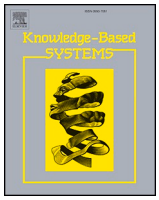
Knowledge-Based Systems 334 (2026) 115104

Contents lists available at ScienceDirect

Knowledge-Based Systems

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



Multi-plane multi-slice longitudinal MRI for deep ensemble progression detection based on enhanced residual multi-head self-attention

Nasir Rahim [a], Shaker El-Sappagh [b,e,f], Mustaqeem Khan [c], Maria Bashir [d], Younhyun Jung [a,*], Tamer Abuhmed [b,*]

[a] School of Computing, Gachon University, Seongnam 13120, Republic of Korea
[b] Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, South Korea
[c] College of Information Technology-United Arab Emirates University (UAEU), Al Ain, UAE
[d] Department of Electrical & Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[e] Faculty of Computer Science and Engineering, Galala University, Suez 435611, Egypt
[f] Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, Banha 13518, Egypt

ARTICLE INFO

Keywords:
Alzheimer's disease
Longitudinal MRI
Multi-plane multi-slice volumes
DL & ensemble models

ABSTRACT

Early and accurate detection of Alzheimer's disease (AD) progression remains a critical challenge in neuroimaging, as existing methods often rely on single-plane or cross-sectional MRI data, neglecting the rich spatio-temporal dynamics captured in multi-plane longitudinal imaging. To address this gap, we propose a novel deep ensemble framework that leverages multi-plane volumetric representations of 3D longitudinal MRI data for enhanced AD progression detection. Our approach introduces a unified 3D volumetric representation of longitudinal MRI by integrating spatially aligned slices from axial, coronal, and sagittal planes across four longitudinal time points (Baseline, M06, M12, and M18), preserving both spatial and temporal context. Our proposed framework employs an optimized heterogeneous deep ensemble setup of 3D-CNN models (i.e., 3D-EfficientNet, 3D-DenseNet, and 3D-ResNet) to extract complementary spatio-temporal features from each anatomical plane, followed by a BiLSTM network with an Enhanced Residual Multi-Head Self-Attention (ERMHA) mechanism to model long-range dependencies and emphasize discriminative spatiotemporal patterns. Comprehensive experiments on the ADNI dataset demonstrate that our proposed framework achieves state-of-the-art performance, with a mean accuracy of 93.73%, sensitivity of 91.72%, specificity of 90.36%, and an AUC of 91.58%, significantly outperforming single-plane based models (best mAUC: 68.24%) and homogeneous ensemble approaches (mAUC: 82.75%). External validation on the NACC cohort further confirms generalizability, with performance metrics improving consistently as data from more longitudinal time steps are incorporated (mAUC: 86.37% at M18). Furthermore, explainability analysis using gradient-weighted attention maps reveals that model predictions are driven by neuroanatomically plausible patterns, with attention focused on hippocampal and entorhinal regions in early progression and extending to temporo-parietal cortices in advanced stages in AD, aligning with established neuropathological trajectories. The proposed framework advances intelligent decision support systems in clinical neuroimaging by combining multi-plane feature fusion, temporal modeling, and ensemble learning, offering a robust and generalized solution for early AD progression detection. Its modular design and computational efficiency make it suitable for integration into knowledge-based diagnostic systems. The dataset and code used in this study are available to the research community at the following link: https://github.com/InfoLab-SKKU/mpms-mri-progression.git

1. Introduction

Alzheimer's disease (AD) is a highly challenging neurodegenerative disease characterized by progressive cognitive decline and irreversible memory loss, ultimately resulting in full dementia [1]. To date, no treatment has demonstrated efficacy in reversing or preventing neurodegeneration, underscoring the critical importance of early diagnosis in delaying cognitive decline [2]. With its complex etiology, cognitive

* Corresponding authors.
E-mail addresses: younhyun.jung@gachon.ac.kr (Y. Jung), tamer@skku.edu (T. Abuhmed).

https://doi.org/10.1016/j.knosys.2025.115104
Received 4 May 2025; Received in revised form 19 November 2025; Accepted 6 December 2025
Available online 8 December 2025

impairment, and memory deterioration, AD has emerged as a significant global health concern [3]. The number of diagnosed AD patients worldwide exceeds 40 million, and this figure is projected to reach 132 million within the next three decades [4]. Mild Cognitive Impairment (MCI) represents a transitional stage between normal aging and the onset of AD, where a diagnosis of dementia based solely on cognitive impairment is insufficient [5]. Notably, approximately 10∼12 % of individuals with normal cognitive function will progress to MCI annually, and within three years, around 44 % of MCI patients will transition to AD [6]. AD follows a long and slow process, ending in a dementia diagnosis at a late stage, which increases frustration for the patients, families, and healthcare professionals, negatively affecting quality of life and causing major financial costs [7]. Even a minor delay in implementing AD treatment strategies can contribute to a decrease in dementia prevalence and alleviate the socioeconomic impact by mitigating self-management challenges faced by patients [8]. Considering that it takes approximately 20∼30 years for AD to manifest as dementia, preventive interventions targeting the earlier stages of the disease offer an opportunity to identify individuals who can benefit the most [9].

Neuroimaging plays a crucial role in AD diagnosis, with structural magnetic resonance imaging (sMRI) being widely used to classify AD patients and identify distinct disease stages based on corresponding pathological changes [10]. MRI represents a significant milestone in neuroimaging, offering exceptional soft tissue contrast and the ability to depict dynamic physiological changes through 3D tomographic visualization. It is a non-invasive and highly effective tool for analyzing and diagnosing structural brain alterations, making it invaluable for AD diagnosis. In recent years, deep learning (DL) computational models, particularly convolutional neural networks (CNNs), have emerged as influential architectures in AD classification compared to other existing approaches [11,12]. Previous CNN-based AD diagnosis methods have primarily relied on features extracted from a single-plane, such as the coronal plane [13]. However, leveraging multi-plane feature fusion may enhance diagnostic accuracy, as cortical atrophies exhibit distinct characteristics across different imaging planes. CNN models employed in neuroimaging studies can be categorized into 2D slice-level, 3D subject-level, 3D patch-level, and region-of-interest imaging [14]. Slice-based methods entail selecting standardized slices from original MRI images and training models on these selected slices. This approach simplifies the model training process by employing 2D image classification models directly for transfer learning purposes [15]. Researchers have proposed numerous slice-based methods for AD diagnosis. For instance, Wang et al. [16] introduced a single-slice approach that achieved higher accuracy than existing systems by incorporating an improved biogeography-based optimization method, a multilayer perceptron, and wavelet entropy. In another study, Leon et al. [17] investigated the application of supervised switching autoencoders (SSA) using a single 2D slice of sMRI for AD classification. They utilized local patch-based methods to identify disease regions and fused neurodegeneration patterns with disease information. Furthermore, ensemble learning techniques have been employed to integrate multiple CNNs or MRI slices, improving stability and classification accuracy. In one such approach, Kang et al. [18] proposed a CNN system that combined eleven 2D slice-level models from three CNNs using ensemble learning, achieving an 81.3 % classification accuracy in distinguishing between cognitively normal individuals and AD patients. Moreover, Shmulev et al. [19] developed a 3D-ResNet model for classifying MCI subjects with a 62 % accuracy in differentiating between MCI converters and non-converters. Additionally, Aderghal et al. [20] incorporated multi-plane data from anatomical planes (axial, sagittal, and coronal) obtained from diffusion tensor imaging and MRI as inputs. By employing a LeNet-like CNN pretrained on the MNIST dataset, they achieved classification accuracies of 86.83 %, 71.75 %, and 69.85 % for AD vs. CN, AD vs. MCI, and MCI vs. CN, respectively.

Longitudinal analysis utilizing MRI has demonstrated exceptional results in the early detection of neurodegenerative diseases, offering the potential to monitor disease progression in AD patients [21]. Researchers have explored various techniques for longitudinal MRI prediction. For instance, Song et al. [22] proposed a GAN-based method that leverages complete information from longitudinal MRI to predict missing data in AD-affected brain scans. Similarly, Pathan et al. [23] presented a predictive regression model based on Large Deformation Diffeomorphic Metric Mapping for longitudinal images with missing data. Their methodology focuses on capturing linear changes within the image sequence. Longitudinal data plays a crucial role in enabling researchers to quantify the duration of various occurrences, capture temporal information, and facilitate the measurement of changes within a sample over time. While longitudinal data analysis has received relatively less attention, the integration of neuroimaging with longitudinal data analysis holds significant promise for the early detection of AD [24]. A study conducted by Juan et al. [25] demonstrated that employing both short-term and long-term longitudinal analysis spans a period exceeding three years. The authors concluded that, particularly in the context of long-term longitudinal data analysis, a substantial research gap exists in the field of computer-aided diagnosis of AD. In another study, Mofrad et al. [26] employed a method that utilized measurements of lateral and hippocampal ventricle volumes derived from longitudinal brain MRI data obtained from the AIBL and ADNI databases. By analyzing these measurements over a 15-year timespan, the authors examined the differences and changes observed. Similarly, Hojjati et al. [27] adopted a feed-forward multilayer perceptron approach, employing longitudinal neuroimaging data obtained from PET and sMRI. Through quantitative analysis, the authors examined the progression of AD. The study used both single and multimodal approaches, revealing significant associations between neuropsychological scores, PET, and sMRI in the detection of severe AD compared to normal aging. While longitudinal MRI data offer advantages for detecting AD progression, such as capturing disease dynamics over time, its limited sample size can be overcome by employing ensemble models, which combine multiple models to improve generalization, reduce overfitting, and enhance the accuracy and robustness of predictions.

Ensemble learning (EL) techniques play a pivotal role in enhancing predictive performance by combining multiple weak models into a single robust model using strategies such as bagging, stacking, and boosting. By aggregating outputs and reducing individual model errors, the EL model demonstrates superior robustness and consistently achieves higher accuracy compared to standalone models [28]. In the field of AD classification, EL has been widely adopted to improve diagnostic accuracy. Khan et al. [29] proposed an EL model that integrates a Decision Tree, extreme gradient boosting (XGBoost), and a polynomial kernel-based support vector machine (SVM). Through cross-validation and comparative analysis with other ML models, their approach achieved an AUC of 95.75 %, outperforming many existing methods. Similarly, Liu et al. [30] introduced a multi-level classifier approach that employs a hierarchical framework to incorporate brain features extracted from MRI scans at different levels. By leveraging imaging parameters from multiple scans, their method achieved an accuracy of 92 %. Building upon these advancements, ML researchers have increasingly explored the integration of EL with DL methodologies for AD diagnosis using MRI data. Several DL based techniques have been developed specifically for medical image analysis and automated feature extraction, enabling improved disease detection, classification, and progression modeling [31,32]. These approaches have been particularly applied to AD-related MRI datasets to extract meaningful diagnostic biomarkers. Hedayati et al. [33] proposed a pretrained autoencoder for feature extraction within an ensemble framework where deep features are first extracted from 3D MRI scans, and CNNs are subsequently employed for AD classification, capturing both spatial structures and temporal variations. Extending this concept, Battineni et al. [34] introduced a hybrid EL framework by merging four distinct classifiers i.e., k-nearest neighbors (kNN), SVM, neural networks, and Naïve Bayes to enhance predictive accuracy. All these studies demonstrate that the

EL approach significantly enhances AD diagnosis performance when applied to MRI-based classification models. By integrating EL frameworks or multiple CNN architectures, researchers have consistently achieved higher accuracy, improved feature extraction, and better predictive modeling for AD progression. These findings underscore the increasing significance of ensemble-based DL approaches in medical imaging and the detection of neurodegenerative diseases.

However, limitations of the existing AD progression detection approaches are that they either process a single representative slice from the entire 3D MRI volume or analyze multiple slices, but only from a single anatomical plane. This simplified approach neglects the rich spatial and temporal information embedded within the 3D MRI volumes, which is crucial for accurately modeling disease progression. Furthermore, most studies employ a binary or ternary classification framework, categorizing patients as CN vs. AD or CN vs. MCI vs. AD, while neglecting the nuanced pathological changes that occur over time. To address these challenges, this study processes longitudinal 3D MRI scans at four time points: baseline (BL), month 6 (M06), month 12 (M12), and month 18 (M18) for each patient, enabling a more comprehensive analysis of AD progression. The main aim is to capture a fine-grained temporal context of AD's progression. The number of four longitudinal time steps were chosen based on the high volume of available patient data in the ADNI and NACC datasets and their ability to reflect short to mid-term morphological changes in key brain regions. Existing studies have shown that adding sequential MRI data over approximately a three-year timespan substantially improves the predictive performance of an ML model, particularly for conversion and disease progression tasks [35,36]. For instance, incorporating MRI history into predictive models yields the largest performance gains around the third annual visit, after which returns diminish [37]. Moreover, anatomical changes measured across intervals of 6∼24 months reliably capture brain atrophy and progression signals [38].

We propose a novel AD progression detection model which is comprised of a hybrid DL model that efficiently processes high-dimensional MRI data while maintaining critical structural and temporal features necessary for accurate disease modeling. Our proposed approach starts with the selection of the most informative and medically validated 2D slices from each anatomical plane (axial, sagittal, and coronal) to ensure comprehensive coverage of critical brain regions associated with AD. This process is performed across all longitudinal time points to maintain spatial and temporal consistency. Next, the corresponding slices from each time step are stacked sequentially, forming structured 3D input volumes that effectively encode disease progression patterns (as illustrated in Fig. 1). To extract spatiotemporal representations, we employ an optimized 3D-CNN, which processes structural 3D volumes independently for different anatomical planes. The 3D-CNN captures fine-grained spatial features while preserving the volumetric context, which is critical for distinguishing AD-related abnormalities. Following this, the extracted multi-plane feature maps are integrated and processed using a bidirectional long short-term memory (BiLSTM) network, which models the longitudinal dependencies in the disease trajectory. BiLSTM allows the network to learn complex temporal patterns across MRI time points, enhancing its ability to differentiate between normal aging and pathological progression. To further refine the extracted features, we introduce an enhanced residual multi-headed self-attention (ERMHA) module, which dynamically reweighs important features while suppressing redundant or less informative regions. This mechanism enhances feature representation by emphasizing disease-specific biomarkers, ensuring robust predictions.

We conducted a comprehensive set of experiments (as illustrated in Fig. 5) to evaluate the effectiveness of integrating multiple MRI planes in a longitudinal manner for disease identification. Additionally, we examined how the diagnostic capabilities of the proposed framework improved with increasing architectural complexity by incorporating enhanced feature extraction and processing mechanisms. To further validate clinical viability and trustworthiness, we thoroughly evaluate the proposed framework through two complementary analyses. First, we assess the model's generalizability through external validation on an out-of-distribution dataset (i.e., the NACC cohort), evaluating its cross-dataset robustness under domain shift conditions. Second, we generate explainable attention maps for the proposed framework, demonstrating that model predictions are guided by neuroanatomically plausible

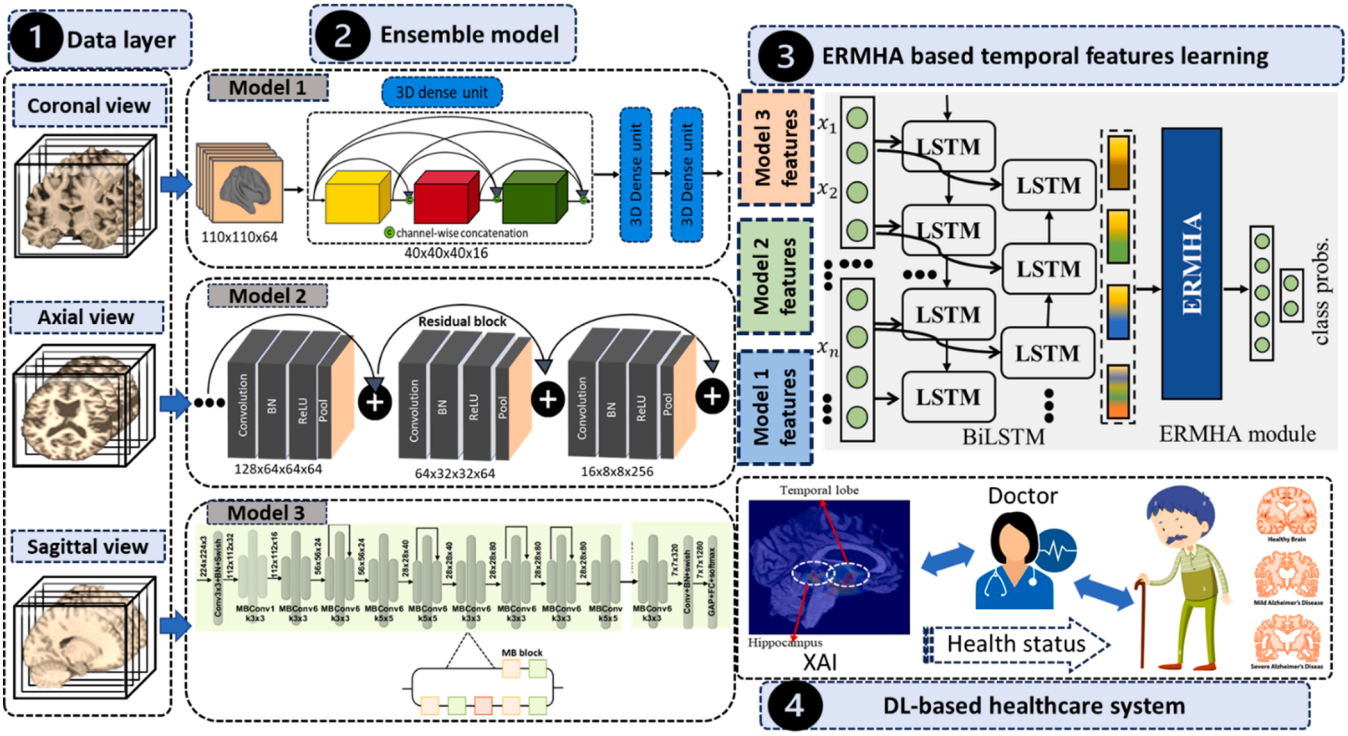


Fig. 1. Proposed framework for AD progression detection.

patterns associated with early-stage neurodegeneration in AD patients. Together, these analyses reinforce the framework's reliability and highlight its potential for clinical translation in real-world healthcare settings.

To summarize, the key contributions of this study are as follows:

- We propose a novel heterogeneous deep ensemble model that uniquely integrates multi-plane spatial fusion, multi-timepoint temporal modeling, and plane-specific architectural optimization for AD progression detection. Unlike existing approaches that employ homogeneous ensembles or single-timepoint analysis, our framework combines three key innovations: (1) a spatially-guided slice sampling strategy across four longitudinal time points, (2) an Enhanced Residual Multi-Head Self-Attention (ERMHA) mechanism tailored for medical imaging, and (3) a heterogeneous ensemble that assigns optimal 3D-CNN architectures to each anatomical plane. This unified combination achieves state-of-the-art performance and significantly outperforms existing models in literature.
- We investigate the impact of incorporating multi-slice and multi-plane fusion (i.e., axial, coronal, and sagittal) in longitudinal 3D MRI for detecting the progression of AD. By leveraging multiple spatial perspectives of the MRI, we ensure a more comprehensive representation of structural brain changes associated with AD.
- We propose a novel 3D volumetric representation of the 2D MRI slices from the longitudinal 3D MRI at four critical time points (BL, M06, M12, and M18). This specialized 3D volume is created by concatenating spatially aligned slices from each MRI plane at different time points, effectively targeting vulnerable brain regions such as the hippocampus, amygdala, and their subregions, which are significantly impacted during AD progression.
- We empirically evaluated five prominent 3D-CNNs as feature extractors to model high-level spatiotemporal dependencies from the proposed 3D volumetric representation. These models leverage 3D convolutional kernels to learn hierarchical feature representations, capturing intricate disease-related structural and temporal changes. Furthermore, the extracted deep features are refined and processed using BiLSTM, to model the sequential progression patterns of AD with greater temporal fidelity.
- We enhance the architectural complexity of the proposed AD progression detection framework by integrating an ERMHA module into the BiLSTM network. This modification selectively emphasizes salient temporal features, improving the model's ability to capture long-range dependencies and subtle disease progression patterns with greater interpretability.
- We explore homogeneous and heterogeneous ensemble learning strategies to further improve model robustness. The homogeneous ensemble involves aggregating multiple instances of the same 3D-CNN model, while the heterogeneous ensemble combines diverse 3D-CNN architectures to capture complementary feature representations. Additionally, we integrate ERMHA into the LSTM-based sequence learning framework to weigh critical spatiotemporal features, optimizing AD progression detection across multiple viewpoints.
- We conduct a comprehensive experimental evaluation of the proposed framework on the ADNI dataset and report state-of-the-art performance with mAcc of 93.73 %, mSen of 91.72 %, mSpe of 90.36 %, and mAUC of 91.58 %. These results demonstrate that proposed heterogeneous deep ensemble network with increased longitudinal time steps in patient's data substantially outperforms other models including single-plane models (mAUC: 68.24 %) and homogeneous deep ensemble approaches (mAUC: 82.75 %).
- We further validate the clinical aspect of the proposed framework using two complementary analyses: (1) external validation on the independent NACC cohort, and (2) model explainability using gradient-weighted class activation mapping. External validation demonstrates robust cross-dataset generalizability, achieving mAUC of 86.37 % on NACC data despite its inherent domain shift challenges. Furthermore, explainability analysis using M3d-CAM reveals that model predictions are driven by neuroanatomically plausible patterns, including hippocampal and entorhinal cortex atrophy in progressive MCI patients, as well as atrophy in the posterior cingulate and temporo-parietal regions in AD patients. These findings align with established models of neuropathological progression and further support the clinical trustworthiness of our proposed approach.

