Alexandria Engineering Journal 82 (2023) 484–502

Contents lists available at ScienceDirect

Alexandria Engineering Journal

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

Original Article

Time-series visual explainability for Alzheimer's disease progression detection for smart healthcare

Nasir Rahim [a], Tamer Abuhmed [a,*], Seyedali Mirjalili [b,c], Shaker El-Sappagh [a,d,e], Khan Muhammad [f,*]

[a] Information Laboratory (InfoLab), Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, South Korea
[b] Center for Artificial Intelligence Research and Optimization, Torrens University Australia, Brisbane, QLD 4006, Australia
[c] University Research and Innovation Center, Obuda University, 1034 Budapest, Hungary
[d] Faculty of Computer Science and Engineering, Galala University, Suez 435611, Egypt
[e] Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, Banha 13518, Egypt
[f] Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul, South Korea

ARTICLE INFO

Keywords:
Alzheimer's disease
Deep learning models
Longitudinal multi-model data
Explainable AI
Smart healthcare

ABSTRACT

Artificial intelligence (AI)-based diagnostic systems provide less error-prone and safer support to clinicians, enhancing the medical decision-making process. This study presents a smart and reliable healthcare framework for detecting Alzheimer's disease (AD) progression. Early detection of AD before the onset of clinical symptoms is the most crucial step in starting timely treatment. To predict the conversion of cognitively normal patients to those with AD, three-dimensional 3D magnetic resonance imaging (MRI) whole-brain neuroimaging methods have been extensively studied. However, depending on the 3D volume, this method is computationally expensive. To solve this problem, we used an approximate rank pooling method originally designed for video action recognition with a 3D MRI volume to obtain a compressed representation of multiple two-dimensional (2D) MRI slices. This study proposes a hybrid multimodal CNN-BiLSTM deep model for AD progression detection, in which the resulting dynamic 2D images are fused with cognitive features. Moreover, a novel explainable AI approach is proposed to provide visual explanations using the resulting longitudinal 2D dynamic images. Temporal explanations were provided by visualizing the affected brain regions captured using longitudinal 2D MRIs. By utilizing a sample of 1,692 subjects with multimodal data from the Alzheimer's Disease Neuroimaging Initiative dataset, our method was assessed using a 10-fold cross-validation process. The model achieved an area under the receiver operating characteristics curve (AUC) of 94% using longitudinal 2D three-time-step dynamic image data. The fusion of 2D dynamic images with cognitive features enhanced the performance by 2% in terms of the AUC. Patients who gradually develop AD, show changes in various brain regions. For such patients, our system highlights the critical role of the hippocampus, medial amygdala, caudal hippocampus, and lateral amygdala at the initial time steps. In the late stages of AD, the system detects abnormalities in extra brain regions such as the medial temporal gyrus, superior temporal gyrus, fusiform gyrus, and caudal hippocampus; indicating that patients have completely progressed to AD.

1. Introduction

Cognitive decline and memory loss are prominent hallmarks of Alzheimer's disease (AD) that develop over time. AD has a significant impact on approximately 50 million individuals worldwide [1]. The number of AD cases is projected to increase to 75 million by 2030 and reach a 135.5 million by 2050 [1]. AD has a high financial burden, with global expenses estimated at $604 billion in 2010 [2]. In addition, caring for people with AD can have a physical and emotional impact on families and caregivers. The pathology of AD is characterized by several years of preclinical progression before the onset of clinical symptoms, which makes a timely diagnosis difficult. Mild cognitive impairment

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

https://doi.org/10.1016/j.aej.2023.09.050
Received 17 July 2023; Received in revised form 29 August 2023; Accepted 19 September 2023
Available online 20 October 2023

(MCI), a precursor of AD, has an annual conversion rate of 10-25% [3]. Early identification of MCI before the onset of irreversible cognitive damage is imperative for preventive interventions. However, the high subjectivity and variability of cognitive and neuroimaging biomarkers make diagnosis challenging. Researchers have successfully used machine learning (ML)-based methodologies to address this challenge. Equipped with sophisticated computational capabilities and the ability to learn from complex data patterns, ML algorithms can efficiently identify subtle alterations in cognitive and neuroimaging biomarkers that may go unnoticed by human assessment [4]. Such applications of ML have demonstrated encouraging potential for the early diagnosis of MCI and the prediction of progression to AD. Bron et al, [5] organized the CADDementia challenge, which aimed to benchmark and compare various classification algorithms for three diagnostic groups, namely AD, MCI, and normal cognition (NC), employing a multicenter dataset. The algorithms were assessed using a test set of 354 magnetic resonance imaging (MRI) scans, and the algorithm with the best performance yielded an area under the receiver operating characterestics (ROC) curve (AUC) of 78.8%. The study involved 15 research teams and leveraged feature extraction strategies that encompassed voxel-based morphometry or a fusion of features including volume, cortical thickness, shape, and intensity. Zang et al. [6] proposed a hybrid diagnostic method based on a deep convolutional neural network (CNN) and a support vector machine (SVM) to classify early (E)MCI versus NC based on 120 three-dimensional (3D) sMRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The VGG16 CNN was trained on two-dimensional (2D) slice images using a transfer learning technique. Deep features were extracted and fused for least absolute shrinkage and selection operator (LASSO) feature selection and SVM classification, resulting in improved classification performance and reduced training time. This method overcomes the limitations of traditional ML and promotes the development of computer-aided EMCI diagnostics using volumetric sMRI features. Chang et al. [7] presented an eight-layer deep CNN model that addressed the issue of insufficient data samples by separating the training data from the original data. The model also utilized batch normalization technology to normalize the input of each layer into a mini-batch, which improved the gradient reliability and accelerated learning convergence. The study implemented a dropout method to alleviate overfitting and computational consumption and outperformed six state-of-the-art approaches in terms of accuracy.

The primary basis for AD research is neuroimaging, especially MRI [8,9]. Although MRI is a crucial neuroimaging modality for detecting AD, combining it with other modalities can provide a complete picture of the patient's condition and improve the accuracy of the progression model. Disease diagnosis based on unimodal medical data may affect the overall diagnosis of a patient's potential cognitive behavior because of the lack of available knowledge for clinicians. However, multimodal systems can produce complete insights, comprehensive results, and consistent behaviors [10], making them more accurate and acceptable diagnostic systems for the medical community than single-mode systems. Numerous studies have employed multimodal data, utilizing cognitive scores such as the Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Score (ADAS), along with neuroimaging, to detect the progression of AD [11]. The image fusion approach proposed by Song et al. [12] combines gray matter information from different types of neuroimaging data such as MRI and positron emission tomography (PET) to generate an integrated input training data. To assess the effectiveness of various modalities in AD classification tasks, simple and multi-scale 3D CNN models were developed. Experiments from this study demonstrated that multimodal fusion, with its robust representation of information, can enhance the disease identification process and accommodate a diverse set of deep neural networks. This model achieves the highest classification accuracy (95%). Xu et al. [13] proposed a weighted multimodal SRC system that combined training data from multiple sources such as MRI and other variants of PET scans and investigated its accuracy and robustness for people with cognitive impairments. The experimental results reported in this study suggest that our method outperforms all other state-of-the-art multimodal classification algorithms in the AD domain. Huang et al. [14] suggested a sparse composite linear discriminant analysis (LDA) model to identify disease-related brain regions using multimodal data such as MRI and PET. In this approach, the LDA parameters are divided into two parts: a common parameter shared by all data sources and a parameter unique to each data source. In this way, they were able to analyze all data sources together and use their strengths. Thus, we obtained highly accurate diagnostic results for the AD domain. Gray et al. [15] trained a random forest model (RF) for normal cognition (NC)vs. mild cognitive impairment (MCI) vs. AD classes using four modalities: fluorodeoxyglucose FDG-PET, MRI, cerebrospinal fluid (CSF), and genetic features. While all these studies obtained data from only one baseline visit with no further data collection, the utilization of multimodal data for the disease diagnostic process could be enhanced if researchers considered the time dimension of the collected data. This allows them to investigate the impact of changes in a progressive manner over time and potentially improve classification efficacy. The most essential information characterizing the progress of a disease is eliminated if subsequent time intervals are omitted from a given dataset [16].

The assessment of neurodegenerative diseases, particularly AD, which is a severe form of chronic cognitive impairment, greatly benefits from the management and analysis of time-series data. Moreover, distinguishing between the CN and AD based solely on baseline or single-visit data poses a significant challenge in the analysis of degenerative brain diseases [17]. For instance, Alvi et al. [18] proposed a novel approach for detecting patients with MCI by integrating conventional ML and deep learning (DL) algorithms. Their framework included the use of gated recurrent unit and long short-term memory (LSTM) models as feature extractors. The extracted features were then fed into the support vector machine (SVM) and k-nearest neighbors (KNN) models to distinguish between patients with CN and those with MCI. They used a publicly available electroencephalogram (EEG) dataset and pre-processed the data by applying an average filter to remove unwanted signals. This study reported a classification accuracy of 95% for distinguish between the MCI and CN classes. Lei et al. [19] proposed a DL-based model to identify patients with early MCI based on longitudinal data. To accomplish this, they first constructed a brain function connectivity network consisting of a similarity group network to effectively reconstruct brain networks. The data collected from the brain networks were further processed using an LSTM network with self-attention to utilize more refined features, thereby improving the detection of diseased patients. The authors utilized the ADNI longitudinal data with two-fold steps covering 1 year. Lee et al. [20] predicted the transformation of a patient from a cognitively impaired to AD state using multimodal longitudinal data with varying time steps. They proposed a gated recurrent unit (GRU) model that effectively captures the temporal relationships within each modality in longitudinal data and predicts the progression to AD. This study reported a maximum accuracy of 81%. El-Sappagh et al. [21] introduced a two-stage hybrid deep neural network (DNN) model that utilized an LSTM module. The initial stage involved classifying the health status of the patients into NC, MCI, or AD. Subsequently, the second stage involved utilizing a regression model to predict the conversion time for patients with progressive (p)MCI to AD. Multimodal data in building DL-based diagnostic systems lead to the development of medically intuitive models, as demonstrated in previous studies. Also, 3D volumetric data such as CT and MRI have been used in many studies since the introduction of DL technologies into the medical domain. The volumetric nature of such data carries a large amount of useful information that can be beneficial for identifying disease pathology. However, models designed to process 3D data are computationally intensive, making them impractical for many real-world scenarios [22]. Furthermore, previous studies have mainly focused on enhancing the performance of relevant systems, ignoring the interpretability of decisions made by these systems.

Medical professionals are reluctant to accept "black-box" models from the ML community that present high accuracy using test data; however may not perform as well using real-world data [23]. The model must justify a specific diagnosis, making explainable artificial intelligence (XAI) systems an essential development in this domain [24]. A fully integrated XAI system can clarify the internal workings of decision-making processes with the aim of engaging a wider community. New European data protection legislation prohibits the use of black box models in several areas, particularly in the medical field, and experts in the field oppose decisions made by systems that cannot be retracted [25]. In order to gain the trust of physicians and encourage professionals to consider the recommendations of an artificial intelligence (AI) system, transparency is a crucial aspect. Transparency allows medical professionals to make treatment decisions based on their experience and judgment. Sometimes, people are unable to explain or justify their decisions, as many scholars have suggested. However, explainability plays a vital role in ensuring the safe, reliable, and fair use of AI while enabling its practical application in real-life scenarios. Medical research utilizing XAI has indicated that a visual explanation of selected features used in the decision-making process of the model provides impressive results [26]. AD diagnosis in existing studies has focused on classifying the condition as a task, with little attention paid to the time aspect of the data, using a single data modality or BL data. Existing studies on developing DL models that can provide explanations include the identification of explanatory feature maps via methods such as saliency maps or Class Activation Maps (CAMs). However, these techniques remain limited in representing the temporal dynamics associated with sequential data [27]. Gradient base-class activations such as CAMs are not readily applicable in the medical field because of their inability to provide voxel-level details of infected brain regions over time, which is a critical point in diagnosing neurodegenerative diseases such as AD.

This paper presents a hybrid DL framework comprising a lightweight deep convolutional neural network (DCNN) to extract deep features and combine them with bidirectional (Bi)LSTM to detect the progression of AD. The DCNN module provides a 2D summary image of the entire 3D MRI volume that represents the anatomical structure of the brain tissue of a patient. A summary image of the 3D MRI volume was generated by applying the temporal rank pooling technique[28], which compresses the spatial and inter-slice relationships of the 3D MRI volume into a single 2D image, known as a dynamic image. A 2D dynamic image was initially extracted from the 3D MRI volume at each time step of the patient's longitudinal data such as baseline (BL), month 6 (M6), and month 12 (M12), which were further processed using a CNN module to extract high-level representative features. The extracted sequence of deep features was subsequently passed through a Bi-LSTM module to learn the progressive deterioration of brain tissues across multiple time steps and predict the patient's health status at 48 months. In addition to the DL framework, we explore the effect of using multimodal data by incorporating the patient's cognitive scores from the baseline time step alongside the CNN-BiLSTM deep features. This integration aims to leverage additional information to obtain more accurate predictions. Furthermore, our study introduces XAI's unique solution for the chronological display of attention maps, which highlights the decline in brain tissue over time. This feature allows for a visual representation of the focused regions of the model and provides interpretability. Despite the limited existing literature addressing the longitudinal interpretability of networks, we extended the capabilities of our framework by incorporating a dedicated module that monitors spatial differences in the brain over time. This enhancement helps improve the accuracy of diagnosing AD cases. To achieve this, we employed a guided Grad-CAM technique that produced voxel-specific activation maps during each time step of 2D MRI slices. These maps provide valuable insights into areas of the brain that contribute to the predicted progression of AD. By combining DL techniques, multimodal data integration, and XAI-driven visualization, this study contributes to advancing the understanding and diagnosis of AD. Our findings can be summarized as follows:

- We propose a hybrid CNN-BiLSTM model that incorporates the concept of dynamic 2D images in a smart healthcare system to detect Alzheimer's disease progression in the time domain using longitudinal MRI volumes.
- Our DL architecture combines a summary of the 2D images extracted from each 3D MRI volume at longitudinal time steps (BL, M06, and M12). We fused these images with the patient's cognitive scores to predict AD progression at 48 months. By integrating both structural and functional changes, our multimodal approach enhances accuracy and captures the intricate interactions between brain regions and cognitive functions.
- To analyze longitudinal MRI data and gain insight into AD progression, we developed a novel XAI technique. This technique enables the generation of visual representations of time-related features, and aids physicians in tracking changes in patients over time. Additionally, our model exhibited high diagnostic performance, making it suitable for implementation as a decision support system in the healthcare industry.

Our model was extensively tested on the ADNI dataset using various settings. We performed comprehensive analyses to compare its performance with well-known DL architectures, such as ResNet50 [29], VGG16 [30], DenseNet121 [31], and EfficientNetB0 [32]. The results demonstrate that our model consistently outperforms all other models across multiple evaluation metrics and different scenarios. This paper proceeds as follows: In Section 2, we present the materials and methods; Section 3 outlines our proposed framework; Section 4 presents and analyzes our experimental outcomes; in Section 5, we demonstrate and evaluate our XAI method for detecting AD progression. Section 6 compares the proposed framework with existing state-of-the-art techniques; Section 7 discusses the limitations of the proposed system;finally, Section 8 concludes the discussion and suggests directions for future research.

