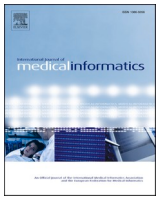
International Journal of Medical Informatics 173 (2023) 105027

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



An unsupervised learning approach to diagnosing Alzheimer's disease using brain magnetic resonance imaging scans

Yuyang Liu *, Suvodeep Mazumdar, Peter A. Bath, for the Alzheimer's Disease Neuroimaging Initiative[1]

Information School, University of Sheffield, 211 Portobello, Sheffield S1 4DP, UK

ARTICLE INFO

Keywords:
Alzheimer's disease
MRI
Machine learning
Deep learning
Unsupervised learning

ABSTRACT

Background: Alzheimer's disease (AD) is the most common cause of dementia, characterised by behavioural and cognitive impairment. Due to the lack of effectiveness of manual diagnosis by doctors, machine learning is now being applied to diagnose AD in many recent studies. Most research developing machine learning algorithms to diagnose AD use supervised learning to classify magnetic resonance imaging (MRI) scans. However, supervised learning requires a considerable volume of labelled data and MRI scans are difficult to label.

Objective: This study applied a statistical method and unsupervised learning methods to discriminate between scans from cognitively normal (CN) and people with AD using a limited number of labelled structural MRI scans.

Methods: We used two-sample t-tests to detect the AD-relevant regions, and then employed an unsupervised learning neural network to extract features from the regions. Finally, a clustering algorithm was implemented to discriminate between CN and AD data based on the extracted features. The approach was tested on baseline brain structural MRI scans from 429 individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI), of which 231 were CN and 198 had AD.

Results: The abnormal regions around the lower parts of limbic system were indicated as AD-relevant regions based on the two-sample *t*-test ($p < 0.001$), and the proposed method yielded an accuracy of 0.84 for discriminating between CN and AD.

Conclusion: The study combined statistical and unsupervised learning methods to identify scans of people with AD. This method can detect AD-relevant regions and could be used to accurately diagnose AD; it does not require large amounts of labelled MRI scans. Our research could help in the automatic diagnosis of AD and provide a basis for diagnosing stable mild cognitive impairment (stable MCI) and progressive mild cognitive impairment (progressive MCI).

1. Introduction

Dementia is receiving increasing attention because there are now over 50 million people all living with dementia globally [1]. Dementia is a clinical condition leading to cognitive and behavioural impairment, i.e., language or memory loss, typically among people aged 65 years old and over [2]. Alzheimer's disease (AD) is the most common and prevalent cause of dementia, accounting for an approximate 60% to 80% of dementia cases [3].

AD is commonly diagnosed using structural MRI scans [4], which generate a 3D image and capture changes in the structure of the brain through using magnetic fields and radio waves [5]. However, analysing MRI scans is difficult because traditional medical imaging scan analysis for AD diagnosing is undertaken manually by doctors. The structure of the brain is complex, and each scan contains millions of voxels (units in a 3D medical imaging scan) and massive amounts of information, so analysing it manually by doctors is a time-consuming task. Machine learning (and deep learning) techniques have been used for diagnosing various disease in previous research, e.g., Coronavirus disease 2019 (COVID-19) [6], cancer [7], AD [8], etc. As a result, computer-aided

* Corresponding author at: Information School, University of Sheffield, 211 Portobello, Sheffield S1 4DP, UK.
E-mail address: yliu369@sheffield.ac.uk (Y. Liu).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 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.

https://doi.org/10.1016/j.ijmedinf.2023.105027
Received 20 October 2022; Received in revised form 22 February 2023; Accepted 23 February 2023
Available online 2 March 2023

diagnosis of AD based on using machine learning (including deep learning) to analyse medical imaging scans has emerged as an area of interest. Zhang et al. [9] employed the improved framework TResNet of residual network with regional attention mechanism to diagnose AD. They fed data from structural MRI scans of grey matter into the network and achieved an accuracy of 90%, sensitivity of 92.8%, and specificity of 87.5% in the classification of AD and CN. Zeng et al. [8] applied a novel deep belief network (DBN) based multi-task learning algorithm to MRI scans of grey matter to discriminate between CN and AD, and achieved an accuracy of 98.62%. Nigri et al. [10] proposed explainable CNN models to better interpret the results in addition to achieving high performance with an AUC (Area Under the Curve) of 0.923. Baskar et al. [11] first extracted hippocampus and posterior cingulate cortex from the MRI scans and then used fuzzy c-means to eliminate suspicious, i.e., irrelevant, samples, before employing a 4-layer backpropagation artificial neural network (BANN) to classify the data into CN, mild cognitive impairment (MCI), and AD. The performance of classification achieved an accuracy of 98.63%, sensitivity of 85.63%, and specificity of 79.56%.

These studies used supervised learning models, which requires a large number of MRI scans that have been labelled beforehand (often manually) for training. Although crowdsourcing can solve the problem of labelling data, labelling medical imaging data is still difficult as it requires qualified practitioners with extensive experience and professional knowledge; however, recruiting these professionals is difficult and resource intensive. Furthermore, learning discriminative features from labelled data also means that the algorithm is guided by people on what to do [12], which means that it is likely to ignore some hidden patterns. In contrast, unsupervised learning methods can capture hidden patterns within the data with the help of feature learning [13]. Therefore, methods using limited amounts of, or no, labelled data (i.e., semi-supervised and unsupervised learning methods) have become important.

In recent years, some studies have used unsupervised learning to undertake diagnosis of AD. Farouk & Rady [14] utilised k-means clustering with a combination of ROI features and the whole brain features to cluster data, achieving an accuracy of 76.3%. Bi et al. [15] applied a PCANet on three slices from three planes of view for each brain MRI scan to extract features and then used k-means to cluster the data, achieving an accuracy of 89.15% in the diagnosis of AD and CN. Escudero et al. [16] applied k-means clustering to data features in five clinical scenarios and one of the scenarios used MRI scans. In their method, 73.6% of the AD scans were clustered into the pathological cluster. However, compared with supervised learning methods, there have been far fewer studies using unsupervised methods. Further work is therefore required to understand the potential of unsupervised learning methods for diagnosing AD, as well as their limitations.

