# ADINSIGHT: A MULTIMODAL AND EXPLAINABLE FRAMEWORK FOR ALZHEIMER'S DISEASE PROGRESSION AND CONVERSION PREDICTION

A Dissertation
IN
Computer Science
and
Biomedical and Health Informatics

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
MATTHEW VELAZQUEZ

MBA, Texas A&M University-Commerce, USA, 2013
B. A., Ottawa University, USA, 2010

Kansas City, Missouri
2023

ADINSIGHT: A MULTIMODAL AND EXPLAINABLE FRAMEWORK FOR ALZHEIMER'S DISEASE PROGRESSION AND CONVERSION PREDICTION

Matthew Velazquez, Candidate for the Doctor of Philosophy Degree

University of Missouri–Kansas City, 2023

## ABSTRACT

ADInsight represents the crux of this dissertation, introducing an integrated and explainable framework centered on predicting Alzheimer's disease (AD) conversion, particularly for those at the early stage of mild cognitive impairment (EMCI). Beginning with an examination of models grounded in individual research modalities, such as clinical data and advanced imaging, the research underscores the potential and limitations of singular approaches. As a response to these findings, this dissertation introduces a multimodal ensemble conversion prediction model that combines Diffusion Tensor Imaging (DTI) scans with clinical data. This ensemble not only increases the accuracy of predictions but is also notable for its dedication to explainability, bridging the gap between intricate neural network predictions and understandable medical interpretations. Upon further exploration a unique framework is revealed, combining the advantages of Random Forest Regression alongside the latest over-sampling methods. This framework

unravels the intricacies of AD's nonlinear progression, leading to the formulation of patient progression groupings. The dissertation is then concluded with the Cognitive Visual Recognition Tracker (CVRT) application. This application marks an exploration into cognitive focus and visual identification, which are essential elements in the development of Alzheimer's disease. Benefiting both clinicians and patients, CVRT paves the way for innovative treatment strategies. In summary, our ADInsight framework provides a novel approach to understanding and predicting the progression of AD, providing a beacon of hope and knowledge in the ongoing struggle against this debilitating condition.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "ADInsight: A Multimodal and Explainable Framework for Alzheimer's Disease Progression and Conversion Prediction," presented by Matthew Velazquez, candidate for the Doctor of Philosophy degree, and hereby certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Yugyung Lee, Ph.D., Committee Chair
Computer Science, School of Science & Engineering

Monica Gaddis, Ph.D.
Department of Biomedical & Health Informatics, School of Medicine

Yusuf Sarwar Uddin, Ph.D.
Computer Science, School of Science & Engineering

Zhu Li, Ph.D.
Electrical and Computer Engineering, School of Science & Engineering

Peter Koulen, Ph.D.
Department(s) of Biomedical Sciences, Ophthalmology, School of Medicine

CONTENTS

ILLUSTRATIONS

# TABLES

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my advisor, Dr. Yugyung Lee, for her unwavering support, guidance, and encouragement throughout the course of my PhD journey. Her insights and advice have been invaluable, shaping not only this dissertation but also my perspective on academic research.

I am also thankful to Dr. Monica Gaddis, my co-disciplinary advisor, whose expertise in Biomedical Informatics has provided a robust interdisciplinary approach to my research. Her constructive feedback and thoughtful questioning have enriched my academic experience immeasurably.

I would like to express sincere thanks to my dissertation committee members, Dr. Yusuf Sarwar Uddin, Dr. Zhu Li, and Dr. Peter Koulen, for their time, expertise, and invaluable contributions to this research project. Their rigorous questions and insightful comments have significantly improved the quality of this dissertation, and I am grateful for their involvement.

The inspiration for this research came from a deeply personal experience with my late grandmother. Her battle with Alzheimer's not only inspired this work but also ignited my passion for developing a predictive model for this devastating disease. I dedicate this work to her memory.

My heartfelt gratitude goes to my peers and colleagues for providing an intellectually stimulating environment and offering invaluable advice and emotional support. Special thanks to the administrative staff for facilitating the resources and conditions needed

to complete this research.

I am profoundly indebted to my wife, Dr. Salma Velazquez, for her constant love, emotional support, and intellectual companionship. Her unwavering support and sacrifices, particularly while balancing her own career and our family life, have made this academic journey far more manageable. I also extend my thanks to my son, Benjamin, whose infectious laughter and pure joy serve as a constant reminder of the world beyond academia. I must also mention Daisy, our devoted dog, whose loyal companionship was a steady source of comfort and happiness during the long and often challenging PhD journey.

# CHAPTER 1

# INTRODUCTION

In this chapter, we provide an introduction to Alzheimer's Disease conversion prediction and the various challenges and approaches that have been implemented thus far. The potential and vital role of explainability in prediction models is emphasized, followed by our contributions and the rationale for our ADInsight framework.

## 1.1 Background and Significance

Alzheimer's disease (AD) is a progressive neurological disorder that causes the brain to diminish and leads to nerve cell death. It is the most common type of dementia and presents with memory loss, impaired daily living activities, and cognitive decline. This can eventually lead to a significant burden on the individuals themselves, as well as caregivers and our healthcare systems. Preventing the progression of AD is difficult as there are no effective treatment plans. However, early detection could significantly alter the course of disease management and potentially enhance the effectiveness of treatments when provided as early as possible [1]. This early stage, classified as Mild Cognitive Impairment (MCI), represents the onset of problems with memory recall, language, thinking, or judgment. Within the MCI umbrella, this dissertation focuses on the Early Mild Cognitive Impairment (EMCI) subjects as they represent the furthest possible MCI subclass from AD. Given that approximately 32% of patients diagnosed with MCI will develop

Alzheimer's Disease [1], it becomes crucial to devise an accurate, explainable tool to identify which patients will convert, given the life-changing implications of an inaccurate prediction. Furthermore, in a healthcare landscape that is increasingly reliant on data-driven decision-making, the value of an interpretable prediction model extends beyond individual patient care. It can potentially shape policy, resource allocation, and the future direction of AD research.

## 1.2 Problem Statement

AD conversion prediction from MCI subjects has been an area of intense research, hoping to create models that can effectively determine individuals who are more likely to progress to AD. However, this area of research faces the challenge of balancing accuracy and explainability. Many different techniques have been explored for this problem across multiple modalities including genetic biomarkers, neuropsychological assessments, neuroimaging data, and different assortments of clinical variables. These approaches often have promising results for accuracy, but often lack explainability.

This deficiency in explainability for prediction models poses a challenge when seeking adoption in real-world clinical environments. Since the factors that generate these predictions are often in a "black-box," it is more difficult for clinicians to establish trust and confidence without having a transparent view into the underlying prediction rationale. Explainable artificial intelligence (XAI) has emerged as a domain to attempt to address this issue by providing prediction explanations that can be interpreted by other researchers or clinicians.

## 1.3 Dissertation Objectives

The primary objectives of this work are as follows:

- To develop an accurate and explainable tool for the prediction of Alzheimer's Disease conversion from the Early Mild Cognitive Impairment (EMCI) stage.
- To enhance the transparency and trust in predictive models, aiding their integration into clinical settings.
- To provide a thorough analysis of the various methods and approaches used in Alzheimer's Disease conversion prediction.
- To develop a multimodal framework that combines different types of data for improved accuracy and explainability.
- To present the application of the developed framework in mobile and web platforms for wider accessibility and use.

## 1.4 Summary of Contributions

This dissertation seeks to address these challenges with Alzheimer's disease conversion prediction by developing a prediction framework (ADInsight) that focuses on explainability. The main goal is to develop a framework that can accurately identify which EMCI subjects are likely to progress to AD, estimate the timeframe of this progression,

while also offering explanations that are interpretable. An explainable framework developed with this criteria can help the transition from black-box AI models to more transparent and reliable clinical decision support systems.

The main contributions of this work are as follows:

- An Alzheimer's Disease Conversion Prediction model to identify the MCI conversion population from clinical data (Chapter 2);
- A Diffusion Tensor Imaging (DTI) approach for visual AD conversion prediction with Convolutional Neural Networks (Chapter 3);
- An ensemble multimodality model that combines clinical and FMRI data to predict AD conversion alongside improved explainability methods (Chapter 4);
- Pattern analysis for longitudinal data is proposed that determines the timeline of a patient's AD conversion alongside categorization into novel groups (Chapter 5);
- Clinical Decision Support Applications that are presented in both mobile and web application formats. These can be used with dynamic inputs across multiple modalities (Chapter 6)

## 1.5 Dissertation Structure

The remaining chapters of this dissertation are organized as follows. In Chapter 2 we develop a random forest model from clinical data that accurately predicts AD conversion. This provides a foundation for subsequent analysis. We then explore Convolutional

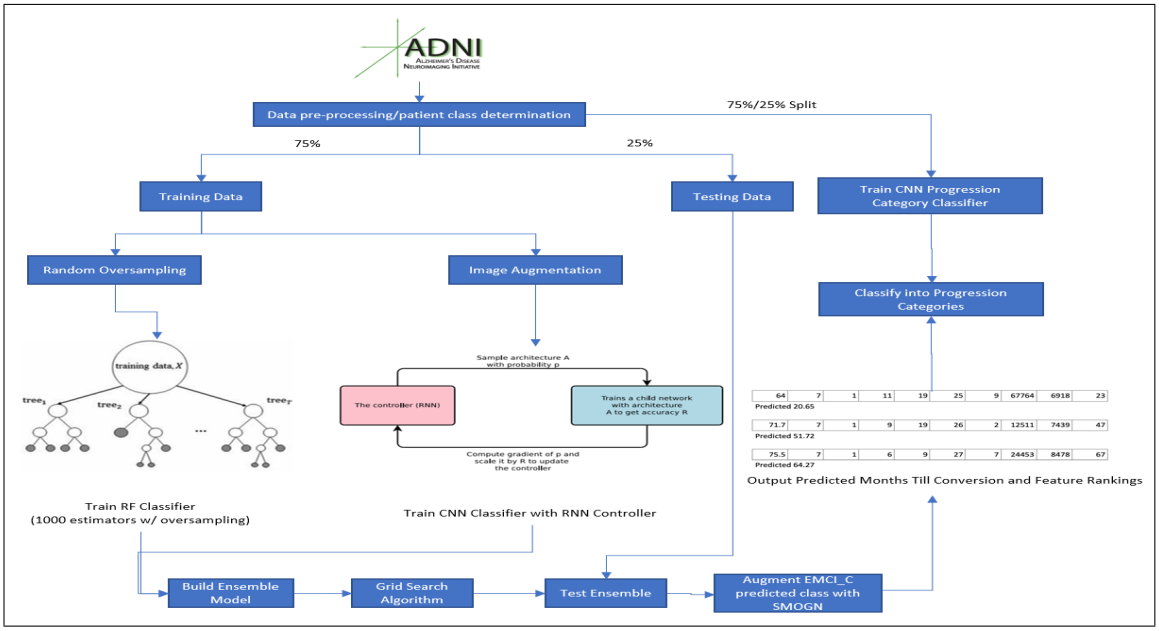



Figure 1: ADInsight Workflow

Neural Network (CNN) explainability via the generation of a visual model (Chapter 3). We then combine our clinical data and fMRI approaches into a single ensemble model capable of explainability across multiple modalities (Chapter 4). This step represents an integration of distinct data sources, enhancing the predictive capacity. Moreover, we present patient timelines that determine when a patient is likely to convert and similarity to other patient' progressions (Chapter 5). These patient timelines provide dynamic insights into the progression of AD. Furthermore, in Chapter 6, we extend this framework to web and mobile solutions so that our results can be reproduced dynamically. This also enables wider reach of our developed framework as well as a practical application. Finally, Chapter 7 serves as our conclusion by providing a summary of the dissertation and potential future work.

