Information Fusion 92 (2023) 363–388

Contents lists available at ScienceDirect

Information Fusion

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

Prediction of Alzheimer's progression based on multimodal Deep-Learning-based fusion and visual Explainability of time-series data

Nasir Rahim [a], Shaker El-Sappagh [a,b,c,*], Sajid Ali [d], Khan Muhammad [e,*], Javier Del Ser [f,g], Tamer Abuhmed [a,*]

[a] Information Laboratory (InfoLab), Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, South Korea
[b] Faculty of Computer Science and Engineering, Galala University, Suez 435611, Egypt
[c] Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, 13518, Banha, Egypt
[d] Information Laboratory (InfoLab), Department of Electrical and Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[e] Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Republic of Korea
[f] TECNALIA, Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain
[g] Department of Communications Engineering, University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain

ARTICLE INFO

Keywords:
AD progression detection
3D CNN
Multimodal information fusion
Time-series data analysis
Explainable AI

ABSTRACT

Alzheimer's disease (AD) is a neurological illness that causes cognitive impairment and has no known treatment. The premise for delivering timely therapy is the early diagnosis of AD before clinical symptoms appear. Mild cognitive impairment is an intermediate stage in which cognitively normal patients can be distinguished from those with AD. In this study, we propose a hybrid multimodal deep-learning framework consisting of a 3D convolutional neural network (3D CNN) followed by a bidirectional recurrent neural network (BRNN). The proposed 3D CNN captures intra-slice features from each 3D magnetic resonance imaging (MRI) volume, whereas the BRNN module identifies the inter-sequence patterns that lead to AD. This study is conducted based on longitudinal 3D MRI volumes collected over a six-months time span. We further investigate the effect of fusing MRI with cross-sectional biomarkers, such as patients' demographic and cognitive scores from their baseline visit. In addition, we present a novel explainability approach that helps domain experts and practitioners to understand the end output of the proposed multimodal. Extensive experiments reveal that the accuracy, precision, recall, and area under the receiver operating characteristic curve of the proposed framework are 96%, 99%, 92%, and 96%, respectively. These results are based on the fusion of MRI and demographic features and indicate that the proposed framework becomes more stable when exposed to a more complete set of longitudinal data. Moreover, the explainability module provides extra support for the progression claim by more accurately identifying the brain regions that domain experts commonly report during diagnoses.

1. Introduction

To date, Alzheimer's disease (AD) has remained perhaps the most degenerative chronic brain disease and primarily affects people over the age of 65 [1]. The World Health Organization [2] recently reported that 50 million people are living with AD, and this figure is predicted to triple by 2050. Unfortunately, there is no cure for AD at the time, and existing therapies can only slow its progression [2]. Early diagnosis is critical because the available treatment options for AD are most successful during the early stages of the disease [3]. Mild cognitive impairment (MCI) is often regarded as a transitional stage in the progression from normal health to AD [4]. According to recent studies, 10–15% of people with MCI progress to AD every year [5]. Numerous machine learning (ML) techniques have been proposed by the research community in the medical domain to distinguish between different classes of cognitive impairments, including stable MCI and progressive MCI [6]. However, many of these techniques mainly focus on a single modality of data or baseline (BL) data. Disease diagnosis based on a single modality may

* Corresponding authors.
E-mail addresses: shaker@skku.edu (S. El-Sappagh), khan.muhammad@ieee.org (K. Muhammad), tamer@skku.edu (T. Abuhmed).

https://doi.org/10.1016/j.inffus.2022.11.028
Received 4 June 2022; Received in revised form 23 November 2022; Accepted 27 November 2022
Available online 5 December 2022

increase the chances of misdiagnosis. In addition, analysis of degenerative brain diseases (e.g., Alzheimer's and Parkinson's) based on a BL or data from a single visit makes it difficult to distinguish between cognitively normal (CN) and MCI due to the high resemblance of the brain tissues [7].

AD is a severe form of cognitive impairment, in which domain experts rely on analyzing the patient's cognitive health using multimodalities [8]. Most studies on AD diagnosis are based on a single modality, such as positron emission tomography (PET), magnetic resonance imaging (MRI), or cognitive scores (CSs) [9]. In many of these studies, the problem of AD diagnosis is formulated as a binary classification task (e.g., CN vs. AD) to simplify the training process. For instance, Bron et al. [10] organized a computer-aided diagnosis competition of dimensia1 (CADDementia1) to compare traditional ML techniques for AD diagnoses. They evaluated 29 ML algorithms that only used MRI data gathered from a single visit to predict the outcomes. The best accuracy was only 63% for the classification of the three-class problem, namely, CN, MCI, or AD. Jiang et al. [11] proposed an eight-layer deep convolutional neural network (DCNN) model that utilizes batch normalization and dropout layers to deal with the problem of overfitting. This model relied on neuroimaging modalities, specifically MRI images, for the input data and resulted in an outperforming classification accuracy using 7399 CN subjects and 7399 patients with AD. Zhang et al. [12] proposed a novel and lightweight binary classifier to distinguish CN from AD by using whole-brain 3D MRI volumes. They utilized only the axial slice from the 3D MRI volume and preprocessed them through a standard preprocessing step, such as spatial normalization, skull stripping, and stationary wavelet entropy. A neural network (NN) with a single layer was utilized, and the network parameters were trained using a particle swarm optimizer, achieving approximately 93% accuracy.

The diagnostic procedure for AD performed by the different studies mentioned above is mainly based on MRI modality alone. However, fusing multimodal data has been proven to enhance the accuracy of ML models in the medical domain [14,15]. In the case of AD diagnosis, it has been shown that combining neuropsychological battery results, CSs, demographics, and neuroimaging data, leads to significant improvement in model performance while reducing negative effects related to noise [15]. In addition, the resultant models are widely acceptable in the medical environment because of the multimodal diagnostic procedure. For instance, Zhang et al. [12] proposed a multimodal neural network based on MRI scans, cerebrospinal fluid (CSF) tests, and fluorodeoxyglucose PET(FDG-PET) scans to distinguish CN, MCI, and AD. Xu et al. [13] used volumetric MRI, FDG-PET, and florbetapir PET modalities to classify MCI and AD. Huang et al. [16] proposed combining MRI and CSF modalities to differentiate CN individuals from those with MCI. Gray et al. [17] incorporated four modalities (FDG-PET, MRI, CSF, and genetic features) and trained a random forest (RF) model for CN, MCI, and AD. However, these studies utilized only a single time step (i.e., all data came from a single BL visit, with no follow-up data collected). Furthermore, no attention was given to a possible time-series aspect of the data, which would have allowed researchers to examine the influence of how sequential features change over time, allowing for improvements in classification performance. In addition, ignoring subsequent time steps from a given dataset eliminates the most critical information that defines the disease progression [8,14].

Time-series data management and analysis are crucial for the assessment of brain-related diseases [60]. However, very limited research based on time-series data analysis has been conducted, particularly in the field of AD progression detection. For example, Chincarini et al. [18] performed MRI-based AD progression detection using time-series data from the Alzheimer's disease neuroimaging initiative (ADNI) dataset. They used four MRI volumes as input data for each patient (i.e., two volumes from the BL visit, one volume taken at a 12-month follow-up appointment, and one from an appointment 24 months after the initial visit). Their focus was on capturing the morphometry of the hippocampal subregions to track the physical progression of AD. They formulated their problem as two binary classification tasks (CN vs. AD and CN vs. MCI). They achieved an area under the receiver operating characteristic (AUROC) curve of 93% for the CN vs. AD task and 88% for the CN vs. MCI task. Similarly, Moradi et al. [19] used MRI data to train a semi-supervised ML-based algorithm to predict the transition of patients from MCI to AD over the next three years. Moore et al. [20] trained a classic RF model to examine the correlation between pairs of data points as they changed over different time steps. In this experiment, demographic features, along with physical data from scans of the patient's brain and their CSs, were utilized to predict AD. The use of multimodal data for building DL-based diagnostic systems has been highly encouraged by domain experts. This is because it enhances the possibility of building accurate, stable, and medically intuitive models, as is evident from the aforementioned studies [21].

In the medical domain, physicians do not accept diagnoses from DL-based algorithms that provide accuracy based on a test dataset [21]. The ability of a model to justify why a specific diagnosis is being made is critical in the medical domain. Systems with this capability are called explainable artificial intelligence (XAI) systems [22]. A fully XAI system can explain the inner mechanisms involved in making specific decisions with the intention of involving a community. According to the European General Data Protection regulations, the use of black-box models is strictly prohibited in various domains, particularly in healthcare systems, and the retractability of decisions made by such systems is discouraged by medical experts [23]. An artificial intelligence (AI) system for the medical domain should have a certain level of transparency to convince medical practitioners and encourage human specialists to take on board the system's recommendations, as they use all their judgment and experience when making treatment decisions. Many scholars have suggested that humans cannot explain or justify their decisions at certain times [24]. Explainability is crucial for AI to be used in a secure, ethical, fair, and trustworthy manner while being a key enabler for its real-world application. Many XAI-based studies in the medical domain have shown excellent results by visually explaining selected features utilized in the decision-making process of a model [22]. However, these studies deal with the diagnostic process of AD as a classification task using either a single modality of data or BL data, without considering the time dimension of the data. The existing explanation methods provided by deep-learning (DL) systems are mainly designed to specify explanatory feature maps, such as saliency maps, class activation maps (CAMs), and gradient-weighted CAMs (Grad--CAMs). By default, these systems are not capable of representing the temporal aspects of features found in 2D time-series data [25]. Despite the successful deployment of CAMs and Grad-CAMs in various domains, they cannot be easily adopted in the medical domain. This is because these types of visual representations do not provide exact pixel-level information specifying the exact tissue location being damaged over time, which is crucial in neurodegenerative diseases such as AD.

In this study, we propose a framework that uses a deep 3D CNN followed by a bidirectional recurrent neural network (3D-CNN-BRNN) to predict the progression of AD. Fig. 1 presents a generic overview of the proposed framework, and further details of the model architecture are provided in Section 3.3. We optimize a hybrid DL model that combines the capabilities of 3D CNNs, recurrent neural networks (RNNs), and feed-forward neural networks, outperforming models in previous studies [26–28]. Our architecture relies on longitudinal 3D MRI volumes recorded over three time steps (BL, month 6 (M06), and month 12 (M12)). Given these data, our model can predict the medical condition of a patient three years later (i.e., at month 48 (M48)). The designed framework is an end-to-end DL model in which the 3D CNN module captures inter- and intra-slice features of 3D volumes by analyzing a patient's time-series volumes to produce a latent representation of convolutional features. Unlike previous studies, the BRNN module was precisely designed to track inter-volumetric relationship features extracted from a 3D CNN module that evolves over different time steps.

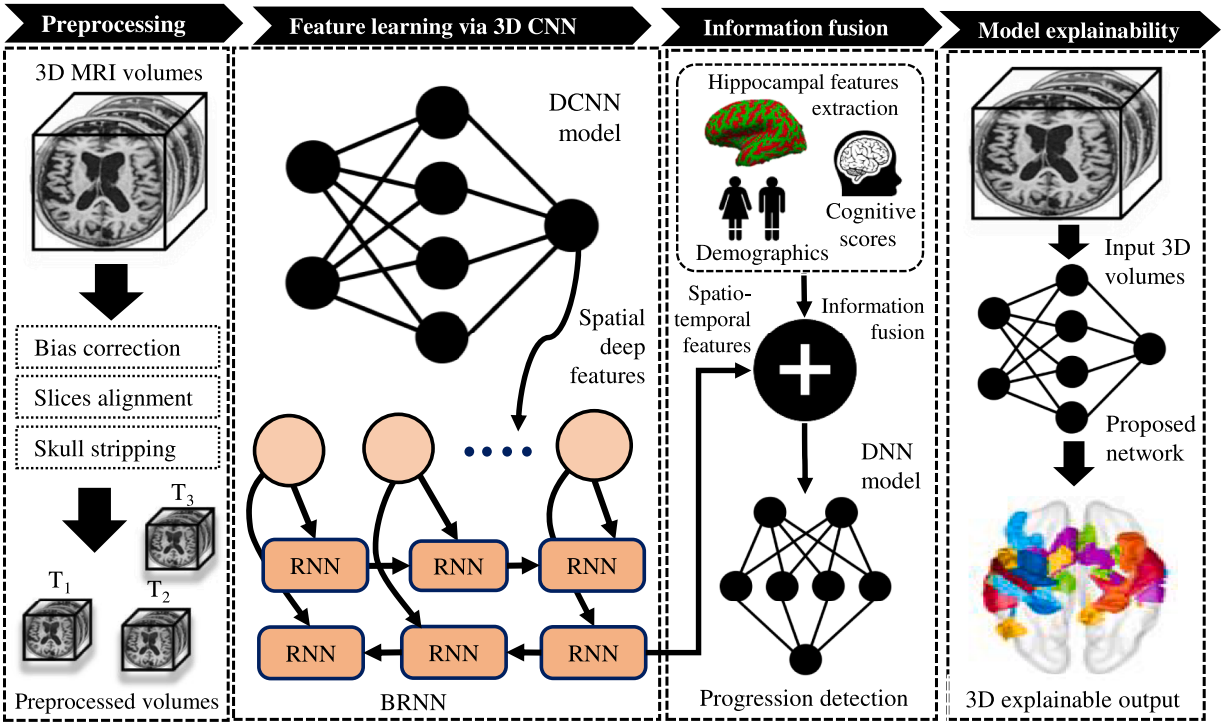

Fig. 1. Birds eye view of the proposed framework.

Furthermore, we also studied the effect of fusing other critical modalities, such as demographic and cognitive biomarkers recorded at the BL, with the deep features extracted from the 3D-CNN-BRNN model for AD progression detection. Medical experts are interested in determining reasons for specific medical decisions. XAI helps models elicit understandable explanations of their output to different audiences [22]. To the best of our knowledge, the time-series visual explainability of 3D MRI neuroimages has not been explored in the literature on AD. After optimizing the DL-based model to achieve its best performance, we extended the model to provide time-series explainability using visual representations over 3D images to further justify the model's decision. This module tracks the regions in the brain tissue over time to help the model recognize patients with AD. The proposed approach addresses both issues by choosing the guided Grad-CAM, which provides voxel-level activation maps as we go through each time step of the 2D MRI slices.

The key contributions of our study are summarized as follows.

- We propose a novel DL model (3D-CNN-BRNN) to address the limitations of the existing deep models for detecting AD progression. Our model predicts the progression of AD 3 years after the final data collection visit, that is, at month 48, which is an acceptable period for the patient's caregivers and family to appropriately respond to the patient's various situations. The model performs predictions based on data from patients' longitudinal MRIs, along with their CSs and demographic features recorded at the BL.
- We propose a novel XAI technique for the 3D visualization of time-series data. We visualize the time-dependent aspects of longitudinal MRI data collected for AD progression detection. To the best of our knowledge, no study in the field of AD progression detection has provided a temporal XAI approach for classification tasks based on longitudinal 3D MRI input data. This enables domain experts to visually inspect the progressive patterns of AD in any given patient detected by the proposed framework. Furthermore, the proposed framework can be adopted in a healthcare system with acceptable detection accuracy, being classified as a white-boxed model.
- We investigate the relationship between using time-series data modalities and improvements in the true positive detection rate to further enhance the trustworthiness of the proposed model.
- Extensive experiments on the ADNI dataset are conducted in various settings. A thorough analysis of our model's performance and comparison with other classic models, that is, 3D VGG13 and 3D ResNet18, showed that our model outperformed all other models on a variety of evaluation metrics, even under different conditions.

The remainder of this paper is structured as follows. Section 2 includes a detailed discussion of the materials and methods of this study. Section 3 explains the implementation of the proposed framework, while Section 4 presents the experimental results and analysis. Section 5 discusses the proposed explaninability approach. Section 6 discusses the limitations of this study, and Section 7 concludes the study by highlighting the key findings along with some suggestions for future research.

2. Materials and methods

The proposed framework is presented in detail in this section. Our method uses MRI along with demographic features and CSs to detect the progression of AD. Fig. 2 presents an overview of the proposed framework. The proposed DL model is composed of a 3D CNN followed by a BRNN for the detection of AD progression. The 3D CNN is responsible for capturing the intra-slice features among the N inputs given for each time step, whereas the BRNN takes as input N sets of time-series data (i.e., N sequences) to calculate AD progression via the binary scores given as its output. These scores specify whether the patient will progress to AD.

The proposed 3D-CNN-BRNN framework comprises four stages. Stage 1 performs essential preprocessing steps on each MRI volume to remove unwanted artifacts and to transform the data into a standard format. Subsequently, the 110 most informative coronal slices are extracted from the preprocessed input volumes. Slice selection is performed after calculating the total half-volume size measured on either side of the central slice in the 3D volume. In this way, we use only the

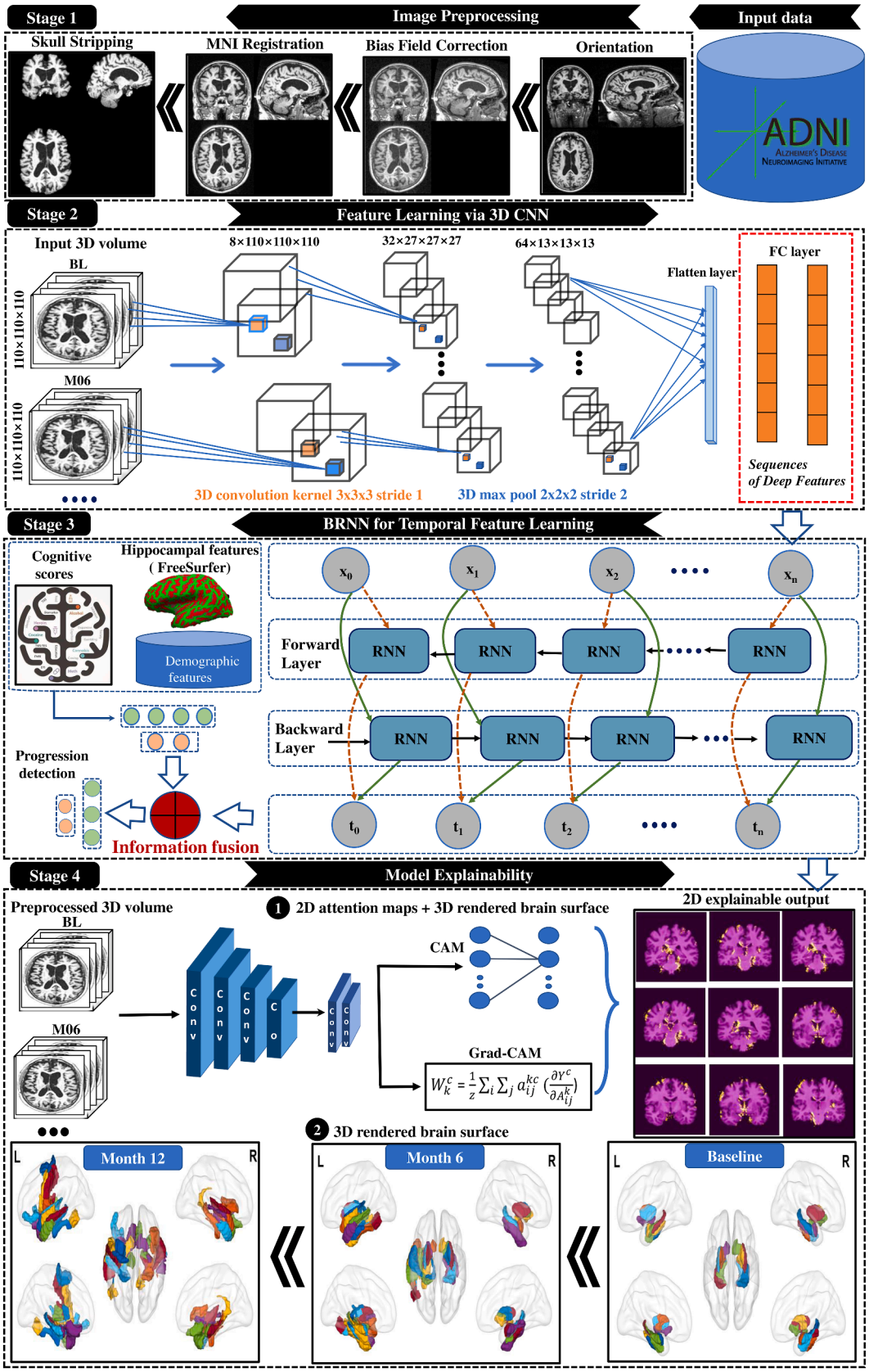

Fig. 2. Proposed 3D-CNN-BRNN framework for AD progression detection and time-series visual explainability.

most informative 2D slices, which enable the proposed network to capture the most representative feature vectors for AD. In Stage 2, the 3D CNN module extracts deep features from the input 3D MRI volumes. Then, to extract deep features from the longitudinal data in the BL∼M6 and BL∼M12 scans, volumes of the same brain for each time step are passed to the 3D CNN to calculate deep features between these time-separated scans. A 3D CNN is designed to extract spatial and temporal features simultaneously, as shown in [29]. However, in the proposed framework, we leverage the capability of 3D CNN models to capture spatial features and the intra-slice relation of the 3D MRI volume at every single time step, that is, BL, M06, and M12. Because we are using time-series MRI volumes, the extracted spatial and intra-slice deep features of the 3D CNN model are used by the BRNN to capture inter-volume relationships for 3 time steps. By the model design, a 3D CNN is dedicated to learning the deep spatial and inter-slices features while the BRNN is designed to learn the spatial and inter-volume features. This is the main reason for dedicating the BRNN to learning deep spatial and intra-volume features. The extracted features are then forwarded to Stage 3, where the extracted features from the three volumes are used as the input to the BRNN model to learn the inter-volume temporal features. A BRNN was selected because the time-series was not very long. The features extracted from the BRNN are concatenated first from both forward and backward directions and then passed through further 256-unit and 128-unit fully connected layers. Separately from the MRI data being processed through the 3DCNN and BRNN, other modalities such as CS and demographic features are processed through two stacked, fully connected layers with 32 and 16 units, respectively. The output feature vector is then fused with the output feature vector from the RNN module, forming 128- and 16-dimensional feature vectors. The concatenated feature vector is then further processed through two fully connected layers to classify each patient as either converted or normal. In addition to classification, we provide further visual representations of the spatial and temporal features from Stage 4 to help interpret and understand the classifications. This step further validated the reliability of the proposed network for real-world use, making it more medically acceptable to domain experts.