This paper contributes to this emerging area by proposing a novel approach that combines statistical methods, an unsupervised neural network, and a clustering algorithm to support clinicians in diagnosing AD. The statistical method locates the AD-relevant regions using a limited number of labelled MRI scans; the unsupervised neural network converts the AD-relevant regions (voxel level), which are low-level features, to higher-level features. Higher-level features refer to a more complex and a more abstract representation of the input raw data that is learned by neural networks. However, neural networks are effectively black boxes and higher-level features computed by neural networks are often difficult to perceive and interpret for humans: this is one of the disadvantages of using these methods. Although higher-level features can indeed be helpful for discriminating between different classes, defining these higher-level features, and what they represent, is not always possible. In this paper we seek to represent these higher-level features, not by defining them precisely, but by plotting them in relation to each other according to their class or group (as shown later in Fig. 3 (c) and (d)). Finally, the clustering algorithm divides the data into two groups: AD and cognitively normal (CN). This process is important in that it involves only limited amounts of labelled data and does not require clinical input. Our proposed method therefore makes an original contribution to this field and it performed well in the clustering. Additionally, it could potentially be used in future work to predict the development of AD.

2. Materials and methodology

Here we describe the data used in our study and our methodology.

2.1. Dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test 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 (MCI) and early Alzheimer's disease (AD). We used T1-weighted MRI scans from 429 participants in ADNI, specifically, a baseline dataset consisting of 231 CN and 198 AD scans (details of these participants are available from https://adni.loni.usc.edu/wp-content/uploads/2012/08/ADNI_Enroll_Demographics.pdf).

2.2. Data pre-processing

The MRI scans were pre-processed using anterior commissure-posterior commissure alignment (AC-PC alignment), motion correction, intensity normalisation, template registration, and skull stripping. The AC-PC alignment was undertaken manually using Statistical Parametric Mapping (SPM) software and the other pre-processing steps were followed using SPM and Computational Anatomy Toolbox (CAT) automatically. The pre-processing workflow finally resampled all the scans to $121 \times 145 \times 121$ voxels, with voxel sizes of 1.5 mm across all planes (sagittal, coronal, and axial).

2.3. Abnormality detection between two groups

This phase aims to detect regions that are visibly different in AD and CN groups at the voxel level, so- called the regions of interest (ROIs). The commonly-used two-sample *t*-test was employed to test for ROIs between the two groups within AD-relevant regions in MRI scans [171819]. The two-sample *t*-test was applied to 60 labelled scans, of which 30 were AD and 30 were CN, to obtain ROIs between the two groups. The choice of the numbers of scans was based on the *t*-test generally being used for small amounts of data (e.g., often a sample size of less than 30) [20]. In addition, based on the experiments we conducted previously, when the number of MRI scans of each group was increased, the distinguishing ROIs that were detected using the two-sample *t*-test were almost the same. Therefore, it was determined that 30 labelled scans for a two-sample *t*-test respectively from each group were sufficient to detect AD-relevant regions, and this reduced the need for labelled MRI scans.

2.4. Feature extraction

Although, following the ROIs detection stage, in which the regions indicating the existence of AD have been recognised, it is still difficult to accurately cluster data using only the intensity values of the voxels within the regions directly, because the voxel intensity values are still low-level features. Hence, to obtain higher level features from unlabelled data, an unsupervised neural network, PCANet, a combination of Principal Component Analysis (PCA) method and Convolutional Neural Network (CNN) [21], was employed, as shown in Fig. 1. This has three stages: the first two stages both learn convolution filters by PCA for

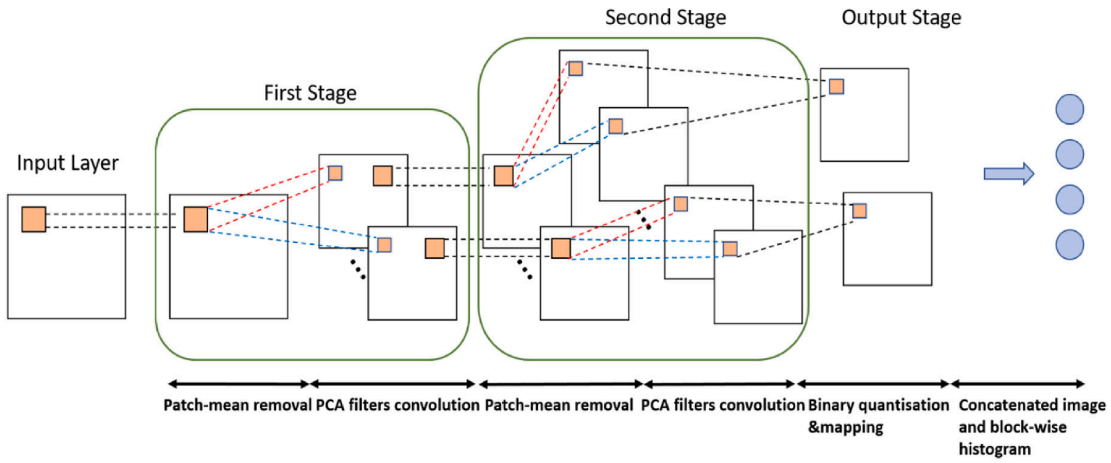


Fig. 1. The structure of PCANet.

feature mapping, and the final stage binarises and generates a block-wise histogram for feature output.

2.4.1. First stage of a PCANet

Given N input images $\{I_i\}_{i=1}^{N}$ of size $m \times n$ and assuming the size of the patch is $k_1 \times k_2$ at all stages. All overlapping patches within the i_{th} image were collected and there are $m \times n$ vectorised patches in each image. We then subtracted the patch mean from each patch and obtained $\overline{X}_i = [\overline{x}_{i,1}, \overline{x}_{i,2}, \cdots, \overline{x}_{i,mn}]$, where $\overline{x}_{i,j}$ is a mean-removed patch. Therefore, all images were operated on in the same way and we obtained $\boldsymbol{X} = [\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_N] \in \mathbb{R}^{k_1 k_2 \times Nmn}$.

