Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PROGRESS TRANSFORMER ON ALZHEIMER'S DISEASE PROBABILITY FINDER FROM MILD COGNITIVE IMPAIRMENT

SARANYA RATHINAM[1*], KRISHNAN NALLAPERUMAL[1], KALIDASS SUBRAMANIAM[2]

[(1)] Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, 627012, Tamil Nadu, India

[(2)]Department of Animal Science, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, 627012, Tamil Nadu, India

E-mail: *saranyadass88@gmail.com, krishnan17563@gmail.com, kallidass@gmail.com

ABSTRACT

The prodromal phase of Alzheimer's disease (AD) is called mild cognitive impairment (MCI). Effective treatments depend on identifying MCI patients who have a high chance of transforming AD. This paper proposes a temporal magnetic resource imaging (MRI) slice feature analysis using transformers to predict the chance of AD. The proposed Progress Transformer (ProgTransAD) model finds the relative changes in the MRI slices with the help of encoding the convolutional backbone feature maps and their corresponding cosine similarity. The proposed deep learning approach forecasts whether someone would develop Alzheimer's disease (AD) after receiving a diagnosis of MCI with three years of analysis. The performance of this unique deep learning network which can accurately diagnose AD progression is analyzed using the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) dataset and this ProgTransAD achieves 94% accuracy one year ahead.

Keywords: *Alzheimer's disease, Mild Cognitive Impairment, Neuroimaging, Magnetic Resource Imaging, ProgTransAD*

1. INTRODUCTION

In Alzheimer's disease (AD), which accounts for about 60% of dementia cases, symptoms such as disorientation, linguistic disorders, and progressive memory loss are common. The ailment would lead to the patient's death at some point [1]. This is a progressive disorder, therefore it will only become worse with time. In 1906, the first formal discovery of Alzheimer's disease was made by Dr. Alois Alzheimer when one of his patients died from a strange mental illness [2]. No established medications or treatments have been able to slow down or halt the progression of AD since there is currently no cure for the disease. From what we can tell from the studies [3-5], significant memory concern (SMC) could be a harbinger of AD and MCI. Early detection of Alzheimer's disease is strongly associated with the detection of mild cognitive impairment (MCI), a precursor stage of the disease. It has been shown that moderate cognitive impairment (MCI) greatly increases the risk of evolving into Alzheimer's disease (AD) or another form of dementia, even while the memory problems and complaints experienced by MCI patients do not substantially impact their daily lives. An accurate and timely diagnosis is crucial for patients with AD because it alerts them to the severity of their illness and allows them to take preventive measures, especially in cases where there is a risk of mild cognitive impairment progressing to Alzheimer's disease. Therefore, there is substantial therapeutic value in MCI's capacity to predict AD.

The development of therapy regimens to slow the course of Alzheimer's disease depends on early illness detection. Between 10% and 15% of people with mild cognitive impairment progress to Alzheimer's disease annually, making it the transitional period between normal brain function and dementia [6-7]. Reports indicate that both mild cognitive impairment (MCI) and Alzheimer's disease (AD) are linked to brain grey matter loss. As a consequence, changes in neuropathology may be seen before an official diagnosis of AD [8].

The use of neuroimaging biomarkers for disease phase classification or progression prediction from mild cognitive impairment to Alzheimer's disease has been explored in several previous studies [9-10]. Among the most popular imaging techniques, structural magnetic resonance imaging (MRI) stands out for its low cost, high resolution, and lack of invasiveness. The researchers in this study divided the MCI patients into two groups according to whether they had AD within three years or not. Patients with moderate cognitive impairment (MCI) who progress to Alzheimer's disease (AD) and those who do not have similar, but much milder, pathology changes. Consequently, the patient who progresses from mild cognitive impairment to Alzheimer's disease is much more difficult to identify. This MRI prediction is challenging because the pathological changes linked with AD development are delicate and vary across subjects, making it difficult to distinguish between individuals who progress from mild cognitive impairment (MCI) to Alzheimer's disease (AD) and those who do not. Performance was improved by removing the impact of aging, which is associated with a decline in prediction accuracy due to the normal shrinking of the brain.

Using machine learning approaches, automatic predictions of MCI to AD conversion are effectively implemented [11]. Contemporary complicated image analysis procedures generate vast amounts of data. Image processing techniques provide tools for making sense of imaging data. The rapid development of deep learning has allowed it to surpass more traditional machine learning methods in some domains, including computer vision and medical image analysis, two areas where it has achieved remarkable progress. Hence, this technology may be used for research and applications by everyone, including those without a medical background, especially in medical image processing. In several visual activities, including those requiring medical analysis, the vision transformer mechanism has recently shown superior performance compared to other mechanisms.

In this work,

- A novel model is developed to analyze MRI slices of individuals with mild cognitive impairment in later years to determine the likelihood of Alzheimer's disease.
- A transformer is used to process features from two layers of the Deep CNN model for the temporal slices, and their progress is measured using the similarity measure approach.
- By using dense characteristics and a transformer-based temporal slice analysis, the likelihood of AD from MCI may be efficiently determined.

2. RELATED WORKS

Since it was shown to be a slowly progressing illness, many researchers have taken an interest in AD. It is essential to regularly monitor AD to grasp the illness's progression and correctly capture predictive variability. For this reason, researchers started to think about longitudinal data instead of one-time-point data when trying to predict when MCI would turn into AD. Machine learning methods work well when used to forecast when mild cognitive impairment would progress to Alzheimer's disease. There are a few studies that attempt to categorize people with MCI who are at a higher risk of developing AD in this body of research.

The research was carried out to assess the feasibility of using principal magnetic resonance imaging (MRI), biomarker candidates, cerebrospinal fluid (CSF), and neuropsychological tests to predict which patients would move from mild cognitive impairment (MCI) to amyotrophic lateral sclerosis (AD), as reported by Ewers et al. [12]. The detection of MCI-to-AD transmission is studied by Ritter et al. [13] using huge multimodal data with different degrees of missing values. They expected AD to evolve within three years, encompassing all available modalities, based on data from MCI patients in the ADNI. Different categorization algorithms and approaches to missing data filling are compared. To evaluate performance, both manually selected and expert-prioritized attributes were used.

Cognitive performance, cognitive reserve on neuropsychological tests, APOE genotype, hippocampal sizes, and MEG (magnetoencephalography) power sources were examined by López et al. [14] to detect the onset of AD in a cohort of 33 individuals with mild cognitive impairment (MCI). Two groups of MCI patients were followed for two years based on the results: those with stable MCI (sMCI, 21 participants) and those with progressing MCI (pMCI, 12 people). A model for predicting AD conversion was developed using hierarchical logistic regression analysis of baseline multifactorial data.

Using magnetic resonance imaging (MRI) and fludeoxyglucose (FDG) PET, Ding et al. [15] developed a model to detect the progression of mild cognitive impairment (MCI) to amyotrophic lateral sclerosis (AD). Laplace eigenmaps were taught to

patients with Alzheimer's disease, healthy controls, and mild cognitive impairment. To forecast the MCI prognosis based on eigenmaps, a support vector machine was used. In addition, the prediction looked at how easy it was to combine different types of data.

