



Structural MRI Texture Analysis for Detecting Alzheimer's Disease

Joana Silva¹ · Bruno C. Bispo² · Pedro M. Rodrigues¹ · for the Alzheimer's Disease Neuroimaging Initiative

Received: 28 November 2022 / Accepted: 6 April 2023
© The Author(s) 2023

Abstract

Purpose: Alzheimer's disease (AD) has the highest worldwide prevalence of all neurodegenerative disorders, no cure, and low ratios of diagnosis accuracy at its early stage where treatments have some effect and can give some years of life quality to patients. This work aims to develop an automatic method to detect AD in 3 different stages, namely, control (CN), mild-cognitive impairment (MCI), and AD itself, using structural magnetic resonance imaging (sMRI).

Methods: A set of co-occurrence matrix and texture statistical measures (contrast, correlation, energy, homogeneity, entropy, variance, and standard deviation) were extracted from a two-level discrete wavelet transform decomposition of sMRI images. The discriminant capacity of the measures was analyzed and the most discriminant ones were selected to be used as features for feeding classical machine learning (cML) algorithms and a convolution neural network (CNN).

Results: The cML algorithms achieved the following classification accuracies: 93.3% for AD vs CN, 87.7% for AD vs MCI, 88.2% for CN vs MCI, and 75.3% for All vs All. The CNN achieved the following classification accuracies: 82.2% for AD vs CN, 75.4% for AD vs MCI, 83.8% for CN vs MCI, and 64% for All vs All.

Conclusion: In the evaluated cases, cML provided higher discrimination results than CNN. For the All vs All comparison, the proposed method surpasses by 4% the discrimination accuracy of the state-of-the-art methods that use structural MRI.

Keywords Alzheimer's disease · Mild-cognitive impairment · Early detection · Magnetic resonance imaging · Co-occurrence matrix · Texture analysis

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

✉ Pedro M. Rodrigues
pmrodrigues@ucp.pt

Joana Silva
joana.rsilva@hotmail.com

Bruno C. Bispo
bruno.bispo@ufsc.br

¹ CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Rua de Diogo Botelho, 1327, Porto 4169-005, Portugal

² Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, Florianópolis, SC 88040-370, Brazil

1 Introduction

Approximately 70% of all dementia cases worldwide are caused by Alzheimer's disease (AD), a progressive neurodegenerative illness. During its early stages - mild-cognitive impairment (MCI) - the condition is asymptomatic. Even though several studies have been conducted, a cure has not yet been discovered [1]. In general, people aged 65 and older live 4 to 8 years after being diagnosed with AD. Nonetheless, some people can live up to 20 years with AD. This extended duration before death significantly impacts public health as a considerable part of that period is spent in a state of dependence and disability [2]. It is therefore imperative to find more precise and reliable means of diagnosing AD to minimize its impact.

AD has 3 stages: (1) *the pre-clinical AD*, distinguished by the asymptomatic period that occurs between the initial brain lesions and the appearance of the first symptoms; (2) *MCI*, the pre-dementia state, in which individuals have cognitive deficits greater than those that naturally emerge with age, but do not fit the criteria imposed for the diagnosis of

AD; (3) *Dementia due to AD* (or simply AD in this study), characterized by severe symptoms.

Dementia due to AD has 3 different phases. The mild phase corresponds to the period where the individual is still operational in several areas, but, for safety reasons, may need help in certain activities. The moderate phase is distinguished by the difficulty in communicating and performing routine tasks. In advanced stages of the disease, individuals require 24-hour care as damage to the areas of the brain responsible for movement emerges [2].

The diagnosis of the disease can be performed in numerous ways. Usually, the main risk factors are pondered through physical and history examinations from the individual and his family. With these risks and through neurological and cognitive exams, it is possible to discard other causes of dementia and evaluate the stage of AD. The most common cognitive test is the Mini-Mental State Examination (MMSE). Its scores range between 0 and 30. Scores on the higher end indicate a higher cognitive function, while lower scores mean more severe cases of dementia [3]. Additionally, there are other employed methods to identify both neurodegeneration and amyloid deposition, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Electroencephalogram (EEG), and Cerebrospinal Fluid (CSF) Analysis [1].

Imaging techniques are used as non-invasive means for AD diagnosis. The imaging modalities are currently focusing on the identification of amyloid deposition or neurodegeneration, e.g., structural MRI allows the computation of atrophy and changes measurements in tissue [4]. MRI-based atrophy measurements are considered valid markers of disease state and progression since atrophy seems to be an inevitable and intrinsic factor of progressive neurodegeneration. Besides that, changes in structural measures, such as ventricular enlargement, hippocampus, entorhinal cortex, whole brain, and temporal lobe volumes, can be associated with modifications in cognitive performance [5]. In general, atrophy progression assessed by MRI is being widely used as an efficiency and safety outcome measure in clinical trials. Nonetheless, out of all the MRI markers, AD hippocampal atrophy is pondered as the best established and validated [6, 7].

Regarding MRI state-of-the-art studies done to diagnose AD, Ruiz et al. [8] proposed an automated computer-aided diagnosis (CAD) system using MRI to extract features from regions of interest (ROI). Several machine learning classifiers were used, but VAF-FS, Random Forest (RF), and XGBoost classifiers were the ones that suit better the problem with an accuracy of 85.86% in the Healthy Controls (CN) vs AD comparison, 71.92% in CN vs MCI, and 68.92% in MCI vs AD.

Thapa et al. [9] used neuropsychological testing coupled with MRI. The machine learning classifier that performed best

was the Support Vector Machine (SVM) feed with information from left and right hippocampal volume and MMSE scores. The obtained discrimination accuracies were 99.2% for CN vs AD, 78.5% for CN vs MCI, and 91.3% for MCI vs AD.

Hon and Khan [10] used MRI images and extracted their entropy to characterize AD activity. Two Convolution Neural Network (CNN) architectures were used (VGG and Inception) and the reached discrimination accuracy was 96.5% for the CN vs AD comparison. Amini et al. [11] used functional MRI (fMRI) images and extracted the average and the standard deviation of cortical thickness, cortical parcel volume, white matter, and surface area. These features were used to feed both machine learning and CNN algorithms. It was found that the proposed CNN obtained a discrimination accuracy of 96.7% for the CN vs AD comparison.

