

## Neuroradiology



## Automated classification of Alzheimer's disease, mild cognitive impairment, and cognitively normal patients using 3D convolutional neural network and radiomic features from T1-weighted brain MRI: A comparative study on detection accuracy

Amin Zarei<sup>a</sup>, Ahmad Keshavarz<sup>a,\*</sup>, Esmail Jafari<sup>b</sup>, Reza Nemati<sup>c</sup>, Akram Farhadi<sup>d</sup>, Ali Gholamrezaezhad<sup>e</sup>, Habib Rostami<sup>f</sup>, Majid Assadi<sup>b</sup>

<sup>a</sup> IoT and Signal Processing Research Group, ICT Research Institute, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran

<sup>b</sup> The Persian Gulf Nuclear Medicine Research Center, Department of Nuclear Medicine, Molecular Imaging, and Theranostics, Bushehr Medical University Hospital, School of Medicine, Bushehr University of Medical Sciences, Bushehr, Iran

<sup>c</sup> Department of Neurology, Bushehr Medical University Hospital, Faculty of Medicine, Bushehr University of Medical Sciences, Bushehr, Iran

<sup>d</sup> The Persian Gulf Tropical Medicine Research Center, The Persian Gulf Biomedical Sciences Research Institute, Bushehr University of Medical Sciences, Bushehr, Iran

<sup>e</sup> Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

<sup>f</sup> Computer Engineering Department, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran

## ARTICLE INFO

## Keywords:

Alzheimer's disease  
Mild cognitive impairment  
Convolutional neural networks  
Radiomic  
MRI

## ABSTRACT

**Objectives:** Alzheimer's disease (AD) is a common neurodegenerative disorder that primarily affects older individuals. Due to its high incidence, an accurate and efficient stratification system could greatly aid in the clinical diagnosis and prognosis of AD patients. Convolutional neural networks (CNN) approaches have demonstrated exceptional performance in the automated stratification of AD, mild cognitive impairment (MCI) and cognitively normal (CN) participants using MRI, owing to their high predictive accuracy and reliability. Therefore, we aimed to develop an algorithm based on CNN and radiomic features derived from ROIs of bilateral hippocampus and amygdala in brain MRI for stratification between AD, MCI and CN.

**Methods:** In this study, we proposed a CNN and radiomic features-based algorithm using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. T1-weighted images were used. We utilized three datasets, including AD (199 cases, 602 images), MCI (200 cases, 948 images), and CN (200 cases, 853 images), to perform binary classification (AD vs. CN, AD vs. MCI, and MCI vs. CN). Finally, we obtained the accuracy (ACC) and the area under the curve of the receiver operating characteristic curve (AUC) to evaluate the performance of the algorithm.

**Results:** Our proposed algorithm achieved acceptable overall discrimination accuracy. In the term of AD vs CN, radiomic-based algorithm alone obtained ACC of 82.6 % and AUC of 88.8, CNN-based algorithm obtained ACC of 80 % and AUC of 87.2 and their fusion showed ACC of 84.4 % and AUC of 90. In the term of MCI vs CN, radiomic-based algorithm alone obtained ACC of 71.6 % and AUC of 77.8, CNN-based algorithm obtained ACC of 69 % and AUC of 75 and their fusion showed ACC of 72.7 % and AUC of 80. In the term of AD vs MCI, radiomic-based algorithm alone obtained ACC of 57 % and AUC of 57.5, CNN-based algorithm obtained ACC of 56.6 % and AUC of 57.7 and their fusion showed ACC of 58 % and AUC of 59.5.

**Conclusion:** In conclusion, it has been determined that hippocampus and amygdala-based stratification using CNN features and radiomic features-based algorithm is a promising method for the classification of AD, MCI, and CN participants.

**Advances in knowledge:** This study proposed an automated procedures based on MRI-derived radiomic features and CNN for classification between AD, MCI and CN.

\* Corresponding author at: Electrical Engineering Department, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran.  
E-mail address: [a.keshavarz@pgu.ac.ir](mailto:a.keshavarz@pgu.ac.ir) (A. Keshavarz).

## 1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that causes problems with behavior, memory, and thinking, ultimately leading to the death of patients.<sup>1</sup> This disease is one of the most expensive diseases in developed countries. As of 2006, Alzheimer's disease has affected over 26.6 million people worldwide, and it is predicted that by 2050, 1 in 85 people will be affected by this condition.<sup>2</sup> Currently, there is no definitive cure for Alzheimer's disease, and the goal is only to reduce the progression of the disease, improve symptoms, solve behavioral problems, and improve the quality of life of affected individuals. If the disease is diagnosed in the early stages, current treatments can temporarily reduce the progression of the disease. However, more effective treatments and ultimately prevention or even a complete cure are essential goals.<sup>3</sup>

Biomarkers of this disease can be considered an indicator of cognitive decline. As the disease progresses over time, biomarker levels also reach abnormal levels at a predictable stage. These biomarkers include beta-amyloid, neuronal degeneration, brain atrophy, and neuronal destruction in sensitive areas, leading to memory loss and decreased cognitive ability.<sup>4</sup> The first three indicators, especially tissue degeneration, which is also traceable in imaging, occur before the full onset of the disease, and their examination can be helpful in early diagnosis. MRI imaging, due to its high resolution and ability to detail brain tissue anatomy, is a useful and powerful tool for tracking neuronal destruction.<sup>5</sup> Currently, artificial intelligence (AI) algorithms are employed in computer-assisted techniques to achieve the diagnosis of Alzheimer's disease.<sup>6</sup>

As AI continues to rapidly advance, researchers are utilizing AI methodologies, particularly deep learning, to tackle intricate issues in various fields, notably medicine. In the realm of Alzheimer's disease, researchers have expanded the application of multiple deep learning models to diagnose different stages of the disease. With the aid of computer-assisted neuroimaging studies, significant progress has been made in the classification of Alzheimer's disease (AD) and cognitively normal (CN) participants. However, while binary classification of AD and CN participants has been successful, it is not as informative as predicting the early-stage transition from mild cognitive impairment (MCI) to AD. Most research has focused solely on binary categorization and has not predicted whether a patient has MCI or the likelihood of progression to AD.<sup>7</sup> In patients with MCI, the brain structure exhibits characteristics of both a healthy individual and a person with Alzheimer's disease. Detecting this stage and determining the extent of the possibility of an individual with MCI progressing to Alzheimer's has always been a major challenge. In these patients, brain changes continue for a while, and some symptoms begin to appear, leading to difficulties that significantly affect daily activities and may ultimately lead to cognitive decline.<sup>8</sup> Early detection of Alzheimer's disease during its prodromal stage or even predicting its likelihood is crucial for effective treatment, just as it is for other illnesses. Treatment outcomes for AD patients are more favorable if they receive prompt care upon suspicion of AD biomarkers or symptoms. Delaying the progression of Alzheimer's disease by one year can lead to a 10 % reduction in the number of affected individuals. The data indicates that detecting Alzheimer's disease in its early stages is essential in reducing the global number of patients with the disease.<sup>9</sup>