Minutes et al. [16] developed a machine learning method to detect the progression from mild cognitive impairment (MCI) to amyotrophic lateral sclerosis (AD) over a span of two to three years. Unidentified longitudinal biomarker values may be found by first transforming the baseline and first follow-up measurements of a sample of multimodal biomarkers into autoregressive parameters trained on longitudinal data. One of the three suggested methods is used to compute the linear prediction coefficients, which may be used with a single predictor or with many predictions. Predictions of future clinical changes may be improved by combining many variables with an SVM classifier.

As a means of accurately predicting when mild cognitive impairment may progress to Alzheimer's disease, Lin et al. [17] developed a CNN-based deep learning method. Age correction and other modifications are applied to MRI images as a first step in processing. The second step is to combine the local patches seen in these photos to generate 2.5 dimensions. Next, the convolutional neural network (CNN) is taught to identify the deep learning features of MCI patients by using NC and AD patches. Then, CNN is trained by mining structural brain image properties using FreeSurfer. The last step is to feed both types of data into an extreme learning machine classifier to predict the AD conversion.

By combining medical assessments, neuroimaging data, and cerebrospinal fluid (CSF) biomarkers, Huang et al. [18] offered a personalized MCI-to-AD transition prediction using a Multi-predictor Nomogram. To generate the Radiomics signature (Rad-sig), the Least Absolute Shrinkage and Selection Operator (LASSO) method was used to identify seventeen cerebral cortex features. Between the transformed and non-transformed subjects, clinical parameters and amyloid-beta peptide (A) concentrations were selected using Spearman correlation. Building and testing a nomogram with image properties, a clinical component, and an A concentration was the next stage.

To detect the transition from mild cognitive impairment (MCI) to Alzheimer's disease (AD) within three years, Li et al. [19] proposed a subtyping-based prediction method using subgroups of MCI patients. To determine whether a patient with mild cognitive impairment (MCI) would develop Alzheimer's disease (AD) within three years, researchers developed a method based on multiple kernel learning, variational Bayes approximation.

Minhas et al. [20] proposed a decision-support method to identify patients who may show signs of progressing from mild cognitive impairment to Alzheimer's disease. We monitor the continuous trajectories of valid predictors in the MCI population that are generated from multivariate MRI biomarkers and cognitive assessments. Assuming piecewise linear illness development, this study established a novel method based on weighted gradient offset that allows for the prediction of the future indicator rate using data from at least two prior follow-up examinations. An SVM classifier is required to identify all of the prediction trajectories before they can be used as features for MCI-to-AD progressors.

Rye et al. [21] created a model to identify basic neurodegenerative processes early on, allowing for treatment to begin before the disease might spread across the brain. Using continuous data from the ADNI dataset, this research examined the course of AD in a group of patients identified with mild cognitive impairment at the standard assessment. There was a group that remained constant throughout time (sMCI, n = 357) and another group that transitioned to AD (cAD, n = 321). We used a Random Forest (RF) classifier that took into account parameters such as Hippocampus volume, cognitive function, and genetic APOE status, in addition to a model that relied on ensembles formed by combining five separate methods. After comparing the results of the two models, the RF approach was used to determine the significance of the features.

Early Alzheimer's disease diagnosis may be very accurate even with small sample numbers; a two-stage strategy integrating contrastive and transfer learning was proposed by Lu et al. [22]. To be more precise, the 3D convolutional neural network (CNN) approach was pre-trained using publically available medical picture data to grasp basic medical features to understand more detailed components of magnetic resonance imaging (MRI) images. Then, contrastive learning was used. All of the conventional methods were outperformed by the two-step method.

To detect aberrant brain activity linked to different stages of Alzheimer's disease (AD), Jiao et al. [23] used electroencephalograms (EEGs). To fully comprehend EEG's usefulness in the precise diagnosis and assessment of AD and its precursor, moderate cognitive impairment (MCI), further study is required. This work is necessary for the discovery

of significant EEG biomarkers to distinguish between persons with early-stage AD and to monitor the progression of the illness. A three-tiered classification of HC, MCI, and AD was developed using biomarkers derived from resting-state EEG recordings. The classification findings of Random Forest Regression were then used to discover the optimal EEG biomarkers.

Neural networks have recently been the focus of a plethora of studies aimed at improving the diagnosis of mild cognitive impairment and Alzheimer's disease.

Basaia et al. [24] developed and validated a deep learning approach for predicting the probability of AD and MCI that might progress to AD (c-MCI) using a structural MRI scan of the brain. Convolutional neural networks (CNNs) were used to 3D T1-weighted images from ADNI and their institute's participants (418 AD, 407 HC, 533 s-MCI, and 280 c-MCI). This study tested CNN's capacity to distinguish between s-MCI, c-MCI, and AD.

Lin et al. [25] developed a system that uses extreme learning machine grading to aggregate multimodal data and predict when MCI will progress to AD. First, LASSO was used to select only relevant MRI features from the obtained magnetic resonance imaging (MRI) pictures. Secondly, all imaging modalities employed on MCI patients, including positron emission tomography, magnetic resonance imaging (MRI), gene data, and biomarkers found in cerebrospinal fluid, were evaluated using the ELM approach. The classifier was finally Over the course of 1–5 years, these grading scores were obtained from many modalities to differentiate between subjects with stable MCI and those with advancing MCI.

For 150 patients with stable mild cognitive impairment (sMCI), 150 healthy controls (NC), and 157 patients with transformed mild cognitive impairment (cMCI), Wu et al. [26] provided the first magnetic resonance imaging (MR) scans and three-year follow-up data. Deep convolutional neural networks were used to predict the time it would take for mild cognitive impairment (MCI) to progress to Alzheimer's disease (AD) and to differentiate between the stages of MCI in the normal control (NC) group. They used five-fold cross-validation and transfer learning from fine-tuned ImageNet to assess conversion risk and evaluate the performance of two convolutional neural network (CNN) schemes, GoogleNet and CaffeNet, in different classifications.

Wegmayr et al. [27] used deep learning algorithms to detect the transition from mild cognitive impairment (MCI) to Alzheimer's disease (AD) during a 48-month follow-up. A state-of-the-art approach that divides the conversion prediction process into discriminative and generative phases. From a baseline image, they create a synthetically aged brain picture using the newly created Wasserstein-GAN model. To predict the future sickness state, an MCI/AD discriminator is fed the aged image.

Using demographic data and cerebrospinal fluid biomarkers in addition to baseline cross-sectional neuroimaging data and longitudinal cognitive presentation, Lee et al. [28] proposed a multi-modal deep learning approach to investigate the prediction of MCI to AD conversion. Several GRUs were applied to the longitudinal multi-domain data as well as data from all subjects with every modality.