Assuming that the number of filters in layer is L_1, a PCA was used to learn the filter bank in this stage. The solution is known as the L_1 principal eigenvectors of $\boldsymbol{XX^T}$. Therefore, the PCA filters are expressed as:

$$W_l^1 = mat_{k_1,k_2}\left(q_l\left(\boldsymbol{XX^T}\right)\right) \in \mathbb{R}^{k_1 k_2}, l = 1, 2, \cdots, L_1$$

where $mat_{k_1,k_2}(\boldsymbol{v})$ is a function that maps $\boldsymbol{v} \in \mathbb{R}^{k_1 k_2}$ to a matrix $\boldsymbol{W} \in \mathbb{R}^{k_1 \times k_2}$, and $q_l\left(\boldsymbol{XX^T}\right)$ denotes the l_{th} principal eigenvector of $\boldsymbol{XX^T}$. This captures the main variation across all of the mean-removed patches.

2.4.2. Second stage of a PCANet

The process of the second stage almost repeats the same process as the first stage. Let the l_{th} filter output of the first stage be:

$$I_i^l = I_i * \boldsymbol{W}_l^1, i = 1, 2, \cdots, N$$

where * denotes 2D convolution and the boundary of I_i is zero-padded before convolving with $\boldsymbol{W}_l^1$ in order to make the I_i^l having the same size of I_i. Similar to the first stage, all the overlapping patches of I_i^l were collected, and the patch mean was subtracted from each patch. This formed $\overline{Y}_i^l = [\overline{y}_{i,l,1}, \overline{y}_{i,l,2}, \cdots, \overline{y}_{i,l,mn}] \in \mathbb{R}^{k_1 k_2 \times mn}$, where $\overline{y}_{i,l,j}$ is the j_{th} mean-removed patch in I_i. We further defined $Y^l = [\overline{Y}_1^l, \overline{Y}_2^l, \cdots, \overline{Y}_N^l] \in \mathbb{R}^{k_1 k_2 \times Nmn}$ for the matrix that collects all mean-removed patches of the l_{th} filter output and concatenate Y^l for the outputs of all filters:

$$\boldsymbol{Y} = \left[Y^1, Y^2, \cdots, Y^{L_1}\right] \in \mathbb{R}^{k_1 k_2 \times L_1 Nmn}$$

The PCA filters of the second stage were obtained as:

$$\boldsymbol{W}_l^2 = mat_{k_1,k_2}\left(q_l\left(\boldsymbol{YY^T}\right)\right) \in \mathbb{R}^{k_1 k_2}, l = 1, 2, \cdots, L_2$$

where L_2 is the number of filters in the second stage. For each input I_i^l of the second stage, there were L_2 outputs, and each convolves $\boldsymbol{W}_l^2$ for $l = 1, 2, \cdots, L_2$:

$$\mathscr{O}_i^l = \{I_i^l * \boldsymbol{W}_l^2\}_{l=1}^{L_2}$$

The number of outputs in the second stage was $L_1 L_2$.

2.4.3. Output stage of a PCANet

In this stage, a process of hashing and histogram generating was implemented. In the process of hashing, L_2 outputs in $\mathscr{O}_i^l$ were converted into a single integer-valued "image":

$$\mathscr{T}_i^l = \sum_{l=1}^{L_2} 2^{l-1} H(\mathscr{O}_i^l)$$

whose every pixel is an integer in the range [0, $2^{l-1}-1$]. $H(\bullet)$ is a Heaviside step function whose value is one for positive and zero otherwise.

Then, for each of the L_1 images $\mathscr{T}_i^l$, $l = 1, 2, \cdots, L_1$, we partitioned it into B blocks. We computed the histogram (with 2^{L_2} bins) of the decimal values in each block. After concatenating all the B histograms into one vector and denote as $Bhist(\mathscr{T}_i^l)$, the feature vector of the input image I_i was then defined to be the set of block-wise histograms:

$$f_i = \left[Bhist\left(\mathscr{T}_i^1\right), Bhist\left(\mathscr{T}_i^2\right), \cdots, Bhist\left(\mathscr{T}_i^{L_1}\right)\right]^T \in \mathbb{R}^{(2^{L_2})L_1 B}$$

In addition, the local blocks could be either overlapping or non-overlapping.

2.4.4. PCANet in this study

In this study, the number of stages was set to two because a two-stage PCANet is sufficient to achieve good performance [21]. The size of the patches $k_1 \times k_2$ was set to 3×3 in both two stages; the number of filters L_1 and L_2 were both set to eight, respectively, which indicated that there were eight principal eigenvectors in each PCA process of the two stages; the block size for the histograms was set to 15×15; the overlap ratio of the local blocks was set to zero in this study for reducing the number of features. After feeding into the PCNet, the original normalised voxel intensity values (zero to one) were mapping to higher-level features.

2.5. Clustering

Data were divided into two groups: AD and CN, which means that the clustering tasks needed to split the data into two clusters. A refined version of k-means cluster analysis, the k-means++ algorithm [22], was used to cluster data from CN and AD patients using the features extracted by PCANet: this was performed for the whole sample of 429 scans, i.e., 231 scans from CN patients and 198 from AD patients. This

method partitions a collection of data into k disjointed clusters without any labels of the data. The detailed k-means++ algorithm is as follows:

(1) Select the first centroid at random;
(2) Calculate the distance from each data point x_i to the nearest, previously chosen centroid, and we note it as $dist(x_i)$;
(3) Choose the new data point from the n data points as the next centroid with the help of the maximum probability: $\frac{dist(x_i)^2}{\sum_{i=1}^{n} dist(x_i)^2}$;
(4) Repeat steps (2) – (3) until k centroids are selected;
(5) Form k clusters by assigning each point to its closest centroid;
(6) Re-compute the centroid of each cluster;
(7) Repeat until centroids do not change [2223].

The k value was set to 2 in this study, because the data had to be divided into two groups: CN and AD. The Euclidean distance, $d(x, y)$, was selected in step (3) of the algorithm, and was calculated by:

$$d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

where n is the number of dimensions and x_k and y_k are the k^{th} features of data points x and y, respectively. Therefore, the distance between a certain data point x_i and a centroid c in the step (2) above, $dist(x_i)$, could be calculated by:

$$dist(x_i) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - c_k)^2},$$

where the n is the number of dimensions and $x_{i,k}$ and c_k are the k_{th} features of x_i and c, respectively.

In the k-means++ algorithm, the first centroid was selected at random, such that the algorithm may have different results for different first centroids: we therefore repeated the k-means++ algorithm ten times, and the mean value was used. Clustering algorithms do not need to train, thus cross-validation was not necessary in this study.