The remainder of this paper is organized as follows. Section 2 details the materials and methods, including dataset characteristics, preprocessing steps, and the architectural components of the proposed heterogeneous deep ensemble framework. Section 3 presents comprehensive experimental results. Section 4 establishes clinical viability through external validation on the NACC cohort and explainability analysis via attention visualization. Section 5 compares proposed approach with state-of-the-art methods. Section 6 analyzes computational complexity and deployment considerations. Section 7 discusses limitations and future directions, and Section 8 concludes the paper.

2. Materials and methods

The proposed study is conducted using longitudinal MRI data spanning four timesteps. First, a 3D volume was created from each MRI plane by stacking the most crucial MRI slices from the longitudinal time steps. Next, a comprehensive set of experiments was conducted to evaluate various DL based 3D models using the newly created 3D volumes. Finally, we analyzed the effectiveness of individual MRI planes and the fusion of multiple planes, incorporating increasing architectural complexity, to assess their impact on AD progression detection. Fig. 1 illustrates the workflow of the proposed AD progression detection framework.

2.1. Dataset

Data used in the preparation of this study were obtained from the ADNI [39]. ADNI was launched in 2003 as a public-private partnership. The primary goal of ADNI has been to test whether different neuroimaging modalities, biological and clinical biomarkers, can be combined to measure the progression of AD. In this study, we employed 2248 (562 $\times$ 4) MRI volumes collected at four longitudinal time points (BL, M06, M12, and M18). Our model predicts changes in patients' health status after 2.5 years, based on the final assessment visit, which occurred in month 48 (M48). Table 1 summarizes the demographic characteristics of the participants, including age, gender, years of education, hippocampus and mini-mental state examination scores (MMSE). Moreover, Fig. 2 illustrates cognitive state transitions of a subject across longitudinal time steps used in our predictive framework. The diagram visualizes the progression of subjects from BL to M18, and their predicted cognitive status at M48.

2.1.1. Image pre-processing

The dataset used in this study consists of 3T T1-weighted anatomical sequences. These sequences were acquired using the volumetric 3D

Table 1
Demographic features of the available subjects.

Feature	Normal subjects	Converted to AD
Available subjects	282	280
Age	72.70±05.80	74.50±04.02
Gender	0.41±00.50	0.49±00.51
Education	16.58±02.53	15.77±02.81
Hippocampus	11.70 ± 29.18	93.82 ± 61.60
MMSE	04.13 ± 05.87	03.63 ± 02.48

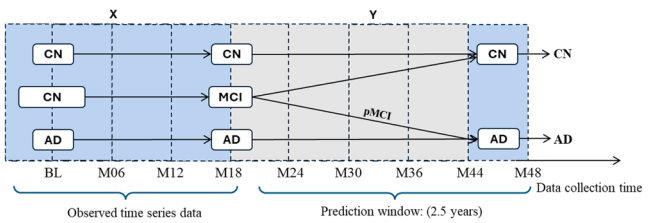


Fig. 2. Illustration of cognitive state transitions across longitudinal time steps used in our predictive framework. The diagram visualizes the progression of subjects from BL to M18, and their predicted cognitive status at M48.

MPRAGE protocol. The voxel size of the acquired images was $1 \times 1 \times 1$ mm, indicating a high-resolution acquisition. From the total of 562 longitudinal volumes, 282 participants were initially classified as cognitively normal across all four time points, while 280 subjects were cognitively normal during the initial visits but later developed AD three years after their last visit at M18 (at M48). All MRI volumes underwent several preprocessing steps. First, each MRI volume was visually inspected and adjusted for left and right orientation according to the anatomical coordinate system. This step ensured that the subsequent analyses were conducted with accurate anatomical alignment. Next, the N4 bias field correction algorithm from the advanced normalization tools (ANT) was employed to correct for inhomogeneities within each MRI volume. This algorithm effectively compensated for intensity variations caused by magnetic field inhomogeneities, resulting in more accurate and reliable image data. Following the bias field correction, the brain extraction tool (BET2) from the FSL software package was utilized for skull stripping. This process effectively removed non-brain tissues and preserved the brain structures of interest in the MRI volumes.

Finally, to enable meaningful comparisons across subjects and studies, all slices of the MRI volumes were registered onto the Montreal Neurological Institute (MNI) 152 template. We employed a rigid/affine registration approach using linear image registration tool (FLIRT) from the FSL tool, executed via the command line. The registration was configured with 12 degrees of freedom, allowing for translation, rotation, scaling, and skew transformations to account for inter-subject spatial variability. To improve intensity-based alignment, we used 256 histogram bins for robust histogram matching between the source MRI and the MNI152 template. Voxel intensities were interpolated using spline interpolation, which minimizes interpolation artifacts while preserving fine anatomical detail. For the similarity metric, we selected the correlation ratio cost function, which is effective for multimodal and intramodal registration by maximizing the statistical dependency between corresponding voxel intensities. This combination of settings ensured precise anatomical alignment while maintaining the fidelity of structural features essential for downstream voxel-wise analyses. By transforming the MRI data into the MNI-152 template space, it becomes possible to align and compare brain structures across different individuals and studies. Through these preprocessing steps, the MRI data was prepared for subsequent analyses, ensuring accurate anatomical alignment, intensity correction, removal of non-brain tissues, and alignment to a common reference space.

2.1.2. 2D-slice selection from the 3D-MRI

We conducted a comprehensive analysis of AD's progression detection by sampling the most crucial 2D slices from each step of the longitudinal T1-weighted MRI. The sampling process involved selecting multiple 2D middle slices from each volumetric scan across all three anatomical planes: axial, coronal, and sagittal at each longitudinal time step (BL, M06, M12, and M18). The ADNI T1-weighted MPRAGE volumes were originally acquired with voxel sizes of $1.0 \sim 1.2$ mm isotropic resolution (with minor variations up to 1.5 mm in some protocols across ADNI phases but standardized to $\sim$1 mm^3 for consistency in volumetric analysis). We selected 16 middle slices ($\pm$8 around the mid slice) at 3.0 mm intervals as a balanced trade-off between anatomical coverage and computational efficiency. While the native isotropic resolution allows for finer sampling, the 3.0 mm spacing provides sufficient separation to reduce redundancy between highly correlated adjacent slices (e.g., consecutive 1 mm slices often exhibit >95 % similarity in structural features due to the smooth nature of brain anatomy), while maintaining consistent coverage of AD-relevant regions, such as the hippocampus, medial temporal lobe, and adjacent cortical structures, key areas where atrophy and volumetric changes occur on a scale of several millimeters. This approach minimizes potential loss of structural information by focusing on spatially diverse yet representative slices, as subsampling at this interval has been shown to preserve diagnostic utility in DL based models for AD without introducing significant information gaps, particularly since finer details below 3 mm (e.g., microvascular changes) are not primary indicators for progression detection in standard T1-weighted imaging.

Furthermore, the proposed slice selection strategy was further supported by a prior study conducted by Silva et al. [40], emphasizing the relevance of selecting at 3.0 mm intervals and the corresponding brain regions they cover in the context of early progression detection in AD patients. In contrast, using the entire 3D MRI volume, while feasible in principle, substantially increases the computational demands (e.g., processing hundreds of slices per volume per time step across multiple planes would require an ensemble of 3D CNNs processing, leading to exponentially higher memory usage often exceeding 24$\sim$64 GB GPU VRAM for batch sizes > 4 and training times up to 10x longer due to the cubic scaling of 3D operations, without achieving any proportional benefits in the diagnostic accuracy, as full-volume approaches often suffer from overfitting to noise or irrelevant peripheral brain regions. Our multi-plane slice-based approach, in contrast, enables efficient 3D CNN processing (e.g., via channel concatenation), which is a common and validated technique in medical imaging DL to approximate full 3D context with reduced complexity. In addition, in a supplementary evaluation, we found that increasing the number of slices beyond 16 at each time step or decreasing the slice spacing to less than 3.0 mm (e.g., to 1.0 mm) resulted in only marginal gains (e.g., < 2 % gain in the mAUC score across validation sets, with no statistically significant difference in sensitivity/specificity for early-stage progression), while leading to a substantial increase in computational and memory requirements (e.g., $3\sim5 \times$ higher GPU memory and training time due to increased input dimensions and parameter counts). Therefore, the adopted configuration represented a well-structured and anatomically informed decision, optimized for both performance and feasibility, making it a precise sampling approach in the sense of efficiently targeting diverse, clinically relevant anatomical features without redundant or low-value data.

The extracted slices corresponding to each plane and longitudinal time step were concatenated along the channel dimension in chronological order (BL, M06, M12, and M18). This process yielded a new 3D volume with dimensions $110 \times 110 \times 64$. Fig. 3 demonstrates how selective slices from each anatomical plane are extracted and integrated to create a comprehensive representation of disease progression over time. Each volume incorporates temporal information by aligning the most crucial slices from each time point, thus preserving the spatio-temporal context critical for detecting progressive disease patterns. These volumes were further processed using an optimized 3D-CNN model tailored for each anatomical plane. The 3D-CNN employed 3D convolutional kernels to extract hierarchical features that encompassed both spatial (x, y) and temporal (z) dimensions. This approach leveraged the depth of the model to capture complex patterns indicative of disease progression.

Following the feature extraction step, the outputs from the 3D CNN models were further processed through time series models. Specifically, we evaluated the performance of LSTM integrated with an ERMHA module. This combination was chosen for its ability to capture temporal dependencies and dynamic changes across the longitudinal time steps. The multi-headed self-attention enabled the model to focus on different

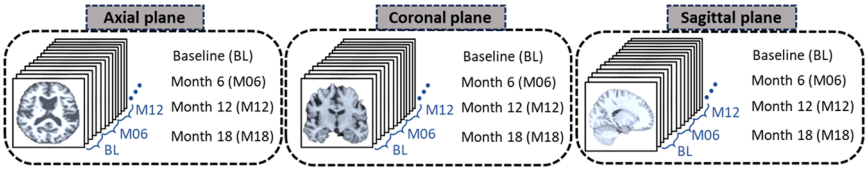


Fig. 3. A visual representation of the steps taken for generating a unified 3D volume using selective slices from the longitudinal MRI at multiple time steps.

parts of the input sequence, thereby enhancing its ability to detect subtle, progressive patterns in the data. It is important to note that the choice of extracting 16 slices from each longitudinal time step was empirically determined to be optimal. Increasing the slice count to 32 or 64 per MRI plane was found to significantly elevate computational demands and the risk of overfitting the model, thereby compromising the generalizability of the model. This balance ensured efficient computation while maintaining model robustness. To validate the efficacy of the proposed AD progression detection framework, a comprehensive set of experiments was conducted. These experiments assessed the framework's ability to distinguish between different stages of AD progression using the prepared dataset. Detailed results and a thorough discussion of these findings are presented in Section 3 **"Results and Discussion"**, where insights on model performance are provided.

2.2. Deep feature learning via an ensemble of 3D convolution neural networks

As part of the proposed framework, we evaluated an ensemble of 3D-CNN architectures for deep feature extraction from 3D MRI volumes. Each model possesses a unique architectural design, varying in depth, convolutional kernel size, and feature aggregation mechanisms, which influence its ability to learn and represent complex medical imaging patterns. The 3D-CNN models integrated into our framework include 3D-VGG, 3D-ResNet, 3D-DenseNet, 3D-InceptionNet, and 3D-EfficientNet, each offering distinct advantages in terms of feature abstraction, computational efficiency, and representational capacity. For instance, 3D-VGG is a straightforward and deep architecture that employs stacked convolutional layers with small ($3 \times 3 \times 3$) kernels, enabling hierarchical feature extraction. Despite its simplicity, it serves as a strong baseline model, effectively capturing local spatial features while maintaining computational efficiency. 3D-ResNet incorporates residual connections, which help mitigate the vanishing gradient problem in deeper networks. This enables efficient gradient flow and improves training stability, making it well-suited for capturing intricate anatomical patterns across multiple MRI slices while preserving essential low-level and high-level spatial features. 3D-DenseNet utilizes dense connections where each layer is directly connected to all subsequent layers, promoting feature reuse and reducing redundant computations. This enhances gradient propagation, encourages efficient learning of discriminative spatial features, and improves model generalization. 3D-InceptionNet features multi-scale convolutional filters within the Inception module, allowing the network to learn both fine and coarse-grained spatial features simultaneously. This architecture is particularly effective in capturing complex anatomical structures by processing spatial variations at multiple scales. Finally, 3D-EfficientNet employs compound scaling to optimize network depth, width, and resolution, achieving a balance between model complexity and computational efficiency. By leveraging depth-wise separable convolutions and squeeze-and-excitation modules, it enhances feature extraction while reducing computational overhead, making it a powerful yet efficient choice for volumetric medical imaging tasks.

Notably, heterogeneous deep EL approaches combine multiple diverse CNN models into a single model, offering significant advantages over relying on a single model. By harnessing the complementary strengths of different networks, heterogeneous EL models typically achieve superior generalization (as demonstrated in Section 3.6) and robustness, reduce variance (as shown in **Exp. 7**), and compensate for individual model weaknesses (by comparing **Exp. 1** and **Exp. 5**). Theoretical and empirical studies have consistently shown that EL models improve predictive performance compared to individual models [41]. Furthermore, in neuroimaging tasks, including AD detection, deep EL frameworks have shown significantly improved performance than their component deep models [42]. In one study, an EL model has achieved 4 % higher classification accuracy compared to standard stacking or single-model baselines [43]. In another study, a dual-level ensemble combining vision transformer (ViT) based feature extraction and classifier-level fusion significantly outperformed prior state-of-the-art methods in brain tumor MRI classification tasks [44]. In contrast, non-ensemble models, though simpler to design, often suffer from higher variance (as shown in **Exp. 1**), overfitting, and limited generalization, especially in complex domains such as medical imaging. They lack the complementary benefits inherent in ensemble approaches, which are particularly vital for capturing subtle anatomical variations in MRI data [41].

2.3. Temporal features learning via BiLSTM

Bidirectional LSTM (BiLSTM) is an extension of the simple LSTM architecture that introduces an additional layer of complexity by processing the input sequence in both forward and backward directions. This architecture enables the BiLSTM to capture information from both past and future contexts, thereby enhancing its contextual understanding, particularly in tasks involving different classes of cognitively impaired individuals. The LSTM unit consists of input, hidden, forget, and output gates, along with a cell state. The computations of the forward LSTM unit can be expressed as follows:

$$\Phi_t = \sigma\left(K_f x_t + V_f h_{t-1} + \beta_f\right)$$

$$\psi_t = \sigma(K_i x_t + V_i h_{t-1} + \beta_i)$$

$$\omega_t = \sigma(K_o x_t + V_o h_{t-1} + \beta_o)$$

$$\mathrm{Y}_t = \tanh(K_c x_t + V_c h_{t-1} + \beta_c)$$

$$s_t = \Phi_t . s_{t-1} + \psi_t{}^* \mathrm{Y}_t$$

$$z_t = \omega_t{}^* \tanh(s_t)$$

In these equations, Φ_t represents the forget gate output, ψ_t represents the input gate output, ω_t represents the output gate output, Y_t represents the candidate cell state output, s_t represents the cell state output, and z_t represents the hidden state output. The σ denotes the sigmoid activation function, and *tanh* denotes the hyperbolic tangent activation function. x_t is the input at time step t, h_{t-1} is the previous hidden state, and s_{t-1} is the previous cell state. K_f, K_i, K_o, K_c, V_f, V_i, V_o, V_c are weight matrices, and β_f, β_i, β_o, β_c are bias vectors.

Similar computations are performed for the LSTM unit in a backward direction. The overall output of the hidden state in the backward direction can be denoted as:

$$h_{t_{backward}} = f_{backward(x_t,\ h_t+1_{backward},\ LSTM_{backward})}$$

Here, $f_{backward}$ represents the computations performed by the LSTM unit

in the backward direction, taking into account the input x_t, the subsequent hidden state $h_t + 1_{backward}$, and the $LSTM_{backward}$ label. We initially combine the feature vectors from both forward and backward LSTM layers and process it via a multilayer perceptron to classify the output into CN vs. progressed to AD classes.

As part of the proposed AD progression detection framework, we iteratively optimized the hyperparameters of the BiLSTM subnetwork using the Grid Search approach [45]. The following is the list of optimized hyperparameters (a) Number of LSTM layers: Defines the number of LSTM layers (b) Number of LSTM cells: Defines the number of memory units in each LSTM layer. (c) Dropout rate: Determines the fraction of hidden units randomly deactivated during training to prevent overfitting. The initial value was set at 0.1. (d) Regularization: Helps control model complexity and reduce overfitting. In this study, L1 regularization was chosen as the initial method. The final optimized set of hyperparameters consisted of two Bidirectional LSTM (BiLSTM) layers, with the first layer containing 128 LSTM units and the second layer containing 64 LSTM units. Each BiLSTM layer utilized a dropout rate of 0.1 to mitigate overfitting by randomly deactivating a fraction of neurons during training. To enforce sparsity and improve generalization, L1 regularization (Lasso) was applied to the recurrent weight matrices, with values set to 0.01 for the first BiLSTM layer and 0.03 for the second layer. These values were selected to balance model complexity and performance, ensuring robust feature extraction while preventing overfitting.

2.4. Temporal features modeling via enhanced residual multi-head self attention

A crucial aspect of human perception is its selective nature when processing external stimuli. Rather than processing all inputs simultaneously, humans tend to prioritize and focus on the most essential parts to extract the desired information. Similarly, in the diagnostic process of disease progression using MRI data, the significance of different MRI planes is crucial. While some information may be redundant, other details can be critical for an accurate diagnosis. Therefore, in disease prediction based on patients' medical data, it becomes necessary to prioritize key features and discard redundant ones. This enables the model to leverage effective information from various MRI planes to build an efficient predictive model and inform decision-making processes. The ERMHA proposed in this study is inspired by the BERT encoder [46], where the bidirectional self-attention structure of the transformer encoder has demonstrated superior performance over unidirectional self-attention in capturing long sequential patterns.