Stage 1: Image Preprocessing: In the first step, extraneous information is removed from the raw MRI volume, allowing for easier comparison of the various brain scans. To achieve this, all MRI volumes were passed through a standard preprocessing pipeline [30]. Image reorientation, bias field correction, skull stripping, and registration to the standard template space are among the preprocessing steps used in the pipeline.

Reorientation to standard space: Preprocessing operations, such as setting the image orientation or flipping left and right, are critical. These operations form the basis of this essential preprocessing step, which allows us to maintain consistency in voxel processing and interpretation across a variety of software and systems. Many full-featured open-source visualization tools can be used for free, including FSLEyes [31] and FreeView [32]. We used FSLEyes to visualize our data, which allowed us to find that some images had been rotated by 180°. We used the fslreorient2std tool from FSL to reorient these images before further processing.

Bias field correction: Numerous factors, such as the patient's position in the MRI scanner or the version of the scanner itself, can cause brightness issues in the resulting MR image. When these brightness issues are seen in MR images, they are caused by low-frequency, smoothed, and undesirable signals inside the scanner. This can affect the overall quality of each image. Failure to correct such inhomogeneities can negatively affect the effectiveness of any subsequent steps, including skull stripping and tissue segmentation. In our scenario, the N4 bias field correction algorithm was utilized from the advanced normalization tools (ANTs) [33]. It is an advanced N3 bias-field correction algorithm that uses an improved B-spline fitting routine to allow multiple resolutions in the correction process.

Skull stripping: Skull stripping is the process of isolating brain tissue from non-brain tissue in an MRI image of the brain. Once a brain mask is accurately calculated, the remaining body parts (e.g., residual neck voxels) can easily be eliminated from the brain region. This type of extra information acts as noise and contributes to the high dimensionality of the training data, while making the classification task more complex. In this study, we employed a software tool called the Brain Extraction Tool (BET2) [34], implemented in FSL, for skull stripping. Because considerable parts of the neck region were present in our scans, this approach failed to accurately identify the brain regions in our earlier investigation. Instead, we employed a strategy that incorporates segmentation phases to provide more robust outcomes to overcome this issue.

MNI152 standard template registration: Template registration is the process of aligning images based on brain structures, to simplify the process of comparing different images. In this study, the MNI152 [35] template space was used to perform the affine transformation. This process does not deform the image because it involves only basic transformation steps such as rescaling, rotation, translation, and shearing. Registration was performed using the FLIRT tool in FSL, and correlation ratios were used as the similarity metric.

Stage 2: Feature Learning via 3D CNN: In Stage 2, we extracted spatial and temporal deep features from the 3D MRI volumes using the proposed DCNN module.

Convolution operation: Most existing CNN architectures are designed for 2D images, indicating that these architectures cannot efficiently encode spatial information from the 3D volumes that we must use as inputs. However, alternative architectures, such as 3D ConvNets, model temporal information much better than 2D ConvNets. They can process both spatial and temporal features simultaneously, which is not the case for 2D ConvNets. Temporal information from the image is typically lost immediately after every convolution operation in the 2D ConvNets. A typical 3D CNN comprises alternatively stacked 3D convolution operations followed by a rectified linear unit (ReLU) activation function to generate feature maps for each filter. The feature maps obtained by the convolution layer comprise the discriminatory information from the input image. These feature maps were further downsampled using a 3D max-pooling layer.

3D Max-pooling: In 3D ConvNets, max-pooling is a downsampling strategy that not only reduces the spatial dimension of the given input but also reduces the depth dimension of the input 3D volume. This is performed by calculating the maximum value from an $h \times w \times d$-size sub-block from an input volume of size $h \times w \times d$ as specified by a kernel of predefined size k and stride s. The output of such an operation is $\frac{(h-k)}{(s+1)} \times \frac{(w-k)}{(s+1)} \times \frac{(d-k)}{(s+1)}$. The max-pooling operation helps preserve the most crucial features that help distinguish input volumes. As input data move from lower layers to higher levels, the max-pooling layer compresses and optimizes the features, making them more robust to distortions and geometric variances.

Network regularization via dropout layer: Dropout is a popular regularization approach that works by randomly choosing neurons at different layers of a neural network. This concept was first presented by Hinton et al. [36] to address the problem of network overfitting. The dropout sets the output of a particular neuron in the hidden layer to zero based on the predefined dropout ratio. The ratio provides the probability of a single neuron being dropped during the training process. Consequently, if the probability value is set to 1, all neurons in the hidden layer will become 0. During the back-propagation step, the dropped neurons participate neither in a forward pass nor in the backward pass. The neural network becomes simpler and more generalizable to the input data when a dropout layer is introduced into the hidden layers.

Stage 3: Temporal Feature Learning via BRNN and Information Fusion: RNNs are a special type of neural network used to analyze sequential inputs; they were specifically designed for natural language processing tasks [37]. RNNs are effective in capturing temporal features from sequences; however, they suffer from a vanishing gradient problem [38]. Long short-term memory (LSTM) models are known as an

advanced variant of RNNs that utilize a mechanism that preserves 'cell state' to learn the long-term dependencies while forgetting the less important information [39]. However, LSTMs are slow when used with many parameters, which makes them prone to overfitting. To circumvent these issues, we used an RNN with a bidirectional architecture to efficiently learn long-range temporal dependencies and overcome the vanishing gradient problem. Fig. 3 presents an unfolded BRNN architecture for N time steps.

One RNN instance receives the input sequence in the normal order, whereas the other RNN receives the same input sequence in reverse order. The outputs of both RNNs were concatenated and forwarded to the next hidden layer. By leveraging the simplicity and effectiveness of BRNNs, we can choose to use BRNNs to appropriately capture features of the progression of AD by considering information from both the future and past time steps.

Let us suppose that x represents an input having length T and the recurrent network consists of I input units, H hidden or intermediate units, and K output units. x_i^t is the $i^{\{th\}}$ input at t time step. Let a_j^t and b_j^t denote the input of the network and the nonlinear identifiable activation function of the j^{th} element at t time, respectively. We initiate with $t = 1$ and iteratively call Eqs. (1) and (2) to obtain the entire sequence of implicit units.

$$a_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} b_{h'}^{t-1} \tag{1}$$

$$b_h^t = \theta\left(a_h^t\right) \tag{2}$$

where θ represents the hyperbolic tangent function, which is a nonlinear activation function. a_h^t represent the hidden state of the forward pass. Similarly, Eq. (3) can be used to compute the output unit of the network.

$$a_k^t = \sum_{h}^{H} w_{hk} b_n^t \tag{3}$$

The activation function influences the objective function of the RNN model, not only on the network's output layer but also on the hidden layer of the following time step, as shown in Eq. (4).

$$\frac{\partial_o}{a_j^t} = \theta'\left(a_h^t\right)\left(\sum_{k=1}^{K} \delta_k^t w_{hk} + \sum_{h'=1}^{H} \delta_{h'}^{t+1} w_{hh'}\right) \tag{4}$$

In each step, the weights of the inputs and outputs of the hidden layer units are equal. As stated in Eq. (5), we can sum the series to obtain the derivative of each layer weight.

$$\frac{\partial_o}{\partial_{w_{ij}}} = \sum_{t=1}^{T} \frac{\partial_o}{\partial a_j^t} \frac{\partial a_j^t}{\partial_{wt}} = \sum_{t=1}^{T} \delta_j^t b_j^t \tag{5}$$

AD is characterized by the progressive deterioration of cognitive abilities [41]. Moreover, age and education level are important factors in accurately assessing the progression of AD [7,13]. In the current study, we investigated the effect of fusing patient demographics and CSs using neuroimaging data to detect AD. In the proposed pipeline, the system learns deep features from the patient's CSs and demographic data and fuses them with the deep features coming from the BRNN. Furthermore, the output from the BRNN is classified as converted or normal status.

Stage 4: Model Explainability: We briefly explain the existing techniques available for highlighting and localizing the important regions in the input image that lead to the classification of the system to help visually explain the CNN's result.

Class activation map (CAM): Class activation maps were originally designed for a specific network architecture, where the global average pooling (GAP) layer is applied to each feature map at the final convolution layer to help interpret prediction decisions made by the CNN.

Let $F^k \in \mathbb{R}^{u \times v}$ denote the GAP layer from a CNN architecture with a total of k feature maps and width u and height v from the last layer. The weight matrix of the k-th feature map may be w_k^c of class c. Then, a prediction score at the output layer, S_c, may be computed as a sum of the weight matrix of the GAP layer, as shown in Eq. (6).

$$S_c = \sum_k w_k^c F^k = \sum_k w_k^c \sum_{x,y} f_x(x,\ y) = \sum_{x,\ y} \sum_k w_k^c f_k(x,\ y) \tag{6}$$

where $f_x(x,\ y)$ expresses the activation from spatial locations $(x,\ y)$ in the k-th feature map. The CAM computes the activation of all feature maps, as represented by $M_c \in \mathbb{R}^{u \times v}$ for class c in Eq. (7).

$$M_{c(x,\ y)} = \sum_k w_k^c f_k(x,\ y) \tag{7}$$

Because the prediction score, S_c, is the sum of all $M_{c(x,\ y)}$, the CAM reflects the relevance of each spatial location $(x,\ y)$ in the final class activation map. Consequently, we can highlight ROIs inside the input image that show the most significant areas to a specific class by simply up-sampling the CAM equal to the input image.

Grad-CAM: Grad-CAM refers to the use of the gradient flow in the backward direction to the input layer to produce visual representations. In contrast to the CAM, the Grad-CAM does not require a GAP layer. Instead, it uses a single fully connected or dense layer. Therefore, the Grad-CAM visualization method does not require the modification of the existing CNN architecture to produce visualization maps. The Grad-CAM for a particular class, C, can be calculated as the weighted sum of the k feature maps. Furthermore, the ReLU activation function is employed on the feature maps to remove the effect of negative gradients for a given class, C. As described mathematically in Eq. (8), the spatial components in the feature maps linked with negative weights are likely to correspond to other classes in the image.

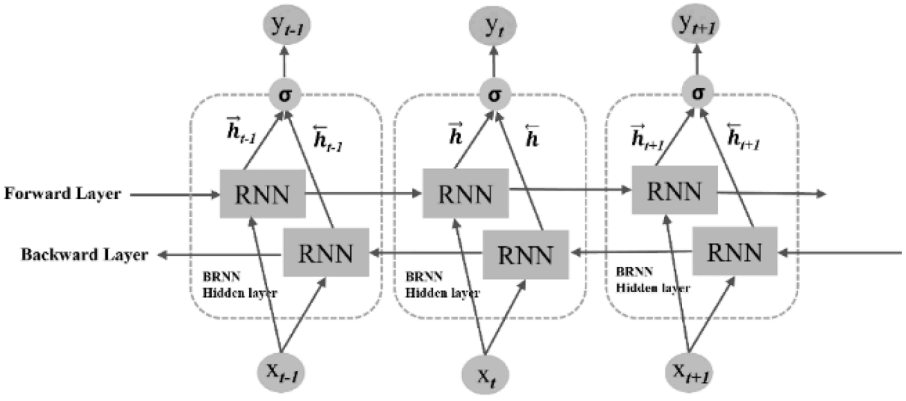

Fig. 3. Unfolded BRNN architecture for N time steps [40].

$$Grad\_M_c(x,\ y) = ReLu\left(\sum_k \alpha_k^c f_k(x,\ y)\right) \tag{8}$$

Here, α_k^c represents the weights computed by the S_c gradients with respect to the k-th feature map, as in Eq. (9).

$$\alpha_k^c = \sum_{x,\ y} \frac{\partial S_c}{\partial f_{k(x,\ y)}} \tag{9}$$

According to Eqs. (8) and (9), α_k^c is equal to w_k^c in the CAM approach. As a result, applying ReLU as an activation to omit spatial pixels with negative weights makes Grad-CAM different from CAM in such cases. Fig. 4 shows some Grad-CAM activation maps from the proposed network that specify the salient brain regions that influence the AD detection result.

Guided Grad-CAM: Despite its popularity, the Grad-CAM technique has one drawback: it does not provide any notable indication of the exact pixel involved. To obtain exact voxel location information in terms of their influence on the decision-making process, we chose to use the guided backpropagation approach [42]. Eq. (10) can be applied to evaluate the gradient of the predicted score of class C, y^c, with respect to input image x.

$$m_c = \frac{dy^c}{dx} \tag{10}$$

In guided backpropagation, the positive values of the gradient are backpropagated and the negative values are set to zero. Let G_i^l be a gradient backpropagated through layer l and $f_i^{l+1} = ReLu(f_i^l)$, then Eq. (11) defines the backpropagation process.

$$G_i^l = (f_i^l) > 0)\ (G_i^{l+1}) \tag{11}$$

After generating the highlighted region associated with the particular backpropagation, we utilize attention mapping according to $m_c = \mathrm{ReLu}(m_c)$ that visualizes each pixel; it only has a positive impact on score y^c. Fig. 4 shows the different attention maps provided by the CAM and guided Grad-CAM algorithms.

3. Experimental setup

The proposed framework was implemented using the PyTorch (1.7.0) library. All experiments were conducted on a workstation with an NVIDIA GeForce GTX TITAN X GPU with 12 GB RAM. The proposed 3D-CNN-BRNN model was trained in an end-to-end manner, and an Adam optimizer was used for parameter optimization. With an extensive search over various hyperparameter settings for optimization, the learning rate used for training was 0.00027, momentum was set to 0.9, and weight decay rate was set to 0.1×10^{-6}. The number of training and validation batches varied based on the number of time steps in the input data. We used training and validation batch size sets for five input volumes when training with only the BL data. The batch size was reduced by one as the number of input time steps increased. For two time steps, the mini-batch size was further reduced to four volumes by passing 4×2 volumes [BL, M06] for each patient. In the case of three time steps, 3×3 volumes [BL, M06, M12] were passed for each patient. Stratified cross-validation (CV) was used in each fold to provide a balanced batch during the training phase. The number of epochs was set to 120, where the loss reached the minimum value at each fold in the 5-fold CV process.

3.1. Implemented 3D CNN architectures

We compared the proposed model with two well-known CNN architectures: 3D ResNet18 [43] and 3DVGG13 [44]. As we are dealing with a 3D version of CNN models, to the best of our knowledge, no official standard architecture can be altered from a 2D deep model to a 3D model. Note that we implemented 3D ResNet18 and 3D VGG13 from scratch and used them for comparison with our proposed architecture. We selected these models because they have approximately the same number of trainable parameters and relatively similar network architectures. In particular, our model consists of ten layers, which is comparable to 3D VGG13 and 3D ResNet18. In addition, we empirically observed during the optimization of the proposed model that, as the model's number of layers increases, it can easily overfit to the training data. This observation suggests that any benchmark comparison of our model with large-sized architectures, such as DenseNet 121 or InceptionNet, should be avoided. A brief explanation of each comparative model is provided below.

3D ResNet: This architecture resolves the problem of performance degradation as network depth increases. He et al. [45] addressed the issue of deeper networks by introducing a concept called a residual connection. The ResNet architecture directly connects two or more layers via residual connections, allowing a certain portion of the information from the previous network layers to be preserved. This mechanism ensures stability in network performance and prevents the gradients in the deeper layers from becoming very small during the backpropagation step. For instance, if x is the input parameter of the network and $H(x)$ is the target output, then $F(x) = H(x) - x$ is the residual connection that allows ResNet to learn across the sub-module such that the target output can be $F(x) + x$. Mathematically, the residual connection can be defined as in Eq. (12).

$$x_{l+1} = f[x_l + F(x_l,\ k_t)] \tag{12}$$

where x_l is the input, x_{l+1} is the output of the l-th residual block, f is the activation function, F is the residual function, and k represents the convolution kernel. Fig. 5 shows the ResNet's residual connection [45].

In this study, we utilized 3D ResNet18 as the backbone architecture to extract the deep features from each 3D MRI volume. The obtained feature maps were utilized to train a BRNN for the progression-detection task. The input size of the 3D volume was set to $110 \times 110 \times 110$, and the Adam optimizer was used for parameter tuning. Other hyperparameters were set, such as the hyperparameters used for training the proposed DCNN model, that is, the learning rate and batch size.

3D Visual Geometric Group (VGG): VGG is a well-known DCNN designed and evaluated on the ImageNet dataset. It is an improved version of the classic AlexNet model, where, in the VGG model, the 11×11 and 5×5 filters in AlexNet are replaced with a series of 3×3 kernel sizes. This mechanism significantly reduces the number of learnable parameters in the network without affecting its generalization ability during the training phase. A max-pooling layer of 2×2 is applied after each convolutional layer to reduce the spatial dimension of the input feature map. The rectified linear unit (ReLU) is an activation function utilized in all hidden layers of this network. Owing to the successful use of the VGG model in various computer vision tasks, we chose to use the 3D VGG model as the backbone feature extractor in the proposed framework. Feature maps from the final convolution layers were obtained for each 3D volume at each time step, that is, BL, M06, and M12. The obtained feature maps were flattened and forwarded to the BRNN to analyze the progressive patterns of AD. A detailed discussion of the proposed BRNN is provided in Table 1. The backbone VGG model was trained in an end-to-end manner by inputting a $110 \times 110 \times 110$ volume. All the hyperparameters were kept similar to those used for the training and testing of the proposed CNN-BRNN architecture, that is, batch size, learning rate, number of epochs, etc.

