



OPEN Explainable early detection of Alzheimer's disease using ROIs and an ensemble of 138 3D vision transformers

Lyes Saad Saoud¹ & Hasan AlMarzouqi²✉

Early detection and accurate diagnosis of brain morphological abnormalities are essential for the effective management and treatment of Alzheimer's disease (AD) and mild cognitive impairment (MCI). Structural magnetic resonance imaging (MRI) is a powerful support tool to aid in disease diagnosis and prediction. In this research study, we present an innovative approach to predict Alzheimer's disease (AD) and mild cognitive impairment (MCI) using MRI data, which integrates regional interest (ROI)-based methodology and deep learning within a comprehensible framework. The proposed method involves dividing the brain into 138 predetermined sections based on anatomical information. Next, we apply three-dimensional vision transformers (3D-ViTs) to each ROI individually, harnessing the power of deep learning. To improve prediction accuracy, we employ a deep belief network (DBN) as an ensemble learning model. Evaluating our approach on the baseline structural MRI dataset obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, and comparing it against five other competing models, we demonstrate its performance across four binary classification tasks and a three-class classification test (AD vs MCI vs CN (Cognitively Normal)). The proposed system outperforms existing models and provides interpretable insights into the brain regions that significantly contribute to solving each classification problem. Our findings align with the existing body of literature and hold promise for guiding future research directions in this domain.

Keywords Alzheimer's disease, MRI, Three-dimensional vision transformers (3D-ViTs), Mild cognitive impairment, Region of interest (ROI)

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that is the most common cause of dementia. Early and accurate diagnosis is crucial for patient care and the development of new treatments. Changes in the brain caused by Alzheimer's disease are known to begin 20 or more years before the disease appears, highlighting the importance of early detection for better patient management and treatment outcomes¹⁻³. Early detection allows for timely interventions that can delay the progression of symptoms, improving the quality of life for patients and easing the burden on caregivers. However, the standard diagnosis process, which is primarily based on neuropsychological assessments, is often subjective and less reproducible, necessitating the development of more objective and reliable methods.

Recent advances in neuroimaging and machine learning have opened new avenues for diagnosing and predicting the progression of Alzheimer's disease. Among the various neuroimaging techniques, Magnetic Resonance Imaging (MRI) plays a critical role in detecting structural and functional brain changes associated with AD. In particular, deep learning (DL) techniques have shown promise in automating the analysis of MRI data to aid in the diagnosis and classification of Alzheimer's disease.

Deep learning approaches have gained considerable traction in medical image analysis, particularly in the classification of AD using MRI data⁴. DL techniques are capable of directly learning complex and high-level features from MRI datasets, enabling accurate and timely diagnosis, which is crucial for appropriate patient treatment and care⁴. Studies have demonstrated the potential of DL in the automatic diagnosis of AD and mild cognitive impairment (MCI) using T1-weighted MRI images, achieving excellent performance in the accurate diagnosis of AD and MCI⁵. The use of Convolutional Neural Networks (CNNs) has been prevalent in classifying different stages of AD, such as MCI and cognitive normal (CN), achieving high accuracy rates for classification^{6,7}.

¹Department of Mechanical Engineering of Khalifa University, Abu Dhabi PO Box 127788, UAE. ²Department of Electrical Engineering of Khalifa University, Abu Dhabi PO Box 127788, UAE. ✉email: hasan.almrzouqi@ku.ac.ae

Despite these advances, challenges remain in implementing deep learning for AD diagnosis, including the need for large, annotated datasets, model interpretability, and integration into clinical workflows⁸. Deep learning techniques contribute to the classification of MCI, AD, and controls by automatically learning hierarchical features from MRI data, enabling the identification of intricate patterns associated with different AD stages⁸. Deep CNNs have been adopted to distinguish different stages of MCI from normal controls, stable MCI, and converted MCI, achieving high accuracy in discrimination among the groups and in the prediction of conversion risk from MCI to AD⁹. However, challenges such as the need for large datasets, computing resources, and careful parameter setting to prevent overfitting or underfitting have been identified in the implementation of deep learning for AD classification using MRI scans¹⁰.

To address these limitations, we propose the integration of 3D Vision Transformers with Region of Interest (ROI)-based analysis for AD diagnosis. Vision Transformers (ViT) have demonstrated success in encoding long-range relationships in images, making them potentially valuable in AD diagnosis. Studies have utilized ROI analysis for AD diagnosis, showing that 3D texture features from ROIs could detect subtle texture differences between tissues in AD patients and normal controls¹¹. These findings suggest that the integration of 3D Vision Transformers with ROI-based analysis could lead to more accurate and reliable diagnostic tools for Alzheimer's disease.

In this study, we propose a novel strategy that combines the advantages of ROI-based methods, Vision Transformers, and ensemble methods for the early prediction of AD. We segment MRI brain scans into 138 volumes containing specific subregions and then train the complex relationship between those ROIs and AD classes using the 3D ViTs separately. A deep belief network (DBN) is then applied as an ensemble meta-learning model that combines the ViT predictions and generates the class labels. Using brain ROIs and ensemble techniques helps avoid overfitting and increases the robustness of the model while keeping the dataset size and computational resources reasonable.

Knowing which parts of the brain are affected and how various areas of the brain relate to symptomatic information is clinically significant. Motivated by such interpretability demands, we present an ensemble 3D ViT-based model that can be used to infer regional abnormalities and provide insight into the complex relationship between these abnormalities and disease progression. Compared to other recently proposed models for AD classifiers, our framework is technically and methodologically different.

The main contributions of this study are as follows.

1. We propose a novel integration of 3D Vision Transformers with a Region of Interest (ROI)-based analysis approach, specifically designed for Alzheimer's disease classification. This integration allows for the extraction and analysis of fine-grained features from specific brain regions, enhancing the model's interpretability and diagnostic precision.
2. We introduce a new ensemble learning framework that leverages Deep Belief Networks (DBNs) to aggregate predictions from multiple 3D Vision Transformer models. This framework not only improves classification accuracy but also enhances the robustness and generalization capability of the model across different datasets.
3. Our method enhances explainability by segmenting MRI brain scans into 138 distinct volumes corresponding to clinically relevant subregions. This segmentation enables a detailed examination of regional brain abnormalities, offering insights into how specific brain areas contribute to Alzheimer's disease progression.
4. Through extensive experiments, we demonstrate that our model not only achieves state-of-the-art performance in Alzheimer's disease classification but also provides a significant improvement in explainability and interpretability, making it a valuable tool for clinicians. The proposed framework offers insights into the complex relationships between different brain regions and Alzheimer's disease, thereby advancing both the scientific understanding and practical application of deep learning in medical imaging.

Dataset and preprocessing

The AD dataset used in this work is from¹². This dataset was constructed using samples from the Alzheimer's disease neuroimaging project database (ADNI) (<http://adni.loni.usc.edu/>). A public-private partnership led to the creation of the ADNI database in 2003. The main goal of ADNI has been to determine whether serial magnetic resonance imaging, PET, other biological markers, clinical, and neurophysical evaluations can be used in conjunction to follow the progression of early Alzheimer's disease and mild cognitive impairment.

