the AlexNet architecture for 3-class classification. The model was assessed using a 10-fold cross validation method, presenting results that improved upon current reviewed literature within the proposed field.

During the development of this model, alternative signal processing techniques and pretrained DL models were researched and tested. Networks such as 'ResNet-18' and 'GoogLeNet' were investigated and compared to 'AlexNet' leading to surprising results. ResNet-18 has 18 convolutional layers and makes use of residual learning every 2 convolutional layers to skip over layers of the network if required. These address the vanishing gradient problems which occurs when adding a large number of layers to a network (He et al., 2015). GoogLeNet is even larger, with 22 convolutional layers. GoogLeNet introduced a unique model that uses inception modules to reduce the depth of the model by collating convolutional resulting using parallel layers of 1x1, 3x3 and 5x5 kernel filters. This resulted in a model that has 12-time fewer parameters than AlexNet, which has 5,000,000. It also contains two auxiliary classifiers that offer a regularising effect on the network (Szegedy et al., 2014). Despite the increased complexity of both GoogLeNet and ResNet-18 compared to AlexNet, they produced models that took longer to train and had worse accuracy results. This is suggested to be due to more overfitting present due to higher divergence in training and validation loss plots.

Alternative signal processing methods could have been used in this method such as Short-Term Fourier Transform (STFT). The output spectrogram from STFT analysis is very similar to the scalogram from CWT analysis, with the main difference being the resulting frequency and time resolutions. STFT uses a fixed window size resulting in time-frequency

Table 3: Model performance of each fold, detailing each class accuracy and the overall accuracy, including the mean and standard deviation of each category, shown to 1 decimal place.

Fold	AD Class Accuracy (%)	MCI Class Accuracy (%)	HA Class Accuracy (%)	Overall Accuracy (%)
1	99.5	97.7	98.5	98.6
2	99.0	97.9	99.3	98.8
3	99.2	98.1	98.3	98.6
4	99.8	98.1	99.0	99.1
5	99.0	97.2	97.5	98.0
6	99.3	98.6	99.0	99.0
7	99.7	99.1	99.0	99.3
8	99.5	99.1	99.3	99.3
9	98.5	98.1	99.2	98.6
10	99.3	99.1	99.2	99.2
Mean	99.3	98.3	98.8	**98.9**
Std	00.4	00.6	00.5	**00.4**

graphs with uniform time and frequency resolutions. A trade-off between the frequency and time resolutions must be made when choosing the fixed window size for STFT, which is not required for CWT. CWT uses variable, scaled window sizes to create graphs with non-uniform time and frequency resolutions. The non-uniform resolutions allow lower frequency trends in the data to be seen at longer time intervals and higher frequency trends to be seen at shorter time intervals. Due to the non-stationary nature and varied frequency range of EEG signals, CWT was chosen as the most appropriate signal processing method.

In comparison to the literature reviewed in Section 2, the results improve on the accuracy of HA vs. MCI vs. AD presented by the most recent study from Ieracitano et al. (2020), who produced a combined bispectrum and CWT feature model, with a maximum average accuracy of 89%. Their paper uses more subjects (*n*=189) in comparison to the subjects in this study's database (*n*=141) and is balanced between classes which means the data are better suited for DL applications. The results are also calculated over a range of 10 tests which gives reliability to their results. The results also improve on the maximum accuracy found in literature, from Bi & Wang (2019) at 95% accuracy of HA vs. MCI vs. HA using an alternative method and dataset. The input to their bespoke machine learning model uses a unique combination of spectral topography maps to produce this result, however, lacks depth as the subject pool is very small (*n*=16) in comparison to our data set. Both papers included a statistical analysis of the results, such as ROC curves and confusion matrices, further allowing direct discussions and comparison between the papers.