In ERMHA, we adopted the standard scaled dot-product attention mechanism to compute multi-head attention. However, we also proposed four key architectural modifications to this mechanism to better suit biomedical imaging tasks such as processing 3D MRI for disease progression analysis. These changes are specifically designed to improve computational efficiency, reduce parameter overhead, and enhance the model's ability to capture global anatomical relationships across spatial slices as well as longitudinal time points. ***First***, we removed the masking matrix operation that is traditionally used in Transformers for controlling attention flow (e.g., enforcing causality in language models). Since our task involves analyzing spatial and temporal contexts where such masking is unnecessary, this removal enables unrestricted bidirectional attention across all positions which are critical for capturing comprehensive inter-slice dependencies in volumetric medical scans. ***Second***, we removed the Feed-Forward Network (FFN) component, which typically follows each attention block. While FFNs enhance non-linear representational power, they also significantly increase computational cost and parameter count. In our case, empirical observations showed that the self-attention mechanism alone is sufficient to model the rich spatial correlations in medical images. Removing the FFN resulted in a lightweight yet expressive module that prioritizes global context modeling over complex feature transformation. ***Third***, we introduced a direct residual connection immediately after the attention computation, preserving the original input features and reinforcing them with attention-enhanced representations. This residual pathway helped mitigate vanishing gradient issues in deeper architectures and improved convergence, particularly important when training on high-dimensional imaging data. ***Fourth***, we applied a single Layer Normalization after the residual connection, instead of the dual normalization layers (before attention and after FFN) used in the standard Transformer architecture. This simplification reduced computational overhead while maintaining training stability, which is especially beneficial when processing high-resolution 3D medical volumes that demand large memory footprints. Together, all these modifications defined the ERMHA module, which is a streamlined, yet powerful variant of multi-head self-attention tailored for medical imaging tasks. By focusing on unrestricted global attention and minimizing architectural complexity, ERMHA enables efficient and effective modeling of long-range dependencies across slices or time points in volumetric data, making it well-suited for applications such as AD progression detection.

The steps can be mathematically represented as follows:

Deep features extracted by 3D-CNN representing both spatial and inter-slice features for each plane are denoted by a sequence of $X = [x_1, x_2, \ldots, x_n]$, where x_i represents the i^{th} element of the sequence and n is the length of the sequence. The BiLSTM processes this sequence in both forward and backward directions, producing two sets of hidden states: $H_f = \left[h^{1f}, h^{2f}, \ldots, h^{f}_{n}\right]$ and $H_b = \left[h^{1b}, h^{2b}, \ldots, h^{b}_{n}\right]$. Here, h^{f}_{i} and h^{b}_{i} represent the forward and backward hidden states at position i, respectively.

The output of h^{f}_{i} and h^{b}_{i} is concatenated and processed through multiheaded self-attention.

$$AttenOutput = MultiHead(Q, K, V) + AttenInput$$

The *AttenOutput* is concatenated hidden state from both directions, and the operation helps mitigate information degradation during the processing. Following this, a Layer Normalization operation is applied:

$$NormOutput = LayerNorm(AttenOutput)$$

The final attention vector $A = [a^1, a^2, \ldots, a_n]$ captures the importance of different hidden states based on their relationships with other states, while also applying the significance of residual connections and layer normalization. Fig. 4 illustrates the architectural design of the proposed ERMHA module.

The output features from the BiLSTM layers are further processed through the ERMHA module. The ERMHA enables the model to attend to multiple aspects of the input features simultaneously, allowing for richer feature extraction. In addition, the residual connections further mitigate the risk of vanishing gradients and facilitate stable gradient flow,

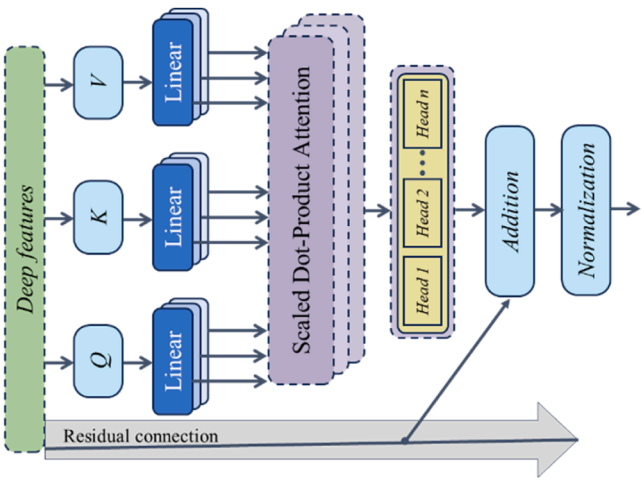


Fig. 4. Enhanced residual multi-head self-attention.

ensuring better convergence during training. In the proposed AD progression detection framework, we stacked two consecutive ERMHA layers. The first ERMHA layer enhances the feature embeddings by emphasizing critical dependencies, while the second ERMHA layer further refines and stabilizes these representations, ensuring that both local and global contextual dependencies are well-preserved. The output of the second ERMHA layer is passed through a dense neural network with 128 units, followed by an output layer of two neurons each producing a probability score corresponding to the likelihood of a given input belonging to either the CN class or the AD progression class.

2.5. Gradient-based attention map analysis

Deep learning models have demonstrated remarkable performance in medical image analysis, their "black-box" nature poses significant challenges for clinical adoption [47]. In healthcare applications, particularly for high-stakes decisions such as neurodegenerative disease diagnosis, clinicians require transparent, interpretable explanations that align with established medical knowledge. Explainable XAI addresses this critical gap by providing insights into which features, patterns, or regions of input data most strongly influence model predictions [48,49]. Several XAI techniques exist in literature for DNN models. For instance, gradient-based methods, including Grad-CAM [50] generate visual explanations by computing gradients of the target class with respect to feature maps in convolutional layers, producing spatial attention maps that highlight discriminative regions. Perturbation-based approaches, including LIME [49] and SHAP [48], explain predictions by systematically perturbing input features and measuring impact on model output. However, these methods face significant limitations when applied to complex ensemble architectures. LIME and SHAP become computationally prohibitive for high-dimensional 3D medical imaging data, particularly when processing multiple longitudinal time points across multiple anatomical views. Moreover, heterogeneous deep ensemble models with non-shared weights across different CNN architectures lack the architectural uniformity required for direct gradient-based or perturbation-based explanation methods.

To address these challenges while maintaining interpretability, we adopted a view-decomposition strategy combined with gradient-weighted class activation mapping. Rather than attempting to explain the full heterogeneous ensemble as a single entity, we generated explanations from simplified view-specific models that preserve the core feature extraction mechanisms. This approach is justified by two key observations: (1) the final ensemble prediction is constructed from view-specific feature representations, making individual view analysis clinically meaningful, and (2) each anatomical view provides complementary diagnostic information that can be interpreted independently by radiologists. For attention map generation, we employed M3d-CAM [51], which extends Grad-CAM to 3D volumetric medical data. M3d-CAM computes the gradient of the predicted class score with respect to the feature maps of the final convolutional layer, weighted by global average pooling to produce a coarse localization map highlighting important regions. Mathematically, for a given class c and feature map A at spatial location (i, j, k), the attention map L is computed as:

$$L^{c}_{i,\,j,\,k} = ReLU\left(\sum_{n} w^{c}_{n} \cdot A^{n}_{i,\,j,\,k}\right)$$

Where $w^{c}_{n} = \frac{1}{Z}\sum_{i,\,j,\,k}\frac{\partial y^{c}}{\partial A^{n}_{i,\,j,\,k}}$ represents the importance weight of feature map n for class c and Z is the total number of spatial locations. The $ReLU$ activation ensures that only features with positive influence on the target class are visualized. Attention maps were generated for each view-specific model using the validation set subjects representing three diagnostic categories i.e., CN, Progressed to AD, and AD. For each category, we selected representative cases based on (1) highest activation magnitude, and (2) anatomical coverage of known AD-vulnerable regions including hippocampus, entorhinal cortex, posterior cingulate cortex, and precuneus [52,53]. Attention values were initially normalized to the range [0, 1] and overlaid on original MRI slices using a jet colormap, where warmer colors (bluish/yellow) indicate stronger model attention. This normalization enables quantitative comparison of attention patterns across different subjects and diagnostic categories.

2.6. Model implementation

The experimental setup for this study is based on an NVIDIA TITAN GTX GPU equipped with 12 GB of VRAM. The proposed framework was developed using TensorFlow 2.0, a highly optimized ML framework for parallel processing and tensor computations. Training was conducted in an end-to-end manner, leveraging the Adam optimizer with an initial learning rate of 1×10^{-3}. The training loss was computed using binary cross-entropy loss. Each input MRI was initially resized to 110×110 spatial dimension with 64 channels as the depth dimension. The input 3D MRI volume was then processed through the proposed framework with a batch size of 8 vol in each iteration. To enhance robustness and mitigate model bias, a 10-fold stratified cross-validation technique was used in the training, ensuring that 70 % of the data was allocated for model training while the remaining 30 % was reserved for validation. The model was initially trained for 50 epochs with a learning rate decay rate of 1×10^{-5} to gradually reduce the learning rate. After each epoch, the learning rate was adjusted by a small factor based on the decay rate. Once the minimum loss was calculated, the optimum learning rate was subsequently adapted for retraining the entire model for 150 epochs. To prevent excessive training that could lead to overfitting, an early stopping mechanism was integrated, which halted the training process when no significant improvement is observed. For each fold in the 10-fold cross-validation, various performance metrics were calculated, including accuracy, sensitivity, specificity, and AUC. The mean value for each metric was then reported. These evaluation metrics provide a comprehensive understanding of the model's performance, especially in terms of balancing false positives and false negatives. Finally, the mean value of each metric was computed across all folds, providing insight into the model's ability to generalize effectively.

2.7. Model evaluation

Evaluation metrics are essential for analyzing the effectiveness of a DL model. These metrics are a numerical representation of measuring performance, allowing for fair comparisons and informed decisions when selecting models. To assess the performance and generalizability of our proposed model, we calculated a list of evaluation metrics including accuracy, sensitivity, specificity, and AUC. The definition of

Table 2
Evaluation metrics along with the respective mathematical formula.

Metric	Description	Formula
Accuracy	Measures overall model performance by evaluating how often the classifier makes correct predictions. It is calculated as the ratio of correctly classified instances to the total number of instances.	$\frac{(TN + TP)}{TP + TN + FR + FN}$
Sensitivity	Also known as recall, it measures the proportion of actual positives that are correctly identified by the model.	$\frac{(TN + TP)}{TP + FN}$
Specificity	Also known as the true negative rate, it indicates how well the model identifies true negatives. It is calculated as the proportion of true negatives out of all actual negatives.	$\frac{TN}{TN + FP}$
AUC	Area Under the ROC Curve, representing the model's ability to distinguish between classes. A higher AUC indicates better classification performance.	

each metric along with its mathematical formula is summarized in Table 2.

3. Results and discussion

This study addresses the challenge of detecting AD progression by leveraging the most crucial, medically validated 2D slices extracted from 3D longitudinal MRI scans. Our investigation demonstrates that enhancing input data with greater architectural complexity significantly improves the overall performance of ML models in AD progression detection. Furthermore, we examine the impact of incorporating an additional layer of complexity into the predictive framework, analyzing how performance evolves as an extra processing layer is introduced. To comprehensively evaluate both aspects of the proposed framework, we conducted a series of diverse experiments. Fig. 5 illustrates the experimental workflow for this study. In **Exp. 1**, we explored a range of 3D-CNN architectures to process spatiotemporal features extracted from 3D volumetric MRI data, assessing their capability in capturing structural and temporal patterns. **Exp. 2** investigates the effect of integrating information from multiple MRI planes to improve disease progression detection. **Exp. 3** increases the architectural complexity of the proposed model by incorporating a BiLSTM network alongside the 3D-CNN backbone. In **Exp. 4**, an ERMHA module was integrated into the BiLSTM model. This addition emphasized the extraction of deep, discriminative features, enabling the generation of a more refined feature set that focuses on highly correlated attributes relevant to disease identification. Finally, **Exp. 5** and **Exp. 6** examined the effectiveness of homogeneous and heterogeneous deep EL approaches in enhancing disease identification performance.

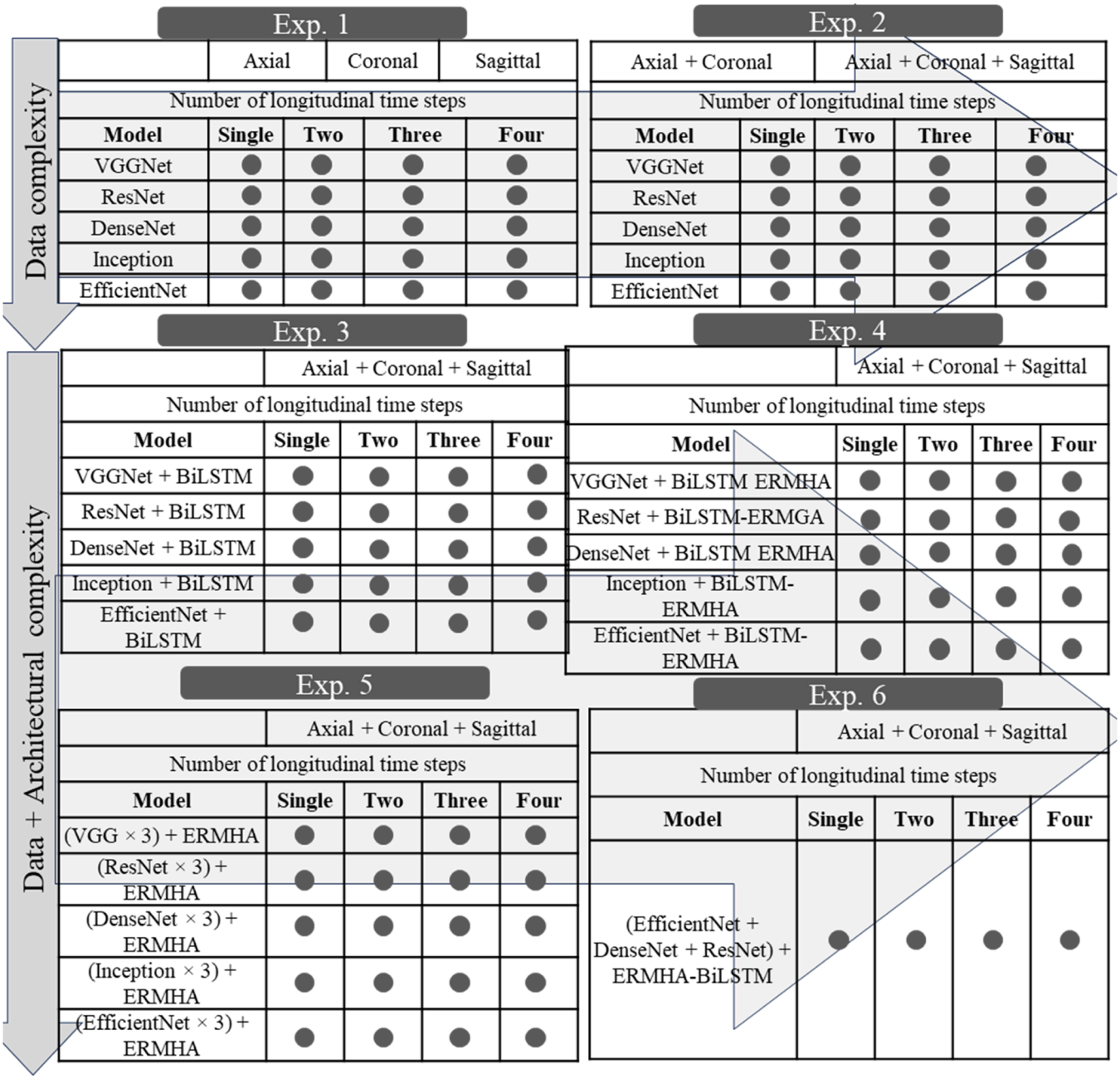


Fig. 5. Experimentation pipeline of the proposed study.

3.1. Experiment 1: optimization of 3D-CNN using single plane of the longitudinal 3D MRI

In **Exp. 1**, we investigate the effectiveness of MRI planes (i.e., axial, coronal, and sagittal) in identifying diseased patterns in brain tissues associated with AD. To achieve this, we first evaluate a comprehensive set of 3D-CNN models, applying them to the 3D volumes derived from the preprocessing step discussed in Section 2.1.1. These volumes are made of 2D slices that contain the anatomical variations of the brain tissues across the three MRI planes, ensuring a robust dataset for comparative evaluation. Additionally, we examined the impact of integrating longitudinal time steps into the training data, aiming to enhance the detection of AD progression over time. This approach leverages temporal patterns in brain morphology changes, providing the models with richer contextual information about disease trajectory. By identifying the optimal plane, we aimed to improve the spatial representational power of 3D-CNN models for subsequent analyses, ultimately enhancing their performance in clinical prediction tasks. Moreover, we placed particular emphasis on analyzing how the incorporation of additional longitudinal time steps impacted model performance and stability.

The results presented in Table 3 shows that the sagittal plane was the most effective plane in highlighting brain regions affected by AD, with 3D-DenseNet achieving the highest Accuracy (i.e., mAcc=67.74±3.75, mSen=68.12±3.93, mSpe=62.42±4.21, and mAUC=68.24±4.62) at BL-M18. For the coronal plane, 3D-ResNet outperformed other models, with an average performance of mAcc=63.42±4.51, mSen=62.33 ±5.42, mSpe=59.71±3.72, and mAUC=65.21±4.29, while the axial plane yielded the best results with 3D-EfficientNet, reporting mAcc=65.72±4.39, Sen=62.89±5.72, Spe=58.34±4.52, and AUC=64.73±3.71. These results indicate that different planes offer complementary insights, with the sagittal plane providing the most critical spatial information for detecting AD progression. A notable improvement in performance was observed across all models and planes as longitudinal time steps were increased from BL to BL∼M18, underscoring the importance of incorporating temporal data to capture disease progression. For instance, 3D-VGG on the axial plane showed an accuracy increase from 50.18±5.34 at BL to 59.37±4.34 at BL∼M18, and 3D-InceptionNet on the sagittal plane improved from 49.14±2.15 at BL to 57.12±2.72 at BL∼M18. These results highlight that including temporal data enables the models to capture inter and intra-slice features more effectively, leading to enhanced classification accuracy and robustness. Model stability, as reflected by reduced standard deviations across metrics, also improved with the inclusion of longitudinal time steps. For instance, 3D-EfficientNet on the axial plane reduced variability in mSpe from ±6.43 at BL to ±4.39 at BL∼M18, demonstrating increased reliability.

Fig. 6 illustrates the effectiveness of different MRI planes and the impact of incorporating longitudinal data. For clarity and consistency, we focused our discussion and comparison on the mAUC metric only, this is because, all evaluation metrics are consistent for all models; therefore, using mAUC is a representative indicator for the rest of the evaluation metrics. The mAUC values for various models were analyzed and compared using different data combinations: BL, BL∼M06, BL∼M12, and BL∼M18. Among all the comparative models, 3D-EfficientNet demonstrated the best mAUC scores when utilizing the axial plane. Starting at BL, it achieved an mAUC score of 57 %, with further improvements as the number of longitudinal time steps increased. The model reported a 1 %, 6 %, and 7 % improvement with the addition of longitudinal time step, ultimately reaching 64 % at BL∼M18. Following closely, 3D-InceptionNet achieved the second-highest performance, reporting mAUC scores of 51 %, 53 %, 55 %, and 61 % with each additional time step in the training data. On the other hand, models such as 3D-VGG, 3D-ResNet, and 3D-DenseNet yielded fluctuating results, lacking an upward trending pattern. Using the coronal plane, 3D-ResNet achieved the highest performance at BL∼M18, achieving mAUC score of

Table 3
Performance comparison of various 3D-CNN models on each plane of 3D-MRI.