Neurologists currently rely on manual examination of brain scans and cognitive assessments to accurately diagnose the symptoms and progression of Alzheimer's disease. However, subtle changes in brain structure can be observed years before distinct biomarkers are identifiable by humans. It is recognized that the human visual system may not be able to detect these subtle changes that contain critical information about a patient's disease state, even when examined by experienced neurologists. Therefore, an AI-based computer-aided system can assist neurologists in detecting complex brain disorders while minimizing the potential for misdiagnosis. Additionally, this technology is expected to

reduce the workload for medical professionals and decrease patient visits and waiting times.<sup>7</sup>

In the past few decades, feature extraction methods have been commonly used to extract features from medical images and biological markers. These features were then fed to popular classifiers such as support vector machine (SVM) and random forest (RF) to obtain acceptable results.<sup>10</sup> Recently, neural networks, especially convolutional neural networks (CNN), have been introduced for automatic feature extraction and classification of diseases. They support clinical decision-making by experts and have reported better results.<sup>11</sup>

Furthermore, radiomics is an emerging field in medical research that aims to extract a large number of statistically defined features from medical images using data analysis algorithms. These features have the ability to identify certain characteristics of a disease that may not be visible to the naked eye.<sup>12</sup> Numerous studies in the field of medicine have utilized these statistical features to extract useful information from various imaging modalities, particularly MRI.<sup>12,13</sup>

Therefore, we aimed to develop an algorithm based on CNN and radiomic features derived from ROIs of bilateral hippocampus and amygdala in brain MRI for stratification between AD, MCI and CN.

## 2. Materials and methods

### 2.1. ADNI participants

In this study, we utilized subject data from ADNI with individual Institutional Review Board (IRB) approval (<http://adni.loni.usc.edu/>). Study subjects provided written informed consent at the time of enrollment for the collection, storage, and research use of their data, and completed questionnaires approved by their respective IRBs. Before being shared, the data was anonymized. We randomly selected baseline structural MRI data (T1-weighted images) from a total of 2403 images of 599 different individuals, including 853 images of 200 cases with cognitively normal controls (CN), 602 images of 199 patients with AD, and 948 images of 200 patients with mild cognitive impairment (MCI) at baseline.<sup>14</sup> Table 1 showed imaging protocol was used for MRI acquisition. Table 2 provides a brief summary of the demographic information for the ADNI participants included in this study.

### 2.2. Preprocessing

The preprocessing performed on the selected data in this study includes two important categories of primary and secondary preprocessing.

#### 2.2.1. Primary preprocessing

All selected images have undergone four important preprocessing stages.<sup>15</sup> Each of the preprocessing steps has been performed for a specific purpose and is necessary considering the nature of image acquisition and the presence of noise and image distortions during imaging. The preprocessing stages include:

- 1) Correction of geometric image distortions caused by nonlinear gradients (Gradwrap).
- 2) Correction of non-uniform brightness intensity caused by rf

**Table 1**

Imaging protocol was used for MRI acquisition.

Acquisition Plane: SAGITTAL	Acquisition Type: 3D
Coil: PA	Field Strength: 1.49399995803833 Tesla
Flip Angle: 8.0 degree	Manufacturer: SIEMENS
Matrix X: 192.0 pixels	Mfg Model: Symphony
Matrix Y: 192.0 pixels	Pulse Sequence: IR/GR
Matrix Z: 160.0 pixels	Weighting: T1
TE: 3.609999895095825 ms	Pixel Spacing X: 1.25 mm
TI: 1000.0 ms	Pixel Spacing Y: 1.25 mm
TR: 3000.0 ms	Slice Thickness: 1.2000000476837158 mm

**Table 2**  
The demographic information for the ADNI participants.

Diagnostic group	Number of patients	Number of images	Mean age (std)	Gender (M/F)
AD	199	602	76 (7.62)	103/96
CN	200	853	77 (5.04)	102/98
MCI	200	948	76 (7.24)	144/56

AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment.

transmission (B1 non-uniformity).

3) Correction of non-uniform brightness intensity caused by wave and dielectric effects in imaging devices with magnetic field strengths of 1.5 T and 3 T (N3).

4) Scaling the image based on an available phantom sample (Scaled).

### 2.2.2. Secondary preprocessing

Secondary preprocessing included skull stripping and region of interest (ROI) extraction.

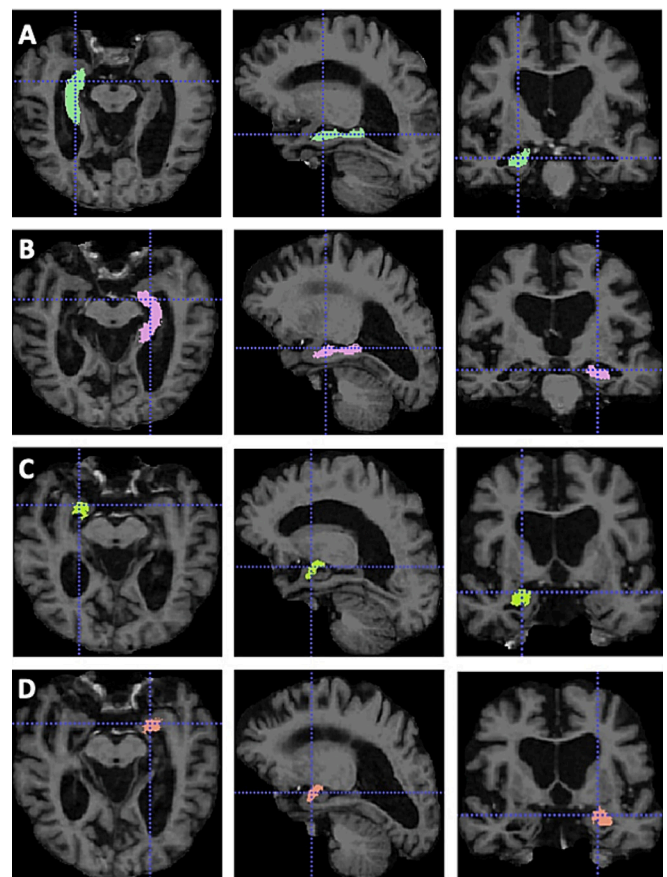
**2.2.2.1. Skull stripping.** Removing the skull from the images is typically the first step in eliminating non-brain regions from the images for various clinical applications and analyses. After undergoing the four primary preprocessing stages, all selected images were skull stripped using FreeSurfer software (<https://surfer.nmr.mgh.harvard.edu/>).<sup>16</sup> Fig. 1 depicts an example of brain image before and after skull stripping processing.

