Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

# Attention Enhanced CNN with LSTM Model for Early Detection of Alzheimer's Disease Using Longitudinal Data

**Rajeswari S.[1], Swathi K.[2*]**

[1]Research Scholar and Assistant Professor, [2]Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India.

**E-mail:** [1]srajeswari@kluniversity.in, [2*]dr.kswathi@kluniversity.in

## Abstract

Early identification was essential in efficient the management of Alzheimer's disease (AD), which is one of the biggest causes of dementia. Conventional Medical Imaging (CMI) methods such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans are used to understand various aspects of health. However, they lack the ability to capture the dynamic progression of Alzheimer's disease without incorporating sequential data series. This research introduces a novel method that has been designed to address these limitations. By combining spatial-temporal analysis with dynamic sequence processing, the work presents the "Attention Enhanced CNN+LSTM" architecture. The proposed dual pipeline architecture combinesLSTeM (Long ShortTerm enhanced Memory) for LSTM processing with an Attention ResNet for CNN processing. This work designs a novel "CS-Attention Block" with "Aggregate Weighted Pooling" for the enhancement of ResNet-50 and integration of LSTeM; further, a QKV (Query-Key-Value) attention mechanism for feature fusion is also proposed. For experiments, the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset containing longitudinal images provides the MRI and PET scan image data used for the analysis of the model. Results*:* The proposed model showed better performance for early AD 12 months before clinical diagnosis. The model also achieved an accuracy of 0.9151, and the AUC reached a value of 0.9784. This result was further improved to be improved by0.9302 for the metric accuracy and 0.9913 for the metric AUC ahead of the 6-month prediction. The model


*DOI: https://doi.org/10.36548/jiip.2025.2.009*
*Received: 03.05.2025, received in revised form: 28.05.2025, accepted: 13.06.2025, published: 24.06.2025*

fuses spatial-temporal features and processes the sequences to predict AD by addressing the limitations of existing models.

**Keywords:** Alzheimer's Disease, Deep Learning, Accuracy, AUC, LSTeM, ResNet.

## 1. Introduction

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia, marked by memory loss, cognitive decline, and behavioral changes. Early detection is crucial for timely intervention and management. Conventional AD detection models largely rely on imaging techniques like MRI and PET scans, with research showing improved diagnostic accuracy through their fusion. However, most models focus on static images and fail to capture the disease’s progressive nature, limiting early prediction capabilities.

To address this, hybrid architectures combining CNN (for spatial feature extraction) and LSTM (for temporal sequence learning) have been explored. Yet, standard CNN+LSTM models struggle—CNNs inadequately model temporal progression, while LSTMs are challenged by the spatial complexity of neuroimages. To overcome this, attention mechanisms are introduced to enhance feature relevance and integration.

This work proposes an Attention-Enhanced CNN+LSTM architecture for early Alzheimer's disease (AD) detection using longitudinal data. The model comprises an Attention ResNet (CNN) pipeline integrated with a Long Short-Term Enhanced Memory (LSTeM) unit. A novel CS-Attention Block with Aggregate Weighted Pooling refines spatial and channel features in ResNet-50, while LSTeM captures temporal progression. The outputs are fused using a QKV attention mechanism and classified using a dual fully connected layer. Experiments using the ADNI dataset demonstrate superior performance in early predictions at 24, 36, and 48 months compared to state-of-the-art models, confirming the model’s robustness in longitudinal AD detection.

### 1.1 Motivation

Identifying Alzheimer's disease in its early stages is crucial for initiating prompt therapeutic interventions meant to halt cognitive decline and improve patients' quality of life. However, the degenerative and progressive characteristics of AD are inadequately represented

by static neuroimaging data, such as single-time-point MRI or PET scans. These scans are essential to conventional diagnostic frameworks. Because the disease progresses slowly, longitudinal imaging data are necessary to detect early, subtle changes that would otherwise go unnoticed. Additionally, integrating metabolic activity from PET and structural data from MRI achieves a comprehensive, multimodal representation of brain state. A sophisticated model that can learn spatial and temporal patterns from multiple imaging modalities will be essential for accurately predicting early-stage Alzheimer's disease.

### 1.2 Contribution

Several important considerations were taken into account when choosing LSTM and hybrid attention-basedCNN architectures for early Alzheimer's disease detection. First, given that the disease predominantly evolves and requires methodologies capable of illustrating time-dependent alterations, LSTM and its more sophisticated variant, LSTeM, are employed. Second, deep convolutional neural networks, such as ResNet-50, can highlight the complex spatial patterns present in neuroimaging data, including MRI and PET scans. However, traditional CNNs often fail to effectively highlight nuanced yet clinically significant information. Spatial attention and channel obstruction augment feature concentration. Integrating temporal and spatial patterns through an appropriate merging approach yields a scaled dot-product attention mechanism. These integrated design choices effectively address critical challenges, including temporal modeling,, spatial accuracy, feature relevance, and sensitivity at the early stages. This makes the chosen method highly suitable for diagnosing longitudinal Alzheimer's disease.

## 2. Literature Review

Recent advancements in machine learning and deep learning have significantly contributed to the early diagnosis and classification of Alzheimer’s Disease (AD). A variety of models and approaches have been proposed, each offering unique insights while also facing notable constraints. For instance, Song et al. (2021) developed the GM-PET model by combining MRI and FDG-PET modalities. Although their method provided valuable diagnostic insight, it lacked temporal analysis and focused only on static imaging. Janghel et al. (2021) and Mehmood et al. (2021) employed VGG-16-based architectures, leveraging fMRI and transfer learning techniques respectively, but their models did not incorporate multimodal or longitudinal data, restricting their application in disease progression modeling. Jo et al.

(2020) proposed a 3D CNN for tau PET scans that was effective for specific tasks but lacked general applicability beyond the tau imaging context. Several other studies also focused primarily on static image classification. Raees et al. (2021) utilized traditional SVM and DNN models on MRI scans but did not implement more sophisticated neural architectures. Murugan et al. (2021) introduced DEMNET for dementia staging using MRI, though it was confined to stage-level prediction without dynamic progression tracking. Mohammed et al. (2021) and Kamada et al. (2021) used deep learning models like ResNet-50, AlexNet, adaptive RBM, and DBN, but their evaluations were limited to single-time-point datasets, omitting longitudinal components critical to understanding AD progression. Further work by Hamdi et al. (2022), Basheera et al. (2021), and Salehi et al. (2020) focused on enhancing CNN performance through architectural innovations or transfer learning. However, these models still centered on static imaging and lacked integration of sequential or multimodal data. More recent developments such as BLADNet (Duan et al., 2023), Conv-Swinformer (Hu et al., 2023), and other attention-based architectures showed promise in improving feature extraction, yet many of these studies excluded non-imaging factors like clinical scores, cognitive tests, or genetic data, which are essential for comprehensive AD modeling. Attempts to integrate longitudinal and multimodal data have been made by researchers such as Tripathi et al. (2023), Zhu et al. (2021), and El-Sappagh et al. (2021), who combined imaging with clinical or cognitive assessments. However, these approaches were often hindered by issues such as missing data, dataset specificity, or lack of generalizability. Moreover, while attention mechanisms have been introduced in some models to focus on critical regions in neuroimages, their usage remains limited, particularly in dual-pathway or hybrid networks that could simultaneously process spatial and temporal information.

In summary, although numerous deep learning approaches have contributed to AD diagnosis, a substantial gap remains. Most current models treat AD detection as a static image classification task, failing to capture the complex temporal dynamics and multimodal factors involved in disease progression. There is a pressing need for a unified, attention-augmented hybrid architecture capable of fusing spatial, temporal, and clinical data while preserving contextual relevance. Such a model would enhance diagnostic accuracy and offer meaningful predictions regarding the progression of Alzheimer’s disease, aligning more closely with real-world clinical needs.

## 3. Methodology

In order to increase the model performance as well as maintain consistency of the data, we carefully preprocessed MRI and PET images. Scanner-specific gradient non-linearity induced geometric distortions were corrected with Gradwarp on the MRI data. To correct for intensity variations across the magnetic field B1 nonuniformity was applied afterward. To sharpen images and correct for low-frequency intensity non-uniformity, we then estimated the bias field with the N3 procedure. The PET images were reconstructed by temporally co-registering individual scan frames and then averaging all frames to yield one composite image per scan. The PET scans were subsequently smoothed to 8 mm FWHM through a Gaussian filter, after spatial normalization to a [160 × 160 × 96] voxels format with 1.5 $mm^3$ isotropic voxel size. For the input of the model, all MRI and PET images were rescaled intensity-wise and shape-wise. Longitudinally aligned MRI and PET scans were used to obtain a temporal sequence for each participant. The length of sequences for different individuals was normalized by padding with zeros or truncation.

### 3.1 Aggregate Weighted Pooling (AWP)

In this section, a new technique called “Aggregate Weighted Pooling” is introduced for improving the processing of convolutional feature maps. The traditional types of models that we use, such as global average pooling and max pooling, however, don’t perform wellwith complex patterns of images. The errors of global average pooling at this layer stem from losing important spatial information, while max pooling fails to capture essential features of less salient objects. The AWP approach is developed to combine the strengths of global average and max pooling. AWP is a methodology that is implemented through a 1×1 convolutional layer to strengthen spatial and channel attention in training.