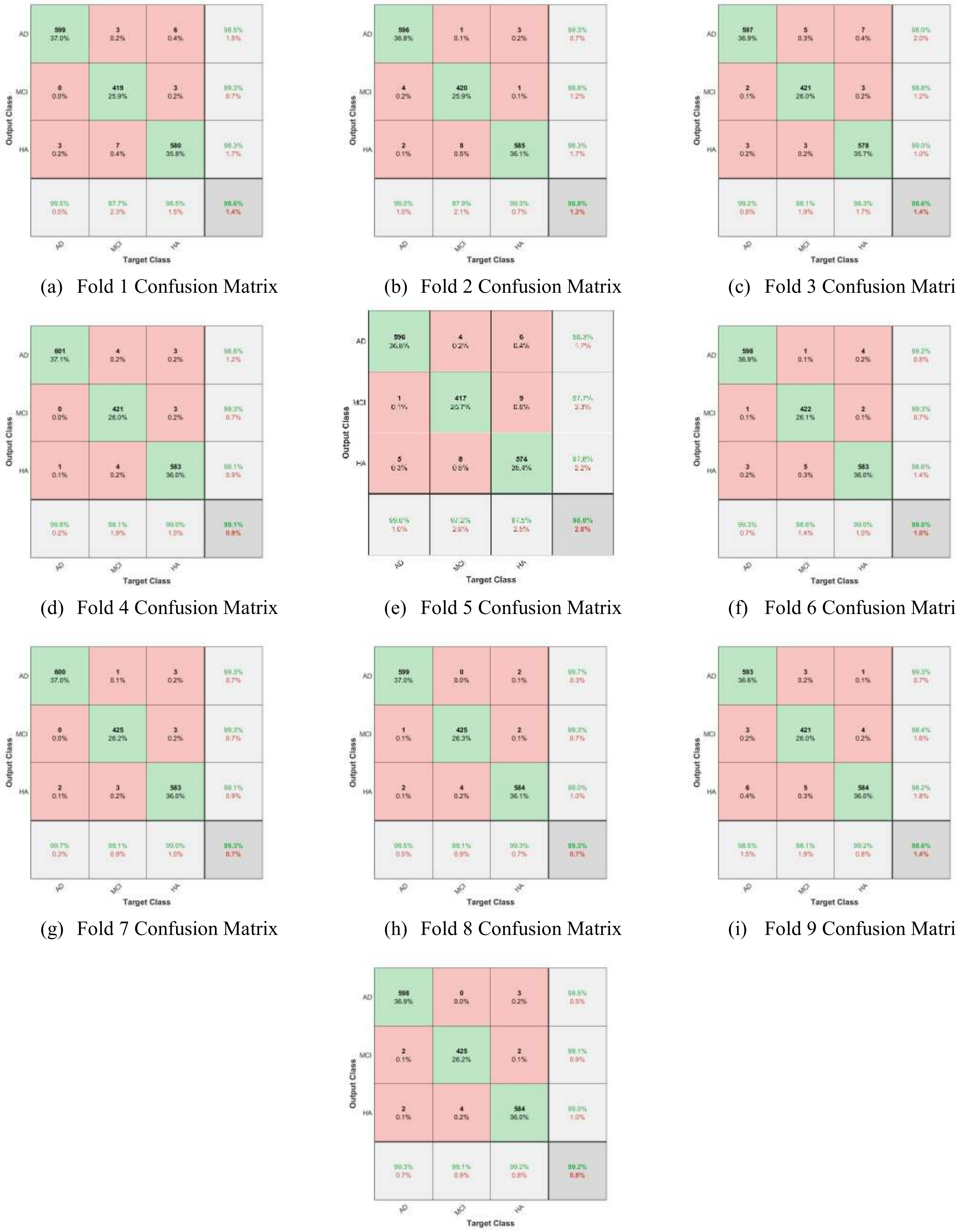
The other papers in Section 2 by Kim & Kim (2018), Fan et al. (2018) and Zhao & He (2015) are all binary classifiers. As this research study produced three-class classification of AD vs. MCI vs. HA, it has a clear impact advantage. The signal processing methods of RP from Kim & Kim (2018) and MSE from Fan et al. (2018) are less computationally complex compared to CWT. Conversely, their respective DNN and LASSO classification models were more complex than the proposed CNN in this study. All three papers also use databases that are much smaller compared to the database used in this study, despite this still produce lower accuracy results.