**2.2.2.2. ROI extraction.** To extract ROI, two regions, including the hippocampus and amygdala, were targeted. The FMRIB Software Library (FSL) was used for this purpose (<https://fsl.fmrib.ox.ac.uk/fsl/>). The FMRIB's Integrated Registration and Segmentation Tool (FIRST)<sup>17</sup> within FSL processed brain images, converting scans to MNI152 format at 1 mm resolution. It utilized a labeled atlas to extract fifteen brain regions from both hemispheres. An example of processed output of the software for extracting each of the mentioned regions in the images is presented in Fig. 2.

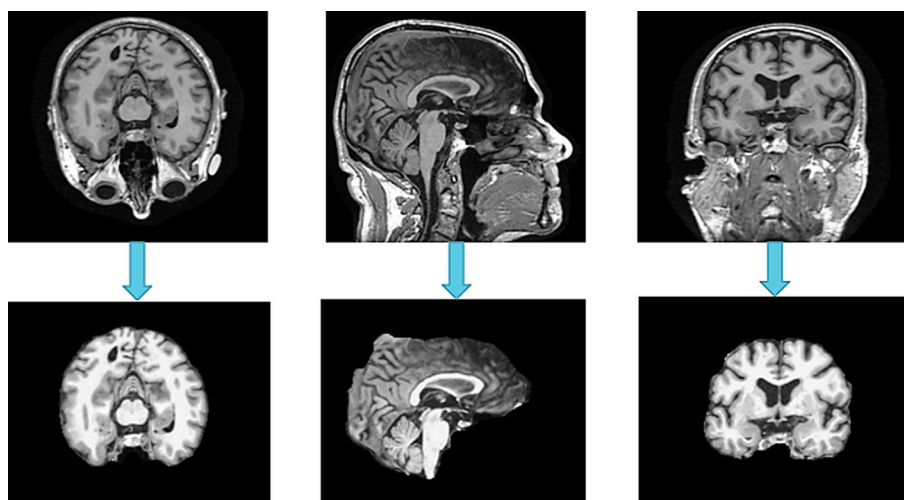
The process of preparing the images, from downloading them from the database to preparing them for input into the proposed algorithm, is displayed in Fig. 3.

### 2.3. Radiomic feature extraction

Radiomic features were extracted using PyRadiomics (<https://pyradiomics.readthedocs.io/en/latest/>) in the Python programming environment (v3.8). A total of 1746 features have been extracted from each ROI region separately obtained from the left and right hippocampus and amygdala. The radiomic features are calculated based on flowchart presented in Fig. 4, first extracted once from the original image and then again after passing through each of the mentioned filters. Table 3 displayed the extracted radiomic features.



**Fig. 2.** An example of processed output of the FSL-FIRST for extracting right hippocampus (A), left hippocampus (B), right amygdala (C) and left amygdala (D).



**Fig. 1.** An example of brain image before and after skull stripping processing with FreeSurfer software.

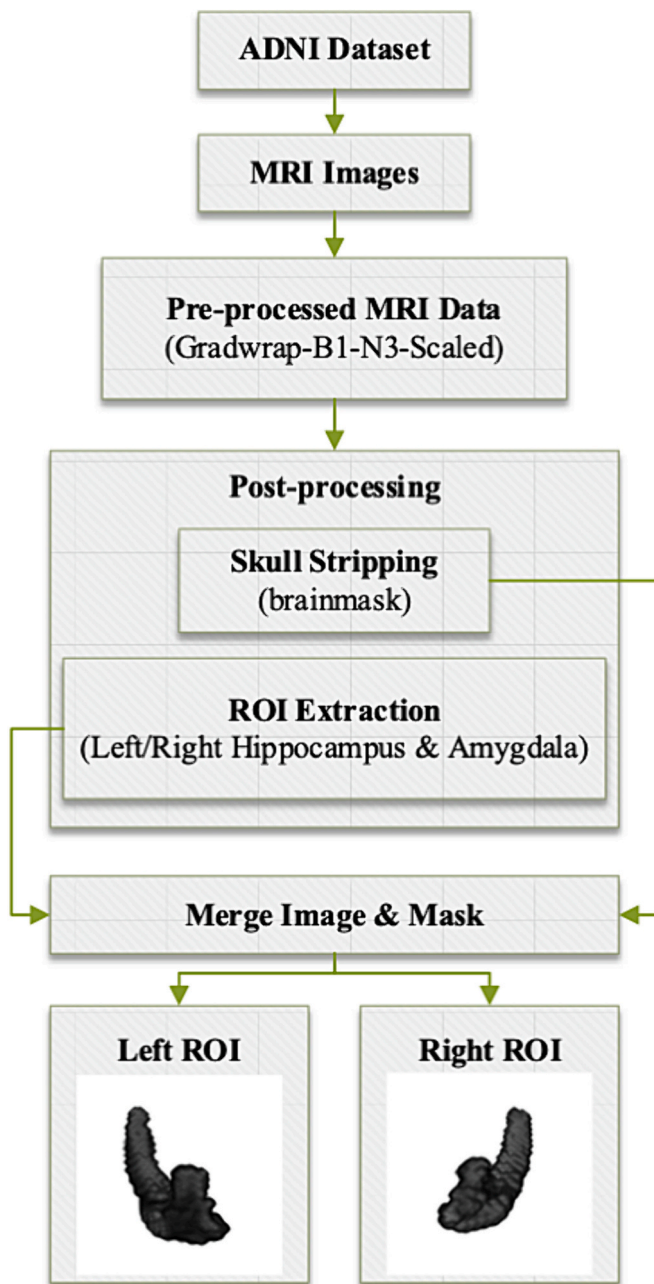


Fig. 3. The flowchart for Image preparation from the download stage to preparation for Input into the proposed algorithm.