In the case of multiple image inputs, the size of the convolutional feature map grows along with redundant information. The AWP strategy can avoid this by forcing the network to concentrate and enhance the discriminative information. This helps AWP to process the complexity introduced by the increased feature maps and emphasize only the most necessary and temporally important features. This will efficiently enhance the feature extraction of CNNs towards the early-stage signs of AD, which would increase the accuracy and dependability of the employed predictive model. Fig. 1 (a) illustrates the method for averaging channel information using variable weights. Let's consider the convolutional feature matrix $X$, which is

defined in the space $\mathbb{R}^{H\times W\times C}$, where $H, W$, and $C$ represent the height, width, and number of channels are represented respectively.

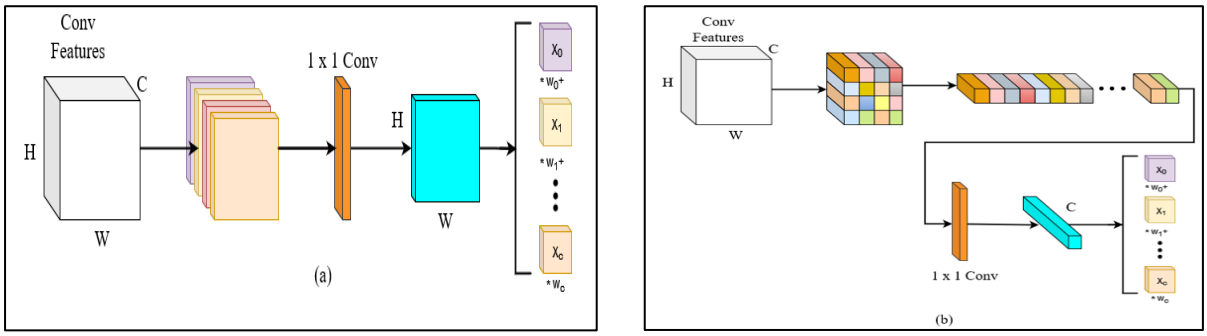



**Figure 1.** (a) Average Weighted Channel Pooling for Channel Information; (b) Average Weighted Spatial Pooling for Spatial Information

### 3.2 Channel and Spatial Attention Blocks

Using the proposed AWP, this work introduces an attention model termed Channel Spatial (CS)-Attention Block; the individual channel and spatial attention blocks are described in the following sections:

#### 3.2.1 Channel Attention Block (CAB)

In CNN, the feature maps are 3D, referred to as height, channel, and width. All channel maps are formed through a convolution kernel, leading to potential information redundancy, particularly with many channels. The approach explores these inter-channel relationships and minimizes redundancy by generating a Channel Attention Matrix (CAM) through a specialized CAB.

In the design, as shown in Fig. 2 (a), for an input feature $X_I$ in $\mathbb{R}^{C\times H\times W}$, the CAB commences by compressing the feature map along its channel axis utilizing the Aggregate Weighted Pooling (AWP). This squeezing process employs a 1×1 convolution kernel $\omega_s$ in $\mathbb{R}^{1\times 1\times C}$, resulting in a projection tensor $X^c_{\text{WAP}}$ in $\mathbb{R}^{H\times W}$. Each element $x_{(i,j)}$ in $X$ signifies a linear combination of all $C$ channels at the spatial location $(i, j)$. Then, a convolutional layer with a 1×1 kernel size is applied, which is then followed by ReLU activation and Batch Normalization, which create an intermediate feature map $X^c$; EQU (1) where:

$$X^c = \text{Conv}\,(X^c_{\text{WAP}}) \tag{1}$$

where $X^c$ is a representation in $\mathbb{R}^{C\times1\times1}$. This process effectively condenses the spatial information and makes the model focus on channel-wise features. Then, the feature map $X^c$ is reshaped and transposed to obtain two new feature maps, one with dimensions C×1 and the other 1×C. Matrix multiplication is employed, followed by a SoftMax operation on these maps to create the CAM, $M_C$, EQU (2)

$$M_C = \text{Softmax}\,(X^c \otimes (X^c)^T) \tag{2}$$

Each element $M_{C_{(i,j)}}$ in the channel attention matrix is further enhanced as following EQU (3):

$$M_{C_{(i,j)}} = \frac{\exp\left((X^c)_i \cdot (X^c)_j^T\right)}{\sum_{k=1}^{C} \exp\left((X^c)_{k^*} \cdot (X^c)_j^T\right)} \tag{3}$$

In this equation, $M_C \in \mathbb{R}^{C\times C}$ represents a 2-D matrix that indicates the inter-channel relationships among channel pairs. In the model, the element $M_{C_{(i,j)}}$ of the CAM denotes the effect of the channel $i$ on the channel $j$ . Subsequently, $X_I$. The feature is refined using this channel's attention. The channel-refined feature $X_C$ in $\mathbb{R}^{C\times H\times W}$, is obtained through a combination of matrix multiplication with the channel attention matrix $M_C$ and a residual shortcut connection, EQU (4).

$$X_C = X_I \oplus \left(\alpha(M_C \otimes X_I)\right) \tag{4}$$

Here, $\alpha$ is a learnable parameter. Initially, $\alpha$ is set to 0 to simplify the model's convergence in the early training epochs. This approach allows the channel attention $M_C$ to act effectively as a kernel selector, identifying the most appropriate filters for the task.

### 3.2.2 Spatial Attention Block (SAB)

The specifics of this spatial attention block are detailed in Fig. 2 (b), providing a clear visualization of its role in the proposed attention module.

For a given channel-refined feature $X_C$ in $\mathbb{R}^{C\times H\times W}$, the SAB begins by compressing the spatial axis's feature map using AWP. This squeezing is accomplished using a $1\times1$ convolution kernel $\omega_s$ in $\mathbb{R}^{1\times1\times C}$, resulting in a projection tensor $X^s_{\text{AWP}}$ in $\mathbb{R}^{H\times W}$. Each element $x_{(i,j)}$ in $X$ is a linear combination of the $C$ channels at the spatial location $(i,j)$. Following channel concatenation, a convolutional layer with a $3\times3$ kernel size is applied. This is

succeeded by a ReLU activation function and Batch Normalization to generate a transitional feature map $X^s$. For maintaining feature map size, a stride of 1 and padding of 1 are employed. Therefore, the intermediate feature map is given by EQU (5):

$$X^s = \text{Conv}\left([X^s_{\text{AWP}}]\right) \tag{5}$$

where $X^s \in \mathbb{R}^{1\times H\times W}$. This spatial attention process allows the model to focus on and amplify significant spatial features in the fused MRI and PET images.

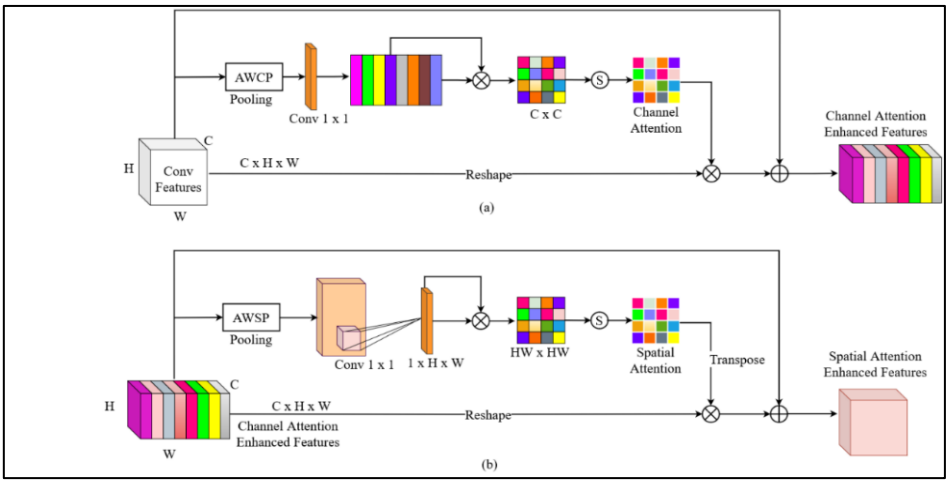



**Figure 2.** (a) Channel Attention Block, (b) Spatial Attention Block

Following the initial processing of the intermediate feature map $X^s$, it is the *'n'* reshaped and transposed to $HW \times 1$ feature map and $1 \times HW$ feature map. The SoftMax operation is applied after matrix multiplication to these maps to create the spatial attention matrix $M_S$, where the SoftMax function is measured across every row of the spatial matrix. This process can be mathematically expressed as EQU (6)

$$M_S = \text{SoftMax}\left(X^s \otimes (X^s)^T\right) \tag{6}$$

For a more detailed explanation, each element $M_{S_{(i,j)}}$ in the spatial attention matrix is calculated as EQU (7)

$$M_{S_{(i,j)}} = \frac{\exp\left((X^s)_i \cdot (X^s)^T_j\right)}{\sum_{k=1}^{HW} \exp\left((X^s)_{k^*} \cdot (X^s)^T_j\right)} \tag{7}$$

In this expression, $M_S$ is a matrix in $\mathbb{R}^{HW\times HW}$, representing the inter-spatial relationships between every two positions in the input feature map that emphasize the most informative spatial regions in the fused MRI and PET images.