Despite the promising results presented, our study has limitations. The recording of scalp EEG signals was conducted in a controlled environment, requiring clinical space and trained professionals which can be time consuming and expensive. Bias is an important topic in medical AI and DL and should be reduced as much as possible. By removing human interaction in the signal processing and classification stages, this study has attempted to avoid bias; however, there were still areas for improvement. The database required initial input of labels for this supervised learning method, which means that there was a possibility of errors within the diagnosis via the standard clinical criteria explained in Section 3.2 (Batum et al., 2015). The educational level of the subjects, which is a known contributor to AD, has not been assessed in the data pre-processing. The similarities and differences between the 5 second epochs within the 10-minute samples have also not been assessed, which could affect the results. Furthermore, it is unrealistic to compare these results directly to the in vivo diagnostic accuracy of AD (77%) as the results within this report rely on pre-defined labels using gold standard techniques. To be used in a clinical setting, this method would need to be validated using additional data, ideally progressing onto a model that produces significant results regardless of the patient's demographical information such as age, location, and educational history.

6. Conclusions

In this study, a signal processing method combined with an optimised DL model has been developed that presents accuracy results of 98.9%, currently higher than any results presented in literature within this field.

Other methods were tested and discussed, including spectrogram time-frequency images, alternative DL architectures GoogLeNet and ResNet-18 and a variety of different hyperparameter values. These different strategies did



(a) Fold 1 Confusion Matrix

(b) Fold 2 Confusion Matrix

(c) Fold 3 Confusion Matrix

(d) Fold 4 Confusion Matrix

(e) Fold 5 Confusion Matrix

(f) Fold 6 Confusion Matrix

(g) Fold 7 Confusion Matrix

(h) Fold 8 Confusion Matrix

(i) Fold 9 Confusion Matrix

(j) Fold 10 Confusion Matrix

Figure 6: (a-j) Confusion matrices showing the model predictions (Output Class) against the ground truth (Target Class) from Folds 1-10 respectively, relating to Table 3.

not produce superior accuracy results but provided a useful comparison to the final model.

This study has shown some very promising results but further work is required to progress this research, with the end goal to produce a clinically accurate tool that could be used to aid healthcare professionals in providing better diagnosis for elderly patients with symptoms that are typical of MCI or AD. Recommendations for future work could encompass the following:

- Test this model as a 2-class problem by comparing the AD vs. MCI accuracies against the MCI vs. HA accuracies. This leads to a hypothesis that a lower number of classification categories would result in improved accuracies.
- Explore alternative ways to create RGB images from EEG signals, possibly by combining other signal processing methods quantifying changes in these signals related to AD.
- Explore additional pre-trained networks or bespoke DL architectures.
- Use the model on other databases (out of sample data) to further understand its reliability and accuracy.
- Test the use of automatic hyperparameter selection using methods such as Bayesian optimisation.
- Create an 'International Standard' data set for training and testing similar systems for the diagnosis of AD.

The findings presented in this study have significantly added to the continued knowledge surrounding DL of EEG signals for AD diagnosis. The increased accuracy results show promising outcomes for future applications of DL to the classification and diagnosis of AD. This could eventually help combat the resource-intensive and human dependant methods that are currently used, ultimately providing a quantitative probability value for the diagnosis of a patient.

Acknowledgements

We would like to acknowledge the input of Viktor Markovskij for the age-matching of the EEG database between AD, MCI and HA subject groups.

References

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrill, M.C., Thies, B. & Phelps, C.H., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), pp.270-79.