3.2. Evaluation metrics

The generalization capability of a trained model must be assessed by using multiple evaluation metrics. Stratified k-fold CV is a technique that preserves the percentage of each class and reduces the biased learning behavior of any model. This study used a stratified 5-fold CV technique to assess the model. In addition, various evaluation metrics

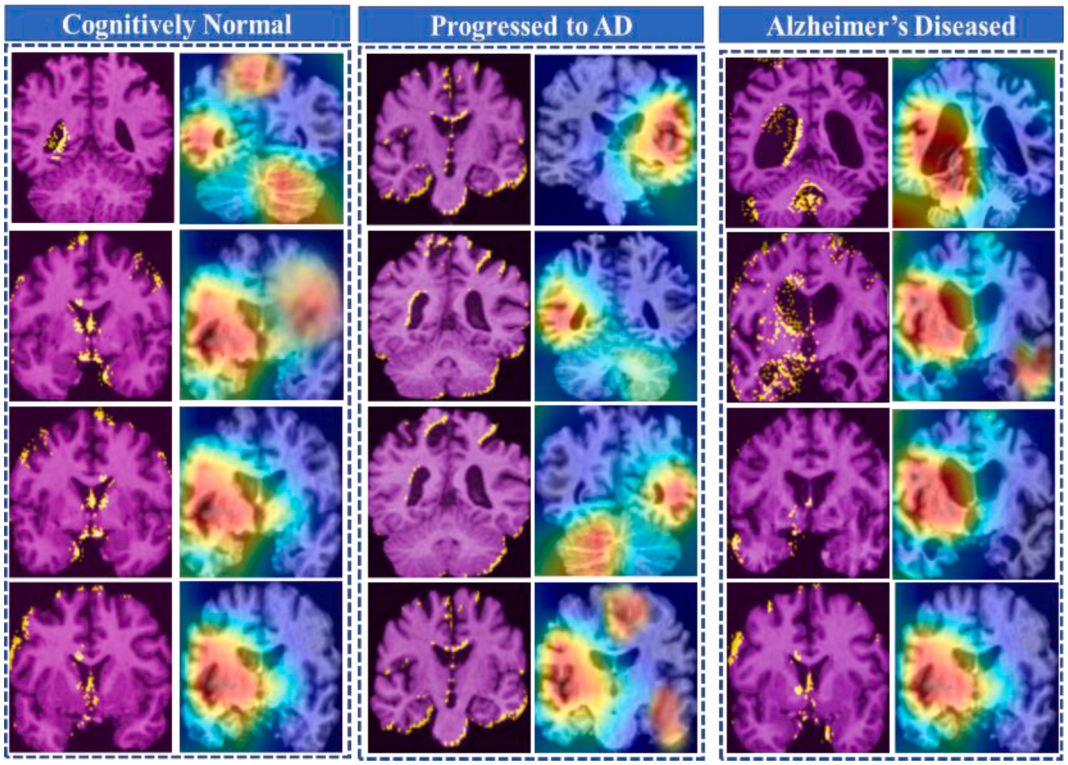

Fig. 4. Guided-Grad-CAM- and Grad-CAM-enabled voxels.

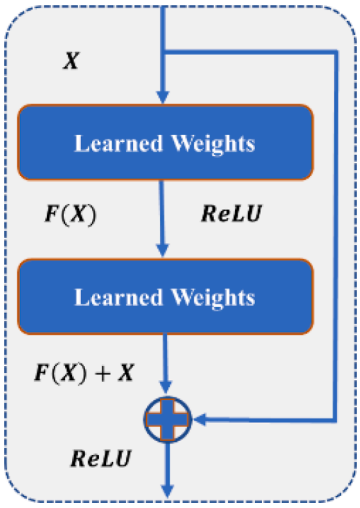

Fig. 5. Example of residual block learning.

were employed to demonstrate the quality-of-fit of the proposed model. The most common evaluation metrics used are accuracy, precision, recall, and AUROC. We also evaluated our model using the defined metrics because these metrics are commonly used measures in bioinformatics literature, helping to compare our results with those of other studies [46,15].

Accuracy: Percentage of instances successfully classified out of the total number of test instances.

Precision: Percentage of accurately classified positive instances out of all predicted positive instances.

Recall: Percentage of accurately classified positive instances out of all instances in the actual class.

Area under the curve (AUC): AUC is another evaluation metric designed to evaluate the classification performance of a ML model at various classification thresholds. The AUC score is a balanced evaluation metric that considers both true-positive and false-positive rates. The mean area under the curve (mAUC) form of the AUC result takes all ordered pairings of categories (i, j) and then calculates the likelihood that a randomly picked element from category i has a better estimated chance of being categorized as category i than a randomly selected element from category j and then all these probabilities are averaged. A perfect classifier has an AUC score of 1, whereas a classifier that makes random guesses has an AUC of 0.5.

3.3. Model architecture

In our experiments, the proposed architecture was optimized by using a well-known grid-search hyperparameter-tuning technique. We conducted many experiments to optimize the most critical hyperparameters of the proposed model, including the number of convolutional layers, the size of the convolutional kernel (i.e., 3×3, 5×5), the layer regularization parameter using a dropout layer in the range between 0.1–0.001, the number of BRNN layers, the number of BRNN units, the number of dense layers, and the number of neurons in each dense layer. An optimized 3D CNN was used to extract spatiotemporal features from the 3D MRI volume at each time step. The 3D CNN has 10 convolution layers and four max-pooling layers. The max-pooling layers were employed after the 2nd, 4th, and 7th convolution layers. The reason for not using a max-pooling layer after each convolutional layer was to preserve the spatial and temporal dimensions of the input volume to learn a wide range of features in both directions. However, the disadvantage of not using polling layers after each convolutional layer is that the training time can be increased. In the proposed network, we address this problem by determining the optimal number of kernels and kernel size at each layer [47,48]. We used eight $3 \times 3 \times 3$ kernels in the 1st and 2nd layers, and the spatial and depth dimensions of the feature maps with a $2 \times 2 \times 2$ max-pooling layer were reduced by a stride of two, followed by a 30% dropout layer. We increased the number of kernels as the spatial dimensions of the feature maps decreased and the

Table 1
Architectural details of the proposed network.

Layer ID	Layer Name	Number of Kernels	Kernel Size/Stride	Output Size
0	Input	–	–	110 × 110 × 110
1	Conv1	8	3 × 3 × 3/1	8 × 110 × 110 × 110
2	Conv2	8	3 × 3 × 3/1	8 × 110 × 110 × 110
4	MaxPool3D	–	2 × 2 × 2/2	8 × 110 × 110 × 110
3	Dropout (0.3)	–	–	8 × 55 × 55 × 55
4	Conv3	16	3 × 3 × 3/1	8 × 55 × 55 × 55
5	Conv4	16	3 × 3 × 3/1	16 × 55 × 55 × 55
6	MaxPool3D	–	2 × 2 × 2/2	16 × 55 × 55 × 55
7	Dropout (0.3)	–	–	16 × 27 × 27 × 27
8	Conv5	32	3 × 3 × 3/1	16 × 27 × 27 × 27
9	Conv6	32	3 × 3 × 3/1	32 × 27 × 27 × 27
10	Conv7	32	3 × 3 × 3/1	32 × 27 × 27 × 27
11	MaxPool3D	–	2 × 2 × 2/2	32 × 27 × 27 × 27
12	Dropout (0.4)	–	–	32 × 13 × 13 × 13
13	Conv8	64	3 × 3 × 3/1	32 × 13 × 13 × 13
14	Conv9	64	3 × 3 × 3/1	64 × 13 × 13 × 13
15	Conv10	64	3 × 3 × 3/1	64 × 13 × 13 × 13
16	MaxPool3D	–	2 × 2 × 2/2	64 × 13 × 13 × 13
17	Dropout (0.4)	–	–	64 × 6 × 6 × 6
18	Flatten	–	–	13,824
19	BRNN Layer	# of layers =1, # of recurrent units =512	–	13,824 + 17
20	Fully Connected	–	–	128
21	Fully Connected	–	–	64
22	Softmax	–	–	2

depth dimensions of the feature maps increased. Sixteen kernels were used in the 3rd and 4th convolutional layers, and the outputs of these layers was passed through the ReLU activation function, max pooling, and a 30% dropout layer. Thirty-two kernels were used in the 5th, 6th, and 7th layers with the ReLU activation function, max-pooling layer, and 40% dropout layer. Finally, 64 kernels were used at the 8th, 9th, and 10th layers with the ReLU activation function, max-pooling layers, and a 40% dropout layer. Throughout the network, the RELU activation function was used. The output of the last convolutional layer was flattened to a 1D feature vector that represents the input volumes from each time step, and the vector is passed to a BRNN to capture the temporal features. The proposed BRNN comprises a single layer with 512 recurrent units in both directions. To benefit from our multimodal input data, we passed combined demographic and CS vectors to a 2-layered fully connected neural network with layer sizes of 32 and 16 neurons. The output of the last layer of this deep neural network (DNN) (i.e., 16 feature vectors) was then concatenated with the output of the BRNN before further processing a network with three dense layers having 128, 64, and 2 hidden units, respectively. The output from the final layer's two units describes the chance of the patient progressing to AD or not, and it does this by outputting probabilities between 0 and 1. The selection of convolution and dense layers in the proposed network architecture is a result of extensive experimentation with various architectures. In addition, the proposed network was further evaluated by inspecting its internal mechanisms using XAI techniques. We generate attention maps for each slice from the input 3D volumes, and it does this for a total of 110 slices corresponding to the number of channels in the input volume. These maps show the attention from the network at specified locations that contribute to the context of the final prediction results. Therefore, we present 2D attention maps and 3D-rendered brain volumes that show progressive atrophy of brain tissues during the progression of AD. The overall technical steps are visually represented in Fig. 6.

4. Results

We evaluated our model using different modalities and compared its accuracy with that of other state-of-the-art deep models in the literature. We performed four different experiments to investigate the effects of adding multimodal input data: 1) progression detection using MRI modality only, 2) progression detection using MRI and demographic features, 3) progression detection using MRI and CSs, and 4) progression detection using MRI, demographics, and CSs. Moreover, we investigated the effect of adding more time steps to the input data on model accuracy. A stratified 5-fold CV technique was employed for training and testing. To find the best classifiers and avoid data leakage, MRI scans from the training datasets were not reused during the testing procedure. We calculated and compared the average of each model achieved using four different evaluation metrics: accuracy, precision, recall, and AUC.

4.1. Dataset

The ADNI dataset was used for this study, which is an open-source platform [49]. The ADNI program was initiated in 2003 as a public-private partnership with a capital budget of $60 million for five years. The main purpose of the ADNI was to determine whether collecting and recording serial MRI, PET, and other clinical tests, biomarkers, and neuropsychological assessment data would allow researchers to track the progression of MCI and pinpoint the development of AD as early as possible. Identifying crucial biomarkers of the early progression of AD will prove beneficial to researchers and doctors attempting to develop novel therapies while monitoring their efficacy. Progress in this area would also reduce the time and expense that clinical trials would incur on subjects.

The ADNI dataset comprises various critical modalities (i.e., patient demographics, CSs, and MRI). It has been used extensively in the literature [42–44]. Most studies focus on BL data without considering the sequential aspect of time-series data. In addition, the ADNI dataset collects patient data on a regular basis (i.e., every six months). We noticed that the timespan between the longitudinal time steps in other datasets such as the National Alzheimer's Coordinating Center (NACC) [51], Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) [52], and Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) [52] is not regular (e.g., one year or six months). Our study was based on 564 MRI volumes at three time steps, that is, a total of 1692 (564 × 3) MRI volumes. We used 3T T1-weighted anatomical sequences recorded using the volumetric 3D MPRAGE protocol, with a 1 × 1 × 1 mm voxel resolution. All MRI volumes in our study were passed through a standard preprocessing pipeline, as shown in Stage 1 of Fig. 2. Among the 564 MRI volumes, 282 subjects were classified as CN at all time points. One hundred subjects were CN at their BL visit but converted to AD within 3 years following the final visit in M12 (i.e., at M48). At the same time, 182 participants were classified as having AD during all visits. We combined 100 converted participants with AD. In this way, the AD class had 182 subjects who had AD from BL until M48, while the 100 subjects were those who were CN at the beginning and converted to AD at M48. These converted subjects are referred to as *converted/progressed to AD*. Similar to other related studies

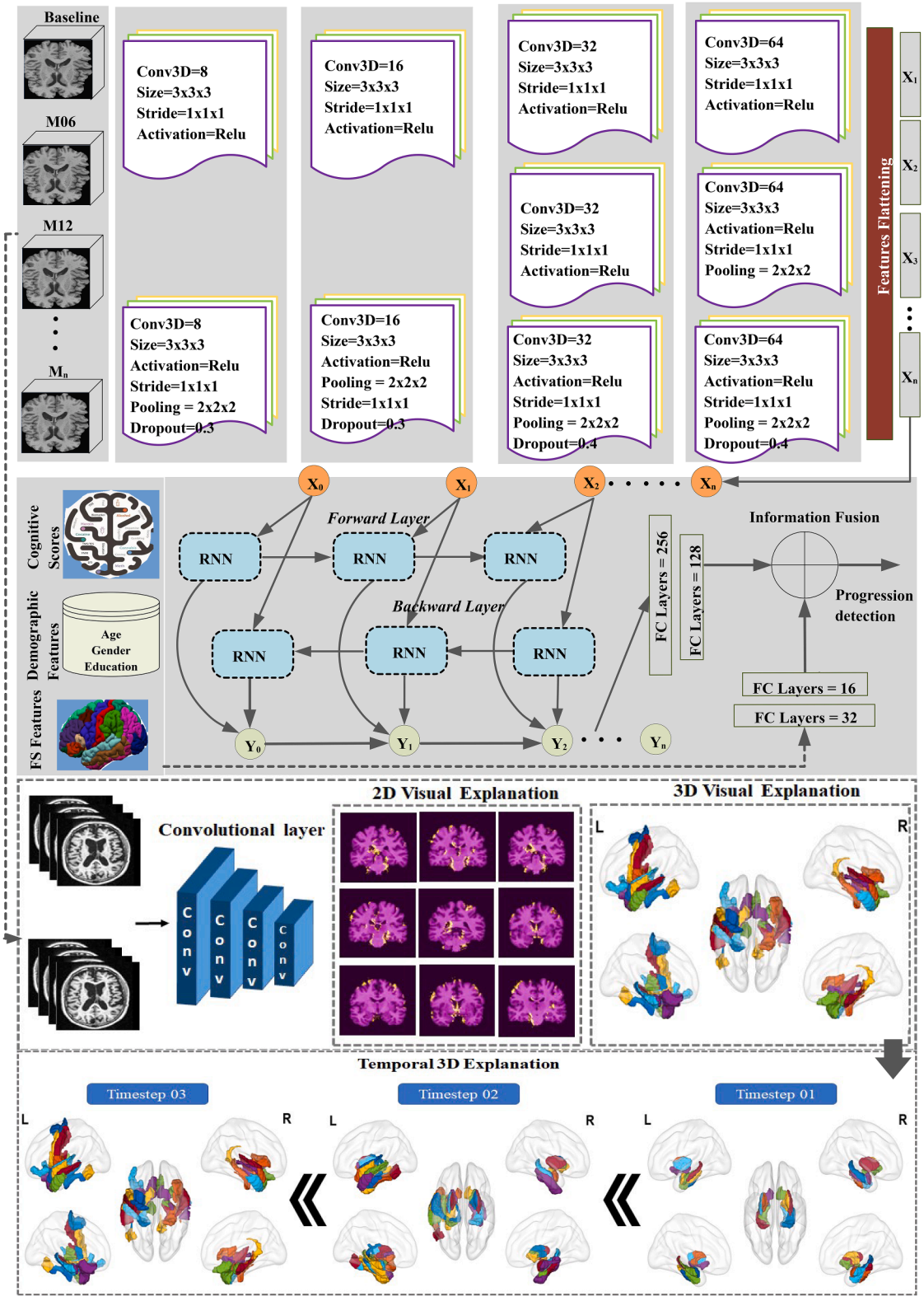

Fig. 6. Proposed framework for AD progression detection using longitudinal data and time-series visual explainability.

on AD progression detection [27,15], we combined converted and AD subjects. The resulting dataset was class-balanced and comprised 282 subjects in the CN and AD classes. We utilized three modalities from the ADNI dataset. The first is demographic data, which include three features: gender, age, and education. The second modality is the CS and selected critical biomarkers, including 13 features (e.g., APOE4, ADAS13, FAQ, MMSE, and four different types of RAVLT scores). The third modality is 3D MRI neuroimaging, which is prepared by employing the ADNI dataset using the FreeSurfer[1] brain image analysis package.

[1] FreeSurfer: http://surfer.nmr.mgh.harvard.edu

Table 2
Mean standard deviation of the 17 selected demographic and CSs at BL visit of CN, converted, and AD patients.

Variable	Meaning	CN ($n = 282$)	Converted ($n = 100$)	AD($n = 182$)
ADAS-13	13-item AD assessment scale	05.34±03.16	07.15±03.14	20.64±7.58
FDG	Fluorodeoxyglucose	0.30±00.50	0.50±0.50	0.88±0.72
TAU	A Protein	937.27±663.5	488.15±535.1	532.0 ± 401
PTAU	A Protein	225.61±248.9	128.96±122.64	300±209.16
CDRSB	Clinical Dementia Rating	41.36±139.11	12.53±12.48	35.72±49.0
MMSE	Mini-Mental State Examination	04.13±05.87	03.63±02.03	9.62±03.99
RAVLT Immediate	Rey Auditory Verbal Learning Test	28.01±08.20	29.42±00.81	22.78±03.03
RAVLT Learning	Rey Auditory Verbal Learning Test	43.36±14.07	42.59±07.64	21.37±08.83
RAVLT Forgetting	Rey Auditory Verbal Learning Test	08.65±10.22	05.57±02.12	02.57±04.46
RAVLT Percentage Forgetting	Rey Auditory Verbal Learning Test	06.02±10.48	03.52±02.65	07.80±17.20
FAQ	Functional Activities Questionnaire	69.26±40.48	103.34±51.58	159.74±107.62
MOCA	Montreal Cognitive Assessment	07.62±31.41	0.37±00.99	15.77±24.64
Hippocampus	Hippocampus Volume	11.70±29.18	93.82±61.61	31.29±92.92
AGE	Age	72.70±05.70	74.50±04.02	74.57±08.18
PTGENDER	Gender	0.41±00.49	0.49±00.50	0.54±00.49
PTEDUCAT	Education	16.59±02.54	15.78±02.80	15.54±02.78
APOE4	No. of ε4 alleles of APOE	01.10±00.75	0.30±00.68	0.27±00.76

Table 3
Comparison of the proposed network with other deep models using only MRI input data.

DL Model	Times Steps	Mean Accuracy	Mean Precision	Mean Recall	mAUC
3DResNet18 [43]	BL	0.71±0.01	0.79±0.06	0.62±0.12	0.71±0.02
	BL∼M06	0.74±0.02	0.86±0.02	0.60±0.06	0.74±0.03
	BL∼M12	0.77±0.04	0.86±0.01	0.67±0.09	0.74±0.03
3D VGG13 [44]	BL	0.81±0.02	0.85±0.04	0.76±0.08	0.81±0.02
	BL∼M06	0.82±0.08	0.85±0.09	0.78±0.07	0.82±0.14
	BL∼M12	0.99±0.07	0.99±0.05	0.99±0.03	0.99±0.07
Proposed Network	BL	0.84±0.01	0.87±0.01	0.83±0.03	0.84±0.01
	BL∼M06	0.84±0.02	0.86± 0.04	0.84±0.04	0.84±0.01
	BL∼M12	0.86±0.02	0.89±0.03	0.85±0.03	0.86±0.01

The first two modalities were collected only during the BL visit. These features have been extensively used in the literature and are mainly used by experts in the medical domain [21,54]. All chosen features have been linked to the development of AD in the literature and are summarized in Table 2.

4.2. Experiment 1: progression detection using MRI modality

In Experiment 1, we utilized only the MRI modality in the training and testing processes of each model (i.e., the proposed network, 3D ResNet18, and 3D VGG13). We then recorded the output of several evaluation metrics, namely, the average accuracy, average precision, average recall, and average AUC, to comprehensively evaluate each model. Furthermore, we compared the AUC of each model to investigate the effect of adding more training data from longitudinal time steps and its impact on the stability of a network.

Results using MRI modality: Table 3 shows the experimental results of the proposed network compared to 3D ResNet18 and 3D VGG13. We conducted various experiments to analyze the effect of adding more data in subsequent time steps, that is, data from BL, BL∼M6, and BL∼M12 visits. Currently, there are three time steps from the available longitudinal data for each patient. Missing data points were handled using forward and backward filling techniques.