3D Model	T-S	Axial Plane				Coronal Plane				Sagittal Plane			
		mAcc	mSen	mSpe	mAUC	mAcc	mSen	mSpe	mAUC	mAcc	mSen	mSpe	mAUC
3D-VGG	BL	50.18 ±5.34	52.35 ±4.37	45.25 ±3.07	54.08 ±4.32	56.24 ±7.15	53.76 ±8.26	47.47 ±5.24	49.20 ±6.37	56.84 ±5.25	57.22 ±6.32	42.47 ±7.14	50.94 ±6.31
	BL∼M6	51.72 ±4.36	46.43 ±4.19	42.51 ±3.64	52.62 ±4.45	55.24 ±5.34	56.32 ±7.21	52.77 ±7.54	54.32 ±5.24	55.31 ±5.32	54.12 ±4.11	51.57 ±5.14	54.12 ±4.24
	BL∼M12	56.72 ±4.33	54.45 ±4.53	52.56 ±5.39	64.37 ±4.24	57.96 ±4.45	58.45 ±5.34	56.37 ±6.47	55.42 ±4.65	57.26 ±3.25	60.45 ±5.11	54.27 ±4.12	60.71 ±4.15
	BL∼M18	59.37 ±4.34	58.31 ±5.50	55.95 ±3.67	58.89 ±4.52	60.91 ±3.32	62.57 ±5.23	55.64 ±4.14	59.64 ±3.51	67.87 ±3.15	66.32 ±4.23	56.69 ±4.12	65.32 ±3.41
3D-ResNet	BL	45.44 ±5.84	46.32 ±4.35	44.35 ±6.78	48.01 ±5.37	53.14 ±6.25	52.72 ±6.35	48.43 ±7.67	52.82 ±5.26	53.44 ±6.21	51.32 ±5.12	45.42 ±7.13	51.22 ±7.32
	BL∼M6	49.17 ±4.31	48.80 ±5.37	46.58 ±4.92	47.41 ±5.32	55.54 ±6.54	54.53 ±7.21	46.88 ±5.72	55.37 ±6.23	52.34 ±5.24	52.43 ±7.82	44.88 ±4.52	50.42 ±6.43
	BL∼M12	57.79 ±5.02	53.15 ±4.34	48.64 ±6.17	51.37 ±4.87	59.49 ±4.26	56.74 ±5.20	51.66 ±4.37	60.15 ±4.28	56.28 ±4.32	52.26 ±5.21	51.22 ±4.28	59.76 ±3.21
	BL∼M18*	59.20 ±4.17	50.53 ±3.14	48.35 ±4.34	52.82 ±5.01	**63.42 ±4.51**	**62.33 ±5.42**	**59.71 ±3.72**	**65.21 ±4.29**	63.48 ±4.32	62.37 ±3.43	56.12 ±3.42	52.33 ±4.23
3D-Densenet	BL	52.34 ±6.45	49.73 ±4.19	47.56±5.24	51.34 ±6.13	52.14 ±7.25	54.62 ±6.43	45.04 ±5.14	50.12 ±7.69	56.44 ±5.65	57.12 ±5.23	47.43 ±4.14	54.32 ±6.25
	BL∼M6	50.36 ±6.57	48.58 ±5.50	46.69 ±6.48	50.37 ±5.53	52.24 ±6.15	55.99 ±5.64	48.27 ±6.77	52.43 ±5.44	56.11 ±6.45	56.55 ±7.23	53.47 ±5.15	55.36 ±5.56
	BL∼M12	54.67 ±4.25	52.47 ±5.21	49.87 ±4.72	56.73 ±5.38	55.72 ±5.48	53.81 ±4.60	47.97 ±5.08	50.84 ±5.66	57.26 ±5.32	56.76 ±5.13	55.77 ±6.13	64.43 ±4.45
	BL∼M18*	53.36 ±4.43	52.35 ±4.36	50.24 ±5.42	54.86±4.63	57.29 ±6.24	56.32 ±5.34	53.36 ±4.62	57.32 ±4.26	**67.74 ±3.75**	**68.12 ±3.93**	**62.42 ±4.21**	**68.24 ±4.62**
3D-InceptionNet	BL	53.38 ±5.44	47.96 ±5.37	44.37 ±6.08	51.51 ±4.83	47.64 ±7.35	54.83 ±8.42	45.33 ±6.24	53.49 ±7.21	49.14 ±2.15	49.12 ±2.41	45.32 ±2.34	52.49 ±3.32
	BL∼M6	53.51 ±4.38	52.31 ±5.37	46.57 ±4.34	53.45 ±4.43	55.24 ±5.12	52.31 ±6.32	44.36 ±4.31	54.66 ±6.28	54.27 ±2.42	54.31 ±3.22	47.22 ±3.71	55.16 ±2.18
	BL∼M12	54.18 ±4.58	52.86 ±5.43	53.83 ±4.52	55.61 ±5.18	60.12 ±2.42	59.54 ±3.14	54.42 ±2.71	56.12 ±2.14	59.42 ±3.16	57.44 ±3.24	55.42 ±2.67	57.24 ±2.14
	BL∼M18	63.35 ±4.54	60.47 ±5.37	58.57 ±5.53	61.47 ±4.37	62.21 ±2.31	60.12 ±2.62	56.23 ±3.41	56.25 ±2.35	57.12 ±2.72	59.42 ±2.32	56.31 ±2.11	59.12 ±2.12
3D-EfficientNet	BL	56.48 ±6.43	57.38 ±5.42	49.60 ±6.47	57.67 ±5.48	53.44 ±7.62	52.32 ±8.24	47.63 ±5.32	52.13 ±6.32	51.32 ±5.43	50.12 ±6.23	47.24 ±7.23	53.32 ±7.17
	BL∼M6	58.77 ±5.10	58.66 ±5.48	52.29 ±6.36	58.71 ±5.34	55.27 ±4.35	54.32 ±5.35	48.67 ±5.22	54.42 ±5.34	53.43 ±5.43	50.12 ±6.21	45.31 ±5.34	58.14 ±4.71
	BL∼M12	63.54 ±5.53	59.71 ±6.36	54.78 ±5.07	63.58 ±4.09	56.77 ±5.12	66.15 ±6.22	52.34 ±5.21	57.57 ±4.13	54.71 ±4.12	53.42 ±5.41	49.21 ±6.45	55.72 ±5.72
	BL∼M18*	**65.72 ±4.39**	**62.89 ±5.72**	**58.34 ±4.52**	**64.73 ±3.71**	62.37 ±4.35	60.74 ±5.19	58.42 ±5.04	63.37 ±4.65	65.12 ±5.25	57.52 ±7.24	54.33 ±4.26	58.37 ±5.42

T-S: Longitudinal Timesteps, Bold text indicates the highest achieved Accuracy for the corresponding 3D models and MRI planes.

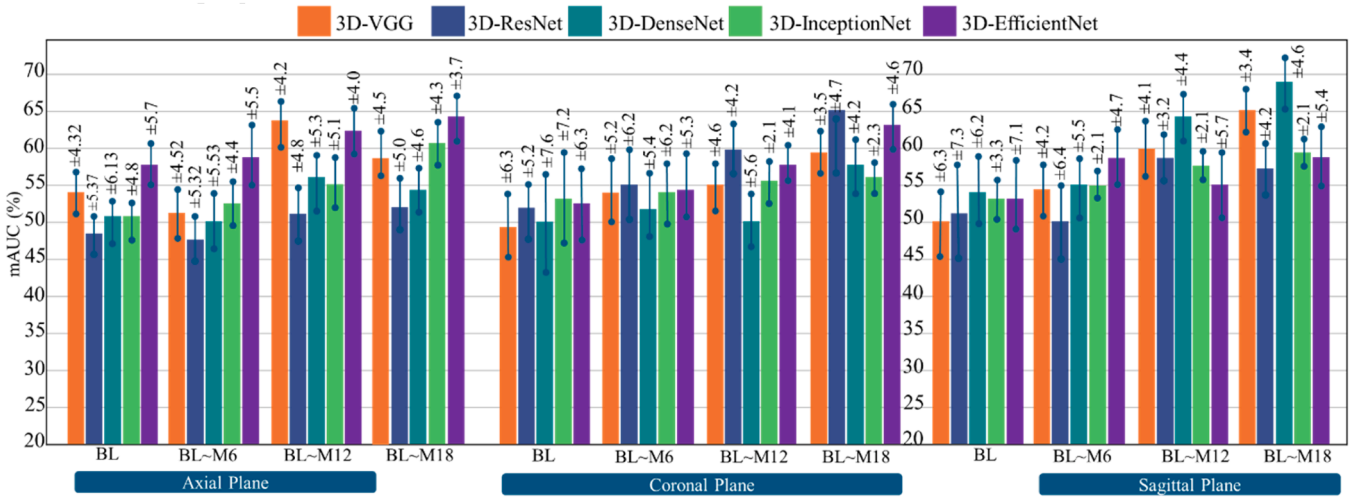


Fig. 6. Evaluating the effectiveness of different MRI planes at a longitudinal timestep.

65 %. Meanwhile, 3D-VGG and 3D DenseNet attain 57 % and 59 % mAUC respectively at the same timestep. Notably, 3D-InceptionNet and 3DEfficientNet demonstrate the most stable results throughout the training process with longitudinal time steps, consistently showcasing improved mAUC with each new addition of longitudinal time steps. In terms of the sagittal plane, 3D-DenseNet outperforms all other comparative models by reporting an mAUC score of 54 % at BL. Notably, it shows significant improvement in subsequent time steps of the training data. Similarly, 3D-VGG and 3D-InceptionNet also exhibit stable and improved performance in achieved accuracies as the longitudinal steps increase.

In conclusion, **Exp. 1** demonstrates that most of the 3D-CNN models report improvements in both aspects, i.e., achieved performance and stability as the longitudinal timesteps increase. This can be attributed to the utilization of the 3D volumetric design, which captures spatial as well as temporal features simultaneously. Building upon these findings, **Exp. 2** will explore the impact of fusing information from multiple planes in a longitudinal manner on the detection of AD progression.

3.2. Experiment 2: optimization of 3D-CNN using multi-plane from the longitudinal MRI

In **Exp. 2**, we explore the potential improvement in the overall performance of AD progression detection by fusing information from multiple MRI planes during the training process. Table 4 reports a comparative analysis on the performance of each model using different plane configurations i.e., single plane (axial), two planes (axial + coronal), and three planes (axial + coronal + sagittal). This configuration allows the models to capture significant patterns of disease progression within a shared learning environment. Subsequently, the convolutional output from each plane is combined and classified using a dense neural network. The dense network incorporates contextual information from multiple MRI planes to enhance the disease identification process.

From the results reported in Table 4, we observed that increasing the number of longitudinal time steps and the fusion of multiplane MRI data significantly increase the overall performance of the models across all evaluation metrices. This improvement is attributed to the capabilities of 3D data processing models in capturing progressive structural changes in the brain which is a hallmark to AD. Moreover, incorporating data from multiple MRI planes further improves classification performance, especially when all three planes (axial, coronal, sagittal) are combined in model's training. This multiplanar fusion strategy leverages complementary anatomical perspectives, enabling the models to detect subtle pathological features that may not be apparent in a single plane. For instance, 3D-InceptionNet demonstrated robust and consistent improvements in the learning curve. From BL to BL~M18, mAcc increased from 55.20 % to 67.27 %, and mAUC from 60.90 % to 70.70 % using all three-plane input. 3D-EfficientNet also exhibited strong performance, with its mAcc rising from 54.20 % to 69.20 % (all three plane), and mAUC from 64.30 % to 71.70 %. The model showed early stability, and its performance steadily improved as both timesteps and imaging views increased. 3D-ResNet showed similar performance trajectories. 3D-VGG performed the best with the combination of two planes (axial + coronal), achieving peak performance at BL~M18 with mAcc 65.54 % and mAUC 64.87 %. However, the performance slightly decreased at BL~M18, when the third plane was added, suggesting potential overfitting or architectural limitations in handling high-dimensional input.

The improvements shown by most of the comparative models using three planes suggest that the fusion of multiplane MRI data and increasing the number of longitudinal timesteps provides a more nuanced representation of the brain's structural changes. The progression from single time point (BL) to multiple time points (BL~M18) allowed models to capture disease trajectories more accurately. Simultaneously, the fusion of two or three MRI planes enriched the spatial context, aiding in more comprehensive pattern recognition.

In Fig. 7, the mAUC comparison of the 3D models trained on single and multiple planes of the longitudinal 3D MRI is presented. By comparing mAUC scores at different time steps, the performance achieved using a combination of all three planes outperformed the performance achieved using single-plane and two-plane models.

For instance, at BL, the mAUC scores for 3D-VGG, 3D-ResNet, and 3D-EfficientNet using three planes were reported as 63 %, 63 %, and 66 %, respectively. These scores represent improvements of (9 %, 11 %), (15 %, 12 %), and (10 %, 7 %) compared to the mAUC scores achieved using single-plane and two-plane combinations. Similarly, at the timesteps BL~M6 and BL~M12, the same models achieve significant improvements in mAUC scores when using three planes compared to single-plane and two-plane setups. Finally, at the time step BL~M18, using three planes yielded the highest mAUC score of 71 % with the 3D-EfficientNet model, outperforming all other combinations and time steps.

By analyzing the overall performance of the trained 3D models in **Exp. 2**, it was concluded that each model demonstrated significant improvements in both accuracy and stability when using all three planes along with the highest number of longitudinal time steps in the training data.

Table 4
Performance comparison of various 3D-CNN with multiple MRI planes from longitudinal data.

3D Model	T-S	Axial				Axial + Coronal				Axial + Coronal + Sagittal			
		mAcc	mSen	mSpe	mAUC	mAcc	mSen	mSpe	mAUC	mAcc	mSen	mSpe	mAUC
3D-VGG	BL	50.18 ±5.34	52.35 ±4.37	45.25 ±3.07	54.08 ±4.32	52.43 ±5.12	51.12 ±6.33	46.14 ±5.16	52.12 ±5.51	58.23 ±6.11	61.51 ±7.09	51.02 ±5.31	63.82 ±7.32
	BL~M6	51.72 ±4.36	46.43 ±4.19	42.51 ±3.64	52.62 ±4.45	57.45 ±4.46	49.33 ±3.56	46.67 ±5.53	55.36 ±4.72	61.31 ±5.21	63.21 ±6.16	52.68±5.30	60.40 ±6.13
	BL~M12	56.72 ±4.33	54.45 ±4.53	52.56 ±5.39	64.37 ±4.24	60.53 ±5.36	57.42 ±4.51	48.46 ±6.23	61.12 ±5.32	64.21 ±5.25	64.52 ±4.23	55.61 ±5.11	62.61 ±4.71
	BL~M18	59.37 ±4.34	58.31 ±5.50	55.95 ±3.67	58.89 ±4.52	**65.54 ±6.65**	**64.71 ±5.31**	**56.78 ±6.13**	**64.87 ±6.41**	62.22 ±5.41	63.54 ±5.10	53.42 ±4.22	60.62 ±5.13
3D-ResNet	BL	45.44 ±5.84	46.32 ±4.35	44.35 ±6.78	48.01 ±5.37	51.13 ±5.43	53.35 ±6.72	49.17 ±4.81	51.26 ±5.37	59.10 ±6.51	60.40 ±7.92	51.40 ±6.32	63.80 ±5.86
	BL~M6	49.17 ±4.31	48.80 ±5.37	46.58 ±4.92	47.41 ±5.32	54.66 ±5.11	51.25 ±6.42	53.12 ±4.42	52.13 ±5.33	60.40 ±7.62	61.50 ±6.12	53.10 ±6.84	62.40 ±5.22
	BL~M12	57.79 ±5.02	53.15 ±4.34	48.64 ±6.17	51.37 ±4.87	56.57 ±4.55	55.67 ±5.34	49.57 ±5.67	58.64 ±5.39	63.68 ±5.38	59.49 ±5.08	54.20 ±6.74	63.54 ±6.37
	BL~M18	59.20 ±4.17	50.53 ±3.14	48.35 ±4.34	52.82 ±5.01	61.13 ±4.24	62.74 ±5.01	55.01 ±5.73	59.31 ±4.23	**65.30 ±4.15**	**67.20 ±5.64**	**63.20 ±5.92**	**66.50 ±4.51**
3D-DenseNet	BL	52.34 ±6.45	49.73 ±4.19	47.56±5.24	51.34 ±6.13	55.34 ±5.15	59.42 ±6.63	46.74 ±6.13	54.32 ±5.25	62.38 ±6.48	60.46 ±6.28	57.47 ±5.93	61.56 ±4.63
	BL~M6	50.36 ±6.57	48.58 ±5.50	46.69 ±6.48	50.37 ±5.53	58.64 ±5.55	55.75 ±4.31	49.27 ±5.34	54.66 ±4.43	63.29 ±5.83	60.20 ±6.50	54.50 ±4.04	66.10 ±5.12
	BL~M12	54.67 ±4.25	52.47 ±5.21	49.87 ±4.72	56.73 ±5.38	56.43 ±5.57	54.67 ±5.46	51.73 ±3.87	55.94 ±4.31	62.34 ±4.01	61.20 ±3.21	52.50 ±4.32	64.60 ±5.72
	BL~M18	53.36 ±4.43	52.35 ±4.36	50.24 ±5.42	54.86±4.63	59.61±4.51	57.62 ±5.36	55.46 ±4.83	61.73 ±3.29	**65.38 ±5.65**	**62.34 ±5.21**	**58.98 ±5.76**	**65.73 ±4.53**
3D-InceptionNet	BL	53.38 ±5.44	47.96 ±5.37	44.37 ±6.08	51.51 ±4.83	51.32 ±5.54	50.12 ±4.35	45.40 ±5.24	49.32 ±5.31	55.20 ±4.10	60.50 ±5.39	52.30±5.79	60.90 ±5.76
	BL~M6	53.51 ±4.38	52.31 ±5.37	46.57 ±4.34	53.45 ±4.43	52.76 ±4.47	55.67 ±4.42	49.71 ±4.64	54.48 ±4.37	61.81 ±4.85	62.23 ± 5.48	52.55 ±6.46	63.70±5.34
	BL~M12	54.18 ±4.58	52.86 ±5.43	53.83 ±4.52	55.61 ±5.18	57.16 ±5.13	55.42 ±5.17	51.63 ±6.43	51.77 ±3.03	62.74 ±4.10	63.43 ±6.27	54.09 ±5.76	61.73 ±4.67
	BL~M18	63.35 ±4.54	60.47 ±5.37	58.57 ±5.53	61.47 ±4.37	64.56 ±4.57	62.85 ±5.38	59.76 ±4.73	63.75 ±5.65	**67.27 ±5.63**	**69.20 ±3.59**	**65.43 ±4.38**	**70.70 ±4.12**
3D-EfficientNet	BL	56.48 ±6.43	57.38 ±5.42	49.60 ±6.47	57.67 ±5.48	57.66 ±5.53	56.65 ±4.57	51.36 ±5.41	58.46 ±4.59	54.20 ±5.99	59.50 ±6.27	54.50±5.01	64.30±5.28
	BL~M6	58.77 ±5.10	58.66 ±5.48	52.29 ±6.36	58.71 ±5.34	59.47 ±3.58	58.62 ±5.45	56.55 ±4.92	60.44 ±5.45	62.20±5.17	61.20±4.39	51.50±6.37	63.60±4.64
	BL~M12	63.54 ±5.53	59.71 ±6.36	54.78 ±5.07	63.58 ±4.09	64.82 ±3.24	61.27 ±5.11	58.23 ±5.34	64.31 ±4.68	62.90 ±6.75	60.20±5.22	53.20±3.37	65.40±5.37
	BL~M18	**65.72 ±4.39**	**62.89 ±5.72**	**58.34 ±4.52**	**64.73 ±3.71**	67.64 ±3.98	65.71 ±4.57	61.19 ±4.64	68.39 ±4.80	**69.20 ±3.13**	**67.20±5.73**	**66.40±4.12**	**71.70±4.17**

T-S: Longitudinal Timesteps.

