Multimodal Classification for Alzheimer's Disease Diagnosis and Progression Prediction using Deep Learning

Prasanth Bathala
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
pbathala3@gatech.edu

Janavi Khochare
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
janavikhochare@gatech.edu

Yadhu Kartha
Machine Learning
Georgia Institute of Technology
Atlanta, GA, USA
yadhukartha@gatech.edu

Hemanth Tammana
Interactive Computing
Georgia Institute of Technology
Atlanta, GA, USA
htammana3@gatech.edu

ABSTRACT

Alzheimer's Disease (AD) is a neurodegenerative disorder that leads to cognitive decline and is the leading cause of dementia in the elderly. Accurate and early diagnosis is crucial to improve patient outcomes, and computer-aided diagnosis is becoming an essential tool for screening at-risk individuals. In this study, we propose a deep-learning based classifier that can classify multi-modal imaging data. Our approach involves using a modified ResNet50 architecture, combined with late fusion of Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) images, to classify patients into one of three stages of AD. We also developed two separate models to identify the risk of AD and monitor disease progression. By addressing these challenges, our research aims to provide clinicians with advanced tools for accurate diagnosis and effective monitoring of disease progression, ultimately contributing to improving patient outcomes and advancing our understanding of this devastating disease.

CCS CONCEPTS

• Applied Computing • LIfe and Medical Sciences • Deep Learning • Medical Image Processing

KEYWORDS

Alzheimer's Disease, MRI, PET, EHR, ResNet50, LSTM, GRU, BiLSTM, BiGRU

ACM Reference format:

Prasanth Bathala, Janavi Khochare, Hemanth Tamanna, and Yadhu Kartha. 2023. **Multimodal Classification for Alzheimer's Disease Diagnosis and Progression Prediction using Deep Learning**

1 Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that affects cognitive function and daily activities, with a significant impact on patients and caregivers [1]. Current estimates indicate that over 50 million people worldwide may have AD or other forms of dementia, and the total cost of healthcare for individuals with dementia aged 65 and over has reached $305 billion in 2020 [2]. By 2050, the number of AD patients is projected to reach 115 million, underscoring the critical need for early and accurate diagnosis and treatment of this condition.

The pathogenesis of AD is not fully understood, but it is thought to be linked to the accumulation of extracellular amyloid-β (Aβ) and neurofibrillary tangles, resulting in neuronal and synaptic loss or damage [3]. Neuropsychological and neuroimaging examinations are the primary clinical tools for assessing AD [4]. Computer-aided diagnosis is an essential tool for screening at-risk individuals. Structural MRI is a widely used neuroimaging technique in AD diagnosis due to its high resolution for soft tissue and ability to display brain anatomical details[5]. PET imaging plays a crucial role as a functional technique, enabling clinicians to observe human brain activities with particular applications in early AD detection[6, 7, 8].

To address the challenges associated with accurately predicting AD in patients and identifying the risk of AD and monitoring disease progression, we propose a deep-learning-based approach that can classify multi-modal imaging data. Our approach involves using a modified ResNet50 architecture, combined with late fusion of MRI and PET images, to classify patients into one of three stages of AD. Additionally, we developed two separate models to predict the risk of AD at the next time step and for future multiple time visits using an encoder-based approach and an encoder-decoder approach, respectively. For disease progression, we utilized Electronic Health Record (EHR) data of the patient.

Integrating multiple data types can improve AD prediction. Causal inference can identify risk factors, aiding targeted interventions. Real-time ML models can aid in personalized treatment plans, leading to better patient outcomes. By addressing both of these challenges, our research aims to provide clinicians with advanced tools for accurate and early diagnosis of AD, as well as effective monitoring of disease progression. Ultimately, our hope is that this work will contribute to improving patient outcomes and advancing our understanding of this devastating disease.

1.1 Previous Related Works

In terms of machine learning models, previous studies [23] have implemented classification methods based on combination of multi-model 3D convolutional networks to learn the various features from MRI brain images. The features obtained from the 3D CNN network are combined with the set of features extracted using 3D convolutional autoencoders (3D CAEs). The model structure was complex involving the usage of multiple architectures and was implemented only for normal and AD patients. To address this issue, recent studies [22] have explored the use of more efficient models such as UNet for AD classification and prediction. However, these models often require a large number of parameters and are computationally expensive. In this study, we propose a novel approach for AD classification that combines both MRI and PET imaging modalities. Specifically, we use a modified version of ResNet with fewer parameters than previous studies, while achieving high accuracy. Our results demonstrate the potential benefits of multimodal imaging and efficient machine learning models for AD diagnosis and prediction. While most of the previous research works highly depend on image preprocessing steps such as skull stripping, segmentation and registration which requires domain expert knowledge, our proposed method does not require any complex preprocessing steps to get high accuracy.

Hu et. Al. discussed about using the latest state-of-the-art transformers and recurrent network with attention, and has achieved around 77.30% accuracy which is pretty less compared to benchmark results till now, which indicates the MRI biomedical images are not sufficient enough to train the transformer. Guan et.al discussed the prediction of Alzheimer's stages using multimodal data (MRI, EHR, SNP) and gave details about the preprocessing methods for all modalities [9]. Using 3D Brain MRI images as input for 3D CNN architectures offer better transferability of the CNNs and have shown improvement in performance and better generalization ability [10]. Other studies include a deep convolutional neural network (CNN) based method, could accurately predict cognitive decline in patients and achieve better generalization[11].