To differentiate between individuals with pre-symptomatic AD and other moderate cognitive impairment (MCI) patients, Shen et al. [29] developed a new architecture for deep belief networks (DBNs). Photos captured by 109 subjects utilizing 18F-fluorodeoxyglucose-PET as part of the continuing longitudinal AD investigation. The patients were categorized into two groups: those with stable cognitive impairment and those with progressive mild cognitive impairment. Picture preprocessing, normalization, and smoothing; region-of-interest identification; feature learning using deep neural networks; and classification with support vector machines with three kernels are the four operations that comprise the suggested framework. To foretell when mild cognitive impairment (MCI) would progress to Alzheimer's disease (AD) three years after diagnosis, Ocasio et al. [30] developed a novel deep-learning approach. The longitudinal data set includes T1-weighted 3D MRI images from both the first scan, which included an examination of MCI, and the follow-up scan, which was taken one year later. To detect the transition of mild cognitive impairment (MCI) patients to Alzheimer's disease (AD) three years after diagnosis, a deep learning classification system was developed using MRIs from the AD and NC cohorts. This approach could then be used in transfer learning. Two methods of transferring knowledge were evaluated.

A computer-aided approach for differentiating Alzheimer's disease (AD) from cognitively normal and moderate cognitive impairment (MCI), its early phase, was devised by Lim et al. [31] using just structural magnetic resonance imaging (sMRI).

Axial brain images obtained from 3D MRI were fed into CNN for multiclass categorization. A convolutional neural network (CNN) built from scratch, VGG16, and ResNet-50 were also considered. To extract features, two convolutional neural network (CNN) models, ResNet-50 and VGG-16, were used. The classification was accomplished by developing a new densely connected classifier.

Using a cascaded DNN (Deep neural network) architecture, Akhtar et al. [32] were able to predict which individuals with MCI will progress to AD in the next year. To forecast the future value of each biomarker using two previous follow-up measurements, a DNN regression approach is trained and calibrated after sorting and normalizing longitudinal data. Next, a second DNN classifier approach is used with the three time-domain window data to identify MCI progressors (MCIp) and MCI stables (MCIs).

Using machine learning and deep learning networks, many automated methods have been shown for predicting the progression of mild cognitive impairment to Alzheimer's disease. Although there are approved approaches in the literature, none of them have shown sufficient performance. Therefore, a reliable model that can improve forecast accuracy is still required. To that end, this research suggests a deep learning model that incorporates an attention mechanism to enhance the precision of predictions.

3. PROPOSED PROGRESS TRANSFORMER MODEL

The architecture of the proposed work is depicted in Figure 1. The proposed deep learning network with a transformer mechanism used in an automated way to forecast the development of mild cognitive impairment to Alzheimer's disease visibly increases the prediction accuracy. Two periods of MRI screening data were delivered into two separate Deep CNN layers and each Deep CNN has three different convolution sizes in our proposed technique. Following the Deep CNN layer, two vision transformers were employed for convolution in two different levels of feature maps. This model utilized a cosine similarity measure to anticipate how similar the data from the diagnosis of MCI and its regular monitoring data with a minimum of 2 to 3 years. After concatenating cosine similarity along with the dense features from the transformer block, a Fully Connected Layer was employed to classify the targeted data. The classification is described as 0 for still in the mild stage, which suggests no progress, and 1 for Alzheimer's disease progress. The following sections explain the modules used in the ProgTransAD.

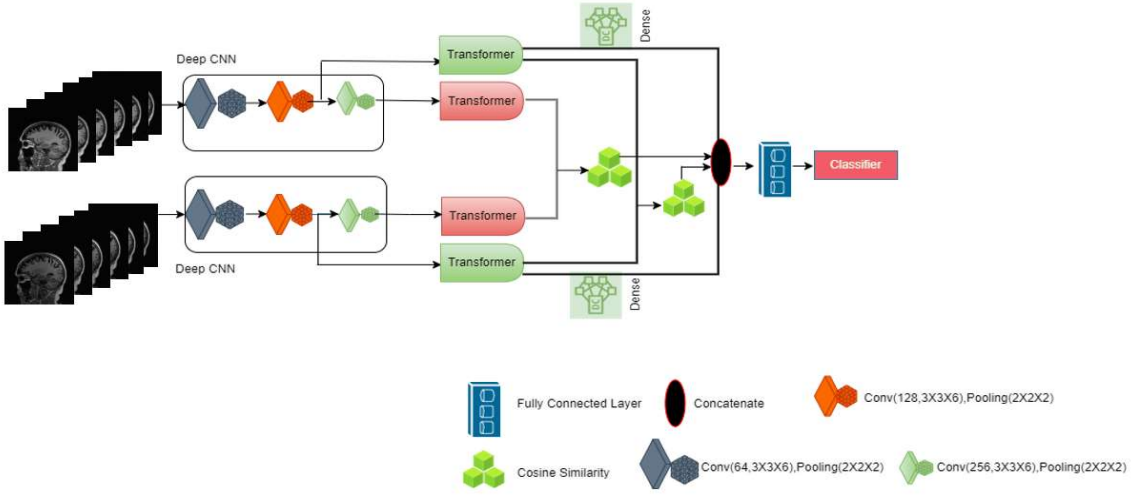

Figure 1: The architecture of the proposed ProgTransAD

3.1 Deep CNN

The majority of the calculation is done in the convolution layer, which is the base layer of a ConvNet. It is a configuration of feature maps that includes neurons. The layer's parameters are a collection of adaptable kernels or filters. The aforementioned filters are convolved with feature maps to generate a distinct 2-dimensional activation map. To construct the resultant volume, this map is subsequently stacked over the depth dimension. The complicated nature of the network is decreased by minimizing the number of parameters by allowing neurons in the same feature map to share parameters.

For each of the two Deep CNN layers in our proposed work, we used three 3D convolution layers with a filter size of 3 x 3 x 6 followed by max-pooling with the size 2 x 2. The first convolution uses 64 filters, the second with 128 feature maps, and the third uses 256 filters to produce feature maps. The transformer block receives the output of convolution layers with filter sizes 256, and 128 from both the two different period MRI slices.

3.2 Vision Transformer

When employed for machine translation, transformers [33] achieve remarkable performance by completely replacing recurrence and convolutions with self-attention methods [34-39]. Transformers eventually overtake other models as the model of choice for a variety of natural language processing (NLP) tasks. Recent studies attempted to duplicate the performance of CNNs on NLP tasks by incorporating the self-attention mechanism into computer vision challenges. These successes raise community interest in developing models for vision tasks that are solely transformer-based, devoid of convolutions and inductive bias. Images must first be separated into patches before every one of them is subjected to the calculation to use Transformers and other approaches successfully. Vision Transformer (ViT) uses the Transformer Encoder to extract features in classification problems. We used the Vision transformer (ViT) design to compete with CNNs on image classification. In our proposed work, the transformer outcome of the convolution layer with filter size 256 from both Deep CNN layers is sent into the Cosine similarity layer. Similar to this, the result of the convolution layer with filter size 128 also from both Deep CNN layers is sent into another Cosine similarity layer.

The architecture of the Vision Transformer is shown in Figure 2. The excellent-quality input image is translated into a not high-quality feature map using a set of transformer blocks comprising self-attention and feed-forward layers for feature encoding, a linear layer for classification score estimation, and a linear layer for patch embedding. A feed-forward Multi-Layer Perceptron (MLP) and a multi-head self-attention (MHSA) layer make up each transformer block. The attention transforms the input linearly to produce a trainable association memory that outputs a query (Qu) and two key-value pairs (L, W). The output of attention is determined mathematically by,

$$Attention\ (Qu, L, W) = Softmax(QuL^T/\sqrt{d})V \quad (1)$$

Where $\sqrt{d}$ Is a scaling factor determined by the network's depth. The MLP then creates the input for the following block using the normalized output of the MHSA. In the self-attention example above, Qu and L are multiplied to create the attention map, that illustrates the correlation among each layer's tokens. To integrate the embeddings in the value V, it is used to retrieve them.