We evaluated 532 brain volumes, which are baseline MRI scans labeled according to change in diagnosis over two years. The baseline dataset includes 168 cognitively normal people (CN), 247 who have mild cognitive impairment (MCI), and 117 who have Alzheimer's disease (AD). Within the MCI group, participants were further classified as 107 with stable MCI (sMCI) and 140 with progressive MCI (pMCI) based on their disease development over 24 months. A longitudinal analysis problem was solved by training the network to differentiate between sMCI and pMCI scans. Table 1 presents some of their clinical features, including median age, gender, and

Feature	CN	sMCI	pMCI	AD
Median Age (range)	74.2 [59.8; 89.6]	74.4 [55.9; 91.4]	74.3 [48.1; 88.3]	75.8 [55.1; 91.4]
Gender (M/F)	84/84	63/44	82/58	64/53
ApoE4 carriers (non-carriers)	293/101/9	93/60/13	57/91/29	104/155/62

Table 1. Clinical and Biological Features.

ApoE4 allele distribution¹². AD subjects had to meet the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association criteria (NINCDS-ADRDA) for probable AD¹³. All MCI cases were amnesic¹⁴.

The images needed further processing to extract the individual ROIs and correct for non-uniformities.¹⁵.

1. **Brain Extraction:** Initially, we applied a brain extraction procedure using PinCrani, which is a versatile method of accurately labeling the adult brain on 3D T1-weighted magnetic resonance head images¹⁶. This step removed non-brain tissues from the MRI scans, ensuring that only the brain regions were analyzed.
2. **Registration to a Standard Template:** To ensure consistent ROI extraction across subjects, all MRI images were registered to a standard template using the MALPEM4D pipeline. This pipeline incorporates symmetric affine intra-subject registration and corrects for differential bias between intra-subject acquisitions using unweighted differential bias correction¹².
3. **ROI Identification:** MALPEM4D¹⁷ was used to create structural segments for 138 brain regions. Each region in the mask volumes was assigned a distinct voxel value corresponding to the region number. To evaluate our model's performance, we divided the ADNI dataset into training and validation sets using an 80/20 split. All reported results reflect the model's performance on the validation set.

The proposed approach

In this section, we present our novel approach for predicting and describing regional anomalies to determine the clinical state of a brain along the spectrum of progression of AD. As illustrated in Fig. 1, we propose the ROI-3DViT-DBN ensemble model for the early prediction of AD. We begin with an extension of ViT¹⁸ that models pairwise interactions between all 3D spatial MRI-ROI tokens and then apply DBN to fine-tune the classification performance. In what follows, we will describe the different components of the proposed system.

Transformers expect tokens as input, and these are encodings, typically equal in size, of the input signal. The proposed system uses a 3D Embed-to-Tokens block¹⁹, which uses tubelet embeddings to extract and linearly embed non-overlapping tubelets that span the 3D MRI input volume (Fig. 2).

The input of each transformer in Fig. 3 receives the sequence of reshaped input tokens $z \in \mathbb{R}^N$, which is obtained using

$$z = [\mathbf{E}x_1 \ \mathbf{E}x_2, \dots, \ \mathbf{E}x_N] + \mathbf{p} \quad (1)$$

where \mathbf{E} is the projection operation and is obtained using a learned 3D convolution filter followed by flattening to a 1D vector. $\mathbf{p} \in \mathbb{R}^N$ is a positional embedding index that ranges from 1 to N . Next, a sequence of $L = 8$ encoder transformer layers is used to process the tokens. Layer normalization (LN), Multi-Headed Self-Attention (MSA)¹⁸, and MLP blocks are incorporated into each layer (l) as follows:

$$y^l = \text{MSA} (\text{LN} (z^l)) + z^l \quad (2)$$

$$z^{l+1} = \text{MLP} (\text{GELU} (\text{MLP} (\text{LN} (y^l)))) + y^l \quad (3)$$

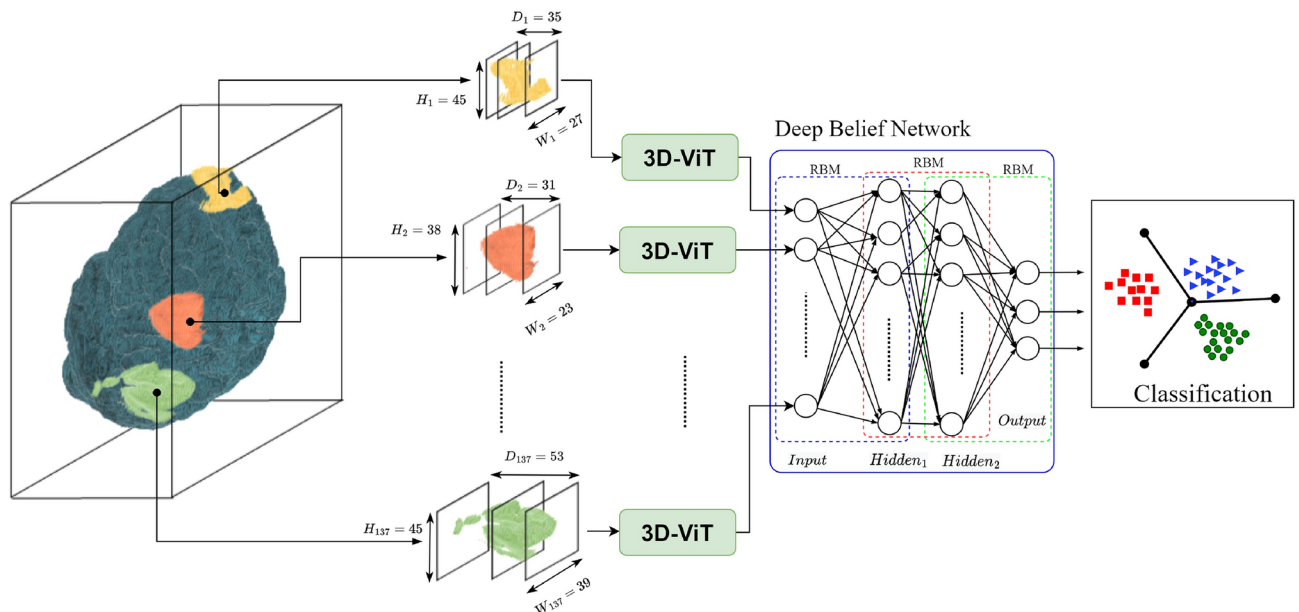


Fig. 1. Overall approach for detecting regional anomalies and forecasting clinical statuses throughout Alzheimer's disease progression.