2. Materials and methods

The development of our proposed AD progression detection method involved fusing 2D dynamic images produced from the 3D MRI volume of a patient's cognitive scores (CS) into a DL-based model consisting of a CNN combined with an LSTM model. Fig. 1 illustrates the workflow of the proposed framework.

2.1. Dataset

The dataset used in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a widely recognized open-source platform for research purposes [33]. Established in 2003 as a public–private partnership, the ADNI received an initial capital budget of $60 million allocated for a 5-year period. The primary objective of this program was to explore the feasibility of using serial MRI and PET scans along with other clinical assessments, biomarkers, and neuropsychological evaluations to track MCI progression and identify early indicators of neurodegenerative diseases. The early detection of AD progression using crucial biomarkers is benificial for physicians and researchers to develop novel therapies and enhance treatment effectiveness. It is worth noting that, unlike widely known datasets, such as the National Alzheimer's Coordinating Center (NACC) [34], Open Access Series of Imaging Studies (OASIS) [35], and Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) [36], the ADNI program gathered patient data at regular 6-month intervals. In addition, all MRI volumes underwent standard pre-processing, as depicted in the image pre-processing section shown in Fig. 1.

In this study, 1,692 (564×3) MRI volumes were collected at three distinct time points: BL, M06, and M12. Our model aimed to predict the change in patient status after a 3-year period, based on the final assessment visit, which occurred in month 48. The dataset consisted of

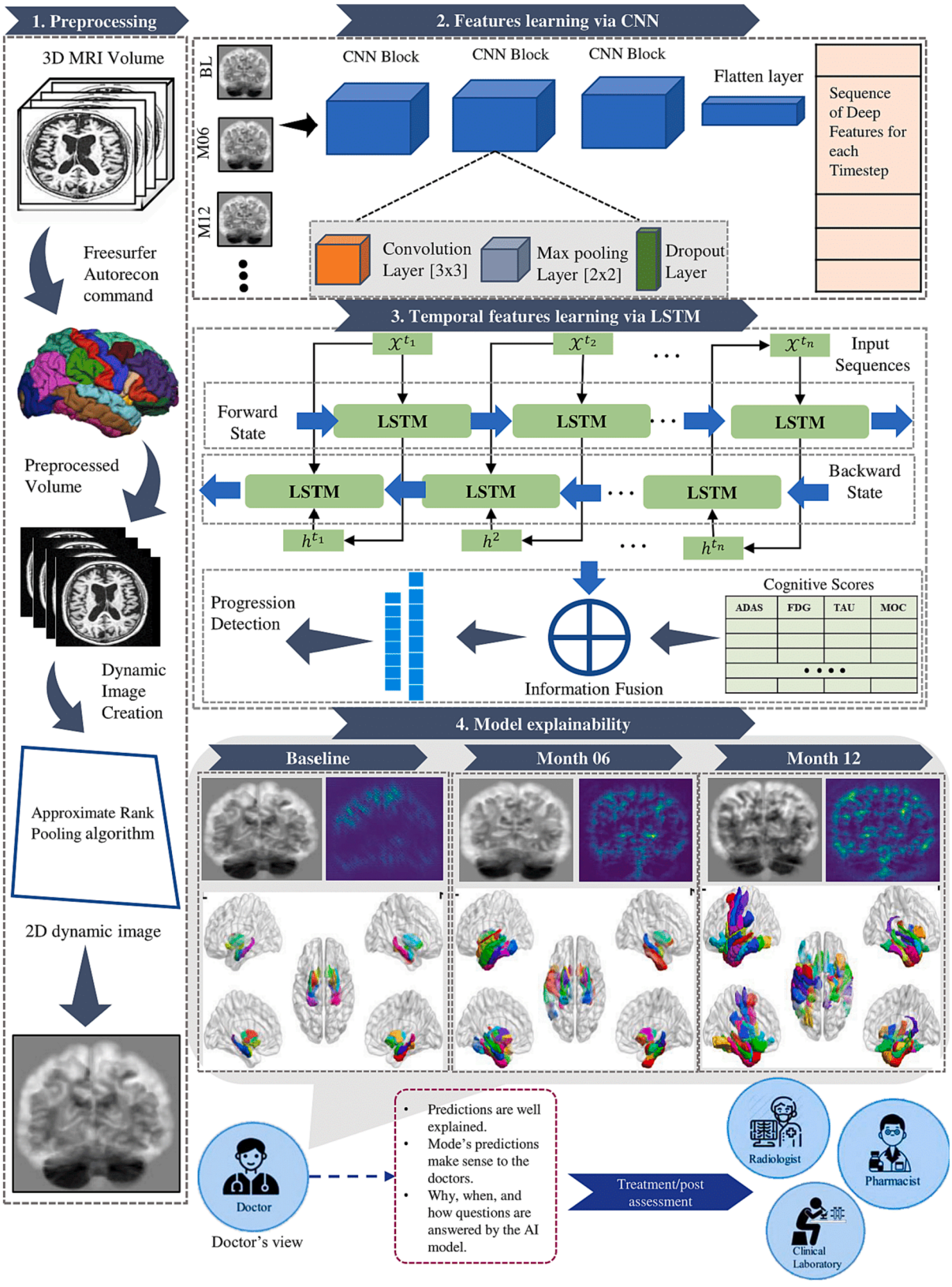

Fig. 1. A hybrid deep CNN-BiLSTM model for AD progression detection.

3T T1-weighted anatomical sequences acquired using the volumetric 3D MPRAGE protocol, featuring a voxel size of $1 \times 1 \times 1$ mm. The study included 564 participants, of which 282 consistently exhibited cognitive normality (CN) at each time point. Among these subjects, 100 individuals initially classified as CN at baseline experienced progression to AD within a three-year period, specifically at month 48 (M48) and starting from month 12 (M12). In contrast, 182 subjects consistently displayed AD symptoms throughout all visits. To form a comprehensive

AD class, we combined 100 participants who converted from CN to AD with 182 participants diagnosed with AD. The collective AD class is comprised of 282 individuals. Within this group, 182 subjects had AD from baseline to M48, and an additional 100 subjects transitioned from CN to AD at M48. This characteristic of the dataset aligns with similar studies [37,38] that designed their datasets to capture the progressive patterns of AD across longitudinal time steps. This sets these studies apart from many existing ones that treat AD as a binary classification task, distinguishing between CN and AD subjects.

While the goal of this study may align with existing research, our approach is distinct in terms of detecting whether a patient will progress towards AD or remain cognitively normal, based on the longitudinal time steps of their data. This unique perspective distinguishes our analysis and provides important insights for predicting AD progression. In addition to structural brain imaging (sMRI), this study explored the significance of CS as vital biomarkers. Several key scores were evaluated, including the unweighted sum of 13 items from the ADAS Cognitive Subscale (ADAS-Cog13), Functional Assessment Questionnaire (FAQ), Mini-Mental State Examination (MMSE), and Rey's Auditory Verbal Learning Test (RAVLT). Moreover, well-established biomarkers, such as apolipoprotein E4 (APOE4) and hippocampal volume were considered. These features have been widely employed in academic studies and clinical practice owing to their relevance and diagnostic value. Table 1 presents the initial statistical features of the study participants.

2.2. Generating dynamic 2D images

Recent advancements in AI-based health-care systems are mainly centered on comprehending the content within 3D MRI slicesbut do not necessarily prioritize the modeling of inter-slice dynamics [9]. This is accomplished using a strategy in which 3D volumes are treated as a sequence of frames. Subsequently, specialized models are trained to refine the essential information vital for resolving specific problems, such as diagnosing a particular brain disorder. A notable proportion of studies harness the capabilities of the 3DCNN model to facilitate the acquisition of spatiotemporal filters, which play a crucial role in optimizing the system's recognition ability [6,39]. These methods have achieved remarkable accuracy in disease identification. Their primary objective was to distinguish between healthy controls and diseased patients rather than to create a concise representation of a patient's volumetric data without losing valuable information. In contrast, dynamic images encode data in a general and content-agnostic manner, resulting in a long-term, robust representation of pixel-level changes [40]. In the case of MRI volumes, this novel representation condenses the voxel information encompassed in all 3D MRI slices into a single image [41]. Our experiments demonstrate the potency, efficiency, and simplicity of dynamic image representation in the context of 3D MRI volumes, particularly within the DL domain. To generate a dynamic 2D image for the entire 3D MRI volume, we adopted the temporal rank pooling (TRP) technique.

Table 1
Initial statistical features of the participating subjects in this study.

Features (Mean ± SD)	CN	Converted	AD
FAQ	69.26 ± 40.48	103.34 ± 51.58	159.74 ± 107.62
FDG	0.30 ± 00.50	0.50 ± 0.50	0.88 ± 0.72
MMSE	04.13 ± 05.87	03.63 ± 02.48	9.62 ± 03.64
MoCA	07.60 ± 31.41	0.37 ± 00.99	15.77 ± 24.63
APOE4	01.10 ± 00.75	0.30 ± 00.68	0.27 ± 00.76
ADAS-Cog13	06.34 ± 04.09	07.15 ± 03.14	20.64 ± 7.59
RAVLT	28.01 ± 08.20	29.42 ± 00.81	22.64 ± 07.47
RAVLT learn	43.36 ± 14.07	42.59 ± 07.64	21.37 ± 08.80
RAVLT forgetting	08.65 ± 10.22	05.57 ± 02.12	02.57 ± 04.45
CDRSB	41.36 ± 139.11	12.53 ± 12.48	35.72 ± 49.10
TAU	937.21 ± 663.5	488.15 ± 535.1	532.0 ± 401
PTAU	225.61 ± 248.8	128.96 ± 122.64	300 ± 209.16
Hippocampus	11.70 ± 29.18	93.82 ± 61.60	31.29 ± 92.90

FAQ = Functional assessment questionnaire; MMSE = Mini-mental state examination; FDG = F-fluorodeoxyglucose; MoCA = Montreal cognitive assessment; APOCE = apolipoprotein E gene; ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognitive subscale; RAVLT = Rey Auditory Verbal Learning Test; CDRSB = Clinical Dementia Rating scale Sum of Boxes; TAU = Tubulin associated unit; PTAU = Phosphorylated TAU.

TRP is a technique proposed by Fernando et al. [28] to obtain dynamic 2D images for video-based action recognition task. According to their work, a video is represented as a ranking function for its $I_1, I_2, I_3 \ldots I_{\mathscr{T}}$ frames. $\psi(I_t) \in \mathbb{R}^d$ represents the d-dimensional feature vector for each individual I_t frame in the input video. $V_T = \frac{1}{t}\sum_{\mathscr{T}=1}^{t}\psi(I_{\mathscr{T}})$ represents the average time of the features to time t and the score is assigned by the rank function $\mathbf{S}(\mathrm{t}|\mathbf{d}) = \langle \mathbf{d}, \mathrm{V_t} \rangle$ to each time t, where $d \in \mathbb{R}^d$ represents a vector of trainable parameters. The significance of each frame in the video is reflected by the ranking assigned to the learnable parameters, **d**. Later times are associated with larger scores as more frames are accumulated for the average so that $\mathrm{q} > \mathrm{t}$ implies $\mathrm{q} > \mathrm{t} \rightarrow \mathrm{S(q|d)} > \mathrm{S(t|d)}$. The constraints for the ranking problem can be represented as convex optimization problems using the RankSVM formula:

$$d^{*} = \rho(I_1, I_2, I_3 \cdots, I_t) = argmin\, E(d) \tag{1}$$

$$E(d) = \frac{\lambda}{2}||d||^2 + \frac{2}{T(T-1)}\sum_{q>t}\max\{0,\ 1 - \mathrm{S(q|d)} + \mathrm{S(t|d)}\} \tag{2}$$

The RankSVM formula involves a quadratic regularization term and a hinge loss term for computing incorrect rankings assigned to $\mathrm{q} > \mathrm{t}$ pairs. Specifically, the quadratic regularization term minimizes the complexity of the SVM model by penalizing large coefficients, whereas the hinge loss term ensures that the SVM model produces the correct rankings for any given pair. In other words, the hinge loss term counts the number of incorrect rankings and penalizes the SVM model. Note that a pair is correctly classified if the difference in scores between the two frames is greater than or equal to one. This is known as minimum margin constraint. The RankSVM formula can be used to generate dynamic images; however, it is computationally expensive. To address this issue, Bilen et al. [28] proposed a fast approximate rank-pooling strategy that uses a simple rank function to approximate the exact rank pooling. Specifically, they use a modified version of the temporal rank pooling formula as shown in the equation below:

$$\widehat{\rho}(I_1,\ I_2,\ I_3 \cdots,\ I_t; \psi) = \sum_{t=1}^{T}\alpha_t \cdot \psi\,(I_t) \tag{3}$$

The given equation shows how to calculate the temporal average of frames to time t using the symbol $\psi\,(I_t)$. The coefficients α_t associated with each frame are calculated as $2t - T - 1$. This method allows the efficient and effective extraction of temporal features from video data, which is useful for tasks such as action recognition and object tracking in the AI/DL domain. In the proposed study, we used a fast approximate rank polling strategy to generate dynamic 2D images from a 3D MRI volume. In the 3D MRI volumes, the z dimension served as the temporal dimension in the video file. Along with many other biomarkers for AD diagnosis, one such biomarker is the abnormal state of peptides and proteins in the cerebrospinal fluid (CSF) of AD patients [42]. In the AD diagnostic process, CSF is collected and tested using a medical process called a lumber puncture. When analyzing AD using MRI, the disturbed CSF is irregular in shape and size compared with the CSF in a normal person's brain. Fig. 2 shows an example of dynamic 2D image from a 3D MRI volume for each class of patient, that is, CN, patients who progressed from a CN state to AD (i.e., converted patients), and AD. The AD patient in Fig. 2 shows an irregular shape in the CSF compared to the CN and converted patients. The same applied to the dynamic image of the Converted Patient.

Table 2
Architectural design of the proposed CNN-BiLSTM model.