In the model, the element $M_{S_{(i,j)}}$ in the spatial attention matrix signifies the impact of the $i^{th}$ position on the $j^{th}$ position. Subsequently, the matrix is applied to the channel-refined feature $X_C$ through matrix multiplication to obtain a spatially-refined feature $X_S$ in $\mathbb{R}^{C\times H\times W}$, enhanced by residual shortcut learning. This process can be formulated as EQU (8):

$$X_S = X_C \oplus \left(\beta(X_C \otimes (M_S)^T)\right) \tag{8}$$

The parameter $\beta$ is a learnable factor, initially set to 0 to simplify the early stages of the training process and facilitate smoother convergence. Through this mechanism, the spatial attention matrix $M_S$ effectively acts as a positional mask, highlighting the most crucial areas within the feature maps for analysis.

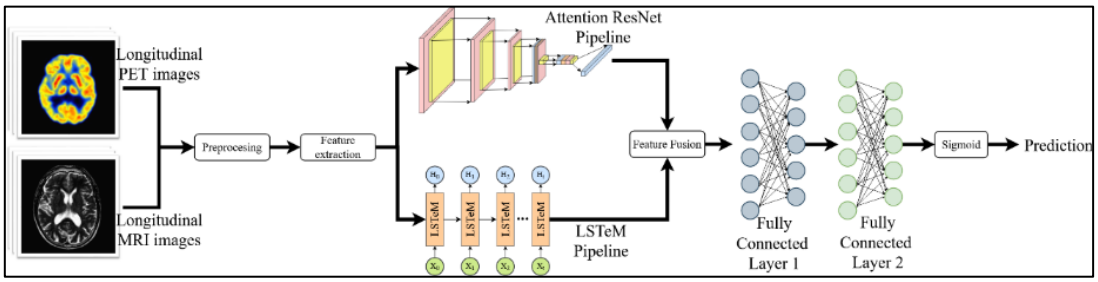


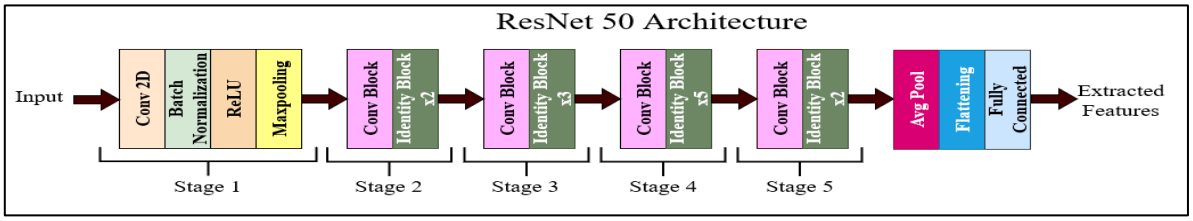

**Figure 3.** Proposed Architecture for Early Prediction of AD

### 3.3 Proposed Attention CNN+LSTM Architecture for Early Detection of AD

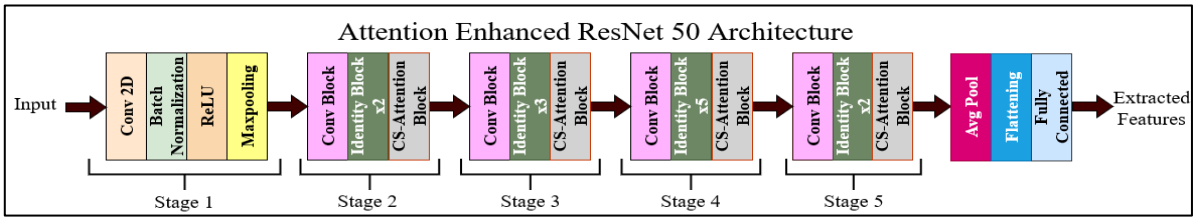
In this section, we propose a Hybrid Neural Network Model (HNNM) which consists of an CNN for spatial analysis and a LSTM for temporal processing, complemented with attention mechanisms for the early detection of Alzheimer's. It uses Attention ResNet for capturing spatio-temporal features and an LSTM module for temporal changes in longitudinal MRI and PET images. These features are then concatenated, and this combined feature set is fused using scaled dot-product attention mechanism, followed by fully connected layers and a sigmoid function to predict the probability of AD, resulting in effective early diagnosis.

### 3.3.1 Focused Spatial FE Using Channel and Spatial Attentive ResNet

The CNN pipeline we use is ResNet-50 (Fig. 4(a)), a type of deep network with skip connections introduced to prevent the training problem of vanishing gradients. It consists of 50 layers and it starts with a 7×7 convolutional layer, batch normalization, ReLU activation, and max pooling. The network is designed with bottleneck blocks, which include two 1×1 convolutions for dimensionality reduction and a 3×3 convolution for feature extraction, followed by batch normalization and ReLU. Mayank, Yadav et al./ JATIT 11 (3) (2016) 298 – 305 299 Skip connections and identity blocks help in the training of deeper layers. Following the last residual block average pooling decreases the spatial dimensions, followed by a FC layer with SoftMax activation to produce classification scores. ResNet-50's strong feature learning capabilities render it highly suitable for challenging image recognition tasks, and an optimal starting point for more advanced deep learning models.




(a)




(b)

**Figure 4.** (a) ResNet 50 architecture, (b) Attention Enhanced ResNet 50

Though the deep-architecture, including stacked convolutions, batch normalization, and residuals, makes ResNet-50 proficient in extracting features, the same treatment for the features at all spatial and channel dimensions is inadequate for the task of MI, where fine-grained feature differences are essential. To overcome this, we include CS-Attentive Blocks following all the residual blocks (Fig. 4(b)). These attention modules selectively weight

diagnostically useful patterns and automatically recalibrate feature maps on both spatial and channel dimensions.

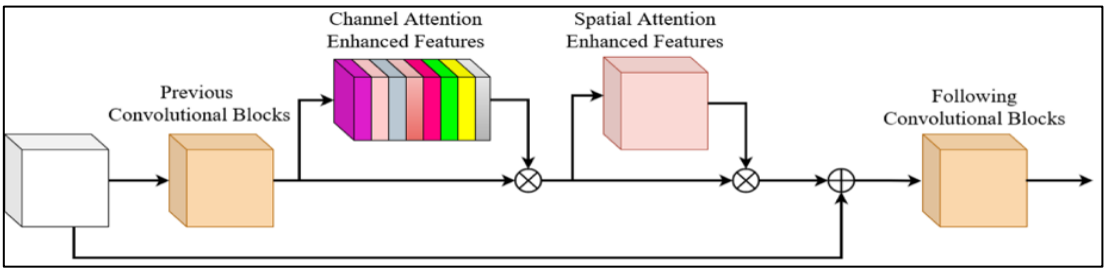



**Figure 5.** CS-Attentive Block in ResNet-50

Previous studies have indicated a preference for placing channel attention ahead of spatial attention (Fig. 5), which was also confirmed in our experimental analysis. Embracing this insight, similar formats have adapted to this model accordingly. In the context of the proposed work, focusing initially on channel attention allows the network to better discern and prioritize the varied and intricate features across different channels in the fused image data. This is particularly important in medical imaging, where additional channels can represent distinct aspects of the brain's structure and function, crucial for early Alzheimer's detection.

Following channel attention, spatial attention then refines the focus within the spatial dimensions of each channel, further improving the model's capacity to pinpoint areas of interest in the sequential image data. The feature map is denoted as $F$ in $\mathbb{R}^{C\times H\times W}$ and $F^{\mathrm{c}}$ represent output after applying the channel attention map to $F$ and $F^{\mathrm{s}}$ represents the output following the application of spatial attention to $F^{\mathrm{c}}$. These steps are mathematically described as EQU (9)

$$\begin{aligned} F^{\mathrm{c}} &= M_C(F) \otimes F \\ F^{s} &= M_S(F^{c}) \otimes F^{c} \end{aligned} \tag{9}$$

Where $M_C(F)$ is the channel attention map applied to $F$, and $M_S(F^c)$ is the spatial attention map applied to $F^c$. The symbol $\otimes$ denotes the element-wise multiplication operation. Finally, the output feature vector of $8 \times 512 \times 1$ dimension from the last FC layer is used as the ResNet feature, EQU (10)

$$r_v = f_{\text{resnet}}\left(v_f\right) \tag{10}$$

Here, $r_v$ is the final feature vector obtained from the ResNet model, and $v_f$ refers to the input vector.

### 3.3.2 Temporal Sequence FE using LSTeM

The LSTM model is a type of RNN that processes data sequences by passing items through a chain of repeating cell modules one at a time. Each cell in an LSTM is equipped with a three-gate structure: input, forget, and output gates. These gates regulate the flow of information.