Al-Khuzai et al. [12] used MRI images and fed the proposed CNN with the 2D image slices. Thus, the discrimination accuracy achieved was 99.3% for the CN vs AD comparison. Liu et al. [13] used MRI images to extract hippocampal features. The chosen classifier was a 3D Densely CNN (DenseNet 3D). The discrimination accuracies obtained were 88.9% for CN vs AD and 76.2% for CN vs MCI. Qiu et al. [14] used MRI images and fed a Fully CNN with AD probability maps. A discrimination accuracy of 87.0% was obtained for CN vs AD.

Vaithinathan and Parthiban [15] extract ROI-based texture measures from MRI images. For the classification, they used several algorithms such as RF, linear SVM, and k-nearest neighbors (KNN). The discrimination accuracy achieved was 87.39% for CN vs AD, 64.74% for CN vs MCI, 63.41% for MCI vs AD, and 66.38% for converter MCI (cMCI) vs stable MCI (sMCI). A Multi-slice ensemble learning was designed by Kang *et al.* [16] to obtain spatial features to train CNNs models. This approach achieved accuracy values of 90.36%, 77.19%, and 72.36% when classifying AD vs CN, AD vs MCI, and MCI vs CN, respectively. Ebrahimi *et al.* [17] applied several deep sequence-based CNN models for AD vs CN with 91.78% accuracy.

In this sense, the main purpose of the present work is to develop an artificial intelligence system that enables to detect AD in MCI and Dementia Stages (AD) stages, using sMRI texture features. The paper is structured as follows: Sect. 2 describes the used MRI database; Sect. 3 focuses on the image processing methodology and the classification process; Sect. 4 discusses the obtained results, lastly, Sect. 5 concludes the work.

2 Materials

The data used in this work are the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a

public–private partnership with the aim of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild-cognitive impairment and early Alzheimer's disease.

Regarding the MRI scan, time overall was about 45 min per subject and session. Each exam undergoes quality control so that in case of, for example, subject motion or poor anatomic coverage, the scan is considered unusable. The database, released in February 2021, consists of 89 subjects scanned longitudinally at 3T with a 3-year follow-up, in which 24 are healthy control subjects, 44 are MCI patients, and 21 are AD patients (patients diagnosed with dementia due to AD). The demographic data of the 3 groups are summarized in Table 1.

3 Methods

The proposed methodology is divided into 3 main steps: (1) preprocessing, (2) wavelet decomposition and feature extraction, and (3) feature selection and classification. Figure 1 summarizes the methodology implementation steps.

3.1 Preprocessing

The dataset was loaded on FreeSurfer 7.1.1 software (freely available online at <https://surfer.nmr.mgh.harvard.edu/>) to decompose each 3D subject data into 2D slices

comprising 3 different anatomical planes, namely, coronal, sagittal, and axial, and then to execute the skull stripping process on the 2D slice MR images. An example of skull stripping is illustrated in Fig. 2.

The resulting 2D slice images were loaded to Matlab® 2019b software. These images were then filtered by the median filter with a 3×3 kernel to remove noise [18]. Subsequently, they were filtered by the imadjust filter to adjust the image intensity values to all scales according to [19]

$$P_{adj}(m, n) = B + \frac{P(m, n) - L}{H - L} * (T - B), \quad (1)$$

where $P(m, n)$ is the input image, $P_{adj}(m, n)$ is the output image, m and n are the image pixel indices, and H and L are the maximum and the minimum pixel level in the original image, respectively, and $T = 255$ and $B = 0$ are the maximum and the minimum pixel levels in the desired image.

3.2 Wavelet Decomposition

The discrete wavelet transform (DWT) was chosen to describe the input images because it is possible to maintain higher resolution at low-frequency bands [20]. It can be obtained by restraining scale (s) and translation (τ) parameters to a discrete lattice with $s = 2^{-m}$ and $\tau = n \cdot p2^{-m}$, where m and n are integers. Hence, for a discrete-time

Table 1 Database demographic data overview

Group	# subjects	Age average \pm SD	Age Range	Gender		MMSE average \pm SD
				F	M	
CN	24	78.2 \pm 4.64	71–88	14	10	29.50 \pm 0.66
MCI	44	73.8 \pm 8.92	56–90	10	34	24.52 \pm 2.91
AD	21	76.2 \pm 8.34	59–89	14	7	19.81 \pm 6.46

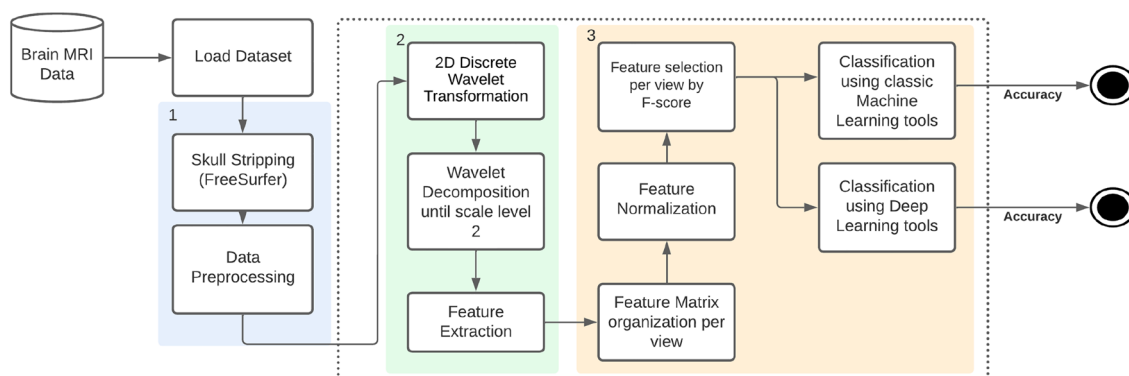
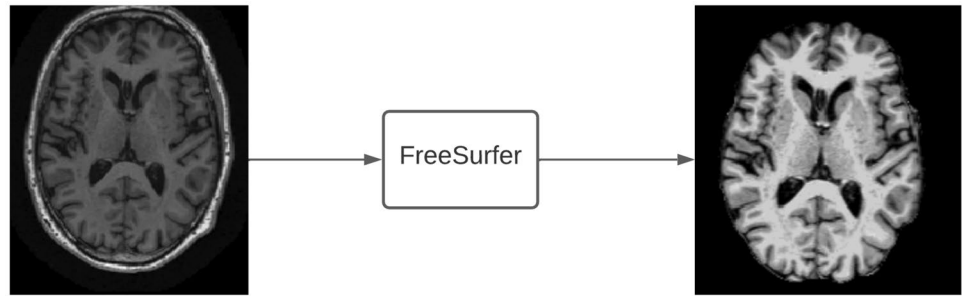