# CHAPTER 2

## FEATURE-BASED RANDOM FOREST MODEL

Alzheimer's Disease (AD) conversion prediction from the mild cognitive impairment (MCI) stage has been a difficult challenge. This chapter focuses on providing an individualized MCI to AD conversion prediction using a balanced random forest model that leverages clinical data. In order to do this, 383 Early Mild Cognitive Impairment (EMCI) patients were gathered from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Of these patients, 49 would eventually convert to AD (EMCI_C), whereas the remaining 334 did not convert (EMCI_NC). All of these patients were split randomly into training and testing data sets with 95 patients reserved for testing. Nine clinical features were selected, comprised of a mix of demographic, brain volume, and cognitive testing variables. Oversampling was then performed in order to balance the initially imbalanced classes prior to training the model with 1000 estimators. Our results showed that a random forest model was effective (93.6% accuracy) at predicting the conversion of EMCI patients to AD based on these clinical features. Additionally, we focus on explainability by assessing the importance of each clinical feature. Our model could impact the clinical environment as a tool to predict the conversion to AD from a prodromal stage or to identify ideal candidates for clinical trials.

As previously mentioned, Alzheimer's Disease (AD) is a progressive, degenerative brain disorder that leads to nerve cell death and tissue loss in the brain. Currently,

there are no treatment plans that prevent the progression of AD, and this has led to increased emphasis on being able to predict AD at an earlier stage. Mild Cognitive Impairment (MCI) is an intermediary stage between being cognitively normal and having AD where 32% of MCI patients will go on to develop Alzheimer's Disease [1]. This makes the MCI stage an ideal target for early prediction as studies point to early diagnosis as being key to potentially delaying the overall progression of AD [1]. Early detection at the MCI stage can assist in clinical trial enrollment and provide more specific treatment plans when more effective ones do become available. Our focus in this chapter was to target the earliest subset of MCI patients (EMCI), as that subset is the furthest from an AD diagnosis and would thus provide a more beneficial prediction. As a result, it is of high importance to accurately determine which EMCI patients will develop AD.

For this reason, an accurate, ensemble learning model that can aid in clinical decision making is necessary to help ascertain the patient's prognosis. A random forest algorithm is a supervised learning algorithm that randomly creates and merges multiple decision trees and that has been proficient with classification problems [2]. In this chapter, a random forest model is used to determine which patients will convert to AD (our EMCI_C class) and those that will not convert (EMCI_NC) against an imbalanced data set. As well as determining how to best balance the data, assessing which clinical features are most relevant for conversion prediction is fundamental to our problem. Through our random forest model we are able to see which of our clinical features has the most significant impact at both the model and the individual prediction levels. This allows us to interpret the individual results better to provide more clinical significance. In this work,

we sought to (1) identify significant features from clinical data; (2) build a random forest classification model from an imbalanced data set of those features; (3) determine the prediction accuracy of our model.

Also, we observed the associations between individual predictors and their importance to the problem. By attempting different feature groupings, we were able to distinguish the most crucial feature types. As a result of this approach, our work provides a clinical decision-making tool that can predict MCI-to-AD conversion with high accuracy and interpret the results meaningfully. We envision that this work will provide an accurate tool for predicting conversion probability from MCI to AD and further understand the impact of neuropsychological, biomarker, and demographic features.

## 2.1 Related work

A review on the use of random forest models in classifying Alzheimer's Disease was provided by Sarica et al. [3]. Their review consisted of 12 studies that were primarily focused on the classification of Alzheimer's Disease stages from MRI images. The accuracy across these studies ranged from 53% to 96%, depending on whether they were performing multiclass classification or not. These studies were also focused on the direct stage classification of AD vs. Normal Controls vs. MCI, rather than the prediction of AD from an earlier stage.

Another review by Weiner et al. [4] summarized 49 ADNI papers. These papers did target the prediction of AD but were also focused mainly on MRI data. These were

occasionally supplemented by clinical data or other imaging data (PET), with most studies using a support vector machine (SVM) model. A few of the studies [5][6][7] did use the random forest algorithm as well and will be compared, alongside the SVM implementations, against our model's performance.

Huang et al. [8] proposed a predictive nomogram that combined image features, clinical factors, and AB concentration to predict the conversion of MCI to AD. They also explored the associations between the different selected features and reported on their significance. Their goal was to examine the associations at both a macro and micro level to better understand the underlying patterns.

Moore et al. [9] proposed using a pairwise selection from time-series data to predict AD conversion. The authors analyzed the relationships between data point pairs at different times using a random forest algorithm. They leveraged a mix of demographic and genetic data and achieved a classification accuracy of 73% as a result.

Lebedev et al. [6] used a combination of structural MRI scans along with a few clinical features from the ADNI data set to achieve an MCI-to-AD conversion accuracy of 81.3%. Their work also saw a sharp increase in accuracy by using a Random Forest algorithm rather than a Support Vector Machine. One advantage in their study is that they validated the model extensively outside of the ADNI data set and found no substantial drop in accuracy, suggesting a good foundation for clinical implementation.

Rana et al. [10] created a model deemed MudNet, which combined both clinical

data and MRI imaging for MCI-to-AD conversion prediction. They used many neuropsychological assessment scores alongside T1-weighted structural MRIs to achieve a conversion accuracy of 69.8%. Their work also provided a time-to-AD conversion classification which differentiated between high-risk (AD conversion within two years), and low-risk people (AD conversion greater than two years) at a 66.9% accuracy.

Thushara et al. [11] used a random forest algorithm for multi-class Alzheimer's classification. Their work sought to distinguish between AD, MCI, cMCI (Converted MCI), and normal controls using largely biomarker features. They achieved a multi-class classification accuracy of 69.33% with an MCI-to-AD conversion prediction (cMCI class) accuracy of 47.19%.

## 2.2 Methods

### 2.2.1 Alzheimer's disease neuroimaging initiative data

All data used for this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and included patients from their ADNI-1, ADNI-2, and ADNI-GO studies [12]. The ADNI was launched in 2003 as a public-private partnership with the primary goal of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) [12]. Early Mild Cognitive Impairment (EMCI) patients were eligible for our study as long as they had follow-up appointments for greater than a year. The EMCI subset consists of patients that are 5-7

years before a possible AD diagnosis and are identified by the results of the Wechsler Memory Scale Logical Memory II test. These EMCI patients were then subdivided into two groups based on whether they would eventually be diagnosed with Alzheimer's Disease or not. We chose to represent these groups as EMCI_C, for our AD conversion group, and EMCI_NC for our stable group. From the ADNI variables, the Clinical Dementia Rating was used to make this determination based on the value of their last visit's diagnosis. The remaining 1806 EMCI visits were then used as a starting point for training prior to augmentation. Of these, 198 belonged to the EMCI_C class while 1608 visits were from EMCI_NC subjects. Overall, our study consisted of 383 EMCI patients (shown in Figure 2), with 49 belonging to the EMCI_C group and the remaining 334 within the EMCI_NC group. These patients were then randomly split such that 75% (288 patients) of our selected patients were used to train the random forest model, with the remaining 25% (95 patients) used for validation testing (shown in Table 1).

Table 1: EMCI Data Set for Machine Learning

| | EMCI_C | EMCI_NC |
|---|---|---|
| Subject# | 49 | 334 |
| Visit# | 198 | 1608 |
| Record# after Oversampling | 1608 | 1608 |
| Training Data | 1206 | 1206 |
| Testing Data | 402 | 402 |

### 2.2.2 Clinical features selection

The clinical features that were used to train our random forest model included a mix of genetic biomarkers (APOE4), physical biomarkers (hippocampal and ventricular

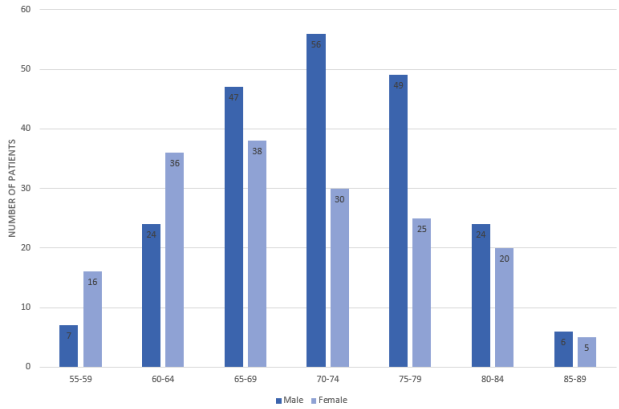



Figure 2: Participants' Age and Gender Distribution

volume), four neuropsychological scale scores (ADAS13, ADAS11, FAQ, MMSE), and the patient's demographic information (age, race). Many different variations of ADNI features were tested for model inclusion; however, these nine features were found to provide the best overall fit. Additionally, related studies have used similar features and found the mix of biomarker and neuropsychological scale scores to be an ideal selection for AD prediction [13]. The ADNI features that we have used per training group can be seen in Table 2.