Given an input $x_t$ at time step $t$, along with the previous hidden state $h_{t-1}$ and memory cell state $c_{t-1}$, the LSTM operates as follows: EQU (11) to EQU (16)

$$\text{Input Gate } (i_t): i_t = \sigma_s\left(U^{(i)}x_t + W^{(i)}h_{t-1}\right) \tag{11}$$

$$\text{Forget Gate } (f_t): f_t = \sigma_s\left(U^{(f)}x_t + W^{(f)}h_{t-1}\right) \tag{12}$$

$$\text{Output Gate } (o_t): o_t = \sigma_s\left(U^{(o)}x_t + W^{(o)}h_{t-1}\right) \tag{13}$$

$$\text{Memory Cell } (\tilde{c}_t): \tilde{c}_t = \sigma_t\left(U^{(c)}x_t + W^{(c)}h_{t-1}\right) \tag{14}$$

$$\text{Update Memory Cell State } (c_t): c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{15}$$

$$\text{Update Hidden State } (h_t): h_t = o_t \cdot \sigma_t(c_t) \tag{16}$$

In these equations, $\sigma_s$ and $\sigma_t$ represent the sigmoid and hyperbolic tangent functions, respectively. The gates and the current memory cell are calculated as activated sums of the weighted current input $x_t$ and the previous hidden state $h_{t-1}$. The forget gate $f_t$ removes data from the memory cell state while the input gate $i_t$ decides which new information is added. The output gate $o_t$ then uses the updated memory cell state $c_t$ to compute the current hidden state $h_t$. The parameters $U^{(*)}$ and $W^{(*)}$ represent the model's weights.

The standard LSTM has limitations when applied to problems that need enhanced memory capabilities and advanced temporal pattern recognition. To address these limitations, the LSTeM for the sequential recommendation model (Duan et al. 2023) is employed in this work.

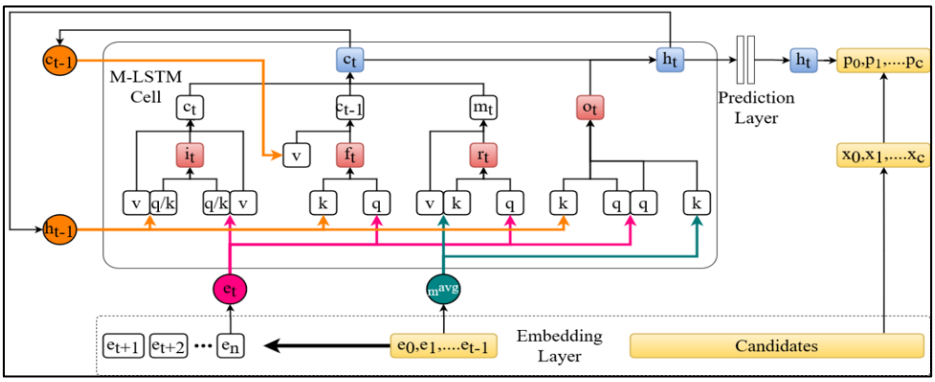



**Figure 6.** LSTeM Architecture

In the proposed architecture (fig 6), the LSTeM architecture is integrated into the LSTM pipeline to process temporal features of the fused MRI and PET longitudinal images. This novel LSTM variant introduces a recovery gate in addition to the standard input, forget, and output gates, thereby enhancing the model's ability to interpret complex temporal sequences.

- **Embedding Layer for MRI and PET Image Data**

Each series of images is transformed into a fixed-length embedding sequence, denoted as $E^{(img)} = \left[e_1^{(img)}, e_2^{(img)}, \ldots, e_n^{(img)}\right]$, where $n$ is the predetermined sequence length. For series with fewer images than $n$, padding (represented as zero vectors) is added at the beginning to maintain a consistent length across all sequences. In cases where a series has more than $n$ images, only the most recent $n$ images are retained for analysis. An embedding matrix $E \in \mathbb{R}^{N \times d}$ is constructed, where $d$ is the dimensionality of the embeddings. This matrix is used in conjunction with an embedding lookup table to map each image series $S^{(img)}$ to its embedding representation $E^{(img)} \in \mathbb{R}^{n \times d}$.

- **Adapting Gate Structures with "Q-K-V" Mechanism for Alzheimer's Detection**

The LSTeM augments the standard LSTM's input, forget, and output gates, which are typically expressed as $U^{(*)}e_t^{(img)} + W^{(*)}h_{t-1}$. The "Q-K-V" mechanism offers an improved method for processing temporal sequences from MRI and PET image data, emphasizing the correlation among the current image input $e_t^{(img)}$ and the previous hidden state $h_{t-1}$.

The Q-K-V Self-Attention model encodes each sequence element $x_i$ into a triplet of query $q_i$, key $k_i$, and value $v_i$. It determines $\alpha_{ij}$ correlation among $x_i$ and $x_j$ through the normalized product of the query of $x_i$ and the key of $x_j$ , resulting in a weighted value $\alpha_{ij}v_j$.

The attention for the entire sequence is calculated as,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{17}$$

In this EQU (17), $Q = \{q_i\}_i, K = \{k_i\}_i, V = \{v_i\}_i$, and $d_k$ is the dimensionality of the key. For normalization, the SoftMax function is employed on the attention weights, ensuring their sum equals one, and reflects the relative importance of each element in the sequence.

The self-attention identifies the correlation among the current image embedding$e_t^{(img)}$, the previous hidden state $h_{t-1}$, and memory $c_{t-1}$, along with the global memory $m_{t-1}^{avg}$. This mechanism maps these elements into each gate's query, key, and value groups. Unlike Self-Attention, which computes the overall relevance of each pair, "Q-K-V" in LSTeM focuses solely on capturing correlations, using the sigmoid function instead of SoftMax to normalize gate weights between 0 and 1.

The input gate manages the data accepted into memory at each time step. It derives information from two sources: the current image embedding $e_t^{(img)}$ and $h_{t-1}$. The operation performed in the gate involves two key products: the first is between the query of $e_t^{(img)}$ and the key of $h_{t-1}$, determining the relevance of short-term memory to the current input; the second is between the query of $h_{t-1}$ and the key of $e_t^{(img)}$, identifying which aspects of the current input are significant. A single vector is employed as a query and key for $e_t^{(img)}$ and $h_{t-1}$, leading to the EQU (18):

$$i_t = \text{Sigmoid}\left(\frac{qk_{(t,i,e)} \cdot qk_{(t,i,h)}}{\sqrt{d}}\right),\ \tilde{c}_t = i_t \cdot \left(v_{(t,i,e)} + v_{(t,i,h)}\right) \tag{18}$$

Here, $qk_{(t,i,e)}$ and $qk_{(t,i,h)}$ are the combined query and key for the current embedding and previous hidden state, while $v_{(t,i,e)}$ and $v_{(t,i,h)}$ are their values. The forget gate, which determines the information to be discarded from the long-term memory $c_{t-1}$, the "Q- K − V " mechanism is again employed. This gate's function is formulated as EQU (19):

$$f_t = \text{Sigmoid}\left(\frac{q_{(t,f,e)} \cdot k_{(t,f,h)}}{\sqrt{d}}\right),\ \tilde{c}_{t-1} = f_t \cdot c_{t-1} \tag{19}$$

Here, $q_{(t,f,e)} = U^{(f)} e_t^{(img)}$ and $k_{(t,f,h)} = W^{(f)} h_{t-1}$ are the query and key for the forget gate with $U^{(f)}$ and $W^{(f)}$ as the trainable parameters. The key of $h_{t-1}$ in place of $c_{t-1}$ the key is used considering $h_{t-1}$ as a selective projection of $c_{t-1}$.

- **Global Memory-Based Recovery Gate**

The recover gate employs a global memory vector, formulated using Self-Attention from prior image embeddings in the sequence. For a sequence $X = \{x_i\}_{i=1}^n$, the global memory vector for an element $x_q$ is computed as EQU (20)

$$\hat{x}_q = \sum_{i=1}^n \alpha(x_q, x_i) x_i \tag{20}$$

Here, $\alpha(x_q, x_i)$ is the attention weight assigned to each element $x_i$ about $x_q$ which is based on their mutual dependencies. This Self-Attention mechanism allows the recover gate to integrate a more holistic view of the fused image sequence by generating the global memory vector at time $t$. For the embedding matrix $E_{t-1} \in \mathbb{R}^{(t-1)\times d}$ of previous images, it first computes the dependencies among items, and then Global Average Pooling (GAP) is employed to create a global memory embedding $m_{t-1}^{avg}$.

The steps are*, EQU (21) to EQU (23)*

Compute Affinities: $\alpha = \text{softmax}\,(E_{t-1} E_{t-1}^T)$ (21)

Generate Weighted Sequence Embedding: $\bar{E}_{t-1} = \alpha E_{t-1}$ (22)

Create Global Memory Embedding: $m_{t-1}^{avg} = \text{GAP}\,(\bar{E}_{t-1})$ (23)

For the recover gate, the "Q-K-V" mechanism is employed with the current image embedding $e_t$ and global memory $m_{t-1}^{avg}$ , EQU (24) and EQU (25)

$$r_t = \text{Sigmoid}\left(\frac{q_{(t,r,e)} \cdot k_{(t,r,m)}}{\sqrt{d}}\right),\ m_t = r_t \cdot v_{(t,r,m)} \tag{24}$$

Here, $q_{(t,r,e)} = U^{(r)} e_t, k_{(t,r,m)} = W^{(r)} m_{t-1}^{avg}$, and $v_{(t,r,m)} = P^{(r)} m_{t-1}^{avg}$. (25)

Finally, the memory at step $t$ is updated by combining the outputs of the input, forget, and recover gates, EQU (26)

$$c_t = c_t + c_{t-1} + m_t \tag{26}$$

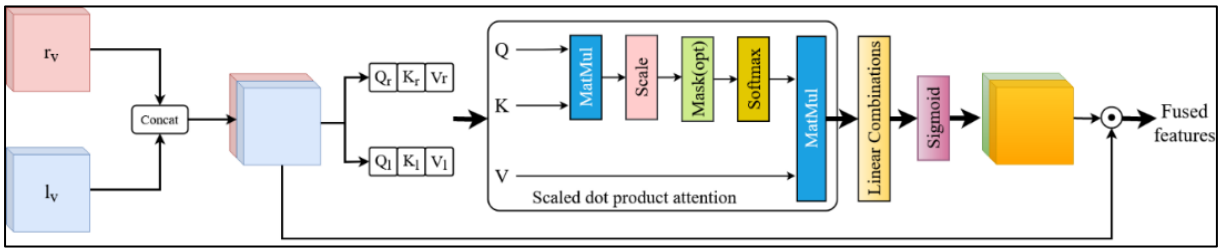



**Figure 7.** Scaled Dot Product Attentive Fusion

- **Output Gate**

The output gate of the LSTeM cell evaluates the relevancy of the cell's output and determines which aspects of $c_t$ should be relayed forward. This gate's function is prejudiced by three query-key pairs that include the reciprocal query-key pairs of the current image embedding $e_t$ and the previous hidden state $h_{t-1}$, denoted as $qk_{(t,o,e)} = U^{(o1)}e_t$ and $qk_{(t,o,h)} = W^{(o1)}h_{t-1}$, where $o$ represents the output gate.