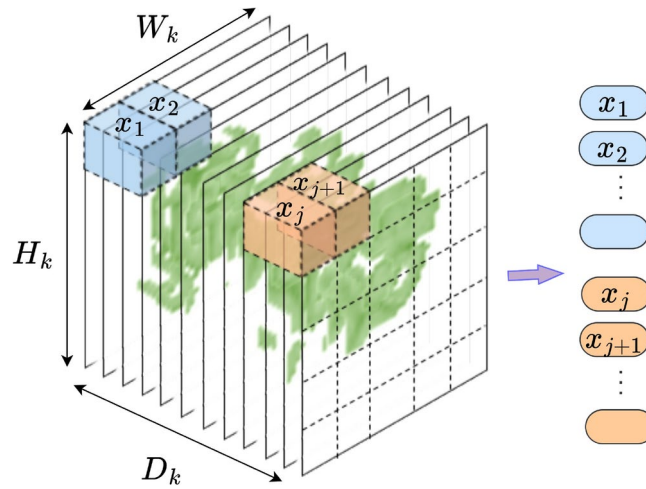


Fig. 2. The tubelet embedding used to encode the ROIs.

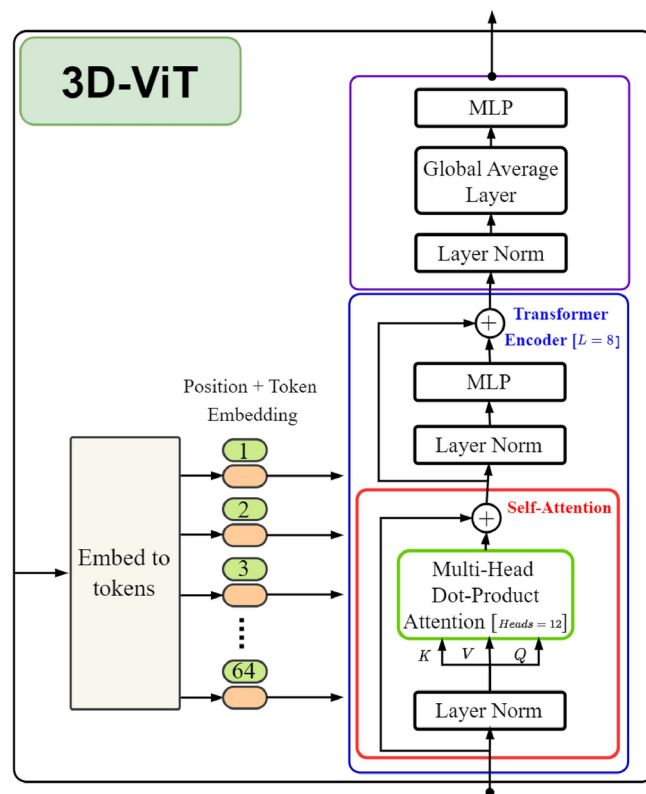


Fig. 3. The proposed 3D-ViT for Alzheimer’s disease detection.

Where GELU is a nonlinear activation function²⁰ that separates the two linear projection layers. The encoded output is then normalized, sent to a global average pooling layer, and another MLP. Finally, a softmax function is used to compute the network output and a sparse categorical cross-entropy loss function is used to train the network.

The predictions from the ROI 3DViT class outputs were combined using ensemble learning to create a forecaster for the entire brain. We experimented with several ensemble learning methods and deep belief networks, which are made up of multiple restricted Boltzmann machines (RBM)²¹, and consistently achieved the best classification performance.

To prevent overfitting during the training of robust 3DViT models and mitigate the impact of class imbalance, we incorporated a dropout rate of 0.1 in the multi-head attention layers. This allowed us to randomly “ignore”

certain features during training. Using spline interpolation, the ROIs were rescaled to a size of $28 \times 28 \times 28$ voxels to allow the use of the same network architecture for the different regions. A kernel and stride size of $8 \times 8 \times 8$ were used in the projection operation E. Due to the class imbalance present in the MRI dataset (e.g., more subjects with Mild Cognitive Impairment (MCI) than with Alzheimer's disease (AD)), we employed accuracy, F1 score, AUROC, and Cohen's kappa statistic as evaluation metrics.

The schematic representation of our proposed approach is illustrated in Figure 1. In the proposed system, we begin by preprocessing the MRI input, as described in Section 2. Specifically, we partition these MRI images into regions of interest (ROIs) using the MALPEM4D anatomical template¹⁷. Next, we train a 3D volumetric image transformer (3DViT) for each of these ROIs and for each classification problem under investigation. The classification tasks addressed in this article include the discrimination of AD vs. CN, MCI vs. CN, AD vs. MCI, and the multiclass classification problem AD vs. CN vs. MCI.

To enhance the robustness and precision of our classification framework, we leverage an ensemble learning approach to combine the predictions from individual 3DViT models. Each ViT model outputs a feature vector representing a Region of Interest (ROI) in the brain. This feature vector then serves as the input to an ensemble model, which acts as a classifier, translating the ROI features into specific disease categories (e.g., AD vs. CN). The field of ensemble learning offers a variety of models well-suited for this task. Common choices include:

Logistic Regression: A statistical method used to predict the probability of a binary outcome based on one or more features.

Decision Trees: Supervised learning algorithms used for classification and regression. They recursively partition data based on features until subsets are pure or a stopping criterion is met.

Bagging or Bootstrap Aggregating: An ensemble method that combines predictions from multiple decision trees to improve accuracy and robustness.

Random Forests: An ensemble method that combine multiple decision trees, each trained on a random subset of data and features, to improve accuracy and robustness.

AdaBoost: An ensemble method that combines weak learners into a strong learner by iteratively training each model on data, adjusting instance weights based on previous model performance.

Gradient Tree Boosting: Ensemble methods that combine multiple decision trees, each trained on the residuals of the previous model, to improve accuracy and robustness.

Histogram-Based Gradient Boosting: A variant of gradient tree boosting that uses histograms for data distribution approximation, making it more efficient and scalable.

Support Vector Machines (SVMs): A supervised learning algorithm used for classification and regression. They find the hyperplane that maximizes the margin between classes.

Multilayer Perceptrons (MLPs): Feedforward neural networks used for classification and regression. They process data through multiple layers of artificial neurons, each applying a non-linear transformation.

Hard Voting: An ensemble method that combine predictions from multiple models to make a final prediction. Hard voting chooses the class with the most votes.

Soft Voting: An ensemble method that combines predictions from multiple models to make a final prediction. In soft voting, we choose the class with the highest predicted probability.

Results

In this section, we present the predictions of the proposed system and compare our results with several proposed methods in the literature. We also use the proposed framework to show the role that each ROI plays in obtaining network prediction. We evaluated three binary classification problems (AD vs. CN, MCI vs. CN, AD vs. MCI, and pMCI vs. sMCI) and one three-class classification problem (AD vs. MCI vs. CN).

All networks were trained for a maximum of 1000 epochs, with early stopping conditions that stopped training after 50 epochs if the loss function did not improve. The proposed architectures have been implemented in Keras with a Tensorflow backend. Before going to the classification experiments and the analysis of ROI regions, we present experiments comparing several ensemble models and illustrating the role played by DBNs in the proposed system.