### 2.2.3 Random forest classification model

Random forests are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is either the mode of the classes, in regards to a classifier, or the mean prediction of the individual trees for a regression model. Since random forests consist of a collection of decision trees that are trained with different data subsets and then averaged, this allows them to be tolerant of the problem of overfitting.

Table 2: ADNI Clinical Features and EMCI Patient Characteristics Used for Random Forest Model Training

| | | **EMCI_C** | | **EMCI_NC** | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Subject#** | 49 | | 334 | | | | |
| | **Time to AD Conversion** | 5.02 years | | – | | | | |
| | **Time in Study** | 12 years | | 12 years | | | | |
| **ADNI Feature** | **Description** | Mean | SD | Mean | SD | **6FT** | **9FT** | **13FT** |
| DX | Diagnosis | – | – | – | – | – | – | – |
| *Demographic Info* | | | | | | | | |
| PTRACCAT | Patient Race | – | – | – | – | | | |
| AGE | Patient Age | 73.5 | 6.47 | 71.1 | 7.49 | | | |
| *Genetic Biomarkers* | | | | | | | | |
| APOE4 | #E4 alleles of APOE | .9 | .71 | .4 | .46 | – | | |
| *Physical Biomarkers* | | | | | | | | |
| Hippocampus | Hippocampal volume | 6875.2 | 947.45 | 7334.1 | 910.20 | – | | |
| Ventricles | Ventricular volume | 39282.7 | 21031.66 | 34504.6 | 21394.49 | – | | |
| *Neuropsychological* | | | | | | | | |
| ADAS13 | 13-item AD Assessment Scale | 15.8 | 6.02 | 13.3 | 5.41 | | | |
| ADAS11 | 11-item AD Assessment Scale | 9.7 | 4.12 | 8.5 | 3.29 | | | |
| FAQ | Functional Activities Questionnaire | 4.1 | 4.38 | 1.82 | 2.50 | | | |
| MMSE | Mini-Mental State Examination | 28.1 | 1.58 | 28.3 | 1.71 | | | |
| RAVLT_immed | #words memorized over 5 trials | 34.5 | 8.39 | 40.3 | 11.40 | – | – | |
| RAVLT_learn | #words learned between trials 1-5 | 4.7 | 2.47 | 5.3 | 2.42 | – | – | |
| RAVLT_forg | #words forgotten between trials 5-6 | 5.1 | 2.54 | 4.1 | 2.64 | – | – | |
| RAVLT_perc_forg | %words forgotten between trials 5-6 | 60.7 | 29.13 | 44.0 | 29.36 | – | – | |

*EMCI_C* the converter group, *EMCI_NC* the stable group, *FT* Feature Training

For our work (as seen in Figure 3), the random forest classifier has two potential classes eligible for its output, EMCI_C (patients that converted to AD) and EMCI_NC (patients that did not convert to AD). These classes are voted on from each individual tree, which is then aggregated to provide an overall probability of AD conversion. Figure 4 shows an example of an individual tree. Random Forest classifiers also allow for individual input variable importance to be evaluated. As part of our work, we built an ad hoc prediction script that evaluates this variable importance at both the model and individual prediction levels. Initially, while training the model, this evaluation helped us determine which variables were most relevant for model inclusion. After the model has been trained, this variable importance ranking then helps to interpret the individual prediction results Table 3 shows the rankings of 6 features, 9 features and 13 features as well as the fractional ranks that are the average of the ordinal ranks for these three feature groups.

## 2.3 Results

### 2.3.1 Demographic and clinical characteristics

As can be seen across Table 2, 383 EMCI patients were gathered from the ADNI database, of which 49 would convert to AD (EMCI_C), and 334 would not convert (EMCI_NC). The patients' average age was 71.4, and 55.6% of the patients were men. There was a significant difference in age between the two groups ($P < .05$) according to the t-test, however this feature's difference was not statistically significant between classes when measuring feature importance. Additionally, a subsequent model was trained on a

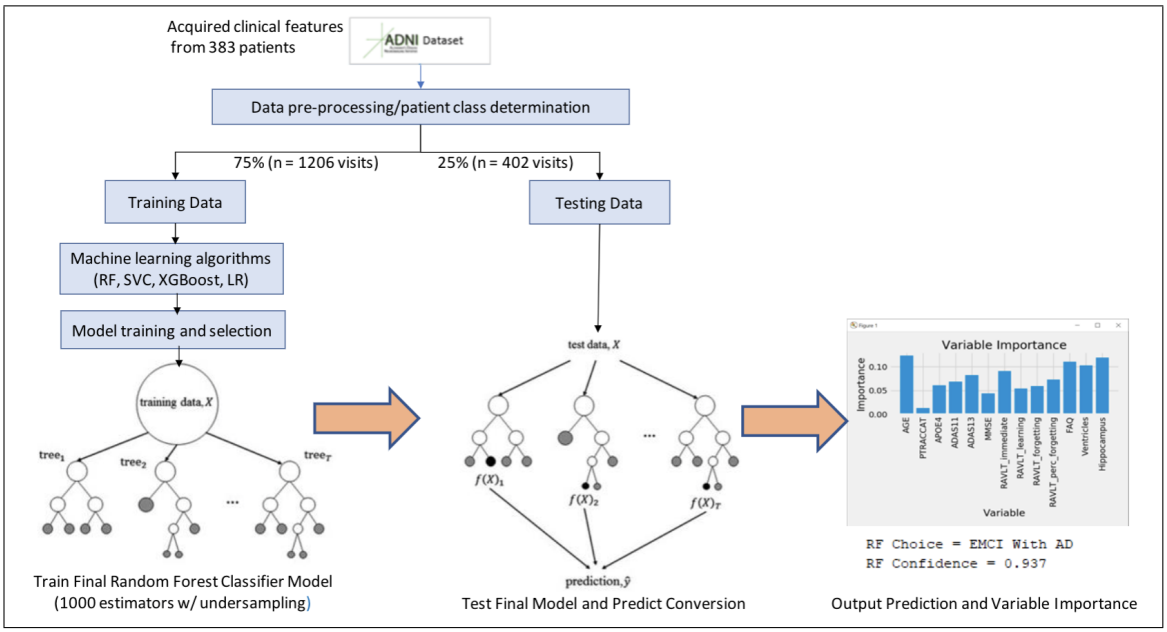



Figure 3: Model Workflow

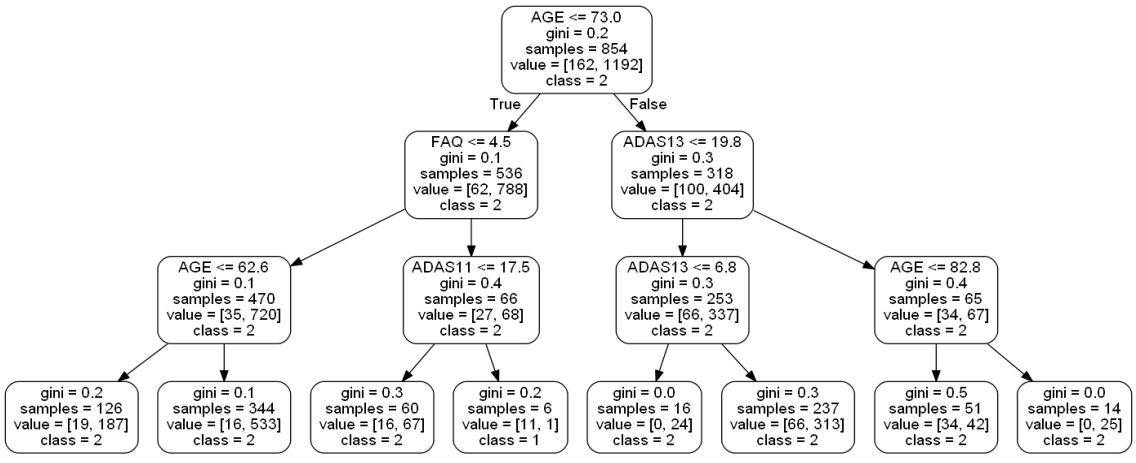



Figure 4: Example of a Small Random Forest Tree Within Our Model

Table 3: Comparison of Feature Importance Ranking by Feature Groups

| **Feature** | **6FT Ranks** | **9FT Ranks** | **13FT Ranks** | **Fractional Ranks** |
|---|---|---|---|---|
| AGE | **1** | **1** | **1** | **1** |
| FAQ | 2 | 4 | 3 | 3 |
| ADAS13 | 3 | 5 | 6 | 4.6 |
| ADAS11 | 4 | 6 | 8 | 6 |
| MMSE | 5 | 8 | 12 | 8.3 |
| PTRACCAT | 6 | 9 | 13 | 9.3 |
| Hippocampus | - | 2 | 2 | 2 |
| Ventricles | - | 3 | 4 | 3.5 |
| APOE4 | - | 7 | 9 | 8 |
| RAVLT_immed | - | - | 5 | 5 |
| RAVLT_perc_forg | - | - | 7 | 7 |
| RAVLT_forg | - | - | 10 | 10 |
| RAVLT_learn | - | - | 11 | 11 |

reduced data set that eliminated Age outliers and accuracy was only reduced by .3%.

Also shown in Table 2 are the genetic and physical biomarkers. The APOE4 and hippocampal volume differences were statistically significant between the EMCI_C and EMCI_NC groups, whereas the ventricular volume was not. For the neuropsychological scale scores, the ADAS13 and the FAQ features were significantly different ($P < .05$). The ADAS11 and the MMSE features were found not to be significantly different. The relationships between our features are further seen in Figure 5 as a Correlation Matrix.

### 2.3.2 Model performance

A workflow of our random forest model can be seen in Figure 3. This summarizes the training methodology as well as the prediction and variable importance output. After our pre-processing steps, we train a 1000 tree random forest model on 2412 exam visits

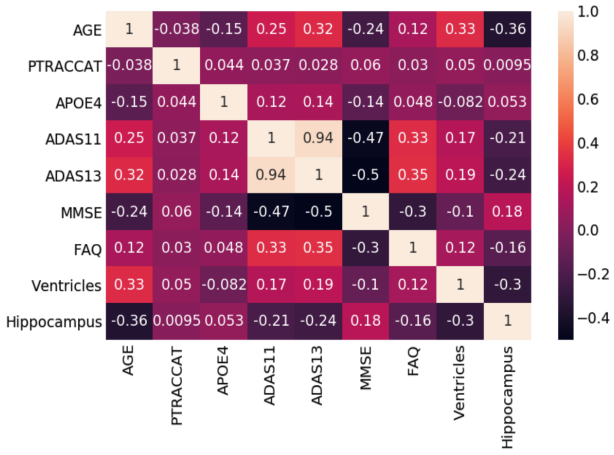



Figure 5: Random Forest Model Correlation Matrix

against different feature groups to compare their results. While initially, using all 13-features seemed to provide the highest accuracy at 91.6%, we found that by removing the Rey Auditory Verbal Learning Test (RAVLT) features, our accuracy rose to 93.6% in the 9-feature group. We also tested a 6-feature group, which removed the RAVLT variables as well as the biomarker data; however, this saw the worst accuracy of the three groups at 89.2%.