The output gate is computed as *EQU (27)*

$$o_t = \text{Sigmoid}\left(\frac{qk_{(t,o,e)} \cdot qk_{(t,o,h)}}{\sqrt{d}}\right),\ h_t = o_t \cdot c_t \tag{27}$$

Here, $h_t$ is the output of the LSTeM cell at time $t$, which represents the model's hidden state. This mechanism ensures that most informative components of the cell state are selectively forwarded alone.

- **Prediction Layers**

The hidden state $h_t$ is then processed through a two-layer Feed-Forward Neural Network (FFNN) to formulate the final representation $\ell_v$ of the patient's neurological condition, EQU (28)

$$\ell_v = W^{(p2)}\left(\sigma\left(W^{(p1)}h_t\right)\right) \tag{28}$$

where, $W^{(p1)}$ and $W^{(p2)}$ in $\mathbb{R}^{d\times d}$ are the trainable parameters of the feedforward layers, and $\sigma$ represents the Rectified Linear Unit (ReLU) activation function that is used to introduce non-linearity between the two layers. The resulting feature vector $l_v$ encapsulates the refined representation of the patient's neurological state derived from the input fused image data.

### 3.3.3 Feature Fusion Using Scaled Dot Product Attention

The fusion process begins by concatenating the feature vectors obtained from the ResNet and LSTeM pipelines, as shown in Fig. 8. Let $r_v$ represent the spatiotemporal features extracted by ResNet and $\ell_v$ denote the dynamic sequence FE by LSTeM. The initial step of fusion is represented by a direct concatenation of these vectors, EQU (29)

$$concat(r_v, \ell_v) = [r_v; \ell_v] \tag{29}$$

Following the concatenation then, the scaled dot-product attention mechanism is used to assess the inter-feature relationships between $r_v$ and $\ell_v$. This transforms each feature set into query ($Q$ ), key ($K$), and value ($V$) vectors through differentiable weight matrices. The queries ($Q_r, Q_l$), keys ($K_r, K_l$), and values ($V_r, V_l$) are computed as following EQU (30)

$$\begin{aligned} Q_r, K_r, V_r &= \text{ LearnableTransform } (r_v) \\ Q_l, K_l, V_l &= \text{ LearnableTransform } (\ell_v) \end{aligned} \tag{30}$$

The similarity between the features is modelled using a function $f(Q, K)$, which calculates the attention weights, EQU (32)

$$f(Q, K) = \frac{[Q_r^T K_l, Q_r^T K_l]}{\sqrt{d}} \tag{32}$$

where $d$ represents the dimensionality of the query and key vectors. The softmax function is applied to these attention weights to normalize using the following:

$$\text{SoftMax } \big(f(Q, K_i)\big) = \frac{\text{Exp } \big(f(Q,K_i)\big)}{\sum_j \text{Exp } \big(f(Q,K_j)\big)}$$

The final step involves calculating the attention values by a weighted summation of the value vectors and the reweighted attention scores, EQU (33)

$$\text{Attention } (Q, K, V) = \sum_i \text{SoftMax } \big(f(Q, K_i)\big) V_i \tag{33}$$

This attention mechanism allows for context-aware fusion of the features. To further refine the feature interaction, a residual model similar to ResNet is employed.

This structure emphasizes the features requiring more attention by adding the input of the attention layer back to its output, EQU (34)

$$y_i = \mu o_i + x_i \tag{34}$$

Here, $o_i$ denotes the output of the attention mechanism and $x_i$ represents the concatenated input, $[r_v; \ell_v]$. This residual connection is followed by sigmoid activation to reweight each feature within the range [0,1], resulting in continuous masks $m_v$ and $m_f$ for $r_v$ and $\ell_v$, respectively, EQU (35)

$$\begin{aligned} m_r &= \text{Sigmoid}\,(y_i^r[r_v; \ell_v]) \\ m_l &= \text{Sigmoid}\left(y_i^l[r_v; \ell_v]\right) \end{aligned} \tag{35}$$

Finally, both features are element-wise multiplied with their corresponding masks, resulting in the reweighted fused vector EQU (37)

$$f_{fused}(r_v, \ell_v) = [r_v \odot m_r; \ell_v \odot m_l] \tag{36}$$

### 3.3.4 Classification Layer

The combined feature vector undergoes further processing in the classification layer. The fused features are first processed using the FC layer, which expands the feature dimensions to allow interpretation and facilitate the learning of complex patterns. The transformation is defined as EQU (38)

$$h_1 = \text{ReLU}\left(W_1 \cdot f_{fused} + b_1\right) \tag{38}$$

where, $W_1$ and $b_1$ represent the weights and biases of this layer, respectively. The ReLU function introduces nonlinearity that enables the model to capture more complex relationships in the data. To prevent overfitting, a dropout layer is included. The network then proceeds to a second FC layer to further refine the features, EQU (39)

$$h_2 = \text{ReLU}\,(W_2 \cdot h_1 + b_2) \tag{39}$$

where, $W_2$ and $b_2$ are the weights and biases of the second layer. The final output layer comprises a single neuron with a sigmoid activation function. This neuron provides a probability score that reflects the likelihood of AD presence, EQU (40)

$$\text{Output} = \text{Sigmoid}\,(W_3 \cdot h_2 + b_3) \tag{40}$$

The sigmoid function is ideal for binary classification.

## 4. Experiment Analysis

The experiments were performed on a machine having Intel Xeon Gold 6230 CPUs, 256 GB DDR4 RAM, and NVIDIA Tesla V100 GPUs with 16 GB HBM2 memory. Implementation was performed using Python 3.8, and neural networks were constructed using TensorFlow 2.4 and Keras 2.4. NumPy version 1.19 and Panda's version 1.2 was responsible for managing data operations, and OpenCV version 4.5 was used to support image operations.

### 4.1 Analysing the proposed Attention ResNet50

#### 4.1.1 Training Attention ResNet

The Attention ResNet model for early Alzheimer's detection incorporates attention mechanisms within a structured ResNet framework. It begins with a 3×3 convolution (stride 2), followed by batch normalization, ReLU activation, and bottleneck blocks with attention applied in three configurations: Parallel (ResNet(P)), Channel-Spatial (ResNet(C-S)), and Spatial-Channel (ResNet(S-C)). A final fully connected layer with 2048 neurons completes the model. Training used binary cross-entropyand hinge loss, with evaluation based on accuracy, precision, recall, and F1-score to compare performance across configurations.

#### 4.1.2 Analysing the Attention ResNet

This section evaluates the Attention ResNet model's performance across different attention configurations compared to the baseline ResNet-50, using Binary Cross Entropy (BCE) and Hinge Loss (HL) (Table 8). As shown in Fig. 8, the baseline ResNet-50 achieved over 93% in all metrics under both loss functions. Introducing attention mechanisms led to notable improvements. The parallel attention configuration (Attn-ResNet(P)) exceeded 98% accuracy with BCE. Sequential placements,spatial before channel (Attn-ResNet(S-C)) and channel before spatial (Attn-ResNet(C-S))yielded further gains. Attn-ResNet(C-S) achieved nearly 99% across metrics, with a peak F1-score of 99.46%, reflecting strong balance between accuracy and recall.