Ensemble model selection: In this study, we applied several ensemble methods to the four AD tasks investigated to select the best performance method. The algorithms examined are logistic regression, bagging meta-estimator, Forests of randomized trees, AdaBoost, Gradient Tree Boosting, Histogram-Based Gradient Boosting, Support Vector Machines, Support vector machines (SVMs), Gaussian Naive Bayes, Decision Trees, Multilayer Perceptron, Hard Voting, Soft voting, and deep belief network.

The results of this comparison are presented in Tables 2 to 5. As can be seen, the DBN consistently outperformed all other ensemble models across all four AD classification tasks (AD vs. CN, MCI vs. CN, AD vs. MCI, pMCI vs. sMCI). This finding establishes the DBN as a promising approach for achieving superior performance in Alzheimer's disease prediction using MRI data.

The results achieved in the four prediction tasks are shown in Table 2 to Table 5. One can see that the deep belief network consistently outperforms all other ensemble models in all AD classification tasks. Therefore, we used this network as part of our system and in the remaining experiments in this paper.

AD/MCI Disease Prediction: We conducted several studies to evaluate the efficacy and validity of our approach by comparing it with current methods in the literature. We specifically investigated the four following strategies.

1. rDNN³: A method for diagnosing Alzheimer's disease (AD) or moderate cognitive impairment (MCI) using magnetic resonance imaging that systematically incorporates voxel-based, region-based, and patch-based methodologies into a cohesive framework. They regarded anatomical forms of areas as atypical patches, compared to other existing approaches that employ cubical or rectangular shapes.

Model	Accuracy	AUROC	F1-Score	Kappa
Logistic Regression	0.76	0.72	0.74	0.47
Bagging Classifier	0.79	0.74	0.76	0.51
Random Forest Classifier	0.76	0.72	0.74	0.47
Adaboost Classifier	0.64	0.59	0.61	0.19
Gradient Tree Boosting	0.71	0.65	0.66	0.33
Histogram-Based Gradient Boosting	0.64	0.63	0.64	0.27
Support Vector Machines	0.76	0.72	0.74	0.47
Decision Trees Classifier	0.62	0.60	0.62	0.21
Multi-layer Perceptron	0.79	0.74	0.77	0.52
Hard Voting	0.76	0.72	0.74	0.47
Soft Voting	0.86	0.83	0.85	0.69
Deep Belief Network	0.90	0.91	0.90	0.81

Table 2. The performance of different ensemble methods in the prediction task AD vs. CN.

Model	Accuracy	AUROC	F1-Score	Kappa
Logistic Regression	0.72	0.56	0.64	0.15
Bagging Classifier	0.72	0.56	0.64	0.15
Random Forest Classifier	0.65	0.56	0.55	0.14
Adaboost Classifier	0.72	0.56	0.64	0.15
Gradient Tree Boosting	0.70	0.53	0.60	0.08
Histogram-Based Gradient Boosting	0.72	0.65	0.71	0.32
Support Vector Machines	0.72	0.56	0.64	0.15
Decision Trees Classifier	0.70	0.56	0.64	0.15
Multi-layer Perceptron	0.74	0.59	0.67	0.23
Hard Voting	0.72	0.56	0.64	0.15
Soft Voting	0.74	0.59	0.67	0.23
Deep Belief Network	0.74	0.60	0.69	0.25

Table 3. The performance of different ensemble methods in the prediction task AD vs. MCI.

Model	Accuracy	AUROC	F1-Score	Kappa
Logistic Regression	0.66	0.58	0.58	0.19
Bagging Classifier	0.63	0.54	0.52	0.09
Random Forest Classifier	0.65	0.57	0.58	0.17
Adaboost Classifier	0.68	0.62	0.64	0.26
Gradient Tree Boosting	0.61	0.53	0.51	0.06
Histogram-Based Gradient Boosting	0.66	0.60	0.62	0.22
Support Vector Machines	0.66	0.58	0.58	0.19
Decision Trees Classifier	0.65	0.59	0.60	0.19
Multi-layer Perceptron	0.74	0.68	0.70	0.40
Hard Voting	0.74	0.69	0.71	0.41
Soft Voting	0.69	0.62	0.63	0.27
Deep Belief Network	0.84	0.81	0.88	0.65

Table 4. The performance of different ensemble methods in the prediction task CN vs. MCI.

2. Full 3D CNN²²: Based on structural MRI images of the brain, they trained a full cubic 3D CNN to identify Alzheimer's disease. Then, they used four distinct gradient-based and occlusion-based visualization approaches to highlight key areas in the input picture to illustrate the network's classification conclusions.
3. Plain and residual CNN²³: They examined two distinct 3D convolutional network designs for brain MRI classification: simple (VoxCNN) and residual convolutional neural networks (ResNet).
4. Sparse autoencoders and 3D CNN²⁴: The authors used sparse autoencoders and 3D convolutional neural networks to predict the disease based on an MRI scan of the brain. Their findings showed that a 3D technique can capture local 3D patterns, which can improve classification performance, though by a slight margin. Ta-

Model	Accuracy	AUROC	F1-Score	Kappa
Logistic Regression	0.57	0.59	0.48	0.22
Bagging Classifier	0.57	0.59	0.48	0.23
Random Forest Classifier	0.54	0.58	0.46	0.19
Adaboost Classifier	0.58	0.62	0.51	0.29
Gradient Tree Boosting	0.49	0.52	0.35	0.06
Histogram-Based Gradient Boosting	0.51	0.56	0.44	0.14
Support Vector Machines	0.61	0.63	0.52	0.32
Decision Trees Classifier	0.52	0.59	0.49	0.19
Multi-layer Perceptron	0.54	0.60	0.48	0.24
Hard Voting	0.59	0.61	0.51	0.27
Soft Voting	0.59	0.62	0.52	0.28
Deep Belief Network	0.60	0.66	0.53	0.32

Table 5. The performance of different ensemble methods in the prediction task AD vs. CN vs. MCI.

Task	Model	Accuracy	AUROC	F1-Score
AD vs. CN	Lee et al. ³	0.86	0.82	0.85
	Rieke et al. ²²	0.72	0.67	0.69
	Khvostikov et al. ²⁵	0.88	0.89	0.89
	Korolev et al. ²³	0.77	0.77	0.77
	Payan and Montana ²⁴	0.77	0.77	0.77
	Proposed*	0.88	0.86	0.90
	Proposed	0.90	0.91	0.90
AD vs. MCI	Lee et al. ³	0.70	0.53	0.60
	Rieke et al. ²²	0.71	0.53	0.60
	Khvostikov et al. ²⁵	0.66	0.49	0.55
	Korolev et al. ²³	0.72	0.59	0.68
	Payan and Montana ²⁴	0.71	0.56	0.65
	Proposed*	0.74	0.58	0.68
	Proposed	0.74	0.60	0.69
CN vs. MCI	Lee et al. ³	0.73	0.66	0.68
	Rieke et al. ²²	0.65	0.59	0.58
	Khvostikov et al. ²⁵	0.79	0.78	0.83
	Korolev et al. ²³	0.65	0.60	0.61
	Payan and Montana ²⁴	0.65	0.64	0.64
	Proposed*	0.71	0.64	0.80
	Proposed	0.84	0.81	0.88
pMCI vs. sMCI	Lee et al. ³	0.88	0.87	0.87
	Rieke et al. ²²	0.80	0.79	0.80
	Khvostikov et al. ²⁵	0.92	0.91	0.90
	Korolev et al. ²³	0.74	0.75	0.74
	Payan and Montana ²⁴	0.77	0.76	0.77
	Proposed*	0.89	0.88	0.86
	Proposed	0.94	0.95	0.94