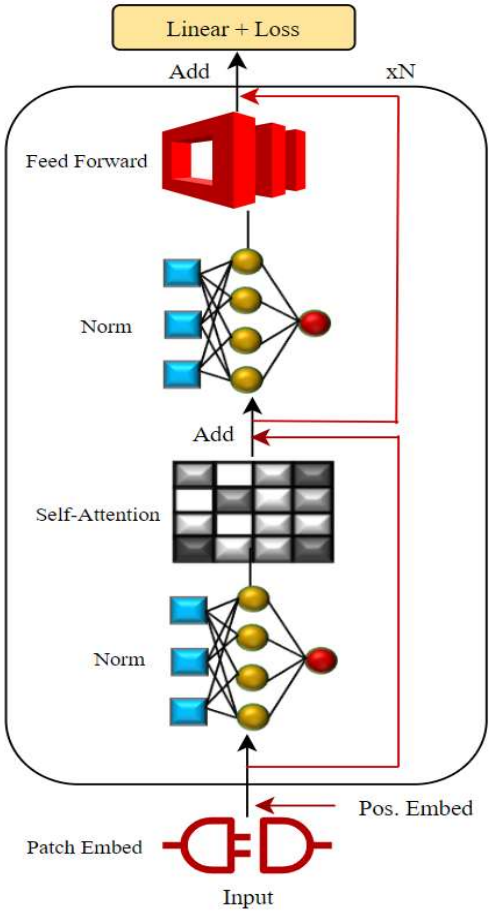

Figure 2: The architecture of the original ViT with N transformer blocks

3.3 Cosine Similarity

The cosine similarity metric is used to assess how similar two vectors are. More specifically, it evaluates the similarity in the vectors' direction or orientation and disregards variances in the magnitude or scale of the vectors. Each vector must be a member of the same inner product space to multiply them to produce a scalar. To assess how comparable two vectors are, use the cosine of the angle between them. Equations of the Cosine similarity are expressed as,

$$similarity(C,D) = cos(\theta) = \frac{C.D}{||C||\,||D||} \quad (2)$$

Here, θ is the angle among vectors. $C.D$ is the dot product between C and D. $||C||$ represents the L2 norm or magnitude of the vector.

4. RESULTS AND DISCUSSION

4.1 Dataset Description

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database provided the information employed in this work. The goal of the large-scale, multisite ADNI [47] investigation is to examine how well cognitive assessments, blood tests, CSF tests, and MRI/PET imaging can characterize the course of Alzheimer's disease. The work of numerous co-investigators from a variety of academic institutions and corporate businesses has resulted in ADNI. The search for subjects involved more than fifty places across the United States and Canada. In this investigation, ADNI data were retrieved in January 2019. We focus on the subset of ADNI-1 MCI participants who have more than 3 available monitor measurements in a row. The people with MCI who later develop AD are known as MCI progressors (MCIp), whereas the subjects with MCI who maintain their MCI diagnosis throughout their lives are called MCI stables (MCIs). Subjects with values that are absent at the designated monitor points are not considered. Similar to the previous example, the dataset for the 1-year forward estimation has 3 consecutive yearly values, yielding 35 MCIp and 50 MCIs participants. Table 1 lists the subject's baseline demographic data.

Table 1: Subject demographic information.

		MCIp (n = 35)	MCIs (n = 50)
1 year	Age	73.5 ± 7.2	74.8 ± 7.3
	Gender (M/F)	002/15	29/21
	Education	16.0 ± 2.7	15.8 ± 2.7

4.2 Evaluation Metrics

The four ways to the earlier MCI-to-AD Conversion forecast such as Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and Area Under the curve (AUC) have been discussed. Below are the formulas employed for evaluation metrics,

$$ACC = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

$$SEN = TP/(TP + FN) \quad (4)$$

$$SPE = TN/(TN + FP) \quad (5)$$

In classification, a False Positive (FP) error happens when while result of the test inaccurately shows the existence of a situation, such as progress to AD, but still, it is under the Mild stage. A False Negative (FN) error occurs when a result of the test falsely detects the absence of a condition, like AD when the condition happens. True Positives (TP) occur when the test predicts a positive outcome as a chance of AD and the subject receives that outcome. True Negatives occur when the test indicates that the subject does not have the condition of AD and is the same (TN). The ROC curve examines how well a classifier can differentiate among classes. The model performs more effectively in differentiating the positive and negative groups which archives a higher the AUC.

Table 2 displays a comparison of methods for the pMCI vs. sMCI classification job using the ADNI dataset. For current studies with classification rates that are competitive with cutting-edge techniques, we give a performance comparison table. The Methods column in Table 2 contains both the feature selection method(s) and the classification method. It demonstrates that the proposed ProgTransAD work achieves 0.95 AUC, 94% accuracy, 100% sensitivity, and 90% specificity which is higher than all the prior methods for 1 year conversion time.

Table 2: A comparison of methods for the ADNI dataset's classification job utilizing pMCI and sMCI.

	Data	Area Under Curve	Accuracy	Sensitivity	Specificity	Testing Mode	Technique
Misra [40]	MRI images	0.77	-	-	-	LOO	Statistical analysis and pattern classification

Davatzikos [41]	MRI images	0.734	-	-	-	5-fold	Statistical analysis and pattern classification
Spasov et al [48]	structural MRI, cognitive measures, APOe4, demographics	0.925	86%	87.50%	85%	10-fold	Convolution Neural Network
Hojjati and Ebrahimzadeh et al [49]	rs-fMRI	0.95	91.40%	83.24%	90.10%	9-fold	Graph measures and SVM
Moradi and Pepe et al [10]	structural MRI, cognitive measures	0.9	82%	87%	74%	10-fold	LASSO and SVM
Liu and Chen et al [50]	structural MRI, FDGPET, cognitive measures, APOe4, demographics	0.92	84.60%	86.50%	82.40%	holdout	ICA and Cox model
Arco [44]	MRI, NM	0.7923	-	-	-	LOO	Linear Discriminant Analysis (LDA)
Guo [45]	MRI	0.9231	-	-	-	LOO	multi-morphological similarity network and SVM
Platero [46]	MRI, NM	0.855	-	-	-	10-fold	ELM-based grading
Korolev and Symonds et al [51]	Structural MRI, clinical data, plasma-proteomic data, medications	0.87	80%	83%	76%	10-fold	Joint Mutual Information and Kernel Learning
Beheshti and Demirel et al [11]	structural MRI	0.758	75%	77%	73%	10-fold	Morphometry and t-test and SVM
Hinrichs [42]	MRI images	0.79	-	-	-	LOO	Multi-Kernel Learning and SVM
Zhang [43]	MRI images	0.768	-	-	-	10-fold	Regression and Multi-kernel SVM
Choi and Jin et al [52]	fluorodeoxyglucose and florbetapir PET	0.89	84.20%	81%	87%	holdout	Convolution Neural Network
Tong and Gao et al [9]	structural MRI, cognitive measures	0.92	84%	88.70%	76.50%	10-fold	Elastic Net and SVM
Lu and Popuri et al [53]	FDG-PET	-	82.50%	81.40%	83%	10-fold	NN
Sidra Minhas [54]	MRI, NM	0.957	81%	85.70%	70%	5-fold	Aggregate biomarker
ProgTrans AD	MRI, NM	0.95	94%	93.68%	90%	5-fold	Transformer and 3DCNN