After running these feature group variations through our random forest training process, we decided to implement the same training data into a support vector classifier (SVC), an XGBoost classifier, and a Logistic Regression model for comparison (See Table 4). As these are commonly used for this problem, we considered this to be a reasonable comparative measure to the efficacy of our random forest model.

Support vector classifiers attempt to find the separating hyperplane that maximizes the distance of the closest points to the boundary of the class. These are typically effective

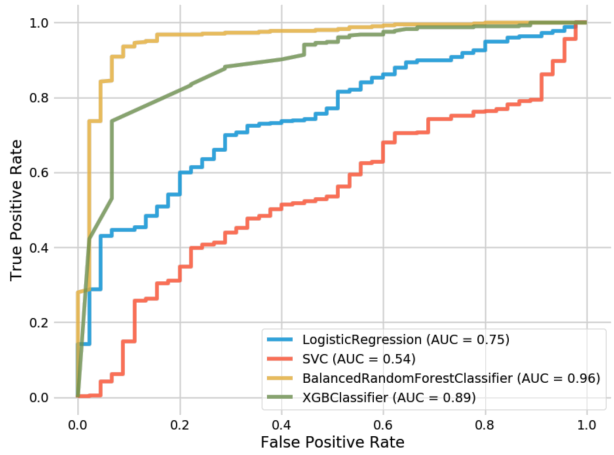



Figure 6: Receiver Operating Characteristic Curves for Random Forest and Comparison Models.

in high dimensional spaces and have seen a fair amount of usage within the AD conversion prediction domain [14] [8]. In both the 9-feature and 13-feature groups, we found that our random forest model outperformed our SVC implementation (93.6% vs. 90% and 91.6% vs. 90%, respectively). The SVC did show higher accuracy than the 6-feature RF model; however, the AUC was inferior on all SVC variations. The difference in AUC between our best RF variation (96% AUC) and our best SVC variation (54% AUC) is shown in Figure 6. One observation when observing our SVC model is that it struggled to predict the negative class (conversion to AD) and predominantly chose the majority class. This was not the case with our balanced random forest model which was able to appropriately distinguish between both classes.

XGBoost is an implementation of gradient boosted decision trees that has seen success in structured data classification. While not being common in the AD conversion prediction space, we wanted to compare how our feature-selection would be handled by

its algorithm. XGBoost resulted in the second-best overall method behind our top RF model and showed significantly better performance than the SVC and Logistic Regression implementations. For the 6-feature group, XGBoost outperformed our RF model (89.8% vs. 89.2%). However, while performing better than SVC and Logistic Regression, the XGBoost model still saw less accuracy than the RF model at both the 9 and 13 feature groups, as seen in Table 4. When comparing AUC, one can see how well XGBoost performed (89%) in relation to SVC (54%) and Logistic Regression (75%).

Table 4: Performance of Random Forest vs Support Vector Classifier

| **Model/Feature** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **AUC** | **p-value** |
|---|---|---|---|---|---|---|
| **Random Forest** | | | | | | |
| 6-Features | 0.892 | 0.907 | 0.980 | 0.942 | 0.88 | 0.91 |
| 9-Features | **0.936** | **0.952** | 0.978 | **0.965** | **0.96** | **0.71** |
| 13-Features | 0.916 | 0.916 | **0.998** | 0.955 | 0.93 | 0.82 |
| **Support Vector** | | | | | | |
| 6-Features | 0.900 | 0.900 | 1 | 0.948 | 0.52 | - |
| 9-Features | 0.900 | 0.900 | 1 | 0.948 | 0.54 | - |
| 13-Features | 0.900 | 0.900 | 1 | 0.948 | 0.55 | - |
| **Logistic Regression** | | | | | | |
| 6-Features | 0.894 | 0.902 | 0.990 | 0.944 | 0.76 | - |
| 9-Features | 0.892 | 0.903 | 0.985 | 0.942 | 0.75 | - |
| 13-Features | 0.896 | 0.904 | 0.990 | 0.945 | 0.75 | - |
| **XGBoost** | | | | | | |
| 6-Features | 0.898 | 0.904 | 0.993 | 0.946 | 0.87 | - |
| 9-Features | 0.920 | 0.930 | 0.985 | 0.957 | 0.89 | - |
| 13-Features | 0.907 | 0.921 | 0.980 | 0.950 | 0.88 | - |

Finally, Logistic Regression was the last method that we leveraged for comparison. Logistic Regression calculates the probability of an event occurrence and can be used when the target variable is categorical. For this model, we trained individual versions of

Table 5: State of the Art MCI-to-AD Prediction

| **Approach** | **Data** | **#Subject** | **Model** | **Estimator** | **MCI-AD(%)** | | **Predict** | **Train** |
|---|---|---|---|---|---|---|---|---|
| | | | | | ACC | AUC | Year | Time |
| Proposed (ours) | Clinical | ADNI(383) | RF | 1000 | **93.6** | **96** | **5-7** | **2.98 sec** |
| Grassi [14] | Clinical | ADNI(550) | SVM | - | - | 88 | 3 | 2 days |
| Huang [8] | Clinical/MRI | ADNI(290) | SVM | 1000 | 80 | 84.6 | 5 | - |
| Albright [13] | Clinical | ADNI(1737) | MLP | - | - | 86.6 | 5 | - |
| Moore [9] | Clinical | ADNI(1627) | RF | 60 | 73 | 82 | 5 | - |
| Ghazi [15] | MRI | ADNI(742) | RNN | 1000 | - | 76 | 5 | 340 sec |
| Rana [10] | Clinical/MRI | ADNI(559) | CNN | 100 | 69.8 | 83 | 5 | - |
| Thushara [11] | Clinical | ADNI(NA) | RF | 100 | 47.2 | - | 5 | - |

6, 9, and 13 features but found them to all exhibit less accuracy than our RF model. Additionally, while its AUC (75%) underperformed in contrast to RF and XGBoost, it did significantly better than our best SVC model (54%). Still, this did not result in a model that was close enough to warrant further consideration for our AD conversion problem.

As mentioned previously, our 9-feature random forest implementation with an accuracy of 93.6% and an AUC of 96% against a 383-patient data set represents our best model. While also using the ADNI data set, Grassi et al. [14] could achieve an AUC of 88% with an SVM that made predictions 3 years prior to AD onset. Huang et al. [8] also attained 80% accuracy and 84.6% AUC with an SVM model against the ADNI data set leveraging both clinical and MRI data looking 5 years prior to AD onset. As our approach differs by using Early Mild Cognitive Impairment patients (EMCI) rather than the broader MCI grouping used by other studies, we can predict conversion from 5-7 years prior to the onset of AD. Our outcome is state-of-the-art when comparing our accuracy and AUC to the previously published work for MCI-to-AD prediction as shown in Table 5.

### 2.3.3 Balancing the data

As a result of our imbalanced data set, where 12.8% of the patients belonged to the minority class (EMCI_C); we perform a random oversampling algorithm that generates new samples with replacement from the EMCI_C class. Replacement ensures that samples can be selected, added to the augmented data set, and then returned to the non-augmented data as eligible for further random sampling. For our study, we choose a minority strategy such that all samples would be generated solely from the existing EMCI_C data. This augmentation provides a balance between the two classes so that the majority class does not take over during model training. Through this process of oversampling our number of minority class visits become equivalent to that of the majority. This allows for the model to be trained against 2412 exam visits (1206 per class) rather than only the 1354 from the original data set train/test split. Table 1 further demonstrates the evolution of the data set after oversampling.

Table 6: Comparison of Imbalanced Data Set Sampling Methods

| **Method** | **Accuracy** |
|---|---|
| Random Over sampler | **93.60** |
| SMOTE | 92.97 |
| Borderline SMOTE | 93.19 |
| ADASYN | 93.30 |

We first compared our random oversampling method against an under-sampling method that targeted the majority class and found a 3.1% increase in accuracy via the oversampling process (see Table 6). Our approach was also compared to class weight

modifications, but they performed poorly in comparison to our minority strategy. After determining that oversampling was the preferred method we began to compare against the established oversampling methods.

SMOTE, or rather Synthetic Minority Oversampling Technique, was the first of these methods that we evaluated against. SMOTE relies on generating new information from the minority class population, rather than duplicating from that population. This is done by pulling from a random minority class sample, and then also finding a random k-nearest neighbor from that sample. The new data is then created in a space between those two samples [16]. However, against our EMCI data set this was found to reduce accuracy by .63% compared to our original technique.

Borderline SMOTE was also considered as this modifies the SMOTE technique to generate new data along the decision boundary of the two classes, rather than randomly between two samples [17]. While we did see a 0.22% improvement over SMOTE, it still fell short of our Random Oversampler.

Finally, we attempted the Adaptive Synthetic Sampling (ADASYN) method as a means of comparison. This deviates from the other SMOTE methods by generating new data based on the density of the data, rather than the decision boundary or k-nearest neighbor. ADASYN focuses its synthetic data creation within the low density feature space regions and creates less data within the high density regions [18]. For our data, this method produced the second best results and outperformed both SMOTE and Borderline SMOTE. The overall accuracy comparison of these oversampling techniques can be seen in Table 6.