3. Results and discussion

3.1. ROIs detection

Using 30 scans from CN patients and 30 scans from AD patients, the two-sample t-test analysis detected the regions pathologically related to AD (increasing the sample size for the t-test did not detect any further regions). Fig. 2 shows a heat map of the ROIs in the sagittal, coronal, and axial view planes of the MRI scan. Red and orange regions indicate that the mean of the intensity value (after being normalised) of the voxel in AD group is greater than that in the CN group and the blue indicates where the mean value of the voxel in AD group is less than CN group.

As shown in Fig. 2, the most conspicuous regions were the lower parts of the limbic system, which includes the parahippocampal gyrus, the amygdala, and the hippocampus, in the left and right cerebra. There were also fragmentary regions involving small parts of the thalamus and frontal lobe (including some tiny parts of the frontal gyrus).

3.2. Clustering results

Since there are two main AD-relevant regions located in the left and right cerebra, we utilised one of them (one ROI) and both of them (two ROIs) separately to discriminate between AD and CN, respectively for the 429 scans. The clustering performance was compared using five metrics: accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), on the clustering task using one ROI and two ROIs (Table 1).

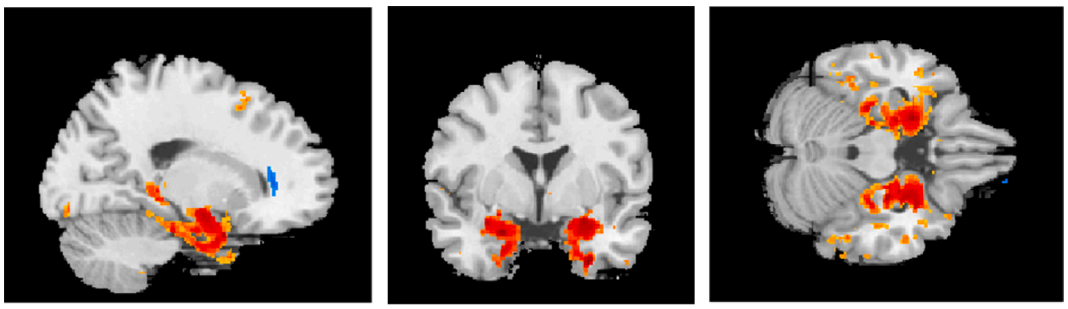
Table 1 shows that, for all five measures, using two ROIs performed better than using one ROI: the accuracy for two ROIs was 6% higher than using one ROI and the sensitivity was 10% higher, indicating that it was better at identifying people with AD; the respective 95% CI indicate that these differences are statistically significant. The ROIs also indicate that there are two diseased areas within AD. The wider CI of sensitivity when using one ROI (64%–75%) indicates a more variable performance. The specificities for one (84%) and two (88%) ROIs indicates that they both performed well in identifying non-AD cases, although the specificity for one ROI is significantly lower than that for two ROIs.

The significantly higher PPV when using two ROIs (86%; 95% CI = 83–88%) versus one ROI (80%; 95% CI = 76–83%) indicates that we can trust that a person identified as having AD actually has the disease, i.e., its diagnosing AD is fairly credible. The significantly higher NPV for two ROIs, i.e., 7% greater than using one ROI, gives greater confidence that someone who receives a negative test result does not actually have the

Table 1
Measures of Clustering using one ROI and two ROIs using PCANet.

Measures		K-means++	
		One ROI	Two ROIs
Accuracy	Mean	0.773	0.842
	95% CI*	(0.764, 0.782)	(0.833, 0.851)
Sensitivity	Mean	0.695	0.797
	95% CI	(0.640, 0.750)	(0.768, 0.826)
Specificity	Mean	0.840	0.881
	95% CI	(0.792, 0.888)	(0.853, 0.909)
PPV	Mean	0.798	0.855
	95% CI	(0.764, 0.832)	(0.829, 0.881)
NPV	Mean	0.767	0.836
	95% CI	(0.747, 0.787)	(0.819, 0.853)
Error Rate		0.227	0.158

* Confidence Interval.



(a) sagittal plane (b) coronal plane (c) axial plane

Fig. 2. A heat map of the regions related to AD noted in the three view planes of the MRI scan.

disease. Overall, using two ROIs was more reliable than using only one ROI, whether a person receives either a "CN" or an "AD" result.

To further demonstrate the effectiveness of the combination of the ROIs selected using the two-sample *t*-test and the PCANet, we compared the distributions of the features in four situations:

(1) features directly extracted from random selected slices (without PCANet);
(2) features directly extracted from the whole slices involving the ROIs (without PCANet);
(3) features directly extracted from the ROIs (without PCANet);
(4) features extracted from ROIs calculated by PCANet in the proposed method.

From the distribution of the features from the randomly-selected slices of the MRI scans (Fig. 3 (a)), we can see that the data are distributed irregularly and the two classes are completely visually inseparable and are hard to split. When we examine the distribution of the features from the whole slices that involve ROIs (Fig. 3 (b)), we can see that, although the two classes are still mixed together, most AD points gather towards the lower right part of the graph and most CN points gather towards the upper left area of the graph. This means that the ROIs do indeed help the two classes of data to split.

From the distribution of the features directly from the ROIs without PCANet (Fig. 3 (c)), we can see that the data further tend to separate compared with Fig. 3 (a) and (b). This means that the ROIs selected by the two-sample *t*-test are truly AD-relevant although there is still room for improvement. When examining the distribution of features calculated by the PCANet using the ROIs (Fig. 3 (d)), it can be seen that, although the two groups are still adjacent to each other, fewer deviants are mixed in the wrong group, compared with the Fig. 3 (c), and there is improved separation of the groups. Therefore, the results in Fig. 3 (c) and (d) generally suggest that the ROIs detected by the two-sample *t*-test are truly related to AD and that the PCANet helps the ROIs convert to high-level features that can be clustered more accurately.