#### 2.4. Clinical parameters

Given the importance of age and gender in the structural differences of the brain among individuals, we also injected these values as two feature vectors into the fully connected layer. To input the age values into the network, we normalized them by dividing them by 100. Male gender was represented by the number 1, while female gender was represented by  $-1$ . These values were then prepared for input into the final network.

#### 2.5. Proposed convolutional neural network (CNN)

CNNs extract the informative features from the input image to classify it. The performance of them depends to the structure of the network, loss function and the quality and number of training samples. Due to the

most informative features of the brain images for Alzheimer diagnosis are texture and shape of hippocampus and amygdala regions,<sup>18</sup> two 3D CNNs were designed to extract these features. The general schematic of this network is shown in Fig. 5. The radiomics features are fused with the output of last convolutional layer, in order to direct the network to focus on the extraction of texture and shape features. In fact, the radiomics features of Table 3 are related to shape and texture and it is supposed using of them, the network focus less on low informative features like edge or using of blurring filters. The obtained results can show this issue. Furthermore, age and gender are injected to the network structure to improve the results. The selected 3D CNN architecture comprises three convolutional layers, each with a stride of two, and without using pooling. The ReLU activation function was chosen for each convolutional layer. Additionally, after output of each convolutional layer pass through the activation function to nonlinearized the values in the network, a dropout layer with rates of 0.25, 0.25, and 0.3 was utilized.

After the final convolutional layer, a batch normalization layer has also been employed. This layer redefines the minimum and maximum range of values after the operations conducted in the previous layers to a narrower and smaller range (1 and  $-1$ ), thereby enhancing the training process's speed and improving the results.

The input images of the left and right hippocampus have been resized to  $16 \times 40 \times 24$  after eliminating the black background areas around them. Therefore, for hippocampus-based classification, the input image size of convolutional layers in the network is  $16 \times 40 \times 24$ . The input images of the left and right amygdala have also been resized to  $20 \times 20 \times 20$  after removing the black background areas around them. Therefore, for amygdala-based classification, the input image size of convolutional layers in the network is  $20 \times 20 \times 20$ . These three convolutional layers will perform the feature extraction operation on the existing brain regions in succession (Fig. 5).

The proposed neural network is designed to use both the extracted radiomic features and features extracted by the convolutional layers during the training process. The overall network includes input layers for inputting ROI images, performing convolutional operations, and input layers for feeding all the extracted radiomic features. After passing the images through the convolutional layers, the outputs are flattened by the Flatten layer, connected to each other, and integrated. The resulting radiomic feature vector is then merged with the features obtained form of a fully connected layer, and the output is transformed into another fully connected layer with 128 neurons and the ReLU activation function to reduce dimensionality. In the last fully connected layer, Dropout with a rate of 0.4 is implemented. Finally, these 128 neurons are input to the last layer with 1 neuron, commonly referred to as the classifier layer. The activation function of the last layer is sigmoid, which is usually used for binary classifiers, and its output represents the probability of belonging to a specific group. If the output is less than 0.5, it indicates belonging to the first group, and if it is higher than 0.5, it indicates belonging to the second group. Table 4 displays the selected parameters for training the final neural network.

To assess the discriminative ability of the proposed classifier against other classifiers, an SVM classifier has also been employed once in the last layer. Furthermore, the radiomics features were fed to the SVM classifier once independently, resulting in three separate classifiers being present simultaneously. To determine the appropriate parameters for the SVM classifier, a grid search method was employed, which selects the best parameters with the highest classification accuracy by substituting various parameters and presenting the outcomes. The overall results obtained from applying the majority vote on these three classifiers are as follows, if two classifiers agree on a group, that group is selected, and the vote of the third classifier is not considered. The implementation details and results are accessible from the GitHub repository: [https://github.com/aminz1995/Alzheimers\\_3DCNN\\_Radiomics](https://github.com/aminz1995/Alzheimers_3DCNN_Radiomics).

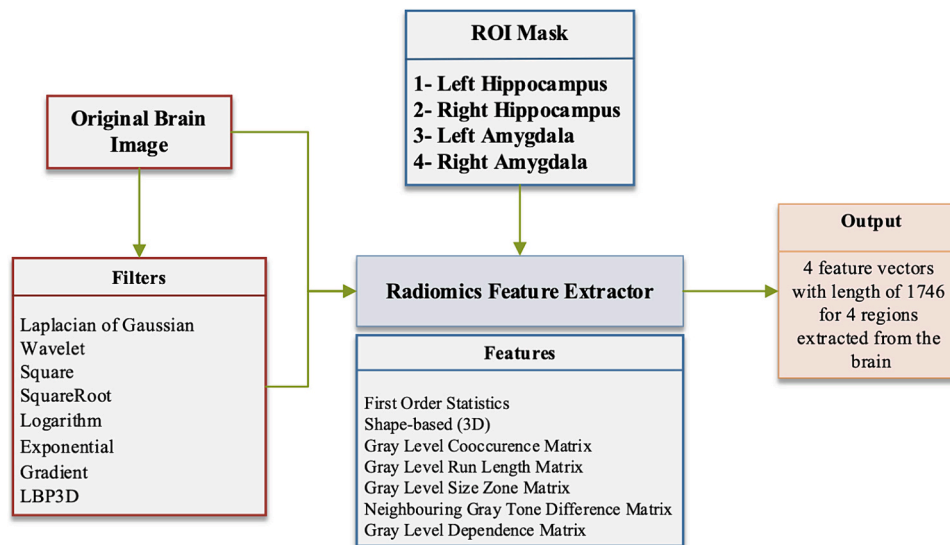


Fig. 4. The flowchart of radiomic features extraction from ROI.