Fig. 1 Image processing methodology workflow

Fig. 2 Example of the skull stripping process



signal $f(n)$, the wavelet decomposition on I octaves is given by

$$f(n) = \sum_{i=1 \text{ to } I} \sum_{k \in \mathbb{Z}} c_{i,k} g[n - 2^i k] + \sum_{k \in \mathbb{Z}} d_{I,k} h_I[n - 2^I k] \quad (2)$$

where $c_{i,k}$ and $d_{I,k}$ correspond to the coefficients of the approximation component and coefficients of the detail component, respectively [21, 22]. These coefficients are given by

$$c_{i,k}(n) = \sum_n f(n) G_i^*[n - 2^i k] \quad (3)$$

$$d_{I,k}(n) = \sum_n f(n) H_I^*[n - 2^I k] \quad (4)$$

The parameters i and k indicate the wavelet scale and translation factors, respectively. Besides that, G_i characterizes the coefficients of the low-pass and H_i the coefficients of the high-pass filters. Every wavelet type and family is different with regard to these filters [21, 23].

Since images are two-dimensional, the DWT is applied to images both vertically and horizontally. The result is four images (subbands) with half the width and the height, one of which is a decimated copy of the image (LL), and the 3 remaining contain information about the details - horizontal (HL), vertical (LH) and diagonal (HH). At each subsequent step of decomposition, the LL subband is replaced by four smaller subbands, so the total number of subbands increases by 3 (see Fig. 3).

In this work, for all participants, in each plane, every image has been decomposed by the DWT until level 2, producing in this way 8 images, as illustrated in Fig. 3.

3.3 Features Extraction

For each of the 89 study participants, 243 images were used for feature extraction: the 27 original plane images (9 images of each of the 3 planes) and the 8 images resulted from the DWT decomposition of each plane image. From each image, 9 texture features were extracted: contrast, correlation, energy, homogeneity, entropy, line and column variances, and line and column standard deviations. Therefore, for each possible mother wavelet used in the DWT decomposition, 2187 features (729 per plane) were computed for each study participant.

The features were computed from the gray level co-occurrence matrix (GLCM), which is a statistical method that considers the spatial relationship of pixels and is employed to describe the texture of an image [24]. Each element $\{i, j\}$ of the GLCM P_{ij} represents the frequency by which the pixel with gray level i is spatially related to the pixel with gray level j [24]. The formula and description of the features are summarized in Table 2, where

$$\mu_i = \sum_{i=1}^N \sum_{j=1}^N iP_{ij} \quad (5)$$

and

Fig. 3 Image Wavelet Decomposition

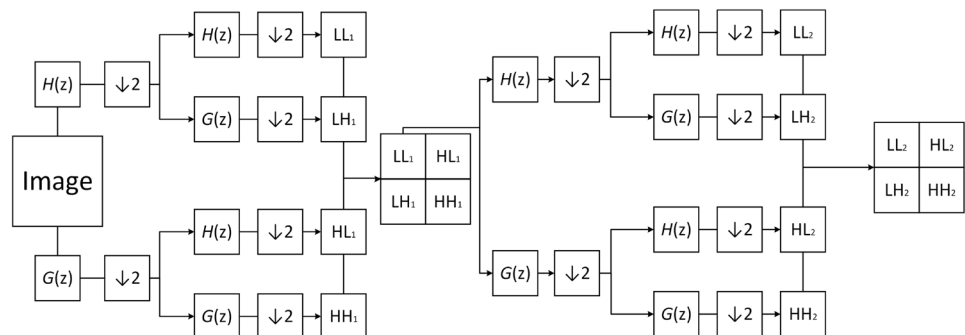


Table 2 Features overview description

Feature	Formula	Description
Contrast	$\sum_{i=1}^N \sum_{j=1}^N P_{ij}(i-j)^2$	Measures the local variations between pixels
Correlation	$\sum_{i=1}^N \sum_{j=1}^N P_{ij} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right]$	Estimates the combined probability occurrence of the indicated pixel pairs
Energy	$\sum_{i=1}^N \sum_{j=1}^N P_{ij}^2$	Specifies the sum of squared elements in the GLCM
Homogeneity	$\sum_{i=1}^N \sum_{j=1}^N \frac{P_{ij}}{1+(i-j)^2}$	Measures the nearness of GLCM elements distribution to the GLCM diagonal
Entropy	$-\sum_{i=1}^N \sum_{j=1}^N P_{ij} \ln P_{ij}$	Assesses the randomness of an intensity image
Variance	$\sigma_i^2 = \sum_{i=1}^N \sum_{j=1}^N P_{ij}(i-\mu_i)^2,$ $\sigma_j^2 = \sum_{i=1}^N \sum_{j=1}^N P_{ij}(j-\mu_j)^2$	Measures the dispersion of the elements of an image
Standard Deviation	$\sigma_i,$ σ_j	Measures the dispersion of the image elements

$$\mu_j = \sum_{i=1}^N \sum_{j=1}^N j P_{ij} \quad (6)$$

are mean values of the GLCM.

For each of the 3 planes (coronal, sagittal, and axial) of each study participant, each feature was averaged considering the 9 original images and the 72 images resulting from their DWT decompositions. This leads to 9 average features (1 value per feature) for each plane of each study participant. These average features were used for the selection processes of mother wavelets and features to improve classification results. The averaging processes per plane were applied to decrease the data dimensionality and consequently improve the execution time of these selection processes.