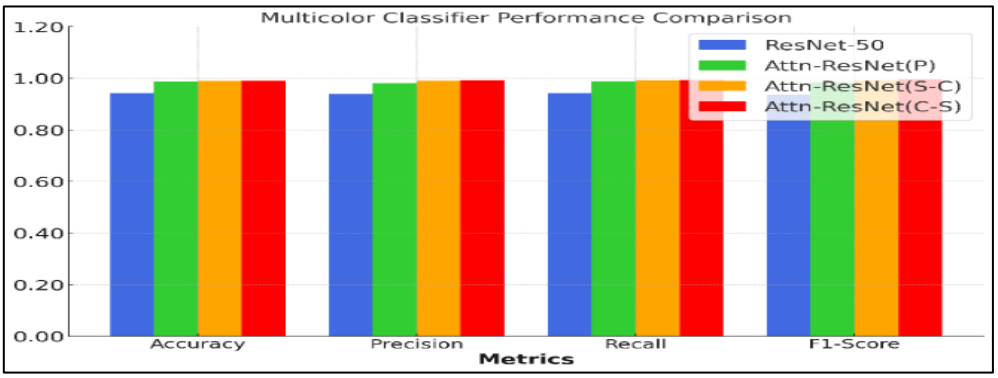



**Figure 8.** Performance Comparison of ResNet Configuration for BCE Loss

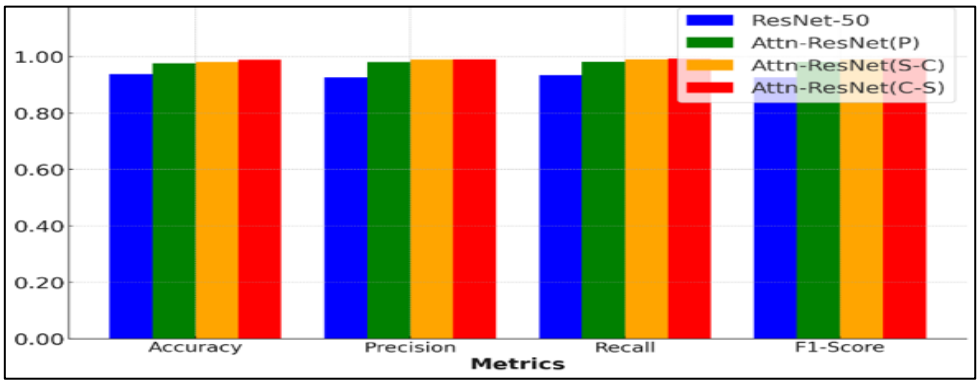



**Figure 9.** Performance Comparison of ResNet Configuration for Hinge Loss

For HL, results (Fig. 9) behaves the same as the BCE, attention based models perform better than ResNet-50. Attn-ResNet (S-C) outperformed the baseline and Attn-ResNet (P), whereas Attn-ResNet (C-S) obtained the highest scores with results close to 99% across all metrics.

Attn-ResNet(C-S) consistently dominated in accuracies across the 25 training epochs, increasing from 41.04% to 98.12% under BCE, and 41.05% to 95.96% under HL. In contrast, ResNet-50 achieved best performance at 89.76% (BCE) and 90.85% (HL). These findings support the success of channel-first attention placement.

### 4.1.3 Visualization of Attention Module

Feature maps of ResNet with and without attention are compared in Fig. 10. The top row (no attention) has the attention relatively evenly spread over the canvas, while on the bottom row, using ResNet(C-S), attention focuses more accurately to important regions. This

attention-embedded mapping contributes to better interpretability and generalization in medical imaging applications by emphasizing clinically important diagnostic features.

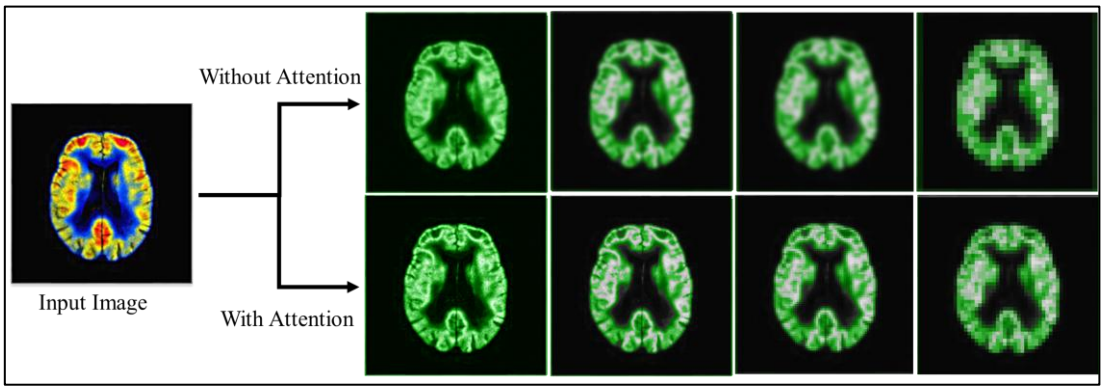



**Figure 10.** Visualization of the Attention Model

### 4.2 Analysis of the Attention-Enhanced CNN-LSTM for Early AD Prediction

The article applies longitudinal MRI and PET data from 381 ADNI subjects (CN, MCI, AD) to predict conversion from MCI-to-AD at 36 and 48 months. For psychometric analyses we expanded temporal sequences by using a 12-month HLRF with appropriately sliding windows for subjects with FF data:

12-month prediction: 160 →220

24-month prediction: 183 →235

The model integrates Attention ResNet for spatial features and LSTeM for temporal learning, with key settings:

ResNet: 3x3 conv, 16 bottleneck blocks, channel-first attention, 2048 FC neurons

LSTeM: 2 layers, 512 hidden units, 128-dim embedding

Training: 50 epochs, batch size 64, learning rate 0.001, dropout 0.5

Classification: 2 FC layers (1024 →512), sigmoid output, BCE loss

The Attn_ResNet + LSTeM achieved better performances, including accuracy, AUC, precision, recall, and F1-score, than ResNet-50/LSTM/LSTeM, and exhibited good predictive capacity for the early stage of AD progression.

### 4.2.1 Performance Evaluation on AD Early Detection

Within each dataset, the testing performance (ACC and AUC) of AD prediction for MCIc was conducted at 24-, 18-, 12-, and 6-month before the diagnostic time point. As shown in Table 11 and Fig. 11, the model was refined from 0.6563 ACC and 0.7026 AUC at 12 months to 0.6779 ACC and 0.7326 AUC at 6 months. Attn_ResNet + LSTeM outperformed ResNet50+LSTM and ResNet50+LSTeM, with the highest ACC and AUC of 0.9302 and 0.9913 at 6 months, verifying its better predictive performance.

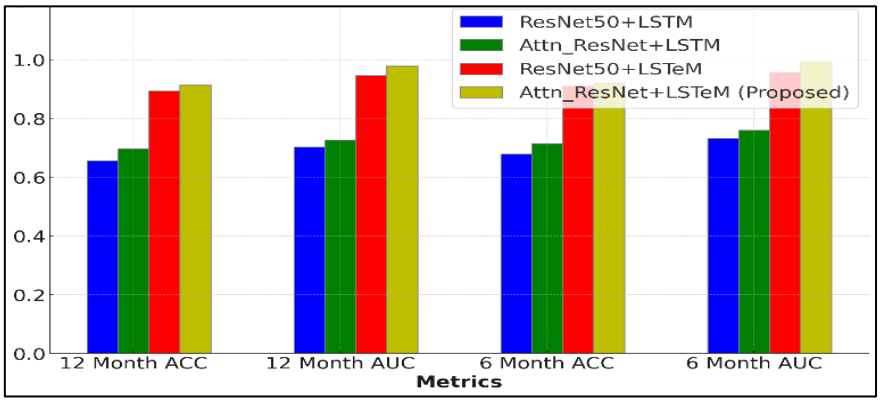



**Figure 11.** ACC and AUC for MCl→AD Subjects at the 24th Month

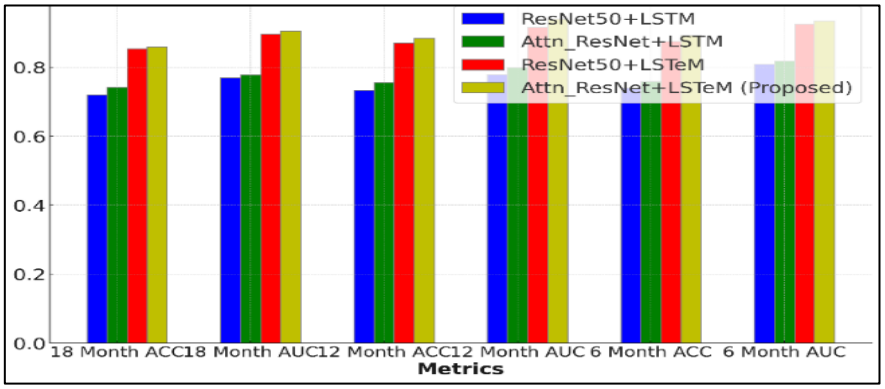



**Figure 12.** ACC and AUC for MCl→AD Subjects at the 36th Month

This comparison also examined models at 18, 12 and 6 months before AD diagnosis in MCI-C subjects who converted within 36 months (Table 12, Fig. 12). ResNet50+LSTM was 0.7407/0.8083(ACC/AUC) and Attn_ResNet+LSTM was 0.7594/0.8173. ResNet50+LSTeM had moderate performance but Attn_ResNet+LSTeM had the best result - 0.9127 ACC and 0.9436 AUC at 6 months. For 48-month converters, the performance of ResNet50+LSTM

improved to 0.6430/0.6841, while that of Attn_ResNet+LSTM increased to 0.6588/0.6892, further demonstrating the superiority of attention-enhanced models for early AD prediction.