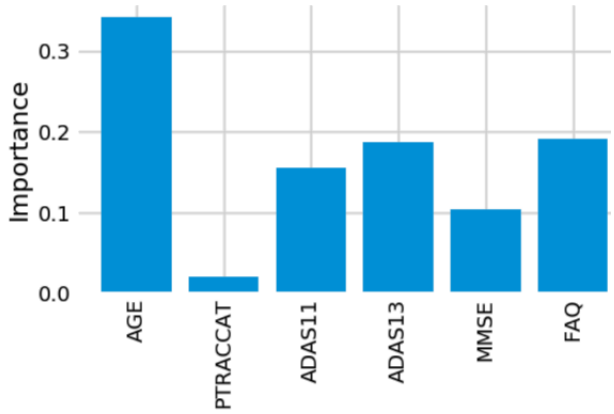



Figure 7: 6-Feature Model Feature Importance

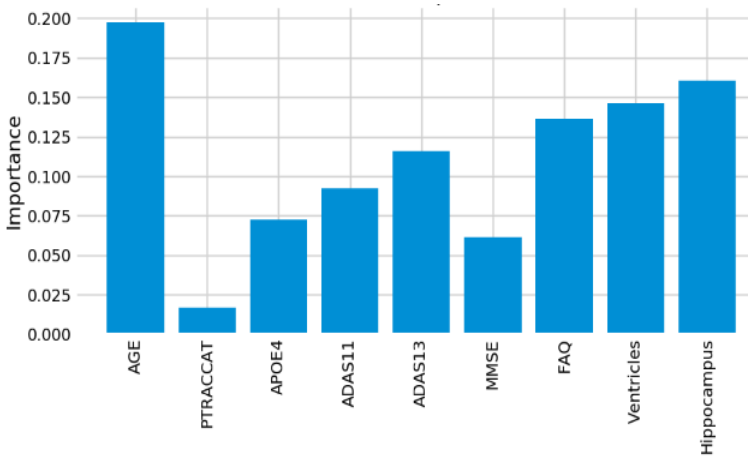



Figure 8: 9-Feature Model Feature Importance

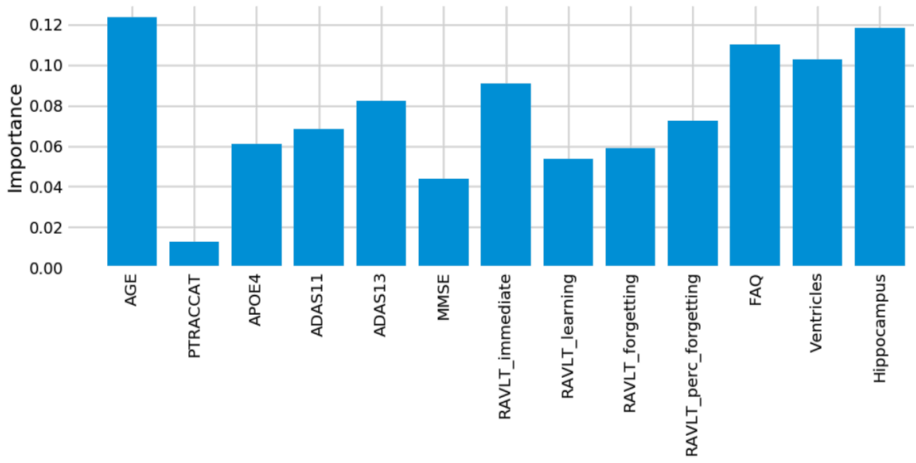



Figure 9: 13-Feature Model Feature Importance

### 2.3.4 Assessment of model feature importance

One advantage of using the random forest algorithm is that feature importance can be assessed at both the model and individual prediction levels. The model feature importance of our three feature groupings can be seen in Figures 7-8 and Figure 9. As a random forest algorithm deals with different combinations of features in each of its' decision trees, this allows for the feature importance to be calculated based on how much the prediction error increases [9]. This is done by first calculating the individual nodes' importance per tree as seen in Equation 2.1. Within this, $ni_j$ represents the importance of node $j$, $w_j$ being the weighted samples reaching node j, and $C_j$ as the impurity value of the node. Once each node's importance has been determined, the feature importance per tree is calculated per Equation 2.2 and is then normalized to a value between 0 and 1 per Equation 2.3. This result is then averaged across the entire forest before being divided by the total number of trees as seen in Equation 2.4 [19].

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{2.1}$$

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k\in\ all\ nodes} ni_k} \tag{2.2}$$

$$normfi_i = \frac{fi_i}{\sum_{j\in\ all\ features} fi_j} \tag{2.3}$$

$$RFfi_i = \frac{\sum_{j\in\ all\ trees} norm\ fi_{ij}}{T} \tag{2.4}$$

For our 6-feature model, the three most important features are Age, FAQ, and ADAS13. For our top-performing 9-feature model, the top features are Age, Hippocampus, and Ventricles. Finally, for our 13-feature model, Age, Hippocampus, and FAQ score are the most important. The presence of hippocampal and ventricular volume towards the top explains why the absence of those features in our 6-feature model resulted in diminished accuracy. By adding in the RAVLT features, our accuracy improved, but these were redundant with the other neuropsychological scale scores, so they were removed for our final model. Age was consistently seen as the best conversion predictor, which corresponds to the increased risk of AD at an older age [20]. Race (PTRACCAT) was routinely at the lowest feature importance between our models, but we did observe a decrease in accuracy upon its' removal. This is likely due to race having very little correlation with the other included features, whereas some neuropsychological scores exhibited signs of possible overlap (RAVLT).

Table 7: Average Standard Deviation Difference for Incorrect Predictions by Ground Truth Class and Feature

| | **Record#** | **Age** | **APOE4** | **ADAS11** | **ADAS13** | **MMSE** | **FAQ** | **Ventricles** | **Hippocampus** |
|---|---|---|---|---|---|---|---|---|---|
| **EMCI_C** | 20 | 1.90 | 2.20 | 1.30 | 1.38 | **2.37** | 0.82 | 0.80 | 1.00 |
| **EMCI_NC** | 9 | 1.74 | 2.26 | **2.87** | 2.46 | 2.68 | 2.61 | 1.13 | 2.76 |

For our best model, we also assess the permutation importance seen in Equation 2.3. This reduces the high cardinality bias seen in the feature importance charts by permuting against a held-out test set. This is done by each feature column being permuted against a baseline metric that was initially evaluated against the data set. The permutation importance is then established as the difference between the baseline metric and the feature column permutation. From this, we see that Age and FAQ maintain their high importance. However, APOE4 is now significantly more relevant in regards to the test set prediction (see Figure 10).

Table 7 shows the average standard deviation differences from the mean out of the subjects that the model predicted incorrectly. We omit our PTRACCAT (Race) feature into this analysis, given that it is not a continuous variable. In total, 20 EMCI_C ground truth subjects and 9 EMCI_NC ground truth subjects were incorrectly classified. By analyzing the standard deviation differences, we can determine which feature was most abnormal compared to the average model prediction for that given class.

We do this by establishing the data set means and standard deviations per feature for both the EMCI_C and EMCI_NC classes. We then take each misclassified patient's feature values and subtract them by the corresponding mean, prior to dividing them by that feature's standard deviation value. This allows us to see which features were the

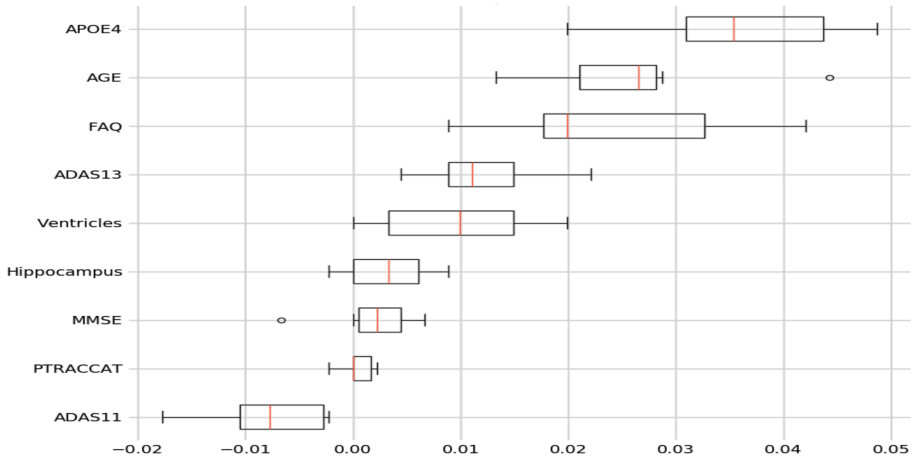



Figure 10: 9-Feature Model Permutation Importance

most unusual at an individual patient level. Coupled with the feature importance ranking this gives us clearer insight into the model's prediction rationale. For example, the MMSE feature was 2.37 standard deviations away from its EMCI_C mean, which contributed to our model misclassifying those cases as EMCI_NC. Of the EMCI_C misclassifications, MMSE proved to be the most misleading feature. However, we still saw higher overall accuracy by including it within our model because of its relatively low feature importance.

Out of the EMCI_NC misclassifications, there was less clarity as to which feature was problematic. However, we do observe a standard deviation increase in our neuropsychological test scores across the nine misclassifications. This is indicative of the model believing these subjects' test scores to be similar to those of the EMCI_C class and thus making the false prediction. In future work, we will explore whether knowing these misleading instances can help our model's accuracy, but currently, it appears that these are outliers within the ADNI dataset.

### 2.3.5 Assessment of individual predictors' feature importance

For the individual level, we can see which features specific to that patient made the largest contribution to their prediction. An example of these prediction contributions can be seen in Figure 11 based on the test patient's features provided in Tables 8 and 9. In this case, our model correctly predicted that this patient would convert to AD with an overall confidence of 90.4%. This confidence is a reflection of the aggregate of all of the individual trees' votes within our forest.

From Figure 11 and the contributions listed in Tables 8 and 9, we see that ventricular volume was the most predictive feature for this patient as it contributed 24.2% towards the model's decision. This was closely followed by their Functional Activities Questionnaire (FAQ) score which contributed 21.2% of the prediction. Race (PTRACCAT) was the only feature that contributed to the wrong prediction, albeit only a 0.3% prediction contribution for this patient. These individual feature importances are calculated in an identical manner to the model feature importance formulas above, however they do not include the model-level aggregation. In this sense, it allows us to interpret precisely why the decision trees chose a certain classification.