Layer Name	Kernels	$\begin{pmatrix}\text{Kernel size}\\ \text{Dropout}\end{pmatrix}$	Output size
Input	–	–	$110 \times 110 \times 1$
*Block*1	$\begin{pmatrix} Conv2D \\ Conv2D \\ MaxPooling \\ Dropout \end{pmatrix}$	$\begin{pmatrix} 3\times3\times1,8 \\ 3\times3\times1,8 \\ 2\times2 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 8\times110\times110 \\ 8\times110\times110 \\ 8\times110\times110 \\ 8\times55\times55 \end{pmatrix}$
*Block*2	$\begin{pmatrix} Conv2D \\ Conv2D \\ MaxPooling \\ Dropout \end{pmatrix}$	$\begin{pmatrix} 3\times3\times1,16 \\ 3\times3\times1,16 \\ 2\times2 \\ 0.2 \end{pmatrix}$	$\begin{pmatrix} 8\times55\times55 \\ 16\times55\times55 \\ 16\times55\times55 \\ 16\times27\times27 \end{pmatrix}$
*Block*3	$\begin{pmatrix} Conv2D \\ Conv2D \\ Conv2D \\ MaxPooling \\ Dropout \end{pmatrix}$	$\begin{pmatrix} 3\times3\times1,32 \\ 3\times3\times1,32 \\ 3\times3\times1,32 \\ 2\times2 \\ 0.3 \end{pmatrix}$	$\begin{pmatrix} 32\times27\times27 \\ 32\times27\times27 \\ 32\times27\times27 \\ 32\times13\times13 \end{pmatrix}$
*Block*4	$\begin{pmatrix} Conv2D \\ Conv2D \\ Conv2D \\ MaxPooling \\ Dropout \end{pmatrix}$	$\begin{pmatrix} 3\times3\times1,64 \\ 3\times3\times1,64 \\ 3\times3\times1,64 \\ 2\times2 \\ 0.3 \end{pmatrix}$	$\begin{pmatrix} 64\times13\times13 \\ 64\times13\times13 \\ 64\times13\times13 \\ 64\times6\times6 \end{pmatrix}$
FlattenLayer	–	–	13826
BiRNN Layer	$\begin{pmatrix} \#LSTM\ layers \\ \#LSTM\ cells \end{pmatrix}$	$\begin{pmatrix} 1 \\ 512 \end{pmatrix}$	$1024 + 14$
Fully connected Layer	$\begin{pmatrix} \#of\ Dense\ units \\ Activation\ funtoin \end{pmatrix}$	$\begin{pmatrix} 128 \\ ReLU \end{pmatrix}$	128
Fully connected Layer	$\begin{pmatrix} \#of\ Dense\ units \\ Activation\ funtoin \end{pmatrix}$	$\begin{pmatrix} 64 \\ ReLU \end{pmatrix}$	64
Output Layer	$\begin{pmatrix} \#of\ Dense\ units \\ Activation\ funtoin \end{pmatrix}$	$\begin{pmatrix} 1 \\ Sigmoid \end{pmatrix}$	1

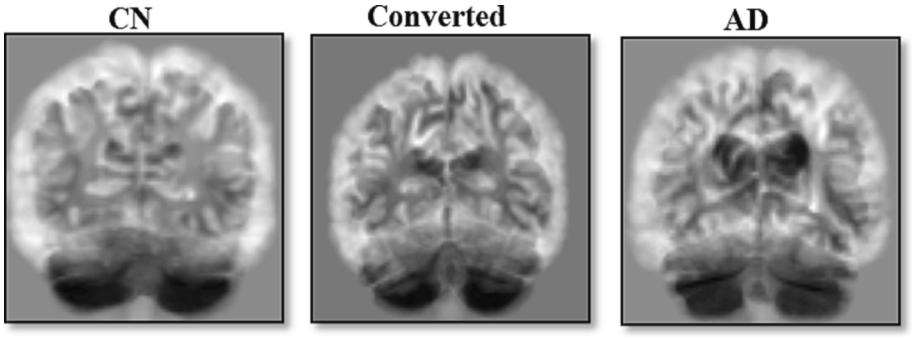

Fig. 2. A dynamic 2D image from CN, Converted and AD classes at BL ~ M12 time step.

2.3. Deep convolutional neural networks

In this study, we used deep convolutional neural networks (DCNN) to generate high-level representative features of the brain tissues from dynamic 2D images. CNNs are an advanced form of artificial neural networks (ANNs) that automatically engineer data features. CNNs combine low-level features with abstract high-level features to create a generalized feature space for the input data samples. Compared with conventional approaches, CNNs have shown better results in several detection and recognition tasks. A CNN imitates the human visual cortex to memorize and learn the visual cues and features appearing in an image, thereby creating an effective representation of data. The CNN model exhibits three main characteristics: local connectivity, parameter sharing, and invariant representation. The nature of a CNN significantly reduces the computational complexity required to analyze high-dimensional data, thus simplifying complex tasks. The main components of the CNN models are convolution, pooling and a fully connected layer. The input image is processed in our model, starting with the extraction of deep features using a convolutional layer. The pooling layer reduces the dimensions of the resulting feature map before it passes through the BiLSTM network. Finally, the output of the LSTM was fed into a dense neural network before producing the final prediction score.

2.4. Long-short term memory

ANNs capable of processing and interpreting sequential input data are known as recurrent neural networks (RNNs), and are commonly employed in natural language processing tasks [43]. These networks are mainly designed to analyze long-text datasets while preserving any contextual information they contain, which is crucial for performing tasks such as time-series prediction, text synthesis, text categorization, and machine translation. Although RNNs are adept at capturing sequential temporal attributes, they face the problem of vanishing gradients, which makes it challenging for them to learn long sequential dependencies from training data. Consequently, learning long sequential dependencies from the training data is a challenging task for RNNs. LSTM is an improved version of RNN that addresses the problem of vanishing and exploding gradients. The internal mechanism of LSTM uses cell units called gates, which regulate the flow of information between layers. These gates are responsible for determining which sequences are important for preservation and which sequences are less important and can be eliminated. In simple LSTM, the hidden state h_t, and memory cell C_t, are functions of their previous statuses, h_{t-1},

and C_{t-1}, and the input vector W_t. The hidden state of each location (h_t) considers only the forward context without considering the backward context, as shown in Eq. (4):

$$C_t, h_t = g^{LSTM}(C_{t-1}, h_{t-1}, W_t) \tag{4}$$

In our proposed framework, we used a bi-directional LSTM that considers forward and backward sequences simultaneously to capture the interdependencies in both directions. BiLSTM uses two parallel channels simultaneously, eventually concatenating hidden states into a single vector. The BiLSTM formula is as follows:

$$\overrightarrow{C_t}, \overrightarrow{h_t} = g^{LSTM}\left(\overrightarrow{C_{t-1}}, \overrightarrow{h_{t-1}}, W_t\right) \tag{5}$$

$$\overleftarrow{C_t}, \overleftarrow{h_t} = g^{LSTM}\left(\overleftarrow{C_{t-1}}, \overleftarrow{h_{t-1}}, W_t\right) \tag{6}$$

The parameters are shared in both the forward and backward LSTMs. $[\overrightarrow{h_n}, \overleftarrow{h_1}]$ represents the entire sequence, and n represents the length of each sequence. $h_t = \overrightarrow{h_t} + \overleftarrow{h_t}$ represents the concatenated feature vector of the forward and backward LSTM, respectively, at position t. Thus, both the forward and backward sequences were considered simultaneously.

2.5. Model explainability

Grad-CAM is a technique used in computer vision and DL algorithms to visualize regions of an image that are relevant for a specific class prediction. In Grad-CAM, gradients flow backward through the network and are used to compute attention weights or relevance scores for each spatial location in a convolutional feature map. These weights are then combined withthe feature map values to obtain class activation maps, highlighting the regions most relevant to the predicted class. However, Grad-CAM treats both positive and negative gradients equally, which can result in noisy and less interpretable visualizations. In contrast, Guided Grad-CAM is an upgraded version of Grad-CAM that addresses the aforementioned limitations by incorporating a guided backpropagation technique [43]. The guided backpropagation process considers only positive gradients during backpropagation and blocks negative gradients. This was achieved by modifying the ReLU activation function in the network by setting the negative gradients to zero. By backpropagating only positive gradients, guided Grad-CAM produces more focused and visually appealing visualizations, emphasizing the regions that positively contribute to the prediction. In neuroimaging analysis, the accurate identification of voxels affectected by a certain disease is crucial. To acquire information on the exact voxels involved in accurate diagnostic prediction in the model, we employ the guided Grad-CAM technique. The following are the mathematical representations of the Grad-CAM and Guided Grad-CAM techniques.

Grad-CAM Equations:

$\alpha(k) = \frac{1}{H\times W}\sum i \sum j \frac{\partial Y^c}{\partial A}(i,j,k)$// Attention weights

$M = ReLU(\sum k\alpha(k)A(k))$// Class activation map

Guided Grad-CAM Equations:

$G(i,j,k) = \frac{\partial Y^c}{\partial A}(i,j,k) if \frac{\partial Y^c}{\partial A}(i,j,k) > 0, else 0$// Guided gradients

$\alpha(k) = \frac{1}{H\times W}\sum i \sum j G(i,j,k)$// Attention weights

$M = ReLU(\sum k\alpha(k)A(k))$// Class activation map

In both Grad-CAM and guided Grad-CAM, $\alpha(k)$ represents the attention weights (or relevance scores) for each spatial location (i, j) in the k^{th} channel of the final convolutional layer. The class activation map M was obtained by multiplying these attention weights by the corresponding feature map values A and applying the *ReLU* activation function.

We calculated attention maps for each time step to determine the extent of brain damage occurring at a specific point in time. To accomplish this, gradients were initially computed using a guided grad-cam technique and then overlaid onto the dynamic images for each time step. Fig. 6 presents the explainable output feature maps for each time step, illustrating the progressive atrophy of brain damage over time. Further details regarding the proposed framework, including the preprocessing of MRI volumes and the architectural design of the model, are discussed in the subsequent sections.

3. Proposed framework

We proposed an efficient network architecture to detect AD progression using compressed 2D representations of 3D MRI volumes from longitudinal data. The proposed network is an end-to-end CNN-BiLSTM model that calculates deep features for each time step and detects a progression pattern in the extracted deep features using bidirectional LSTM. The CNN module for the proposed network extracts the deep features using each time step, thatis , baseline, M06, and M12, and then passes each feature vector to the BiLSTM module, which captures the temporal features in each sequence and detects whether the subject is progressing towards AD or remains cognitively normal. In addition, we investigated the efficiency of multimodal data, that is, CS features fused with the output of the BiLSTM network, before they are classified using a deep neural network. To summarize, the propseod framework comprises of four stages. Stage1 performs the basic preprocessing steps and dynamic image creation from the MRI volume; stage 2 performs deep feature extraction from the dynamic 2D image; stage 3 performs temporal feature learning via the LSTM model and data fusion of multimodalities; and stage 4 performs a visual explanation of the AD progression using the proposed time series visual explainability approach.

3.1. Image preprocessing and dynamic image creation

The pre-removal of irrelevant information from the raw MRI volume during the preprocessing stage facilitates the comparison of different brain scans. To achieve this goal, the autorecon1 command from the FreeSurfer tool [44] was applied to perform the essential pre-processing steps on the raw MRI volumes. FreeSurfer is a software suite that is commonly used to analyze and visualize structural and functional neuroimaging data. FreeSurfer provides an autorecon-all command that performs cortical reconstruction, including white mater and gray mater segmentation, surface generation from segmented data, and spherical or flattened representation of cortical surfaces. Autoerecon1 is a subset of autorecon-all, which performs motion correction, intensity normalization, and skull stripping. In our case, after performing auteorecon1, the output volume was further registered to MNI152 using FLIRT to align each image to the common template space and simplify the process of comparing different images. It has been shown in the literature that non-preprocessed volumes can significantly reduce the performance of the DL model because the skull variance is treated as noise in the MR slices. Next, we focused on the coronal slices of each MRI volume because they carry the most discriminating AD-related information in brain tissues [45]. Of the 256 coronal slices in each 3D MRI volume, 110 were collected from the middle of the 3D volume, eliminating the very top slices that carry less structural information about the tissues involved in AD. We then passed this newly created volume of size ($110 \times 110 \times 110$) through the dynamic rank-pooling algorithm discussed in Section 2.2. In this way, we compressed all temporal information from the 3D volume into a single 2D image while preserving structural information about the brain tissues. The dynamic 2D image was then processed through the DCNN to learn the spatial hierarchies of the deep features at each longitudinal time step of the training data (BL, BL ~ M06, and BL ~ M12).

3.2. Features learning using a convolutional neural network

The proposed framework encompasses an end-to-end CNN-BiLSTM network that serves as the core of the AD progression detection model.

This network was designed to extract deep features from 2D dynamic images and leverage the temporal information captured by the BiLSTM network. The architectural design of each submodule in the proposed hybrid DL framework was designed to ensure efficient processing of the most representative deep features, thereby enabling the accurate detection of AD progression. In designing the proposed hybrid DL framework, we searched for the optimal DL architecture through a combination of experimentation and architectures proposed literature [17,20,46,47]. Our objective was to strike a balance between model complexity and performance improvement. We followed a systematic experimental process to set different parameter values. We performed a wide array of experiments and varied individual parameters such as the number of convolution layers, number of kernels in each convolution layer, and number of LSTM layers in the LSTM module, while observing their effects on different aspects of the model's performance. After completing this initial phase, we employed a Bayesian optimizer to fine-tune additional parameters, including the number of recurrent units within the LSTM head, the dropout rate at different layers, and the learning rate. This sophisticated optimization process contributed significantly in refining the performance of our model.

The CNN module, which was responsible for deep feature extraction, was composed of ten convolution layers and four max-pooling layers organized in blocks (refer to Fig. 1). The purpose of these blocks was to progressively capture and refine the most relevant features at each layer, while also reducing the spatial dimension of the input data as the network deepened. The first block consisted of two consecutive convolution layers, with each layer employing eight kernels. These convolution layers were followed by a max-pooling layer, which was responsible for down-sampling the spatial dimensions, and a dropout layer with a Bayesian optimized threshold value of 0.2, which helped prevent overfitting by randomly dropping a fraction of the connections during training. Similarly, the second block comprised two convolution layers, each equipped with 16 kernels, followed by a max-pooling layer and a dropout layer with a threshold value of 0.2. The third block incorporated three convolution layers, each utilizing 32 kernels. Following the convolution layers, a max-pooling layer and a dropout layer with a threshold value of 0.3 to enhance regularization. The last block, which was designed to capture the most intricate features, was constructed using three convolution layers, each employing 64 kernels. As in the previous blocks, a max-pooling layer and a dropout layer with a threshold value of 0.3 were incorporated.

At each convolutional layer, a 3×3 convolutional kernel was utilized, which has been proven effective in capturing spatial patterns in various computer vision tasks. The ReLU activation function was applied after each convolution layer to introduce non-linearity, thereby enabling the model to learn complex representations. It is noteworthy that the max-pooling layer was not applied after every convolutional layer. Instead, it was used after every two convolutional layers. This configuration aimed to retain the spatial features and enable the network to learn a broader range of spatial information from dynamic 2D images. However, this approach can lead to longer training times and increased risk of overfitting the model to the training data. To address potential overfitting, a dropout layer was introduced after each convolutional block. These dropout layers served as regularization techniques by randomly deactivating a portion of neurons during training, thereby preventing them from relying accessively heavily on specific features and promoting more robust learning.

Following the convolutional layers, the output from the last convolutional layer at each time step was reshaped into a single-dimensional column vector. This column vector, with size N, represented the compressed and abstracted deep features at each time step (BL, M6, and M12). These time-dependent column vectors were then passed through a Bi-LSTM network, which aimed to capture the temporal dependencies and progressive deterioration of brain tissues over time.

3.3. Sequence learning and multimodal data fusion via RNN

The deep sequences obtained from the CNN module were further processed using the BiLSTM network, as discussed in the previous section. The objective of this step was to capture the temporal dependencies across longitudinal time steps. In this study, the Bi-LSTM module was configured with a single layer comprising 512 LSTM cells in the Bi-LSTM layer, which operated in both forward and backward directions. To introduce non-linearity into the CNN output data within the LSTM layer, a hyperbolic tangent (tanh) activation function was applied. This activation function enabled the Bi-LSTM module to learn complex temporal patterns from the deep sequences. After passing through the BiLSTM layers, the outputs from both the forward and backward LSTM layers were concatenated into a single vector. This merging operation resulted in the formation of the output of the overall Bi-LSTM subnetwork. The cumulative output from the BiLSTM module was then fused with the CS feature of the patients as, listed in Table 1. This fusion or combination of the Bi-LSTM outputs with the CS feature allows for the incorporation of additional relevant patient-specific information into the predictive model.