Batum, K., Cinar, N., Sahini, Ş., Cakmak, M.A. & Karsidag, S., 2015. The connection between MCI and Alzheimer disease: neurocognitive clues. *Turkish Journal of Medical Sciences*, 2105(45), pp.1137-40.

Bi, X. & Wang, H., 2019. Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. *Neural Networks*, 114(2019), pp.119-35.

Britton, J.W., Frey, L.C., Hopp, J.L., Korb, P., Koubeissi, M.Z., Lievens, W.E., Pestana-Knight, E.M. & Louis, E.K.S., 2016. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. 1st ed. Chicago: American Epilepsy Society.

Capecci, E., Morabito, F.C., Campolo, M., Mammone, N., Labate, D. & Kasabov, N., 2014. Feasibility Study of Using the NeuCube Spiking Neural Network Architecture for Modelling Alzheimer's Disease EEG Data. In *Italian Workshop on Neural Networks (WIRN)*. Salerno, 2014.

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F.J. & Falk, T.H., 2018. Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment. *Disease Markers*, 2018(5174815), pp.1-26.

Cassani, R., Falk, T.H., Fraga, F.J., Cecchi, M., Moore, D.K. & Anghinah, R., 2017. Towards automated electroencephalography-based Alzheimer'sdisease diagnosis using portable low-density devices. *Biomedical Signal Processing and Control*, 33(2017), pp.261-71.

Craik, A., He, Y. & Contreras-Vidal, J.L., 2019. Deep Learning for EEG Classification Tasks: A Review. *Journal of Neural Engineering*, 16(031001), pp.1-28.

Dauwels, J., Vialatte, F. & Cichocki, A., 2010. Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing? *Current Alzheimer's Research*, 7(6), pp.487-505.

DeTure, M.A. & Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. *Molecular Degeneration*, 14(32), pp.1-18.

DeTure, M.A. & Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration*, (2019) 14(32), pp.1-18.

Dodge, S. & Karam, L., 2017. *A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions*. arXiv.

Dukart, J., Schroeter, M.L. & Mueller, K., 2011. Age Correction in Dementia – Matching to a Healthy Brain. *PLOS ONE*, 6(7), pp.1-9.

Fan, M., Yang, A.C., Fuh, J.-L. & Chou, C.-A., 2018. Topological Pattern Recognition of Severe Alzheimer's Disease via Regularised Supervised Learning of EEG Complexity. *Frontiers in Neuroscience*, 12(685), pp.1-10.

GBD 2016 Dementia Collaborators, 2019. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*, 18(1), pp.88-106.

Greco, A., Costantino, D., Morabito, F.C. & Versaci, M., 2003. A Morlet wavelet classification technique for ICA filtered sEMG experimental data. In *Proceedings of the IJCNN 2003*. Portland, 2003. IEEE.

He, K., Zhang, X., Ren, S. & Sun, J., 2015. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2015.

Hossin, M. & Sulaiman, M.N., 2015. A Review On Evaluation Metrics For Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5(2), pp.1-11.

Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A. & Morabito, F.C., 2019. A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. *Neurocompuuting*, 323(2019), pp.96-107.

Ieracitano, C., Mammone, N., Hussain, A. & Morabito, F.C., 2020. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks*, 123(2020), pp.176-90.

Jack Jr, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M. & Sperling, R., 2018. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), pp.535-62.

Jack Jr, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Feldman, H.H., Frisoni, G.B., Hampel, H., Jagust, W.J., Johnson, K.A., Knopman, D.S., Petersen, R.C., Scheltens, P., Sperlin, R.A. & Dubois, B., 2016. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, 87(5), pp.539-47.

Kim, D. & Kim, K., 2018. Detection of Early Stage Alzheimer's Disease using EEG Relative Power with Deep Neural Network. In *IEEE Engineering in Medicine and Biology Society Annual International Conference*. Honolulu, Hawaii, 2018.