Tables 8 and 9 show the prediction contributions from subjects across both the EMCI_C and EMCI_NC classes. The PC column represents the amount of each feature's contribution to the overall prediction. A positive value indicates the contribution towards the ground truth class, whereas a negative value denotes the contribution towards the incorrect class. The PC's overall sum and average for the EMCI_C class's correctly classified cases are higher than 0.7 and 0.08 while ones for the EMCI_C class are lower

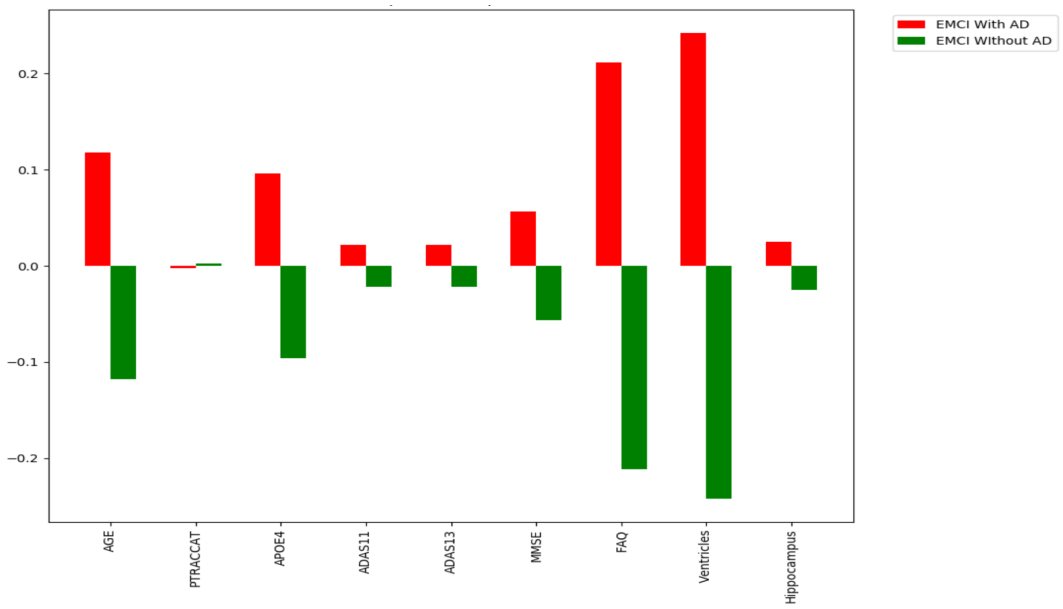



Figure 11: Feature Importance Example for an Individual Patient

than 0.061 and 0.008, respectively. The PC sum and average are strong indicators for the classification. As Table 7 is helpful for interpreting possibly misleading features at the model level, this individual PC metric allows us to better understand the model's decision making on a per subject basis.

## 2.4 Discussion

We have demonstrated that a random forest model can take clinical features and accurately predict MCI-to-AD conversion probability. Our RF classifier showed superior performance compared to competing SVM, XGBoost, and Logistic Regression implementations, including our own. It is also worth noting that the best models looked at all MCI patients, rather than the earlier EMCI subset. This gives our model the strength of predicting from 5-7 years prior to the onset of AD. Our results show that clinical features

Table 8: Example Features and Prediction Contributions (9FT PC) for EMCI_C Cases

| **EMCI_C Class** | | **Correctly Classified** | | | | **Misclassified** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **9 Features** | **Mean** | **Subject C1** | | **Subject C2** | | **Subject C3** | | **Subject C4** | |
| | | Value | PC | Value | PC | Value | PC | Value | PC |
| **Age** | 73.5 | 68 | 0.11 | 77 | **0.118** | 69.1 | 0.008 | 73 | **0.337** |
| **FAQ** | 4.1 | 0 | -0.036 | 10 | **0.212** | 4 | 0.007 | 3 | 0.039 |
| **ADAS13** | 15.8 | 20 | 0.042 | 17 | 0.022 | 10 | -0.008 | 9 | 0.025 |
| **ADAS11** | 9.7 | 13 | 0.039 | 9 | 0.022 | 5 | -0.006 | 4 | 0.013 |
| **MMSE** | 28.1 | 29 | 0.006 | 26 | 0.057 | 26 | 0.023 | 29 | -0.011 |
| **PTRACCAT** | – | 7 | -0.001 | 7 | -0.003 | 7 | 0.003 | 7 | 0 |
| **Hippocampus** | 6875.2 | 7853 | **0.173** | 6901 | 0.025 | 5576 | **0.158** | 7835 | -0.02 |
| **Ventricles** | 39282.7 | 38627 | **0.145** | 24285 | **0.242** | 35280.12 | -0.006 | 32379 | -0.031 |
| **APOE4** | 0.9 | 2 | **0.248** | 1 | 0.096 | 2 | **0.138** | 0 | -0.022 |
| **PC: Sum (AVG)** | | **0.726 (0.08)** | | **0.79 (0.087)** | | **0.317 (0.035)** | | **0.33 (0.036)** | |

can also outperform MRI-based models. This is important as obtaining neuropsychological scores, a significant subset of our chosen features can be far more affordable and less intensive than obtaining a patient's MRI imaging. With a more flexible approach, the expectation is that this predictor would be easier to deploy into a clinical setting.

In our experiments with the feature groupings, we found neuropsychological scores to be the most reliable and essential feature subset as we always experienced lower model accuracy with their exclusion. Performing tests on individual predictors also showed their weaknesses as each predictor demonstrated improved accuracy when coupled with an additional predictor. Even the neuropsychological scores by themselves exhibited signs of subjectivity, which were remediated by including biomarker and demographic features.

Additionally, our methods for oversampling an initially imbalanced data set can be of use throughout the medical research domain. With many medical data sets consisting of similar target class imbalance, our process enhances bagging algorithms by augmenting

Table 9: Example Features and Prediction Contributions (9FT PC) for EMCI_NC Cases

| **EMCI_NC Class** | | **Correctly Classified** | | | | **Misclassified** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **9 Features** | **Mean** | **Subject NC1** | | **Subject NC2** | | **Subject NC3** | | **Subject NC4** | |
| | | Value | PC | Value | PC | Value | PC | Value | PC |
| **Age** | 71.1 | 59 | 0.047 | 81 | 0.001 | 79.8 | **0.106** | 63.6 | 0.009 |
| **FAQ** | 1.82 | 4 | -0.037 | 1 | 0.056 | 5 | 0.099 | 17 | **0.158** |
| **ADAS13** | 13.3 | 16 | -0.001 | 21 | -0.005 | 17 | -0.013 | 37 | **0.346** |
| **ADAS11** | 8.5 | 13 | -0.002 | 14 | 0.009 | 12 | 0.019 | 27 | **0.208** |
| **MMSE** | 28.3 | 29 | 0.015 | 27 | -0.008 | 30 | -0.002 | 19 | -0.025 |
| **PTRACCAT** | – | 7 | 0.001 | 7 | 0.008 | 7 | -0.006 | 7 | -0.002 |
| **Hippocampus** | 7334.1 | 8303 | 0.023 | 6288 | 0.015 | 5437 | **0.256** | 7223.86 | -0.021 |
| **Ventricles** | 34504.6 | 22275 | -0.03 | 30260 | -0.001 | 69583 | -0.043 | 35280.12 | -0.019 |
| **APOE4** | 0.4 | 0 | 0.045 | 1 | -0.002 | 0 | -0.041 | 0 | -0.037 |
| **PC: Sum (AVG)** | | **0.061 (0.006)** | | **0.073 (0.008)** | | **0.375 (0.041)** | | **0.617 (0.068)** | |

more samples for the minority classes. For our purpose, this was only tested within a binary classification problem, however we will be extending this technique to multi-class problems.

One limitation of this study is that all of the patients were from the ADNI data set. While our accuracy was verified by splitting our data across multiple instances, we did not test the population outside of the ADNI participants. The inclusion of other data sets into our model would help account for even more significant variations and will be a target for future work.

In the future, we would like to combine this clinical features dependent model with our prior diffusion tensor imaging model [21] in order to create an ensemble predictor that can handle a large variety of available patient information. This would allow for greater flexibility for patient input data while maintaining high accuracy in the prediction.

Additionally, it is currently difficult to differentiate between the sub-types of dementia when a patient presents with cognitive and memory decline [22]. This can lead to an inaccurate treatment plan if the patient is misdiagnosed. Having the ability to predict additional sub-types at such an early stage would help significantly with pharmacological management [23]. Researching the differences between these sub-types based on this study's clinical features will be a subject of our future work.

In summary, we created a balanced random forest model based on multiple features to predict the MCI-to-AD conversion probability. In addition, we determined which features were most important for the overall model, as well as for individual patient predictions. We also took advantage of oversampling methods to better balance the target classes. As early detection is critical for both clinical trial enrollment and cost-effective treatment plans, we expect our work to help in clinical diagnosis as well as establishing treatment timelines. Our random forest model achieved state-of-the-art performance with an accuracy of 93.6% and showed that the combination of demographic, neuropsychological scores and biomarker features could be used to predict which EMCI patients are at a higher risk of AD.

# CHAPTER 3

# DIFFUSION TENSOR IMAGING (DTI) DEEP NEURAL NETWORK (DNN) MODEL

Building on the foundation laid in prior chapters, Alzheimer's Disease is an irreversible, progressive brain disorder that slowly destroys cognitive abilities. In recent years, the relationship between the prodromal Mild Cognitive Impairment (MCI) stage and the Alzheimer's Disease (AD) stage has been extensively researched in hopes of finding a path towards early diagnosis. Early detection at the MCI stage can help determine appropriate treatment plans as well as assist in clinical trial enrollment as 32% of individuals with MCI will develop AD within 5 years.

Computer vision studies leveraging Magnetic Resonance Imaging (sMRI, fMRI), Diffusion Tensor Imaging (DTI), and Positron Emission Tomography (PET) have led to encouraging results in classifying the different stages of AD. Studies around DTI specifically have shown that structural differences in white matter are prevalent between these stages. Rather than classification between stages, we propose a recurrent neural network model (RNN) based on the DTI modality for identifying the subset (32%) of individuals with Early Mild Cognitive Impairment (EMCI) that will develop AD. Our results demonstrate high accuracy in determining which individuals will develop AD within the next 5-7 years. Additionally, we propose our augmentation methods for DTI data as well as our classification accuracy across the traditional AD stage categories.

AD continues to be a leading cause of dementia, with a growing global prevalence.

In 2017, nearly 44 million people were reported to have Alzheimer's Disease. This number is expected to be over 100 million by 2050 [24]. There are many studies that demonstrate that as AD progresses many regions of the brain experience pathological structural changes. These structural changes are significantly different at varying stages of the disease [25]. A diagnosis of suspected AD is made through a battery of neuropsychologic testing combined with clinical interviews as the current lack of specific radio-diagnostic markers makes diagnosis and clinical research from neuroradiology unreliable. There is an existing need for further technological development of image-based tools for AD research. Previous attempts at image-based detection have focused on classifying the stage of the disease, rather than predicting whether an individual is likely to develop AD in the future [26]. As a result, the focus of our study is to create an image-based detection model that can predict 5-7 years prior which patients will progress to Alzheimer's Disease.