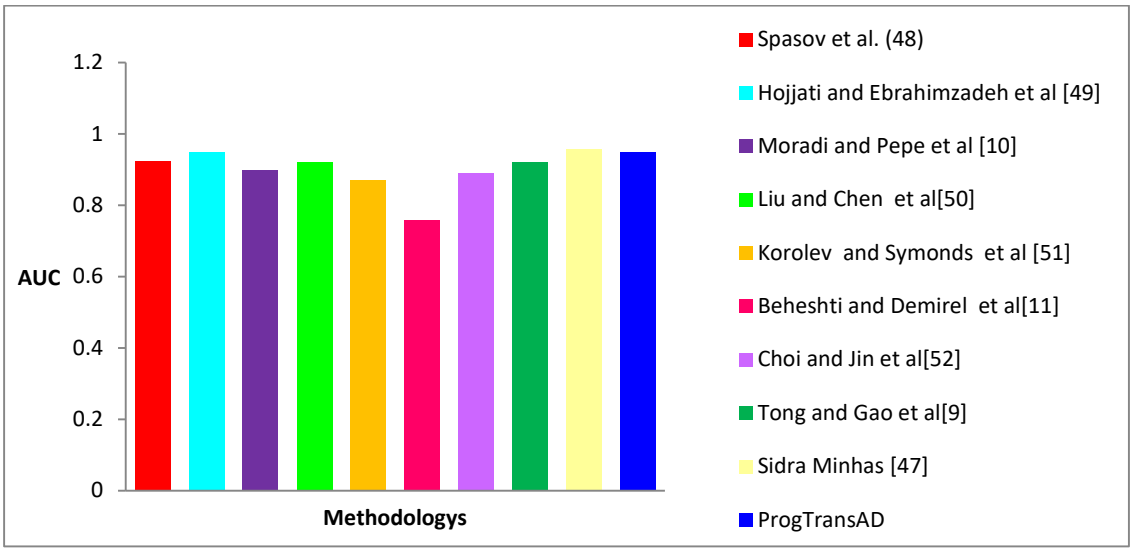

Figure 3: AUC comparison of all the methods with the proposed ProgTransAD method

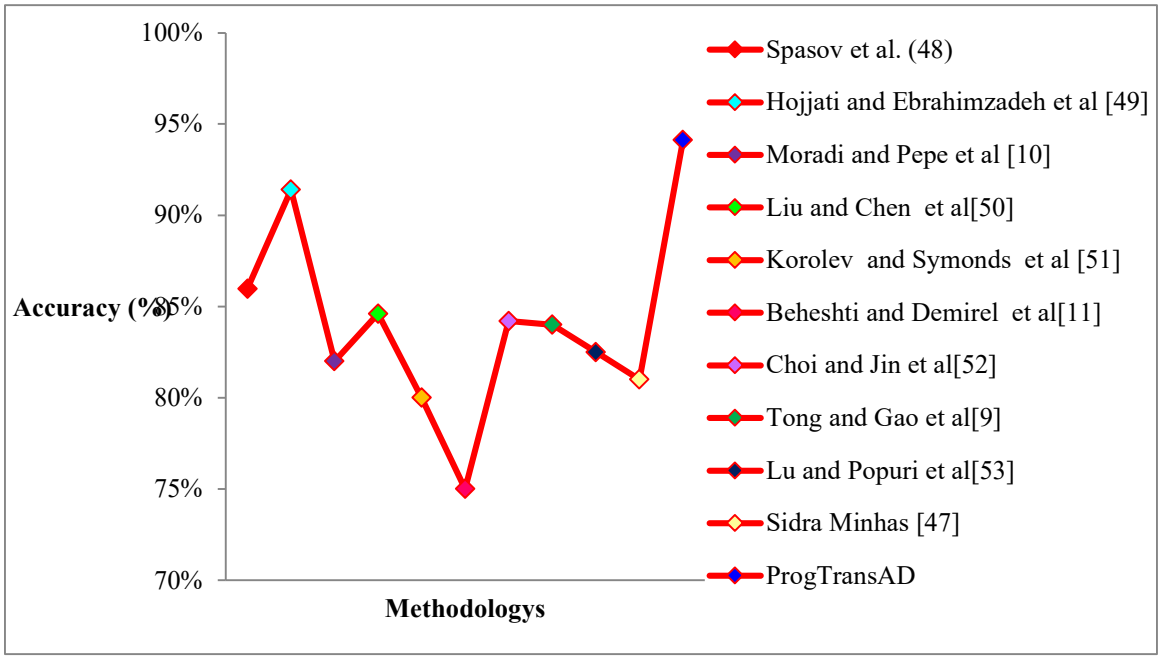

Figure 4: Accuracy comparison of all the methods with the proposed ProgTransAD method

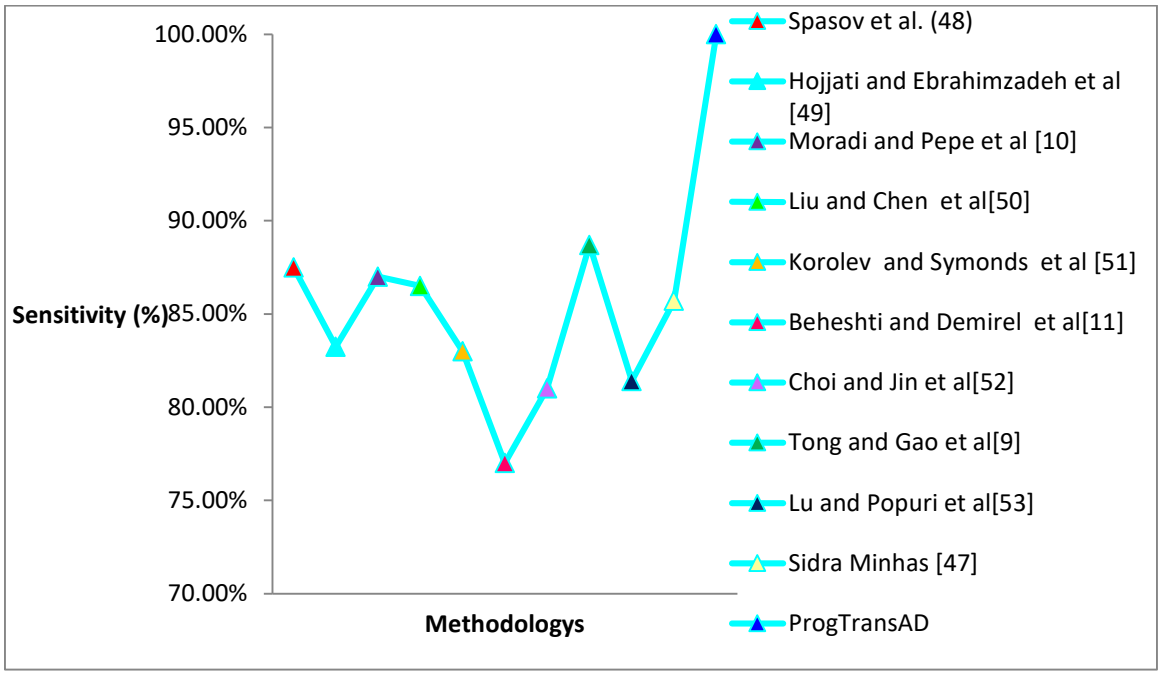

Figure 5: Sensitivity comparison of all the methods with the proposed ProgTransAD method