Table 6. Performance comparison with four existing binary prediction tasks. Proposed* utilizes demographic and clinical features: age, gender, marital status, education, and APOE status.

bles 6 and 7 show the findings of the four competing approaches. The binary prediction, i.e., tasks with two classes, is presented in Table 6, and the generalization to three classes is given in Table 7. The proposed model outperforms other methods in all tasks. The proposed model achieved an accuracy of 90% and an AUROC of 91% in the AD vs. CN task. In the CN versus MCI task, our approach achieved an accuracy of 84% and an AUROC of 81%. In the AD vs. MCI task, the proposed approach achieved an accuracy of 74% and an AUROC of 60%. For the patients who convert from stable MCI (sMCI) to progressive MCI (pMCI) over 24 months, our model achieves an AUROC of 91%. Finally, the proposed model achieves an accuracy of 60% and an AUROC of 66% in the three-class prediction problem AD vs. CN vs. MCI.

Task	Model	Accuracy	AUROC	F1-Score
AD vs. CN vs. MCI	Lee et al. ³	0.57	0.65	0.52
	Rieke et al. ²²	0.48	0.51	0.33
	Khvostikov et al. ²⁵	0.52	0.56	0.41
	Korolev et al. ²³	0.53	0.61	0.48
	Payan and Montana ²⁴	0.52	0.57	0.44
	Proposed*	0.65	0.70	0.56
	Proposed	0.60	0.66	0.53

Table 7. Performance comparison with four existing three-class classification tasks. Proposed* utilizes demographic and clinical features: age, gender, marital status, education, and APOE status.

In Tables 6 and 7, we incorporated additional clinical and demographic features: age, gender, marital status, education, and APOE status. However, for tasks other than AD vs. CN vs. MCI, the results indicate that the inclusion of these features did not lead to improved performance.

Next, we conducted experiments where we used the 138 pre-trained 3D ViTs to solve the prediction problems we studied in this paper. Fig. 4, Fig. 6, and Fig. 9 show bar plots describing the performance of the five ROIs that achieve the highest AUROC scores in the three AD prediction problems investigated in this paper. Visualizations of these regions on MRI scans are shown in Fig. 5, Fig. 7, Fig. 8, and Fig. 10. These ROIs represent crucial areas of the brain highlighted by our model predictions, shedding light on the neural regions essential to distinguish between different stages of the disease.

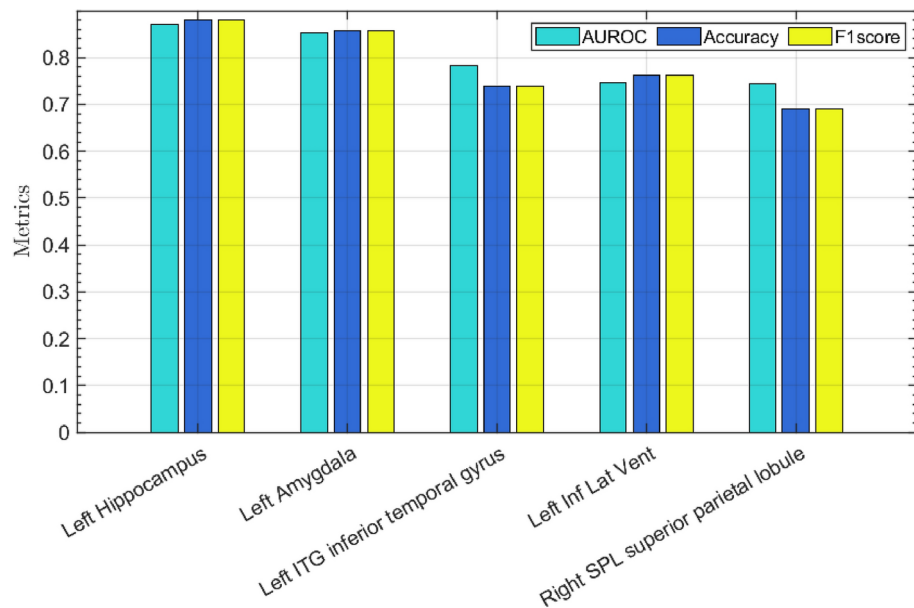


Fig. 4. The top five ROIs in terms of AUROC scores for the task AD vs. CN.

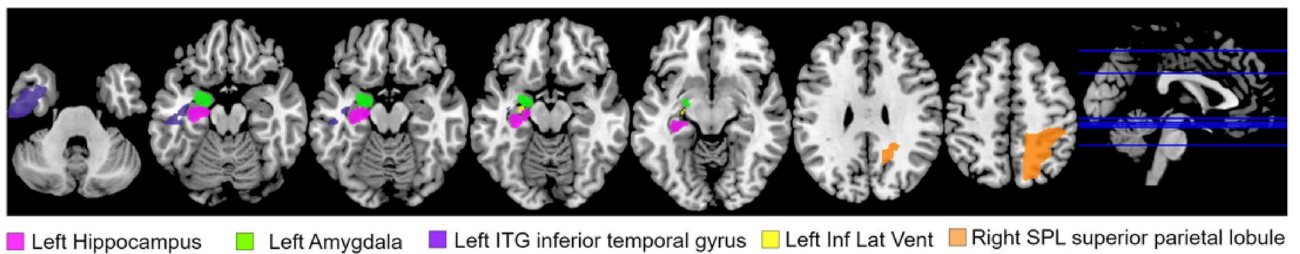


Fig. 5. Visualization of the top five ROIs in terms of AUROC in the AD vs. CN prediction task.