The combined feature set, consisting of the Bi-LSTM output and CS features, was subsequently processed using a dense neural network. This network comprised two hidden layers: one with 128 units and the other with 64 units. The purpose of these hidden layers was to extract higher-level representations and features from the combined input data. The specific number of units in each hidden layer was chosen to strike a balance between model complexity and performance. Finally, the output of the last hidden layer, which was a 64-dimensional vector, was passed through a single-unit output layer. The output layer was equipped with a sigmoid activation function that squashed the output value between zero and one. This output score served as an indicator of the model's prediction, where values closer to zero suggested a prediction for a patient with CN, whereas values closer to one indicated progression to AD. By employing this comprehensive architecture, the model leveraged the strengths of both the CNN and Bi-LSTM modules to learn and represent important temporal and patient-specific features. Additionally, by incorporating state-of-the-art XAI methodologies, we conducted further analysis of the proposed network, facilitating a comprehensive exploration of its internal operations. To achieve transparency in the decision-making process, our XAI method generated attention maps for each stage of the input data, highlighting the active regions of the brain. A detailed technical discussion is presented in subsequent sections.

3.4. Spatial and temporal explainability

The attention maps generated by the proposed network played a vital role in explaining the DL model. Attention maps play a pivotal role in understanding DL models [48] and offer insights into their decision-making processes [49]. These maps visually highlight the areas of input data on which the model focuses, providing the reasoning behind its decisions. Used in domains like image recognition and medical diagnostics [50,51], attention maps reveal what influences the model's predictions. They act as interpretable guides, fostering transparency and trust in the model's decisions by exposing biases and enhancing the alignment with human intuition [52]. For more details on attention map explainability, readers are referred to the following study [51]: Utilizing the guided Grad-CAM technique, the proposed network generates attention maps that provide insights into the explainability of the DL model. At each stage of the analysis, involving 2D compressed representations of the MRI slices, the attention maps provided detailed voxel-level representations of the activation map. Furthermore, in addition to the 2D representation of the highlighted brain tissues, we showcased a 3D view of the same activated brain regions, highlighting different brain regions from a 3D perspective. The 3D-rendered brain surface shown in Fig. 6 was generated using a technique called surface reconstruction, which involves combining a set of 3D points to form a surface mesh. This

mesh can then be rendered using various 3D graphics techniques to create realistic visualizations of the brain. Fig. 6 presents a detailed visualization of the 2D attention maps and 3D rendered brain surface, offering an illustrative representation of the infected regions contributing to the disease identification process in the framework.

4. Experiments and results

The experiments in this study were performed using an NVIDIA TITAN GTX GPU with 12 GB of memory, and the proposed models were implemented using the TensorFlow 2.0 library. The proposed model was trained in an end-to-end manner with the Adam optimizer at the best learning rate of 0.0001, as suggested by the Bayesian optimizer, while the remaining parameters such as momentum, and weight decay were maintained at their default values [53]. The total loss was computed using a binary cross-entropy function. To achieve optimal performance, the size of the input image was specified as $110 \times 110 \times 1$ in the gray-scale image, and the number of images per batch was set to 32. The model was trained using a stratified 10-fold cross-validation approach in which the training data were divided into 75 % for the training set and 25 % for validation set at each fold. The model was initially trained for 80 epochs to determine the optimal number of parameters for LSTM units, dropout threshold, and a learning rate. After obtaining the optimal number of parameters using the Bayesian optimizer, the model was trained for 150 epochs using fine-tuned parameters for each fold. The training process was further constrained by applying an early stopping technique to prevent unnecessary training, which could potentially result in the overfitting of the model. The stratified N-fold method ensured a balanced number of samples in each batch. This approach also facilitates the calculation of several performance metrics, including mean accuracy, mean AUC, mean F1 scores, mean precision, and mean recall [54]. Moreover, to ensure a fair comparison, all hyperparameters were set in the same manner for all other comparative models.

To examine in contrast to the proposed network, we compared its results with ResNet50, VGG16, DenseNet121, and EfficientNetB0, which are among the most powerful DL models in the scientific community. The selection criterion for the comparative models was based on as analysis of how different architectural designs of the CNN models affect the detection of AD progression. We observed that a wide range of medical studies, particularly on the classification of medical images, has used shallow models that rely on hand-designed features such as shape, color and texture [55,56]. However, the main problem with these approaches is that the extracted features are low-level and do not represent the concept of a high-level problem domain. In addition, the generalizability of these models is poor. However, deep models have been successful in various domains, including medical and non-medical domains, owning to their significant success. They are particularly known as excellent feature extractors; therefore, using them to classify medical images avoids complicated and expensive feature engineering processes. The comparative models used in this study are state-of-the-art frameworks with excellent performance in image classification and image recognition tasks.

AlexNet first won the challenge of ImageNet, which is an image classification task. Subsequently, the DL-based approach began to explode and achieve breakthroughs in various domains, including speech recognition, language translation, and disease classification. Following AlexNet, other models, such as VGGNet and ResNet, have improved the limitations of deep CNN models, such as the problems of vanishing and exploding gradients. The VGG network introduces blocks of convolutional layers with smaller 3×3 kernels throughout the network. This architectural design achieves a dual purposes of reducing the number of overall learnable parameters and improving network accuracy. The ResNet model further improves accuracy by introducing a skip connection. These connections allow the network to skip one or more layers, which helps to address the vanishing gradient problem. This mechanism creates a shortcut path for the gradient flow through the network. In the DenseNet model, the gradinet flow problem was solved by introducing dense connections.

Dense connections are a key feature of the DenseNet architecture, in which each layer is connected to all preceding layers. This dense connectivity pattern allows feature maps of earlier layers to be used directly by later layers, promoting feature reuse and improving the gradient flow. Specifically, the output feature maps of each layer are combined with the feature maps of all the preceding layers and forwarded as inputs to the next layer. This helps maintain the diversity of feature maps across the network, which can prevent overfitting and improve accuracy. In the case of EfficientNet, the problems of model overfitting and high accuracy are handled differently. EfficientNet employs a complex scaling method that systematically scales the depth, width, and resolution of a network. The network depth increases with the addition of more layers, and its width increases with the number of filters in each layer. This method ensures that the network can learn more complex features without compromising computational efficiency. The unique architectural design of EfficientNet makes it suitable for numerous applications. Because these networks use powerful features and represent the capabilities of deep CNN models, we chose them as the baseline models against which we tested the proposed CNN-Bi-LSTM model.

We conducted multiple experiments to examine the impact of various factors on AD progression. Fig. 3 illustrates the role of longitudinal and multimodal data in AD diagnosis. The experiments conducted in this study encompassed single-modality medical data, multi-modality medical data, longitudinal data analysis of individual patients, and different architectural DNN designs. In the following sections, we present the experimental setup, which addresses the aforementioned key points in the following pipeline: 1) Progression detection utilizing a single modality of longitudinal data such as 2D dynamic images, with diverse architectural designs like backbone feature extractors, and 2) Progression detection using multiple modalities of longitudinal data such as 2D images plus CS with diverse architectural designs as backbone feature extractors.

4.1. Evaluation metrics

To assess the generalizability of the proposed framework to training data, we conducted a thorough evaluation using various performance metrics. These metrics included accuracy, precision, recall, F1 score, and AUC. Each metric is defined mathematically using the following equation:

The accuracy metric quantifies the proportion of correctly identified samples (CN/AD) in the predicted data. It is calculated by dividing the sum of true negatives (TN) and true positives (TP) by the total number of samples (TS).

$$Accuracy = \frac{(TN + TP)}{TS} \tag{7}$$

Precision denotes the ratio of accurately classified patients with AD (TP) to the total number of predicted positive samples (AD). Its computation is achieved by dividing TP by the sum of TP and false positives (FP):

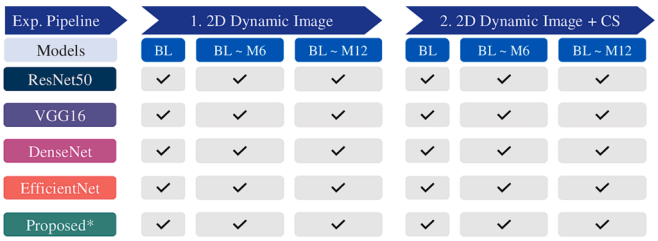

Fig. 3. The experimental route map with single and multimodal data of the proposed framework.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The recall metric, often referred to as the sensitivity or true-positive rate, represents the ratio of correctly classified patients with AD (TP) to the total number of patients with AD in the dataset. This was determined by dividing TP by the sum of TP and false negatives (FN):

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The F1 score served as a combined measure of precision and recall, providing a weighted average that accounted for both metrics. Its calculation involves multiplying precision and recall and then dividing the result by the sum of precision and recall, multiplied by 2:

$$F1 - score = 2^{*}\frac{Precision^{*}Recall}{Precision + Recall} \tag{10}$$

Finally, the AUC was used to assesses the performance of the model across various classification thresholds, characterizing the relationship between the true positive and false positive rates. The evaluation results of the model at different thresholds were represented by the AUC score. In the equation above, TP refers to true positives, FP refers to false positives, TN represents true negatives, and TS corresponds to the total number of samples.

4.2. Single modality-based detection of AD progression

In Experiment 1, several well-known DL models were evaluated to determine their effectiveness in detecting AD progression. The DL models used in this study included ResNet50, VGG16, DenseNet121, EfficientNet, and the proposed network. Each model was used as a feature extractor for the time steps (BL, M6, and M12), with deep features from each time step processed using Bi-LSTM to detect AD progression. Deep feature extraction and sequence learning methods are discussed in detail in Sections 3.2 and 3.3, respectively. We also compared the output performance of each model with that of the proposed network. The primary objective of Experiment 1 was to examine the effectiveness of a single data modality, specifically MRI, in detecting AD progression . To achieve this, we computed several evaluation metrics, including the accuracy, precision, recall, F1 score and, AUC averages, to conduct a comprehensive assessment of each model. Furthermore, we have included a separate section dedicated solely to comparing the achieved AUC of each model. This analysis aimed to evaluate the impact of augmenting the training dataset with longitudinal time steps on the network stability.

4.2.1. Evaluation results using a dynamic 2D image

Table 3 presents the statistics of the various evaluation metrics collected for each comparative model trained on a single training modality such as dynamic 2D images. Initially, each comparative model was evaluated using a single time step of the training data, BL only, which was then combined with subsequent time steps, BL + M06 and BL + M06 + M12. In this study, the ADNI dataset was composed of three-time steps of longitudinal MRI data collected for each patient at 6-months intervals. By including additional time points in the training data, the model should theoretically enhance the disease identification process, as it can observe more time-series data for the same patient, reflecting the progressive patterns of the disease diagnostic process.

Table 3 presents the achieved performance of each model at different time steps, namely, BL, BL + M06 and BL + M06 + M12. The results indicated that when trained on BL data, EfficientNet exhibited the highest performance, with a mean performance of accuracy = 0.89 ± 0.02, AUC = 0.91 ± 0.02, F1 score = 0.88 ± 0.03, precision = 0.89 ± 0.04, and recall = 0.87 ± 0.03. The proposed network followed closely with the second-best accuracy, recording a mean accuracy of 0.83 ± 0.03, mean AUC of 0.84 ± 0.02, mean F1 score of 0.82 ± 0.03, mean precision of 0.85 ± 0.03, and mean recall of 0.81 ± 0.04. DenseNet121 outperformed ResNet50 and VGG16, although it reported a lower accuracy in BL than of EfficientNet and the proposed network.

Table 3
Evaluation results for various backbone models using dynamic 2D images.

Backbone network	Time Steps	Accuracy	AUC	F1 Score	Precision	Recall
ResNet50 [29]	BL	0.75 ± 0.03	0.83 ± 0.04	0.79 ± 0.03	0.83 ± 0.05	0.74 ± 0.04
	BL ~ M06	0.82 ± 0.02	0.82 ± 0.05	0.83 ± 0.04	0.83 ± 0.03	0.84 ± 0.02
	BL ~ M12	**0.85 ± 0.04**	**0.84 ± 0.04**	**0.85 ± 0.06**	**0.87 ± 0.06**	**0.86 ± 0.05**
VGG16 [30]	BL	0.77 ± 0.04	0.76 ± 0.04	0.76 ± 0.03	0.79 ± 0.04	0.73 ± 0.03
	BL ~ M06	0.78 ± 0.02	0.80 ± 0.04	0.76 ± 0.05	0.82 ± 0.05	0.72 ± 0.06
	BL ~ M12	**0.84 ± 0.04**	**0.85 ± 0.03**	**0.83 ± 0.04**	**0.86 ± 0.04**	**0.84 ± 0.04**
DenseNet121 [31]	BL	0.79 ± 0.3	0.79 ± 0.03	0.78 ± 0.02	0.82 ± 0.02	0.75 ± 0.03
	BL ~ M06	0.83 ± 0.03	0.82 ± 0.02	0.81 ± 0.03	0.83 ± 0.02	0.81 ± 0.04
	BL ~ M12	**0.88 ± 0.04**	**0.89 ± 0.03**	**0.87 ± 0.03**	**0.87 ± 0.04**	**0.86 ± 0.30**
EfficientNet [32]	BL	0.89 ± 0.02	0.91 ± 0.02	0.88 ± 0.03	0.89 ± 0.04	0.87 ± 0.04
	BL ~ M06	0.91 ± 0.03	0.92 ± 0.02	0.90 ± 0.03	0.91 ± 0.03	0.89 ± 0.03
	BL ~ M12	**0.90 ± 0.04**	**0.90 ± 0.04**	**0.88 ± 0.02**	**0.88 ± 0.06**	**0.88 ± 0.03**
Proposed Network	BL	0.83 ± 0.03	0.84 ± 0.04	0.82 ± 0.03	0.85 ± 0.03	0.81 ± 0.04
	BL ~ M06	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.04	0.93 ± 0.03	0.87 ± 0.03
	BL ~ M12	**0.92 ± 0.03**	**0.94 ± 0.03**	**0.93 ± 0.03**	**0.95 ± 0.02**	**0.93 ± 0.02**

Bold text indicates the best results.

When trained with two-time steps of the training data, BL + M06 all models demonstrated a notable enhancement in accuracy. Among the comparative models, EfficientNet outperformed all others, reporting an average accuracy of 0.91 ± 0.03, AUC of 0.92 ± 0.02, F1 score of 0.90 ± 0.03, precision of 0.91 ± 0.03 and recall of 0.89 ± 0.04. However, our proposed network outperformed EfficientNet in the average precision score, reporting 0.93 ± 0.03. This suggests that the proposed network may be a better choice when precision is a critical metric. Moreover, with two-time step training data, ResNet50 surpassed VGG16 and achieved an accuracy similar to that reported by DenseNet121. ResNet50 showed a rapid improvement in accuracy when training with two-time steps of the training data, compared to training with BL data alone. The average performance reported by ResNet50 at BL + M12 are as follows: accuracy = 0.82 ± 0.02, AUC = 0.82 ± 0.05, F1 score = 0.80 ± 0.04, precision = 0.83 ± 0.03 and recall = 0.84 ± 0.02. The improvement in the reported accuracies with longitudinal time steps of the training data refers to the model's ability to capture progressive patterns from multiple time steps.