As far as classification, there is significant research going to estimate the risk of AD in the future. Previous studies have used single-modality magnetic resonance imaging (MRI) or positron emission tomography (PET) to classify and predict AD progression. However, the limitations of single-modality imaging, such as low sensitivity and specificity, have motivated the development of multimodal approaches.

In the study by [12], unsupervised and supervised machine learning techniques were applied to the ADNI dataset to learn the progression and clinical subtypes of Alzheimer's disease. The study formed clusters of people with varying levels of Alzheimer's disease progression and classified them into low, moderate, and high progression zones. However, the approach lacks personalization as the output for each cluster indicates the progression level of the cluster, rather than individual patients. Therefore, further research is necessary to incorporate more complex and personalized machine learning approaches for Alzheimer's disease progression modeling.

The study by [13] introduces a novel external memory network (EMN), TC-EMNet, for disease progression modeling using temporal clustering and external memory mechanisms to capture temporal patterns and dynamics. TC-EMNet leverages a variation autoencoder framework and a memory network to address data irregularities and long-term dependency issues in recurrent neural networks. Although the proposed approach compared to baseline studies, it does not compare to human-effective methods, which is crucial in medical data. Therefore, further research is required to evaluate the effectiveness of the proposed method against human-effective methods and address the data limitations in medical applications. Overall, the study presents a promising approach for disease progression modeling, but further research is necessary to validate its effectiveness in medical applications.

In the quest for early diagnosis and intervention of Alzheimer's disease (AD), the study by [14] presents two deep learning architectures, PPAD and PPAD-AE, designed to predict the progression of AD using MRI, cognitive tests, and PET biomarkers. The study utilizes long short-term memory (LSTM) and gated recurrent unit (GRU) to model the temporal sequence of data and an autoencoder/decoder to predict AD progression. However, the paper is limited in its use of restricted data and overlooks crucial features such as the time between patients' visits. To address these limitations, the present study aims to improve the accuracy of AD progression prediction by incorporating hyperparameter tuning and considering the time between patients' visits as a feature. Overall, the paper presents a promising approach to predicting AD progression, but further advancements are necessary to address the limitations in the data and features used for modeling.

2 Methodology and System Design

2.1 Dataset

The data utilized in this project were obtained from theAlzheimer’s Disease Neuroimaging Initiative (ADNI) dataset (https://adni.loni.usc.edu/). ADNI is a longitudinal multicenter study that aims to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. To promote AD diagnosis and treatment, ADNI makes all data and samples available for scientists worldwide (27, 28). The dataset contains multimodal data analyzes mainly from North American participants and covers different AD stages. For this study, we selected subjects who had both T1-weighted MRI and FDG-PET scans captured in the same period. MRI scans labeled as MPRAGE were chosen as they are considered to be of the highest quality. The

clinical data in the ADNI dataset consists of Demographics, Cognitive Assessment Scores, and MRI Biomarkers. The Distribution of the data among the three modalities across the three classes is shown in the Table 1

Table 1: Distribution of classes

	CN	MCI	AD
MRI	200	396	133
PET	393	786	205
EHR	1900	1574	791

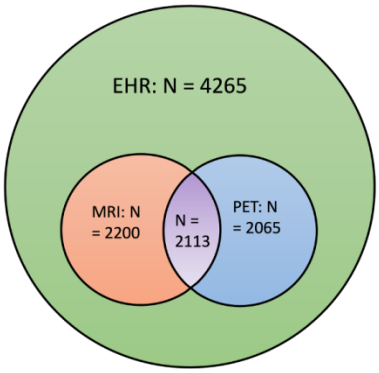

Fig 1. Venn diagram showing the data distribution

2.1.1 MRI Data

MRI imaging has become an essential tool for Alzheimer's disease (AD) research, providing non-invasive and accurate measurements of structural and functional changes in the brain. Alzheimer's disease is a progressive neurological disorder that affects memory, thinking, and behavior, and is characterized by the accumulation of beta-amyloid plaques and tau tangles in the brain. MRI imaging has been used to detect and track these pathological changes, as well as to identify structural changes associated with disease progression. Thus, it is used as an effective representative of Alzhimer's disease progression [16].

One of the most used Structural MRI [17] sequences in AD research is the T1-weighted sequence, which provides high-resolution images of brain anatomy. T1-weighted images can be used to detect changes in gray matter volume, which has been shown to decrease in AD patients as the disease progresses. MRI imaging can also be used to visualize beta-amyloid plaques and tau tangles directly, using specialized imaging agents that bind to these proteins. In recent years, MRI MP Rage (Magnetization Prepared Rapid Gradient Echo) imaging has emerged as a promising technique for AD research. MP Rage is a T1-weighted imaging sequence that uses a magnetization preparation step to increase contrast between gray and white matter. This technique has been shown to provide superior image quality and better detection of small changes in cortical thickness compared to conventional T1-weighted imaging.

2.1.2 PET Data

PET imaging is a powerful tool for visualizing and measuring biological processes in living tissues, including the brain. PET imaging with the radiotracer AV45 has become an essential tool for Alzheimer's disease (AD) research [18], as it helps to detect and track the accumulation of beta-amyloid plaques in the brain, a hallmark feature of AD. AV45 PET imaging has been used to study the progression of beta-amyloid deposition in the brain over time, identify individuals at risk for developing AD, and monitor the effects of interventions aimed at reducing amyloid accumulation.