3.3. Experiment 3: AD progression detection using 3D-CNN-BiLSTM

Building upon the performance achieved in **Exp. 2** using all three planes of the MRI, we further enhanced the architectural complexity of the model while continuing to incorporate all three planes in a longitudinal analysis simultaneously. 3D-CNNs are commonly employed for the analysis of spatio-temporal data, such as 3D volumetric data. These networks are specifically designed to capture both spatial and temporal information by utilizing convolutional and pooling operations. On the other hand, BiLSTM networks excel at modeling long-term dependencies and sequential patterns in data. In **Exp. 3**, we combine the strengths of both 3D-CNNs and LSTMs to enhance the architectural complexity of the existing AD progression detection models. We achieve this by incorporating an LSTM module into the 3D-CNN framework. This approach allows the CNN module to extract fine-grained details from all three planes of the MRI, capturing the spatiotemporal features generated. Together, this design enhances the model's ability to represent disease progression more comprehensively.

As shown in Table 5, the 3D-InceptionNet + LSTM achieves the highest overall performance at the BL~M18 timestep, with mAcc: 76.40 ±3.15, mSen: 75.27 ±5.64, mSpe: 74.26 ±3.52, and mAUC: 73.50 ±3.21. This superior performance demonstrates the model's ability to effectively integrate both spatial and temporal information from the longitudinal data, providing a comprehensive representation of disease progression. The 3D-ResNet + LSTM also performs competitively, achieving the second-highest performance at BL~M18 (mAcc: 76.30 ±2.15, mSen: 74.20 ±2.64, mSpe: 71.20 ±2.92, and mAUC: 75.50 ±1.51). Its performance improves steadily with longer longitudinal intervals, showcasing its ability to capture progressive patterns in the data. However, it demonstrates slightly less sensitivity and specificity compared to the 3D-InceptionNet + LSTM at BL~M18. 3D-EfficientNet + LSTM, and 3D-DenseNet + LSTM also presented notable improvements at BL~M18. Finally, 3D-VGG + LSTM reports the lowest performance compared to other models, with mAcc: 69.30 ±5.15, mSen: 67.20 ±4.64, mSpe: 66.20 ±3.92, and mAUC: 68.50 ±4.51.

Fig. 8 presents the comparison of mAUC scores for 3D-CNN-LSTM models trained on multiple planes extracted from longitudinal 3D MRI. The evaluation of mAUC scores at different time steps demonstrated that combining all three planes (coronal, axial, and sagittal) with an increased architectural complexity results in superior performance compared to the setups in the previous experiments. Comparing the achieved mAUC scores, at the BL and BL~M06, no significant improvement was observed with the new hybrid architecture. However, at the BL~M12, all five models exhibited a notable boost in mAUC scores. In particular, the 3D-InceptionNet + LSTM and 3D-EfficientNet + LSTM models achieved the highest mAUC scores, outperforming all other comparative models.

Moreover, at the BL~M18 timestep, the 3D-ResNet model reported 75 % mAUC score, reflecting a 7 % increase compared to the mAUC score at BL-M12. Additionally, the 3D-ResNet-LSTM, 3D-InceptionNet-LSTM, and 3D-Efficient-LSTM models also showcased significant improvements in mAUC scores. From the visual representation of Fig. 8, it becomes evident that the increased architectural complexity combined with the data complexity resulted in additional performance boosts and enhanced result stability, particularly with an increase in the longitudinal time steps of the data.

3.4. Experiment 4: AD progression detection using 3D-CNN + BiLSTM-ERMHA

In **Exp. 4**, we further increased the architectural complexity of the proposed framework by incorporating ERMHA module in the 3D-CNNN + BiLSTM. This modification enables the model to stay focused on the significant parts of the input sequences that are more informative in the disease identification process.

In Table 6, at the BL timestep, the 3D-EfficientNet + BiLSTM-

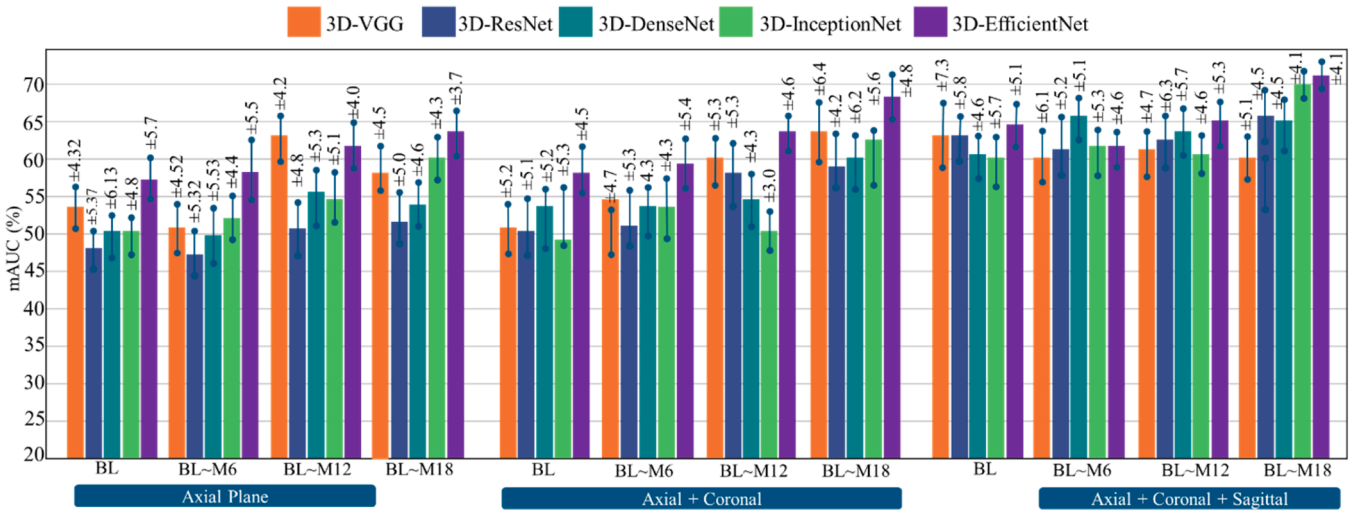


Fig. 7. Evaluating the effectiveness of combining multiple planes in AD progression detection.

Table 5
Performance comparison of various CNN models with multiplane MRI volumes using 3D-CNN + LSTM model.

Timeseries Model	Times Steps	mAcc (%)	mSen(%)	mSpe(%)	mAUC (%)
3D-VGG + LSTM	BL	62.35 ±4.58	61.40 ±5.92	57.40 ±5.32	64.80 ±4.86
	BL~M06	63.4 ±4.62	60.54 ±5.12	59.10 ±3.84	64.40 ±4.22
	BL~M12	67.20 ±3.64	64.50 ±4.44	62.20 ±5.74	67.30 ±4.46
	BL~M18	**69.30 ±5.15**	**67.20 ±4.64**	**66.20 ±3.92**	**68.50 ±4.51**
3D-ResNet + LSTM	BL	59.10 ±4.51	60.40 ±3.92	57.50 ±1.27	62.80 ±4.86
	BL~M06	63.40 ±3.62	61.50 ±3.12	58.10 ±4.84	63.40 ±4.34
	BL~M12	68.40 ±2.04	65.50 ±3.44	61.20 ±4.74	68.30 ±3.46
	BL~M18	**76.30 ±2.15**	**74.20 ±3.64**	**71.20 ±2.92**	**75.50 ±3.97**
3D-DenseNet + LSTM	BL	62.10 ±4.29	61.40 ±6.97	57.40 ±5.23	63.80 ±6.35
	BL~M06	64.40 ±3.63	62.50 ±5.12	56.10 ±5.82	63.40 ±4.52
	BL~M12	66.20 ±5.03	68.50 ±5.45	63.20 ±4.64	68.30 ±5.43
	BL~M18	**71.30 ±4.15**	**70.20 ±3.64**	**67.20 ±5.52**	**73.50 ±3.21**
3D-InceptionNet + LSTM	BL	63.10 ±6.29	62.40 ±5.97	55.40 ±7.23	64.80 ±5.99
	BL~M06	64.40 ±4.62	62.50 ±5.12	57.10 ±6.84	63.40 ±5.22
	BL~M12	73.20 ±5.04	70.20 ±4.45	65.82 ±4.74	74.30 ±4.46
	BL~M18	**76.40 ±3.15**	**75.27 ±5.64**	**74.26 ±3.52**	**73.50 ±3.55**
3D-EfficientNet + LSTM	BL	59.54 ±6.32	60.40 ±5.22	58.31 ±5.32	64.80 ±4.39
	BL~M06	66.40 ±4.62	62.50 ±5.12	57.10 ±5.84	64.40 ±4.17
	BL~M12	73.20 ±3.04	71.50 ±4.42	66.82 ±5.74	71.30 ±4.70
	BL~M18	**76.30 ±3.15**	**77.27 ±4.64**	**75.26 ±3.52**	**76.50 ±4.21**

Best-performing results across time steps are shown in bold.

ERMHA achieves the highest mAcc. (70.10 ±4.51 %) and mSen. (69.40 ±5.92 %), outperforming other comparative models such as 3D-DenseNet (64.11 ±4.33 %) and 3D-InceptionNet (66.20 ±4.17 %) in the same experiment as well as in all previous experiments. This highlights the superior spatial feature extraction capability of 3D-EfficientNet combined with BiLSTM-ERMHA, even at early stages. Other models, such as 3D-VGG and 3D-ResNet, fall behind, possibly due to their less efficient feature extraction capabilities at the same level. Performance improved across all models as more longitudinal data was incorporated. At BL~M06, 3D-EfficientNet and 3D-DenseNet demonstrates significant performance gain, achieving accuracies of 69.40 ±5.09 % and 67.10 ±5.52 %, respectively, indicating their strong temporal modeling capabilities. By BL~M18, 3D-EfficientNet outperforms all models, achieving the highest mAcc. (79.30 ±4.35 %), mSen. (80.27 ±2.27 %), mSpe. (77.26 ±3.12 %), and mAUC (81.50 ±3.19 %). 3D-DenseNet followed closely, with an mAcc of 79.30 ±3.15 %. In comparison, 3D-InceptionNet and 3D-ResNet also shows competitive performance, though they were slightly lower in stability (standard deviation) and specificity. 3D-VGG, despite improvements compared to previous experiments, consistently underperformed due to its simpler architecture. Overall, the 3D-EfficientNet + BiLSTM+ ERMHA emerged as the most robust model, excelling in leveraging spatial-temporal features for accurate AD progression detection.

Fig. 9 illustrates the comparison of mAUC scores achieved by the 3D-CNN + BiLSTM-ERMHA. By evaluating each model at different time steps, we observed consistent improvements in mAUC scores for all comparative models. As depicted in Fig. 9, it is evident that as the number of longitudinal timesteps increase, each comparative model exhibits enhanced stability and performance in the disease identification process. This observation aligns with the notion that capturing temporal dynamics is crucial for accurate disease classification. In particular, at the BL~M12 time step, we observed a significant boost in the overall performance of each model, indicating that the inclusion of additional temporal information along with attention mechanism led to better disease identification.

3.5. Experiment 5: homogeneous and heterogeneous deep ensemble models for AD progression detection using 3D-CNN + BiLSTM-ERMHA

In the previous experiments, we examined two critical aspects of the proposed AD progression detection framework. Firstly, we demonstrated that incorporating all three MRI planes (axial, coronal, and sagittal) in a longitudinal manner provided a more comprehensive representation of the data, significantly enhancing the model's ability to

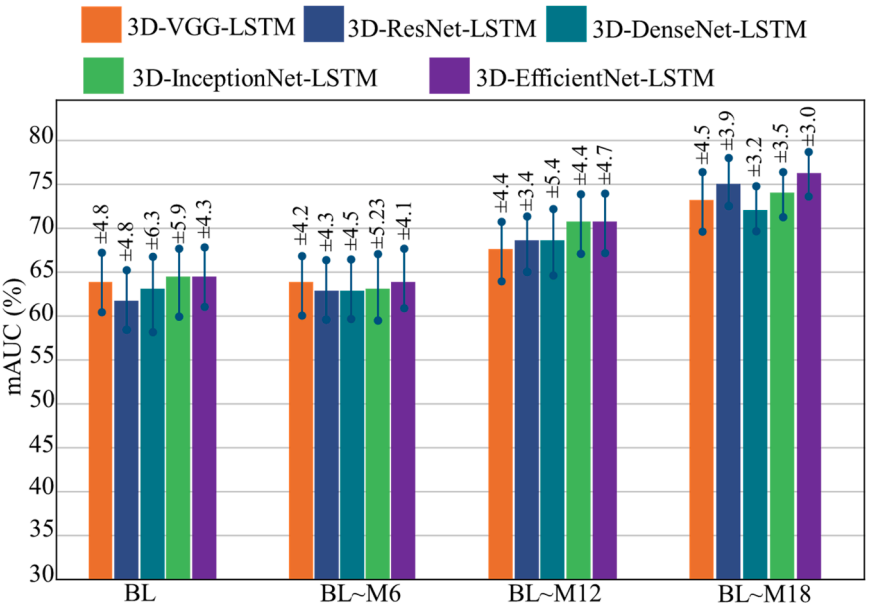


Fig. 8. mAUC comparison of various backbone 3D-CNN combined with LSTM for longitudinal MRI planes.

detect disease progression. This multi-plane approach ensured the integration of spatial information from diverse perspectives, capturing a wider range of anatomical details relevant to AD progression. Secondly, we observed that increasing the architectural complexity of the framework further amplified its performance. By leveraging more complex architectures, such as deeper and more interconnected neural networks, we were able to effectively capture the subtle and intricate patterns embedded in the training data. This capability enabled the model to automatically learn and recognize the progressive patterns of disease deterioration with greater precision. To address the computational demands of the enhanced model, we implemented a weight-sharing strategy. The trainable weights of the 3D-CNN were shared across all input planes, reducing redundancy and optimizing resource utilization.

In this experiment, we investigated the effectiveness of combining multiple 3D models for deep feature extraction through two EL based approaches: *homogeneous and heterogeneous*. In the homogeneous EL strategy, we trained separate branches of 3D-CNN feature extractors, each employing the same architectural design tailored for the axial, coronal, and sagittal planes. These individual models were independently trained on their respective planes to extract features specific to each plane. Subsequently, the output features from each plane were integrated and processed using BiLSTM-ERMHA to detect the progressive deterioration of brain tissues. Conversely, the heterogeneous EL method involved training distinct and independent 3D-CNN models, each with unique architectures, for the three planes. The output feature vectors from these diverse models were collectively processed through BiLSTM-ERMHA to analyze the progression of AD. Table 7 presents a performance comparison between the homogeneous and heterogeneous EL methods. In certain cases, the homogeneous ensemble network outperformed the shared backbone weight-sharing strategy employed in prior experiments. For instance, models based on 3D-DenseNet and 3D-EfficientNet based feature extractors showed significant performance improvements when used within the homogeneous ensemble setup, surpassing previously reported results. Similarly, the 3D-InceptionNet model achieved a slight improvement at the BL compared to prior experiments; however, no notable improvement was observed a later time step in the homogeneous setup. On the other hand, the 3D-VGG model failed to show any performance improvements in the homogeneous EL setup.

In the case of heterogeneous EL setup for the 3D-CNN based feature extractor, the selection of 3D-CNN backbone was based on the individual performance achieved by each model with their corresponding plane in **Exp. 1**. For instance, as demonstrated in Table 3, 3D-efficientNet exhibited the best performance with the axial plane, 3D-DenseNet excelled with the sagittal plane, and 3D-ResNet demonstrated optimal performance with the coronal plane. These networks were subsequently combined in a heterogeneous EL manner, followed by BiLSTM-ERMHA module. The reported results surpassed all other data/architectural based setup and achieved performance in previous experiments. Although the reported results from this setup at the BL were not impressive; however, the model demonstrated increased stability and remarkable improvements in achieved accuracies as the number of longitudinal time steps increased, ultimately surpassing the performance of all other models. At BL~M18, the framework exhibited remarkable performance, reaching at an mAcc of 93.73 ±2.39, mSen of 91.72 ±2.97, mSpe of 90.36 ±2.37, and a mAUC of 91.58 ±2.27.

Across all ensemble configurations and longitudinal time steps, the 95 % confidence intervals (CI) steadily narrow as more follow-up data are incorporated, reflecting growing precision in the model's prediction. At BL, CI spans roughly 6~8 points for accuracy but by BL~M18, the range shrinks down to 3~5 points indicating that predictions became more stable and less variable over time. This progressive tightening of the confidence interval values holds for sensitivity, specificity, and AUC as well, underscoring that late-stage models not only perform better on average but also do so with greater statistical certainty. The notably narrower CI achieved by the heterogeneous EL model at BL~M18 highlights its robustness, suggesting that its superior mean metrics are unlikely to be due to chance and require formal statistical comparison against competing frameworks.

Fig. 10 presents a visual comparison of the performance between homogeneous and heterogeneous EL networks, using the mAUC score as the primary evaluation metric, as detailed in Table 7. In the homogeneous EL approach, only a limited number of models demonstrated performance improvements. On the other hand, the heterogeneous EL approach consistently outperformed all model combinations. This finding underscores the significant advantage of incorporating diverse networks within a heterogeneous ensemble, which facilitates the integration of complementary features and patterns, ultimately enhancing the accuracy of disease identification. At the BL~M18 time step, the heterogeneous EL network achieved an impressive mAUC score of 91.5 %. This outstanding performance highlights the effectiveness of the heterogeneous deep ensemble model in capturing complex temporal

Table 6
Performance comparison of various 3D-CNN models with multiplane MRI using 3D-CNN + BiLSTM-ERMHA.

Timeseries Model	Times Steps	mAcc (%)	mSen (%)	mSpe (%)	mAUC (%)
3D-VGG + BiLSTM-ERMHA	BL	64.11 ±3.33	66.40 ±4.22	59.31 ±4.32	65.80 ±5.67
	BL~M06	65.40 ±4.02	65.54 ±3.12	60.10 ±4.24	65.40 ±5.22
	BL~M12	73.20 ±5.64	71.50 ±5.44	68.20 ±3.98	72.30 ±4.40
	BL~M18	**74.30 ±3.17**	**73.20 ±4.17**	**70.20 ±3.46**	**74.50 ±3.17**
3D-ResNet + BiLSTM - ERMHA	BL	65.11 ±6.33	67.40 ±5.22	62.31 ±4.32	66.80 ±5.43
	BL~M06	69.40 ±3.11	68.50 ±4.49	62.10 ±4.12	68.30 ±5.17
	BL~M12	72.20 ±2.94	70.50 ±4.14	69.20 ±4.63	71.30 ±4.16
	BL~M18	**77.30 ±3.45**	**75.20 ±3.77**	**73.20 ±3.97**	**77.50 ±4.17**
3D-DenseNet + BiLSTM-ERMHA	BL	64.11 ±4.33	66.40 ±4.22	59.31 ±4.56	65.80 ±4.67
	BL~M06	67.10 ±5.52	66.50 ±5.71	59.10 ±5.74	67.40 ±5.22
	BL~M12	75.20 ±4.04	72.50 ±4.42	69.20 ±5.74	74.30 ±4.46
	BL~M18	**79.30 ±3.15**	**77.20 ±4.61**	**71.20 ±3.53**	**76.50 ±4.01**
3D-InceptionNet + BiLSTM-ERMHA	BL	66.20 ±4.17	64.90 ±5.77	58.92 ±5.17	66.40 ±5.52
	BL~M06	67.40 ±4.90	66.50 ±6.08	59.10 ±5.16	66.40 ±4.48
	BL~M12	74.20 ±4.17	71.50 ±5.42	67.82 ±4.36	77.30 ±5.42
	BL~M18	**77.30 ±3.14**	**75.27 ±2.28**	**76.26 ±3.52**	**74.50 ±3.21**
3D-EfficientNet +BiLSTM-ERMHA	BL	70.10 ±4.51	69.40 ±5.92	61.40 ±5.27	67.80 ±3.86
	BL~M06	69.40 ±5.09	67.50 ±4.11	63.10 ±5.49	68.40 ±4.71
	BL~M12	76.20 ±4.09	74.50 ±3.79	70.82 ±4.71	75.30 ±3.34
	BL~M18	**79.30 ±4.35**	**80.27 ±2.27**	**77.26 ±3.12**	**81.50 ±3.19**

Bold text: best Accuracy at a particular time step.

dynamics and leveraging diverse feature representations to improve progression detection performance. The inherent diversity of the heterogeneous EL approach enables the proposed network to address the limitations of individual models, resulting in a more robust and comprehensive analysis of AD progression.