To further test the diagnostic abilities of the models, we trained each

model with three longitudinal time steps, including BL, M06, and a follow-up visit of month 12 (M12) training data (BL + M06 + M12), and evaluated their performance. Our proposed network outperformed all comparative models with an average performance of accuracy = 0.92 ± 0.03, AUC = 0.94 ± 0.03, F1 score = 0.93 ± 0.04, precision = 0.95 ± 0.02 and recall = 0.93 ± 0.03. Interestingly, ResNet50, VGG16, and DenseNet121 continued to improve in accuracy in the same incremental manner, as they showed improved accuracy with single- and two-time steps of training data. However, EfficientNet reported the worst accuracy when combined with three-time steps of training data, indicating unstable behavior. The model was unable to report the expected behavior and showed a significant degradation in the achieved performance. By combining the training data from the three-time steps, the model achieved an accuracy of less than 90%, compared with the accuracies achieved in the previous time steps, which were greater than 90%. The findings indicate that the proposed network outperformed all other comparative models in terms of achieving accuracy using three longitudinal time steps, except for EfficientNet, which displayed unstable behavior. The results of this study suggest that models that can capture longitudinal patterns from multiple time steps provide improved diagnostic capabilities for a given prediction task.

Based on the available context, it is important to highlight that although all models achieved acceptable accuracy, they displayed high variance and stability issues, as indicated by the fluctuations in their standard deviations. This unstable behavior was probably due to the presence of noise in the training data. However, after integrating the CS features with the longitudinal MRI data during the training process, we observed an improvement in the stability of each model, which was evident in the subsequent experiment.

4.2.2. *Model comparison using dynamic 2D images*

The effectiveness of the proposed model is demonstrated in Fig. 4, which shows the impact of incorporating additional time steps such as longitudinal data into the training dataset. We limit our discussion and comparison of our results to the mean AUC metric due to the limitation and homogeneity of the results. Fig. 4 shows the results obtained using only MRI data. We collected and compared the mean AUCs of different models using different combinations of data: BL, BL ~ M06, and BL ~ M12.

This study found that EfficientNet achieved the highest mAUC score with BL of any other comparative model, achieving a 91% AUC score. The second-highest mAUC score was reported by proposedNet, with an mAUC score of 84%. Interestingly, ReNet50 outperformed VGG16 and DenseNet121, with an AUC score of 83%, but it was less accurate than that of EfficientNet and proposedNet. At this point, the results suggest that, while ReNet50 may perform well in single-time-step training data, it may not be the best choice when two-time-step training data are used. Furthermore, the study found that, with two-time steps of training data, all models except ResNet50 achieved an improvement in their AUC scores. VGG16, DenseNet121, and ProposedNet reported noticeable improvements of 4%, 3%, and 6%, respectively, whereas EfficientNet achieved only 1% improvement. However, ResNet50 did not exhibit improved performance with additional training data.

In BL ~ M06, the mAUC scores for VGG16, DenseNet121, and the ProposedNet were 80%, 82%, and 90%, respectively, compared to 76%, 79%, and 84%, respectively, at the baseline time step. This indicates that these models benefited significantly from additional training data. However, EfficientNet achieved only a 1% improvement in its mAUC score, whereas ResNet50 and DenseNet121 achieved modest improvements of 3% and 1%, respectively. In BL ~ M12, all comparative models reported a significant improvement in mAUC scores, except for EfficientNet, which decreased by 2% after combining training data from the three-time steps. By contrast, ProposedNet reported a 4% improvement, from 90% to 94% in its mAUC score. Among the other models, DenseNet121 achieved a significant improvement with three-time steps of training data, improving its performance by 7% in terms of the mAUC score. ResNet50 and VGG16 also reported improvements of 3% and 5% in the BL ~ M12 time step, respectively.

4.3. *Multimodality-based detection of AD progression*

In the second experiment, we explored the effects of the clinical scores on the detection of AD progression by combining MRI and CS. Assessment of a patient's cognitive abilities is critical for detecting AD

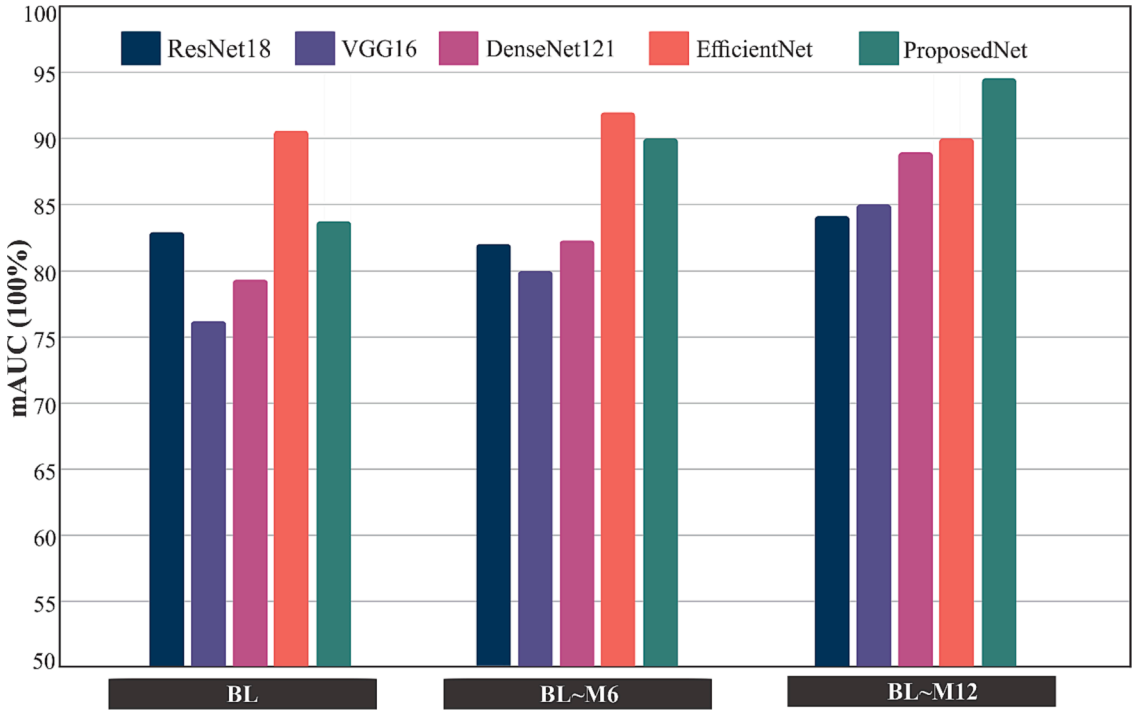

Fig. 4. mAUC comparison of deep models using dynamic image in BL, BL ~ M06, and BL ~ M12.

progression because cognitive decline is one of the major symptoms [38]. The selected features are noteworthy because they are well-known medically and have been extensively studied in scientific research. The proposed multimodal data fusion approach can effectively integrate complementary information from different modalities, thereby enabling a complex modality to convey a more comprehensive and refined depiction of information that surpasses individual inputs. To evaluate the proposed framework using multimodal data, we computed the same set of evaluation metrics used in Experiment 1. This provided insights into the improved accuracy that occurs when analyzing multiple data modalities in disease diagnostic processes. In addition, we examined the effect of multimodal data, including longitudinal MRI, on network stability by comparing the AUC of each model.

4.3.1. Evaluation results using 2D dynamic images + CS

A detailed analysis of the effect of utilizing multiple training data modalities on the detection of AD progression is presented in Table 4. The study involved integrating the CS list with MRI data during the training phase to better understand disease progression. To ensure the reliability and accuracy of the results, the CS scales were acquired at baseline, as summarized in Table 2. CS is a critical indicator of an individual's cognitive ability and is used in conjunction with other diagnostic biomarkers such as genetic biomarkers and behavioral scores. These markers provide valuable insights into disease progression and enable clinicians to develop personalized treatment plans for patients. Moreover, the multimodal data approach used in this study is a promising technique that can significantly improve the accuracy of AD diagnosis and provide reliable progression predictions.

Table 4
Comparative analysis of the proposed network and other deep models using dynamic 2D images + CS.

Backbone network	Time Steps	Accuracy	AUC	F1 Score	Precision	Recall
ResNet50 [29]	BL	0.82 ± 0.01	0.80 ± 0.02	0.83 ± 0.01	0.84 ± 0.02	0.83 ± 0.02
	BL ~ M06	0.84 ± 0.02	0.86 ± 0.01	0.86 ± 0.02	0.87 ± 0.02	0.86 ± 0.02
	BL ~ M12	**0.87 ± 0.01**	**0.88 ± 0.02**	**0.89 ± 0.01**	**0.89 ± 0.01**	**0.88 ± 0.02**
VGG16 [30]	BL	0.85 ± 0.01	0.83 ± 0.02	0.80 ± 0.03	0.88 ± 0.03	0.75 ± 0.04
	BL ~ M06	0.80 ± 0.02	0.81 ± 0.02	0.83 ± 0.01	0.87 ± 0.01	0.78 ± 0.02
	BL ~ M12	**0.86 ± 0.20**	**0.87 ± 0.1**	**0.86 ± 0.20**	**0.86 ± 0.01**	**0.85 ± 0.20**
DenseNet121 [31]	BL	0.87 ± 0.02	0.85 ± 0.2	0.82 ± 0.20	0.86 ± 0.03	0.80 ± 0.03
	BL ~ M06	0.89 ± 0.02	0.88 ± 0.02	0.86 ± 0.02	0.88 ± 0.02	0.86 ± 0.02
	BL ~ M12	**0.91 ± 0.02**	**0.93 ± 0.02**	**0.92 ± 0.02**	**0.95 ± 0.01**	**0.89 ± 0.04**
EfficientNet [32]	BL	0.94 ± 0.02	0.96 ± 0.02	0.93 ± 0.03	0.95 ± 0.02	0.93 ± 0.02
	BL ~ M06	0.94 ± 0.02	0.91 ± 0.02	0.92 ± 0.01	0.95 ± 0.01	0.89 ± 0.01
	BL ~ M12	**0.89 ± 0.03**	0.88 ± 0.02	**0.88 ± 0.02**	**0.89 ± 0.04**	**0.87 ± 0.05**
Proposed Network	BL	0.83 ± 0.02	0.84 ± 0.01	0.82 ± 0.01	0.87 ± 0.02	0.79 ± 0.02
	BL ~ M06	0.91 ± 0.02	0.91 ± 0.02	0.92 ± 0.01	0.91 ± 0.02	0.95 ± 0.01
	BL ~ M12	**0.95 ± 0.01**	**0.96 ± 0.01**	**0.96 ± 0.02**	**0.97 ± 0.02**	**0.96 ± 0.01**

Bold text indicates the best results.

Table 4 presents the scores obtained for multiple evaluation metrics of each comparative model. In the BL, EfficientNet outperformed all other comparative models with an average performance of accuracy = 0.94 ± 0.02, AUC = 0.96 ± 0.02, F1 score = 0.93 ± 0.03, precision = 0.95 ± 0.02, and recall = 0.93 ± 0.02. DenseNet and VGG16 achieved the second-highest accuracy. However, at the BL timestep, our proposed network did not perform well and reported a lower accuracy than that of EfficientNet, DenseNet, and VGG16. This can be attributed to the fact that the proposed framework was designed to model temporal dependencies from longitudinal data, which can not be captured well in a single training data time step. Furthermore, an important observation is that the results shown in BL were highly stable, exhibiting minimal variance ($\leq$2), which highlights the stability of the network with the training dataset.

After training each network with two time steps of longitudinal data collected from BL + M06, we observed a significant improvement in the overall performance of the proposed model. Our model achieved an average accuracy of 0.91 ± 0.02, AUC of 0.90 ± 0.02, F1 score of 0.92 ± 0.01, Precision of 0.91 ± 0.02, and recall of 0.95 ± 0.01, exceeding the 90% threshold for each evaluation metric. This significant improvement in performance demonstrates the importance of using multi-modal data in the disease identification process. EfficientNet showed no significant improvement in its overall performance compared with that of the other models. The results obtained with EfficientNet in BL ~ M06 were either similar to previous time steps (i.e., in BL where Accuracy = 0.94 ± 0.02 and precision = 0.95 ± 0.01) or degraded in BL ~ M06 (i.e., AUC = 0.91 ± 0.02, F1 score = 0.92 ± 0.01 and recall = 0.89 ± 0.01), indicating unstable behavior of this model. Alternatively, DenseNet121 reported the best performance in BL ~ M06 compared to that of ResNet50 and VGG16, achieving an accuracy of 0.89 ± 0.02, AUC of 0.88 ± 0.02, F1 score of 0.86 ± 0.02, Precision of 0.88 ± 0.02 and recall of 0.86 ± 0.02. Performance improvements in BL ~ M06 indicate that the use of longitudinal data from multiple time steps can improve the accuracy of disease identification models.

In the BL~M12, our proposed network once again outperformed other comparative models, demonstrating a gradual improvement in performance, surpassing a $\geq$95 % value for each metric. Among DenseNet121, VGG16, and ResNet50, DenseNet121 achieved the best performance, passing the 90% threshold for every metric except recall, which was 89%. However, the performance of the EfficientNet model continued to degrade, reaching the worst accuracy when three-time steps of the training data were combined with the CS. Although this model performed well in experiments 1 and 2 in BL, it was unable to maintain stable performance when presented with longitudinal data. This may be because the model cannot adequately differentiate the anatomical differences in brain tissues occurring at different time steps and treats the data as a single time step. Additionally, the model showed a very high variance with precision and recall ranging between 4 and 5.

4.3.2. Model comparison using dynamic 2D images + CS

Fig. 5 presents the results of our study on the effectiveness of different comparative models using multimodal longitudinal training data (MRI + CS) and the mAUC as a comparison metric. This figure provides a clear overview of the performance of the models and the improvements obtained using multimodal data in the process of disease diagnosis.