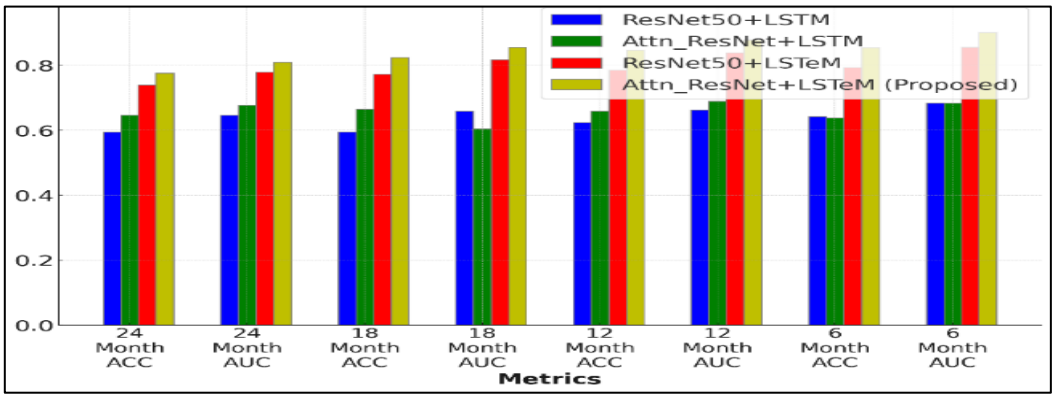



**Figure 13.** ACC and AUC for MCl→AD Subjects at the 48th month after Baseline Scan

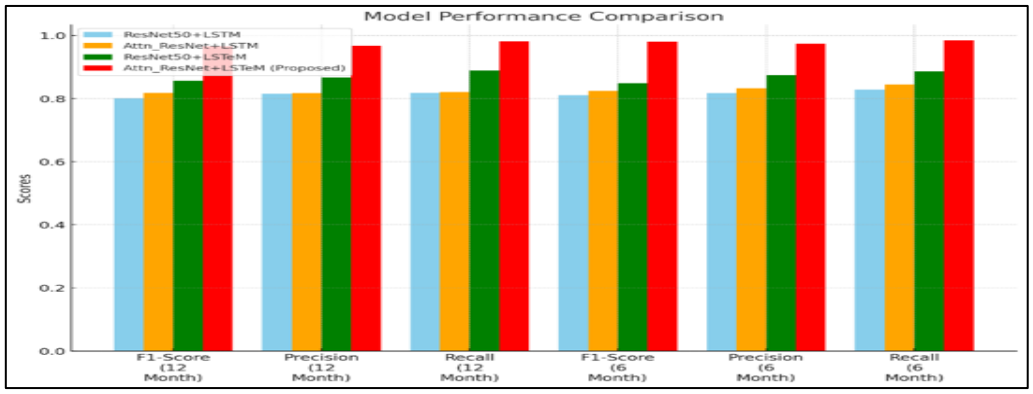



**Figure 14.** F1, Precision and Recall for MCl→AD Subjects at the 24th Month after Baseline Scan

The ResNet50+LSTeM model performed well, and the ACC/AUC increased from 0.7396/0.7788 (24-month) to 0.7930/0.8563 (6-month). The Attn_ResNet+LSTeM performed better than it and achieved 0.9051 ACC and 0.9119 AUC at 6 months. These findings demonstrate the benefit of the attention mechanism combined with spatial-temporal architectures for better prediction of early AD and future clinical applications. Performance (Tables 14–16, Figs. 14–16) provides additional suport for the utility of the model. Both reached their highest values 6 months in for the 24-month dataset (0.9801 for F1, 0.9736 for precision, and 0.985 for recall, confirming its high predictive capacity).

The proposed model is highly competitive on all the datasets which validated the utility of the attentions. At 18 months, it reached the values of 0.9004 (F1), 0.9077 (precision) and 0.9112 (recall), increasing to 0.9101, 0.9377and 0.942 at 12 months.

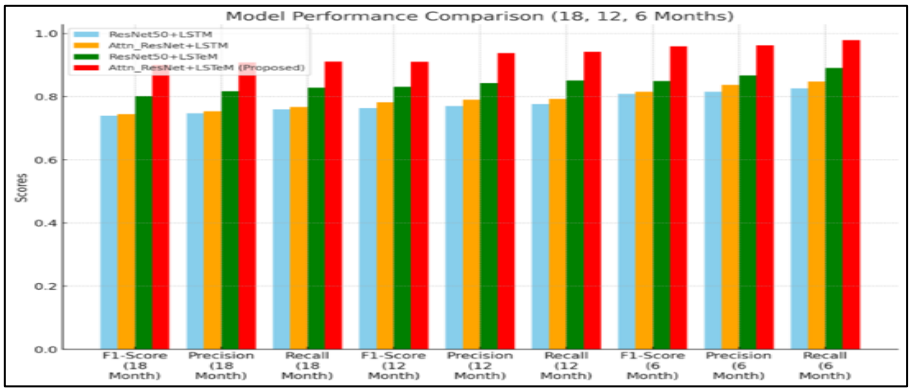



**Figure 15.** F1, Precision and Recall for MCl →36$^{th}$ month

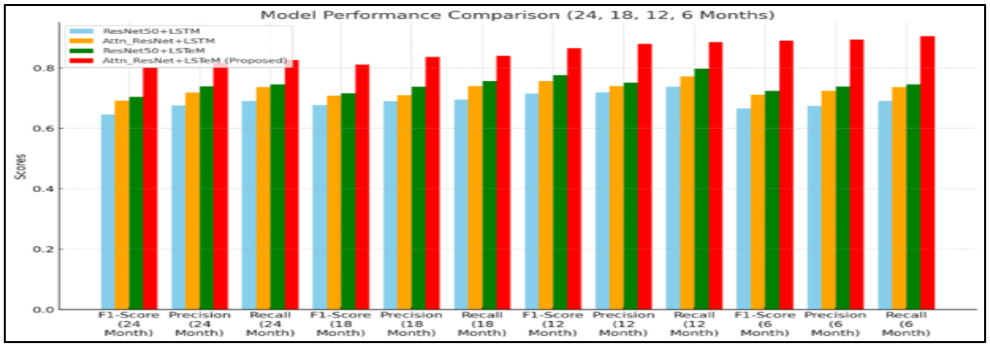



**Figure 16.** F1, Precision and Recall for MCl→48th Month

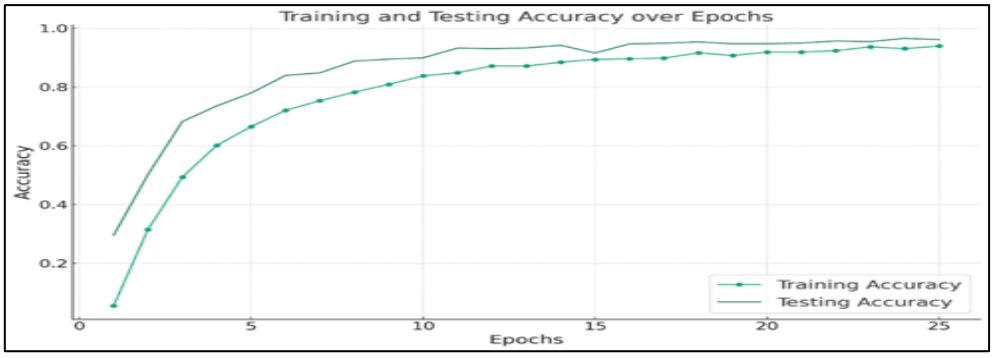



**Figure 17.** Training vs Testing Accuracy

## 4.3 Training and Validation Loss of the Proposed Model

The Attention-Enhanced CNN+LSTM model achieved significant learning gains after 25 epochs (Figs. 17 & 18), the training accuracy is increased from 5.47% to 93.97% and the

testing accuracy is increased from 29.47% to 96.20%. Losses were drastically reduced indicating that the model is able to predict effectively. In comparison (Fig. 19), the model achieved higher scores than five other state-of-the-art AD prediction methods: accuracy of 0.9302, AUC of 0.9913, precision of 0.9736, recall of 0.985, and F1 score of 0.9801, demonstrating its outstanding prediction ability.