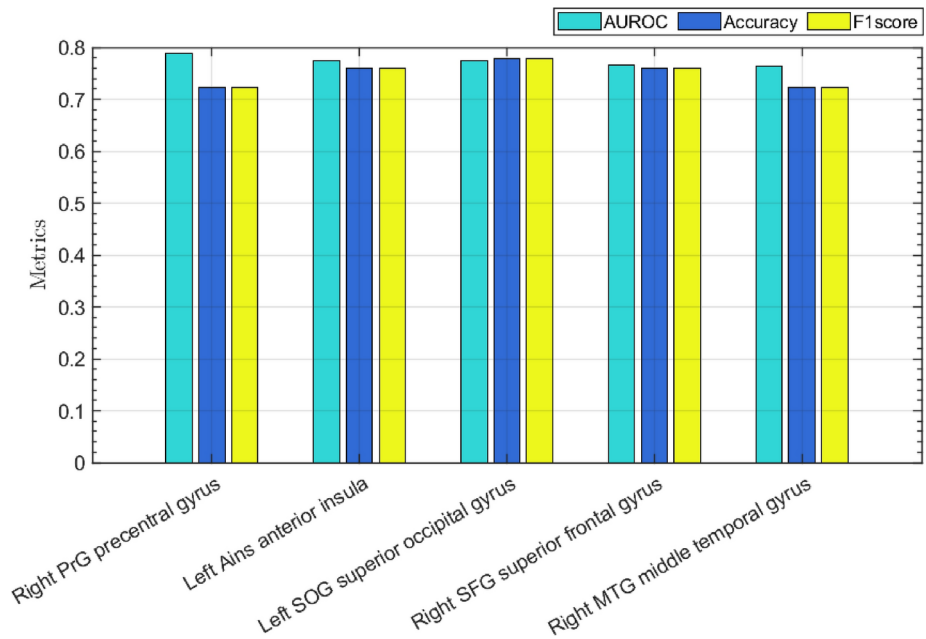


Fig. 6. The top five ROIs in terms of AUROC scores for the task AD vs. MCI.

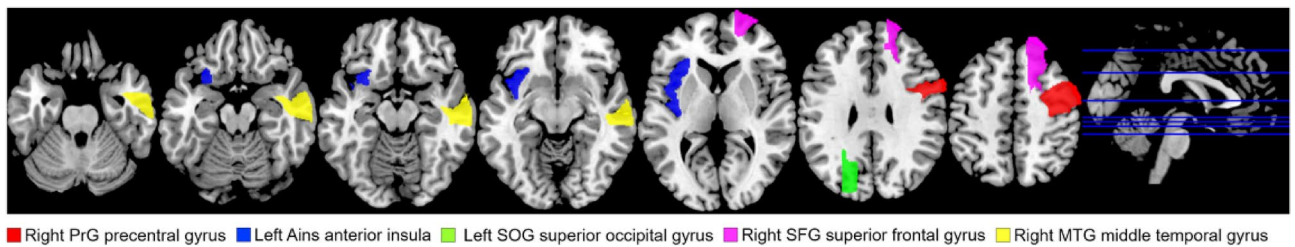


Fig. 7. Visualization of the top five ROIs in terms of AUROC in the AD vs. MCI prediction task.

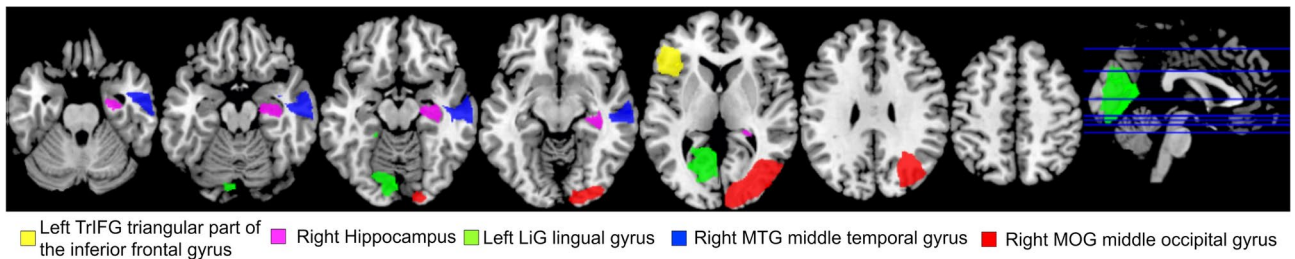


Fig. 8. Visualization of the top five ROIs in terms of AUROC in the CN vs. MCI prediction task.

The supplementary material includes a series of comprehensive lists, each detailing the 138 brain regions ordered by their AUROC scores for the three different diagnostic tasks. This detailed breakdown offers insights that can be of significance for both clinical applications and neuroscience research.

Discussions

Our study develops a new method for the prediction of Alzheimer’s disease (AD) using magnetic resonance imaging (MRI) data. We compared our proposed method with several established approaches from the literature and evaluated its performance across various classification tasks.

A key contribution of our work lies in the comprehensive evaluation of various ensemble methods for AD prediction. We compared a range of models, including logistic regression, bagging meta-estimators, random forests, AdaBoost, gradient tree boosting, histogram-based gradient boosting, support vector machines (SVMs),

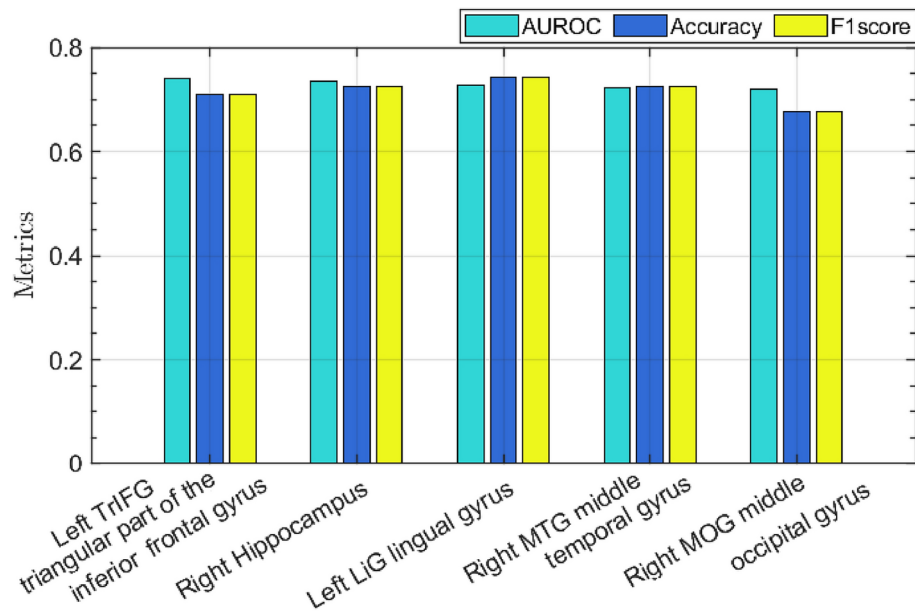


Fig. 9. The top five ROIs in terms of AUROC scores for the task CN vs. MCI.