At the BL, our proposed network outperformed 3D ResNet18 and 3D VGG13 by achieving a mean accuracy of 0.84±0.01%, mean precision of 0.87±0.01%, mean recall of 0.83±0.03%, and mAUC of 0.84±0.01%. At the BL, 3D ResNet18 achieved a mean accuracy of 0.71±0.01%, mean precision of 0.79±0.06%, mean recall of 0.62±0.12%, and mAUC of 0.71±0.02%; 3D VGG13 achieved a mean accuracy of 0.81±0.02%, mean precision of 0.85±0.04%, mean recall of 0.76±0.08%, and mAUC of 0.81±0.02%. By analyzing MRI data from two time steps, that is, adding the longitudinal data from BL∼M6 visits, the proposed network again outperformed all other networks by achieving a mean accuracy of 0.84±0.02%, mean precision of 0.86±0.04%, mean recall of 0.84 ±0.04%, and mAUC of 0.84±0.01%. By contrast, 3D ResNet18 achieved a mean accuracy of 0.74±0.02%, mean precision of 0.86±0.02%, mean recall of 0.60±0.06%, and mAUC of 0.74±0.03%. 3D VGG13 performed better than 3D ResNet18, achieving a mean accuracy of 0.82±0.08%, mean precision of 0.85±0.09%, mean recall of 0.78±0.07%, and mAUC of 0.80±0.14%. Finally, analysis of MRI data from all three time steps revealed that 3D VGG13 achieved a significant improvement in all evaluation metrics, achieving a mean accuracy of 0.99±0.07%, mean precision of 0.99±0.05%, mean recall of 0.99±0.02%, and mAUC of 0.99±0.07%. Our proposed network achieved a mean accuracy of 0.86 ±0.02%, mean precision of 0.89±0.03%, mean recall of 0.85±0.03%, and mAUC of 0.86±0.01%. 3D ResNet18 achieved a mean accuracy of 0.77±0.04%, mean precision of 0.86±0.01%, mean recall of 0.67 ±0.09%, and mAUC of 0.74±0.03%. Using three time steps, 3D VGG13 achieved the best results, but the model variance remained high. The explanation for these results could be that adding more time steps to the other models, especially 3DVGG13, causes the model's performance to fluctuate owing to overfitting to the noise presented from these extra time steps. However, this was not the case in our model. The statistics show that the proposed network achieves stability over all the time steps. This indicates that the proposed network can identify affected brain regions and become more confident as it is exposed to new data of the same subjects collected sometime after the BL data.

Comparison of the DL models using MRI modality: Fig. 7 shows a performance comparison of the proposed network when using only the MRI modality with AUC as the metric. The performance of the proposed network was compared with that of other classic deep neural networks, 3D ResNet18 and 3D VGG13. The evaluation matrix used in the performance comparison was based on the mAUC. The performance of each network was investigated based on longitudinal input data.

As shown in Fig. 7, the proposed model achieves 84% mAUC using the BL data compared with 3D ResNet18 and 3D VGG13, which achieve

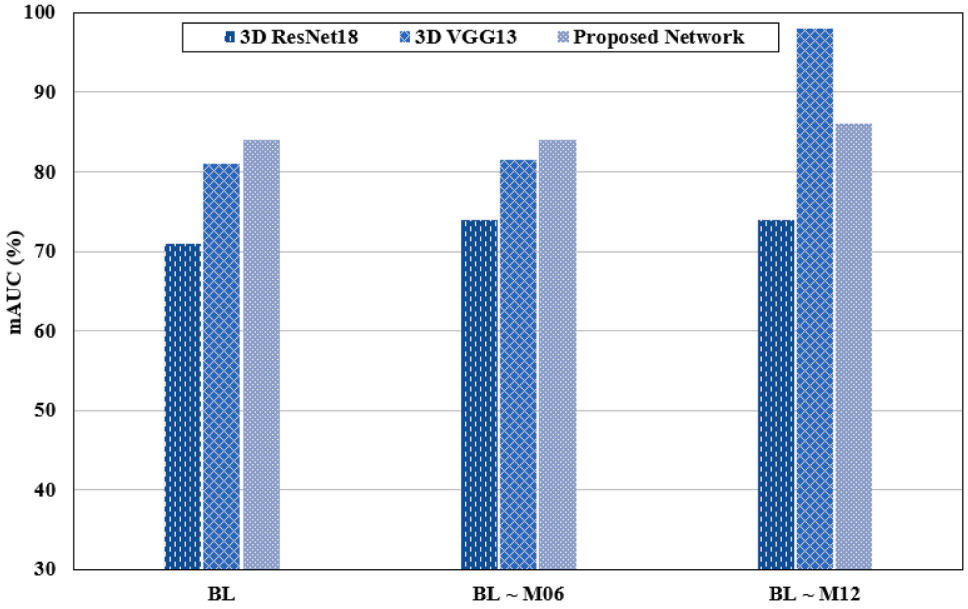

Fig. 7. Performance comparison of the proposed framework with 3D VGG13 and 3D ResNet18 using only MRI modality.

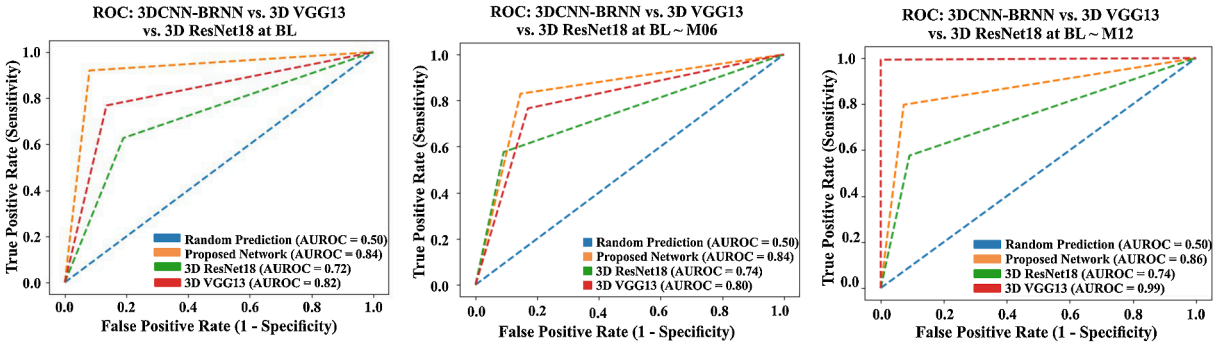

Fig. 8. ROC curves for the proposed framework, 3D VGG13, and 3D ResNet18 using only MRI modality.

71% and 81%, respectively. We also investigated the effect of longitudinal data on the performance of AD progression detection models. In the proposed model, we passed the input volume from the 3D CNN for each time step, that is, BL + M06, and flattened the feature vector after the final convolution layer. After obtaining the N feature vector for each time step, we forwarded these two sequences to the BRNN to compute the progression of AD over a time span of six months. Thus, we improved the mAUC for 3D ResNet18 by 4%, achieving 74%. We did not observe any significant improvements in the proposed network or 3D VGG13. Finally, by evaluating each model using input data from three steps, that is, BL + M06 + M12, over a 12-month time span, we found that our model achieved a 2% improvement to 86% mAUC. With this input data setup, 3D VGG13 significantly improved and outperformed all models by achieving a 99% mAUC.

We further studied the diagnostic ability of each model by estimating class-specific ROC curves. Fig. 8 shows the cross-validated AUCs of the models using the MRI modality at only BL, BL + M06, and BL + M06 + M12. The reason for a single deviated point in the AUC curve is the logits obtained in the final layer of the proposed network. We used the cross-entropy loss function from the PyTorch library, which applies the Sigmoid or SoftMax activation function at the last layer based on the number of classes during the training step [55]. We used the same logits to calculate the AUC curve, as shown in Fig. 8. The sensitivity of the proposed network was 84.2% compared to that of 3D VGG13 and 3D ResNet18, which achieved 81.7% and 71.9%, respectively. The true positive rate is expected to improve as the network starts receiving data

Table 4
Comparison of proposed network and other deep models using multimodal data (MRI + demographics).

DL Model	Times Steps	Accuracy	Precision	Recall	mAUC
3D ResNet18 [43]	BL	0.74±0.04	0.79±0.04	0.70±0.09	0.74±0.05
	BL~M06	0.75±0.04	0.84±0.04	0.73±0.04	0.74±0.04
	BL~M12	0.73±0.04	0.77±0.08	0.76±0.05	0.73±0.03
3D VGG13 [44]	BL	0.82±0.03	0.85±0.04	0.77±0.05	0.82±0.02
	BL~M06	0.80±0.07	0.84±0.08	0.75±0.09	0.81±0.07
	BL~M12	0.82±0.04	0.87±0.04	0.75±0.08	0.82±0.04
Proposed Network	BL	0.88±0.13	0.89±0.14	0.88±0.10	0.88±0.1
	BL~M06	0.90±0.07	0.90±0.06	0.92±0.07	0.90±0.03
	BL~M12	0.96±0.02	0.99±0.01	0.92±0.05	0.96±0.01

from subsequent time steps from the same patient. Using BL + M06, the sensitivity of 3D ResNet18 improved to 74.2%, whereas it remained the same for the other two models. Furthermore, using BL + M06 + M12 led to a significant increase in the sensitivity achieved by our network.

4.3. Experiment 2: progression detection using MRI and demographic features

Experiment 2 shows the effect of combining two modalities (i.e., MRI and demographics) to investigate the contribution of demographic features in the progression of AD.

Results using MRI and demographic features: Table 4 shows the effect of utilizing multimodal input data on the detection of AD progression. Three demographic features were combined with MRI features during the training phase when only BL MRI images were used.

The three demographic features were patient age, gender, and education level. Combining demographic features with MRI features significantly improves the proposed network compared with using only the input data from the MRI modality. At the BL, our proposed model outperformed 3D ResNet18 and 3D VGG13 by achieving a mean accuracy of 0.88±013, mean precision 0.89±0.14, mean recall 0.88±0.10, and mAUC of 0.88±0.1. 3D ResNet18 achieved a mean accuracy of 0.74 ±0.04, mean precision 0.79±0.04, mean recall 0.70±0.09, and mAUC of 0.74±0.05. 3D VGG13 achieved a mean accuracy of 0.82±0.03, mean precision 0.85±0.04, mean recall 0.77±0.05, and mAUC of 0.82±0.02. Furthermore, by using longitudinal data from BL~M6, only 3D ResNet18 and the proposed network improved their overall accuracy. The proposed model again outperformed all other models by achieving a mean accuracy of 0.90±0.07%, mean precision of 0.90±0.06%, mean recall of 0.92±0.07%, and mAUC of 0.90±0.03%. Finally, with three time steps of data, that is, (BL~M12), 3D ResNet18, and 3D VGG13 could not improve and were affected by the noise in the training data. Particularly 3D VGG13 In particular, the overall accuracy of 3D VGG13 degraded compared with the performance on MRI modality only because of either interpreting the fused demographic features as noise or lacking epochs to converge for better performance. For BL~M12, 3D ResNet18 achieved a mean accuracy of 0.73±0.04%, mean precision of 0.77±0.08%, mean recall of 0.76±0.05%, and mAUC of 0.73±0.03, whereas 3D VGG13 achieved a mean accuracy of 0.82±0.04%, mean precision of 0.87±0.04%, mean recall of 0.75±0.08%, and mAUC of 0.82±0.04%. However, the response of our model for BL ~ M12 surpassed all other models by achieving a mean accuracy of 0.96 ±0.02%, mean precision of 0.99±0.01%, mean recall of 0.92±0.05%, and mAUC of 0.96±0.01%. The proposed model significantly improved its accuracy and reduced the variance at (BL+M12) by demonstrating a more accurate and reliable performance.

Comparison of the DL models using MRI and demographic modalities: Fig. 9 shows the effect of multimodal data on detecting the progression of AD. These models used MRI data fused with the patients' demographic features. The demographic features included marital status, age, and education level.

Overall, after fusion, the mAUC increased, and each model improved the mAUC as it was exposed to more time-step data of the same patents. The proposed model achieved 88% mAUC compared with 3D ResNet18 and 3D VGG13, which achieved 74% and 82%, respectively. Using data from two time steps, that is, BL and M06, 3D ResNet18 achieved 74%, 3D VGG13 achieved 81%, and the proposed network achieved 90%. Finally, using data from three time steps, our model outperformed all other models by achieving a 96% mAUC compared with 3D ResNet18 and 3D VGG13, which achieved mAUCs of 73% and 84%, respectively.

Fig. 10 illustrates the diagnostic ability of each model by estimating class-specific ROC curves. The figure shows the cross-validated AUC results based on the longitudinal MRI data, that is, BL, BL + M06, and BL + M06 + M12 MRI data fused with demographic features based on data collected at the BL. The sensitivity of the proposed network was 88%, whereas those of 3D VGG13 and 3D ResNet18 were 82.1% and 74.4%, respectively. Furthermore, the true positive rate of the proposed network improved when using training data from two time steps, that is, BL + M06 fused with demographic features from the BL. Our model achieved 89.8% accuracy using data from two time steps, which is a 1.89% improvement compared with using only BL + demographic data. Moreover, the proposed model using BL + M06 input data outperformed the other networks, 3D VGG13 and 3D ResNet18, which achieved 80.7% and 74.4%, respectively. Finally, using the BL + M06 + M12 + demographic features, the proposed network outperformed all other models and data combinations, achieving a significant improvement of approximately 6%, reaching an AUROC of 96%. The AUROC from the other models, that is, 3D VGG13 and 3D ResNet18, was not stable and achieved values of 82.2% and 73%, respectively.

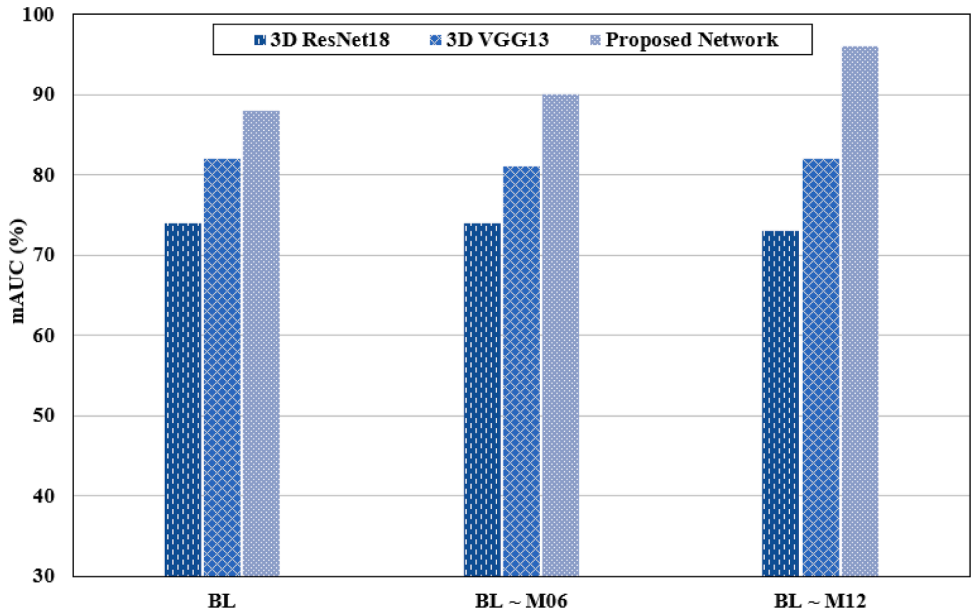

Fig. 9. Performance comparison of the proposed framework with 3D VGG13 and 3D ResNet18 using two modalities (MRI + demographics).

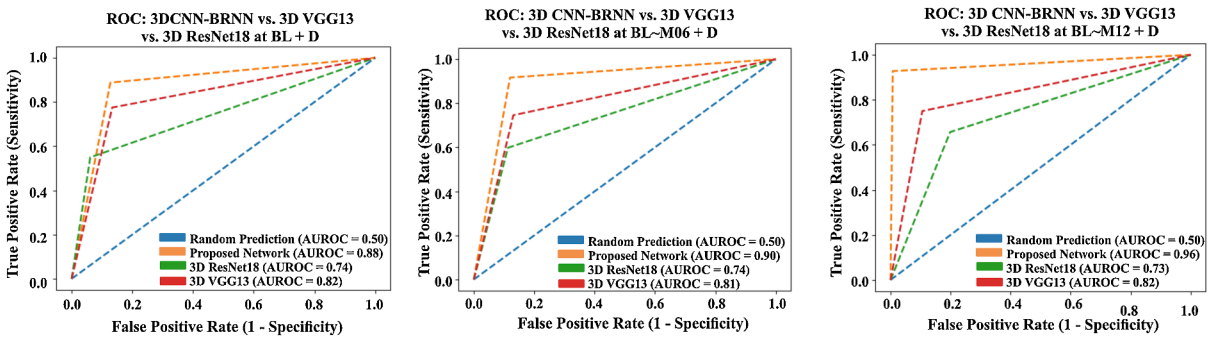

Fig. 10. ROC curves of the proposed framework, 3D VGG13, and 3D ResNet18 using MRI and demographic features.

Table 5
Comparison of proposed network with other deep models using multimodal data (MRI + CSs).

DL Model	Times Steps	Accuracy	Precision	Recall	mAUC
3D ResNet18 [43]	BL	0.74±0.04	0.88±0.04	0.70±0.05	0.73±0.05
	BL~M06	0.75±0.02	0.85±0.07	0.71±0.06	0.73±0.02
	BL~M12	0.78±0.04	0.88±0.06	0.76± 0.05	0.78±0.02
3D VGG13 [44]	BL	0.82±0.02	0.89±0.04	0.77±0.05	0.82±0.02
	BL~M06	0.89±0.04	0.87±0.04	0.85±0.04	0.89±0.03
	BL~M12	0.88±0.01	0.91±0.03	0.85±0.03	0.89±0.02
Proposed Network	BL	0.89±0.05	0.93±0.04	0.88±0.05	0.88±0.05
	BL~M06	0.90±0.04	0.90±0.03	0.96±0.04	0.94±0.02
	BL~M12	0.96±0.02	0.95±0.01	0.92±0.03	0.91±0.03

4.4. Experiment 3: progression detection using MRI and CSs

To investigate the progression of AD by analyzing the cognitive abilities of patients, we trained our model with CSs fused with MRI data. This combination of data allowed the model to detect essential patterns from each modality known for disease progression. Furthermore, the detection abilities of different models were compared using the AUC metric, as discussed in the previous sections.

Results using MRI and CSs: Table 5 shows the effects of combining CSs with MRI as input data. The CSs were recorded at the BL and are summarized in Table 2. These include genetic biomarkers (APOE4), physical biomarkers (e.g., hippocampal volume), and behavioral test statistics (e.g., ADAS13, FAQ, MMSE, and four different types of RAVLT scores). By fusing CSs with MRI longitudinal data, our proposed model outperformed all the other models, that is, 3D ResNet18 and 3D VGG13, in all data combinations. Our model achieved a mean accuracy of 0.89 ±0.05%, mean precision of 0.93±0.04%, mean recall of 0.88±0.05%, and mAUC of 0.88±0.5% at the BL. 3D ResNet18 achieved a mean accuracy of 0.74±0.04%, mean precision of 0.88±0.04%, mean recall of 0.70±0.05%, and mAUC of 0.73±0.05%. 3 G VGG13 achieved a mean accuracy of 0.82±0.02%, mean precision of 0.89±0.04%, mean recall of 0.77±0.05%, and mAUC of 0.82±0.02% at the BL.