## 2.6. Experiments

Performance of the CNN was validated and tested with three binary classifications: AD vs CN, AD vs MCI, MCI vs CN. The performance evaluation method employed in this study was the 5-fold cross-validation method. The process for each classification comprised three steps - training, validation, and testing. The initial step involved randomly dividing the MRI data for each classification dataset into a sizeable training and validation set (80 % of images) and a testing set (20 % of images). The learning process comprises both training and validation stages, where training is performed to determine the network parameters, and validation is carried out to progress towards optimal parameters. Therefore, in each of the five repetitions, 20 % of the remaining 80 % were used for validation, and the remainder, randomly, were used for training.

In the final step, CNN was employed to classify the raw images of the testing set. In our study, we evaluated our binary classification method to investigate the impact of incorporating region of interest (ROI) features. We measured the improvement in algorithm performance with and without radiomic features alongside a convolutional neural network. To evaluate the performance of the CNN, various performance measures were used such as accuracy (ACC) and receiver operating characteristic (ROC) curves with the corresponding area under the ROC curve (AUC).

Additionally, proposed algorithm results were compared to a ten-year experienced radiologist. For this comparison, ten AD patients (23 images), ten CN (25 images) and ten MCI (24 images) were selected randomly among testing set and the hippocampus and amygdala images of them were evaluated by the radiologist. The radiologist did not know the clinical information of the patients, except age and gender.

## 3. Results

Table 5 presents the outcomes of the proposed algorithm and the impact of merging radiomic features with features extracted by the CNN for discriminating between AD and CN participants, compared to the scenario without combination. Similarly, Tables 6 and 7 present the same approach for discriminating between MCI with CN and MCI with AD, respectively. In all of these evaluations, the effect of integrating features from two ROI regions extracted from brain tissue was also examined. This was done by incorporating the hippocampus region alone, the amygdala region alone, and finally, the combination of the two regions into the algorithm.

Upon examining each row of the tables, it is evident that integrating the features of the hippocampus and amygdala regions has been effective in improving performance accuracy. For instance, in the case of integrating radiomics features with CNN, this improvement is 2.4 % (ACC), whereas in the case of using only CNN, this improvement is 1.4 %, and in the case of using only radiomics features and SVM classifier, this improvement is 0.1 %. Furthermore, by evaluating the final performance of each table, it can be observed that integrating radiomics features with network features has also enhanced performance accuracy by approximately 1.8 % compared to the non-integrated state.

In contrast to the observations noted in the classification of AD/CN, as well as MCI/CN, in the classification of AD/MCI, we only observed the highest performance accuracy on the amygdala region, as evident by examining the rows of all three tables. Additionally, by evaluating the final performance of each table, it is apparent that for this classification, the CNN alone has demonstrated better performance for diagnosis.

Finally, in comparison of the radiologist and proposed algorithm, the algorithm showed better performance than radiologist in classification of the evaluated patients (Table 8).

## 4. Discussion

Early and accurate diagnosis of AD is crucial for effective treatment, and therefore, many researchers have focused on developing computer-aided systems that can diagnose AD in its early stages and on an individual basis.<sup>10–13</sup> In this study, we developed and validated a CNN that can individual diagnosis of AD and MCI patients, based on a single cross-sectional brain structural MRI scan. In addition, we used radiomic features derived from hippocampus and amygdala for stratification of AD, MCI and CN participants. The results demonstrated that our CNN performed exceptionally well in differentiating AD and MCI patients from healthy controls, especially when CNN derived features combined with radiomic features. Additionally, our proposed algorithm showed better performance than radiologist in classification of the patients.

Deep learning (DL) is a novel approach to extracting discriminative features from images. CNN is a popular type of DL used in various image analysis and computer vision tasks. DL utilizes a deep-layer receptive field combination and pooling mechanism to automatically learn visual features from input pixel images (5). Compared to traditional machine learning programs, DL has been found to deliver outstanding performance.<sup>7</sup>

The identification of bio-indicators for AD, the combination of MRI and machine learning has been particularly effective. In most cases,

**Table 3**  
The extracted radiomics features.

Row	Type	Count	Features list names
1	Shape features (3D)	17	Voxel Volume, Mesh Volume, Surface Area, Surface Volume Ratio, Compactness1, Compactness2, Sphericity, Spherical Disproportion, Maximum 3D Diameter, Maximum 2D Diameter Slice, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Major Axis Length, Minor Axis Length, Least Axis Length, Elongation, Flatness
2	First-order statistical features	19	Energy, Total Energy, Entropy, Minimum, 10th percentile, 90th percentile, Maximum, Mean, Median, Interquartile Range, Range, Mean Absolute Deviation, Robust Mean Absolute Deviation, Root Mean Squared, Standard Deviation, Skewness, Kurtosis, Variance, Uniformity
3	Second-order statistical features	24	Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Imc1, Imc2, Idm, mmc, Idmn, Id, Idn, Inverse Variance, Maximum Probability, Sum Average, Sum Entropy, Sum Squares
		16	Small Area Emphasis, Large Area Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Size-Zone Non-Uniformity, Size-Zone Non-Uniformity Normalized, Zone Percentage, Gray Level Variance, Zone Variance, Zone Entropy, Low Gray Level Zone Emphasis, High Gray Level Zone Emphasis, Small Area Low Gray Level Emphasis, Small Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Large Area High Gray Level Emphasis
		16	Short Run Emphasis, Long Run Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Gray Level Variance, Run Variance, Run Entropy, Low Gray Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Long Run High Gray Level Emphasis
		5	Coarseness, Contrast, Busyness, Complexity, Strength
		14	Small Dependence Emphasis, Large Dependence Emphasis, Gray Level Non-Uniformity, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Gray Level Variance, Dependence Variance, Dependence Entropy, Low Gray Level Emphasis, High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis, Small Dependence High Gray Level Emphasis, Large Dependence Low

**Table 3 (continued)**

Row	Type	Count	Features list names
			Gray Level Emphasis, Large Dependence High Gray Level Emphasis

GLCM: gray-level co-occurrence matrix; GLSZM: gray level size zone matrix; GLRLM: gray-level run length matrix; NGTDM: neighborhood gray-tone difference matrix; GLDM: Gray Level Dependence Matrix.