Our findings demonstrate that EfficientNet outperforms other comparative models, with an outstanding mAUC score of 96% in the BL time step, which is a significant improvement of 5% compared with that

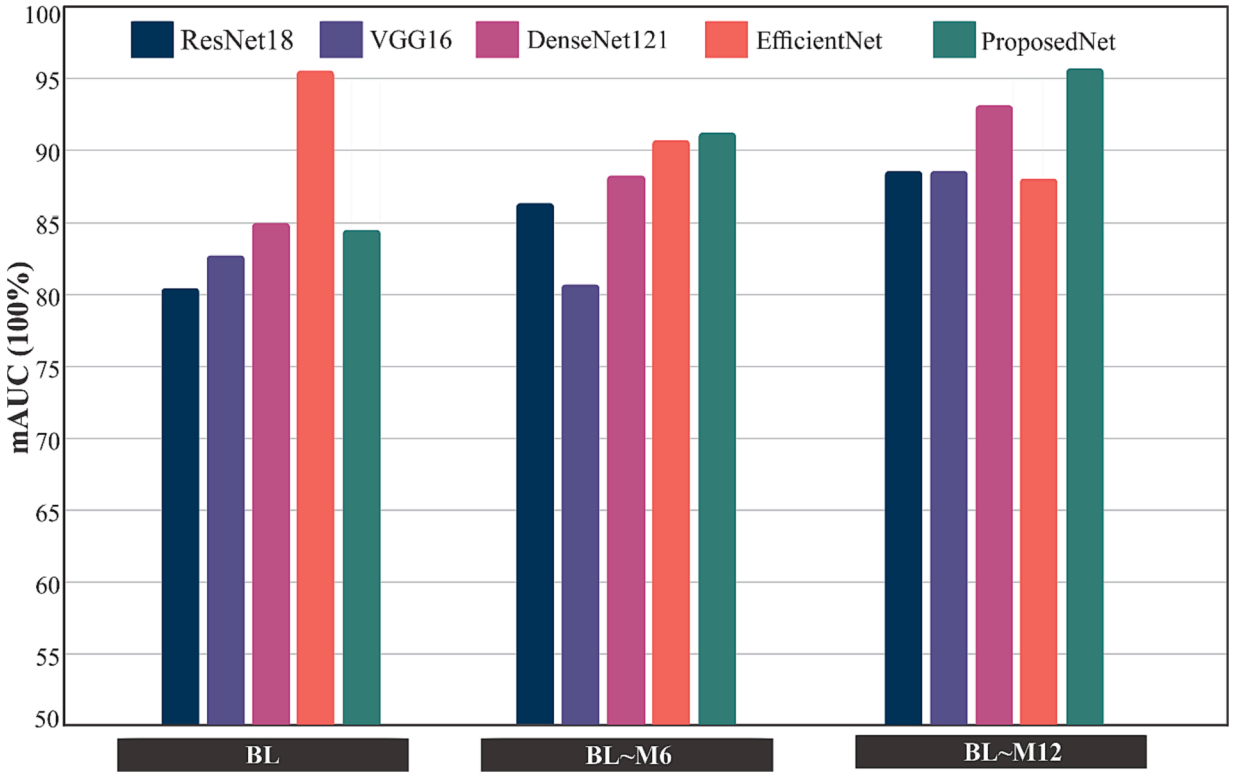

Fig. 5. Performance of various deep models on multimodal medical data (dynamic 2D image + CS).

achieved using a single modality. VGG16 and DenseNet reported improvements of 7% and 5%, respectively, indicating that the use of multimodal data can significantly affect the disease identification process. The proposed network did not show any significant improvement in the mAUC score at the BL time step; however, there was a 2% improvement in the precision metric. After training the models with two-time steps of multimodal longitudinal data, most comparative models demonstrated a significant improvement in the mAUC score compared with that achieved using two timesteps with a single modality. For BL~M6, ResNet reported a 4% improvement from 0.82 ± 0.05 to 0.86 ± 0.01, DenseNet reported a 6% improvement from 0.82 ± 0.02 to 0.88 ± 0.01, and the proposed network reported a 1% improvement, increasing from 0.90 ± 0.02 to 0.91 ± 0.02. VGG16 did not result in any significant improvement in the mAUC score; however, the recall score improved. In BL~M12, all comparative models reported additional improvement in the mAUC score compared to that achieved using single-modality data in BL~M12. The proposed network and DenseNet, VGG16, and ResNet models reported improvements of 2%, 4%, 2%, and 3%, respectively, in the mAUC scores using MRI + CS longitudinal training data.

In conclusion, most of the comparative models in this study achieved significant improvements in overall performance using multimodal longitudinal data, highlighting the importance of incorporating multimodal data for disease identification. In particular, the proposed model exhibited a significant improvement in performance when trained with multimodal longitudinal data, outperforming the other comparative models considered in this study. EfficientNet performed well at baseline but did not maintain a stable performance when presented with longitudinal data, indicating the importance of designing models that can capture temporal dependencies in the data.

We tested our results using both early and late fusion of multimodalities to evaluate the proposed model. In early fusion, the CS features are fused with a feature vector extracted from the CNN model before being passed to the LSTM. In late fusion, the output feature vector from the LSTM model was concatenated with the CS features and fed into a dense layer to distinguish between patients with CN and those with AD. Our observation from early fusion was that the proposed LSTM model considers CS features as part of deep features, as we did not observe any significant effect of these features on the overall accuracy. However, for late fusion, the proposed model exhibited a significant improvement in the detection accuracy of AD progression. We also observed that owing to the lightweight structural design of the proposed network and the carefully chosen number of kernels in each convolution layer, the proposed model identified infected brain regions very well without compromising on the overfitting problem. We also noticed that the proposed model showed an improvement in accuracy and stability, as it saw an increased number of longitudinal time steps during the training process.

5. Explainable deep models

DL-based approaches have become increasingly popular in recent years, owing to their ability to automatically learn features and generalize across a wide range of applications. However, the complexity of such algorithms and, the huge amount of data on which they are trained , make it difficult to understand the underlying information in brain scans, which leads to specific outputs. Consequently, the decision-making process of DL models is often considered a black box. Although several efforts have been made to develop XAI methods in healthcare to interpret machine predictions, an effective method specifically designed to represent visual changes in brain atrophy observed during AD progression using long-term imaging data such as longitudinal MRI is still lacking. Therefore, we propose a comprehensive solution comprising complementary approaches that offer clear visual representations of the reasoning process of the model. Among the proposed approaches, one technique generates activation maps at the voxel level of dynamic 2D images for each time step of the longitudinal data, showing which areas of the brain that are most active during a particular classification in a heat map-like visualization. The second technique constructs a 3D model of brain tissues in a longitudinal manner,

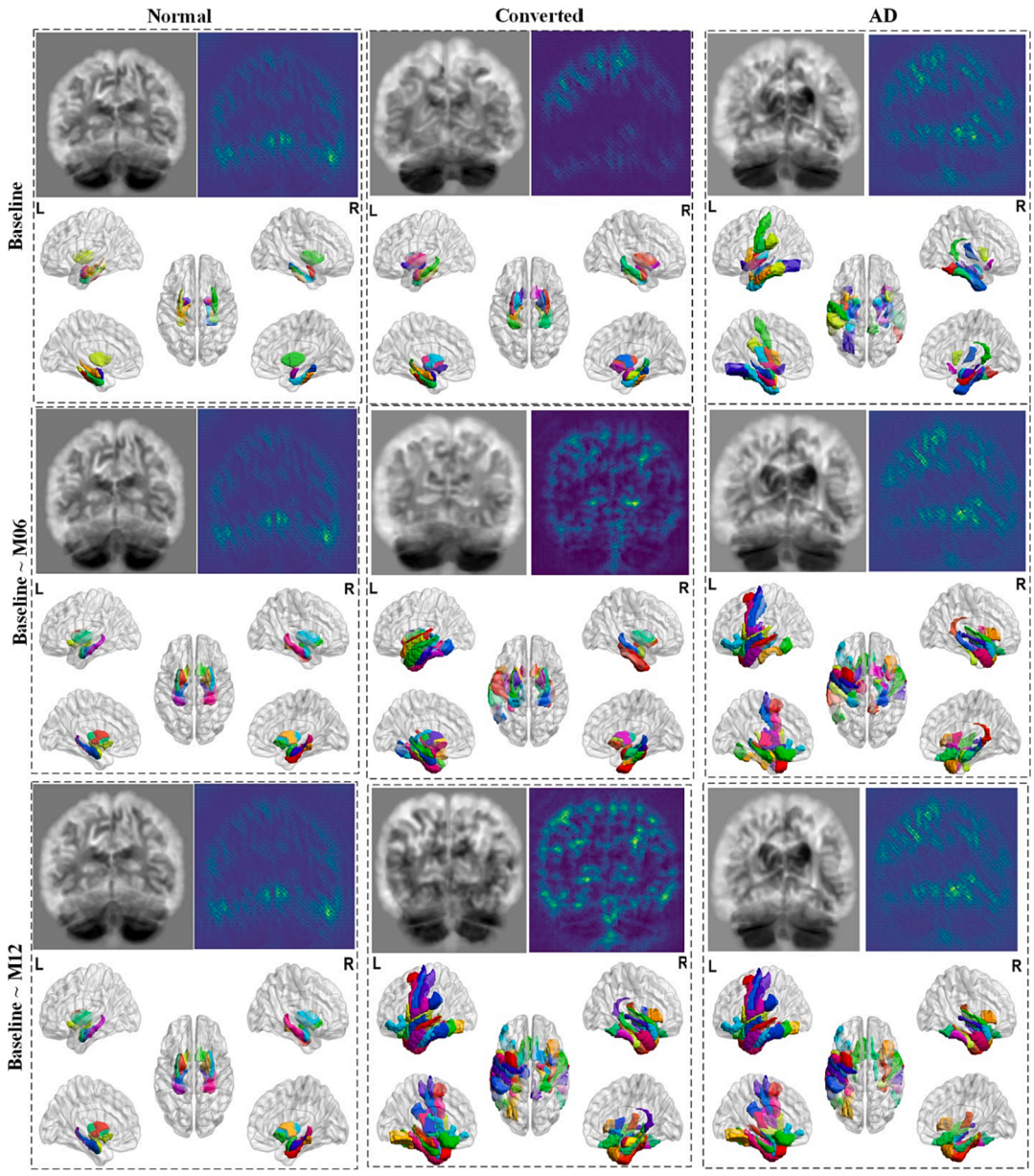

Fig. 6. Proposed visual interpretation of the time series (X-axis represents CN, CN converted to AD and AD patients. Y-axis specifies the patient's health status at different time steps (BL, M06, and at M12).

representing the activated brain regions for the CN, which progresses to AD and AD patients. By combining these techniques, our approach presents a distinctive and comprehensive method for comprehending the decision-making processes of DL models for detecting AD progression. This, in turn, has the potential to aid the development of more efficient diagnostic tools and treatment plans for patients with AD.

5.1. 3D brain reconstruction and voxel activation

CT and MRI yield thin volume slices that are not exactly 2D, making 3D visualization relatively straightforward. Reconstruction and rendering of 3D volumes and surfaces is the most common visualization technique for these neuroimaging modalities [57]. However, the slicing approach does not allow 3D visualization. Two techniques are commonly used to visualize such data: surface and volume rendering. When rendering surfaces, polygonal surfaces are generated from datasets and rendered. Unlike conventional geometry-based visualization techniques, volume rendering utilizes a color mapping system to render elements directly on an image plane, omitting the need for primitive shapes. When it comes to depicting surface structures or organs, surface rendering is often used, while volume rendering is a more versatile method for visualizing internal structures in volume data. Surface rendering is a common method of displaying 3D imaging data obtained via sectional scanning. This technique can be performed either manually or automatically. Manual segmentation involves delineating specific brain structures by manually selecting voxels in a template, whereas automatic segmentation leverages specialized software, such as FreeSurfer [44], to conduct the entire preprocessing workflow and produce detailed anatomical maps. Although manual segmentation offers high precision, it requires substantial time and effort from operators. Automatic segmentation techniques requiring limited user input do not always perform optimally in complex systems. However, these limitations should be carefully considered before making appropriate decisions. Therefore, this study used manual mapping of the brain segments, as depicted in Fig. 6.

Once the 2D activation maps are obtained, the explainer rendered a

3D surface based on the 2D activation maps in several additional processing steps. First, surface acquisition was initiated, which involved gathering brain surface data. This information comprised four categories: vertex ID, vertex position, triangle ID, and vertex index for each triangle. This information is stored in a text file in ASCII format. Using FreeSurfer, the brain surface was extracted by transforming the vertex coordinates into the Montreal Neurological Institute (MNI) template space. In the next stage, brain volume data were converted into the NIFTI format using volume mapping procedures. These steps included processing different types of volume data, ranging from T-maps and Z-maps to atlases. In our investigation of neuroanatomical research and structure–function relationships, we employed the Brainnetome Atlas [58], which consists of 246 areas located in both hemispheres. These areas were parcellated based on the atlas-connectivity and functionally characterized using meta-data labels obtained from the BrainMap database [59]. The labels included the behavioral domain and paradigm class, which were identified using forward and reverse inference techniques. To identify the brain regions that contributed the most to the classification process, the intersection points for each brain were first calculated. Subsequently, volume mapping was performed, which converted the vertex coordinates of the brain surface into image file voxels using various techniques that assigned the vertices to their corresponding values. To superimpose the corresponding voxels between the 3D surface and the image file, an ROI was created, and box-smoothing methods were employed for the ROI volume and brain surface. The surface was rendered using the MATLAB toolbox BrainNetViewer [60]. Fig. 6 displays the activated brain regions for the different classes in a longitudinal manner.

5.2. Time-dependent attention maps of activated brain voxels

To visualize the significant features that contribute to determining the final output class, we leveraged the MedCam Python library [61], which specializes in visualizing the attention maps of 2D and 3D deep models. We provided library with test data containing the recommended labels, which enabled the generation of attention maps. The attention maps were then integrated linearly and normalized to enhance their interpretability. Next, we superimposed these attention maps onto the corresponding dynamic 2D images at each time step. This allowed us to identify the regions that played crucial roles in the decision-making process. By analyzing the attention maps, we gained insights into how the DL model arrived at its findings. Fig. 6 shows 2D slices that incorporate the voxel details obtained from the MedCam library. These slices provide a detailed view of the identified regions. Notably, the distinguishing features between CN individuals, those with AD, and individuals transitioning from CN to AD within a 3-year period are highlighted. Furthermore, to provide a comprehensive representation of AD progression, we included a 3D-rendered brain surface. This visualization effectively demonstrates the affected brain regions that are affected by AD progression. By employing the MedCam library and examining the visual results, we gained a deeper understanding of the decision-making process of the DL model and the specific features that differentiate individuals with CN, AD, and those transitioning between these states.

Activated brain regions of cognitively normal people: The findings of this study revealed that certain brain regions have a higher potential to distinguish between CN individuals and those with AD. Specifically, the rostral Hippocampus [62], medial Amygdala [63], globus Pallidus [63], lateral Amygdala [63], area 28/34 (EC, Entorhinal Cortex) [64] and caudal area 35/36 [62] (also known as the Parahippocampal Gyrus), were found to be the most effective regions for discriminating between the two groups. These regions exhibited consistent patterns across all time steps (BL, M06, and M12), suggesting that they could serve as stable biomarkers for the early diagnosis and monitoring of AD. Moreover, the use of such biomarkers has potential implications in the development of effective treatments and interventions to manage and slow disease progression. By utilizing these specific brain regions, clinicians can better assess disease severity and design treatment plans tailored to the needs of individual patients. Additionally, the findings of this study can be used to develop new diagnostic tools that are more accurate and effective than the current methods.