In order to understand better why some cases were mistakenly discriminated into AD at baseline, when they were actually CN, we looked at how these patients progressed subsequently, as summarised in Table 2. We can observe that 15 of the participants who were CN at baseline, subsequently developed to MCI from CN or developed to AD from CN (excluding those who developed to MCI and returned to CN later) before the 36th month, and 6 out of these 15 participants were clustered into the AD cluster. If we examine the participants who developed to AD before the 48th month, 8 out of 25 were clustered into the AD cluster; if we look at those who developed to AD before the 60th month, 8 out of 31 were clustered into AD; 37 participants progressed to

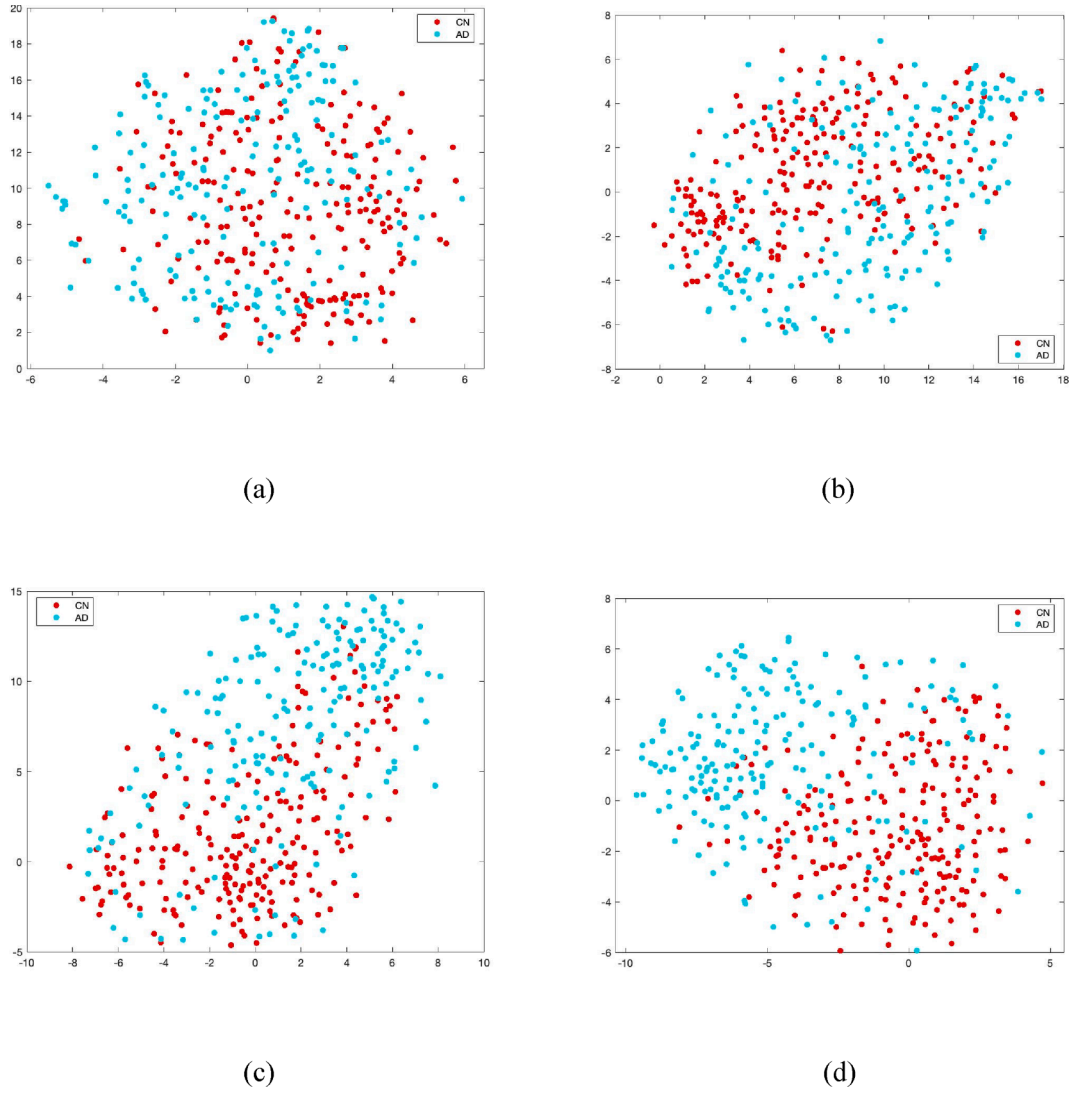


Fig. 3. Distributions of the features from four situations. (a) is the distribution of features from randomly selected slices; (b) is the distribution of features from the slices involving ROIs; (c) is the distribution of features directly from the ROIs; (d) is the distribution of the features computed by PCANet using the ROIs.

Table 2
A summary of the scans that are CN but were mis-clustered into AD. The first column shows the month when the participants developed to MCI from CN or developed to AD from CN; the second column indicates the number of patients who developed to MCI or AD from CN in corresponding months; the third column is the number of participants who are CN at baseline but were mistakenly clustered into AD group by our proposed method; the fourth column indicates the ratio of the third column to the second column.

MCI/AD detected in (months)	No. of patients	Mis-clustered into AD group	Mis-clustered into AD group (%)
36	15	6	40
48	25	8	32
60	31	8	25
72	37	9	24
108	52	12	23

MCI or AD before the 72nd month and 9 of them were clustered into AD; if we prolong the conversion time to 108th month, 12 out of 52 participants were clustered into AD.

From the observations above, our conjecture is that the proposed method could potentially help clinicians to predict the early development and progression of the participants from being CN to having MCI and AD, and the reason is that the unsupervised learning method is based on the features themselves, instead of feedback from the human assessors. Hence, even though some participants were diagnosed as CN at baseline, if their brains more or less show pathological changes related to AD, it is quite natural that they could possibly be clustered into AD based on the changes. This suggests that this method can potentially detect pre-clinical changes in the brain, i.e., before the patient is diagnosed with AD.

However, from Table 2, the performance of the prediction decreases with the time of conversion delaying. This is likely to be because the proposed method only used the baseline scans; however, if the time of the conversion is long after the baseline, the structure of the brain at the conversion time is more likely to have dramatic changes than that at the baseline, which means that the scan of the baseline still shows a healthy status even if it develops into AD in the long term.

In addition, some participants eventually developed only MCI (rather than AD), but they were still clustered into AD. Regardless of the wrong clustering by the proposed methods, it is also likely that the participants developed to progressive MCI and their brain had some AD-related changes but still exhibited symptoms of MCI. We will use this in our future work to explore whether progressive MCI and stable MCI can be discriminated using CN and AD scans. This will involve clustering all MCI, CN, and AD scans into two groups, and if an MCI scan is clustered into AD group or CN group, we will determine whether it is more likely to be progressive MCI or stable MCI, respectively.