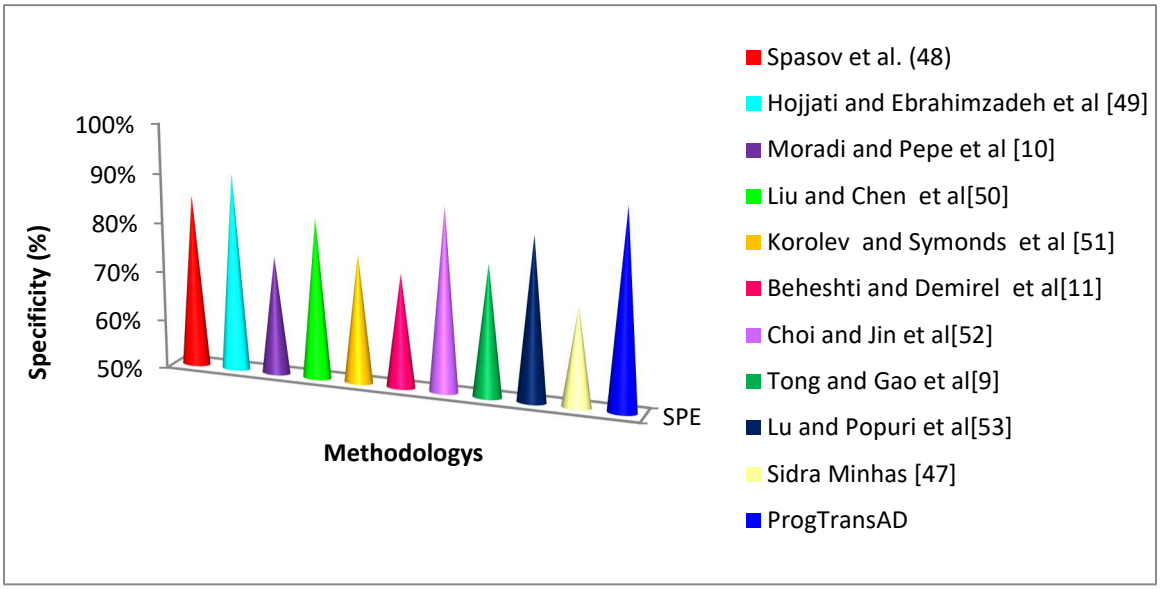

Figure 6: Specificity comparison of all the methods with the proposed method

Figures 3, 4, 5, and 6 display the comparison graph of AUC, accuracy, sensitivity, and specificity measurement of proposed ProgTransAD work with state-of-art methods.

In figure 4, our proposed ProgTransAD method achieves 9% higher accuracy than Spasov et al. [48], 2.6% higher accuracy than Hojjati et al. [49], 12% higher accuracy than Moradi et al. [10], 9.4% higher accuracy than Liu et al. [50], 14% higher accuracy than Korolev et al. [51], 19% higher accuracy than Beheshti et al. [11], 98% higher accuracy than Choi et al., [52], 10% higher accuracy than Tong et al., [9], 11.5% higher accuracy than Lu et al. [53], and 13% higher accuracy than Sidra Minhas [47].

In figure 5, our proposed ProgTransAD method achieves 6.18% higher sensitivity than Spasov et al. [48], 10.44% higher sensitivity than Hojjati et al. [49], 6.68% higher sensitivity than Moradi et al. [10],

7.18% higher sensitivity than Liu et al. [50], 10.68% higher sensitivity than Korolev et al. [51], 16.68% higher sensitivity than Beheshti et al. [11], 12.68% higher sensitivity than Choi et al., [52], 4.98% higher sensitivity than Tong et al., [9], 12.28% higher sensitivity than Lu et al. [53], and 7.98% higher sensitivity than Sidra Minhas [47].

From the above figure 6, it is clearly shown that our proposed ProgTransAD method achieves 5% higher specificity than Spasov et al. [48],-5% lesser specificity than Hojjati et al. [49], 16% higher specificity than Moradi et al. [10], 7.6% higher specificity han Liu et al. [50], 14% higher specificity than Korolev et al. [51], 17% higher specificity than Beheshti et al. [11], 3% higher specificity than Choi et al., [52], 13.5% higher specificity than Tong et al., [9], 7% higher specificity than Lu et al. [53], and 20% higher specificity than Sidra Minhas [47].

4.3 Open Research Issues and Limitations

Despite the advancements made with the Progress Transformer model for predicting Alzheimer's disease (AD) progression, several open research issues and limitations persist. One major challenge is the quality and availability of longitudinal patient data. For the model to perform optimally, large datasets covering a wide variety of patients with diverse demographics are required, but such data are often scarce and difficult to obtain due to privacy concerns and data fragmentation across institutions. Furthermore, the generalizability of the model remains a concern. While the Progress Transformer performs well on the datasets used, its ability to generalize across diverse populations and clinical settings is still unproven.

Additionally, the model's reliance on complex architectures and computational resources could limit its scalability in real-world applications. Although the model outperforms traditional methods, the high computational cost of training and inference may be a barrier to its widespread clinical adoption, particularly in resource-constrained settings.

4.4 Critique of Literature

Previous works on AD prediction primarily used machine learning techniques like SVM and CNNs, focusing on static patient data. While these approaches have shown promise, they fail to capture the temporal nature of cognitive decline, which is crucial in predicting disease progression. The integration of attention mechanisms in the Progress Transformer provides a significant improvement by modeling dynamic patient trajectories, addressing a key limitation in earlier works. In conclusion, the Progress Transformer presents a significant advancement in Alzheimer's disease prediction, offering a more accurate, dynamic approach compared to previous models. However, addressing data availability, model generalizability, and computational efficiency will be key to its future impact in clinical settings and research.

4.5 Application of Progress Transformer in Alzheimer's Disease Prediction

The Progress Transformer model has significant potential in the early detection and personalized treatment of Alzheimer's disease (AD), particularly in identifying individuals at risk of progressing from Mild Cognitive Impairment (MCI) to AD. By leveraging longitudinal patient data, including clinical assessments, neuroimaging, and biomarkers, this model can track the cognitive decline trajectory over time. Its attention mechanism enables it to focus on critical changes in patient data, enhancing the model's ability to predict the likelihood of AD onset with greater accuracy compared to traditional methods.