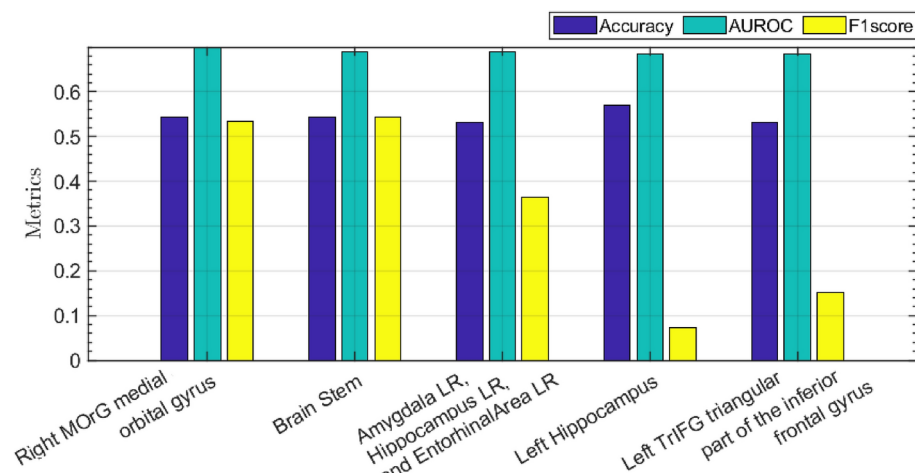


Fig. 10. Examples of the five high predicted ROIs with the ensemble 3D-Vit-DBN model for the task of pMCI vs. sMCI.

decision trees, multilayer perceptrons (MLPs), Hard Voting, and Soft Voting, against a deep belief network (DBN). Our results consistently demonstrated the DBN's superiority across all AD classification tasks.

Additionally, a series of comprehensive lists, each detailing the 138 brain regions ordered by their AUROC scores for the three different diagnostic tasks, are presented in the supplementary material. We want to highlight that the top two regions in the AD versus CN task (Supplementary Fig. 1, Supplementary Fig. 4) are the hippocampus and the amygdala. There is abundant academic literature showing the importance of these two regions in the early prediction of Alzheimer's disease (see, for example,²⁶ and²⁷).

Further insights into the role of individual ROIs can be obtained by comparing the figures in the supplementary material and noting how the relevant importance of brain regions differs between tasks. For example, the brain stem region plays an important role in solving the CN vs. MCI and pMCI vs. sMCI problems, as it achieves high AUROC scores compared to other brain ROIs (Supplementary Fig. 3 and Supplementary Fig. 4). However, in the AD vs. CN problem (Supplementary Fig. 1), the brain stem does not appear to play a similar role, as many other brain regions surpass it in AUROC.

We have also noted that ensemble methods, such as the DBN used in our study, often exhibit superior performance. This advantage can be attributed to the enhanced robustness and generalization capabilities provided by the ensemble learning strategy, particularly when focusing on individual Regions of Interest (ROIs). By concentrating on individual ROIs, the method helps avoid overfitting and reduces the impact of confounding

factors that could negatively affect the model's generalization ability. Furthermore, this approach allows for more accurate extraction of relevant features.

In addition, the role of specific ROIs in distinguishing between MCI, AD, and controls offers valuable insights into the clinical relevance of these regions. By understanding which ROIs contribute most to classification performance, we can better interpret the underlying neurobiological differences associated with these conditions and improve the clinical applicability of our model. In what follows, we provide examples of how several of our ROI results relate to the existing literature on Alzheimer's disease.

Analysis of ROI-based diagnostic results

Differential involvement of brain regions

The ROIs identified for different classifications (AD vs. CN, AD vs. MCI, CN vs. MCI) reflect varying degrees of neurodegeneration and functional impairment associated with each condition. For instance:

- **AD vs. CN:** The left hippocampus and left amygdala are prominent in this classification. These regions are critical for memory formation and emotional processing, and they are often among the first to show pathological changes in AD, such as tau and amyloid-beta accumulation. Their early involvement can lead to significant differences in classification performance, as these areas are susceptible to the neurodegenerative processes of AD²⁸.
- **AD vs. MCI:** The right precentral gyrus and left anterior insula are highlighted here. The precentral gyrus is involved in motor function^{29,30}, while the anterior insula plays a role in emotional awareness and cognitive processing³¹. The transition from MCI to AD often involves not just memory deficits but also changes in executive function and motor skills. This may explain why these regions are more predictive in distinguishing between these two groups³².
- **CN vs. MCI:** The left triangular part of the inferior frontal gyrus and the right hippocampus are significant here. The inferior frontal gyrus is associated with higher-order cognitive functions, including language and decision-making^{33,34}. Its involvement suggests that cognitive decline in MCI may be detected through changes in executive function and language processing, which are less affected in CN individuals^{35,36}.

Pathological correlates

The classification performance of these ROIs can be attributed to their association with specific pathological features of Alzheimer's disease. Notably, regions such as the hippocampus and entorhinal area are recognized as early sites of tau pathology, a hallmark of AD. The presence of neurofibrillary tangles in these areas has been shown to significantly impair memory and cognitive function, making them critical for differentiating AD from CN and MCI³⁷.

Clinical implications

Understanding the predictive power of specific ROIs for Alzheimer's disease classifications can significantly enhance clinical practice. By identifying key regions, clinicians can not only develop focused therapeutic approaches aimed at preserving function in these critical areas³⁸ but also improve the accuracy of early diagnosis. This, in turn, can lead to more effective management of Alzheimer's disease and mild cognitive impairment³⁹.

In summary, the differences in AUROC scores for various ROIs across the classification tasks can be attributed to the distinct roles these regions play in cognitive function and their susceptibility to pathological changes. Discussing these aspects not only provides insight into the validity of the results but also enhances the clinical relevance of the findings.

Conclusions

In this paper, we introduced an innovative approach for early detection of Alzheimer's disease (AD) and mild cognitive impairment (MCI) using MRI data. Our method, which combines regional interest (ROI)-based analysis and 3D vision transformers, outperformed existing models in accuracy and interpretability. By dividing the brain into 138 predetermined sections and leveraging deep belief networks, we achieved superior performance on several classification tasks. Importantly, our model provides insights into the key brain regions that drive classification decisions. Our findings hold promise in improving the diagnosis of AD and MCI and in guiding treatment strategies. Moving forward, this work provides a foundation for further research in this area that ultimately advances diagnostic tools and treatments for these conditions and improves the lives of affected individuals.

Data availability

The dataset used in this study MALPEM-ADNI can be downloaded from <https://doi.gin.g-node.org/10.12751/g-node.aa605a/>

Received: 2 January 2024; Accepted: 14 October 2024

Published online: 12 November 2024

References

1. Reiman, E. M. et al. Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant alzheimer's disease in the presenilin 1 e280a kindred: a case-control study. *The Lancet Neurology* **11**, 1048–1056 (2012).
2. Villemagne, V. L. et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic alzheimer's disease: a prospective cohort study. *The Lancet Neurology* **12**, 357–367 (2013).