Infected brain tissues in CN progressed to AD case: The second column of Fig. 6 portrays the disease progression in converted patients. In contrast to individuals with long-term AD, coverted patients undergo a swift transition from a normal state to AD, in contrast to a gradual decline in cognitive function. Disease severity was indicated by a rapid increase in both the number and volume of affected brain regions over time. To identify patients who initially had a CN status but later progressed to AD, the network examined the same regions used to identify patients with a CN status at time steps 1, 2, and 3, including the Hippocampus [62], Amygdala (medial and lateral) [65] and Parahippocampal regions [66]. However, as cognitive impairment worsened in time step 2, the network detected additional affected regions such as medial Amygdala [65], lateral Amygdala [65], dorsolateral Putamen [65], caudal Hippocampus [62], rostroventral area 20 (the Fusiform Gyrus) [64], Globus Pallidus [63], area 28/34 (EC, Entorhinal cortex) [63] and area TL [66]. These regions are associated with memory, learning, and attention, and their involvement in the progression of AD has been previously reported. Atrophy of more than 50% of the brain tissue occurs by time step 3 in patients who have progressed to AD, encompassing all regions identified for time steps 1 and 2. The atrophy of these brain regions causes a significant decline in cognitive function, resulting in the manifestation of AD symptoms. Additional affected regions in time step 3 include rostral area 21 and superior temporal Sulcus [67], that is, middle temporal Gyrus [68], rostral area 22 [68] and lateral area 38 [68], i.e. superior temporal [67]; lateroventral area 37, i.e., Fusiform Gyrus [64]; and caudoventral of area 20 [63], intermediate lateral area 20 [56], and caudolateral of area 20 [56], that is, the inferior temporal Gyrus [52] and caudal Hippocampus [50]. The network detects the affected regions and attributes them to AD conversion in patients. Identifying these regions and their correlation with AD can help develop effective treatments to slow the disease progression. The discovery of these biomarkers has significant clinical implications for early AD diagnosis and monitoring, delayed disease onset, and enhanced quality of life.

Infected brain tissues in AD patient: In Fig. 6, the third column displays the disease progression of a patient who already had AD on the BL scan. The impact of AD on several regions of the brain is evident, including the Hippocampus [62], which is known to play a critical role in memory and spatial navigation and is typically one of the first regions to show atrophy in AD patients. The medial Amygdala [51], caudal Hippocampus [63], and lateral Amygdala [63] are also affected, indicating that AD has a broad impact on the limbic system, which is responsible for emotion, behavior, and memory processing. The dorsolateral Putamen [63], rostroventral area 20 (Fusiform Gyrus) [62], Globus Pallidus [63], area 28/34 [63], and area TL (lateral PPHC, posterior Parahippocampal Gyrus) [69] are also affected, suggesting that AD significantly affects various brain regions involved in memory processing and decision-making. Moreover, the proposed network identified several regions as influential in BL and BL~M06 in AD patients, including rostral area 21 and the superior temporal Sulcus (middle Temporal Gyrus) [70], rostral area 22 and lateral area 38 (superior temporal Gyrus) [71], Lateroventral area 37 (Fusiform Gyrus) [70], caudoventral area 20 and intermediate lateral area 20, and caudolateral area 20 (inferior Temporal Gyrus) [69] and caudal Hippocampus [63]. These findings have significant implications for the diagnosis and monitoring of AD, as they provide valuable information about the progression of the disease and the affected brain regions.

The proposed approach for visualizing the changes in AD-affected brain regions over time is novel and comprehensive. Fig. 6 illustrates that in patients with AD who display disease symptoms from the beginning, the initial scan reveals most of the brain atrophy. The

proposed 3D visual and temporal explanations provide clinicians with an intuitive method for tracking the changes in a patient's condition over time. By utilizing the proposed explainability approach involving 3D and 2D visual temporal details, physicians can intuitively monitor changes in a patient's condition over time. The apparent worsening of the patient's condition at baseline is indicated by numerous affected brain regions compared with the brain of a healthy person. As time progresses to BL ~ M06, the number of affected regions increases, and by BL ~ M12, the situation continues to worsen. The accurate tracking of a patient's condition is facilitated by a proposed explanation that highlights the newly affected regions at each time step and allows physicians to monitor changes effectively. This approach can help physicians to identify the progressive nature of the disease and develop appropriate treatment strategies. By monitoring a patient's condition over time and identifying newly affected regions, physicians can adjust treatment plans and optimize care to meet a patient's specific needs. Overall, the proposed method has significant implications in improving patient outcomes and advancing our understanding of AD.

6. Comparative analysis with existing techniques

In this section, we compare the best proposed competitive framework for detecting AD progression. It is important to note that all comparative studies were evaluated using multiple metrics, including precision, recall, F1-score, AUC, and accuracy. This approach enabled us to assess the strengths of each study from various perspectives. Additionally, many authors have abstained from publishing the training data used in their studies. Consequently, reproducing their published results was not unfeasible for us, and as a result, our comparison was conducted solely based on the published findings of those authors. The results demonstrate that the proposed framework becomes more stable and accurate when multimodal longitudinal training data are used. To demonstrate the effectiveness of the proposed method, it was compared with the most recent methods in the literature. The results are presented in Table 5. We observed that the majority of published studies only utilized a single slice from the entire MRI volume [70], resulting in the loss of crucial information required to ensure the stability of the model in predicting the disease. Moreover, processing complete 3D MRI volumes as longitudinal data incurs substantial computational cost. The suggested approach overcomes these limitations by not only enhancing the detection of AD progression, but also yielding more accurate and consistent results through the integration of multimodal data.

For instance, El-Sappagh et al. [38] conducted a study using longitudinal data consisting of 15 time-steps to detect AD progression. By introducing a novel approach employing a stacked CNN-LSTM model, they aimed to predict multiple variables. Despite the promising results demonstrated by their proposed model, they overlooked the absence of consideration of the temporal gap between the last observed data point and the prediction time steps. Failure to account for this temporal gap could potentially affect the predictive capabilities of the model, thereby emphasizing the necessity for future investigations to address this limitation. Abuhmed et al. [46] introduced hybrid deep models that leveraged multimodal time-series data spanning 18 months. The primary objectives of their research were to detect the progression of AD and predict future cognitive scores. To accomplish these objectives, researchers incorporated a diverse array of features including MRI, PET, and neuropsychological and cognitive scores. Notably, the authors relied on preprocessed features accessible in the ADNI database. They reported average precision of 84.68%, recall 84.80%, F1-score 84.73%, and accuracy 82.63%. El-Sappagh et al. [47] proposed a cost-efficient time-series model that incorporated a range of patient comorbidities, including cognitive scores, treatment history, and demographics, to detect AD progression. This study employed conventional ML-based methods, such as SVM, RF, KNN, decision tree (DT), and logistic regression (LR), to achieve the objective of identifying AD progression. Zhu et al. [72] employed spatiotemporal features extracted from longitudinal MRI to predict the likelihood of patients with MCI converting to AD before the manifestation of clinical symptoms. The authors suggested the use of temporally structured TS-SVMs to identify spatiotemporal features that capture structural transformations in the brain during AD progression. Dong et al. [73] presented DeepAtrophy, a system that utilizes pairs of longitudinal MRI scans to identify AD progression by

Table 5
Comparison of the proposed model and existing literature techniques.

Study	#S	Mod	LT?	#T	Performance (%)					ML Method
					Pre	Rec	F1-S	AUC	Acc	
El-Sappagh et al. [38], (2021)	1536	MRI, PET, CS, ASD, NPD	Yes	15	94.02	98.42	92.56	-	92.62	Stacked CNN Bi-LSTM
Abuhmed et al. [46], (2021)	1371	MRI, PET, CS, N, NP, D	Yes	4	84.68	84.80	84.73	-	82.63	Bi-LSTM
El-Sappagh et al. [47], (2021)	1536	MRI, PET, CS, ASD, NPD	Yes	15	-	99.99	91.36	-	92.21	RF, DT, LR, SVM, KNN, XGBoost, NB, MLP
Zhu et al. [72], (2021)	151	MRI	Yes	5	-	-	-	86.5	85.4	TS-SVM
Dong et al. [73], (2021)	492	MRI	Yes	6	-	-	-	-	88.00	DeepAtrophy
Ghazi et al. [74], (2021)	1757	MRI, PET, CSF	Yes	3	-	-	-	93.40	-	Modified LR
Kang et al. [75], (2021)	798	MRI	No	1	-	-	-	-	90.36	EL-CNN
El-Sappagh et al. [21], (2022)	1371	MRI, CS, D, CSF, NS Markers	Yes	4	94.07	94.07	94.07	-	93.87	2-staged AD progression detection
Helaly et al. [76], (2022)	1500	MRI	No	1	-	-	-	-	94.34	RESU-Net
Sharma et al. [77], (2022)	2400	MRI	No	1	-	-	-	-	95.00	FLS-TWSVM
Atefe et al. [81], (2022)	210	MRI	Yes	3	-	82.00	-	94.00	87.2	EL-CNN
Kong et al. [39], 2022	370	MRI, PET	No	1	-	-	-	-	87.67	AD fusion model
Zhang et al. [1], (2023)	876	MRI	No	1	-	97.83	-	98.34	96.61	MRN-Net
Guan et al. [78], (2023)	360	MRI	No	1	-	69.46	-	75.70	73.54	IADT
Goel et al. [79], 2023	420	MRI, PET	No	1	92.56	95.33	-	-	95.89	RVFL
Li et al. [80], 2023	446	MRI	No	1	-	-	-	-	92.42	GCM-EB2
Zhentao et al [82]., 2023	275	MRI	Yes	2	-	79.97	-	1.53	77.2	VGG-TSwinformer
Eslami et al. [83], 2023	1123	MRI, PET, CSF	Yes	4	-	-	-	-	91.83	ML4VisAD
Proposed Net (2023)	**1692**	**MRI, CS**	**Yes**	**3**	**97.00**	**96.00**	**96.00**	**96.00**	**95.00**	**2D-CNN Bi-LSTM**

#S = Number of subjects, Mod = Modalities; LT?= Is longitudinal time steps available?; #T = Number of longitudinal time steps; CS = Cognitive scores; D = Demographic features; N = Neuropsychiatric features; ASD = Assessment data; NPD = Neuropathological data; NB = Naïve bayes; GBoost = eXtreme Gradient Boosting; EL = Ensemble learning; RESU-Net = ResNet-UNET; FLS-TWSVM = fuzzy hyperplane based least square twin SVM; GCM-EB2 = global attention mechanism-EfficientNetB2; MRN = Multi relation net; IADT = Interpretable autoencoder model with domain transfer learning; RVFL = Random vector functional link; ML4VisAD = Machine Learning for Visualizing AD.

deducing the temporal information between paired scans. The authors utilized a 50-layer 3D Resnet architecture to construct the proposed network and observed considerably higher activation between scan pairs with more significant alterations. The researchers employed T1-weighted MRI scans and cropped patches solely from the hippocampal regions and reported an accuracy of 88%.

Several studies, [38,46,47,72–74,21] have used longitudinal training data across multiple time steps, including MRI, PET, CS and other essential patient comorbidities. These studies provide comprehensive insights into disease analysis. In contrast, several other studies [75–77,1,78–80] relied solely on baseline data, neglecting the time series aspect of the training data. These studies lacked longitudinal training data for disease analysis and did not adequately explain their results. Among the aforementioned studies, only two [47,1] outperformed the proposed framework in terms of the disease identification accuracy. Of these two, the superior performance of a study by El-Sappagh et al. may be owing to the diverse data modalities and longitudinal time steps used in the training data. All these studies provided acceptable accuracy in the disease diagnosis process, but none reported voxel-level tissue damage in the decision-making process. Table 5 presents a comparison of several metrics, namely, the number of subjects, data modalities, availability of longitudinal training data, number of time steps, performance achieved, and approach taken in each study. Notably, our proposed network surpassed the performances achieved in many studies. Given its superior performance and robustness, the proposed system can serve as a baseline model for the development of advanced healthcare systems targeting the onset of AD.

Additionally, we used visual representations of brain images to explain the time series and justify the model's decision to identify patients with AD. The proposed XAI technique was designed to track brain regions over time, providing valuable information for early detection and diagnosis of AD. Various XAI approaches have proven beneficial in numerous medical domains for clarifying deep models' reasoning processes [84–86]. However, many of these XAI techniques do not provide a clear visual representation of brain atrophy observed on MRI images during long-term studies for detecting AD progression. The proposed technique fills this gap by providing an intuitive and comprehensive visualization of changes in brain regions over time, allowing physicians to track disease progression ease and accuracy. In particular, our proposed XAI technique exhibited superior performance compared to the most advanced methods in the literature, making it a promising tool for AD management. However, further development is necessary to ensure its suitability for implementation in clinical settings. We believe that, through continuous research and refinement, our XAI technique has the potential to revolutionize the diagnosis and management of AD, providing patients with more effective and personalized treatment options. Integrating our model into clinical workflows requires solving various technical and ethical problems. Nevertheless, the potential for improving the precision and effectiveness of AD diagnosis and management using DL tools appears promising.

7. Current limitations

Despite the good performance of the ADNI dataset under various conditions, the proposed framework has limitations. The current study utilized a single neuroimaging modality, such as MRI. However, for the detection of AD progression, the incorporation of multiparametric MRI (PET, diffusion tensor imaging, and functional MRI) provides a wealth of disease-related information in the diagnostic process. We were unable to explore other modalities because of their unavailability. Furthermore, we used multimodal data consisting of longitudinal MRI scans and cross-sectional cognitive scores. Unfortunately, the absence of cognitive scores in a longitudinal capacity hindered our ability to explore them at different time points. Additional factors that hindered the improvement of the learning performance of our proposed approach included compatibility issues related to the longitudinal time steps captured in each subject group (i.e., 6 months (BL, M6, M12, etc.)), as well as limited public access to these datasets, specifically the MIRIAD, OASIS or AIBL datasets. Finally, the primary focus of this study was the visual explainability of 2D MRI images and 3D brain surfaces in a longitudinal manner only, as the second modality (cognitive scores) was based solely on the BL timestep.

8. Conclusions

In medicine and healthcare, the use of machine learning and data mining techniques is immensely beneficial for the early detection and diagnosis of numerous diseases. AD is the most severe form of dementia and leads to memory loss and cognitive decline. Many medical diagnostic systems rely primarily on baseline data acquired during the initial visit, disregarding the dynamic nature of clinical information. Conventional DL models operate as black boxes, making it challenging to explain their decision-making processes. Although current DL models demonstrate high precision, their practical implementation is hindered by the volumetric nature of medical data, which necessitates intensive computation and makes it difficult for physicians and regulators to verify the predictive results of a given system. In this study, we used an approximate rank pooling strategy to generate a single 2D image from a whole-brain 3D MRI volume. The resulting 2D dynamic image was a compressed representation of the entire 3D MRI volume. Dynamic images were generated longitudionally and were used to detect AD progression. We also propose an efficient CNN-Bi-LSTM model that outperformes comparative models in detecting AD progression. The effectiveness of each model was evaluated using single-mode data (dynamic 2D image) and multimodal data (dynamic 2D image + CS) data. In addition, we introduce a new technique that renders our model's decisions medically interpretable and acceptable to medical professionals. For this purpose, we chose a guided grad-cam to visually represent influential features by generating heatmaps that highlighted the location of the exact voxels in the damaged brain regions, showing the features that most influenced the final classification of the system for AD progression in this patient. Localization heatmaps showed a large influence of the lateral ventricle, limbic system, subregions and other disease-affected regions of the cortex. These images are consistent with regions that are commonly affected during AD progression. Our future research will examine the effectiveness of alternative modalities, including cognitive scores in a longitudinal context, and PET neuroimaging for identifying AD progression. Furthermore, we explored the impact of integrating various types of time-series data on model performance. While this study introduced a novel explainable 2D approach to MRI slices and 3D brain surfaces within time series-data, medical professionals tend to favor multiple explanations to increase confidence in the model's results. Therefore, we plan to investigate additional XAI techniques in future research that incorporate complementary modalities such as PET and diffusion neuroimaging. In future research, we will explore the interesting idea of a learnable strategy for temporal rank pooling to create a dynamic image.