3.6. Selecting the best combinations of training data with the best architectural design

Fig. 11 compares the best performing models from the list of experiments and illustrates the progressive improvement in the mAUC score as additional layers of architectural complexity are incorporated into the proposed AD progression detection framework. It also demonstrates the effect of using single versus multiple planes of 3D MRI data in the disease identification process. Notably, all results presented on Fig. 11 were obtained using four longitudinal time steps i.e., **BL~M18**. In addition, the figure compares only the three highest-performing models from each experiment.

In **Exp. 1**, we tested the individual contribution of different planes in a longitudinal 3D MRI for AD progression detection. A list of deep 3D-CNN models was evaluated on the proposed designed 3D volume described in **Section 3.1.2**. The best-performing model using only single MRI plane reported an mAUC score of 65.21 % using the 3D-ResNet model and a coronal plane. 3D-DenseNet reported 68.24 % using the sagittal plane, and 3D-EfficientNet achieved 64.73 % using the axial plane. **Exp. 2** explored the significance of combining information from multiple MRI planes. Each model was first evaluated on individual planes (e.g., axial, coronal, and sagittal), and then retrained on multi-plane combinations (e.g., axial+coronal, and axial+coronal+sagittal) by extracting deep features from each plane via intermediate layers and fusing them before the final classification layer. Results achieved with multiplane led to significant mAUC gains over single-plane models, confirming that multiplane integration improves the diagnostic capacity of the models.

In **Exp. 3**, we added an extra layer of architectural complexity into the proposed ML framework by integrating a BiLSTM module with a 3D-CNN in an end-to-end configuration, processing all three MRI planes in a longitudinal sequence. This addition enabled the model to capture spatiotemporal dependencies available in the training data, improving its ability to detect AD progression detection and lead the mAUC score to 76 %. Following this, **Exp. 4** added further enhancement in the architectural design of the proposed model and added ERMHA module alongside the BiLSTM module. This mechanism further refined deep features flowing through the intermediate layers of the proposed model and enhanced the model's discriminative abilities and resulted in a mAUC score of 81.5 %. Finally, in **Exp. 5** we investigated EL approaches by comparing homogeneous ensembles, which are formed by combining three instances of the same model with non-shared weights, with heterogeneous ensembles that integrate different model architectures. The first two bars in Exp5. correspond to the top two performing models in the homogeneous EL methods, while the third bar shows the heterogeneous EL results. The heterogeneous EL (EB-ERMHA), enhanced with the ERMHA module, achieved the highest performance, with an mAUC of 91.58 %. These results highlight the cumulative benefit of architectural refinement, multiplane integration, and ensemble learning, and underscore the importance of both aspects i.e., well-structured input representations and sophisticated network designs for effective AD progression detection.

The results presented so far demonstrate that the heterogeneous deep ensemble model outperform all other architectural designs on the ADNI dataset; however, establishing high classification accuracy alone is not sufficient for clinical deployment. For the proposed model to gain clinical acceptance and prove trustworthiness in real-world healthcare settings, two critical aspects must be rigorously evaluated. First, the model's ability to generalize beyond the training distribution must be verified through external validation on independent cohorts with different demographic characteristics and imaging protocols. Second, the decision-making process must be made interpretable to clinicians, ensuring that predictions are driven by neuroanatomically relevant features rather than spurious correlations. Consequently, in the following section, we assess the proposed framework's generalizability through cross-dataset evaluation and examine its explainability by visualizing the brain regions that most significantly influence its predictions.

4. Model generalizability and explainability

To assess the clinical viability of the proposed framework, we further evaluate two critical aspects of the proposed model i.e., models' generalizability and models' explainability. First, we validate the model's robustness on an independent set of cohorts to examine cross-dataset performance. Second, we employ visualization of the attentions maps to identify the neuroanatomical regions most influential in the model's decision-making process.

4.1. Evaluating model's generalizability with external validation

The proposed model was initially trained and fine-tuned using the

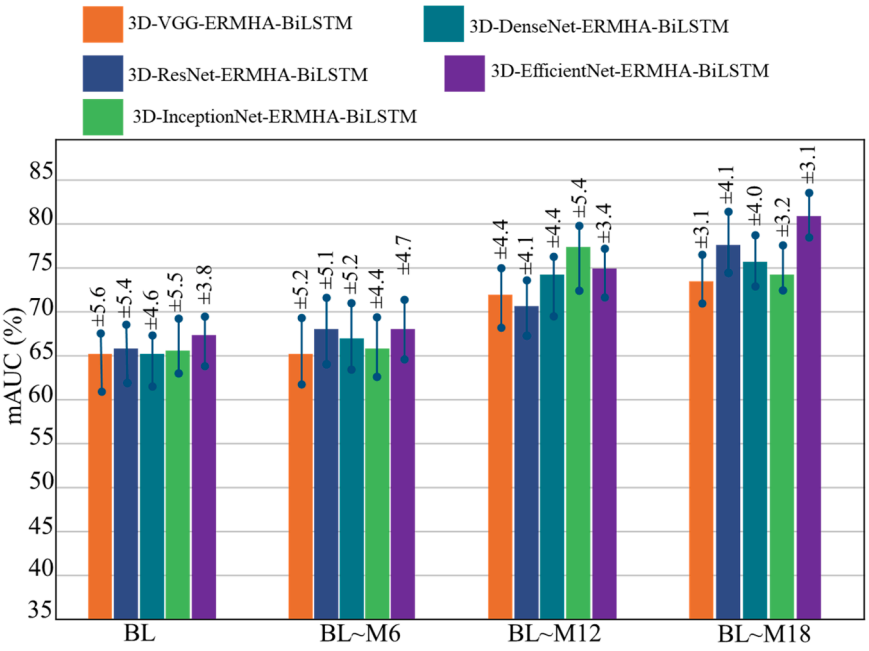


Fig. 9. mAUC comparison of various models incorporating multi-plane MRI fusion using 3D-CNN + BiLSTM-ERMHA.

real-world ADNI dataset. To assess its stability and capacity to generalize beyond the training data, we conducted an external validation with a separate dataset known as NACC, which included patients of diverse ethnic backgrounds from European Union (EU) countries. In this final set of experiments, we evaluated the proposed model's robustness and generalizability when applied to an entirely independent cohort of patients. The robustness of the proposed framework is illustrated in Table 8, which shows the results of training the model on one dataset and testing it on another, allowing for an unbiased evaluation of its performance in disease identification. To test this idea, the top-performing model from earlier experiments (Exp. 1–5), specifically the **3D-EfficientNet + 3D-DenseNet + 3D-ResNet-BiLSTM-ERMHA** configuration, was applied to test subjects from the NACC dataset. This dataset included 31 cognitively normal (CN) subjects and 29 CN subjects who later developed AD. *The relatively small sample size in the external validation was due to the limited number of NACC cohort patients meeting the same criteria as the ADNI training data*, as outlined in Section 2.1 **Dataset**. To track AD progression in the NACC cohort, the same preprocessing steps used for the ADNI dataset were applied to the NACC data. Table 8 presents a comparison of the ADNI-trained model's performance on both the ADNI and NACC test sets, demonstrating that the configuration selected for the ADNI data also performs effectively on the NACC test data.

Table 8 presents the performance metrics of the proposed model across different longitudinal time steps, comparing results when trained on the ADNI dataset and tested on both the NACC and ADNI datasets. At the BL timestep, when the model is trained on ADNI and tested on NACC, it achieves an mAcc of 58.65 ± 3.82 %, mSen of 58.69 ± 4.47 %, mSpe of 59.42 ± 5.34 %, and mAUC of 57.52 ± 5.78 %. In contrast, when tested on the ADNI dataset, the model performs better, with an mAcc of 67.41 ± 4.51 %, mSen of 65.40 ± 4.92 %, mSpe of 65.40 ± 4.27 %, and mAUC of 68.80 ± 4.86 %. With the inclusion of data from the BL to M06, performance improves for both test sets. For ADNI→NACC, the model reports an mAcc of 67.03 ± 3.27 %, mSen of 63.81 ± 4.54 %, mSpe of 61.68 ± 3.35 %, and mAUC of 68.78 ± 4.83 %. For ADNI→ADNI, the model reports improved performance, with exact values of 77.20 ± 3.04 %, 74.50 ± 4.42 %, 69.32 ± 3.14 %, and 76.13 for the same metrics, respectively. At the BL~M12, further improvements are observed. For ADNI→NACC, the model reports 77.67 ± 4.67, 69.72 ± 5.37, 64.97 ± 3.94, and 76.63 ± 4.73, while for ADNI→ADNI, the performance is notably higher, i.e., 86.30 ± 3.20 %, 81.17 ± 3.22 %, 78.66 ± 2.12 %, and 85.27 ± 4.19 % for the same metrics, respectively. Finally, the highest performance is observed at the BL~M18 time step. For ADNI→NACC, the model reaches an mAcc of 85.64 ± 2.39 %, mSen of 82.49 ± 3.76 %, mSpe of 79.71 ± 3.69 %, and mAUC of 86.37 ± 3.58 %. For ADNI→ADNI, the reported metrics are the highest, that is, 93.73 ± 2.39 %, 91.72 ± 2.97 %, 90.36 ± 2.37 %, and 91.58 ± 2.27 %, respectively. Overall, the results indicate that the model's performance improves as more longitudinal time steps are included in the training. However, the model consistently presented higher performance when tested on the ADNI dataset compared to the NACC dataset.

This discrepancy likely reflects the challenges of domain adaptation across different cohorts, resulting from variations in data distribution, patient demographics, imaging protocols, or disease progression patterns, which necessitate robust feature alignment and transfer learning techniques to mitigate covariate shifts and enhance model generalizability.

Fig. 12 presents a comparison of mAUC scores of the proposed model trained on the ADNI dataset and evaluated on the NACC test set. Although the NACC cohorts did not achieve the same performance levels as the ADNI subjects due to the distributional differences, the model still effectively captured temporal patterns from the longitudinal test data. The performance gap was mostly higher at the BL time step; however, stability improved progressively with the increase in longitudinal time steps (e.g., M06 and beyond). As shown in Fig. 12, the accuracy gap between the two datasets reduced over time. Notably, at the BL~M18, the proposed model demonstrated the smallest performance gap and superior modeling performance, underscoring its ability to generalize effectively across cohorts. It is important to note that the NACC test set comprises fewer subjects than the ADNI dataset, i.e., NACC = 60, and ADNI= 562). This is because, the ADNI cohort covers a wide demographic and clinical spectrum, including subjects at various stages of cognitive decline, which provides a more diverse and statistically robust sample for model training and evaluation. In contrast, the relatively small size of the NACC cohort increases susceptibility to statistical noise and variability, which can negatively impact model performance and generalizability.

Fig. 12 also highlights the 2D representative sample MRI slices from both cohorts, highlighting variations in image quality and anatomical presentation. The issue of domain adaptation has been a fundamental

Table 7
Performance comparison of homogeneous and heterogeneous EL models using multiplane longitudinal MRI.

Timeseries Model	TS	mAcc ± std [CI] (%)	mSen ± std [CI] (%)	mSpe ± std [CI] (%)	mAUC ± std [CI] (%)
(3D-VGG + 3D-VGG + 3D-VGG) – BiLSTM-ERMHA	BL	66.61 ± 5.43 [62.73–70.49]	65.40 ± 5.22 [61.67–69.13]	60.31 ± 4.32 [57.22–63.40]	67.80 ± 5.67 [63.74–71.86]
	BL~M06	69.40 ± 6.62 [65.66–73.14]	66.54 ± 4.12 [63.59–69.49]	61.16 ± 5.84 [56.98–65.34]	67.40 ± 6.22 [62.95–71.85]
	BL~M12	73.30 ± 6.37 [68.74–75.86]	71.50 ± 5.54 [67.54–75.46]	67.20 ± 6.74 [64.38–72.02]	72.30 ± 4.46 [69.11–75.49]
	BL~M18	**77.30 ± 5.35 [74.47–81.13]**	**75.20 ± 5.34 [71.38–79.02]**	**70.20 ± 4.92 [66.68–73.72]**	**76.50 ± 6.51 [71.84–81.16]**
(3D-ResNet + 3D-ResNet + 3D-ResNet) - BiLSTM-ERMHA	BL	65.11 ± 6.33 [60.58–69.64]	64.40 ± 5.22 [60.67–68.13]	59.31 ± 5.32 [55.50–63.12]	66.80 ± 5.67 [62.74–70.86]
	BL~M06	69.40 ± 6.62 [64.66–74.14]	68.50 ± 6.12 [64.12–72.88]	62.10 ± 7.84 [56.49–67.71]	68.40 ± 5.22 [64.67–72.13]
	BL~M12	72.20 ± 5.04 [68.59–75.81]	70.50 ± 7.44 [65.18–75.82]	68.20 ± 4.74 [64.81–71.59]	71.30 ± 5.46 [67.39–75.21]
	BL~M18	**76.30 ± 3.15 [74.05–78.55]**	**74.20 ± 4.64 [70.88–77.52]**	**69.20 ± 3.92 [66.40–72.00]**	**75.50 ± 3.51 [72.99–78.01]**
(3D- DenseNet + 3D-DenseNet + 3D - DenseNet) - BiLSTM - ERMHA	BL	65.42 ± 4.51 [62.19–68.65]	67.20 ± 4.97 [63.64–70.76]	61.52 ± 5.37 [57.68–65.36]	69.80 ± 5.07 [66.17–73.43]
	BL~M06	68.87 ± 5.43 [64.99–72.75]	65.96 ± 7.12 [60.87–71.05]	64.75 ± 2.14 [63.22–66.28]	66.76 ± 4.73 [63.38–70.14]
	BL~M12	75.92 ± 4.04 [73.03–78.81]	73.45 ± 4.67 [70.11–76.79]	68.77 ± 5.18 [65.06–72.48]	74.43 ± 4.56 [71.17–77.69]
	BL~M18	**80.42 ± 3.15 [78.17–82.67]**	**76.85 ± 4.66 [73.52–80.18]**	**75.20 ± 4.50 [71.98–78.42]**	**79.50 ± 3.21 [77.20–81.80]**
(3D-InceptionNet + 3D-InceptionNet + 3D-InceptionNet) - BiLSTM-ERMHA	BL	68.43 ± 5.11 [64.77–72.09]	65.76 ± 6.49 [61.12–70.40]	62.92 ± 4.24 [59.89–65.95]	67.30 ± 6.52 [62.64–71.96]
	BL~M06	65.62 ± 5.62 [61.60–69.64]	64.50 ± 4.12 [61.55–67.45]	63.21 ± 5.84 [59.03–67.39]	63.40 ± 5.22 [59.67–67.13]
	BL~M12	72.23 ± 4.73 [68.85–75.61]	70.50 ± 4.97 [66.94–74.06]	66.41 ± 5.64 [62.38–70.44]	70.30 ± 4.42 [67.14–73.46]
	BL~M18	**76.73 ± 4.02 [73.85–79.61]**	**73.48 ± 4.14 [70.52–76.44]**	**68.76 ± 4.91 [65.25–72.27]**	**73.50 ± 3.21 [71.20–75.80]**
(3D-EfficientNet + 3D-EfficientNet + 3D-EfficientNet) - BiLSTM - ERMHA	BL	67.10 ± 5.51 [63.16–71.04]	68.72 ± 4.92 [65.20–72.24]	63.40 ± 5.57 [59.42–67.38]	64.80 ± 5.34 [60.98–68.62]
	BL~M06	71.37 ± 4.37 [68.24–74.50]	69.45 ± 5.68 [65.39–73.51]	65.64 ± 4.89 [62.14–69.14]	67.35 ± 4.28 [64.29–70.41]
	BL~M12	81.32 ± 4.46 [78.13–84.51]	79.50 ± 5.79 [75.36–83.64]	78.82 ± 4.64 [75.50–82.14]	74.30 ± 5.68 [70.24–78.36]
	BL~M18	**85.43 ± 3.48 [82.94–87.92]**	**83.77 ± 4.43 [80.60–86.94]**	**79.72 ± 4.86 [76.24–83.20]**	**82.75 ± 3.73 [80.08–85.42]**
3D-Efficient + 3D-DenseNet + 3D -ResNet - BiLSTM-ERMHA	BL	67.41 ± 4.51 [64.18–70.64]	65.40 ± 4.92 [61.88–68.92]	65.40 ± 4.27 [62.35–68.45]	68.80 ± 4.86 [65.32–72.28]
	BL~M06	77.20 ± 3.04 [75.03–79.37]	74.50 ± 4.42 [71.34–77.66]	69.32 ± 3.14 [67.07–71.57]	76.13 ± 4.02 [73.25–79.01]
	BL~M12	86.30 ± 3.20 [84.01–88.59]	81.17 ± 3.22 [78.87–83.47]	78.66 ± 2.12 [77.14–80.18]	85.27 ± 4.19 [82.27–88.27]
	BL~M18	**93.73 ± 2.39 [92.02–95.44]**	**91.72 ± 2.97 [89.60–93.84]**	**90.36 ± 2.37 [88.66–92.06]**	**91.58 ± 2.27 [89.96–93.20]**

Bold text: Rest achieved accuracy, TS: Longitudinal time step, Std: Standard deviation, CI: Confidence interval.

challenge in deploying ML models across heterogeneous healthcare datasets. Despite applying consistent preprocessing steps to harmonize imaging data across both datasets, inherent inter-dataset differences persist. Which includes variations in MRI acquisition protocols, differences in biomarker availability, and clinical assessment tools between ADNI and NACC datasets. As a result, a model trained exclusively on ADNI data is expected to exhibit degraded performance when applied to NACC due to the domain shift and covariate distribution mismatch. This phenomenon reflects the challenges in cross-cohort generalization, where differences in feature distributions and measurement noise can lead to reduced predictive accuracy, emphasizing the need for advanced domain adaptation techniques or multi-source training strategies to improve robustness in real-world clinical applications.

4.2. Models' explainability

Following the explainability approach described in Section 2.6, we visualized attention maps across different diagnostic groups to identify neuroanatomical regions driving the model's predictions. Fig. 13 presents representative attention maps for CN, progressed to AD, and AD patients across axial, coronal, and sagittal views.

As shown in Fig. 13, attention maps from CN subjects exhibited diffuse, low-magnitude activations across all anatomical views, with minimal focal structure and only weak emphasis on central midline regions. These patterns reflect the absence of discriminative pathological features, consistent with preserved brain architecture. Irregular peripheral activations likely represent model uncertainty rather than meaningful anatomical signal. Progressive cognitively impaired patients showed distinct localization within medial temporal lobes [54], particularly hippocampal and entorhinal cortex implicated in early memory dysfunction [55]. Coronal and sagittal views showed additional engagement of the posterior cingulate cortex and precuneus region which are key nodes of the default mode network vulnerable to prodromal disruption [56]. These findings suggest the model captures selective regional vulnerability characteristic of early neurodegeneration. Finally, attention maps for AD patients demonstrated substantially

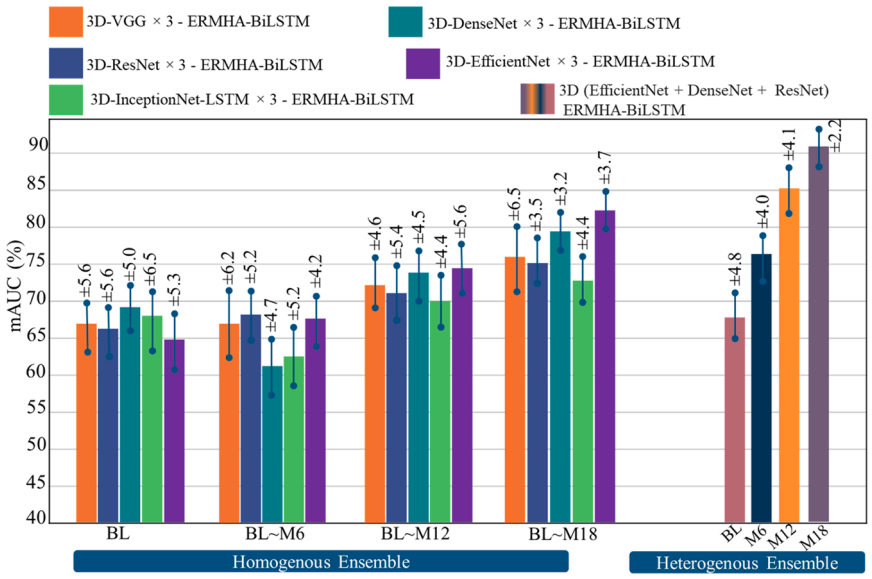


Fig. 10. mAUC comparison of homogenous and hetrogenous ensembling netwrok with longitudional MRI planes.