AV45 is a radiotracer that binds specifically to beta-amyloid plaques and emits positrons that can be detected by PET scanners. By measuring the concentration of AV45 in different regions of the brain, the extent and distribution of beta-amyloid plaques in living AD patients can be determined. This imaging technique has the potential to improve early diagnosis and monitoring of AD, as well as to facilitate the development of new treatments for this devastating disease. PET imaging with AV45 is a valuable tool for AD research, providing a non-invasive and accurate way to visualize and quantify beta-amyloid accumulation in the brain [19].

2.1.3 Multimodal data

MRI and PET provide complementary information [20] about different aspects of brain structure and function. MRI can provide detailed structural information about brain regions and their integrity, while PET can provide information about functional activity, metabolism, and the presence of specific biomarkers associated with AD, such as amyloid plaques and tau tangles. Combining these modalities allows for a more comprehensive assessment of brain changes associated with AD, potentially leading to improved accuracy and reliability of AD detection. By using multiple modalities, the strengths of each modality can be leveraged to compensate for the limitations of others, leading to improved overall performance in detecting AD. For example, PET can provide high specificity in detecting the presence of amyloid plaques, which are a hallmark of AD, while MRI can provide information about brain atrophy, which is associated with disease progression.

Combining multiple modalities can enhance the sensitivity and specificity of AD detection [21, 22]. By using multiple modalities, the strengths of each modality can be leveraged to compensate for the limitations of others, leading to improved overall performance in detecting AD. For example, PET can provide high specificity in detecting the presence of amyloid plaques, which are a hallmark of AD, while MRI can provide

information about brain atrophy, which is associated with disease progression. Different modalities may be affected by different sources of variability, such as noise, artifacts, and biases. By combining multiple modalities, the impact of such variability can be mitigated, leading to more robust and reproducible results. This can improve the reliability and confidence of AD detection in research and clinical settings.

2.1.4 EHR Data

In this study, we utilized the TADPOLE longitudinal cohort3, which is part of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The TADPOLE challenge aimed to monitor the disease progression of individuals with Alzheimer's Disease through the analysis of various biomarkers such as MRI images, cognitive test scores, and clinical measurements. The cohort consists of data from ADNI-1, ADNI-2, and ADNI-GO. By leveraging this dataset, we aimed to gain insights into the disease progression and to develop accurate models that can aid in the early diagnosis and treatment of Alzheimer's Disease. The distribution of the data in the TADPOLE challenge is shown in the Table 2

Table 2: Distribution of ADNI dataset

Clinical Status	# of Subjects		# of Visits		Age (mean ± std dev)	
	Male	Female	Male	Female	Male	Female
CN	930	970	2126	2084	75.10 ± 5.64	74.42 ± 5.38
MCI	993	591	4207	2756	73.711 ± 7.11	71.80 ± 7.85
AD	410	381	884	684	75.56 ± 7.11	73.56 ± 7.11

To predict the AD status using the TADPOLE dataset, we focused on six volumetric features extracted from T1-weighted MRI scans: ventricles, hippocampus, fusiform gyrus, middle temporal gyrus, entorhinal cortex, and whole-brain, as well as cognitive test scores including MMSE, ADAS-cog11, and ADAS-cog13. These specific features were chosen based on previous studies. Despite the wealth of biomarkers available in the TADPOLE dataset, we chose to focus on these specific features based on [15].

2.2 System Block Diagram & Overall Architecture

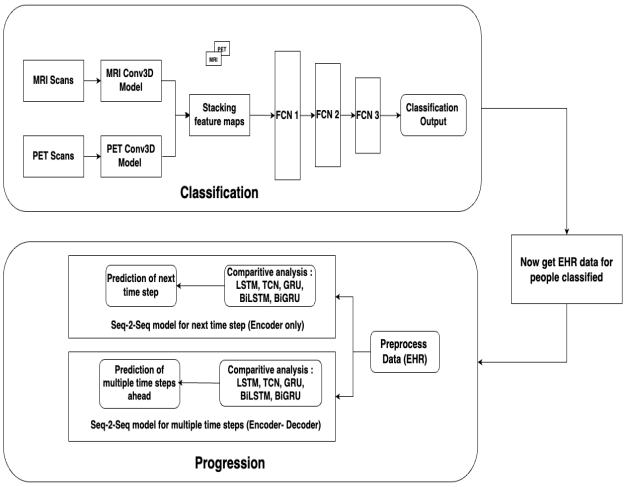

Fig 2. Complete architecture

The proposed architecture in our research paper consists of two modules: classification and progression. The classification module takes as inputs the MRI and PET scans of the patients, and employs 3D convolutional models for feature extraction from each scan. The extracted features are then combined using feature fusion and passed to a fully connected network for classification output prediction. The progression module takes the electronic health records (EHR) data of the patients as input, which is preprocessed using various techniques discussed in Section 2.5.1. The preprocessed data is then fed into two seq-2-seq models: one for predicting the next time step and the other for predicting multiple future time steps. The encoder layer of these models uses comparative analysis of LSTM, TCN, GRU, BiLSTM, and BiGRU, and predicts the progression of the disease accordingly.

2.4 Classification

2.4.1 Data Preprocessing for MRI and PET

ADNI dataset provides the preprocessed images obtained using methods such as N3 scaling and B1 normalization. B1 normalization and N3 scaling are two commonly used techniques in MRI and PET image processing, which aim to correct for intensity inhomogeneity in the images. B1 normalization corrects for variations in the radiofrequency (RF) field across the image. This normalization technique is particularly important for quantitative MRI analyses, such as voxel-based morphometry (VBM), where accurate measurement of gray matter, white matter, and cerebrospinal fluid is critical. N3 scaling corrects for intensity inhomogeneity caused by scanner or acquisition-related factors. These techniques are important preprocessing steps in MRI and PET image analysis as they help to improve the accuracy and reliability of subsequent image analyses. This technique is particularly useful for improving the quality of PET images, which often suffer from intensity inhomogeneity due to the limited sensitivity of the PET scanner. The resulting images were of different dimensions, hence, we rescaled the images to (1,60,128,128) so that it matches the input dimension of the model.