Gaining an increase in popularity with clinicians in recent years [27], Diffusion tensor imaging (DTI) provides a way to explore the micro-architecture of the brain by detecting how water moves along the white matter tracts. This form of magnetic resonance imaging (MRI) has revolutionized diagnostic imaging as it is capable of producing image contrast based on the water molecule diffusion difference within the brain [28]. In previous AD studies, tissue measurement differences based on DTI scans were tested, typically using the fractional anisotropy results [29]. Fractional anisotropy is represented by a zero to one value that describes the degree of anisotropy of a diffusion process. For this chapter, we explore the differences in apparent diffusion coefficient (ADC) DTI images. The ADC represents a measure of the magnitude of water molecule diffusion within tissue.

We then use these images to determine if they are statistically significant in predicting AD from the Early Mild Cognitive Impairment (EMCI) prodromal stage.

This chapter is organized as follows: we will first review the work that is being done with Alzheimer's Disease classification/prediction. We will describe our dataset and pre-processing steps. We will then describe our technical approach for data collection/augmentation as well as the specifics regarding our model and architecture. We will then evaluate the model's performance and report on the results achieved. Finally, we conclude with the chapter's primary insights and potential for future work in addition to the limitations that were present.

## 3.1 Related Work

A review on quantitative methodologies [30] summarized the recent brain studies in Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) while another researched conversion prediction from MCI to AD [31]. Similarly, detecting AD onset in MCI subjects were presented by Amoroso et al., [32].

While our work focused on the DTI modality, other work in this area attempted to use a combination of other modalities such as sMRI, fMRI, PET, and non-clinical data such as sociodemographic and other ancillary information.

Li et al. [33] proposed a deep learning framework for early prognosis of AD based on hippocampal MRI data. They trained a deep learning classifier based on the ADNI-1 cohort by extracting informative imaging features, and built a time-to-event prognostic model on these features. This predicted the progression to AD for MCI subjects of the

ADNI-GO/2 and AIBL cohorts.

Grassi et al. [34] proposed a study where their aim was to develop a new machine-learning algorithm to allow for a three-year prediction for conversion to AD in subjects diagnosed with MCI. Their algorithm aimed to achieve good predictive performance based only on a reduced set of sociodemographic characteristics, clinical information, and neuropsychological test scores. Their goal was to not rely on information coming from procedures that are expensive, invasive, or not readily available in many clinical settings, such as neuroimaging techniques, lumbar puncture, and genetic testing.

Khvostikov et al. [35] proposed an adapted architecture of a CNN for classification of 3D volumes of hippocampal ROIs and explored fusions of two modalities, sMRI and DTI. The authors ran a number of experiments with different configurations: used image modalities (sMRI, DTI), ROI sizes, number of convolutional layers, and number of convolutions in each layer before comparing them. They trained and evaluated 3 binary classifiers: AD-NC, AD-MCI, and MCI-NC. The results that the authors obtained confirmed that training Deep Neural Networks on 3D volumes and fusing different modalities can produce high classification accuracy. They report an accuracy of 90% for AD-NC and 80% for AD-MCI.

## 3.2 Dataset and Pre-Processing Steps

Our work used data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [36] which includes both the T2-weighted images as well as the DTI imaging. The ADNI was launched in 2003 as a public-private partnership with the

primary goal of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD)[36]. In addition to the DTI images, ADNI diagnostic summary data was used to distinguish between which patients, at the MCI stage, would later develop AD and those that would not develop AD.

For pre-processing, the first step was to download the Average DC DTI images from ADNI in DICOM format. After obtaining the set of DICOM images, these were then converted to the NIFTI format using MRIConvert. Once we separated the image folders by patient, we used the Nilearn Python module to extract the central slice from each axial image. After this, we divided the patients into classes based on their level of dementia, i.e. Cognitively Normal (CN), Mild Cognitive Impairment (MCI), Alzheimer's Disease (AD), and of utmost importance Early Mild Cognitive Impairment (EMCI) patients.

These classes are defined by ADNI based on the subject's test scores and clinical diagnosis [37]. Cognitively Normal (CN) is seen as the control group for the study as they show no signs of depression, dementia, or cognitive impairment. Mild Cognitive Impairment (MCI) subjects have reported a memory concern either to a clinician or autonomously, however they do not show any signs of dementia and their daily living activities are largely intact. Determining the level of MCI (early vs. late) is done via the Wechsler Memory Scale Logical Memory II test. The Early Mild Cognitive Impairment subjects represent the earliest stage of memory concern and are typically categorized as such 5-7 years prior to an AD diagnosis [38]. Finally, the Alzheimer's Disease (AD)

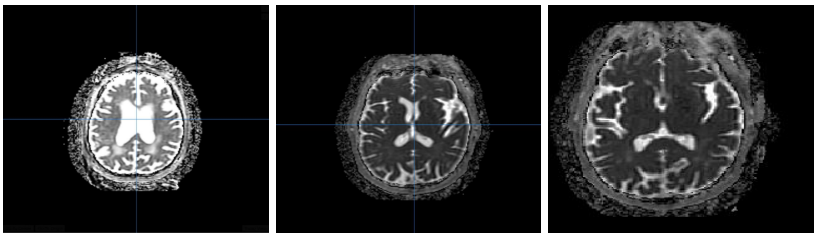


Figure 12: DICOM Format Example Downloaded from ADNI (Left), NIFTI Image Example After MRIConvert Conversion (Middle), and Final Training Image After NiLearn Python Script (Right).

group represents individuals who are experiencing significant mental deterioration that results in both cognitive and behavioral changes.

We further split the EMCI class into two subclasses by cross-referencing the ADNI Diagnostic Summary information in order to determine whether that individual would eventually develop Alzheimer's Disease or not. On average, 32 percent of individuals with Mild Cognitive Impairment will develop Alzheimer's Disease within the next 5 years of their follow-up appointments [39]. Since these EMCI scans are captured 5-7 years prior to AD diagnosis, identifying the 32 percent became our primary goal.

Going forward these subclasses will be referred to as EMCI with AD vs. EMCI without AD. Cross-referencing between the DTI images and the diagnostic summaries was done for each study ADNI1, ADNIGO, ADNI2, and ADNI3 once common variables were determined as they differed slightly between studies.

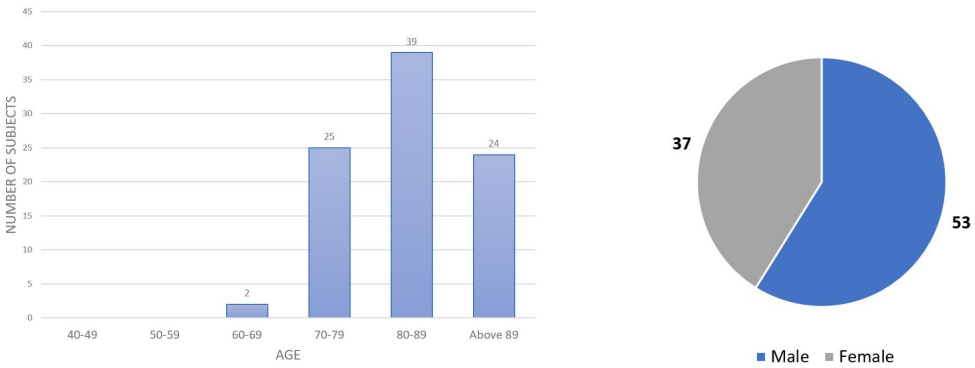



Figure 13: Participant Information (Age and Gender Distribution)

## 3.3 Methods

### 3.3.1 Data collection

As previously mentioned, the Average DC DTI images used for our study were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The Diagnostic Summary Information data that was used to cross-reference the MCI to AD converted patients was also obtained from the same ADNI dataset.

For the EMCI with AD vs. EMCI without AD comparison, 405 DTI scans of both male and female subjects were gathered. After our pre-processing steps, each of our corresponding 405 images represented a singular central slice of the complete DTI scan. These images were then grouped by patient into 90 distinct EMCI subjects. Of these 90 subjects, 16 had eventually been diagnosed with Alzheimer's Disease during the study. 74 of the subjects did not have an AD diagnosis at any point in the study.

### 3.3.2 Data augmentation

Overfitting is a common problem that can occur when training with smaller datasets. To help prevent this, our 405 EMCI images were then augmented with a few methods to improve our training set size. One method used was to mirror the images by flipping them horizontally. After performing only this augmentation, the accuracy of our model increased by eight percent. Building on this success, further augmentations were tested and the randomization of image brightness became another method that we added to our process. The random brightness is particularly useful for brain imaging as typically the input data will vary in brightness level. By inserting 30 random brightness variations per image into our training process, our accuracy improved by an additional six percent. Many other methods were tested such as random cropping or scaling but these techniques lowered the overall accuracy given the nature of our input data.