3.4 Wavelet Selection Process

The extracted features were used for binary classification within the pairs CN vs MCI, AD vs MCI, and CN vs AD, and for multi-class classification All vs All. All binary classifications were performed using exclusively the information of each of the 3 planes (coronal, sagittal, and axial) and also using together the information of the 3 planes.

Since the values of each feature depend on the mother wavelet used in the DWT decomposition, a search to find the five wavelets that result in features with greater discriminant capacity considering all study group pairs (CN vs MCI, AD vs MCI, CN vs AD and All vs All) and all study planes (coronal, sagittal, axial, and 3 planes) was performed. The evaluated wavelet families were Haar, Daubechies (Db), Symlets (sym), Coiflets (Coif), Biorthogonal (Bior), Reverse biorthogonal (rbio), Meyer, and Fejer-Korovkin (fk). The average features were used for this purpose.

The average values of each feature were separated for each combination of study group pair, study plane, wavelet, feature, and subband (or full-band). Each combination that uses only 1 plane leads to 1 value per study participant. Each combination that uses together the 3 planes leads to 3 values per study participant. Within each combination, including all study participants, the average values were normalized using z-score [25] and then applied to the Kruskal-Wallis (KW) test [26]. The KW test was used to determine if the null hypothesis that the data of the study groups come from the same distribution is accepted. In this test, p-values lower than 0.05 indicate that there is a significant difference between the distributions and then the null hypothesis is rejected [26]. It is worth mentioning that, for the multi-class study group All vs All, the p-values were corrected by the Bonferroni method [27].

Figure 4 shows the 15 cases with the highest number of average features that reject the null hypothesis and the corresponding wavelet. It is observed that the five wavelets with the highest number of significant features were Biorthogonal 1.1, Reverse Biorthogonal 1.1, Reverse Biorthogonal 1.3, Reverse Biorthogonal 1.5, and Reverse Biorthogonal 3.1. These wavelets were chosen for feature selection and classification procedure steps.

3.5 Features Selection and Classification

As mentioned earlier, classification within each study group pair (CN vs MCI, AD vs MCI, CN vs AD, and All vs All) was carried out for each study plane (coronal, sagittal, axial, and 3 planes). For improving the execution time and the classification results, for each combination of study group pair and study plane, a search was carried out to find the features, computed through the five selected wavelets, that result in the highest classification accuracy. Once again, the average features were used for selection purposes.

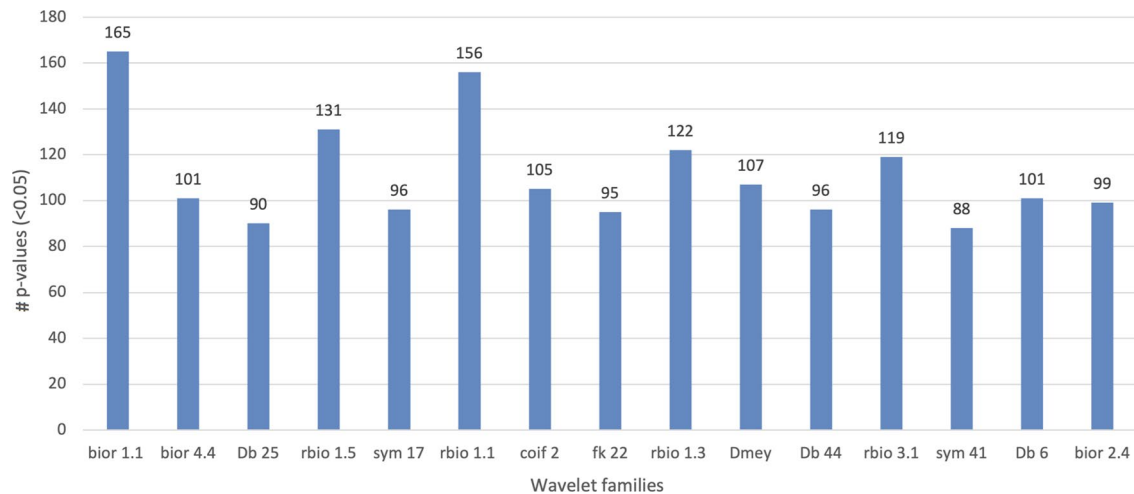


Fig. 4 Best performances in the Kruskal-Wallis test and the corresponding wavelets

The non-normalized average values of each feature were separated for each combination of the study group pair and study plane. Each combination initially had 369 features (9 features \times 8 images resulting from DWT decomposition \times 5 wavelets + 9 features \times 1 original plane image) for each plane of each study participant included in the study group pair. Within each combination, including all study participants belonging to the corresponding study group pair, the average values were normalized using z-score [25]. Then for each combination, including all study participants belonging to the corresponding study group pair, the normalized average values of all features were applied as inputs to a cascade of one F-score algorithm [28] and one classical machine learning (cML) algorithm to select, according to the maximum classification accuracy, the best set of features. The F-score algorithm individually assesses and rates the features based on their F-score. The features with an F-score value above the average are chosen as the relevant features [28].

The number of features selected by the f-score algorithm ranged from 2 to 9 in unit steps and from 10 to all in steps of 5. The cML algorithms were different configurations of decision trees, discriminant analysis, naive-Bayes, support vector machines (SVM), k -nearest neighborhood (KNN), and ensemble. In addition to the cML algorithms, a convolution neural network (CNN) was also applied. For each combination of study group pair and study plane, the CNN was fed with the sets of selected features that, used as inputs to the cML algorithms, led to the best classification result. The classifiers and their configurations are described in Table 3. In all cases, in order to verify the generalization capacity of the classifiers, a leave-one-out cross-validation procedure was used, a well-known process that allows the use of the whole dataset for testing, without leakage between train and test sets [29].

4 Results and Discussion

For each combination of study group pair and study plane, the highest classification accuracy achieved using the cML algorithms, and the corresponding number of selected features (ft), are shown in Table 4. The classification accuracy achieved employing the CNN, and the corresponding number of selected features (ft) and study plane, are shown in Table 5.

Scrutiny of Table 4 reveals that, for the study group pair CN vs AD, the highest classification accuracy achieved through the cML algorithms was 93.3% using 35 features from the sagittal plane and also with 115 features selected from the 3 planes, both with bagged trees classifiers. The lowest classification accuracy achieved through the cML algorithms was 77.8% using the axial plane. For this study group pair, as indicated in Table 5, the highest classification accuracy achieved through the CNN algorithm was 82.2% using the 115 features selected from the 3 planes.