2.4.2 Algorithms for Classification

We drew inspiration from the resnet50 model, we modified the structure by changing the filter sizes to 256 and 512 from 1024 and 2056 to reduce the complexity and training time. We added dense layers in the end of the network with Relu and also sigmoid activation with 3 outputs to match the labels. The model was initialized with random weights. Random initialization is one way of performing symmetry breaking, which is the act of preventing all of the weights. As a result, symmetry is broken, and each neuron no longer performs the same computation, we can get better results each time we train the network.In the machine learning model from being the same. The Resnet model was proposed to solve the issue of diminishing gradient. The idea is to skip the connection and pass the residual to the next layer so that the model can continue to train. With Resnet models, CNN models can go deeper and deeper.This is one of the main reasons we went forward with Resnet. The model has 3 conv blocks , each block has 3 conv layers with increasing sizes of filters, 16,64,128,256 respectively. We have also used average pooling and dropout layers to counter overfitting. After hyper parameterizing epochs and filter sizes, we came to the conclusion to implement the filter sizes of 64,128,256 for deeper conv blocks and train it for 50 epochs. Our main goal was to reduce computation cost and improve performance of the classifier.This heatmap shows the distribution of accuracy based on tuning the hyperparameter for the classification model.

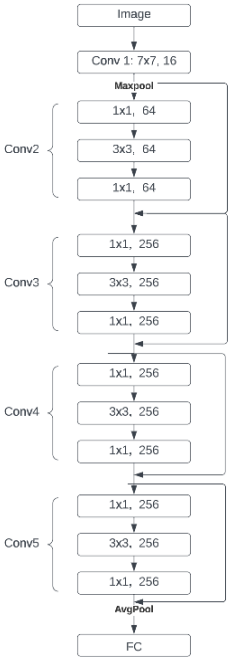

Fig.3 Classification architecture

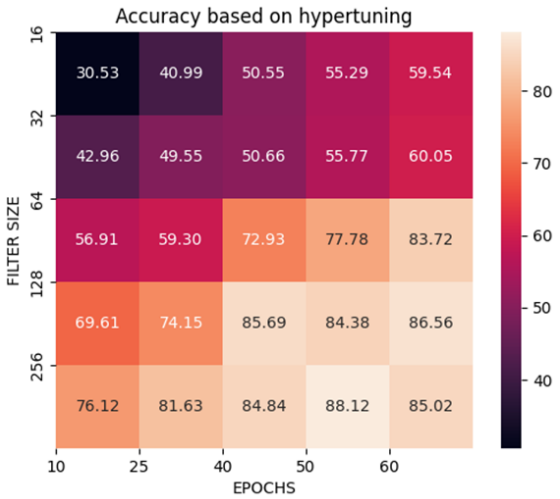

Fig. 4 Epochs vs filter size accuracy values

Feature concatenation [10] is often used when working with multimodal data because it allows us to combine features from multiple modalities into a single feature vector, which can then be used as input to a machine learning model. By doing this, we can leverage information from multiple modalities to improve the performance of the model. By using feature concatenation, the model can learn to recognize patterns in the combined feature vector that are not apparent when looking at the features from each modality separately. Previous works such as [24] proved that it can improve the performance of the model, especially when the modalities are complementary and provide different types of information.

2.4.3 Performance Metrics for classification and Regression

For the AD classification problem, performance metrics such as accuracy, precision, recall, and F1 score are used to evaluate the performance of the model. Accuracy is the most commonly used metric and is defined as the percentage of correctly classified instances. However, since the dataset is imbalanced (different number of images for each of the 3 classes (CN, MCI, AD), accuracy can be misleading and not reflect the true performance of the model.

Accuracy: (TP + TN) / (TP + TN + FP + FN)

Precision is the metric that measures the proportion of correctly predicted instances among the instances that were predicted to belong to a particular class. Recall, on the other hand, measures the proportion of correctly predicted instances among the instances that actually belong to a particular class. Precision and recall are especially important in an imbalanced dataset as they provide insights into how well the model is performing for the minority class.

Precision: TP / (TP + FP)

Recall: TP / (TP + FN)

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

The F1 score is a weighted average of precision and recall, which takes into account both metrics and provides a more balanced evaluation of the model. It is calculated as the harmonic mean of precision and recall, and ranges from 0 to 1, with higher values indicating better performance.

F1 score: 2 * (precision * recall) / (precision + recall)

2.5 Progression (Discuss the algorithms and architectures)

2.5.1 Data Preprocessing for EHR

In our study, we performed several preprocessing techniques to prepare the data for analysis. We normalized the volumetric MRI features using each subject's intracranial volume (ICV) to account for inter-subject variability in brain size. We also conducted feature-wise linear normalization based on the min/max values to standardize the range of the features. To handle irregular time intervals between visits, we added a new feature called Time Difference. We split the data into a training set and a test set in a ratio of 7:3.

We also observed that there were many missing observations in the MRI biomarkers and clinical labels of the TADPOLE longitudinal cohort dataset. To address this issue, we considered the mean imputation method, which was also used by Mohammad Al Olaimat et al. However, we also reviewed and compared different imputation techniques, such as zero imputation, most recent observation imputation, and k-nearest neighbor imputation, which were discussed by Wonsik Jung et al. Ultimately, we decided to try out the Miss-Forest (Random Forest Imputation Algorithm) for our project due to its advantages, which include handling missing values in large datasets, providing unbiased estimates of missing values, and preserving the distribution of the original data.