Compared with some state-of-the-art methods that detect regions pathologically related to AD, our method located similar regions. In their study, Zhang et al. [9] noted the left-hand side of hippocampus from slice 72 to slice 74 of the brain MRI scan as the parts with obvious changes. Nigri et al. [10] highlighted the region within the left hippocampus and ventricles in form of a heat map using Occlusion Test from their proposed neural network, AlexNet 2D plus Channel, similar to the red region in Fig. 2 (a). Nigri et al. [10] used the Swap Test, to interpret their network: the Swap Test highlighted the hippocampus in the coronal plane, which is also similar to the ROI in Fig. 2 (b).

In addition to the detection of pathologically AD-relevant regions, the clustering performance used in our study compared favourably with other state-of-the-art clustering methods (shown in Table 3). Bi et al. [15] used similar methods, PCANet and k-means, to cluster CN and AD data, and achieved overall accuracy of 89.15%. Although the accuracy in their study was 5% greater, accuracy alone is not sufficient to comprehensively evaluate the performance of a method, because it only measures its overall capability to discriminate between CN and AD. Other measures of accuracy, e.g., sensitivity, specificity, etc., as we presented here, together provide a more rounded evaluation of the performance of a method, especially when data are unbalanced, i.e., there are significantly more of one class than another. In other words, if, for example, the number of the data that are correctly clustered into CN is extremely large but the number of the data correctly clustered into AD is small, the overall accuracy still can be good, especially when the two classes of data are unbalanced: thus comparing other metrics beyond accuracy is important.

Farouk et al. [14] used k-means method alone based on the selected ROIs and the whole brain, respectively. They obtained the best performance when using the whole brain (accuracy = 76.3%). Our accuracy was around 8% greater than this, which suggests that the PCANet is a key component that enhances the performance when clustering CN and AD data. Escudero et al. [16] also employed k-means to cluster CN and AD based on the normalised left and right hippocampus volumes from MRI scans, and 73.6% of AD were correctly clustered. Our proposed method achieved 79.65% sensitivity, namely, 79.65% of the AD cases were clustered into the correct cluster, which is more than 6% higher than Escudero et al.'s [16] method. Cabreza et al. [24] applied a GAN to MRI scans and achieved an accuracy of 74.44%, a sensitivity of 73.86%, and a PPV of 82.28%, all of which were lower than our method.

In addition to the studies that employed unsupervised learning methods, our method also outperforms some state-of-the-art supervised learning methods, as listed in the Table 3. From Table 3, we can see that the accuracy of our method is higher than those of some studies [30313233]. Although some studies [252627] achieved higher accuracy, some of the other measures that were achieved using our method were higher, i.e., sensitivity or specificity. Zeng et al. [8], Lin et al. [28], and Lin et al. [29] achieved accuracies of 98.62%, 88%, and 88.79%, respectively, which were much higher than our method. The reason

Table 3
The studies of CN/AD clustering and classification based on structural MRI. Empty values (–) indicate these measures were not reported in the respective publications.

	Study	Accuracy	Sensitivity	Specificity	PPV	NPV	Dataset	Method
Unsupervised learning	[14]	76.3%	–	–	–	–	113 CN, 162 AD	K-means
	[15]	89.15%	–	–	–	–	307 CN, 243 AD	PCANet, k-means
	[16]	–	69%	94%	–	–	200 CN, 200 AD	K-means
	[24]	74.44%	73.86%	–	82.28%	–	Not reported	GAN
Supervised learning	[25]	96.92%	73%	88%	76.73%	82%	36 CN, 38 AD	HPT-TSVM
	[26]	84.4%	83.6%	85.9%	–	–	226 CN, 186 AD	FSNet
	[27]	87.1%	93.3%	85.5%	–	–	61 CN, 37 AD	DBAD CNN based model
	[8]	98.62%	–	–	–	–	92 CN, 92 AD	DBN-based multi-task learning
	[28]	88%	–	–	–	–	330 CN, 336 AD	ROI-based CNN
	[29]	88.79%	–	–	–	–	229 CN, 188 AD	CNN, PCA, extreme learning machine
	[30]	83.7%	79.16%	87.2%	–	–	228 CN, 188 AD	CNN
	[31]	80%	–	–	–	–	61 CN, 50 AD	3D-CNN
	[32]	81.3%	–	–	–	–	233 CN, 188 AD	ResNet
	[33]	81.25%	–	–	–	–	92 CN, 92 AD	SDPSO-SVM
	Our approach	**84.17%**	**79.65%**	**88.05%**	**85.46%**	**83.64%**	**231 CN, 198 AD**	**Two-sample *t*-test, PCANet, k-means++**

could be that their methods learned more information from the labelled data by using supervised neural networks with backpropagation. However, the backpropagation also makes the process of diagnosis more time-consuming. Therefore, this becomes a trade-off between performance and efficiency: although our method achieved a slightly lower performance, it reduced the reliance on the labels of the data and the overall process was quicker. In other words, our method balanced the performance (a relatively advantageous performance) and the efficiency (quicker discrimination) of the diagnosis.

An additional advantage of our using unsupervised learning methods, is that these do not require training data or test data. This means that there is no data leakage, a limitation identified by Wen et al. [34] in a number of earlier studies [35–38].

The advantages of the method proposed in our study can be summarised as follows: Our method reduced the need for labelled MRI scans in the process of diagnosis of AD, which required only 30 scans from CN and AD groups, respectively, for locating the AD-relevant regions. PCANet, as a 2-stage unsupervised neural network, converted the low-level features (voxels) within the ROIs to higher-level features without backpropagation. Although a number of CNN architectures, e.g., ResNet, LeNet, AlexNet, etc., have been developed, most of these architectures are supervised, which requires backpropagation and is time-consuming. However, the process of the feature extraction in the PCANet does not rely on the labels for the data and this makes it much quicker than supervised neural networks. In addition, although there are improved versions of PCA, e.g., kernel PCA, they are still mainly used for reducing the dimensions of data rather than feature extraction. Therefore, PCANet manipulates the unsupervised characteristic of the PCA and the feature extraction of CNN to extract features from the data very well, and without the need for labels. Finally, k-means++ performed well in discriminating between CN and AD without using labels for the data, which is relatively quick and simple. In addition to diagnosing AD, our method also has the potential to predict the development of AD; this requires further development and testing.