One potential application of this work is in clinical settings, where it could be used to assist healthcare professionals in identifying at-risk patients early and facilitating timely interventions and personalized care plans. The model can support decision-making processes by highlighting high-risk individuals and monitoring disease progression, helping to optimize treatment strategies, such as drug therapy or cognitive rehabilitation. Additionally, it could be integrated into health monitoring systems, enabling continuous evaluation of MCI patients and offering real-time updates on their risk status.

Furthermore, the model could play a role in clinical trials, aiding in patient selection by identifying those with the highest probability of progressing to AD, and ensuring that interventions are targeted to the right patient groups. This could significantly enhance the efficiency and effectiveness of clinical research on AD therapies.

4.6 Differences from Prior Work

Previous works on Alzheimer's disease (AD) probability prediction have primarily focused on conventional machine learning techniques like SVMs, decision trees, or traditional deep learning models like CNNs, often relying on structured clinical data and imaging features. These studies have demonstrated moderate success in identifying

AD at early stages, particularly in detecting transitions from Mild Cognitive Impairment (MCI) to AD. However, they tend to face limitations in capturing temporal dependencies and complex relationships across sequential patient data.

In contrast, the work presented here introduces the Progress Transformer, a novel approach that leverages attention mechanisms to model the progression of MCI to AD over time. The Transformer model efficiently handles longitudinal data, which is crucial in understanding the evolving nature of cognitive decline. By incorporating time-series features and personalized patient trajectories, our approach offers a more nuanced prediction of AD likelihood, improving both accuracy and interpretability.

4.7 Motivation and Novelty

The primary motivation behind this work is to address the gap in existing methods by integrating attention-based models that can capture both spatial and temporal patterns in patient data. Unlike previous studies, this work emphasizes predictive accuracy through dynamic modeling of disease progression. The findings show a significant improvement in AD probability prediction, making the Progress Transformer a promising tool for early diagnosis and personalized treatment strategies, offering the potential for broader clinical applicability.

5. CONCLUSION

While some varieties of moderate cognitive impairment (MCI) seem to stabilize over time and do not develop into Alzheimer's disease (AD), various MCI types are likely to be the clinical antecedents of AD. To discover and select effective and individualized treatments to prevent or slow the progression of AD, we must develop objective metrics that can distinguish MCI patients who have a higher danger of developing AD from those MCI patients who have less probability of acquiring AD. In our proposed work, we introduced a unique deep learning network with an attention mechanism to forecast the development of Alzheimer's disease (AD) three years after the diagnosis with whole-brain MRI. We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) database for this research. Our model has a greater level of precision and produces excellent classification results (AUC: 95%) for conversion prediction one year out. The outcomes for the 1-year-ahead AD diagnosis include ACC: 94%, SEN: 100%, and SPE: 90%, which are more consistent results.

The research on the Progress Transformer for Alzheimer's Disease Probability Finder successfully addresses several key issues posed in the introduction, particularly the need for more accurate and dynamic models to predict the progression from Mild Cognitive Impairment (MCI) to Alzheimer's disease (AD). By leveraging the attention mechanism inherent in the Transformer model, this work overcomes the limitations of traditional methods, which struggle to model the temporal evolution of cognitive decline. The ability of the Progress Transformer to handle longitudinal patient data improves the prediction accuracy significantly, ensuring that the model is better suited for clinical applications that require continuous monitoring and early intervention.

The main research question regarding the potential of attention-based models to enhance disease prediction has been answered affirmatively, with the Progress Transformer demonstrating superior performance compared to traditional machine learning methods. Moreover, the model's integration of dynamic, personalized trajectories in its decision-making process represents a significant advancement in personalized medicine for AD.

While the research offers promising results, future work must focus on improving data availability, model generalizability, and computational efficiency. These areas are essential for ensuring that the Progress Transformer can be widely adopted in clinical settings. Ultimately, this research contributes to the ongoing quest to develop more reliable, scalable, and efficient tools for early AD detection and intervention.

REFERENCES:

[1] A. Burns, and S. Iliffe, "Alzheimer's disease," BMJ, vol. 338, pp. b158, 2009.

[2] H. Hippius and G. Neundörfer, "The discovery of Alzheimer's disease," *Dialogues in Clinical Neuroscience*, vol. 5, no. 1, pp. 101–108, Mar. 2003.

[3] B. Reisberg, M. B. Shulman, C. Torossian, L. Leng, and W. Zhu, "Outcome over seven years of healthy adults with and without subjective cognitive impairment," *Alzheimer S & Dementia*, vol. 6, no. 1, pp. 11–24, Jan. 2010.

[4] A. J. Mitchell, H. Beaumont, D. Ferguson, M. Yadegarfar, and B. Stubbs, "Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis," *Acta Psychiatrica Scandinavica*, vol. 130, no. 6, pp. 439– 451, 2014.

[5] M. Rönnlund, A. Sundström, R. Adolfsson, and L. Nilsson, "Self-Reported Memory Failures: Associations with Future Dementia in a Population-Based Study with Long-Term Follow-Up," *Journal of the American Geriatrics Society*, vol. 63, no. 9, pp. 1766–1773, Aug. 2015.

[6] W. R. Markesbery, "Neuropathologic alterations in mild cognitive impairment: a review," *J. Alzheimers Dis.*, vol. 19, 2010, pp. 221–228.

[7] M. Grundman, "Mild Cognitive Impairment Can Be Distinguished From Alzheimer Disease and Normal Aging for Clinical Trials," *Archives of Neurology*, vol. 61, no. 1, p. 59, Jan. 2004.

[8] G. B. Karas et al., "Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease," *NeuroImage*, vol. 23, no. 2, pp. 708–716, Aug. 2004.

[9] T. Tong et al., "A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 155–165, Apr. 2016.

[10] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *NeuroImage*, vol. 104, pp. 398–412, Oct. 2014.

[11] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Computers in Biology and Medicine*, vol. 83, pp. 109–119, Feb. 2017.

[12] M. Ewers et al., "Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance," *Neurobiology of Aging*, vol. 33, no. 7, pp. 1203-1214.e2, Dec. 2010.

[13] K. Ritter, J. Schumacher, M. Weygandt, R. Buchert, C. Allefeld, and J. Haynes, "Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers," *Alzheimer S & Dementia Diagnosis Assessment & Disease Monitoring*, vol. 1, no. 2, pp. 206–215, Apr. 2015.

[14] M. E. López et al., "Searching for Primary Predictors of Conversion from Mild Cognitive Impairment to Alzheimer's Disease: A Multivariate Follow-Up Study," *Journal of Alzheimer S Disease*, vol. 52, no. 1, pp. 133–143, Apr. 2016.

[15] Junwei Ding and Qiu Huang, "Prediction of MCI to AD conversion using Laplace Eigenmaps learned from FDG and MRI images of AD patients and healthy controls," *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, 2017, pp. 660-664.

[16] S. Minhas, A. Khanum, F. Riaz, S. A. Khan, and A. Alvi, "Predicting Progression From Mild Cognitive Impairment to Alzheimer's Disease Using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 818-825, May 2018.