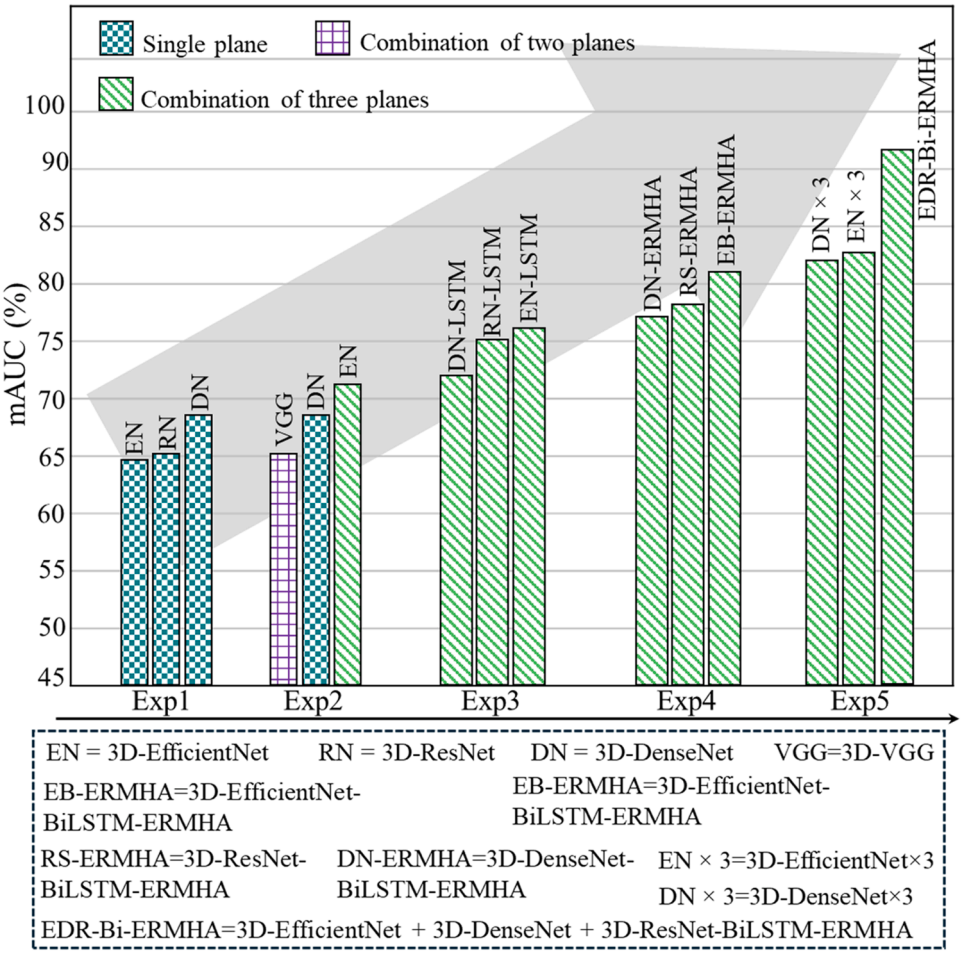


Fig. 11. The best AUC scores achieved at different experiments.

broader cortical involvement, extending from medial temporal regions to lateral temporo-parietal association areas [57]. Strong attention to posterior cingulate, parietal cortex, and precuneus reflected large-scale network breakdown, while dorsolateral frontal involvement in severe cases corresponded to executive dysfunction [58]. This distributed activation pattern indicates the model differentiates AD from earlier stages through widespread cortical degeneration rather than focal atrophy alone.

Table 8
Performance comparison of the proposed model on ADNI and NACC datasets for external validation.

Longitudinal time- step	Model: 3D-EfficientNet + 3D-DenseNet + 3D-ResNet-BiLSTM-ERMHA							
	ADNI→NACC (Train on ADNI, Test on NACC)				ADNI→ADNI (Train on ADNI, Test on ADNI)			
	mAcc.(%)	mSen.(%)	mSpe.(%)	mAUC(%)	mAcc.(%)	mSen.(%)	mSpe.(%)	mAUC(%)
BL	58.65 ±3.82	58.69 ±4.47	59.42 ±5.34	57.52 ±5.78	67.41 ±4.51	65.40 ±4.92	65.40 ±4.27	68.80 ±4.86
BL~M06	67.03 ±3.27	63.81 ±4.54	61.68 ±3.35	68.78 ±4.83	77.20 ±3.04	74.50 ±4.42	69.32 ±3.14	76.13 ±4.02
BL~M12	77.67 ±4.67	69.72 ±5.376	64.97 ±3.94	76.63 ±4.73	86.30 ±3.20	81.17 ±3.22	78.66 ±2.12	85.27 ±4.19
BL~M18	**85.64 ±2.39**	**82.49 ±3.76**	**79.71 ±3.69**	**86.37 ±3.58**	**93.73 ±2.39**	**91.72 ±2.97**	**90.36 ±2.37**	**91.58 ±2.27**

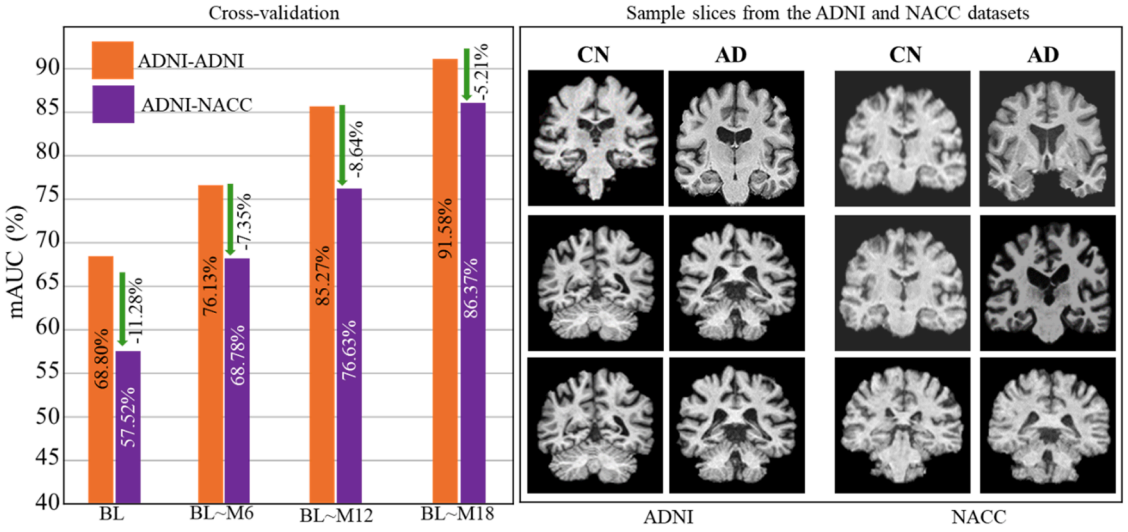


Fig. 12. Cross-validation results and representative 2D midbrain slice images from the ADNI and NACC datasets, showcasing both cognitively normal and progressed Alzheimer's disease subjects.

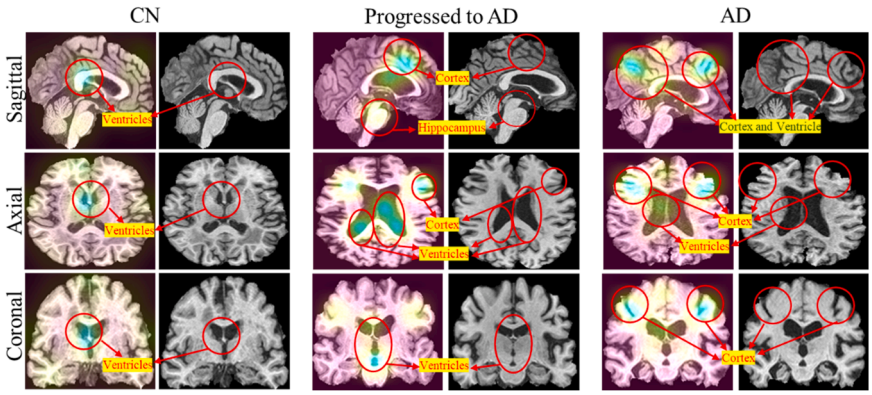


Fig. 13. Attention map visualization for CN, progressive cognitive impairment, and AD patients across axial, coronal, and sagittal views.

5. Performance comparison of the state-of-the-art methods

To ensure a thorough evaluation of our study, we assessed the most relevant studies reported in the literature between 2021 and 2025 and compared their results with the performance of our proposed framework. The comparison was done based on several performance metrics, including accuracy, sensitivity, specificity, and AUC. A notable limitation of these studies is the lack of publicly available implementation details, making it infeasible for us to reimplement the entire framework from scratch. As a result, we compared our framework's performance with the results published in these studies by following the same evaluation strategy used by the authors in the literature [10,59]. Additionally, finding studies using the exact same longitudinal time steps as our (i.e., four longitudinal time steps at six-month intervals) was also challenging. However, all comparative studies included in Table 9 share two critical commonalities: they utilize longitudinal MRI data sourced from the ADNI database, and their primary objective is to detect sequential patterns within this longitudinal data to identify the progression of cognitive impairment in AD patients. Moreover, to strengthen the validity of our evaluation, we conducted extensive ablation studies by implementing a wide range of DL models, as detailed in Table 3 through Table 8.

For instance, Chen et al. [65] introduced a hybrid CNN-Transformer model trained on multiple datasets (ADNI, OASIS, AIBL), yet its best performance was achieved with ADNI. Similarly, Mulyadi et al. [70] and Nallapu et al. [61] reported higher metrics on ADNI than on GARD and EXPEDITION3, respectively, reinforcing its superiority and validating its use in our study. Zheng et al. [64] proposed EffiSwin, using longitudinal

Table 9
Comparison of existing models with the proposed AD progression detection framework in terms of modality, longitudinal support, and performance.

Published Study	Year	DM	Longitudinal timesteps	Dataset	Performance (%)				Published framework	Task
					mAcc	mSen	mSpe	mAUC		
Ours*	**2025**	**MRI**	**Yes (4)**	**ADNI, NACC**	**93.7**	**91.7**	**90.36**	**91.58**	**3D-CNN-BiLSTM-ERMHA**	**AD progression detection**
Ozdemir et al. [60]	2025	MRI	Yes (6)	ADNI, NACC	90.0	-	-	90.5	DyEPAD	AD progression detection
Nallapu et al. [61]	2025	MRI	Yes (2)	ADNI	73.4	74.8	-	74.0	LDA	AD progression detection
Theodorou et al. [62]	2025	MRI	Yes (2)	ADNI	60.5	59.5	-	78.0	MRI2PET	AD classification
Mieling et al. [63]	2025	MRI	Yes (2)	ADNI	77.0	74.0	88.0	-	XGBoost	AD progression
Zheng et al [64].	2025	MRI	Yes (2)	ADNI	81.69	80.27	84.35	82.27	EffiSwin	AD progression detection
Chen et al. [65]	2024	MRI	Yes (2)	ADNI, OASIS, AIBL	93.4	-	-	93.3	Transformer- Deformable Attention	Classification (CN vs. AD)
Bapat et al. [66]	2024	MRI	Yes (3)	ADNI	83.5	89.4	99.2	91.9	3D-DenseNet-1DCNN	AD progression detection
Hao et al [67].	2024	MRI	Yes (4)	ADNI	91.08	89.81	92.04	90.78	Weighted Hypergraph CNN	Classification (CN vs. AD)
Inan et al. [68]	2024	MRI	Yes (1)	ADNI, OASIS	83.64	-	-	-	EfficientNetV2	Classification (CN vs. AD)
Hu et al. [14]	2023	MRI	Yes (3)	ADNI	77.2	79.97	71.59	81.53	VGG-TSwinformer	AD progression
Jahan et al. [69]	2023	MRI	Yes (2)	ADNI	73.0	-	-	-	EfficientNet-B7	Classification (CN vs. AD)
Mulyadi et al. [70]	2023	MRI	Yes (2)	ADNI - GARD	91.83	-	-	94.92	XADLiME Framework	Classification CN vs. AD
Mofrad et al. [26]	2022	MRI	Yes (3)	ADNI	76.10	-	-	-	Ensemble of CNN models	Classification (CN vs. AD)
Hazarika et al. [52]	2022	MRI	Yes (1)	ADNI	90.0	90.60	-	-	Improved DenseNet-121	Classification (CN vs. AD)
Aghai et al. [48]	2022	MRI	Yes (1)	ADNI	87.0	82.0	92.0	94.0	CNN + TL	AD Detection
Ocasio et al. [49]	2021	MRI	Yes (2)	ADNI	79.3	-	-	-	CNN	AD Progression
Rojas et al. [50]	2021	MRI	Yes (1)	ADNI	86.0	86.0	86.0	91.0	Compressed Dense Net	AD Diagnosis
Mofrad et al. [53]	2021	MRI	Yes (2)	ADNI	69.0	60.0	-	-	LME + SVM	Progression (CN - AD)

DM: Data modalities, "– ": indicates that the study did not report this metric.

MRI scans to achieve 81.69 % accuracy and 82.27 % F1-score. Bapat et al. [66] used 3D MRI and cognitive scores in a two-stage CNN model, reaching only 83.5 % accuracy. Hao et al. [67] constructed hypergraphs at four timepoints and applied KNN with hypergraph convolution, but their best AUC was 90.78 %. Inan et al., [68] proposed an automated DL framework for AD diagnosis using structural MRI scans. Their proposed approach integrates a guided ML learning-based slice selection with a DL model combining EfficientNetV2S and dense learned features. Validated on ADNI data, the model achieved 83.64 % classification accuracy, outperforming many existing methods while reducing the need for manual intervention. However, the study showed heavy reliance on predefined slice selection, potential biases in dataset distribution, and the need for further validation on larger, more diverse populations to ensure generalizability. Hu et al., [14] proposed VGG-TSwinformer, a DL framework that combines CNN and Transformer for predicting the progression of MCI to AD using longitudinal MRI. The model extracts spatial features from sMRI images, employs a sliding-window attention mechanism for fine-grained feature fusion, and uses temporal attention to capture brain atrophy patterns over time. VGG-TSwinformer was validated on the ADNI data and reported 77.2 % accuracy, 79.97 % sensitivity, and 81.53 % AUC. Mofrad et al., [26] proposed a DL based method to represent one-dimensional longitudinal measurements as two-dimensional grayscale images, transforming longitudinal data classification problems into image classification tasks. This allows the application of image-based DL to longitudinal data, even in cases of imbalanced datasets or missing data. The method was evaluated on the task of predicting dementia from brain volume trajectories derived from longitudinal MRI data, classifying subjects with stable sMCI versus and converted to AD. Using an ensemble of CNNs trained on ADNI data, the model achieved competitive accuracy on an independent test set.

Due to space constraints, a detailed discussion of every published approach listed in Table 9 is not feasible. Instead, the table provides a structured comparison based on six critical characteristics including: the published framework, year of publication, training data, number of longitudinal time steps, number of performance metrics calculated, and classification task. Our analysis compares the proposed approach against 18 most recently published studies. Fourteen studies employed longitudinal MRI while four studies were based on the baseline MRI. Baseline MRI studies were included to ensure completeness of our study and to provide readers with a broader comparative context, as they represent approaches that remain relevant in clinical practice and highlight the advantages of longitudinal modeling over single timestep analysis. We believe this comparison provides a thorough and up-to-date context for evaluating our method. However, the results of our study show that the proposed framework, through its integration of deep learning, attention, and sequential modeling over four longitudinal timepoints, consistently outperforms prior state-of-the-art approaches across all key metrics. Its robust spatiotemporal learning makes it a strong candidate for scalable, and efficient clinical decision support in AD progression detection.

6. Computational complexity of the proposed models

Table 10 presents a comparative analysis of different 3D-CNN architectures integrated with a BiLSTM-ERMHA module for training and prediction. The key metrics for comparison include trainable parameters, training time, and prediction time. The computational complexity of the proposed model is divided into three categories: shared trainable weights (single CNN backbone with BiLSTM-ERMHA), homogeneous EL (multiple instances of the same CNN model for each plane), and heterogeneous EL (Combination of different CNN models). In the case of shared trainable weights for all three planes, 3D-EfficientNet has the fewest trainable parameters (8.03 M), leading to faster training (10.2 min) and a relatively low prediction time (0.17 sec). The 3D-DenseNet-

Table 10
Computational complexity of the proposed model across different training strategies and ensemble modes.

M	TS	Comparative models	Trainable parameters (in million)	Training time (in minutes)	Prediction time (in seconds)	mAUC
1	Shared trainable weights	3DEfficientNet - BiLSTM ERMHA	8.03	10.2	0.17	81.5 ±3.19
		3DDenseNet - BiLSTM ERMHA	9.04	15.12	0.19	76.5 ±4.01
		3DResNet - BiLSTM ERMHA	26.62	10.11	0.16	77.5 ±4.17
2	Homogenous EL	3DEfficientNet × 3 - BiLSTM ERMHA	24.09	30.5	0.23	82.75 ±3.73
		3DDenseNet × 3 -BiLSTM-ERMHA	25.43	38.78	0.24	79.5 ±3.21
		3DResNet × 3 BiLSTM - ERMHA	78.31	20.03	0.22	75.5 ±3.51
3	Hetrogenous EL	3D-EfficientNet + 3D-DenseNet + 3D-ResNet - BiLSTM-ERMHA	42.05	25.53	0.22	91.58 ±2.27

M:Mode, (1): Best-performing models with different MRI planes, TS:Training strategy.

based feature extraction module has slightly more parameters (9.04 M) and requires more training time (15.12 min), suggesting that it is computationally more intensive than 3D-EfficientNet. 3D-ResNet has the highest trainable parameters (26.62 M) in this category, yet its training time (10.11 min) is nearly the same as EfficientNet, indicating that its architecture efficiently utilizes hardware resources. In the case of homogeneous EL, where multiple instances of the same 3D model are repeated three times for each MRI input, the training time increases significantly compared to single-model versions due to the ensemble architecture. However, the prediction time remains nearly the same for all three models, with ResNet-based ensembles being slightly faster (0.22 sec). Finally, the heterogeneous EL setup offers moderate trainable parameters (42.05 M) and balanced prediction/inference time compared to shared trainable weights and homogeneous EL. The training time (25.53 min) is shorter than the homogeneous EL setup, demonstrating that combining different models improves efficiency. Furthermore, the prediction time (0.22 sec) is as fast as the best homogeneous ensemble (ResNet × 3), making it a strong choice for deployment.

In addition to achieving optimal performance, our proposed framework was designed with computational efficiency in mind to support potential clinical integration. Several architectural choices such as limiting the slice count to 16 per plane, using 3D-CNN backbones, and introducing an ERMHA module help maintain a reasonable memory footprint and inference time while preserving predictive performance. On our experimental hardware (NVIDIA TITAN X, 12 GB VRAM), the mean inference time per subject was 0.173 s for a single model and 0.22 s for the heterogeneous ensemble, which is within the range suitable for offline clinical assessment. Although this study did not experimentally evaluate mixed-precision inference or CPU-only execution, such optimizations are standard in modern AI pipelines and could be readily incorporated into the proposed system. For example, mixed-precision inference (FP16) and more lightweight backbone variants (e.g., 3D-MobileNet, quantized 3D-EfficientNet) are expected to further reduce resource requirements. Likewise, containerized deployment (e.g., via Docker) would facilitate integration into existing hospital picture archiving and communication system (PACS) workstations or cloud platforms, enabling execution on mid-range GPUs (8 GB VRAM) or high-performance CPUs with modest increases in processing time. These considerations indicate that the framework can be adapted for environments with varying computational resources, supporting its potential for real-world clinical use [71].