Further, each model improved in accuracy when using two time steps of training data, and we chose data from BL ∼ M6. Our model outperformed other models by achieving a mean accuracy of 0.90±0.04%, mean precision of 0.90±0.03%, mean recall of 0.96±0.04%, and mAUC of 0.94±0.02%. Moreover, the proposed model showed significant improvements when using three time steps of input data, i.e., BL ∼ M12, by achieving a mean accuracy of 0.96±0.02%, mean precision of 0.95 ±0.01%, mean recall of 0.92±0.03%, and mAUC of 0.91±0.03%. 3D ResNet18 achieved a mean accuracy of 0.78±0.04%, mean precision of 0.88±0.06%, mean recall of 0.76±0.05%, and mAUC of 0.78±0.02%. 3D VGG13 achieved a mean accuracy of 0.89±0.04%, mean precision of 0.87±0.04%, mean recall of 0.85±0.04%, and mAUC of 0.89±0.03% for BL ∼ M12. According to the results in Table 5, the proposed network has more stable performance as the number of steps changes and outperforms the other models with various input data combinations. This indicates that the proposed network can identify brain regions affected by AD, and its output becomes more confident as it is exposed to new data from subsequent time steps. As previously noted, adding more time steps improves the performance of all DL models. In addition, adding more data increased the stability of all the models by reducing the fluctuation in variance. However, using data from only two time steps increased the noise in our model even though the overall model performance was improved.

Comparison of the DL models using MRI and CSs: Fig. 11 shows the effect of multimodal data on the progression of AD. The multimodal data combined the MRI data with the patient's CSs. Our model outperformed all the other models when using MRI data from three time steps, achieving 89%, 94%, and 91% mAUC for BL, BL + M06, and BL + M06 + M12 MRI data, respectively. 3D ResNet18 achieved 73%, 73%, and 78%, whereas 3D VGG13 achieved 82%, 89%, and 88% mAUC when using BL, BL + M06, and BL + M06 + M12 MRI input data, respectively.

Fig. 12 shows the cross-validated AUC results for models using longitudinal MRI data, that is, BL, BL + M06, and BL + M06 + M12 MRI data fused with cognitive features from the BL. The sensitivity of the proposed network was 88% compared with that of 3D VGG13 and 3D ResNet18, which achieved 82.1% and 74.4%, respectively. Furthermore, the true positive rate of the proposed network improved when using training data from two time steps, that is, BL + M06 data fused with demographic features from the BL. Using data from two time steps, our model achieved 89.8%, which is a 1.89% improvement compared with using only BL + demographic data. Moreover, the proposed model using only BL + M06 data still outperformed the other networks, 3D VGG13 and 3D ResNet18, which achieved 80.7% and 74.4%, respectively. Finally, using BL + M06 + M12 data + demographic features, the proposed network outperformed all the other models and input data combinations. It significantly improved by approximately 6%, reaching an AUROC of 96%. By contrast, 3D VGG13 and 3D ResNet18 did not achieve stability, obtaining AUROC scores of 82.2% and 79.9%, respectively.

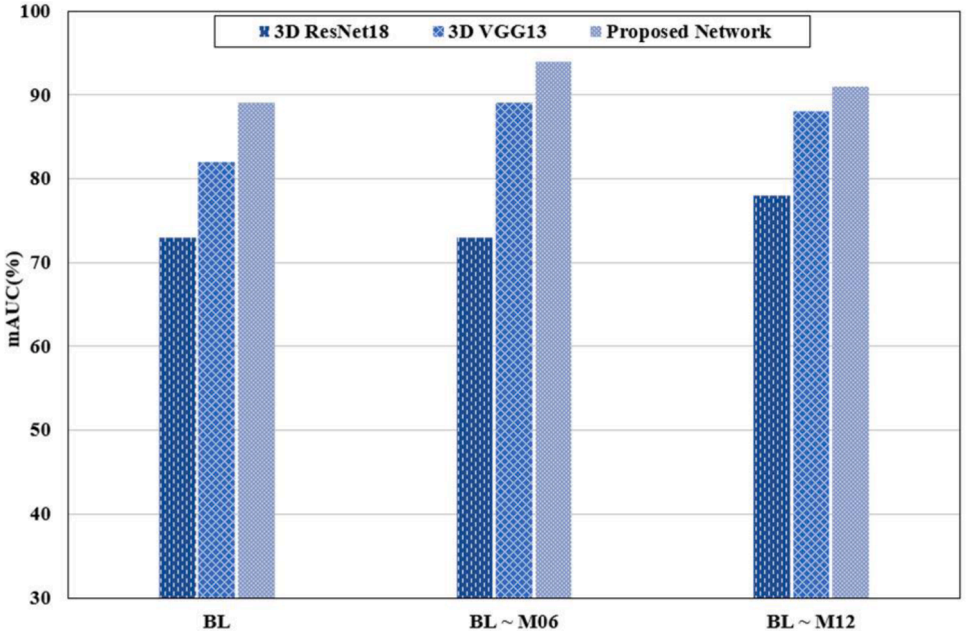

Fig. 11. Performance comparison of the proposed framework with 3D VGG13 and 3D ResNet18 using MRI and CSs.

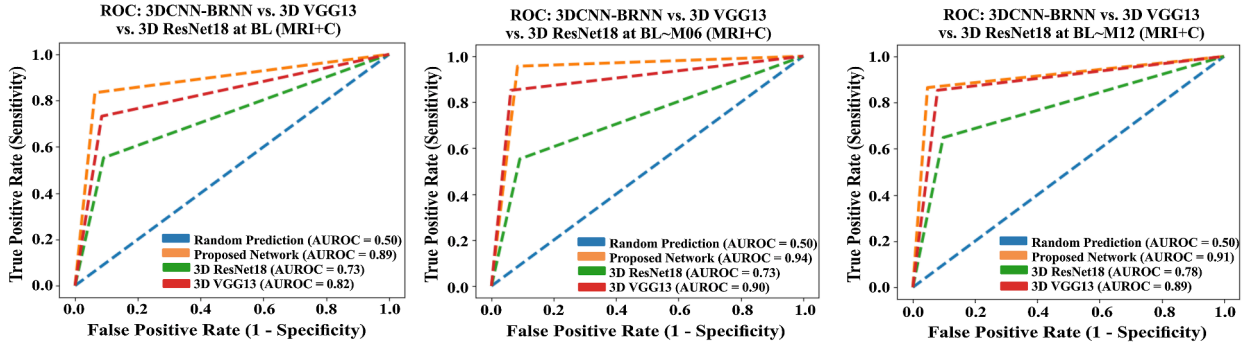

Fig. 12. ROC curves of the proposed framework, 3D VGG13, and 3D ResNet18 using MRI and CSs.

Table 6
Comparison of proposed network with other deep models using the three modalities.

DL Model	Times Steps	Accuracy	Precision	Recall	mAUC
3D ResNet18 [43]	BL	0.72±0.04	0.78±0.02	0.69±0.04	0.72±0.04
	BL~M06	0.74±0.03	0.84±0.02	0.67±0.04	0.73±0.03
	BL~M12	0.75±0.03	0.85±0.06	0.69±0.05	0.75±0.02
3D VGG13 [44]	BL	0.83±0.03	0.87±0.02	0.78±0.06	0.83±0.03
	BL~M06	0.86±0.04	0.91±0.04	0.80±0.07	0.86±0.03
	BL~M12	0.92±0.03	0.98±0.02	0.87±0.03	0.93±0.01
Proposed Network	BL	0.93±0.01	0.96±0.02	0.88±0.03	0.92±0.01
	BL~M06	0.95±0.02	0.95±0.02	0.92±0.04	0.90±0.01
	BL~M12	0.92±0.01	0.92±0.03	0.92±0.05	0.93±0.01

4.5. Experiment 4: progression detection using MRI, demographics, and CSs

In Experiment 4, we fused all three modalities (MRI, demographics, and CSs) and investigated the effect of multimodalities on AD progression detection. We evaluated the performance of the three comparative models, that is, the proposed network, 3D VGG13, and 3D ResNet18. We also investigated the effects of combining multimodal data and longitudinal MRI on the stability and robustness of the model.

Results using MRI, CSs, and demographic: Table 6 shows the combined effect of fusing all input data modalities, including BL demographic and CS data, with the time-series data from the MRI modality. With such a setup, each model drastically boosted the accuracy. However, the proposed model still outperformed all other models with all data combinations.

At the BL, our proposed model achieved a 0.93±0.01 mean accuracy, 0.96±0.02 mean precision, 0.88±0.03 mean recall, and 0.92±0.01 mAUC. 3D ResNet18 achieved a 0.72±0.04 mean accuracy, 0.78±0.02 mean precision, 0.69±0.04 mean recall, and 0.72±0.04 mAUC. 3D VGG13 achieved a 0.83±0.03 mean accuracy, 0.87±0.02 mean

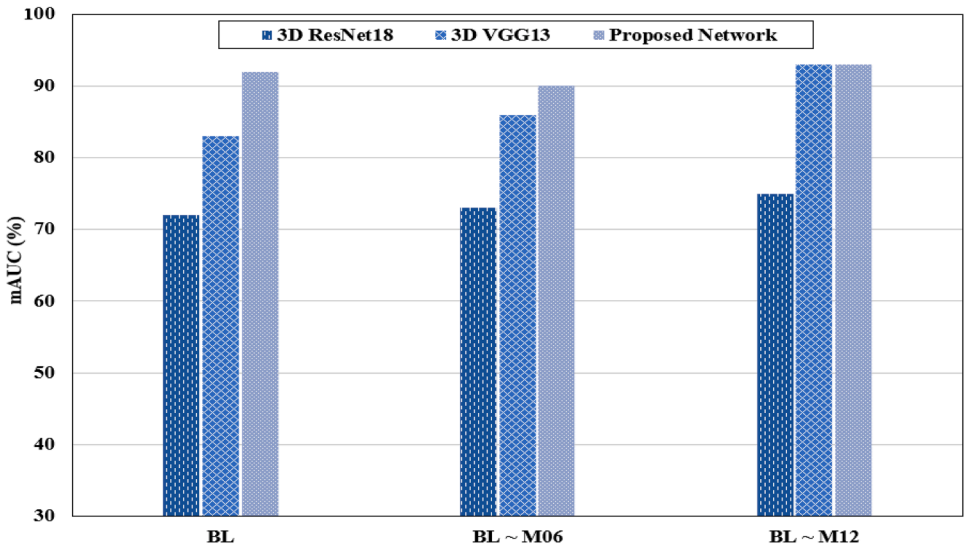

Fig. 13. Comparison of the proposed framework with 3D VGG13 and 3D ResNet18 using MRI, demographics, and CSs.

precision, 0.78±0.06 mean recall, and 0.83±0.03 mAUC at the BL. Each model saw improvements in accuracy with two time steps of training data, that is, BL ~ M6. The proposed model outperformed the other models by achieving a 0.95±0.02 mean accuracy, 0.95±0.02 mean precision, 0.92±0.04 mean recall, and 0.90±0.01 mAUC. At BL ~ M6, 3D ResNet18 achieved a 0.74±0.03 mean accuracy, 0.84±0.02 mean precision, 0.67±0.04 mean recall, and 0.73±0.03 mAUC. 3D VGG13 achieved 0.86±0.04 mean accuracy, 0.91±0.04 mean precision, 0.80 ±0.07 mean recall, and 0.86±0.03 mAUC. Moreover, the proposed model significantly improved with three time steps of data, i.e., BL~ M12, obtaining a 0.92±0.01 mean accuracy, 0.92±0.03 mean precision, 0.92±0.05 mean recall, and 0.93±0.01 mAUC. However, 3D ResNet18 achieved a 0.75±0.03 mean accuracy, 0.85±0.06 mean precision, 0.69 ±0.05 mean recall, and 0.75±0.02 mAUC. 3D VGG13 achieved a 0.92 ±0.03 mean accuracy, 0.98±0.02 mean precision, 0.87±0.03 mean recall, and 0.93±0.01 mAUC at BL~ M12. The described experiments confirmed that the proposed method is stable from the beginning but improves as it is exposed to subsequent time steps in the longitudinal data. Furthermore, adding multimodal training and testing data makes each model more accurate in recognizing AD progression.

Comparison of DL models using MRI, demographic, and CS data: Fig. 13 shows a comparison of the proposed network against the other networks, using MRI, demographics, and CS input data from different time steps. These models consider all numerical features available by including demographics and CSs; that is, three features from demographics and 14 features from CSs. The proposed model achieved very stable results using data from each time step by achieving 92%, 90%, and 93% mAUC when using the BL, BL + M06, and BL + M06 + M12 MRI data, respectively. 3D VGG13 achieved 83%, 86%, and 93%, respectively, and 3D ResNet18 achieved 72%, 73%, and 75% mAUC using BL, BL + M06, and BL + M06 + M12 MRI data, respectively.

Fig. 14 shows the cross-validated AUC results when using longitudinal MRI data, that is, MRI data from BL, BL + M06, and BL + M06 + M12 fused with demographic + cognitive features from the BL. Here, we analyze the effect of using longitudinal MRI data along with aggregated features from demographic and CS data.

The sensitivity of the proposed network using BL MRI data and all aggregated features was 92%, outperforming all other models in our comparison, that is, 3D VGG13 and 3D ResNet18. These achieved AUROC values of 83% and 72.4%, respectively. The AUROC achieved when using BL + M12 MRI data degraded to 90.4% for the proposed network, but it still outperformed the other models: 3D ResNet18 (AUROC = 72.5%) and 3D VGG13 (AUROC = 86.1%). This degradation was probably caused by noise in the input training data. Finally, when using BL + M06 + M12 MRI data + aggregated features, the proposed network outperformed all the other models by significantly improving its AUROC to 93%. The other models, 3D VGG13 and 3D ResNet18, achieved AUROCs of 92.7% and 75.4%, respectively.

4.6. *Selecting the best combination of modalities*

Fig. 15 shows the effectiveness of the proposed network in terms of the mAUC from two dimensions, that is, how the proposed model performs as (longitudinal) input data from subsequent time steps are added to the training data. Furthermore, we investigated the effect of utilizing multimodal data on mAUC. As shown in previous experiments, we used

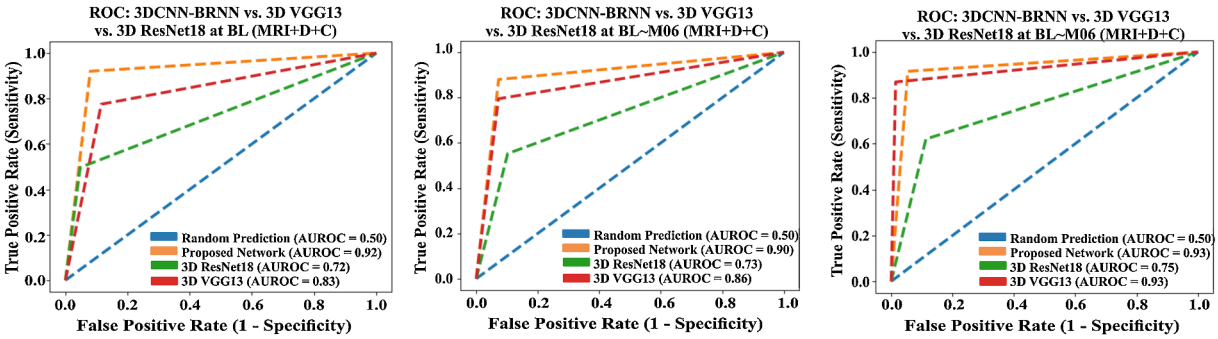

Fig. 14. ROC curves of the proposed framework, 3D VGG13, and 3D ResNet18 using MRI, demographics (D), and CSs (C).

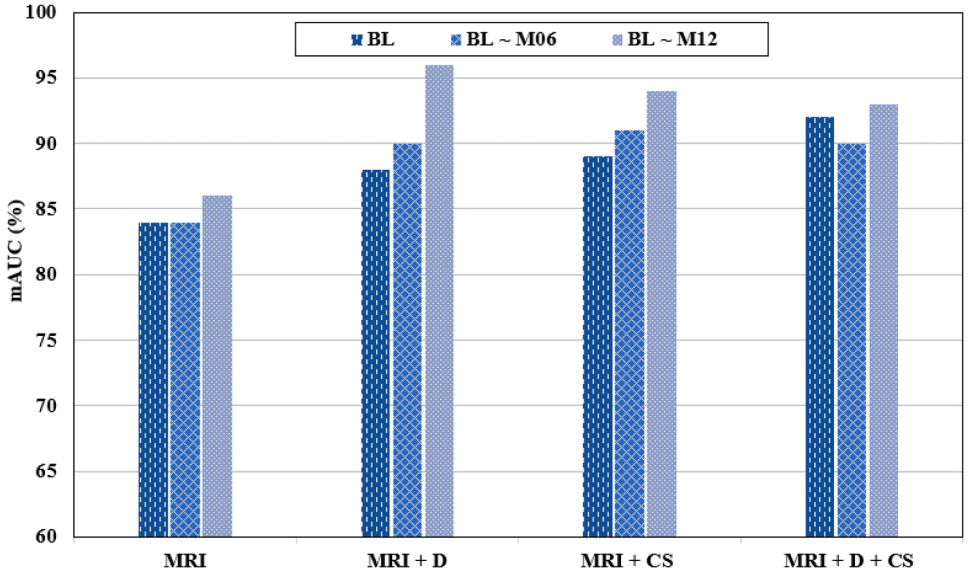

Fig. 15. Performance of the proposed network with respect to multimodal data.

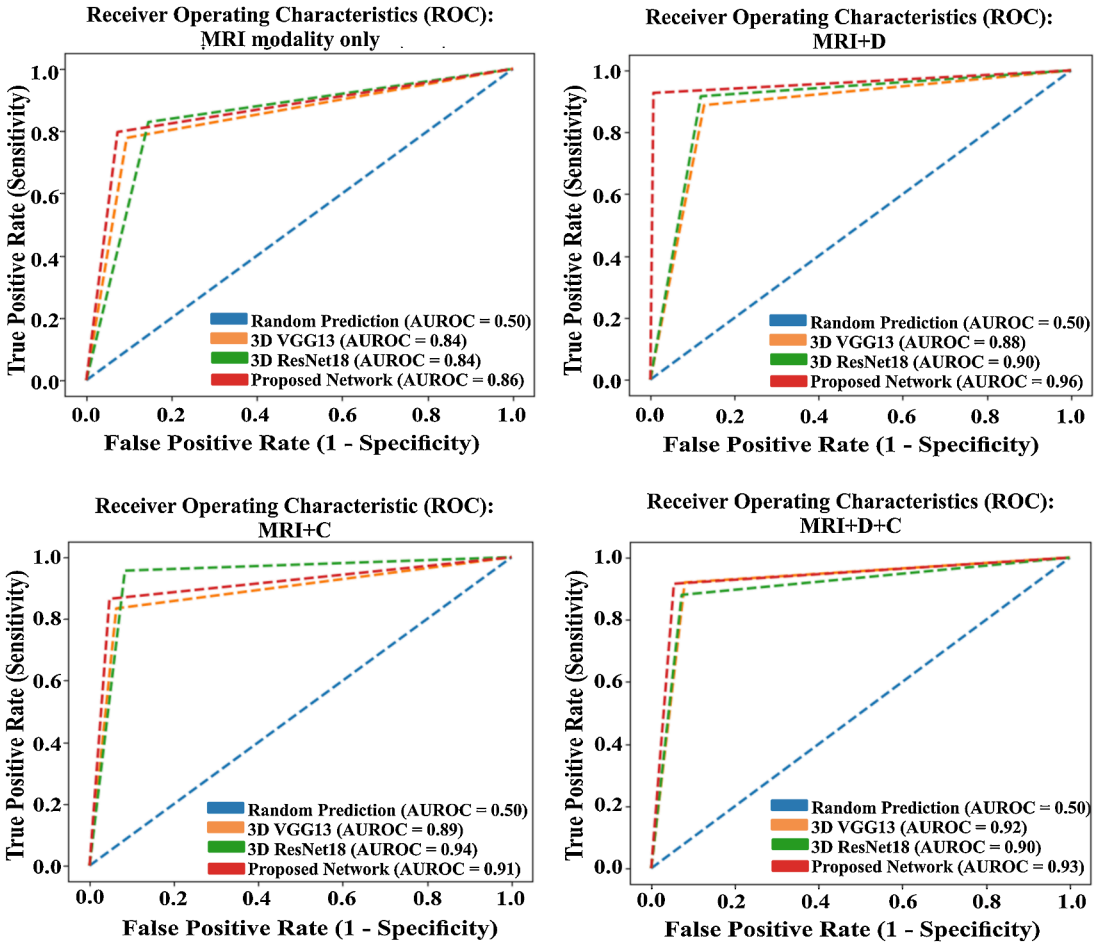

Fig. 16. ROC–AUC curve of the proposed network using multimodal data.

the mAUC as the performance metric for testing and comparing the models because it is proportional to the other metrics.