Kingma, D.P. & Ba, J.L., 2015. ADAM: A Method for Stochastic Optimisation. In *International Conference on Learning Representations*. San Diego, 2015.

Kramer, M.A., Chang, F.-L., Cohen, M.E., Hudson, D. & Szeri, A.J., 2007. Synchronization Measures of the Scalp Electroencephalogram Can Discriminate Healthy From Alzheimer's Subjects. *International Joirnal of Neural Systems*, 17(2), pp.1-9.

Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional. *Advances in neural information processing systems*, 25(2), pp.1-9.

Lilly, J.M. & Olhede, S.C., 2010. On the Analytic Wavelet Transform. *IEEE Transactions on Information Theory*, 56(8), pp.4135-56.

Mallat, S., 2009. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd ed. Burlington, MA: Elsevier.

Mathworks, 2020. *parula*. [Online] Available at: https://www.mathworks.com/help/matlab/ref/parula.html [Accessed 08 Nov 2020].

Morabito, F.C., Campolo, M., Ieracitano, C., Ebadi, J.M., Bonanno, L., Bramanti, A., Salvo, S.D., Mammone, N. & Bramanti, P., 2016a. Deep Convolutional Neural Networks for Classification of Mild Cognitive Impaired and Alzheimer's Disease Patients from Scalp EEG Recording. In *IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI)*. Bologna, 2016a.

Olhede, S.C. & Walden, A.T., 2002. Generalized Morse Wavelets. *IEEE Transactions on Signal Processing*, 50(11), pp.2661-70.

Parra, M.A., Butler, S., McGeown, W.J., Nicholls, L.A.B. & Robertson, D.J., 2019. Globalising strategies to meet global challenges: the case of ageing and dementia. *Journal of Global Health*, 9(2), p.020310.

Petersen, R.C. & Knopman, D.S., 2006. MCI is not a clinically useful concept. *International Psychogeriatrics*, 18(3), pp.393-414.

Petersen, R.C. & Negash, S., 2008. Mild Cognitive Impairment: An Overview. *CNS Spectrums*, 13(1), pp.45-53.

Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T. & Prina, M., 2015. *The Global Impact of Dementia: Summary Sheet*. London: Alzheimer's Disease International.

Rossini, P.M., Iorio, R.D., Vecchio, F., Anfossi, M., Babiloni, C., Bozzali, M., Bruni, A.C., Cappa, S.F., Escudero, J., Fraga, F.J., Giannakopoulos, P., Guntekin, B., Logroscino, G., Marra, C., Miraglia, F., Panza, F., Tecchio, F., Pascual-Leone, A. & Dubois, B., 2020. Early diagnosis of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. Report from the IFCN-sponsored panel of experts. *Clinical Neurophysiology*, 131(6), pp.1287-310.

Ruder, S., 2017. An overview of gradient descent optimization. *arXiv*, 1(1609.04747), pp.1-14.

Sabbagh, M.N., Lue, L.-F., Fayard, D. & Shi, J., 2017. Increasing Precision of Clinical Diagnosis of Alzheimer's Disease Using a Combined Algorithm Incorporating Clinical and Novel Biomarker Data. *Neurology and Therapy*, 6(1), pp.83-95.

Stanford Vision Lab, 2010. [Online] Available at: http://image-net.org/challenges/LSVRC/2010/ [Accessed 24 DEC 2019].

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A., 2014. Going Deeper with Convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2014.

Weller, J. & Budson, A., 2018. Current understanding of Alzheimer's disease diagnosis and treatment. *F100Research*, 7(F1000 Faculty Rev: 1161), pp.1-9.

Zhao, Y. & He, L., 2015. Deep Learning in the EEG Diagnosis of Alzheimer's Disease. In C.V. Jawahar & S. Shan, eds. *Computer Vision – ACCV 2014 Workshops*. 1st ed. Singapore: Springer. pp.340-53.