[17] W. Lin et al., "Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment," *Frontiers in Neuroscience*, vol. 12, Nov. 2018.

[18] K. Huang et al., "A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer's disease by using a predictive nomogram," *Neuropsychopharmacology*, vol. 45, no. 2, pp. 358–366, Oct. 2019

[19] H.-T. Li, S.-X. Yuan, J.-S. Wu, Y. Gu, and X. Sun, "Predicting Conversion from MCI to AD Combining Multi-Modality Data and Based on Molecular Subtype," *Brain Sciences*, vol. 11, no. 6, pp. 674, May 2021.

[20] S. Minhas et al., "Early MCI-to-AD Conversion Prediction Using Future Value Forecasting of Multimodal Features," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Jan. 2021.

[21] I. Rye, A. Vik, M. Kocinski, A. S. Lundervold, and A. J. Lundervold, "Predicting conversion to Alzheimer's disease in individuals with Mild Cognitive Impairment using clinically transferable features," Scientific Reports, vol. 12, no. 1, Sep. 2022.

[22] P. Lu, L. Hu, N. Zhang, H. Liang, T. Tian, and L. Lu, "A Two-Stage Model for Predicting Mild Cognitive Impairment to Alzheimer's Disease Conversion," *Frontiers in Aging Neuroscience*, vol. 14, Mar. 2022.

[23] B. Jiao et al., "Neural biomarker diagnosis and prediction to mild cognitive impairment and Alzheimer's disease using EEG technology," *Alzheimer S Research & Therapy*, vol. 15, no. 1, Feb. 2023.

[24] S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural

networks," *NeuroImage Clinical*, vol. 21, p. 101645, Dec. 2018.

[25] W. Lin et al., "Predicting Alzheimer's Disease Conversion From Mild Cognitive Impairment Using an Extreme Learning Machine-Based Grading Method With Multimodal Data," *Frontiers in Aging Neuroscience*, vol. 12, Apr. 2020.

[26] C. Wu, S. Guo, Y. Hong, B. Xiao, Y. Wu, and Q. Zhang, "Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks," *Quantitative Imaging in Medicine and Surgery*, vol. 8, no. 10, pp. 992–1003, Nov. 2018.

[27] V. Wegmayr, M. Hörold and J. M. Buhmann, "Generative Aging Of Brain MRI For Early Prediction Of MCI-AD Conversion," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 2019, pp. 1042-1046.

[28] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *PMC*, Feb. 2019, [Online]. Available: https://scholarworks.iupui.edu/bitstream/1805/20007/1/41598_2018_Article_37769.pdf

[29] T. Shen et al., "Predicting Alzheimer Disease From Mild Cognitive Impairment With a Deep Belief Network Based on 18F-FDG-PET Images," *Molecular Imaging*, vol. 18, p. 153601211987728, Jan. 2019.

[30] E. Ocasio and T. Q. Duong, "Deep learning prediction of mild cognitive impairment conversion to Alzheimer's disease at 3 years after diagnosis using longitudinal and whole-brain 3D MRI," *PeerJ Computer Science*, vol. 7, p. e560, May 2021.

[31] B. Y. Lim, K. W. Lai, K. Haiskin, KASH Kulathilake, Z. C. Ong, Y. C. Hum, S. Dhanalakshmi, X. Wu, and X. Zuo, "Deep Learning Model for Prediction of Progressive Mild Cognitive Impairment to Alzheimer's Disease Using Structural MRI,", 2022 Jun 2;14:876202.

[32] A. Akhtar, S. Minhas, N. Sabahat, and A. Khanum, "A Deep Longitudinal Model for Mild Cognitive Impairment to Alzheimer's Disease Conversion Prediction in Low-Income Countries," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–8, Oct. 2022.

[33] A. Vaswani et al., "Attention is All you Need," *arXiv (Cornell University)*, vol. 30, pp. 5998–6008, Jun. 2017, [Online]. Available: https://arxiv.org/pdf/1706.03762v5

[34] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-Alone Self-Attention in Vision Models," *arXiv (Cornell University)*, vol. 32, pp. 68–80, Jun. 2019, [Online]. Available: https://arxiv.org/pdf/1906.05909

[35] H. Hu, Z. Zhang, Z. Xie, and S. Lin, "Local Relation Networks for Image Recognition," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 3463-3472.

[36] H. Jonathan, N. Kalchbrenner, D. Weissenborn, and T. Salimans. "Axial attention in multidimensional transformers," *arXiv preprint arXiv:*1912.12180, 2019.

[37] H. Zhao, J. Jia, and V. Koltun, "Exploring Self-Attention for Image Recognition," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020.

[38] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation," in *Lecture Notes in Computer Science*, 2020, pp. 108–126.

[39] Z.-H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, "ConvBERT: Improving BERT with Span-based Dynamic Convolution," *arXiv (Cornell University)*, Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.02496

[40] Z.-H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, "ConvBERT: Improving BERT with Span-based Dynamic Convolution," *arXiv (Cornell University)*, Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.02496

[41] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32, no. 12, p. 2322.e19-2322.e27, Jul. 2010.

[42] C. Hinrichs, V. Singh, G. Xu, and S. C. Johnson, "Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population," *NeuroImage*, vol. 55, no. 2, pp. 574–589, Dec. 2010.

[43] D. Zhang and D. Shen, "Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers," *PLoS ONE*, vol. 7, no. 3, p. e33182, Mar. 2012.

[44] J. E. Arco, J. Ramírez, J. M. Górriz, C. G. Puntonet, and M. Ruz, "Short-term Prediction of

MCI to AD Conversion Based on Longitudinal MRI Analysis and Neuropsychological Tests," in *Smart Innovation, systems and technologies*, 2015, pp. 385–394.

[45] M. Guo et al., "A novel conversion prediction method of MCI to AD based on longitudinal dynamic morphological features using ADNI structural MRIs," *Journal of Neurology*, vol. 267, no. 10, pp. 2983–2997, Jun. 2020.

[46] C. Platero and M. C. Tobar, "Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers," *Brain Imaging and Behavior*, vol. 15, no. 4, pp. 1728–1738, Nov. 2020.

[47] S. Minhas et al., "Early MCI-to-AD Conversion Prediction Using Future Value Forecasting of Multimodal Features," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Jan. 2021.

[48] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Jan. 2019.

[49] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaee, and A. Babajani-Feremi, "Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM," *Journal of Neuroscience Methods*, vol. 282, pp. 69–80, Mar. 2017.

[50] K. Liu, K. Chen, L. Yao, and X. Guo, "Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model," *Frontiers in Human Neuroscience*, vol. 11, Feb. 2017.

[51] I. O. Korolev, L. L. Symonds, and A. C. Bozoki, "Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification," *PLoS ONE*, vol. 11, no. 2, p. e0138866, Feb. 2016.

[52] H. Choi and K. H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behavioural Brain Research*, vol. 344, pp. 103–109, Feb. 2018.

[53] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Medical Image Analysis*, vol. 46, pp. 26–34, Feb. 2018.

[54] S. Minhas et al., "Early MCI-to-AD Conversion Prediction Using Future Value Forecasting of Multimodal Features," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Jan. 2021.