We calculated results using only BL MRI input data, BL + M06 MRI input data, and BL + M06 + M12 MRI input data. The results show that our model achieved a mAUC of 86% with only BL MRI input data ($p < 0.001$) and BL + M06 MRI input data ($p < 0.005$). However, the performance achieved a significant boost as we added data from the third time step, that is, M12, and the mAUC reached with this input data was 86% ($p < 0.005$). We also investigated the effects of adding demographic features to the input data for progression detection. By adding demographic features recorded at the BL to the BL MRI, BL MRI + M06 MRI data, and BL + M06 + M12 MRI input data, the proposed model achieved 88%, 90%, and 96% mAUC, respectively. Demographic features that we used included age, gender, and education level. In addition, we investigated the effect of fusing CSs with the MRI data. The proposed model achieved mAUCs of 89%, 94%, and 91% by fusing CSs with the respective MRI input data from one, two, and three time steps, respectively. Finally, we investigated the effect of combining all modalities, that is, MRI, demographic feature, and CS data. The proposed model achieved mAUCs of 92%, 90%, and 93% by fusing all additional modalities with the respective MRI input data from one, two, and three time steps, respectively. It should be noted that the proposed model achieved a mAUC of 96% when using MRI data from all three time steps fused with demographic features, outperforming all other combinations of input data.

Fig. 16 shows a sensitivity analysis of the proposed network from two perspectives: (1) from a time-series perspective with only MRI input data, that is, scan data from BL, BL∼M06, and BL∼M12 and (2) from a multimodal input data perspective, that is, using MRI, MRI + demographic (D), MRI + CS, and MRI + demographic (D) + CS as the input data. The AUROC achieved by the proposed network using only data from the MRI modality at BL was 84.2%. The accuracy remained the same when adding BL∼M06 and BL∼M12 to the time-series input data, and no significant improvements were made; however, the false positive rate was reduced as the model was given subsequent time step data from the longitudinal MRI data. The model significantly improved the true-positive rate when demographics features were fused with the MRI data. The demographic features utilized here were the patient's age, gender, and education.

The model achieved 88% AUROC by fusing these data with BL MRI data. In addition, with MRI data from two time steps, that is, BL∼M06, the AUROC improved by 1.8% compared with fusing with just BL MRI data that reached an AUROC of 89.8%. This shows that the proposed model can capture temporal features as it sees longitudinal data from subsequent time steps. Using data recorded at three different time steps, the model significantly boosted its true-positive rate and achieved a 96% AUROC. Furthermore, by fusing the MRI data with CSs, the proposed model achieved an AUROC of 88.5% when fused with only BL MRI data. The model achieved a drastic boost when fusing BL∼M06 MRI data, achieving a 93.6% AUROC when receiving data from two time steps. However, the AUROC falls to 90.9% when using all three time steps of MRI data, that is, BL∼M12. Finally, by combining the demographics + CSs recorded at the BL with the MRI data, the model achieved 92%, 90.4%, and 93.1% AUROC when using BL, BL∼M06, and BL∼M12 MRI input data, respectively.

4.7. Comparison with current state-of-the-art methods

In this study, we proposed a novel architecture that improves the performance of existing models for the detection of AD progression. The proposed method utilizes multimodal data to extract critical features from each input modality and utilizes all of them to detect the progression of AD. The 3D CNN used in the proposed network captures intra-slice features from various input MRI volumes of the same patient

Table 7
Comparison of the proposed model with various models from previous studies.

Ref.	Year	Sub#	Modality	Performance				ML Method
				ACC	PRE	SEN	AUC	
[56]	2017	272	FDG-PET	92.16	—	88.46	—	RF-Robust SVM
[52]	2018	1984	50 landmarks extracted and utilized from MRI	93.70	—	94.60	98.60	CNN
[50]	2018	1051	FDG-PET	93.58	—	91.54	—	DNN
[2]	2019	1737	D, CS	—	—	—	96.58	Multilayer perceptron
[57]	2019	1105	FreeSurfer-based numerical features extraction from MRI, PET, and DTI modalities	—	—	—	94.00	LSTM
[26]	2019	830	MRI	91.33	—	86.87	93.22	Stacked-CNN bidirectional gated recurrent unit
[51]	2019	742	Six volumetric MRI biomarkers	—	—	—	90.00	LSTM
[58]	2019	186	Cortical models, cortical metrics, and ADAS cognitive test	100	—	100	—	SVM
[54]	2019	488	66 numerical features (i.e., CS, PET, MRI, and sociodemographic)	—	—	—	87.00	SVM, Kernel Ridge Regression
[59]	2020	449	MRI biomarkers	85.00	—	86.30	94.00	Linear mixed effects
[60]	2020	1677	Hippocampus features and CS	—	—	—	91.00	RNN with filling strategy
[61]	2020	485	Hippocampus segmented regions from MRI volumes	92.00	—	92.00	—	Logistic regression, K-nearest neighbor, SVM, decision tree, and RF
[27]	2020	1536	Statistical measures (MRI, PET, CS, assessment data, and neuropathological data)	92.62	94.02	98.82	—	Stacked CNN bidirectional LSTM
[62]	2020	509	MRI	84.00	—	92	—	CNN-Ensemble learning
[53]	2020	1371	Neuropsychological tests and biomarkers (MRI, PET, CS, neuropathological data, and Comorbidities)	74.55	84.68	84.80	—	Bidirectional LSTM
[7]	2021	1029	Comorbidities, CS, brain disorder and not brain disorder medicine	98.00	99.63	99.70	—	RF, decision tree, logistic regression, SVM, and k-nearest neighbor
[63]	2021	151	MRI	90.00	—	—	96.20	Temporally Structured SVM
[64]	2021	492	Cropped hippocampal regions from MRI	—	—	—	88.00	DeepAtrophy
[65]	2020	216	MRI	88.90	—	—	92.50	3D DenseNet
[66]	2021	1757	16 ADNI and 6 NACC biomarkers (MRI, PET, and CSF)	—	—	—	93.40	Modified Logistic Function
[67]	2022	1371	Neuropsychological tests and biomarkers MRI, CS, Comorbidities, and CSF	93.87	94.07	94.07	—	2-staged AD progression detection
[68]	2022	1500	MRI	94.34	—	—	—	RESU-Net
[28]	2022	400	MRI	90.83	—	95.00	—	FDN-ADNet
[69]	2022	809	MRI, PET, and Single Nucleotide Polymorphism	—	—	—	96.00	Multi-Classification Framework
[70]	2022	559	MRI, PET, CS, neuropathological data, and Comorbidities	98.56	98.56	98.56	98.56	Ensemble Classifier
Ours*	2022	1692	MRI, D, and CS	96.00	99.00	92.00	96.90	3D-CNN-BRNN

(Sub# = Number of subjects, ACC= Accuracy, PRE=Precision, SEN=Sensitivity, AUC=Area under the curve).

at different time steps and combines these with the feature vector extracted from the CSs and demographic data. This combination of a CNN using CSs and demographic input data is enhanced by further processing its outputs using a BRNN to capture intra-slice features over different time steps. The network was evaluated using MRI-only, MRI + CS, MRI + demographics, and MRI + CS + demographics input data. Our framework achieved accurate and stable results using the proposed architecture.

Table 7 shows a comparison of the proposed method with the most recent alternative methods developed for AD progression detection. Most DL-based studies in the field of AD progression detection use time-series data [19,55]. In addition, we found that selecting only a single slice or working with a specific region from the entire MRI volume to detect AD progression is a notable drawback of most existing techniques. Such mechanisms typically result in a wide range of information losses, which are critical for the stability of a model in a disease prediction task. Time-series data analysis is prevalent in AD progression detection [19, 55]; however, the models used in these studies cannot be easily evaluated because of the lack of explainability [50]. In many studies, the problem of AD diagnosis has been addressed as a binary classification task (e.g., CN vs. AD). They use traditional ML-based approaches to distinguish between these two classes. For example, Lu et al. [50] proposed an incomplete RF, robust support vector machine (SVM) approach for the detection of cognitively impaired individuals. They built an incomplete RF model by using FDG-PET image features, modeling its outputs as noise-corrupted feature datasets and minimizing a loss function based on these noisy data within a robust programming framework. They reported an 88.46% sensitivity and 96% specificity with BL data using only a single modality. Liu et al. [52] proposed a multitask multi-channel DL framework by extracting 60 well-known anatomical landmarks from brain MRIs and then using them as image patches to train a DNN model to jointly perform classification and regression. Uysal et al. [61] used targeted data from only the hippocampal brain region by using an ITK-SNAP tool to segment parts of the scans from the rest of the brain volumes. Various ML models for detecting AD progression were then trained using the processed data. Cui et al. [26] proposed a 3D CNN + gated recurrent unit that uses whole-brain MRI as the input data for AD classification. They used time-series data consisting of data collected at six different time steps for each patient and achieved a 92% accuracy using their proposed setup. As mentioned before, they used whole-brain MRI volumes as input data, which are usually considered to contain unnecessary voxels, increasing the network's processing load. In [51], Ghazi et al. proposed a generalized method for training LSTM networks that can handle missing values in both the input and output. They reported that they worked with 63% of the missing data and achieved 90% AUC with their proposed approach. They utilized volumetric features from the MRI modality (i.e., ventricles, hippocampus, whole brain, fusiform, middle temporal gyrus, and entorhinal cortex) from 11 time steps over a one-year period. Albright et al. [2] proposed a technique called the "All-Pairs" method. Using this method, they compared all temporal data-point pairs from each patient and attempted to capture the progression of AD. They conducted their study using 1737 patients from the ADNI database and reported an mAUC of 96%. Hong et al. [57] utilized ROI features, such as volume, cortical thickness, and surface areas extracted by the FreeSurfer tool from MRI, PET, and DTI modalities. They then designed an LSTM model with a fully connected layer and activation layers to encode the temporal relations between the features and the next AD stage. Nguyen et al. [60] proposed an RNN-based AD progression detection model in which missing values were handled using a minimal RNN approach, achieving an accuracy of 91%. El-Sappagh et al. [27] conducted progression detection by using longitudinal data collected at 15 different time steps. However, they did not consider the gap between the last observed data and when the AD classification was made (i.e., it consisted of an estimation problem rather than a forecasting problem). Abuhmed et al. [53] proposed hybrid deep models based on multimodal time-series data from the previous 18 months. They utilized a wide array of features extracted from inputs, including MRI, PET, neuropsychological and CSs, and demographics data, to perform two tasks: AD progression detection and CS prediction. In their work, they used ready-made features provided by the ADNI database. Lahmiri et al. [58] examined several ML models to determine the extent to which the fractal dimensions of the pial, gray/white boundary, and cortical ribbon can be used to classify AD patients and healthy controls. Additionally, they evaluated the feasibility of integrating the fractal dimensions of these cortical models (cortical ribbon, pial surface, and logistic regression gray/white surface) with cortical metrics (cortical thickness and gyrification index). In their study, only the BL data from the ADNI dataset were used, and the accuracy, sensitivity, and specificity were all reported to be 100%. El-Sappagh et al. [7] reported a 98% accuracy by utilizing longitudinal data at four time steps. They used a rich dataset composed of 1029 subjects for each modality, including demographics, CSs, brain disorder medicine, non-brain disorder medicines, and comorbidities or disorders. They investigated the performance of several ML-based algorithms, including the decision tree, logistic regression, SVM, and k-nearest neighbor. Their work was also based on the features already available in the ADNI database.

Table 7 summarizes the five characteristics of each approach described above: number of subjects used in the study, data modality, number of time steps, performance, and architecture used. As noted in this table, the proposed network outperformed most of the other state-of-the-art models. The performance and robustness of the proposed system represent an excellent starting point for building a clinical decision support system for AD progression detection. Furthermore, the model can explain features that influence the final decision. In addition, the proposed model is more intuitive than most state-of-the-art methods. This is because it examines a variety of multimodal longitudinal data during the training process, which is critical for effectively assessing mental health. Although the proposed model outperforms other state-of-the-art DL-based models in the field of AD management, further improvements are still required before it can be deployed to help diagnose actual patients.

Table 8 highlights the uniqueness of the present study compared

Table 8
Singular contributions of the proposed method when compared with state-of-the-art techniques in AD progression detection.

Ref.	Year	A	B	C	D	E	F	G
[56]	2017	×	×	×	×	×	×	×
[52]	2018	×	✓	×	×	×	×	×
[50]	2018	×	×	×	×	×	×	×
[2]	2019	3	×	×	×	×	×	×
[57]	2019	10	✓	×	×	×	×	×
[26]	2019	6	×	✓	✓	×	×	×
[51]	2019	11	×	×	×	×	×	×
[58]	2019	×	×	×	×	×	×	×
[54]	2019	×	✓	×	×	×	×	×
[59]	2020	×	×	×	×	×	×	×
[60]	2020	4	×	×	×	×	×	×
[61]	2020	×	×	×	×	×	×	×
[27]	2020	15	✓	✓	×	×	×	×
[62]	2020	×	×	×	✓	✓	×	×
[53]	2020	4	✓	✓	×	×	×	×
[7]	2021	4	✓	×	×	×	×	×
[63]	2021	5	×	×	×	✓	×	×
[64]	2021	6	×	×	×	×	✓	×
[65]	2021	×	×	×	✓	×	×	×
[66]	2021	3	✓	×	×	×	×	×
[67]	2022	4	✓	×	×	×	×	×
[68]	2022	×	×	×	×	×	×	×
[28]	2022	×	×	×	×	×	×	×
[69]	2022	×	✓	✓	×	✓	×	×
[70]	2022	✓	✓	×	×	×	×	×
Ours*	2022	3	✓	✓	✓	✓	✓	✓

with the studies listed in Table 7. We determined the superiority of the proposed model over the methods proposed in previous studies by evaluating certain criteria, such as:

(A) Does the method use time-series data? How many time steps does the method utilize?
(B) Is the use of information fusion explored?
(C) Is a hybrid model proposed?
(D) Does the method utilize 3D MRI neuroimaging?

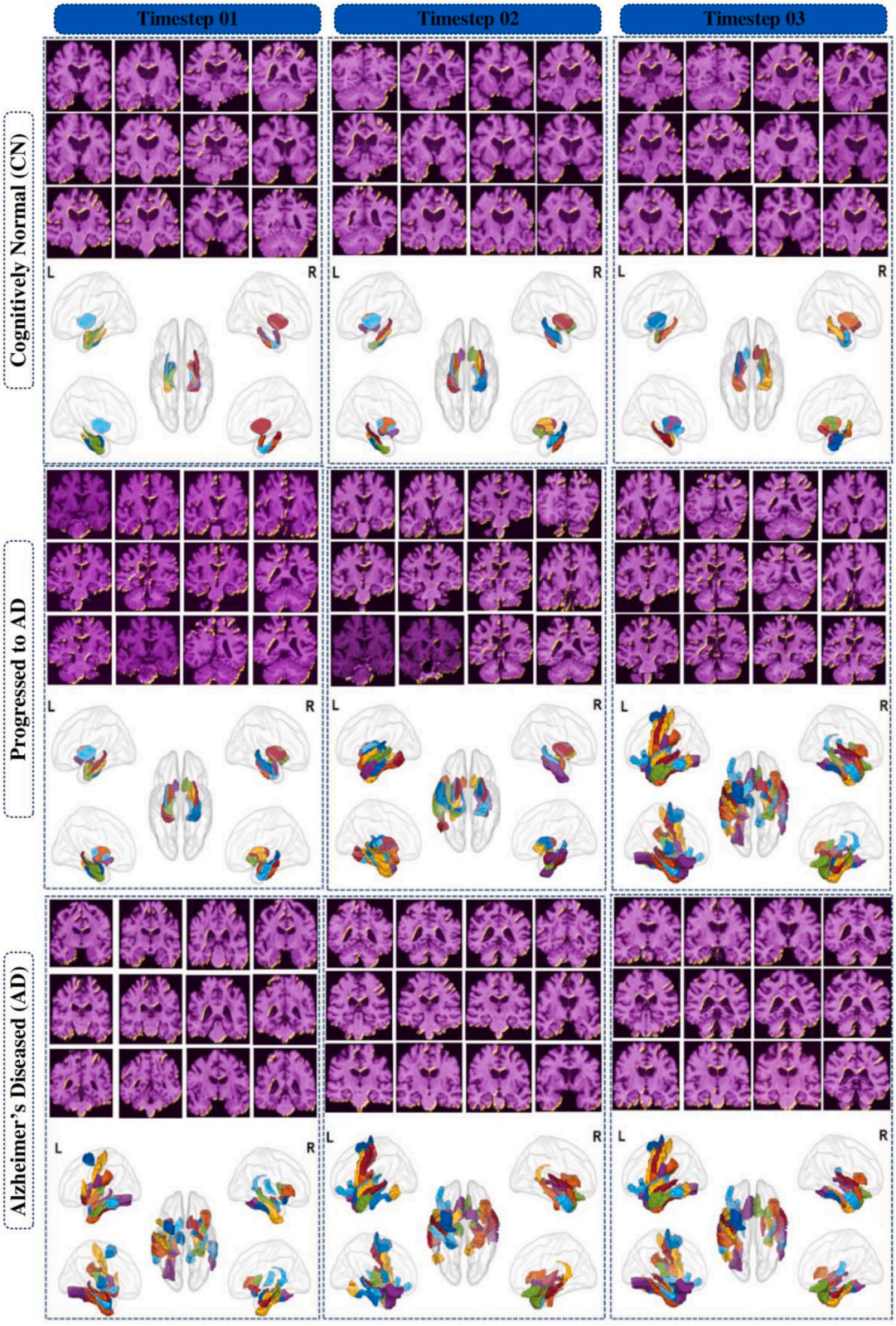

Fig. 17. 2D slices and 3D surface rendering along with activated brain regions for AD progression detection.

(E) Does the method support an MRI-based visual XAI?
(F) Is a time-series XAI proposed?
(G) Does the method track affected brain regions over time using voxel-level visualization?

In [28][51][54][56][58][60][61][63][64][67][70], BL data, including only MRI, PET, and CSs, which are well-known biomarkers for the diagnosis of patients with AD, were utilized. However, any longitudinal training data or explanation of the decision-making process of the proposed approaches was not included. Among all these studies based on BL data only, the methods proposed in [58] and [28] outperformed our proposed method by achieving 100% and 97% accuracy and 100% and 95% specificity, respectively. In [2][7][26][27][53][55][57][62][65][66][68][69], longitudinal data were utilized and acceptable accuracies for AD progression detection were reported. However, voxel-level tissue damage caused by AD progression was not visualized in any of these studies.

5. Model explainability

DL-based approaches have become the "method of choice" owing to their automated feature learning and generalization capability. However, deep models are fundamentally complicated ML algorithms; therefore, what information hidden in the brain scans causes these algorithms to arrive at a specific result is unclear. To date, various XAI techniques have been used to explain the decisions of deep models in the medical domain. However, in longitudinal data analysis for AD progression detection, no appropriate XAI technique is currently available to provide a visual explanation of the brain atrophy found in MRI data. With the proposed model, we provide two visual explanations of the model decision-making process: (1) voxel-level activation maps of 2D MRI slices from each time step of the longitudinal data, and (2) 3D brain surface rendering. Fig. 17 shows these 2D slices and 3D surface renderings along with the activated brain regions in the CN, Converted, and AD patients.

5.1. 3D surface rendering

CT and MR acquisitions provide sectional images that are not exactly 2D. These images are always thin volume slices. Therefore, the 3D visualization of these slices is rather simple owing to their 3D nature. The most common methods used for visualization are slicing, volume rendering, and surface rendering [71]. However, 3D visualization is not possible using the slicing approach. The surface-rendering approach entails creating polygonal surfaces from the datasets and then rendering them. On the other hand, volume rendering entails giving each data element a color and opacity value and then projecting the components directly onto an image plane without the need for geometric primitives. Surface rendering is the depiction of surface structures or organs. For 3D imaging of sectional scan data, surface rendering is a common visualization approach [71]. Surface extraction may be performed in two ways:1) manually, using a standard template to identify the voxel coordinates in a specific brain area of the 3D surface, and 2) automatically, by executing the full preprocessing pipeline using advanced software such as FreeSurfer [72] to visualize specific brain areas. Manual segmentation techniques provide the most precise surface but are time-consuming and difficult for the operator. Unfortunately, in large systems, the currently available automated segmentation solutions that require only basic human intervention cannot always be guaranteed to perform successfully [73]. Hence, this study utilized manual mapping of the brain segments, as shown in Fig. 17.