abnormal brain images are classified as either normal or abnormal in investigations of AD-related brain abnormalities. Once the disease is detected, the next steps typically involve identifying the location of the abnormalities and designing personalized treatment plans. Previous studies have identified several characteristics of AD that can be used for stratification.<sup>19,20</sup> These may include the amplitude of low-frequency fluctuations or hippocampal association with reduced frequency components, regional homogeneity, functional connectivity with strength of ROI in terms of the automated anatomical labeling atlas, whole-brain or selected regional functional correlation connectivity matrices, covariance connectivity matrices, and graph-theoretical measures.<sup>21</sup>

Based on the outcomes, our CNN demonstrated an ACC of 80 % and an AUC of 87 in stratifying between AD and CN, an ACC of 69 % and an AUC of 75 for stratifying between MCI and CN, and an ACC of 56.6 % and an AUC of 57.7 for stratifying between MCI and AD, indicating high and acceptable performance of binary classification, particularly for discriminating AD with CN and MCI with CN. Based on the results the accuracy for distinguishing AD from MCI is notably lower compared to distinguishing MCI from CN. Given that MCI is a transitional stage between normal cognition and Alzheimer's Disease, we would expect the results for AD vs. MCI and CN vs. MCI to be within almost similar range. However, this was not the case, primarily due to the less severe nature of the AD cases in the ADNI dataset.<sup>22</sup> This overlap in symptoms makes the classification of AD from MCI more challenging. Our results are consistent with previous studies. Xiaowen chen et al. developed a deep CNN by using structural brain MRI for stratification between AD, MCI and CN obtained 90.6 % for AD vs NC, AUC of 85.1 % for AD vs MCI.<sup>21</sup> In another study, Zamani et al. developed a multiobjective optimization algorithm for diagnosis of early stage of MCI as compared to CN participants using structural MRI. They achieved classification accuracy of 93 % indicating a high classification accuracy can be achieved using a single modality of biomarkers with an effective optimization method.<sup>23</sup> Silvia Basaia et al. developed and validated a deep learning CNN that utilizes a single structural MRI of the brain to predict Alzheimer's disease (AD) and mild cognitive impairment (MCI). The performance of the CNN was evaluated in distinguishing between AD, MCI patients who will progress to AD (c-MCI), and stable MCI (s-MCI). The CNN achieved a high level of accuracy in all classifications, with the highest accuracy of 99 % in distinguishing AD from CN individuals and accuracy of 75 % in distinguishing c-MCI and s-MCI.<sup>24</sup>

In addition to CNN derived features, we performed binary classification between AD, MCI and CN based on radiomic features extracted from ROIs of hippocampus and amygdala. According to the results, radiomics-based algorithm achieved an ACC of 82.6 % and an AUC of 88.8 in stratifying between AD and CN, an ACC of 71.6 % and an AUC of 77.8 for stratifying between MCI and CN, and an ACC of 57 % and an AUC of 57.5 for stratifying between MCI and AD. In concordance with our study, Feng Feng et al. performed a study to evaluate classification of AD and MCI by using radiomic features extracted from bilateral hippocampus ROI in brain MRI. They achieved an ACC of 86.75 % and AUC of 93 in discriminating AD vs CN, ACC of 70.51 % in discriminating MCI vs CN and 59.15 % in discriminating AD vs MCI.<sup>25</sup> Beheshti et al. performed a study on detection of AD using gray matter derived radiomic features from structural brain MRI and they obtained classification accuracy of 92.48 % for the diagnosis of AD.<sup>26</sup>

Additionally, we performed binary classification using fusion of CNN

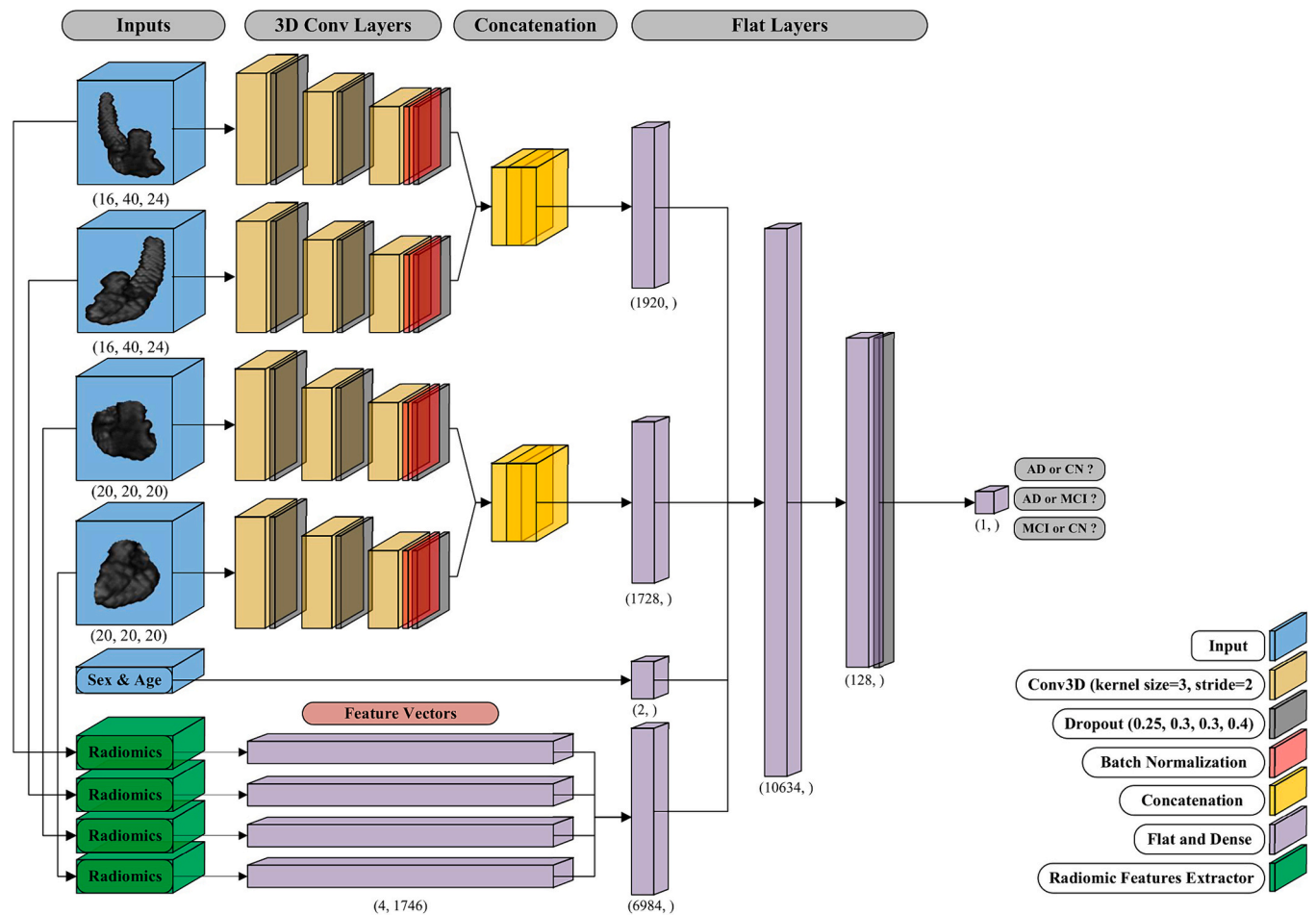


Fig. 5. The characteristics of the proposed 3D convolutional neural network (CNN).