2.5.2 Algorithms for Progression

2.5.2.1 Seq-2-seq for next time step:

The Seq-2-seq model for Next time step prediction is a machine learning system that predicts a patient's next visit's conversion to Alzheimer's disease (AD). A recurrent neural network (RNN) and a multi-layer perceptron (MLP) model are the two key components of the approach. The RNN is trained to learn a latent representation of the patient's longitudinal clinical data up to a specific visit. This is symbolized by the symbol xt and is represented by Eq 1. The MLP model then combines the cross-sectional demographic data (D) and xt to forecast AD conversion in the following visit. Eq 2 represents this prediction, which is denoted as y′. Trainable linear transformation matrices W1 and W2, as well as bias vectors b1 and b2, are used to train the model.The sigmoid function is denoted by σ.

$$\hat{x}t = RNN(X) \quad (1)$$

$$y' = \sigma\,(W1(ReLU\,(W2(\hat{x}t \oplus D) + b2)) + b1) \quad (2)$$

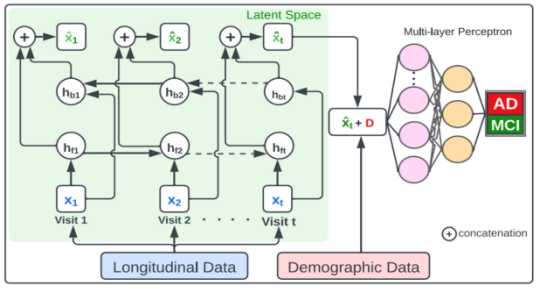

Fig 5:

Seq-2-seq for next time step architecture [14]

2.5.2.2 Seq-2-seq for multiple time steps :

For multiple future time steps prediction, the Seq-2-seq model includes an RNN component that captures a hidden representation (xt) of the longitudinal clinical data up to t visits (as indicated in Eq 1). This hidden representation is used by the decoder component to generate representations for future visits up to n, and the resulting representations are used to train an MLP model with the cross-sectional demographic data (D) to predict the conversion to AD at the (t +n)th visit (as shown in Eq 3). W1 and W2 are the trainable linear transformation matrices, b1 and b2 are the bias vectors, and is the sigmoid function.

$$y' = \sigma\,(W1\,(ReLU(W2(xt{+}(n{-}1) \oplus D) + b2)){+}b1) \quad (3)$$

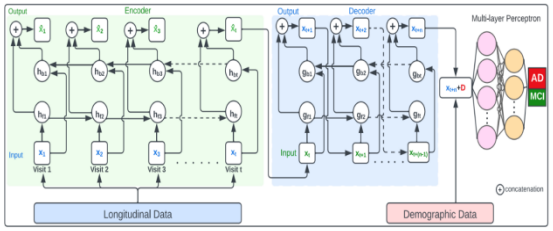

Fig 6:Seq-2-seq for multiple future time steps architecture [14]

2.5.2.3. Other Algorithms

Temporal convolutional Networks:

Temporal Convolution Networks (TCN) are useful for modeling long-term dependencies in sequences. Two important characteristics of the TCN are that (a) the output length is the same as the input length, and (b) there is no data leakage from the present to the past. TCN mixes 1D Fully Convolutional Networks (FCN) with causal convolutions to obtain these characteristics. The dilated convolution process captures long-term interdependence by allowing for a larger receptive field. TCN may be defined using equations (5) and (6), where y represents the input sequence and f represents the filter. In the context of Alzheimer's disease progression prediction, TCN can be used to model longitudinal data from multiple visits of patients. One advantage of TCN over other sequence modeling approaches is that it can capture long-term dependencies without the vanishing gradient problem that can affect RNNs like LSTMs.

$$F(u) = (y * b\, f)(u) = \sum_{j}^{m-1} \quad f(j) y_u - b_j \tag{5}$$

$$o = Activation(y + F(y)) \tag{6}$$

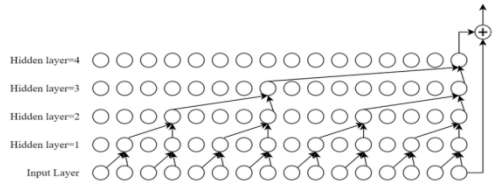

Fig 7: TCN architecture

LSTM :

The Long Short-Term Memory (LSTM) architecture is a gated recurrent neural network design that solves the disappearing and exploding gradient concerns in RNNs. LSTMs may store context information in their internal memory, allowing them to properly interpret variable-length sequences. The gated structure of the LSTM architecture defines it, with input, output, and forget gates controlling the flow of information into and out of the memory cell. Standard LSTMs, on the other hand, assume uniformly distributed elapsed time, which limits their efficacy for longitudinal data with time anomalies.

GRU:

The GRU (Gated Recurrent Unit) is a condensed version of the LSTM that employs fewer parameters and gating techniques to regulate information flow. The GRU cell contains two gates that allow it to selectively update and forget information: update and reset. Equations that update the hidden state and output based on the input and prior hidden state may be used to create the GRU cell, just as LSTM.