For the pair AD vs MCI, it is observed from Table 4 that the highest classification accuracy achieved through the cML algorithms was 87.7% using 80, 95, and 140 features from the coronal plane and the quadratic SVM classifier. The lowest classification accuracy achieved through the cML algorithms was 78.5% using the sagittal plane. For this study group pair, as indicated in Table 5, the highest classification accuracy achieved through the CNN algorithm was 75.4% using the 95 features selected from the coronal plane.

Regarding the pair CN vs MCI, it is observed from Table 4 that the highest classification accuracy achieved through the cML algorithms was 88.2% using 30, 40, 60, 65, 70, 75, 80, 85, and 90 features selected from the coronal

Table 3 Used classifiers and optimal parameters

Classifier		Optimal parameters
Decision Trees	Fine Tree - FT	Maximum number of splits = 150
	Medium Tree - MT	Maximum number of splits = 150
	Coarse Tree - CT	Maximum number of splits = 150
Discriminant Analysis	Linear Discriminant - LD	Covariance structure: Full
	Quadratic Discriminant - QD	Covariance structure: Full
	Logistic Regression - LR	Covariance structure: Full
Naive Bayes	Gaussian Naive Bayes - GNB	-
	Kernel Naive Bayes - KNB	-
SVM	Linear SVM - LSVM	Box constraint level = 3
	Quadratic SVM - QSVM	Box constraint level = 3
	Cubic SVM - CSVM	Box constraint level = 4
	Fine Gaussian SVM - FGSVM	Box constraint level = 3
	Medium Gaussian SVM - MGSVM	Box constraint level = 3
	Coarse Gaussian SVM - CGSVM	Box constraint level = 1
KNN	Fine KNN - FKNN	Number of neighbors = 3
	Medium KNN - MKNN	Number of neighbors = 3
	Coarse KNN - CKNN	Number of neighbors = 3
	Cosine KNN CosKNN	Number of neighbors = 3
	Cubic KNN - CubKNN	Number of neighbors = 3
	Weighted KNN - WKNN	Number of neighbors = 3
	Ensemble	Boosted Trees - BossT
Bagged Trees - BagT		Maximum number of splits = 150
Subspace Discriminant - SubD		Covariance structure: Full
Subspace KNN - SubKNN		Number of neighbors = 3
CNN	RUSBoosted Trees - RUSBT	Maximum number of splits = 150
		ImageInputLayer = 1
		Convolution2dLayer = 1
		ReluLayer = 1
		FullyConnectedLayer = 3
	SoftmaxLayer = 1	
	ClassificationLayer = 1	
	Training algorithm = adam	
	Max Epochs = 1000	

Table 4 Classical machine learning classification per plane

	CN vs AD		AD vs MCI		CN vs MCI		All vs All	
3 planes	93.3%	115 ft	84.6%	80 ft	85.3%	120 ft	69.7%	25 ft
Coronal	88.9%	100 ft	87.7%	80, 95,140 ft	88.2%	30,40,60,65,70,75,80,85,90 ft	75.3%	80,95,105,115 ft
Sagittal	93.3%	35 ft	78.5%	6,170 ft	82.4%	40 ft	65.2%	8,9,10,65,165,175 ft
Axial	77.8%	45 ft	84.6%	85 ft	82.4%	10 ft	69.7%	35,80ft
Best result	93.3%	35,115 ft	87.7%	80,95,140 ft	88.2%	30,40,60,65,70,75,80,85,90 ft	75.3%	80,95,105,115 ft

Table 5 Summary of the DL classification results

	CN vs AD		AD vs MCI		CN vs MCI		All vs All	
Best result	82.2%	115ft (3planes)	75.4%	95ft (Coronal)	83.8%	85,90 ft (Coronal)	64.0%	80,85,95 ft (Coronal)

plane and the Fine KNN. The lowest classification accuracy achieved through the cML algorithms was 78.5% using the sagittal plane. For this study group pair, as indicated in Table 5, the highest classification accuracy achieved through the CNN algorithm was 75.4% using the 95 features selected from the coronal plane.

Concerning the study group pair All vs All, as indicated in Table 4, the highest classification accuracy achieved through the cML algorithms was 75.3% using 80, 95, 105 and 115 features selected from the coronal plane and the subspace KNN classifier. The lowest classification accuracy achieved through the cML algorithms was 65.2% using the sagittal plane. It is observed from Table 5 that, for this study group pair, the highest classification accuracy achieved through the CNN algorithm was 64% using the 80, 85, and 95 features selected from the coronal plane. The lowest classification results were obtained for this study group pair, which indicates that the multi-class classification is the one in which the extracted features and the ML algorithms have more difficulty in discriminating between the groups.

Analyzing the results, it is observed that the CNN algorithm did not obtain classification accuracies higher than the cML algorithms in any of the four study group pairs. In fact, except for the pair CN vs MCI, the best result achieved using the CNN algorithm is worse than the worst result achieved using the cML algorithms. The overall poor performance of the CNN algorithm may be due to a non-optimal selection of the features to be applied on its inputs since the features were selected by applying the f-score algorithm combined with the cML algorithms and not with the CNN.

The only result above 90% was obtained in the pair CN vs MCI. This high classification accuracy is particularly important because, due to the lack of a cure for Alzheimer's disease, early detection plays a key role in medical intervention to reduce brain damage, preserve daily functioning for longer, and give the patient time to plan the future. Despite not having obtained the highest accuracy, the pair CN vs AD was the only one for which classification results above 80% were achieved in all study planes. This overall high performance was expected because CN and AD are the groups that have the greatest anatomical differences in the brain [30].

Among the study planes, the coronal plane was the one in which the best overall classification accuracies were obtained. This result is sustained by previous studies [31] and [32] and can be justified by the fact that the coronal plane enables a clearer view of 3 of the most important tissues for AD, namely, the cerebral cortex, ventricle, and hippocampus. Consequently, it is possible to indicate that the coronal plane allows the best visualization of the differences in the various anatomical regions of the 3 groups studied.