**Table 4**  
Selected parameters for training the neural network integrated with radiomics features.

Input size	Hippocampus images (16 × 40 × 24 × 1) Amygdala images (20 × 20 × 20 × 1) Radiomic features ((4 × 1746) × 1) Age & Sex (2 × 1)
Learning Rate	0.001 (Reduce with 0.99 rate)
Batch Size	40
Epoch	250
Optimizer	Adam
Conv3D layers – Dropout Rate – Batch Normalization	3 × 4 conv layers (4-ROI): 8 filter (3 × 3 × 3)-stride (2 × 2 × 2)-relu-0.25 dropout 16 filter (3 × 3 × 3)-stride (2 × 2 × 2)-relu-0.3 dropout 32 filter (3 × 3 × 3)-stride (2 × 2 × 2)-relu-0.3 dropout- BatchNormalization
Dense layers parameters & Dropout Rate	1920 neurons (Hippo) + 1728 neurons (Amyg)+ (4 × 1746) features (ROIs) + 2 features (Age & Sex) 128 neurons-0.4 dropouts 1 neuron output
Activation function of the last layer	Sigmoid

features and radiomic features derived from ROIs bilateral hippocampus and amygdala for discrimination of AD, MCI and CN. According to the results, feature fusion-based algorithm showed better performance than CNN feature alone and radiomic feature alone in discrimination of AD vs CN, MCI vs CN and AD vs MCI. Feature fusion-based algorithm showed an ACC of 84.4 % and an AUC of 90 in stratifying between AD and CN, an

**Table 5**  
The results of the proposed radiomics-based algorithm alone, the proposed CNN-based algorithm alone and the proposed feature fusion-based algorithm in discriminating between Alzheimer's disease (AD) and cognitive normal (CN) participants for hippocampus, amygdala and fusion of them (mean ± SD).

		ROI Fusion	Hippocampus	Amygdala
Radiomic	ACC	82.6 ± 3.4	82.5 ± 3.9	80.6 ± 4.2
	AUC	88.8 ± 3.6	88.8 ± 4.4	86.4 ± 5.0
CNN	ACC	80.0 ± 4.8	76.0 ± 5.3	78.6 ± 3.1
	AUC	87.2 ± 5.2	84.2 ± 5.1	85.5 ± 4.2
Radiomic+CNN	ACC	84.4 ± 3.2	82.0 ± 2.3	81.0 ± 4.2
	AUC	90.0 ± 4.0	89.0 ± 4.4	88.0 ± 4.6

ACC, Accuracy; AUC, under the receiver operating characteristic (ROC) curve.

**Table 6**  
The results of the proposed radiomics-based algorithm alone, the proposed CNN-based algorithm alone and the proposed feature fusion-based algorithm in discriminating between mild cognitive impairment (MCI) and cognitively normal (CN) participants for hippocampus, amygdala and fusion of them (mean ± SD).

		ROI Fusion	Hippocampus	Amygdala
Radiomic	ACC	71.6 ± 3.0	71.0 ± 1.9	68.0 ± 2.7
	AUC	77.8 ± 4.5	77.6 ± 4.4	74.1 ± 3.7
CNN	ACC	69.0 ± 3.1	66.8 ± 4.0	68.4 ± 2.8
	AUC	75.0 ± 4.3	73.6 ± 5.8	75.3 ± 4.1
Radiomic+CNN	ACC	72.7 ± 3.3	71.5 ± 5.3	70.0 ± 1.6
	AUC	80.0 ± 4.6	79.3 ± 5.3	77.5 ± 3.1

ACC, Accuracy; AUC, under the receiver operating characteristic (ROC) curve.

**Table 7**

The results of the proposed radiomics-based algorithm alone, the proposed CNN-based algorithm alone and the proposed feature fusion-based algorithm in discriminating between Alzheimer's disease (AD) and mild cognitive impairment (MCI) participants for hippocampus, amygdala and fusion of them (mean  $\pm$  SD).

		ROI Fusion	Hippocampus	Amygdala
Radiomic	ACC	57.0 $\pm$ 4.0	56.2 $\pm$ 6.0	57.8 $\pm$ 2.0
	AUC	57.5 $\pm$ 6.0	57.0 $\pm$ 7.0	57.3 $\pm$ 3.9
CNN	ACC	56.6 $\pm$ 3.5	57.8 $\pm$ 3.0	61.5 $\pm$ 1.9
	AUC	57.7 $\pm$ 3.9	56.8 $\pm$ 2.3	62.0 $\pm$ 5.0
Radiomic+CNN	ACC	58.0 $\pm$ 3.1	57.5 $\pm$ 4.2	60.0 $\pm$ 4.9
	AUC	59.5 $\pm$ 3.0	58.5 $\pm$ 6.8	60.3 $\pm$ 5.5

ACC, Accuracy; AUC, under the receiver operating characteristic (ROC) curve.

**Table 8**

The accuracy of the proposed algorithm with sigmoid and SVM classifiers compared to radiologist.

	CN/AD	CN/MCI	MCI/AD
Sigmoid classifier	91.7	69.4	78.7
SVM classifier	92.1	63.3	72.4
Radiologist	76.3	57.1	55.3

MCI: mild cognitive impairment, CN: cognitively normal, AD: Alzheimer's disease, SVM: Support vector machine.

ACC of 72.7 % and an AUC of 80 for stratifying between MCI and CN, and an ACC of 58 % and an AUC of 59.5 for stratifying between MCI and AD.