The 2D activation maps generated by the explainer were further processed to produce a 3D surface rendering. The following steps describe the process of generating 3D surface rendering. 1) Surface acquisition: The brain surface data are collected and consist of four fields in an ASCII text file with the suffix 'nv': a) vertex number, b) vertex coordinates, c) triangle face number, and d) index of vertex making up the triangles. As the vertex coordinates were translated into the MNI space, the brain surface was obtained using the FreeSurfer tool. 2) Volume mapping: The brain volume data, which might be in the form of a T-map, Z-map, atlas, or any other volume data, are converted to the NIFTI format. The Brainnetome Atlas [74] was used to examine the structure-function links and compare neuroanatomical research. There are currently 246 areas in the bilateral hemispheres of the Brainnetome Atlas. Furthermore, by utilizing forward and reverse inference, the atlas-connectivity-based parcellation-yielded areas are functionally characterized according to the BrainMap database's [75] behavioral domain and paradigm class meta-data labels.

The number of intersection points in each brain area was used to identify the brain regions that contributed most to the final classification. The basic idea behind volume mapping is to use various techniques to convert the vertex coordinates of the brain surface into voxels in an image file and assign vertices to the associated values. The idea behind drawing an ROI is to convert the voxels in an image file with the same index into a 3D surface. The ROI volume and brain surface were processed using box-smoothing methods. A MATLAB toolbox, BrainNet Viewer [76], was used to render the surface. Fig. 17 shows the active areas of the brain for each class at each time step. A detailed description of each region is provided in Table S1 in the Supplementary file.

5.2. Voxel activations in 2D slice and 3D surface view

AD is associated with extreme difficulty in performing routine tasks, with patients losing their ability to solve problems, plan, reason, use, judge, and think. This is accompanied by increased confusion and problems with speaking, vision, reading, concentration, and spatial and temporal perception. As a result of these symptoms, a person's mood and personality shifts, and there is a loss of interest in favored pastimes or social activities. A substantial amount of research has linked cognitive, behavioral, and emotional changes to particular anatomical alterations in brain tissue. We discuss how the proposed framework identifies these changes in the context of cognitive deterioration with respect to time (as a progressive problem).

Generating attention maps: We employed the MedCam Python library developed by Gotkowaski [77] to visualize salient features that make significant contributions to the final determination of the output class. Validation data with specified labels were fed to MedCam, and the attention maps obtained were linearly integrated and normalized to achieve understandable attention maps. The total number of attention maps given by MedCam is equal to the shape of the input volume; in this case, $110 \times 110 \times 110$. Subsequently, these attention maps were overlaid on the corresponding slice of the input volume to identify the most salient voxels in the decision-making process. Fig. 17 shows the 2D slices with the corresponding voxel details obtained from MedCam. The visualized salient regions have statistical significance that may help users of the system understand the decision-making process of the deep model. The most discriminative features of a patient with CN and AD status are shown along with those of patients who were initially CN but progressed to AD within three years. Furthermore, multiple 2D slices from each 3D volume were shown to illustrate the activated voxels from different regions in each slice.

Activated brain regions for CN people: The first row in Fig. 17 shows that the brain regions that most easily differentiate CN persons from AD patients are the rostral Hippocampus [78], medial Amygdala [79], Globus Pallidus [79], lateral Amygdala [79], area 28/34 (EC, Entorhinal cortex), and caudal area 35/36, that is, the Parahippocampal gyrus [80]. For CN individuals, these regions remain the same over time (i.e., in scans from time steps 1, 2, and 3). No progression can be noticed in the visual illustrations of CN patients in Fig. 17, indicating that the patient is in a stable state, and no regions deteriorate in time step 1 or later in time steps 2 and 3.

Activated brain regions for AD patients: The third row in Fig. 17

shows disease progression in an AD case. The patient was in a bad condition during the BL scan. Many regions of the brain are affected. The most affected brain regions include the Hippocampus [78], medial Amygdala [79], caudal Hippocampus [79], lateral Amygdala [79], dorsolateral Putamen [81], rostroventral area 20, that is, the Fusiform gyrus [80], Globus pallidus [79], area 28/34 [79], and area Temporal lobe (TL) (lateral PPHC and posterior Parahippocampal gyrus) [82]. Furthermore, the rostral area 21 and superior temporal Sulcus, that is, the middle temporal Gyrus [83]; rostral area 22 and lateral area 38, that is, the superior temporal Gyrus [84]; lateroventral area 37, that is, the Fusiform Gyrus [83]; and the caudoventral of area 20 and intermediate lateral area 20 and caudolateral of area 20, that is, the inferior temporal Gyrus [85] and caudal Hippocampus [79] are the brain regions highlighted as influential by the proposed network at time steps 1 and 2 in AD patients. In the AD patient who was diseased from the beginning, most of the brain atrophy was already present in the very first scan, as shown in Fig. 17. Using the proposed 3D visual and temporal explanations, physicians can easily and intuitively track how the status of the patient has changed over time. At time step 1, the patient's status was clearly negative owing to the number of affected brain regions compared with a CN patient. In addition, by time step 2, the number of affected regions increased compared with that in time step 1. By time step 3, the situation had worsened because the number of affected regions had increased considerably compared with that in time step 2. In our proposed explanation output, physicians can see the newly affected regions at every time step to accurately track the status of the patient.

Activated brain regions for converted patients: The second row in Fig. 17 shows the disease progression in a converted case. Compared with long-term AD patients, converted cases change rapidly from a normal state to AD. The number and volume of the affected brain regions increases rapidly over time. For converted patients, who were CN at the beginning (i.e., at time step 1), the network looked at the same regions used to identify CN status patients (time steps 1, 2, and 3). These regions include the Hippocampus, Amygdala (medial and lateral), and Parahippocampal regions. However, regarding the converted patient, by time step 2, as the cognitive impairment in the patient increased, additional affected regions were detected by the network, including the Hippocampus [78], medial Amygdala [79], caudal Hippocampus [79], lateral Amygdala [79], dorsolateral Putamen [81], and rostroventral area 20, that is, the Fusiform Gyrus [80], Globus Pallidus [79], area 28/34 (EC, Entorhinal cortex) [79], and area TL [82]. For a patient who converts to AD, by time step 3, as the patient succumbs completely to AD, more than half of the brain tissue atrophies, including all the regions specified for time steps 1 and 2. Additional affected regions are detected by the network, including the rostral area 21 and superior temporal Sulcus, that is, the middle temporal Gyrus [86], rostral area 22, and lateral area 38, that is, the superior temporal [86]; lateroventral area 37, i.e., Fusiform Gyrus [80]; and the caudoventral of area 20, intermediate lateral area 20, and caudolateral of area 20, that is, the inferior temporal Gyrus [85] and caudal Hippocampus [79].

Fig. 18 shows a concise representation of the regions presented in Fig. 17. The brain regions most affected by AD are the Basal Ganglia (BG), Parahippocampal Gyrus (PhG), Amygdala (Amyg), and Hippocampus (Hipp). Fig. 17 shows a practical use case of a patient from each category. The circular subregions represent the activation of that region over all three time steps. For CN patients, both the left and right dorsolateral Putamen (dlPu) regions were activated in all three time steps (T1, T2, and T3). By examining this region, the proposed network can discriminate CN individuals from AD patients. Other detected subregions of the Basal Ganglia (BG) include the left and right Globus Pallidus (GP), as well as the Nucleus Accumbens (NAC) at T1 and T2. The A28/34 subregion of the PhG was activated at all three time steps. In addition, A35/36c and the Temporal Lobe showed activations in T2 and T3. The Thalamus region was activated only in T1 and T3. Similarly, the Amygdala (Amyg) and Hippocampus (Hippo) regions were activated at different time steps in CN patients. For converted patients, dlPu and NAC showed activation over all three time steps (T1, T2, and T3). In addition, both the ventromedial Putamens (vmPu) were activated in T2 and T3. Similarly, A28/34 and TL were activated over all three time steps. Most of the subregions of the Amyg and Hippo were also activated. For patients with AD who had AD even at the beginning and continued in their AD status throughout, the main activated subregions were the NAC over all three time steps. In addition, dlPU was activated in the third time step. Similarly, A28/34 and TL were activated over all three time steps in AD patients. In addition to these regions, the Amyg and Hipp were activated over all three time steps for AD patients.

The influential regions identified by the model for both AD and converted patients were medically relevant. For example, under AD, the subcortical areas of the Hippocampus and Amygdala in the medial temporal lobe have repeatedly been identified as the most significant regions [81]. Likewise, structural alterations in the Amygdala, a brain area primarily responsible for emotional experiences and expressions, have been linked to personality changes under AD, including increased irritability and anxiety [72,73]. The Parahippocampal Gyrus, Thalamus, and Putamen were also significantly stimulated in subcortical areas. While the primary purpose of the Thalamus is to transport motor and sensory impulses to the cerebral cortex controlling awareness and sleep, the Dorsal Striatum is thought to have a direct role in subjective

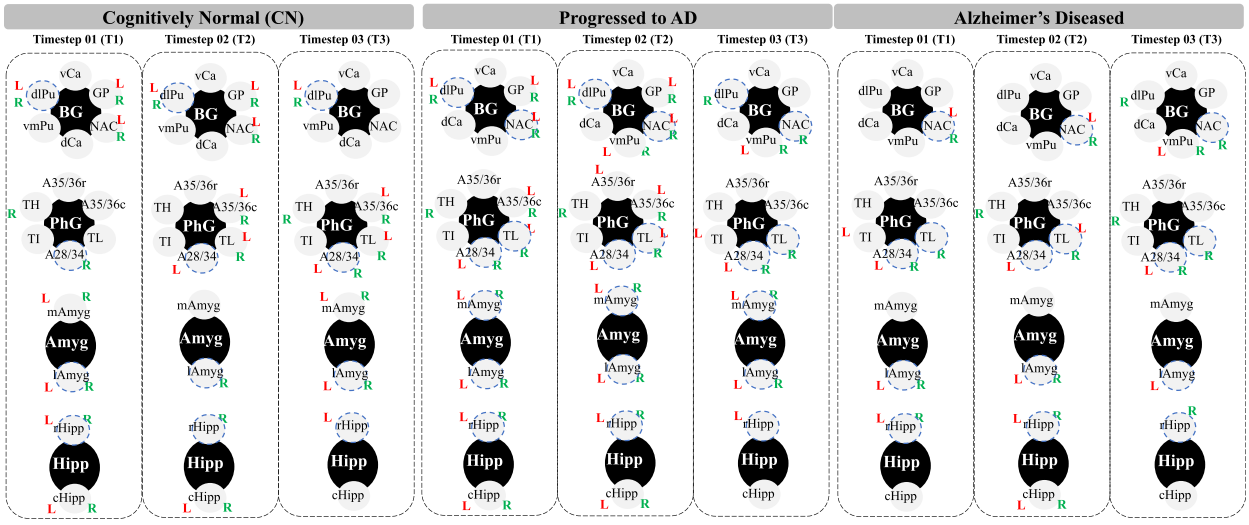

Fig. 18. Activated regions of the brain at different time steps.

decision-making. The Putamen and Thalamus areas of the brain show abnormalities that are typical in subjects with AD [74,75]. In addition to the medial temporal lobe reported by many studies, we found peak activations in the inferior and superior temporal Gyrus as well as in the Fusiform gyrus, which are best known for use in pattern recognition (e.g., face, body, object, or colors) [86].

These areas are also associated with decision-making and problem-solving and are believed to be severely damaged in patients with AD. This leads to increased lethargy, inappropriate behavior, and conditions of being stuck in a loop of compulsive behavior [87]. In addition to the Frontotemporal network, AD is characterized by deterioration of the Precuneus region (see the converted case in Fig. 17 at time step 3), which is a critical part of the Parietal lobe [88]. The Cerebellum has also been increasingly examined as a brain region associated with the progression of AD and has been found to be a direct contributor to cognitive and neuropsychiatric deficits in brain functionality. This region is particularly important for cognitive and behavioral functions [89]. Deterioration of the Cerebellum causes several symptoms, including loss of balance and coordination, tremors, difficulty in speaking, and abnormal eye movement in elderly people. Finally, atrophy in the occipital lobe is associated with being unaware of the surroundings and experiencing hallucinations and misperceptions. In addition, the occipital regions and their subregions, including the Calcarine, Cuneus, and Lingual gyrus, are affected during AD progression (see the converted case in Fig. 17 at time step 3).

5.3. Comparison of the proposed XAI approach with state-of-the-art methods

We provide time-series explainability using visual representations of 3D images to justify the model's decision. This module tracks the regions in the brain tissue over time to help the model recognize patients with AD. Various XAI techniques have been used to explain the decisions of deep models in the medical domain. However, in longitudinal data analysis for AD progression detection, there is currently no available XAI technique to provide a visual explanation of brain atrophy found in MRI data. With the proposed model, we provide two visual explanations of the model decision-making process: (i) voxel-level activation maps of 2D MRI slices from each time step of the longitudinal data and (ii) 3D brain surface rendering.

Fig. 17 displays 2D slices and 3D surface renderings, along with the activated brain regions in CN, Converted, and AD patients. For each category, the model shows the activation of brain regions in a progressive manner that defines tissue damage over time. Fig. 18 shows a concise representation of the regions presented in Fig. 17. The encircled subregions represent the activation of the brain region over the three time steps. Only a limited number of studies on AD had been conducted to explore the visual XAI features for explaining model decisions in a medically relevant manner. To the best of our knowledge, the visual explainability of longitudinal 3D MRI has not been explored in any other studies to highlight damaged brain regions. In addition, specific MRI regions have not been tracked over time or their role in AD progression has not been examined in any previous study. In this study, we proposed novel solutions to both challenges.

Table 9 shows a list of the XAI approaches used when defining the decision-making process of a model-based diagnostic framework. In [90] and [91], SHAP values from well-known AD biomarkers were determined, specifying their contribution to the decision-making process of a model. In [92,93], and [94], saliency-based XAI approaches were proposed by visualizing the feature maps of the hidden layers of a DCNN. In [95], a 3D explainable residual block was proposed for the BL 3D MRI volume, whereas in [96], a GAN-based XAI approach for producing multi-way counterfactual maps was proposed to explain the decision of the proposed framework. Unlike the system proposed in this study, none of these methods provide voxel-wise 2D and 3D explanations for each slice in the longitudinal MRI, making them less adaptable to the actual AD diagnosis environment of a health care system.

6. Current limitations

Although the proposed framework performed well on the ADNI dataset under a variety of settings, some limitations still exist. First, the parameters of the DCNN model, such as the number of layers, their size, and the number of kernels in each layer, are optimally determined. However, the proposed framework cannot be evaluated over other datasets that either have an extremely limited number of longitudinal subjects or are not publicly accessible. Second, we explored multimodal data for AD progression detection. Our dataset was composed of longitudinal MRIs with cross-sectional biomarkers at the BL only. We could not investigate the effects of CSs and demographic features at different time steps because of the unavailability of these features in a longitudinal manner. Third, our model was trained from scratch. Owing to compatibility issues with other available datasets (NACC, AIBL, and MIRIAD), we could not fine-tune the existing framework trained on other large-scale 3D medical image datasets that can further improve the learning performance of the proposed framework. Finally, we mainly focused on the visual explainability of the 2D slices of the proposed framework and the 3D brain surface and ignored the visualization of other modalities, such as demographics and CSs, in the progression detection process. This is because our dataset was exclusively composed of longitudinal MRI data.

7. Conclusion and future work

Alzheimer's disease is the most severe form of dementia, and there is currently no medically approved cure for this disease. The available medical diagnostic systems are mainly based on cross-sectional data collected from an initial BL visit, without the longitudinal aspect of the available clinical data being considered. Conventional DL models function as black-box models without explaining their decision-making.

Table 9
Comparison of our proposal with existing explainability techniques.

Ref.	Year	XAI technique	Time series	Modality	Tracking voxel level affected brain regions in the longitudinal MRI	XAI category
[94]	2019	Saliency map	No	MRI	No	Gradient-based
[97]	2020	Rule extraction	No	MRI	No	Argumentation-based
[92]	2020	Saliency map	Yes	MRI	No	Gradient-based
[95]	2021	Class activation map	No	MRI	No	Gradient-based
[90]	2021	SHAP	No	Demographics, MRI, genetics, lab tests, CS, and neuropsychological battery	No	Fuzzy rule
[91]	2022	SHAP	Yes	Demographic, clinical, and neuropsychological assessment	No	Game theory
[93]	2022	Saliency map	No	MRI, PET, and neuropsychology test	No	Graph-based
[96]	2022	Counterfactual map	Yes	MRI	No	Visual explanation
Ours*	2022	Time-series-guided Grad-CAM	Yes	MRI, Demographics, and CSs	Yes	Time-series visual explanation

Existing DL models are extremely accurate; however, their real-world adoption is hindered because doctors and regulators cannot verify their results. This study proposed a novel framework for AD progression detection using longitudinal MRI input data. We also investigated the effect of multimodal input data by adding patients' demographics and CSs. The proposed network is composed of a 3D CNN followed by a BRNN that outputs a decision based on the temporal features from longitudinal MRI data. Our experiments show that the proposed model achieves better results than existing state-of-the-art techniques for AD progression detection by incorporating longitudinal and cross-sectional data. We further proposed a novel explainability approach to help doctors understand the decisions of the proposed network in a medically acceptable way. The proposed network was optimized using a well-known grid-search hyperparameter optimization technique. We achieved promising results that outperformed existing studies and other DL models.

The main aim of this study was to explore the role of MRI time-series data in predicting AD progression. The proposed model achieved high and very stable results. However, fusing MRI with other modalities, such as CSs and demographics, can further improve model performance. In the future, we will explore the effect of fusing multimodal time-series data on model performance. In addition, our current study proposed a novel time-series visual explainability for 2D slices and 3D brain surfaces. However, medical experts prefer multiple explanations to trust the model's results. Thus, other XAI techniques that use additional modalities, such as demographics and CSs, will be explored in future research. We also plan to investigate the time-series aspects of other explainability methods, such as SHAP [98], in combination with visual temporal explainability. This will enable us to compare the contribution of each modality in the decision-making process. Finally, we will compare the performance of the proposed architecture with avant-garde 3D DL architectures, with the potential to yield better performance scores than those reported here. In addition, we will utilize state-of-the-art data augmentation techniques to prepare MRI data for better model training, as discussed in [99].

CRediT authorship contribution statement

Nasir Rahim: Conceptualization, Software, Writing – original draft. **Shaker El-Sappagh:** Methodology, Writing – review & editing. **Sajid Ali:** Conceptualization, Validation. **Khan Muhammad:** Investigation, Writing – review & editing. **Javier Del Ser:** Investigation, Writing – review & editing. **Tamer Abuhmed:** Methodology, Project administration, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgement

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the ICT Creative Consilience Program (IITP-2021–2020–0–01821), supervised by the Institute for Information & communications Technology Planning & Evaluation (IITP), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1011198).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.inffus.2022.11.028.