7. Research limitations and future directions

This study has several limitations and presents opportunities for future research. Firstly, although the proposed framework was initially trained and validated using ADNI dataset, external validation on the NACC cohort revealed a consistent performance gap, with higher results achieved on ADNI compared to NACC. While the framework demonstrated robustness when tested on two independent datasets, the limited sample size in the NACC cohort constrains the statistical power of the external validation. Expanding validation to larger and more diverse cohorts such as OASIS-3, AIBL, and MIRIAD would provide stronger evidence of generalizability and clinical utility. Our future work will explore domain adaptation techniques, such as adversarial training or style transfer approach, to minimize distributional shifts between cohorts and improve cross-dataset generalizability. Secondly, the proposed framework relied exclusively on neuroimaging data. Incorporating multimodal information, including genetic biomarkers, clinical records, and cognitive assessments, may capture complementary disease signatures and further strengthen predictive performance, particularly for early-stage AD progression. Thirdly, the proposed AD progression detection framework was developed retrospectively. To ensure its practical applicability, it requires evaluation in a prospective setting where patient outcomes are unknown. Fourthly, the feature extraction backbone was limited to 3D CNNs. Emerging architectures such as vision transformers or hybrid CNN-Transformer backbones could improve spatiotemporal representation learning. Exploring these alternatives represents a promising research direction. Fifth, while our computational analysis showed inference times suitable for offline clinical use, further work is needed to optimize deployment through mixed-precision inference, quantization, and lightweight backbone architectures such as 3D-MobileNet. Finally, explainability remains a critical aspect. Our future work will also focus on incorporating end-to-end XAI methods to provide transparent and clinician-friendly interpretations of the model's predictions.

8. Conclusion

AD is a neurodegenerative disease, and existing medical diagnostic systems often rely on baseline data obtained during the initial visit, neglecting the dynamic nature of clinical information. Most literature studies in the AD domain focus solely on a partial set of MRI (a single slice or a single plane), which overlooks a significant amount of valuable information contained within the entire 3D MRI volume. In this study, we addressed these limitations by considering multiple 2D middle slices from longitudinal MRI volumes across different time steps. This approach allows for capturing the most crucial slices that encompass the progressive anatomical changes occurring in the brain tissues over time. To analyze these complex relationships and temporal dynamics within the AD-affected brain tissues, we employed a two-step framework. First, we utilized a 3D-CNN to capture inter-slice relationships and extract spatial features. Subsequently, an attention-based BiLSTM-ERMHA was

used to model the temporal dependencies and highlight the most salient features relevant to AD progression detection. We evaluated the effectiveness of our proposed AD progression detection pipeline through various experiments. The framework exhibited promising performance across diverse conditions on the ADNI dataset and further demonstrated robust generalizability in external validation on the NACC cohort, supporting its effectiveness for longitudinal clinical assessment.

Declaration of competing interest

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

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (RS-2025-00554526, 2021R1A2C1011198), the Institute for Information & Communication Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) under the ICT Creative Consilience Program (IITP-2021-2020-0-01821), and the AI Platform to Fully Adapt and Reflect Privacy-Policy Changes (RS-2022-II220688).

References

[1] M.S. Rafii, P.S. Aisen, Detection and treatment of Alzheimer's disease in its preclinical stage, Nature Aging 3 (5) (2023) 520–531, https://doi.org/10.1038/s43587-023-00410-4.
[2] Y. Liu, et al., Diffusion tensor imaging and Tract-based Spatial statistics in Alzheimer's disease and mild cognitive impairment, Neurobiol. Aging 32 (9) (2011) 1558–1571, https://doi.org/10.1016/j.neurobiolaging.2009.10.006. Apr.
[3] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease, J. Neurosci. Methods 339 (2020) 108701, https://doi.org/10.1016/j.jneumeth.2020.108701. Apr.
[4] S. Gauthier, C. Webster, S. Servaes, J.A. Morais, P. Rosa-Neto, Alzheimer's Disease International: World Alzheimer Report 2022 – Life after diagnosis, Alzheimer's Dis. Int. (2022) 361–364.
[5] N. Rahim, N. Ahmad, W. Ullah, J. Bedi, Y. Jung, Early progression detection from MCI to AD using multi-view MRI for enhanced assisted living, Image Vis. Comput. 157 (March) (2025) 105491, https://doi.org/10.1016/j.imavis.2025.105491.
[6] C.Z.L. Ong, et al., Classification of cognitive syndromes in a Southeast Asian population: interpretable graph convolutional neural networks, Knowledge-Based Syst. 309 (2025) 112855, https://doi.org/10.1016/j.knosys.2024.112855. June 2024.
[7] J. Liu, D. Zeng, R. Guo, M. Lu, F.X. Wu, J. Wang, MMHGE: detecting mild cognitive impairment based on multi-atlas multi-view hybrid graph convolutional networks and ensemble learning, Cluster. Comput. 24 (1) (2021) 103–113, https://doi.org/10.1007/s10586-020-03199-8. Apr.
[8] A. Giovannetti, et al., Deep-MEG: spatiotemporal CNN features and multiband ensemble classification for predicting the early signs of Alzheimer's disease with magnetoencephalography, Neural Comput. Appl. 33 (21) (2021) 14651–14667, https://doi.org/10.1007/S00521-021-06105-4/TABLES/4. Apr.
[9] 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.
[10] 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. Apr.
[11] M.L.F. Jumaili, E. Sonuç, An attention-based CNN framework for Alzheimer's disease staging with multi-technique XAI visualization, Comput. Mater. Contin. 83 (2) (2025) 2947–2969, https://doi.org/10.32604/cmc.2025.062719.
[12] I. Nagarajan, G.G. Lakshmi Priya, A comprehensive review on early detection of Alzheimer's disease using various deep learning techniques, Front. Comput. Sci. 6 (2024), https://doi.org/10.3389/fcomp.2024.1404494.
[13] 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.
[14] 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.
[15] N. Mahendran, D.R.V.P. M, A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease, Comput. Biol. Med. 141 (2022) 105056, https://doi.org/10.1016/J.COMPBIOMED.2021.105056. Apr.
[16] S.H. Wang, et al., Single slice based detection for alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization, Multimed. Tools Appl. 77 (9) (2018) 10393–10417, https://doi.org/10.1007/s11042-016-4222-4.
[17] R. Mendoza-Léon, J. Puentes, L.F. Uriza, M.H. Hoyos, Single-slice Alzheimer's disease classification and disease regional analysis with Supervised Switching Autoencoders, Comput. Biol. Med. 116 (2019) 2020, https://doi.org/10.1016/j.compbiomed.2019.103527. October.
[18] 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. Apr.
[19] Y. Shmulev, M. Belyaev, A. Disease, N. Initiative, Y. Shmulev, and M. Belyaev, "Predicting conversion of mild cognitive impairments to Alzheimer's disease and exploring impact of neuroimaging," 2018.
[20] K. Aderghal, K. Afdel, J. Benois-Pineau, G. Catheline, Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities, Heliyon. 6 (12) (2020), https://doi.org/10.1016/j.heliyon.2020.e05652.
[21] B. Lei, et al., Longitudinal study of early mild cognitive impairment via similarity-constrained group learning and self-attention based SBi-LSTM, Knowledge-Based Syst 254 (2022), https://doi.org/10.1016/J.KNOSYS.2022.109466. Apr.
[22] L. Song, Q. Wang, H. Li, J. Fan, B. Hu, Longitudinal structural MRI data prediction in nondemented and demented older adults via generative adversarial convolutional network, Neural Process. Lett. 55 (2) (2023) 989–999, https://doi.org/10.1007/s11063-022-10922-6.
[23] S. Pathan and Y. Hong, "Predictive image regression for longitudinal studies with missing data".
[24] J.Y. Namgung, E. Noh, Y. Jang, M.J. Lee, B. Park, A robust multimodal brain MRI - based diagnostic model for migraine : validation across different migraine phases and longitudinal follow - up data, J. Headache Pain (2025), https://doi.org/10.1186/s10194-024-01946-5.
[25] 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, Comput. Methods Programs Biomed. 189 (2020) 105348, https://doi.org/10.1016/j.cmpb.2020.105348. Apr.
[26] S.A. Mofrad, H. Bartsch, A.S. Lundervold, S.A. Mofrad, A.S. Lundervold, From longitudinal measurements to image classiication: application to longitudinal MRI in Alzheimer's disease From longitudinal measurements to image classification: application to longitudinal MRI in Alzheimer's disease, Res. Sq. (2022) 0–9.
[27] S.H. Hojjati, A. Babajani-Feremi, Prediction and modeling of neuropsychological scores in Alzheimer's disease using multimodal neuroimaging data and artificial neural networks, Front. Comput. Neurosci. 15 (January) (2022) 1–16, https://doi.org/10.3389/fncom.2021.769982.
[28] N. Rahim, S. El-Sappagh, H. Rizk, O.A. El-serafy, T. Abuhmed, Information fusion-based Bayesian optimized heterogeneous deep ensemble model based on longitudinal neuroimaging data, Appl. Soft Comput. 162 (2024) 111749, https://doi.org/10.1016/j.asoc.2024.111749. December 2023.
[29] Y.F. Khan, B. Kaushik, C.L. Chowdhary, G. Srivastava, Ensemble model for diagnostic classification of Alzheimer's disease based on brain anatomical magnetic resonance imaging, Diagnostics 12 (12) (2022) 3193, https://doi.org/10.3390/diagnostics12123193. Dec.
[30] J. Liu, M. Li, W. Lan, F.X. Wu, Y. Pan, J. Wang, Classification of Alzheimer's disease using whole brain hierarchical network, IEEE/ACM Trans. Comput. Biol. Bioinforma. 15 (2) (2018) 624–632, https://doi.org/10.1109/TCBB.2016.2635144.
[31] C. Xu, et al., Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture, Med. Image Anal. 50 (2018) 82–94, https://doi.org/10.1016/J.MEDIA.2018.09.001. Dec.
[32] Y. Cai, et al., Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: A prediction model development and validation study, BMJ Open. 11 (1) (2021) 1–7, https://doi.org/10.1136/bmjopen-2020-041139.
[33] R. Hedayati, M. Khedmati, M. Taghipour-Gorjikolaie, Deep feature extraction method based on ensemble of convolutional auto encoders: application to Alzheimer's disease diagnosis, Biomed. Signal Process. Control 66 (2021) 102397, https://doi.org/10.1016/J.BSPC.2020.102397. Apr.
[34] G. Battineni, N. Chintalapudi, F. Amenta, E. Traini, A comprehensive machine-learning model applied to Magnetic resonance imaging (MRI) to predict Alzheimer's disease (AD) in older subjects, J. Clin. Med. 2020 9 (2020) 2146, https://doi.org/10.3390/JCM9072146. *Page*vol. 9, no. 7, p. 2146Jul.
[35] C. Platero, M.C. Tobar, Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers, Brain ImAging Behav. 15 (4) (2021) 1728–1738, https://doi.org/10.1007/s11682-020-00366-8.
[36] K. Li, W. Chan, R.S. Doody, J. Quinn, S. Luo, Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data, J. Alzheimer's Dis. 58 (2) (2017) 361–371, https://doi.org/10.3233/JAD-161201.
[37] B.K. Karaman, M.R. Sabuncu, Assessing the significance of longitudinal data in Alzheimer's Disease forecasting, no. stable MCI (2024) 1–14.
[38] X. Hua, et al., Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals, Neuroimage 51 (1) (2010) 63–75, https://doi.org/10.1016/j.neuroimage.2010.01.104.
[39] 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.

[40] J.A. Castro-Silva, M.N. Moreno-García, L. Guachi-Guachi, D.H. Peluffo-Ordóñez, Novel hippocampus-centered methodology for informative instance selection in Alzheimer's disease data, Heliyon. 10 (19) (2024) e37552, https://doi.org/10.1016/j.heliyon.2024.e37552.
[41] A. Alotaibi, Ensemble deep learning approaches in health care: A review, Comput. Mater. Contin. 82 (3) (2025) 3741–3771, https://doi.org/10.32604/cmc.2025.061998.
[42] D. Muller, I. Soto-Rey, F. Kramer, An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks, IEEe Access. 10 (2022) 66467–66480, https://doi.org/10.1109/ACCESS.2022.3182399.
[43] N. An, H. Ding, J. Yang, R. Au, T.F.A. Ang, Deep ensemble learning for Alzheimer's disease classification, J. Biomed. Inform. 105 (2020) 103411, https://doi.org/10.1016/j.jbi.2020.103411. Apr.
[44] Z. Ullah and J. Kim, "Hierarchical deep feature fusion and ensemble learning for enhanced brain tumor MRI classification," pp. 1–46, 2025.
[45] H. Alibrahim, S.A. Ludwig, Hyperparameter optimization: comparing genetic algorithm against grid search and bayesian optimization, 2021 IEEE Congr. Evol. Comput. CEC 2021 - Proc. (2021) 1551–1559, https://doi.org/10.1109/CEC45853.2021.9504761.
[46] I. Tenney, D. Das, E. Pavlick, BERT rediscovers the classical NLP pipeline, ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (2020) 4593–4601, https://doi.org/10.18653/v1/p19-1452.
[47] 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.
[48] A. Aghaei, M.E. 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.
[49] E. Ocasio, T.Q. Duong, Deep learning prediction of mild cognitive impairment conversion to Alzheimer's disease at 3 years after diagnosis using longitudinal and wholebrain 3D MRI, PeerJ Comput. Sci. 7 (2021) 1–21, https://doi.org/10.7717/PEERJ-CS.560.
[50] B. Solano-Rojas, R. Villalón-Fonseca, A low-cost three-dimensional densenet neural network for alzheimer's disease early discovery†, Sensors 21 (4) (2021) 1–17, https://doi.org/10.3390/s21041302. Feb.
[51] K. Gotkowski, C. Gonzalez, A. Bucher, A. Mukhopadhyay, M3d-CAM: a PyTorch library to generate 3D attention maps for medical deep learning. Informatikaktuell, Springer Vieweg, Wiesbaden, 2021, pp. 217–222, https://doi.org/10.1007/978-3-658-33198-6_52.
[52] R.A. Hazarika, D. Kandar, A.K. Maji, An experimental analysis of different Deep Learning based models for Alzheimer's Disease classification using brain Magnetic Resonance images, J. King Saud Univ. - Comput. Inf. Sci. 34 (10) (2022) 8576–8598, https://doi.org/10.1016/J.JKSUCI.2021.09.003. Apr.
[53] S.A. Mofrad, A. Lundervold, A.S. Lundervold, A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease, Comput. Med. Imaging Graph. 90 (2021) 101910, https://doi.org/10.1016/j.compmedimag.2021.101910. Jun.
[54] B.C. Dickerson, D.H. Salat, J.F. Bates, C.E. Stern, D. Blacker, and M.S. Albert, "Medial temporal lobe function and structure in mild cognitive impairment," vol. 56, no. 1, pp. 27–35, 2015, doi: 10.1002/ana.20163.Medial.
[55] C. Pennanen et al., "Hippocampus and entorhinal cortex in mild cognitive impairment and early AD," vol. 25, pp. 303–310, 2004, doi: 10.1016/S0197-4580(03)00084-8.
[56] M. Bailly, et al., Precuneus and cingulate cortex atrophy and hypometabolism in patients with Alzheimer's disease and mild cognitive impairment: MRI and 18F-FDG PET quantitative analysis using FreeSurfer, Biomed Res. Int. 2015 (2015), https://doi.org/10.1155/2015/583931.
[57] L. Chauveau et al., "Medial temporal lobe subregional atrophy in aging and Alzheimer ' s Disease : A longitudinal study," vol. d, no. October, pp. 1–15, 2021, doi: 10.3389/fnagi.2021.750154.
[58] T. Yokoi, H. Watanabe, H. Yamaguchi, and E. Bagarinao, "Involvement of the Precuneus /posterior cingulate cortex is significant for the development of Alzheimer ' s Disease : a PET (THK5351, PiB) and resting fMRI study," vol. 10, no. October, pp. 1–15, 2018, doi: 10.3389/fnagi.2018.00304.
[59] J. Zhang, X. He, L. Qing, X. Chen, Y. Liu, H. Chen, Multi-relation graph convolutional network for Alzheimer's disease diagnosis using structural MRI, Knowledge-Based Syst. 270 (2023), https://doi.org/10.1016/j.knosys.2023.110546.
[60] C. Ozdemir, M. Al Olaimat, S. Bozdag, A dynamic model for early prediction of Alzheimer's disease by leveraging graph convolutional networks and tensor algebra, Pac. Symp. Biocomput. 30 (2025) 675–689.
[61] B.T. Nallapu, et al., A machine learning approach to predict cognitive decline in Alzheimer disease clinical trials, Neurology. 104 (8) (2025) e213490, https://doi.org/10.1212/WNL.0000000000213490.
[62] B. Theodorou, A. Dadu, B. Avants, M. Nalls, and J. Sun, "MRI2PET : Realistic PET image synthesis from MRI for automated inference of brain atrophy and Alzheimer ' s," 2025.
[63] M. Mieling, M. Yousuf, and N. Bunzeck, "Predicting the progression of MCI and Alzheimer ' s disease on structural brain integrity and other features with machine learning," 2025, doi: 10.1007/s11357-025-01626-5.
[64] G. Zheng, Y. Lu, H. Chen, Deep learning-based framework for predicting mild cognitive impairment progression in neurology using longitudinal MRI, IEEE Access. 13 (April) (2025) 68903–68919, https://doi.org/10.1109/ACCESS.2025.3562432.
[65] Q. Chen, Q. Fu, H. Bai, Y. Hong, LongFormer: longitudinal transformer for Alzheimer's Disease classification with structural MRIs, in: Proc. - 2024 IEEE Winter Conf. Appl. Comput. Vision, WACV 2024, 2024, pp. 3563–3572, https://doi.org/10.1109/WACV57701.2024.00354.
[66] R. Bapat, D. Ma, T.Q. Duong, Predicting four-year's Alzheimer's disease onset using longitudinal neurocognitive tests and MRI data using explainable deep convolutional neural networks, J. Alzheimer's Dis. 97 (1) (2024) 459–469, https://doi.org/10.3233/JAD-230893.
[67] X. Hao, J. Li, M. Ma, J. Qin, D. Zhang, F. Liu, Hypergraph convolutional network for longitudinal data analysis in Alzheimer's disease, Comput. Biol. Med. 168 (2024) 1–10, https://doi.org/10.1016/j.compbiomed.2023.107765. July 2023.
[68] M.S.K. Inan, et al., A slice selection guided deep integrated pipeline for Alzheimer's prediction from Structural Brain MRI, Biomed. Signal Process. Control 89 (2023) 2024, https://doi.org/10.1016/j.bspc.2023.105773. November.
[69] S. Jahan, M.S. Kaiser, An Explainable Alzheimer's Disease Prediction Using EfficientNet-B7 Convolutional Neural Network Architecture, Springer Nature Singapore, 2023, https://doi.org/10.1007/978-981-19-8032-9_53.
[70] A.W. Mulyadi, W. Jung, K. Oh, J.S. Yoon, K.H. Lee, H.Il Suk, Estimating explainable Alzheimer's disease likelihood map via clinically-guided prototype learning, Neuroimage 273 (2023) 120073, https://doi.org/10.1016/j.neuroimage.2023.120073. December 2022.
[71] P. Rajpurkar, E. Chen, O. Banerjee, E.J. Topol, AI in health and medicine, Nat. Med. 28 (1) (2022) 31–38, https://doi.org/10.1038/s41591-021-01614-0.