It is worth noting that the results presented and discussed above were obtained by using all study participants on wavelet and feature selection. Although easily found in literature,

this is not the most rigorous way to select features because it may introduce a risk factor of overfitting. The selection was performed in this way due to the small size of the database, but this risk was reduced by the cross-validation employed on the evaluation performance.

A comparison between the classification results obtained in the present work and those found in the literature also using the ADNI image database is depicted in Table 6. It is observed that not all state-of-art methods performed the three binary classifications made in the present work, focusing on the pair CN vs AD. And, more importantly, only three of the state-of-art methods carried out the multi-class classification All vs All.

For the pair CN vs MCI, crucial for early detection, the sMRI-based method proposed in the present work outperformed the methods developed in Ruiz et al. [8], Lebedev et al. [33], and Zhang et al. [34], Thappa et al. [9], and Liu et al. [35], by 21, 19, 16, 14, and 14%, respectively. However, the 88% accuracy achieved in the present work is 1% lower than that obtained in Lee et al. [36].

In the AD vs MCI case, the 88% achieved in the present work is 19, 18, 16, 12% higher than that obtained in Ruiz et al. [8], Lebedev et al. [33], Lee et al. [36], and Zhang et al. [34], respectively, but 3% lower than that obtained in Thappa et al. [9], where all are sMRI-based methods.

For the pair CN vs AD, compared with only sMRI-based methods, the 93% achieved in the present work is 14, 10, 7 and 7% higher than that obtained in Lebedev et al. [33], Qiu et al. [14], Zhang et al. [34] and Ruiz et al. [8], respectively, but 6% lower than that obtained in Thappa et al. [9]. Regarding the multi-class classification All vs All, the proposed method stands out for achieving the highest accuracy, outperforming the methods developed in Lebedev et al. [33], Zhang et al. [34] and Lee et al. [36] by 34, 23 and 4%, respectively.

Compared with diagnosing methods through images techniques other than sMRI, the proposed method outperformed the methods developed in Liu et al. [35] and Cheng et al. [37] by 2 and 1%, respectively, but it is surpassed by 4% by the fMRI-based method's developed in Amini et al. [11]. Although the above performance comparisons are evidence of the proposed method's ability to discriminate the different stages of AD, they should be carefully analyzed since different works may use different amounts of subjects, or the same amount but different subjects, even if the database is the same.

In addition to the ADNI database, the sMRI-based method developed in Qiu et al. [14] was also originally evaluated using other image databases and these results are summarized in Table 7. It is observed that, for the pair CN vs AD, the classification accuracy achieved by applying the proposed method to the ADNI database also outscores those obtained by applying the method developed in [14] to the

Table 6 Comparison with previous works with ADNI database

Study	Database	Exam Type	Features extracted	Best Classifier	Feature Selection	Accuracy
[8]	ADNI	sMRI	Voxels of ROIs	VAF-FS, Random Forest, XGBoost	ROIs	CN vs AD - 86% CN vs MCI - 72% MCI vs AD - 69%
[9]	ADNI	sMRI	MMSE, hippocampus left and right volume	SVM	Filter-based	CN vs AD - 99% CN vs MCI - 79% MCI vs AD - 91%
[14]	ADNI	sMRI	Disease probability maps	CNN	Not applied	CN vs AD - 83%
[11]	ADNI	fMRI	Average, standard deviation of cortical thickness, volume of cortical parcelling, white matter, and surface area	CNN	PCA	CN vs AD - 97%
[16]	ADNI	sMRI	Spatial features	CNNs ensemble model	ROIs	CN vs AD - 90% CN vs MCI 72% MCI vs AD 77%
[37]	ADNI	PET	Voxel-wise intensity	CNN	Not applied	CN vs AD - 92%
[17]	ADNI	sMRI	MRI Slices	CNN	Not applied	CN vs AD - 92%
[35]	ADNI	FDG-PET	Latent and generic features	CNN - BGRU	Not applied	CN vs AD - 91% CN vs MCI - 79%
[34]	ADNI	sMRI	ROI Volume of gray matter tissue	SVM	t-test	CN vs AD - 86% CN vs MCI - 72% MCI vs AD - 76% All vs All - 52%
[33]	ADNI	sMRI	Cortical thickness, Sulcal depth, Jacobian, tissue volumes	Random Forest ensembles	Recursive feature elimination - Gini-criterion	CN vs AD - 79% CN vs MCI - 69% MCI vs AD - 70% All vs All - 41%
[36]	ADNI	sMRI	Complex nonlinear relationships among voxels	CNNs ensemble model	Not applied	CN vs AD - 93% CN vs MCI - 89% MCI vs AD - 82% All vs All - 71%
Present work	ADNI	sMRI	Statistical and textural features	BagT FGSVM QSVM SubKNN	F-score	CN vs AD - 93% CN vs MCI - 88% MCI vs AD - 88% All vs All - 75%

Table 7 Comparison with previous imaging works with different databases

Study	Database	Exam Type	Features extracted	Best Classifier	Feature Selection	Accuracy
[14]	AIBL	sMRI	Disease probability maps	CNN	Not applied	CN vs AD - 87%
[14]	FHS	sMRI	Disease probability maps	CNN	Not applied	CN vs AD - 77%
[14]	NACC	sMRI	Disease probability maps	CNN	Not applied	CN vs AD - 82%
Present work	ADNI	sMRI	Statistical and textural features	BagT FGSVM QSVM SubKNN	F-score	CN vs AD - 93% CN vs MCI - 88% MCI vs AD - 88% All vs All - 75%

Table 8 Comparison with non-imaging works

Study	Database	Exam Type	Features extracted	Best Classifier	Feature Selection	Accuracy
[39]	University Hospital Center of São João	EEG	Cepstral and Lacstral Distances	ANN	KW test	CN vs ADM - 96% CN vs MCI - 98% MCI vs ADM-ADA - 94% All vs All - 96%
[38]	University Hospital Center of São João	EEG	Relative Power, Spectral Ratios, Maxima, Minima and Zero Crossing	ANN	KW test	CN vs AD - 95% CN vs MCI - 77% MCI vs AD - 83% All vs All - 90%
[41]	IRCCS	EEG	Wavelet Coefficients	Decision Tree	Not applied	CN vs AD - 83% CN vs MCI - 92% MCI vs AD - 79%
[42]	Centre for Biomedical Technology	MEG	Measures of synchronization	Randomized 2D CNN	Not applied	CN vs MCI - 91%
[40]	University Institute of Psychiatry	Biomarkers	$A\beta_{1-42}$ /P-tau concentrations	ML	Not applied	CN vs AD - 83% MCI vs AD - 78%
Present work	ADNI	sMRI	Statistical and textural features	BagT FGSVM QSVM SubKNN	F-score	CN vs AD - 93% CN vs MCI - 88% MCI vs AD - 88% All vs All - 75%

AIBL, FHS, and NACC databases. Besides the different features computed from the images, a factor that may be contributing to the better general performance of the proposed method is the feature selection, a procedure not performed in [14]. Although enriching, these comparisons need to be carefully analyzed because different image databases were employed in the studies.