References

[1] Y. Liu, et al., Diffusion tensor imaging and tract-based spatial statistics in Alzheimer's disease and mild cognitive impairment, Neurobiol. Aging 32 (9) (2011) 1558–1571, https://doi.org/10.1016/j.neurobiolaging.2009.10.006. Sep.
[2] J. Albright, Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm,", Alzheimer's Dement. Transl. Res. Clin. Interv. 5 (2019) 483–491, https://doi.org/10.1016/j.trci.2019.07.001.
[3] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease, J. Neurosci. Methods 339 (2020), 108701, https://doi.org/10.1016/j.jneumeth.2020.108701. Jun.
[4] H.T. Shen, et al., Heterogeneous data fusion for predicting mild cognitive impairment conversion, Inf. Fusion 66 (2021) 54–63, https://doi.org/10.1016/j.inffus.2020.08.023. Feb.
[5] S. Qiu, G.H. Chang, M. Panagia, D.M. Gopal, R. Au, V.B. Kolachalama, Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment, Alzheimer's Dement. Diagnosis, Assess. Dis. Monit. 10 (2018) 737–749, https://doi.org/10.1016/j.dadm.2018.08.013.
[6] Z. Liu, T.S. Johnson, W. Shao, M. Zhang, J. Zhang, K. Huang, Optimal transport- and kernel-based early detection of mild cognitive impairment patients based on magnetic resonance and positron emission tomography images, Alzheimer's Res. Ther. 14 (1) (2022) 1–12, https://doi.org/10.1186/s13195-021-00915-3. Dec.
[7] S. El-Sappagh, et al., Alzheimer's disease progression detection model based on an early fusion of cost-effective multimodal data, Futur. Gener. Comput. Syst. 115 (2021) 680–699, https://doi.org/10.1016/j.future.2020.10.005.
[8] Y. Zhang, S. Wang, K. Xia, Y. Jiang, P. Qian, Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion, Inf. Fusion 66 (2021) 170–183, https://doi.org/10.1016/j.inffus.2020.09.002. Feb.
[9] Y. Fan, N. Batmanghelich, C.M. Clark, C. Davatzikos, Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, Neuroimage 39 (4) (2008) 1731–1743, https://doi.org/10.1016/j.neuroimage.2007.10.031. Feb.
[10] E.E. Bron, et al., Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge, Neuroimage 111 (2015) 562–579, https://doi.org/10.1016/j.neuroimage.2015.01.048. May.
[11] X. Jiang, L. Chang, Y.-.D. Zhang, Classification of Alzheimer's disease via eight-layer convolutional neural network with batch normalization and dropout techniques, J. Med. Imaging Heal. Informatics 10 (5) (2020) 1040–1048, https://doi.org/10.1166/jmihi.2020.3001. Feb.
[12] Y. Zhang, et al., Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization, J. Alzheimer's Dis. 65 (3) (2018) 855–869, https://doi.org/10.3233/JAD-170069. Jan.
[13] L. Xu, X. Wu, K. Chen, L. Yao, Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment, Comput. Methods Programs Biomed. 122 (2) (2015) 182–190, https://doi.org/10.1016/j.cmpb.2015.08.004. Nov.
[14] G. Muhammad, F. Alshehri, F. Karray, A. El Saddik, M. Alsulaiman, T.H. Falk, A comprehensive survey on multimodal medical signals fusion for smart healthcare systems, Inf. Fusion 76 (2021) 355–375, https://doi.org/10.1016/j.inffus.2021.06.007. Dec.
[15] S. El-Sappagh, T. Abuhmed, S.M. Riazul Islam, K.S. Kwak, Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data, Neurocomputing 412 (2020) 197–215, https://doi.org/10.1016/j.neucom.2020.05.087.
[16] S. Huang, et al., Identifying Alzheimer's disease-related brain regions from multi-modality neuroimaging data using sparse composite linear discrimination analysis, Adv. Neural Inf. Process. Syst. 24 (2011).
[17] K.R. Gray, P. Aljabar, R.A. Heckemann, A. Hammers, D. Rueckert, Random forest-based similarity measures for multi-modal classification of Alzheimer's disease, Neuroimage 65 (2013) 167–175, https://doi.org/10.1016/j.neuroimage.2012.09.065. Jan.
[18] A. Chincarini, et al., Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer's disease, Neuroimage 125 (2016) 834–847, https://doi.org/10.1016/j.neuroimage.2015.10.065. Jan.
[19] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398–412, https://doi.org/10.1016/j.neuroimage.2014.10.002. Jan.
[20] P.J. Moore, T.J. Lyons, J. Gallacher, Random forest prediction of Alzheimer's disease using pairwise selection from time series data, PLoS ONE 14 (2) (2019), https://doi.org/10.1371/journal.pone.0211558. Feb.
[21] A. Holzinger, et al., Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence, Inf. Fusion 79 (2022) 263–278, https://doi.org/10.1016/j.inffus.2021.10.007. Mar.

[22] A. Barredo Arrieta, et al., Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115, https://doi.org/10.1016/j.inffus.2019.12.012. Jun.

[23] A. Holzinger, C. Biemann, C.S. Pattichis, and D.B. Kell, "What do we need to build explainable AI systems for the medical domain?," arxiv.org, 2017, [Online]. Available: http://arxiv.org/abs/1712.09923.

[24] P. McCarthy, FSLeyes, 1470761, Zenodo, 2019, https://doi.org/10.5281/zenodo.

[25] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, and N. Díaz-Rodríguez, "Explainable Artificial Intelligence (XAI) on TimeSeries data: a survey," 2021, [Online]. Available: http://arxiv.org/abs/2104.00950.

[26] R. Cui, M. Liu, RNN-based longitudinal analysis for diagnosis of Alzheimer's disease, Comput. Med. Imaging Graph. 73 (2019) 1–10, https://doi.org/10.1016/j.compmedimag.2019.01.005.

[27] S. El-Sappagh, T. Abuhmed, K.S. Kwak, Alzheimer disease prediction model based on decision fusion of CNN-BiLSTM deep neural networks, in: Advances in Intelligent Systems and Computing, 1252, AISC, 2021, pp. 482–492, https://doi.org/10.1007/978-3-030-55190-2_36.

[28] R. Sharma, T. Goel, M. Tanveer, R. Murugan, FDN-ADNet: fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans, Appl. Soft Comput. 115 (2022), 108099, https://doi.org/10.1016/J.ASOC.2021.108099. Jan.

[29] S. Ji, W. Xu, M. Yang, K. Yu, 3D Convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231, https://doi.org/10.1109/TPAMI.2012.59.

[30] M. Liu, D. Cheng, W. Yan, Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images, Front. Neuroinform. 12 (June) (2018) 1–12, https://doi.org/10.3389/fninf.2018.00035.

[31] P. McCarthy, FSLeyes, Zenodo, 2020, https://doi.org/10.5281/zenodo.3937147. Jul.

[32] "FreeSurfer_freeview - Free Surfer Wiki." [Online]. Available: https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/OutputData_freeview.

[33] "Advanced Normalization Tools." [Online]. Available: http://stnava.github.io/ANTs/.

[34] "BET - FslWiki - Skull Stripping." [Online]. Available: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET.

[35] "MNI Atlases - FslWiki." [Online]. Available: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases.

[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

[37] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, https://doi.org/10.1038/323533a0.

[38] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, Int. J. Uncertainty, Fuzziness Knowlege-Based Syst. 6 (2) (1998) 107–116, https://doi.org/10.1142/S0218488598000094. Nov.

[39] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, Neural Comput 12 (10) (2000) 2451–2471, https://doi.org/10.1162/089976600300015015.

[40] S. Zhang, K. Bi, T. Qiu, Bidirectional recurrent neural network-based chemical process fault diagnosis, Ind. Eng. Chem. Res. 59 (2) (2020) 824–834, https://doi.org/10.1021/acs.iecr.9b05885.

[41] M. Mullan, C. Bennett, C. Figueredo, F. Crawford, Clinical features of early onset, familial Alzheimer`s disease linked to chromosome 14, Am. J. Med. Genet. 60 (1) (1995), https://doi.org/10.1002/ajmg.1320600109.

[42] J.T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: the all convolutional net," arXiv Prepr. arXiv1412.6806, 2014.

[43] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer leaming to 3D ResNet-18 for Alzheimer's disease detection on MRI images," in International Conference Image and Vision Computing New Zealand, Nov. 2020, vol. 2020-Novem. doi: 10.1109/IVCNZ51579.2020.9290616.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. Accessed: Oct. 06, 2021. [Online]. Available: http://www.robots.ox.ac.uk/.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-Decem, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[46] W. Zhang, et al., Multi-task learning with multi-view weighted fusion attention for artery-specific calcification analysis, Inf. Fusion 71 (2021) 64–76, https://doi.org/10.1016/J.INFFUS.2021.01.009. Jul.

[47] J. Dolz, C. Desrosiers, I. Ben Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study, Neuroimage 170 (2018) 456–470, https://doi.org/10.1016/j.neuroimage.2017.04.039. Apr.

[48] G. Muhammad, M. Shamim Hossain, COVID-19 and Non-COVID-19 classification using multi-layers fusion from lung ultrasound images, Inf. Fusion 72 (2021) 80–88, https://doi.org/10.1016/j.inffus.2021.02.013. Aug.

[49] R.C. Petersen, et al., Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization, Neurology 74 (3) (2010) 201–209, https://doi.org/10.1212/WNL.0b013e3181cb3e25.

[50] D. Lu, K. Popuri, G.W. Ding, R. Balachandar, M.F. Beg, Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease, Med. Image Anal. 46 (2018) 26–34, https://doi.org/10.1016/j.media.2018.02.002.

[51] M.Mehdipour Ghazi, et al., Training recurrent neural networks robust to incomplete data: application to Alzheimer's disease progression modeling, Med. Image Anal. 53 (2019) 39–46, https://doi.org/10.1016/j.media.2019.01.004.

[52] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2019) 1195–1206, https://doi.org/10.1109/TBME.2018.2869989.

[53] T. Abuhmed, S. El-Sappagh, J.M. Alonso, Robust hybrid deep learning models for Alzheimer's progression detection, Know.-Based Syst 213 (2021), 106688, https://doi.org/10.1016/j.knosys.2020.106688.

[54] M. Bucholc, et al., A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual, Expert Syst. Appl. 130 (2019) 157–171, https://doi.org/10.1016/j.eswa.2019.04.022.

[55] "CrossEntropyLoss — PyTorch 1.11.0 documentation." https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html#torch.nn.CrossEntropyLoss (accessed May 26, 2022).

[56] S. Lu, Y. Xia, W. Cai, M. Fulham, D.D. Feng, Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and FDG-PET imaging, Comput. Med. Imaging Graph. 60 (2017) 35–41, https://doi.org/10.1016/j.compmedimag.2017.01.001.

[57] X. Hong, et al., Predicting Alzheimer's disease using LSTM, IEEE Access 7 (2019) 80893–80901, https://doi.org/10.1109/ACCESS.2019.2919385.

[58] S. Lahmiri, A. Shmuel, Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease, Biomed. Signal Process. Control 52 (2019) 414–419, https://doi.org/10.1016/j.bspc.2018.08.009.

[59] C. Platero, M.C. Tobar, Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease, J. Neurosci. Methods 341 (March) (2020), 108698, https://doi.org/10.1016/j.jneumeth.2020.108698.

[60] M. Nguyen, T. He, L. An, D.C. Alexander, J. Feng, B.T.T. Yeo, Predicting Alzheimer's disease progression using deep recurrent neural networks, Neuroimage 222 (September 2019) (2020), 117203, https://doi.org/10.1016/j.neuroimage.2020.117203.

[61] G. Uysal, M. Ozturk, Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods, J. Neurosci. Methods 337 (November 2019) (2020), 108669, https://doi.org/10.1016/j.jneumeth.2020.108669.

[62] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, X. Song, Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: a Novel Approach Combining Convolutional Neural Networks and Ensemble Learning, Front. Neurosci. 14 (May) (2020) 1–19, https://doi.org/10.3389/fnins.2020.00259.

[63] Y. Zhu, M. Kim, X. Zhu, D. Kaufer, G. Wu, Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data, Med. Image Anal. 67 (2021), 101825, https://doi.org/10.1016/j.media.2020.101825.

[64] M. Dong, et al., DeepAtrophy: teaching a neural network to detect progressive changes in longitudinal MRI of the hippocampal region in Alzheimer's disease, Neuroimage 243 (September 2020) (2021), 118514, https://doi.org/10.1016/j.neuroimage.2021.118514.

[65] M. Liu, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, Neuroimage 208 (2020), 116459, https://doi.org/10.1016/j.neuroimage.2019.116459. Mar.

[66] M. Mehdipour Ghazi, et al., Robust parametric modeling of Alzheimer's disease progression, Neuroimage 225 (June 2020) (2021), 117460, https://doi.org/10.1016/j.neuroimage.2020.117460.

[67] S. El-Sappagh, Hager Saleh, F. Ali, E. Amer, Tamer Abuhmed, Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time, Neural Comput. Appl. (2022) 1–23, https://doi.org/10.1007/s00521-022-07263-9.

[68] H.A. Helaly, M. Badawy, A.Y. Haikal, Toward deep MRI segmentation for Alzheimer's disease detection, Neural Comput. Appl. 34 (2) (2022) 1047–1063, https://doi.org/10.1007/s00521-021-06430-8.

[69] F. Nan, et al., A Multi-classification Accessment framework for reproducible evaluation of multimodal learning in Alzheimer's disease, IEEE/ACM Trans. Comput. Biol. Bioinforma. XX (Xx) (2022) 1–14, https://doi.org/10.1109/TCBB.2022.3204619.

[70] S. El-Sappagh, F. Ali, T. Abuhmed, J. Singh, J.M. Alonso, Automatic detection of Alzheimer's disease progression: an efficient information fusion approach with heterogeneous ensemble classifiers, Neurocomputing 512 (2022) 203–224, https://doi.org/10.1016/J.NEUCOM.2022.09.009. Nov.

[71] A.G. Schreyer, S.K. Warfield, Surface rendering. 3D Image Processing, Springer, 2002, pp. 31–34.

[72] B. Fischl, FreeSurfer, Neuroimage 62 (2) (2012) 774–781, https://doi.org/10.1016/j.neuroimage.2012.01.021.

[73] S.A.M. Silverman, D.L. Janssen, Diagnostic imaging, Reptil. Med. Surg. 2 (2006), 471489.

[74] L. Fan, et al., The Human Brainnetome Atlas: a New brain Atlas based on connectional architecture, Cereb. Cortex 26 (8) (2016) 3508–3526, https://doi.org/10.1093/cercor/bhw157. Aug.

[75] A.R. Laird, J.L. Lancaster, P.T. Fox, BrainMap: the social evolution of a human brain mapping database, Neuroinformatics 3 (1) (2005) 65–77, https://doi.org/10.1385/ni:3:1:065.

[76] M. Xia, J. Wang, Y. He, BrainNet viewer: a network visualization tool for human brain connectomics, PLoS ONE 8 (7) (2013) e68910, https://doi.org/10.1371/journal.pone.0068910. Jul.

[77] K. Gotkowski, C. Gonzalez, A. Bucher, A. Mukhopadhyay, M3d-CAM: a PyTorch library to generate 3D attention maps for medical deep learning. Informatik Aktuell, 2021, pp. 217–222, https://doi.org/10.1007/978-3-658-33198-6_52.
[78] S.J. Greene, R.J. Killiany, Hippocampal subregions are differentially affected in the progression to Alzheimer's disease, Anat. Rec. 295 (1) (2012) 132–140, https://doi.org/10.1002/ar.21493.
[79] P.T. Nelson, et al., The amygdala as a locus of pathologic misfolding in neurodegenerative diseases, J. Neuropathol. Exp. Neurol. 77 (1) (2018) 2–20, https://doi.org/10.1093/jnen/nlx099.
[80] G.W. Van Hoesen, J.C. Augustinack, J. Dierking, S.J. Redman, R. Thangavel, The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates, Ann. N. Y. Acad. Sci. 911 (2000) 254–274, https://doi.org/10.1111/j.1749-6632.2000.tb06731.x.
[81] S. Reeves, M. Mehta, R. Howard, P. Grasby, R. Brown, The dopaminergic basis of cognitive and motor performance in Alzheimer's disease, Neurobiol. Dis. 37 (2) (2010) 477–482, https://doi.org/10.1016/j.nbd.2009.11.005. Feb.
[82] D.P. Devanand, et al., Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease, Neurology 68 (11) (2007) 828–836, https://doi.org/10.1212/01.wnl.0000256697.20968.d7. Mar.
[83] E. Castro, A. Ulloa, S.M. Plis, J.A. Turner, V.D. Calhoun, Generation of synthetic structural magnetic resonance images for deep learning pre-training, Proc. - Int. Symp. Biomed. Imaging 2015-July (2015) 1057–1060, https://doi.org/10.1109/ISBI.2015.7164053.
[84] A. Ulloa, S. Plis, E. Erhardt, V. Calhoun, Synthetic structural magnetic resonance image generator improves deep learning prediction of schizophrenia, IEEE Int. Work. Mach. Learn. Signal Process. MLSP 2015-Novem (2015) 1–6, https://doi.org/10.1109/MLSP.2015.7324379.
[85] S.W. Scheff, D.A. Price, F.A. Schmitt, M.A. Scheff, E.J. Mufson, Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease, J. Alzheimer's Dis. 24 (3) (2011) 547–557, https://doi.org/10.3233/JAD-2011-101782.
[86] G. Karas, et al., Amnestic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease, in Am. J. Neuroradiol. 29 (5) (2008) 944–949, https://doi.org/10.3174/ajnr.A0949. May.
[87] S.P. Poulin, R. Dautoff, J.C. Morris, L.F. Barrett, B.C. Dickerson, Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity, Psychiatry Res. - Neuroimag. 194 (1) (2011) 7–13, https://doi.org/10.1016/j.pscychresns.2011.06.014.
[88] J.D. Schmahmann, Cerebellum in Alzheimer's disease and frontotemporal dementia: not a silent bystander, Brain 139 (5) (2016) 1314–1318, https://doi.org/10.1093/brain/aww064. BrainMay 01,.
[89] D. Chan, N. Fox, M. Rossor, Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia [6], Neurology 58 (5) (2002) 838, https://doi.org/10.1212/WNL.58.5.838.
[90] S. El-Sappagh, J.M. Alonso, S.M.R. Islam, A.M. Sultan, K.S. Kwak, A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease, Sci. Rep. 11 (1) (2021) 2660, https://doi.org/10.1038/s41598-021-82098-3.
[91] A. Lombardi, et al., A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of Mild Cognitive Impairment and Alzheimer's Disease, Brain Informatics 9 (1) (2022) 1–17, https://doi.org/10.1186/s40708-022-00165-5. Dec.
[92] A. Essemlali, E. St-Onge, M. Descoteaux, P.-.M. Jodoin, Understanding Alzheimer disease's structural connectivity through explainable AI, Proc. Mach. Learn. Res. 121 (2020) 217–229.
[93] S. Anjomshoae, S. Pudas, A.D.N.I (ADNI), Explaining graph convolutional network predictions for clinicians - an explainable AI Approach to Alzheimer's Disease Classification, SSRN Electron. J. (2022), https://doi.org/10.2139/SSRN.4194675. Sep.
[94] K. Oh, Y.-.C. Chung, K.W. Kim, W.-.S. Kim, and I.-.S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning", doi: 10.1038/s41598-019-54548-6.</bib.
[95] X. Zhang, L. Han, W. Zhu, L. Sun, D. Zhang, An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE J. Biomed. Heal. Informatics (2021), https://doi.org/10.1109/JBHI.2021.3066832.
[96] K. Oh, J.S. Yoon, H. Il Suk, Learn-explain-reinforce: counterfactual reasoning and its guidance to reinforce an Alzheimer's Disease diagnosis model, IEEE Trans. Pattern Anal. Mach. Intell. (2022), https://doi.org/10.1109/TPAMI.2022.3197845.
[97] K.G. Achilleos, S. Leandrou, N. Prentzas, P.A. Kyriacou, A.C. Kakas, C.S. Pattichis, Extracting explainable assessments of Alzheimer's disease via machine learning on brain MRI imaging data, in: Proc. - IEEE 20th Int. Conf. Bioinforma. Bioeng. BIBE 2020, 2020, pp. 1036–1041, https://doi.org/10.1109/BIBE50027.2020.00175. Oct.
[98] S.M. Lundberg and S.I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, 2017, vol. 2017-Decem, pp. 4766–4775. Accessed: Dec. 16, 2021. [Online]. Available: https://github.com/slundberg/shap.
[99] M. Choe, J. Yoo, G. Lee, W. Baek, U. Kang, and K. Shin, "MiDaS: representative sampling from real-world hypergraphs," 2022, doi: 10.1145/1122445.1122456.