However, the proposed method has some limitations: First, the PCANet used a basic (linear) PCA to learn convolutional filters, which means that it cannot capture nonlinear structures within the data. Second, although the PCANet is simply structured (with only two convolution layers) and is quicker in extracting features due to not having backpropagation, it increases the dimension of the original features. In other words, the dimension of the output from the PCANet is much higher than that of the input. This suggests that, if the input of the PCANet has high dimensions, the extracted features will occupy a larger volume of computational memory. Therefore, when a PCANet is applied to MRI scans, ROIs or slices containing ROIs need to be selected beforehand. In this way, the performance of the PCANet also depends on the detection of the ROIs in this study. Despite these limitations, the methods used in this study make an important contribution to the detection of AD in MRI scans as is summarised in the next section.

4. Conclusion

AD is the most common type of dementia, and it is becoming a worldwide problem. The available AD diagnosis that uses machine learning and deep learning often requires a large amount of labelled data but labelling medical imaging data is quite difficult. Methods involving a limited amount of labelled data or data without labels should therefore be considered.

In this paper, we proposed a method to discriminate between CN and AD using a limited number of labelled brain structural MRI scans. We first employed a two-sample *t*-test to detect ROIs between CN and AD using the labelled scans; we then used PCANet to extract high-level features from the ROIs; finally, we utilised k-means++ to divide the MRI scans into two clusters using the high-level features. The proposed method achieved accuracy of 84.17%, sensitivity of 79.65%, specificity of 88.05%, PPV of 85.46%, and NPV of 83.64%. In addition, the method potentially predicts the progression of the CN participants, and it suggests that stable MCI and progressive MCI can be discriminated by using AD and CN data. This could be of potential benefit to clinicians in diagnosing people with dementia.

The contribution of the study can be summarised as follows. This study proposed an unsupervised approach that combines a statistical method, a neural network, and a clustering algorithm, to locate AD-relevant regions and to discriminate between CN and AD based on only a limited number of labelled MRI scans. The method in this study reduced the reliance on the labels of the data and this could potentially save much of the time for the diagnosis of AD. In addition, this study also provides evidence that unsupervised learning methods have the potential to predict the development of AD and thus they could be used to help with the diagnosis of stable MCI and progressive MCI.

Based on the limitations of this study and proposed method, some improvements could be developed in future work. In recent years, MCI has been increasingly studied due to its being an intermediate stage in the development of AD. The two substages, stable MCI (sMCI) and progressive MCI ((pMCI), are potentially indicative of a person's likelihood of worsening and progressing to AD (pCMI) or not (sMCI) [394084142]. Therefore, future work could use MRI scan data to cluster and differentiate between stable MCI and progressive MCI. In addition, the grey matter in the brain also has an essential role in more accurately diagnosing AD in previous studies [9264344], this could be used in future work to improve the accuracy of the diagnosis. As for the method used in this study, a two-sample z-test could be potentially used for detecting ROIs based on a larger number of labelled data (over 30 scans in each groups); other unsupervised neural networks, e.g., anchor neighbourhood discovery (AND) [45], could be considered for extracting features. For the clustering stage, other optimised k-means methods e.g., quantum-inspired ant lion optimised hybrid k-means [6], could be used in future work, and, in addition to k-means algorithms, Gaussian mixture models could also be considered to cluster the data in future research.

Summary table.

What was already known on the topic?

- Millions of people suffering from Alzheimer's disease (AD) globally; however, manual diagnosis of AD by doctors is time-consuming;
- Machine learning techniques can help doctors to identify patients with AD using magnetic resonance imaging (MRI) scans.
- Most previous studies employed supervised learning methods and MRI scans to diagnose AD, which means that their methods required a large number of labelled MRI scans. However, medical imaging data are very difficult to label. Therefore, the methods for diagnosing AD require further development and improvement.

What this study added to our knowledge?

- This study proposed a new approach that combines a statistical method and unsupervised learning models to diagnose AD using a limited number of labelled MRI scans.
- The method achieved a relatively good performance for the diagnosis of AD.
- The proposed method also detected the AD-associated regions in the brain and validated them.
- The study also provided a basis that the unsupervised learning methods have the potential to predict the progression of AD.
- Future studies should report sensitivity, specificity, positive and negative predictive values, as well as overall accuracy. Ninety-five percent CI should also be included.

CRediT authorship contribution statement

Yuyang Liu: Conceptualization, Methodology, Visualization, Data

curation, Formal analysis, Investigation, Writing – original draft. **Suvodeep Mazumdar:** Conceptualization, Methodology, Writing – review & editing, Supervision, Validation. **Peter A. Bath:** Conceptualization, Methodology, Writing – review & editing, Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Grant) 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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2023.105027.