BiLSTM:

An addition to LSTM that involves bidirectional processing is called Bidirectional LSTM (BiLSTM). The hidden state in a typical LSTM solely depends on the previous inputs in the sequence at each time step. BiLSTM, on the other hand, analyzes the sequence both forward and backward, combining the resultant hidden states to provide a more accurate representation of the sequence. Since both forward and backward context information can be valuable in a variety of applications, this enables BiLSTM to capture both in the sequence.

BiGRU:

Bidirectional GRU (BiGRU) is a similar architecture to BiLSTM, but uses GRU cells instead of LSTM cells. Like BiLSTM, BiGRU processes the sequence in both directions to capture forward and backward context information. BiGRU is computationally less expensive than BiLSTM, making it a good choice for applications where efficiency is a concern.

3 Results

3.1 Classification Results

In order to identify the importance of each of the modality i.e, MRI and PET on the overall performance, the model was implemented three times i.e using MRI alone, PET alone, using both MRI and PET images. The values corresponding to each of the testing methods were shown in the Table. xxx. It is evident from the results that MRI provides better results compared to PET when implemented alone. However, when using multimodal data, it is observed that adding PET images helps in the improvement of the model.

Table 3: Comparing the performance of our proposed method vs single modal

	MRI	PET	MRI+PET
Accuracy	0.8549	0.6010	**0.8858**
Precision	0.8537`	0.4982	0.8858
Recall	0.8539``	0.6010	0.8858
F-1 Score	0.8538	0.4890	0.8832

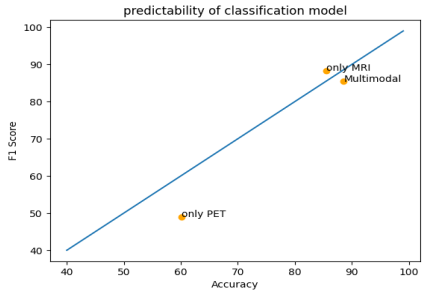

Fig. 8 Predictability for different models in classification

To assess the performance of our model, we compared the results of our proposed method with that of the baseline model implemented in [bb]. Juan Song, et al implemented the image fusion technique to combine both the MRI and PET modalities and compared the performance with the other methods such as feature fusion, MRI alone, PET alone showing that their model provides better accuracy when implemented using UNet. However, using a simpler architecture such as ResNet50 and modifying the kernel size and dropout rates, we were able to get better accuracy as shown in Table. aa compared to their image fusion technique and UNet.

Table 4. Comparing the performance of our proposed method vs other research works

	MRI	PET	MRI+PET
Image fusion + 3D CNN	0.6486	0.6010	0.7454
Image fusion + multiscale 3D CNN	0.6645	0.5876	0.7152
Our method (Feature fusion + modified resnet)	**0.8539**	**0.6066**	**0.8858**

3.2 Progression Results

3.2.1 Results without time difference and additional normalization:

3.2.1.1 Seq2Seq Model for Next Visit Results:

In this study, we evaluated the performance of different models for predicting the next visits in longitudinal healthcare data. The baseline Seq2Seq model achieved an F1 score of 0.74. However, the TCN model outperformed both the baseline and other models, achieving an F1 score of 0.75. This can be attributed to the architecture's ability to capture long-term dependencies in the data without suffering from the vanishing gradient problem that can affect recurrent neural networks. The use of causal convolutions and residual layers allowed for effective history sizes to be achieved, enabling the model to look far into the past to make accurate predictions. Overall, our results suggest that TCN is a promising approach for predicting the next visit in healthcare data.

Table 5 : Seq2Seq Model for Next Visit Results with time difference feature and additional normalization.

Models	**LSTM**	**TCN**	**GRU**	**BiLSTM**	**BiGRU**
Best Confusion Matrix	[[164 52] [20 99]]	**[[170 46] [20 99]]**	[[161 55] [18 101]]	[[169 47] [22 97]]	[[157 59] [14 105]]
Precision	0.655	**0.683**	0.647	0.674	0.640
Recall	0.832	**0.832**	0.849	0.815	0.882
F1- Score	0.733	**0.750**	0.7345	0.7376	0.7420

3.2.1.2 Seq2Seq Model for Multiple Future Visit Results:

The research investigated the performance of the Seq2Seq architecture for multiple future visits in predicting health risk. The baseline model achieved a F1 score of 0.74. However, the BiLSTM model outperformed the baseline model, achieving a F1 score of 0.72. This is because the BiLSTM model has the advantage of decoding for dynamic time steps, which is useful for predicting risk for multiple future time steps. Additionally, it considers the past and future time steps while encoding a sequence, resulting in a more accurate prediction of health risks. The results of this study suggest that the BiLSTM model is an effective approach for predicting health risks using sequential data.

Table 6 : Seq2Seq Model for Next Visit Results with time difference feature and additional normalization.

Models	**LSTM**	**TCN**	**GRU**	**BiLSTM**	**BiGRU**
Best Confusion Matrix	[[124 57] [12 82]]	[[123 58] [11 83]]	[[6 175] [3 91]]	**[[140 41] [18 76]]**	[[133 48] [15 79]]
Precision	0.590	0.589	0.342	**0.649**	0.622
Recall	0.872	0.883	0.968	**0.809**	0.840
F1- Score	0.703	0.7063	0.5055	**0.720**	0.715

3.2.2 Results with time difference feature and additional normalization

Table 7 : Seq2Seq Model for Next Visit Results with time difference feature and additional normalization.