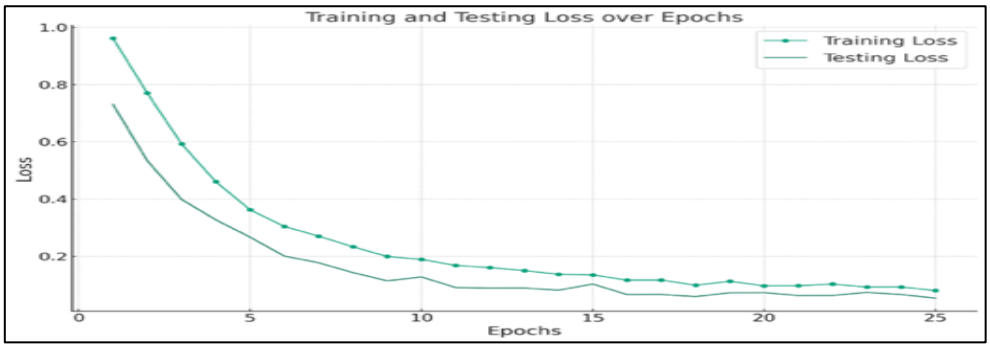



**Figure 18.** Training vs Testing Loss

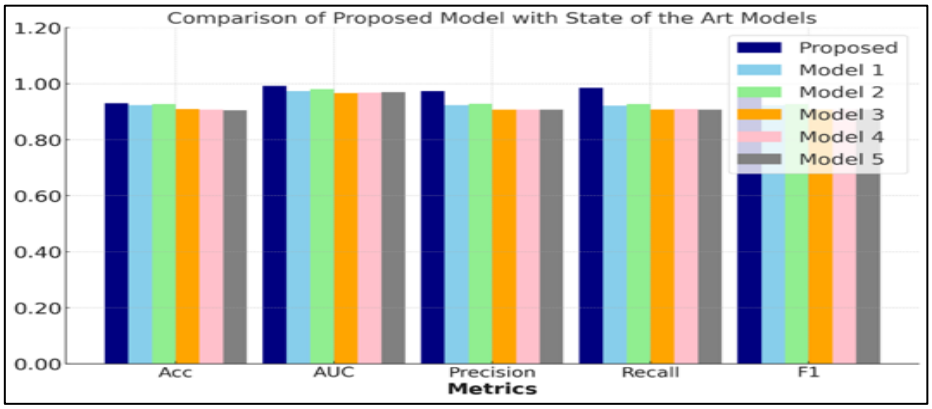



**Figure 19.** State-of-the-Art Comparison

## 5. Conclusion and Future Work

In this research, an innovative "Attention Enhanced CNN+LSTM" structure was introduced for the early detection of Alzheimer's disease (AD) based on longitudinal MRI and PET images. The two-pipeline model combines features with deep learning to represent spatio-temporal brain patterns and reduces the deficiencies of static image-based methods. Through integration of spatial MRI features with metabolic information derived from PET, the model yields additional insight into the evolution of the disease. The results showed high accuracy and AUC for the predictionof AD in the months prior to clinical diagnosis. Training LSTMs

with attention mechanisms developed the ability to interpret complex medical imaging data beyond what could be achieved using CNNs alone.

## References

[1] van Oostveen, Wieke M., and Elizabeth CM de Lange. "Imaging techniques in Alzheimer's disease: a review of applications in early diagnosis and longitudinal monitoring." International journal of molecular sciences 22, no. 4 (2021): 2110.

[2] Song, Juan, Jian Zheng, Ping Li, Xiaoyuan Lu, Guangming Zhu, and Peiyi Shen. "An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis." Frontiers in digital health 3 (2021): 637386.

[3] Janghel, R. R., and Y. K. Rathore. "Deep convolution neural network based system for early diagnosis of Alzheimer's disease." Irbm 42, no. 4 (2021): 258-267.

[4] Mehmood, Atif, Shuyuan Yang, Zhixi Feng, Min Wang, AL Smadi Ahmad, Rizwan Khan, Muazzam Maqsood, and Muhammad Yaqub. "A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images." Neuroscience 460 (2021): 43-52.

[5] Jo, Taeho, Kwangsik Nho, Shannon L. Risacher, Andrew J. Saykin, and Alzheimer's Neuroimaging Initiative. "Deep learning detection of informative features in tau PET for Alzheimer's disease classification." BMC bioinformatics 21 (2020): 1-13.

[6] Raees, PC Muhammed, and Vinu Thomas. "Automated detection of Alzheimer's Disease using Deep Learning in MRI." In Journal of Physics: Conference Series, vol. 1921, no. 1, IOP Publishing, 2021, 012024.

[7] Murugan, S., Venkatesan, C., Sumithra, M. G., Gao, X. Z., Elakkiya, B., Akila, M., & Manoharan, S. "DEMNET: a deep learning model for early diagnosis of Alzheimer's diseases and dementia from MR images." Ieee Access, 9, (2021),90319-90329.

[8] Mohammed, Badiea Abdulkarem, Ebrahim Mohammed Senan, Taha H. Rassem, Nasrin M. Makbol, Adwan Alownie Alanazi, Zeyad Ghaleb Al-Mekhlafi, Tariq S. Almurayziq, and Fuad A. Ghaleb. "Multi-method analysis of medical records and MRI images for

early diagnosis of dementia and Alzheimer's disease based on deep learning and hybrid methods." Electronics 10, no. 22 (2021): 2860.

[9] Kamada, Shin, Takumi Ichimura, and Toshihide Harada. "Image-Based Early Detection of Alzheimer's Disease by Using Adaptive Structural Deep Learning." In Intelligent Decision Technologies: Proceedings of the 13th KES-IDT 2021 Conference, pp. 595-605. Springer Singapore, 2021.

[10] Hamdi, Mounir, Sami Bourouis, Kulhanek Rastislav, and Faizaan Mohmed. "Evaluation of neuro images for the diagnosis of Alzheimer's disease using deep learning neural network." Frontiers in Public Health 10 (2022): 834032.

[11] Basheera, Shaik, and M. Satya Sai Ram. "Deep learning based Alzheimer's disease early diagnosis using T2w segmented gray matter MRI." International Journal of Imaging Systems and Technology 31, no. 3 (2021): 1692-1710.

[12] Patil, Vijeeta, Manohar Madgi, and Ajmeera Kiran. "Early prediction of Alzheimer's disease using convolutional neural network: a review." The Egyptian Journal of Neurology, Psychiatry and Neurosurgery 58, no. 1 (2022): 130.

[13] Salehi, Ahmad Waleed, Preety Baglat, Brij Bhushan Sharma, Gaurav Gupta, and Ankita Upadhya. "A CNN model: earlier diagnosis and classification of Alzheimer disease using MRI." In 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2020, 156-161.

[14] Khagi, Bijen, and Goo-Rak Kwon. "3D CNN design for the classification of Alzheimer's disease using brain MRI and PET." IEEE Access 8 (2020): 217830-217847.

[15] Pan, Dan, An Zeng, Longfei Jia, Yin Huang, Tory Frizzell, and Xiaowei Song. "Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning." Frontiers in neuroscience 14 (2020): 259.

[16] Ebrahimi, Amir, Suhuai Luo, and for the Alzheimer'S. Disease Neuroimaging Initiative. "Convolutional neural networks for Alzheimer's disease detection on MRI images." Journal of Medical Imaging 8, no. 2 (2021): 024503-024503.

[17] Alshammari, Majdah, and Mohammad Mezher. "A modified convolutional neural networks for MRI-based images for detection and stage classification of alzheimer disease." In 2021 National Computing Colleges Conference (NCCC), IEEE, 2021, 1-7.

[18] Abugabah, Ahed, Atif Mehmood, Sultan Almotairi, and Ahmad AL Smadi. "Health care intelligent system: A neural network based method for early diagnosis of Alzheimer's disease using MRI images." Expert Systems 39, no. 9 (2022): e13003.

[19] Duan, Junwei, Yang Liu, Huanhua Wu, Jing Wang, Long Chen, and CL Philip Chen. "Broad learning for early diagnosis of Alzheimer's disease using FDG-PET of the brain." Frontiers in Neuroscience 17 (2023): 1137567.

[20] Ebrahim, Doaa, Amr MT Ali-Eldin, Hossam E. Moustafa, and Hesham Arafat. "Alzheimer disease early detection using convolutional neural networks." In 2020 15th international conference on computer engineering and systems (ICCES), IEEE, 2020, 1-6.

[21] Rana, Md Masud, Md Manowarul Islam, Md Alamin Talukder, Md Ashraf Uddin, Sunil Aryal, Naif Alotaibi, Salem A. Alyami, Khondokar Fida Hasan, and Mohammad Ali Moni. "A robust and clinically applicable deep learning model for early detection of Alzheimer's." IET Image Processing 17, no. 14 (2023): 3959-3975.

[22] Sun, Haijing, Anna Wang, Wenhui Wang, and Chen Liu. "An improved deep residual network prediction model for the early diagnosis of Alzheimer’s disease." Sensors 21, no. 12 (2021): 4182.

[23] Ebrahimi, Amir, Suhuai Luo, and Raymond Chiong. "Introducing transfer learning to 3D ResNet-18 for Alzheimer’s disease detection on MRI images." In 2020 35th international conference on image and vision computing New Zealand (IVCNZ), IEEE, 2020, 1-6.

[24] Odusami, Modupe, Rytis Maskeliūnas, Robertas Damaševičius, and Tomas Krilavičius. "Analysis of features of Alzheimer’s disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network." Diagnostics 11, no. 6 (2021): 1071.

[25] Roy, Projapoti, Md Main Oddin Chisty, and HM Abdul Fattah. "Alzheimer’s disease diagnosis from MRI images using ResNet-152 Neural Network Architecture." In 2021

5th International Conference on Electrical Information and Communication Technology (EICT), IEEE, 2021, 1-6.

[26] Odusami, Modupe, Rytis Maskeliūnas, Robertas Damaševičius, and Sanjay Misra. "ResD hybrid model based on ResNet18 and DenseNet121 for early Alzheimer disease classification." In International Conference on Intelligent Systems Design and Applications, Cham: Springer International Publishing, 2021, 296-305.

[27] Liu, Sheng, Chhavi Yadav, Carlos Fernandez-Granda, and Narges Razavian. "On the design of convolutional neural networks for automatic detection of Alzheimer's disease." In Machine Learning for Health Workshop, PMLR, 2020, 184-201.

[28] Nguyen, Minh, Tong He, Lijun An, Daniel C. Alexander, Jiashi Feng, BT Thomas Yeo, and Alzheimer's Disease Neuroimaging Initiative. "Predicting Alzheimer's disease progression using deep recurrent neural networks." NeuroImage 222 (2020): 117203.