References

[1] "Dementia," World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dementia (Accessed: 16-Jun-2022).
[2] L. Caroline dos Santos Picanço, et al., Alzheimer's disease: a review from the pathophysiology to diagnosis, new perspectives for pharmacological treatment, Curr. Med. Chem. 25 (2018) 3141–3159.
[3] Alzheimer's Association, 2021 Alzheimer's Disease Facts and Figures, 2021.
[4] G.B. Frisoni, N.C. Fox, C.R. Jack, P. Scheltens, P.M. Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. (2010).
[5] S.M. Atabo, A.A. Umar, A review of imaging techniques in scientific research/clinical diagnosis, MOJ Anat. Physiol. 6(5) (2019).
[6] J. Chen, X. Qi, L. Chen, F. Chen, G. Cheng, Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection, Knowledge-Based Syst. 203 (Sep. 2020).
[7] H. Li, P. Wu, Z. Wang, J. Mao, F.E. Alsaadi, N. Zeng, A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis, Comput. Biol. Med. (Dec. 2022) 106265.
[8] N. Zeng, H. Li, Y. Peng, A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease, Neural Comput. Appl. (2021).
[9] Y. Zhang, Q. Teng, Y. Liu, Y. Liu, X. He, Diagnosis of Alzheimer's disease based on regional attention with sMRI gray matter slices, J. Neurosci. Methods 365 (October 2021) (2022).
[10] E. Nigri, N. Ziviani, F. Cappabianco, A. Antunes, A. Veloso, Explainable deep CNNs for MRI-based diagnosis of Alzheimer's disease, Proc. Int. Jt. Conf. Neural Networks (2020).
[11] D. Baskar, V.S. Jayanthi, A.N. Jayanthi, An efficient classification approach for detection of Alzheimer's disease from biomedical imaging modalities, Multimed. Tools Appl. 78 (10) (2019) 12883–12915.
[12] K. Raza, N.K. Singh, A tour of unsupervised deep learning for medical image analysis, Curr. Med. Imaging Former. Curr. Med. Imaging Rev. 17 (9) (2021) 1059–1077.
[13] J. Latif, C. Xiao, A. Imran, S. Tu, Medical imaging using machine learning and deep learning algorithms: a review, in: 2019 2nd Int. Conf. Comput. Math. Eng. Technol. iCoMET 2019, 2019.
[14] Y. Farouk, S. Rady, Early diagnosis of Alzheimer's disease using unsupervised clustering, Int. J. Intell. Comput. Inf. Sci. 20 (2) (2020) 112–124.
[15] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, X. Ma, Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology, Neurocomputing (2019).
[16] J. Escudero, J.P. Zajicek, E. Ifeachor, Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means, Conf. Proc. IEEE Eng. Med. Biol. Soc. 2011 (2011) 6470–6473.
[17] I. Beheshti, H. Demirel, Feature-ranking-based Alzheimer's disease classification from structural MRI, Magn. Reson. Imaging (2016).
[18] S. Leandrou, D. Lamnisos, P.A. Kyriacou, S. Constanti, C.S. Pattichis, Comparison of 1.5 T and 3 T MRI hippocampus texture features in the assessment of Alzheimer's disease, Biomed. Signal Process. Control 62 (July) (2020) 3–8.
[19] S. Seyedi, et al., Comparing VBM and ROI analyses for detection of gray matter abnormalities in patients with bipolar disorder using MRI, Middle East Curr. Psychiatry 27 (1) (2020).
[20] B. Gerald, A brief review of independent, dependent and one sample t-test, Int. J. Appl. Math. Theor. Phys. 4 (2) (2018) 50.
[21] T.H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, PCANet: a simple deep learning baseline for image classification? IEEE Trans. Image Process. 24 (12) (2015) 5017–5032.
[22] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2006, pp. 1027–1035.
[23] P.-N. Tan, M. Steinbach, A. Karpatne, V. Kumar, Introduction to Data Mining, Second, Person Education, New York, NY, 2019.
[24] J.N. Cabreza, G.A. Solano, S.A. Ojeda, V. Munar, Anomaly detection for Alzheimer's Disease in brain MRIs via unsupervised generative adversarial learning, in: 4th Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2022 - Proc., 2022, pp. 230–234.
[25] R. Kumari, S. Goel, S. Das, Using SVM for Alzheimer ' s Disease detection from 3D T1MRI, in: 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON), 2022, pp. 600–604.
[26] H. Li, X. Shi, X. Zhu, S. Wang, Z. Zhang, FSNet: Dual interpretable graph convolutional network for alzheimer's disease analysis, IEEE Trans. Emerg. Top. Comput. Intell. (2022) 1–11.
[27] J.S. Kim, et al., Deep learning-based diagnosis of Alzheimer's disease using brain magnetic resonance images: an empirical study, Sci. Rep. 12 (1) (2022) 1–8.
[28] J. Wen, et al., Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation, Med. Image Anal. 63 (Jul. 2020).
[29] W. Lin et al., Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment, Front. Neurosci., 12(NOV) (2018).
[30] K. Aderghal, M. Boissenin, J. Benois-Pineau, G. Catheline, K. Afdel, "Classification of sMRI for AD diagnosis with convolutional neuronal networks: a pilot 2-D+ε study on ADNI, in: Multimedia Modeling International Conference, 2017, vol. 1, no. January, pp. 226–237.
[31] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, in: Proc. - Int. Symp. Biomed. Imaging, 2017, pp. 835–838.
[32] A. Valliani, A. Soni, Deep residual nets for improved Alzheimer's diagnosis, in: ACM-BCB 2017 - Proc. 8th ACM Int. Conf. Bioinformatics, Comput. Biol. Heal. Informatics, 2017. p. 615.
[33] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing (2018).
[34] J. Wen, et al., Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation, Med. Image Anal. 63 (Jul. 2020), 101694.
[35] E.H. Asl, et al., Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network, Front. Biosci. - Landmark 23 (3) (2018) 584–596.
[36] K. Backstrom, M. Nazari, I.Y.H. Gu, A.S. Jakola, An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images, in: Proc. - Int. Symp. Biomed. Imaging, vol. 2018-April, no. Isbi, 2018. pp. 149–153.
[37] S. Basaia et al., Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, NeuroImage Clin., 21 (Jan.) (2019).
[38] A. Farooq, S. Anwar, M. Awais, S. Rehman, A deep CNN based multi-class classification of Alzheimer's disease using MRI, in: 2017 IEEE Inter- national Conference on Imaging Systems and Techniques (IST), 2017. pp. 3–8.
[39] S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, Neuroimage 189 (2019) 276–287.
[40] Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang, Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network, Front. Neurosci. 13(MAY) (2019).

[41] X. Zhang, L. Han, W. Zhu, L. Sun, D. Zhang, An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE J. Biomed. Heal. Informatics 26 (11) (2021) 5289–5297.
[42] B. Zheng, A. Gao, X. Huang, Y. Li, D. Liang, X. Long, A modified 3D EfficientNet for the classification of Alzheimer's disease using structural magnetic resonance images, IET Image Process. (2022).
[43] A. Mehmood, et al., A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images, Neuroscience 460 (Apr. 2021) 43–52.
[44] X. Long, L. Chen, C. Jiang, L. Zhang, Prediction and classification of Alzheimer disease based on quantification of MRI deformation, PLoS One 12 (3) (2017).
[45] J. Huang, Q. Dong, S. Gong, X. Zhu, Unsupervised deep learning by neighbourhood discovery, in: 36th Int. Conf. Mach. Learn. ICML 2019, vol. 2019-June, 2019. pp. 5090–5099.