Models	**LSTM**	**TCN**	**GRU**	**BiLSTM**	**BiGRU**
Best Confusion Matrix	[[150 66] [14 105]]	[[150 66] [16 103]]	[[164 52] [21 98]]	[[149 67] [17 102]]	[[165 51] [20 99]]
Precision	0.614	0.609	0.653	0.603	0.66
Recall	0.882	0.865	0.823	0.857	0.83
F1- Score	0.724	0.715	0.728	0.7083	0.736

Table 8: Seq2Seq Model for Multiple Future Visits Results with time difference feature and additional normalization.

Models	**LSTM**	**TCN**	**GRU**	**BiLSTM**	**BiGRU**

Best Confusion Matrix	[[145 36] [23 71]]	[[138 43] [19 75]]	[[120 61] [12 82]]	[[126 55] [13 81]]	[[125 56] [12 82]]
Precision	0.664	0.635	0.573	0.595	0.594
Recall	0.755	0.797	0.872	0.862	0.872
F1- Score	0.706	0.707	0.691	0.704	0.706

For the next time step risk prediction, the BiGRU model achieved an F-1 score of 73%, while other models performed similarly. However, for risk prediction in multiple time steps, all models showed similar performance. Based on these results, it was found that the time difference feature did not significantly contribute to the prediction of AD risk, while MRI biomarkers, PET biomarkers, and cognitive assessment scores had a significant impact on predicting the risk of AD. These findings suggest that incorporating these biomarkers and assessment scores can lead to more accurate predictions of AD risk.

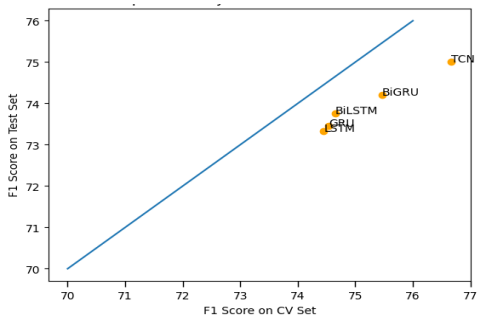

Fig. 9: Predictability Plot for Risk Prediction for next visit

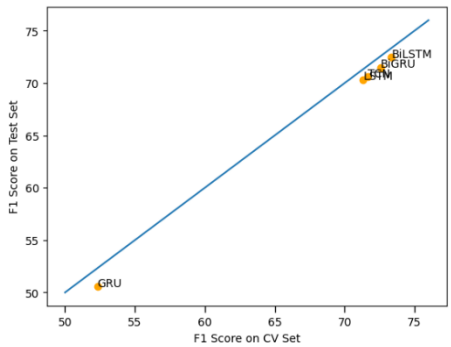

Fig. 10: Predictability Plot for Risk Prediction for multiple time visits

To evaluate the performance of multiple models implemented for a specific task, a predictability plot was generated. The plot visualizes the accuracy of each model on a validation dataset and an evaluation dataset and allows for easy comparison between the models. The results showed that the TCN model outperformed the other models, as it achieved higher accuracy on the evaluation dataset compared to the other models. This finding suggests that the TCN model may be more suitable for the given task and highlights its potential as a promising model for future studies.

3.3 MockGUI Setup

In addition to developing a machine learning model for Alzheimer's disease (AD) classification and progression, we have also implemented a graphical user interface (GUI) to enable patients and clinicians to easily interact with the model. The Screenshot of the GUI is shown in Fig. 11.

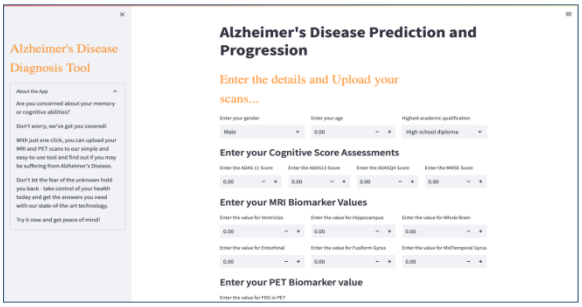

Fig 11: GUI for the better usage of the model

The GUI has been designed with a patient-friendly interface, allowing patients to input their electronic health record (EHR) data, including MRI and PET biomarkers, cognitive assessment scores, and upload their scan reports for MRI and PET. The model then generates a prediction of the patient's current AD status and assesses their risk for developing AD in the future. To generate risk predictions for future visits, the user is prompted to input the number of future visits for which they would like to receive risk predictions. The model then outputs the predicted risk for AD for each future visit.

The GUI provides an accessible and user-friendly platform for patients and clinicians to receive accurate AD predictions and risk assessments based on their individual data. This approach is particularly valuable for early diagnosis and intervention, as it enables patients to track their disease progression and receive timely treatment. Moreover, this technology can facilitate the efficient management of healthcare resources, as clinicians can use predictions to prioritize patient care and optimize treatment plans.

4 Conclusion:

In this research, we aimed to investigate the effectiveness of multimodal data (MRI + PET) compared to single modality (MRI or PET) in classifying patients with MCI and AD. Our findings demonstrated that classification performance was efficient when considering multimodal data. Furthermore, MRI and PET provided complementary features that helped differentiate patients with MCI and AD. We also implemented two Seq2Seq architectures to predict the progression of AD in the next time step and multiple time steps. The TCN model performed best for the next time step prediction, while the BiLSTM model performed best for multiple time steps. Our study also showed that the addition of the time difference feature did not significantly contribute to the prediction of AD risk. However, MRI biomarkers, PET biomarkers, and cognitive assessment scores were found to be important predictors of AD risk. These results highlight the importance of

multimodal data and specific biomarkers in accurately predicting the risk of AD.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Dr. May Wang, the teaching assistants, our fellow classmates, and all those who contributed to our course project on Medical Image Processing at the Georgia Institute of Technology. Their invaluable guidance, support, and feedback have been crucial to our success, and we are grateful for their contributions and for making this project a valuable learning experience.