3. Lee, E., Choi, J.-S., Kim, M. & Suk, H.-I. Toward an interpretable alzheimer's disease diagnostic model with regional abnormality representation via deep learning. *NeuroImage* **202**, 116113 (2019).
4. Ifiti, S. A., Ahmed, R., Rahman, A. & Reza, A. Innovative method for alzheimer's disease detection using convolutional neural networks. In *Lecture Notes in Networks and Systems*[SPACE]https://doi.org/10.1007/978-3-030-87903-2_45 (2023).
5. Li, J., Wei, Y., Wang, C. & Xu, L. 3-d cnn-based multichannel contrastive learning for alzheimer's disease automatic diagnosis. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–10. <https://doi.org/10.1109/TIM.2021.3124787> (2022).
6. Sujathakumari, B., Kulkarni, S. & Hallikeri, V. Brain magnetic resonance imaging image classification for alzheimer's disease and its hardware acceleration. *IAES International Journal of Artificial Intelligence* **13**, 1–11. <https://doi.org/10.11591/ijai.v13i1.pp1-11> (2024).
7. Savas, S. Detecting the stages of alzheimer's disease with pre-trained deep learning architectures. *Arabian Journal for Science and Engineering* **47**, 919–929. <https://doi.org/10.1007/s13369-021-05769-5> (2022).
8. Rahat, I., Hossain, T., Ghosh, H. & Ravindra, J. Exploring deep learning models for accurate alzheimer's disease classification based on mri imaging. *EAI Endorsed Transactions on Pervasive Health and Technology* **10**, e4. <https://doi.org/10.4108/eai.10-10-2023.178412> (2024).
9. Wu, C., Guo, S., Hong, Y. & Zhang, Q. Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. *Quantitative Imaging in Medicine and Surgery* **8**, 400–410. <https://doi.org/10.21037/qims.2018.05.15> (2018).
10. Oktavian, M., Yudistira, N. & Ridok, A. Classification of alzheimer's disease using the convolutional neural network (cnn) with transfer learning and weighted loss. *IAENG International Journal of Computer Science* **50**, 391–396. <https://doi.org/10.1142/S021821302350024X> (2023).
11. Kadri, R., Bouaziz, B., Tmar, M. & Gargouri, F. Comprehensive strategy for analyzing dementia brain images and generating textual reports through vit, faster r-cnn and gpt-2 integration. *Digital Signal Processing* **138**, 103084. <https://doi.org/10.1016/j.dsp.2023.103084> (2023).
12. Ledig, C., Schuh, A., Guerrero, R. & Heckemann, R. A. Dataset - structural brain imaging in alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *G-Node* (2018).
13. Petersen, R. C. et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology* **74**, 201–209 (2010).
14. Petersen, R. C. Mci criteria in adni: Meeting biological expectations. *Neurology* **97**, 597–599. <https://doi.org/10.1212/WNL.00000000000012588> (2021).
15. Ledig, C., Schuh, A., Guerrero, R., Heckemann, R. A. & Rueckert, D. Structural brain imaging in alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Scientific Reports* **8**, 11258 (2018).
16. Heckemann, R. A., Ledig, C., Gray, K. R. & Aljabar, P. Brain extraction using label propagation and group agreement: PinCram. *PLOS ONE* **10**, e0132192 (2015).
17. Ledig, C., Shi, W., Makropoulos, A. & Koikkalainen, J. Consistent and robust 4d whole-brain segmentation: Application to traumatic brain injury. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 673–676 (2014).
18. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv (2020).
19. Arnab, A. et al. Vivit: A video vision transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021).
20. Hendrycks, D. & Gimpel, K. Gaussian error linear units (gelus). arXiv (2016). et al.1606.08415.
21. Sohn, I. Deep belief network-based intrusion detection techniques: A survey. *Expert Systems with Applications* **167**, 114170 (2021).
22. Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D. & Ritter, K. Visualizing convolutional networks for mri-based diagnosis of alzheimer's disease. In et al., D. S. (ed.) *Understanding and Interpreting Machine Learning in Medical Image Computing Applications* (Springer International Publishing, 2018).
23. Korolev, S., Safiullin, A., Belyaev, M. & Dodonova, Y. Residual and plain convolutional neural networks for 3d brain mri classification. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 835–838 (2017).
24. Payan, A. & Montana, G. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. arXiv (2015). et al.1502.02506.
25. Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A. & Catheline, G. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. arXiv (2018). et al.1811.07782.
26. Göschel, L., Kurz, L., Dell'Orco, A. & Köbe, T. 7t amygdala and hippocampus subfields in volumetry-based associations with memory: A 3-year follow-up study of early alzheimer's disease. *NeuroImage: Clinical* **38**, 103439 (2023).
27. Coupé, P., Manjón, J. V., Mansencal, B. & Tourdias, T. Hippocampal-amygdalo-ventricular atrophy score: Alzheimer disease detection using normative and pathological lifespan models. *Human Brain Mapping* **43**, 3270–3282 (2022).
28. Ahmed, S., Kim, B. C., Lee, K. H., Jung, H. Y. & for the Alzheimer's Disease Neuroimaging Initiative. Ensemble of roi-based convolutional neural network classifiers for staging the alzheimer disease spectrum from magnetic resonance imaging. *PLOS ONE* **15**, e0242712, <https://doi.org/10.1371/journal.pone.0242712> (2020).
29. Tanji, K., Sakurada, K., Funiu, H. & Suzuki, K. Functional significance of the electrocorticographic auditory responses in the premotor cortex. *Frontiers in Neuroscience* (2015).
30. Silva, A., Liu, J., Zhao, L. & Chang, E. A neurosurgical functional dissection of the middle precentral gyrus during speech production. *Journal of Neuroscience* (2022).
31. Pavuluri, M. & May, A. I feel, therefore, i am: The insula and its role in human emotion, cognition and the sensory-motor system. *AIMS Neuroscience* (2015).
32. Cai, S., Peng, Y., Chong, T. & Huang, L. Differentiated effective connectivity patterns of the executive control network in progressive mci: A potential biomarker for predicting ad. *Current Alzheimer Research* (2017).
33. Fadiga, L., Craighero, L. & D'Ausilio, A. Broca's area in language, action, and music. *Annals of the New York Academy of Sciences* (2009).
34. Uddén, J. & Bahlmann, J. A *rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus* (Philosophical Transactions of the Royal Society B, Biological Sciences, 2012).
35. Cui, L., Zhang, Z., Zac Lo, C.-Y. & Guo, Q. Local functional mr change pattern and its association with cognitive function in objectively-defined subtle cognitive decline. *Frontiers in Aging Neuroscience* (2021).
36. Liang, P., Wang, Z., Yang, Y. & Li, K. Three subsystems of the inferior parietal cortex are differently affected in mild cognitive impairment. *Journal of Alzheimer's Disease* (2012).
37. Liu, M., Cheng, D., Yan, W. & Initiative, A. D. N. Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. *Frontiers in Neuroinformatics* **12** (2018).
38. Schwarz, C. G. et al. A large-scale comparison of cortical thickness and volume methods for measuring alzheimer's disease severity. *NeuroImage: Clinical* **11**, 802–812 (2016).
39. Turner, H. C. et al. Analyses of the return on investment of public health interventions: A scoping review and recommendations for future studies. *BMJ Global Health* **8**, e012798. <https://doi.org/10.1136/bmjgh-2023-012798> (2023).

Author contributions

Hasan AlMarzouqi: proposed the paper idea and wrote the paper Lyes Saoud: Implemented the approach, improved the model, and wrote the paper.

Declarations

Competing interests

The authors declare no competing interests

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-76313-0>.

Correspondence and requests for materials should be addressed to H.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024