CRediT authorship contribution statement

Nasir Rahim: Conceptualization, Data curation, Software, Formal analysis, Visualization, Writing – original draft. **Tamer Abuhmed:** Conceptualization, Methodology, Supervision, Formal analysis, Writing – review & editing, Funding acquisition. **Seyedali Mirjalili:** Conceptualization, Investigation, Data curation, Writing – review & editing, Visualization. **Shaker El-Sappagh:** Methodology, Formal analysis, Validation, Writing – review & editing. **Khan Muhammad:** Supervision, Formal analysis, Validation, Writing – review & editing.

Declaration of Competing Interests

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.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No. 2021R1A2C1011198), (Institute for Information & communications Technology Planning & Evaluation) (IITP) grant funded by the Korea government (MSIT) under the ICT Creative Consilience Program (IITP-2021-2020-0-01821), and AI Platform to Fully Adapt and Reflect Privacy-Policy Changes (No. 2022-0-00688). Data used in preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this article.

References

[1] S. Qasim Abbas, L. Chi, Y.P.P. Chen, Transformed domain convolutional neural network for Alzheimer's disease diagnosis using structural MRI, Pattern Recognit. 133 (2023), 109031, https://doi.org/10.1016/j.patcog.2022.109031.
[2] 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.
[3] B. Lei, et al., Predicting clinical scores for Alzheimer's disease based on joint and deep learning, Expert Syst. Appl. 187 (2022), 115966, https://doi.org/10.1016/j.eswa.2021.115966.
[4] H.K. Bharadwaj, et al., A review on the role of machine learning in enabling IoT based healthcare applications, IEEE Access 9 (2021) 38859–38890, https://doi.org/10.1109/ACCESS.2021.3059858.
[5] 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.
[6] 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.
[7] 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.
[8] D. Gupta, U. Kose, V.H.C. Albuquerque, Editorial: Computational methods for neuroimaging: challenges and future trends, Front. Comput. Neurosci. 17 (2023), https://doi.org/10.3389/fncom.2023.1181169.
[9] M. EL-Geneedy, H.E.D. Moustafa, F. Khalifa, H. Khater, E. AbdElhalim, An MRI-based deep learning approach for accurate detection of Alzheimer's disease, Alex. Eng. J. 63 (2023) 211–221, https://doi.org/10.1016/j.aej.2022.07.062.
[10] T. Hrishikesh Jaware, V. Ramesh Patil, C. Nayak, A. Elmasri, N. Ali, P. Mishra, A novel approach for brain tissue segmentation and classification in infants' MRI images based on seeded region growing, foster corner detection theory, and sparse autoencoder, Alex. Eng. J. 76 (2023) 289–305, https://doi.org/10.1016/j.aej.2023.06.040.
[11] S.K. Teh, I. Rawtaer, A.H. Tan, Predictive self-organizing neural networks for in-home detection of Mild Cognitive Impairment, Expert Syst. Appl. 205 (2022), 117538, https://doi.org/10.1016/j.eswa.2022.117538.
[12] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, P. Shen, An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis, Front. Digit. Heal. 3 (2021) 19, https://doi.org/10.3389/fdgth.2021.637386.
[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.
[14] 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).
[15] 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.
[16] A.V. Savchenko, N.S. Belova, Sequential analysis in Fourier probabilistic neural networks, Expert Syst. Appl. 207 (2022), 117885, https://doi.org/10.1016/j.eswa.2022.117885.
[17] A. Rayan, et al., Utilizing CNN-LSTM techniques for the enhancement of medical systems, Alex. Eng. J. 72 (2023) 323–338, https://doi.org/10.1016/j.aej.2023.04.009.
[18] A.M. Alvi, S. Siuly, H. Wang, K. Wang, F. Whittaker, A deep learning based framework for diagnosis of mild cognitive impairment, Knowledge-Based Syst. 248 (2022), https://doi.org/10.1016/j.knosys.2022.108815.
[19] B. Lei et al. , Longitudinal study of early mild cognitive impairment via similarity-constrained group learning and self-attention based SBi-LSTM☆, 254 (2022) 109466, doi: 10.1016/j.knosys.2022.109466.
[20] G. Lee et al. , Predicting Alzheimer's disease progression using multi-modal deep learning approach, Sci. Rep. 9(1) (2019), doi: 10.1038/s41598-018-37769-z.
[21] S. El-Sappagh, H. Saleh, F. Ali, E. Amer, T. 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.
[22] A. Elhence, V. Kohli, V. Chamola, B. Sikdar, Enabling cost-effective and secure minor medical teleconsultation using artificial intelligence and blockchain, IEEE Internet Things Mag. 5 (1) (2022) 80–84, https://doi.org/10.1109/iotm.001.2100142.
[23] S. Aras, P.J.G. Lisboa, Explainable inflation forecasts by machine learning models, Expert Syst. Appl. 207 (2022), 117982, https://doi.org/10.1016/J.ESWA.2022.117982.
[24] P.R. Magesh, R.D. Myloth, R.J. Tom, An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery, Comput. Biol. Med. 126 (2020), 104041, https://doi.org/10.1016/j.compbiomed.2020.104041.
[25] A. Holzinger, C. Biemann, C.S. Pattichis, 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.
[26] A. Barredo Arrieta et al. , Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115, doi: 10.1016/j.inffus.2019.12.012.
[27] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, N. Díaz-Rodríguez, Explainable artificial intelligence (XAI) on TimeSeries data: a survey, 2021, [Online], Available: http://arxiv.org/abs/2104.00950.
[28] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, Action recognition with dynamic image networks, IEEE Trans. Pattern Anal. Mach. Intell. 40 (12) (2018) 2799–2813, https://doi.org/10.1109/TPAMI.2017.2769085.
[29] F. Wang et al. , Residual attention network for image classification, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, 2017-Janua, no. 1, 2017, pp. 6450–6458, doi: 10.1109/CVPR.2017.683.
[30] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015. Accessed: Oct. 06, 2021. [Online], Available: http://www.robots.ox.ac.uk/.
[31] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
[32] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: 36th International Conference on Machine Learning, ICML 2019, 2019, pp. 10691–10700.
[33] 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.
[34] D.L. Beekly, et al., The National Alzheimer's Coordinating Center (NACC) database: the uniform data set, Alzheimer Dis. Assoc. Disord. 21 (3) (2007) 249–258, https://doi.org/10.1097/WAD.0b013e318142774e.
[35] D.S. Marcus, A.F. Fotenos, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults, J. Cogn. Neurosci. 22 (12) (2010) 2677–2684, https://doi.org/10.1162/jocn.2009.21407.
[36] I.B. Malone, et al., MIRIAD-public release of a multiple time point Alzheimer's MR imaging dataset, Neuroimage 70 (2013) 33–36, https://doi.org/10.1016/j.neuroimage.2012.12.044.
[37] N. Rahim, S. El-Sappagh, S. Ali, K. Muhammad, J. Del Ser, T. Abuhmed, Prediction of Alzheimer's progression based on multimodal Deep-Learning-based fusion and visual Explainability of time-series data, Inf. Fusion 92 (2023) 363–388, https://doi.org/10.1016/j.inffus.2022.11.028.
[38] 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, 2021, pp. 482–492, https://doi.org/10.1007/978-3-030-55190-2_36.
[39] Z. Kong, M. Zhang, W. Zhu, Y. Yi, T. Wang, B. Zhang, Multi-modal data Alzheimer's disease detection based on 3D convolution, Biomed. Signal Process. Control 75 (2022), https://doi.org/10.1016/j.bspc.2022.103565.
[40] J. Wang, A. Cherian, F. Porikli, Ordered pooling of optical flow sequences for action recognition, in: Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, 2017, pp. 168–176, https://doi.org/10.1109/WACV.2017.26.
[41] X. Xing, et al., Dynamic image for 3D MRI image Alzheimer's disease classification, in: A. Bartoli, A. Fusiello (Eds.), Computer Vision – ECCV 2020 Workshops, Springer International Publishing, Cham, 2020, pp. 355–364.
[42] N.J. Ashton, et al., Cerebrospinal fluid p-tau231 as an early indicator of emerging pathology in Alzheimer's disease, EBioMedicine 76 (2022), https://doi.org/10.1016/J.EBIOM.2022.103836.
[43] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for simplicity: the all convolutional net, in: 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings, 2015.

[44] B. Fischl, FreeSurfer, Neuroimage 62 (2) (2012) 774–781, https://doi.org/10.1016/j.neuroimage.2012.01.021.
[45] G. Martí-Juan, G. Sanroma-Guell, G. Piella, A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease, in: Computer Methods and Programs in Biomedicine, Vol. 189, Elsevier, Jun. 01, 2020, pp. 105348, doi: 10.1016/j.cmpb.2020.105348.
[46] T. Abuhmed, S. El-Sappagh, J.M. Alonso, Robust hybrid deep learning models for Alzheimer's progression detection, Knowledge-Based Syst. 213 (2021), 106688, https://doi.org/10.1016/j.knosys.2020.106688.
[47] 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.
[48] W. Wang, J. Shen, Deep visual attention prediction, IEEE Trans. Image Process. 27 (5) (2018) 2368–2378, https://doi.org/10.1109/TIP.2017.2787612.
[49] A.M. Hafiz, S.A. Parah, R.U.A. Bhat, Attention mechanisms and deep learning for machine vision: a survey of the state of the art, 2021, [Online], Available: http://arxiv.org/abs/2106.07550.
[50] S.M. Muddamsetty, M.N.S. Jahromi, A.E. Ciontos, L.M. Fenoy, T.B. Moeslund, Visual explanation of black-box model: Similarity Difference and Uniqueness (SIDU) method, Pattern Recogn. 127 (2022), 108604, https://doi.org/10.1016/j.patcog.2022.108604.
[51] S. Ali, et al., Explainable Artificial Intelligence (XAI): what we know and what is left to attain Trustworthy Artificial Intelligence, Inf. Fusion (2023), https://doi.org/10.1016/j.inffus.2023.101805.
[52] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy Artificial Intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation, Inf. Fusion 99 (June) (2023), 101896, https://doi.org/10.1016/j.inffus.2023.101896.
[53] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.
[54] 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.
[55] C. Barata, M. Ruela, M. Francisco, T. Mendonca, J.S. Marques, Two systems for the detection of melanomas in dermoscopy images using texture and color features, IEEE Syst. J. 8 (3) (2014) 965–979, https://doi.org/10.1109/JSYST.2013.2271540.
[56] W.V. Stoecker, et al., Detection of granularity in dermoscopy images of malignant melanoma using color and texture features, Comput. Med. Imaging Graph. 35 (2) (2011) 144–147, https://doi.org/10.1016/J.COMPMEDIMAG.2010.09.005.
[57] A.G. Schreyer, S.K. Warfield, Surface rendering, in: 3D Image Processing, Springer, 2002, pp. 31–34.
[58] 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.
[59] 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.
[60] 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.
[61] K. Gotkowski, C. Gonzalez, A. Bucher, A. Mukhopadhyay, M3d-CAM: a PyTorch library to generate 3D attention maps for medical deep learning, in: Informatik aktuell, Springer Vieweg, Wiesbaden, 2021, pp. 217–222, https://doi.org/10.1007/978-3-658-33198-6_52.
[62] 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.
[63] 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.
[64] 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, in: Annals of the New York Academy of Sciences, New York Academy of Sciences, 2000, pp. 254–274, doi: 10.1111/j.1749-6632.2000.tb06731.x.
[65] 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.
[66] 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.
[67] G. Karas, et al., Amnestic mild cognitive impairment: Structural MR imaging findings predictive of conversion to Alzheimer disease, AJNR Am. J. Neuroradiol. (2008) 944–949, https://doi.org/10.3174/ajnr.A0949.
[68] A. Ulloa, S. Plis, E. Erhardt, V. Calhoun, Synthetic structural magnetic resonance image generator improves deep learning prediction of schizophrenia, in: IEEE Int. Work. Mach. Learn. Signal Process. MLSP, 2015-Novem, 2015, pp. 1–6, https://doi.org/10.1109/MLSP.2015.7324379.
[69] 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.
[70] S. Risacher, A. Saykin, J. Wes, L. Shen, H. Firpi, B. McDonald, Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort, Curr. Alzheimer Res. 6 (4) (2009) 347–361, https://doi.org/10.2174/156720509788929273.
[71] 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. - Neuroimaging 194 (1) (2011) 7–13, https://doi.org/10.1016/j.pscychresns.2011.06.014.
[72] 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.
[73] 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, doi: 10.1016/j.neuroimage.2021.118514.
[74] M. Mehdipour Ghazi et al. , Robust parametric modeling of Alzheimer's disease progression, Neuroimage 225(June 2020) (2021) 117460, doi: 10.1016/j.neuroimage.2020.117460.
[75] W. Kang, L. Lin, B. Zhang, X. Shen, S. Wu, Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis, Comput. Biol. Med. 136 (2021), 104678, https://doi.org/10.1016/j.compbiomed.2021.104678.
[76] 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.
[77] 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.
[78] H. Guan, L. Yue, P.-T. Yap, S. Xiao, A. Bozoki, M. Liu, Attention-guided autoencoder for automated progression prediction of subjective cognitive decline with structural MRI, IEEE J. Biomed. Heal. Informatics (2023) 1–10, https://doi.org/10.1109/jbhi.2023.3257081.
[79] T. Goel, R. Sharma, M. Tanveer, P.N. Suganthan, K. Maji, R. Pilli, Multimodal neuroimaging based Alzheimer's disease diagnosis using evolutionary RVFL classifier, IEEE J. Biomed. Heal. Informatics (2023) 1–9, https://doi.org/10.1109/JBHI.2023.3242354.
[80] H. Li, et al., Attention-based and micro designed EfficientNetB2 for diagnosis of Alzheimer's disease, Biomed. Signal Process. Control 82 (2023), 104571, https://doi.org/10.1016/j.bspc.2023.104571.
[81] A. Aghaei, M. Ebrahimi Moghaddam, H. Malek, Interpretable ensemble deep learning model for early detection of Alzheimer's disease using local interpretable model-agnostic explanations, Int. J. Imaging Syst. Technol. 32 (6) (2022) 1889–1902, https://doi.org/10.1002/ima.22762.
[82] Z. Hu, Z. Wang, Y. Jin, W. Hou, VGG-TSwinformer: transformer-based deep learning model for early Alzheimer's disease prediction, Comput. Methods Programs Biomed. 229 (2023), https://doi.org/10.1016/j.cmpb.2022.107291.
[83] M. Eslami, S. Tabarestani, M. Adjouadi, A unique color-coded visualization system with multimodal information fusion and deep learning in a longitudinal study of Alzheimer's disease, Artif. Intell. Med. 140 (March) (2023), 102543, https://doi.org/10.1016/j.artmed.2023.102543.
[84] K. Oh, Y.C. Chung, K.W. Kim, W.S. Kim, I.S. Oh, Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, Sci. Rep. 9 (1) (2019), https://doi.org/10.1038/s41598-019-54548-6.
[85] 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.
[86] 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, pp. 1036–1041, Oct. 2020, doi: 10.1109/BIBE50027.2020.00175.