A comparison between the classification results obtained in the present work and those found in the literature using signal and biomarkers techniques is summarized in Table 8.

It is observed that the proposed sMRI-based method did not present the best performance in any of the analyzed study group pairs. For the pair CN vs MCI, it outperformed the method developed in [38] by 11% but it is surpassed by the method introduced in [39] by 10%.

In the MCI vs AD case, the proposed method outscored the methods developed in [40, 41], and [38] by 10, 9, and 5%, respectively, but is outperformed by the method elaborated in [39] by 6%. For CN vs AD, the proposed method outperformed both the methods developed in [40] and [41] by 10%, but is outscored by the method produced in [38] by 2%. In the multi-class All vs All case, the proposed method did not outperform the EEG-based methods developed in [38] and [39], being surpassed by 21% by the latter.

5 Conclusion

Alzheimer's disease is one of the neurodegenerative diseases with the highest prevalence, affecting millions of people worldwide. This work aimed to detect AD on the stages CN, MCI, and AD itself using sMRI. A set of co-occurrence matrix and texture statistical measures (contrast, correlation, energy, homogeneity, entropy, variance, and standard deviation) were extracted from a two-level DWT decomposition of sMRI images. The discriminant capacity of the measures was analyzed and the most discriminant ones were selected to be used as features for feeding classical machine learning algorithms and a CNN. The classical algorithms achieved the following classification accuracies: 93.3% for AD vs CN, 87.7% for AD vs MCI, 88.2% for CN vs MCI, and 75.3% for All vs All. The CNN achieved the following classification accuracies: 82.2% for AD vs CN, 75.4% for AD vs MCI, 83.8% for CN vs MCI, and 64% for All vs All. For the All vs All comparison, the proposed method outperformed by 4% the highest classification accuracy of the state-of-art sMRI-based methods.

The accuracies achieved for AD vs CN, AD vs MCI, and CN vs MCI indicate that the evaluated measures have a great ability to distinguish within these binary groups. However, despite surpassing the state-of-the-art, additional research should be conducted to improve the accuracy of the challenging multi-class classification All vs All. Despite the promising results, the database size was a limitation for the

present study because all study participants needed to be used for wavelet and feature selection tasks. In future, the work should be updated with a larger sMRI database that can be divided into training and testing subsets.

Acknowledgements This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, through the project UID/50016/2020. Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author Contributions JS and PMR contributed to conceptualization; JS contributed to methodology; BCB and PMR contributed to validation; JS contributed to writing—original; BCB and PMR contributed to writing—review and editing; and PMR contributed to supervision. All authors read and approved the final manuscript.

Funding Open access funding provided by FCTIFCCN (b-on). This work received no funding support.

Data Availability Statement Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Declarations

Conflicts of interest No potential conflict of interest was reported by the authors.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Figueira, M. (2014). *Manual de psiquiatria clínica*. Lisboa: Lidel, Edições Técnicas.
2. Alzheimer's Disease Association. (2021). 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 17(3), 327–406. <https://doi.org/10.1002/alz.12328>
3. Galea, M., & Woodward, M. (2005). Mini-mental state examination (MMSE). *Australian Journal of Physiotherapy*, 51(3), 198. [https://doi.org/10.1016/s0004-9514\(05\)70034-9](https://doi.org/10.1016/s0004-9514(05)70034-9)
4. van Oostveen, W. M., & de Lange, E. C. M. (2021). Imaging techniques in Alzheimer's disease: A review of applications in early diagnosis and longitudinal monitoring. *International Journal of Molecular Sciences*, 22(4), 2110. <https://doi.org/10.3390/ijms22042110>
5. Worthoff, W. A., Yun, S. D., & Shah, N. J. (2018). Introduction to Magnetic Resonance Imaging. In: Hybrid MR-PET Imaging (pp. 1–44). Royal Society of Chemistry.
6. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2), 67–77. <https://doi.org/10.1038/nrneurol.2009.215>
7. Chandra, A., Dervenoulas, G., & Politis, M. (2018). Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. *Journal of Neurology*, 266(6), 1293–1302. <https://doi.org/10.1007/s00415-018-9016-3>
8. Ruiz, E., Ramírez, J., Górriz, J. M., & Alzheimer's, J. C. (2018). Disease computer-aided diagnosis: Histogram-based analysis of regional MRI volumes for feature selection and classification. *Journal of Alzheimer's Disease*, 65(3), 819–842. <https://doi.org/10.3233/jad-170514>
9. Thapa, S., Singh, P., Jain, D. K., Bharill, N., Gupta, A., & Prasad, M. (2020). Data-driven approach based on feature selection technique for early diagnosis of Alzheimer's disease. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE
10. Hon, M., & Khan, N. M. (2017) Towards Alzheimer's disease classification through transfer learning. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1166–1169).
11. Amini, M., Pedram, M. M., Moradi, A., & Ouchani, M. (2021). Diagnosis of Alzheimer's disease severity with fmri images using robust multitask feature extraction method and convolutional neural network (CNN). *Computational and Mathematical Methods in Medicine*, 2021, 1–15. <https://doi.org/10.1155/2021/5514839>
12. Al-Khuzai, F. E. K., Bayat, O., & Duru, A. D. (2021). Diagnosis of Alzheimer disease using 2D MRI slices by convolutional neural network. *Applied Bionics and Biomechanics*, 2021, 1–9. <https://doi.org/10.1155/2021/6690539>
13. Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Shen, L., et al. (2020). A multi-model deep convolutional neural network for automatic