REFERENCES

1. Carrion C, Folkvord F, Anastasiadou D, Aymerich M. Cognitive therapy for dementia patients: a systematic review. Dement Geriatr Cogn Disord. (2018) 46:1–26. doi: 10.1159/000490851
2. Alzheimer's Association. Alzheimer's disease facts and figures. Alzheimers Dement. (2020) 16:391–460. doi: 10.1002/alz.12068
3. Wang C, Saar V, Leung KL, Chen L, Wong G. Human amyloid β peptide and tau co-expression impairs behavior and causes specific gene expression changes in Caenorhabditis elegans. Neurobiol Dis. (2018) 109:88–101. doi: 10.1016/j.nbd.2017.10.003
4. Dai Z. Applications, opportunities and challenges of molecular probes in the diagnosis and treatment of major diseases. Chin Sci Bull. (2017) 62:25–35. doi: 10.1360/N972016-00405
5. Wenk GL. Neuropathologic changes in Alzheimer's disease. J Clin Psychiatry. (2003) 64(Suppl 9):7–10. Available online at: https://www.psychiatrist.com/ JCP/article/Pages/neuropathologic- changes- alzheimers-disease.aspx
6. Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, de Leon MJ. Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. J Alzheimers Dis. (2010) 20:843–54. doi: 10.3233/JAD-2010-091504
7. Camus V, Payoux P, Barré L, Desgranges B, Voisin T, Tauber C, et al. Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment. *Eur J Nucl Med Mol* Imaging. (2012) 39:621–31. doi: 10.1007/s00259-011-2021-8
8. Riederer I, Bohn KP, Preibisch C, Wiedemann E, Zimmer C, Alexopoulos P, et al. Alzheimer disease and mild cognitive impairment: integrated pulsed arterial spin-labeling MRI and18F-FDG PET. *Radiology*. (2018) 288:198–206. doi: 10.1148/radiol.2018170575
9. Guan, Hao, et al. "Learning transferable 3D-CNN for MRI-based brain disorder classification from scratch: An empirical study." *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer International Publishing, 2021.
10. Choi, Hongyoon, Kyong Hwan Jin, and Alzheimer's Disease Neuroimaging Initiative. "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging." *Behavioural brain research* 344 (2018): 103-109.
11. Venugopalan, Janani, et al. "Multimodal deep learning models for early detection of Alzheimer's disease stage." *Scientific reports* 11.1 (2021): 3254.
12. Satone, Vipul, et al. "Learning the progression and clinical subtypes of Alzheimer's disease from longitudinal clinical data." *arXiv preprint arXiv:1812.00546* (2018).
13. Zhang, Zicong, Changchang Yin, and Ping Zhang. "Temporal Clustering with External Memory Network for Disease Progression Modeling." *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2021.
14. Al Olaimat, Mohammad, et al. "PPAD: A deep learning architecture to predict progression of Alzheimer's disease." *bioRxiv* (2023): 2023-01.
15. Jung, W., Jun, E., Suk, H. I., & Alzheimer's Disease Neuroimaging Initiative. (2021). Deep recurrent model for individualized prediction of Alzheimer's disease progression. *NeuroImage*, *237*, 118143.
16. R. Cuingnet, E. Gerardin, J. Tessieras et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database," NeuroImage, 2nd ed, vol. 56, 2011, pp.766-781.
17. Wenk GL. Neuropathologic changes in Alzheimer's disease. J Clin Psychiatry. (2003) 64(Suppl 9):7–10. Available online at: https://www.psychiatrist.com/ JCP/article/Pages/neuropathologic-changes-alzheimers-disease.aspx
18. Clark CM, Schneider JA, Bedell BJ, et al. Use of Florbetapir-PET for Imaging β-Amyloid Pathology. JAMA. 2011;305(3):275–283. doi:10.1001/jama.2010.2008
19. Hongyoon Choi, Kyong Hwan Jin, et al. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. 2018. ISSN 0166-4328, https://doi.org/10.1016/j.bbr.2018.02.017.
20. Youssofzadeh V, McGuinness B, Maguire LP and Wong-Lin KF (2017) Multi-Kernel Learning with Dartel Improves Combined MRI-PET Classification of Alzheimer's Disease in AIBL Data: Group and Individual Analyses. Front. Hum. Neurosci. 11:380. doi: 10.3389/fnhum.2017.00380
21. Cheng, D., & Liu, M. (2017). CNNs based multi-modality classification for AD diagnosis. In International Congress on Image and Signal Processing. https://doi.org/10.1109/cisp-mei.2017.8302281
22. Song J, Zheng J, Li P, Lu X, Zhu G and Shen P (2021) An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer's Disease Diagnosis. Front. Digit. Health 3:637386. doi: 10.3389/fdgth.2021.637386
23. F. Li, D. Cheng and M. Liu, "Alzheimer's disease classification based on combination of multi-model convolutional networks," 2017 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 2017, pp. 1-5, doi: 10.1109/IST.2017.8261566.
24. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X and the Alzheimer's Disease Neuroimaging Initiative (ADNI) (2019) Diagnosis of Alzheimer's Disease via Multi-Modality 3D Convolutional Neural Network. Front. Neurosci. 13:509. doi: 10.3389/fnins.2019.00509
25. Hu, Z., Wang, Z., Jin, Y., & Hou, W. (2023). VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. Computer Methods and Programs in Biomedicine, 229, 107291.