In this study, we developed and validated a CNN and radiomic features based algorithm by using ROIs of bilateral hippocampus and amygdala in structural brain MRI which achieved acceptable accuracy and AUC for classification of AD, MCI and CN. In addition, our method offers practical benefits including reduced image processing time and storage space savings. The reduced data size due to feature extraction helps streamline processing and storage management, making it more suitable for large-scale clinical applications.

The major limitation of this study was the single-center limitation. We obtained data from single-center datasets, which may limit the reproducibility of findings. To overcome this limitation, using larger multi-center datasets is recommended for future studies. This will help address challenges in reproducibility and statistical power by taking testing data from independent centers. We suggest that further research be carried out on other anatomical sites and using different imaging techniques to develop a fully automated medical image stratification system that can be used in both clinical and research contexts.

## 5. Conclusion

In conclusion, our study has shown that stratifying participants based on hippocampal and amygdala features through a combination of CNN and radiomic features is a promising approach for classifying individuals with AD, MCI, and CN status. This method enhances classification accuracy and provides insights into the neuroanatomical changes associated with these conditions. By integrating advanced machine learning techniques with neuroimaging data, we can improve early diagnosis and intervention, ultimately benefiting patient outcomes. Future research should focus on validating these findings in larger, diverse populations and exploring their clinical applicability.

## CRediT authorship contribution statement

**Amin Zarei:** Methodology, Investigation, Data curation. **Ahmad Keshavarz:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Esmail Jafari:** Writing – review & editing, Writing – original draft, Data curation. **Reza Nemati:** Supervision, Methodology, Investigation. **Akram Farhadi:** Writing –

review & editing. **Ali Gholamrezaezhad:** Writing – review & editing, Supervision. **Habib Rostami:** Conceptualization. **Majid Assadi:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

## Declaration of competing interest

The author(s) declare that they have no conflict of interest.

## References

- Kocahan S, Doğan Z. Mechanisms of Alzheimer's Disease pathogenesis and prevention: the brain, neural pathology, N-methyl-D-aspartate receptors, tau protein and other risk factors. *Clin Psychopharmacol Neurosci* 2017;15(1):1–8.
- Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* 2007;3(3):186–91.
- Yiannopoulou KG, Papageorgiou SG. Current and future treatments in Alzheimer Disease: an update. *J Cent Nerv Syst Dis* 2020;12:1179573520907397.
- Kocahan S, Doğan Z. Mechanisms of Alzheimer's disease pathogenesis and prevention: the brain, neural pathology, N-methyl-D-aspartate receptors, tau protein and other risk factors. *Clinical Psychopharmacology and Neuroscience* 2017;15(1):1.
- Vu T-D, Ho N-H, Yang H-J, Kim J, Song H-C. Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection. *Soft Computing* 2018;22:6825–33.
- Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal* 2020;63:101694.
- Lim BY, Lai KW, Haskin K, et al. Deep learning model for prediction of progressive mild cognitive impairment to Alzheimer's Disease using structural MRI. *Frontiers in aging. Neuroscience* 2022:14.
- Arvesen E. Automatic Classification of Alzheimer's Disease from Structural MRI. 2015.
- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7(3):263–9.
- de Mendonça LJC, Ferrari RJ. Alzheimer's Disease Neuroimaging I. Alzheimer's disease classification based on graph kernel SVMs constructed with 3D texture features extracted from MR images. *Expert Systems with Applications* 2023;211:118633.
- Xu X, Lin L, Sun S, Wu S. A review of the application of three-dimensional convolutional neural networks for the diagnosis of Alzheimer's disease using neuroimaging. *Rev Neurosci* 2023;(0).
- Leandrou S, Lamnisos D, Bougias H, et al. A cross-sectional study of explainable machine learning in Alzheimer's disease: diagnostic classification using MR radiomic features. *Front Aging Neurosci* 2023;15:1149871.
- Du Y, Yu J, Liu M, et al. The relationship between depressive symptoms and cognitive function in Alzheimer's disease: the mediating effect of amygdala functional connectivity and radiomic features. *J Affect Disord* 2023;330:101–9.
- Jack Jr CR, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008;27(4):685–91.
- ADNI | MRI Pre-Processing 2023 [Available from: <https://adni.loni.usc.edu/data-samples/adni-data/neuroimaging/mri/mri-pre-processing/>].
- Fischl B. FreeSurfer Neuroimage 2012;62(2):774–81.
- Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 2011;56(3):907–22.
- Khatri U, Kwon G-R. Alzheimer's Disease diagnosis and biomarker analysis using resting-state functional MRI functional brain network with multi-measures features and hippocampal subfield and amygdala volume of structural MRI. *Frontiers in aging. Neuroscience* 2022:14.
- Llano DA, Bundela S, Mudar RA, Devanarayan V, Alzheimer's Disease Neuroimaging I. A multivariate predictive modeling approach reveals a novel CSF peptide signature for both Alzheimer's Disease state classification and for predicting future disease progression. *PLoS One* 2017;12(8):e0182098.
- Popuri K, Ma D, Wang L, Beg MF. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. *Hum Brain Mapp* 2020;41(14):4127–47.
- Chen X, Tang M, Liu A, Wei X. Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. *Annals of Translational Medicine* 2022;10(14).
- Birkenbihl C, Salimi Y, Domingo-Fernández D, Lovestone S, AddNeuroMed consortium, Fröhlich H, Hofmann-Apitius M. Japanese Alzheimer's Disease Neuroimaging initiative; and the Alzheimer's Disease Neuroimaging initiative. Evaluating the Alzheimer's disease data landscape. *Alzheimers Dement (N Y)* 2020 Dec 16;6(1).
- Zamani J, Sadr A, Javadi AH. Diagnosis of early mild cognitive impairment using a multiobjective optimization algorithm based on T1-MRI data. *Sci Rep* 2022;12(1):1020.

- 24 Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical* 2019;21:101645.
- 25 Feng F, Wang P, Zhao K, et al. Radiomic features of hippocampal subregions in Alzheimer's disease and amnesic mild cognitive impairment. *Front Aging Neurosci* 2018;10:290.
- 26 Beheshti I, Demirel H, Farokhian F, Yang C, Matsuda H. Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error. *Comput Methods Programs Biomed* 2016;137:177–93.