- hippocampus segmentation and classification in Alzheimer's disease. *NeuroImage*, 208, 116459. <https://doi.org/10.1016/j.neuroimage.2019.116459>
14. Qiu, S., Joshi, P. S., Miller, M. I., Xue, C., Zhou, X., Karjadi, C., et al. (2020). Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*, 143(6), 1920–1933. <https://doi.org/10.1093/brain/awaa137>
 15. Vaithinathan, K., & Parthiban, L. (2019). A novel texture extraction technique with T1 weighted MRI for the classification of Alzheimer's disease. *Journal of Neuroscience Methods*, 318, 84–99. <https://doi.org/10.1016/j.jneumeth.2019.01.011>
 16. Kang, W., Lin, L., Zhang, B., Shen, X., & Wu, S. (2021). Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis. *Computers in Biology and Medicine*, 136, 104678. <https://doi.org/10.1016/j.compbiomed.2021.104678>
 17. Ebrahimi, A., Luo, S., & Chiong, R. (2021). Deep sequence modelling for Alzheimer's disease detection using MRI. *Computers in Biology and Medicine*, 134, 104537. <https://doi.org/10.1016/j.compbiomed.2021.104537>
 18. Isa, I. S., Sulaiman, S. N., Mustapha, M., & Darus, S. (2015). Evaluating denoising performances of fundamental filters for T2-weighted MRI images. *Procedia Computer Science*, 60, 760–768. <https://doi.org/10.1016/j.procs.2015.08.231>
 19. Tan, L., & Jiang, J. (2019). Image processing basics. *Digital signal processing* (pp. 649–726). Amsterdam: Elsevier.
 20. Zhang, D. (2019). Wavelet transform. *Texts in computer science* (pp. 35–44). New York: Springer.
 21. Chaplot, S., Patnaik, L. M., & Jagannathan, N. R. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control*, 1(1), 86–92. <https://doi.org/10.1016/j.bspc.2006.05.002>
 22. Gonzalez, R., & Woods, R. (2018). *Digital image processing*. New York, NY: Pearson.
 23. Nayak, D. R., Dash, R., & Majhi, B. (2016). Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing*, 177(C), 188–197. <https://doi.org/10.1016/j.neucom.2015.11.034>
 24. Kumar, PSS., & Dharun, VS. (2016) Extraction of texture features using GLCM and shape features using connected regions. *International Journal of Engineering and Technology*, 8(6), 2926–2930. <https://doi.org/10.21817/ijet/2016/v8i6/160806254>
 25. Priddy, K. L., & Keller, P. E. (2005). *Artificial neural networks: An introduction*. SPIE.
 26. Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>
 27. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
 28. Sevani, N., Hermawan, I., & Jatmiko, W. (2019). Feature selection based on F-score for enhancing CTG data classification. In *2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)* (pp. 18–22). IEEE. <https://doi.org/10.1109>
 29. Haykin, S. O. (2008). *Neural networks and learning machines* (3rd ed.). Pearson.
 30. Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., et al. (2019). Study of brain morphology change in Alzheimer's disease and amnesic mild cognitive impairment compared with normal controls. *General Psychiatry*, 32(2), e100005. <https://doi.org/10.1136/gpsych-2018-100005>
 31. Raza, M., Awais, M., Ellahi, W., Aslam, N., Nguyen, H. X., & Le-Minh, H. (2019). Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques. *Expert Systems with Applications*, 136, 353–364. <https://doi.org/10.1016/j.eswa.2019.06.038>
 32. Ren, F., Yang, C., Qiu, Q., Zeng, N., Cai, C., Hou, C., et al. (2019). Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification. *IEEE Access*, 7, 181423–181433. <https://doi.org/10.1109/access.2019.2920241>
 33. Lebedev, A. V., Westman, E., Westen, G. J. P. V., Kramberger, M. G., Lundervold, A., Aarsland, D., et al. (2014). Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, 6, 115–125. <https://doi.org/10.1016/j.nicl.2014.08.023>
 34. Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 55(3), 856–867. <https://doi.org/10.1016/j.neuroimage.2011.01.008>
 35. Liu, M., Cheng, D., & Yan, W. (2018). Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Frontiers in Neuroinformatics*. <https://doi.org/10.3389/fninf.2018.00035>
 36. Lee, E., Choi, J. S., Kim, M., & Suk, H. I. (2019). Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning. *NeuroImage*, 202, 116113. <https://doi.org/10.1016/j.neuroimage.2019.116113>
 37. Cheng, D., & Liu, M. (2017). Classification of Alzheimer's disease by cascaded convolutional neural networks using PET images. *Machine learning in medical imaging* (pp. 106–113). Springer.
 38. Rodrigues, P. M., Freitas, D. R., Teixeira, J. P., Alves, D., & Garrett, C. (2018). Electroencephalogram signal analysis in Alzheimer's disease early detection. *International Journal of Reliable and Quality E-Healthcare*, 7(1), 40–59. <https://doi.org/10.4018/ijrqeh.2018010104>
 39. Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P., & Freitas, D. (2021). Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, 25(9), 3384–3395. <https://doi.org/10.1109/jbhi.2021.3069789>
 40. Forlenza, O. V., Radanovic, M., Talib, L. L., Aprahamian, I., Diniz, B. S., Zetterberg, H., et al. (2015). Cerebrospinal fluid biomarkers in Alzheimer's disease: Diagnostic accuracy and prediction of dementia. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(4), 455–463. <https://doi.org/10.1016/j.dadm.2015.09.003>
 41. Fisco, G., Weitschek, E., Cialini, A., Felici, G., Bertolazzi, P., Salvo, S. D., et al. (2018). Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC Medical Informatics and Decision Making*. <https://doi.org/10.1186/s12911-018-0613-y>
 42. Lopez-Martin, M., Nevado, A., & Carro, B. (2020). Detection of early stages of Alzheimer's disease based on MEG activity with a randomized convolutional neural network. *Artificial Intelligence in Medicine*, 107, 101924. <https://doi.org/10.1016/j.artmed.2020.101924>