### 3.3.3 Architecture

For our work we decided to implement the NASNet architecture [40] as the backbone. The NASNet architecture was developed in 2017 by Google researchers and represents an automated machine learning structure which gathers and trains an entirely new neural network NASNet built upon Google's AutoML by redesigning the search space so that ideal layers for ImageNet classification, COCO object detection, and CIFAR-10 could be established efficiently [41]. It is the combination of these two layers that result in the new architecture. To substantiate our chosen architecture we performed comparisons between Inception V3 [42], NASNet [40], and PNASNet [43]. We initially began

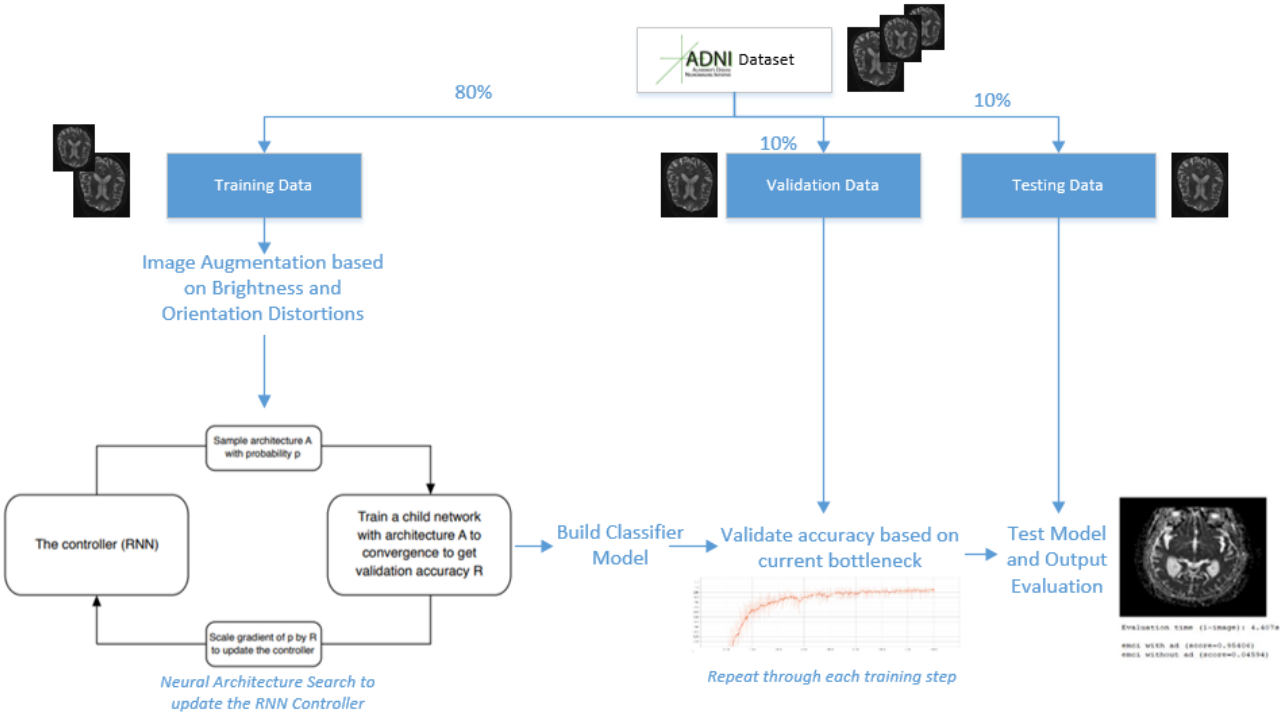



Figure 14: Model Workflow

our model on the Inception V3 architecture, but discovered that switching to NASNet boosted our accuracy 2% (96.4% vs. 94.7%). We also ran our model against the PNASNet architecture but found that it was 5.2% less accurate than our NASNet model (96.4% vs. 91.2%). As a result, we felt confident proceeding with our chosen architecture.

### 3.3.4 Training process

Transfer Learning is a technique that takes a model that has already been trained on a related task and reuses it for a new model. For our work, we reuse the feature extraction capabilities from powerful image classifiers that were trained on ImageNet [44]. We

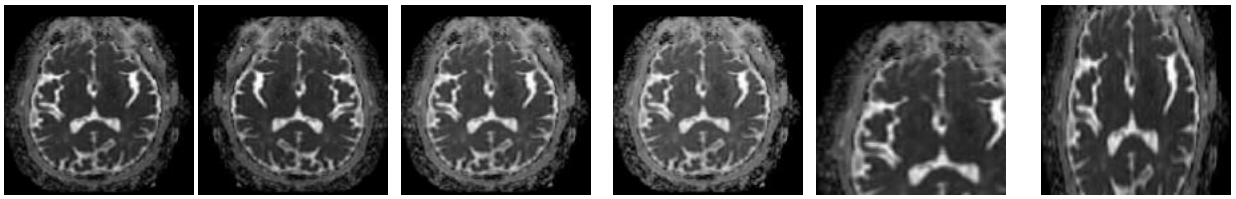
Table 10: Model Accuracy by Demographic

| Category | Group | Subject(%) | EMCI with AD | | EMCI without AD | |
|---|---|---|---|---|---|---|
| | | | Subject# | Accuracy(%) | Subject# | Accuracy(%) |
| Gender | Male | 59% | 11 | 96.8% | 42 | 97.3% |
| | Female | 41% | 5 | 95.8% | 32 | 96.1% |
| Age | 60-69 | 2% | 2 | 100% | 0 | N/A |
| | 70-79 | 28% | 8 | 95.5% | 17 | 96.3% |
| | 80-89 | 43% | 5 | 96.9% | 34 | 97.2% |
| | $> 89$ | 27% | 1 | 100% | 23 | 96.2% |

then retrain the final classification layer on our training set using TensorFlow. Starting in 2010, ImageNet is a visual database that consists of over 14 million hand-annotated images with over 20,000 categories [44]. This image database has been useful for prior medical classification in the radiological and pathological domains so we determined that it would fit well with our work [45].

For the training process, we use 80% of our images for the primary training set, with 10% being reserved for validation testing, and the remaining 10% reserved for the testing set. The validation set is then used while training in order to see the accuracy of the model as it progresses through each epoch.

In this work we test multiple hyperparameter configurations and image distortions as seen in Table 11. While initially our accuracy was low (67.7%), we are able to increase this significantly through parameter modifications. At first we thought cropping could increase our model's performance but given the consistent structure of DTI scans, we decide to exclude it. This exclusion along with our modified learning rate of .005 helps bring our model to greater than 80% accuracy.



(1) Normal (2) Flipped (3) 30% Brightness (4) 50% Brightness (5) 30% Crop (6) Scaled Wide Screen

Figure 15: DTI Input Distortion Examples

Table 11: Compared Model Configurations

| Model | **Accuracy** | **Arch.** | **Train. Steps** | **Scale Dist.** | **Bright. Dist.** | **Crop Dist.** | **Flipped Img.** | **Learn. Rate** |
|---|---|---|---|---|---|---|---|---|
| Model 1 | **96.4%** | NASNet [40] | 8000 | 0 | 30 | 0 | True | .005 |
| Model 2 | 94.7% | Inception V3 [42] | 8000 | 0 | 30 | 0 | True | .005 |
| Model 3 | 91.2% | PNASNet [43] | 8000 | 0 | 30 | 0 | True | .005 |
| Model 4 | 89.6% | NASNet [40] | 8000 | 0 | 0 | 0 | True | .005 |
| Model 5 | 79% | NASNet [40] | 10000 | 30 | 30 | 30 | True | .005 |
| Model 6 | 77.1% | NASNet [40] | 8000 | 30 | 30 | 30 | False | .005 |
| Model 7 | 75% | NASNet [40] | 10000 | 50 | 50 | 50 | True | .003 |
| Model 8 | 73.7% | Inception V3 | 8000 | 50 | 50 | 50 | True | .005 |
| Model 9 | 67.7% | PNASNet [43] | 8000 | 50 | 50 | 50 | True | .01 |

From there, we find that dropping the scale distortions gives a similar bump in performance as did the cropping exclusion. Randomly changing the brightness on the training data also proves to be key as it helps handle the scope of variations we see in our test data. Finally, by flipping the input images, we consistently see better performance ($\sim 2\%$ increase) so that parameter is left at True for the majority of our configurations. After optimizing these augmentation distortions, we train on 8000 steps and a learning rate of .005 to achieve our final model.

## 3.4 Results

In this chapter we assess the accuracy of using MRI DTI scans, taken at the initial EMCI diagnosis, to predict whether an individual will develop Alzheimer's Disease. We run many different configurations for our experiments: different architectures, distortions, and learning rates to arrive at our final model with an accuracy of 96.4%. This accuracy is based on the percent of the images in the test set that are given the correct label after the model is fully trained. We group our findings by gender and by age group to determine if there is any significant performance difference between them. In Table 10 we see that our age groups were ranged as follows: 70-79, 80-89, and greater than 89. As far as we are aware, our accuracy is state of the art as other MCI-to-AD prediction models can be compared in Table 12.

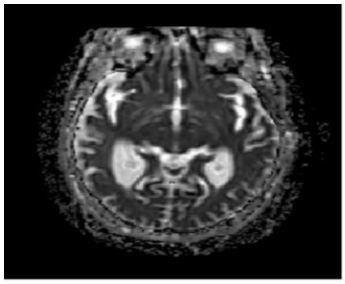
In addition to our goal of AD prediction based on the EMCI stage, our deep learning model also performs well with classification between the standard categories (NC, MCI, AD). For differentiating between AD and MCI subjects our accuracy is 98.1%. When comparing MCI to NC subjects, our accuracy is 95.2% as the water molecule diffusion is slightly less indicative between those classes. Finally, when comparing AD to NC subjects, our accuracy is 99.9% as the difference in DTI imaging is at it's greatest between those categories. Related classification methods and their accuracy for these standard categories can be seen in Table 13.

Figure 16 shows an example of our model's classification output for a single image. The confidence score is displayed across our two classes, EMCI with AD and EMCI

without AD. On a per image basis, this score would originally range between 71.4%-98.2% confidence with the lower scores being tied to darker images. However, after our brightness augmentations, the overall confidence range improved to 82.5%-98.8% as our model was more confident in low-light inputs. These score ranges do not impact our overall reported accuracy of 96.4% as that accuracy is derived on a binary basis. That is, did the model predict the correct class, rather than the question of how confident was the model in predicting the correct class?

Looking at Figure 18 we see additional samples of correctly classified test inputs. As seen in these examples, our model predicts the patient's AD trajectory accurately despite differing qualities of input. This was important to achieve as DTI scan quality can vary greatly depending on the process and machine used. Figure 17 provides a case where our model misclassified the image. The confidence of the model in this instance was 52% as it was near the border of the two classes. Lack of clarity and axial centrality contributed to the lower score on this image.

Since we use TensorFlow as our deep learning library, we are able to output the Tensorboard graphs seen in Figure 19. These graphs show the loss (Fig. 19(a)) and accuracy (Fig. 19(b)) of our model as it trains. This helps us determine the relevant amount of steps that we should use as we can see how the accuracy interval diminishes over time. In our experimentation, we tried multiple step configurations ranging from 1000-20000 but found 8000 to be the best point for loss minimalization. Additionally, by stopping at 8000 steps we take an extra measure to avoid potential overfitting of the model. For our work, we perform these experiments on an Intel Core i5-9600K with an Nvidia RTX 2080 using



```
Evaluation time (1-image): 4.407s

emci with ad (score=0.95406)
emci without ad (score=0.04594)
```

Figure 16: Example Code Output for AD Prediction for an EMCI Subject.

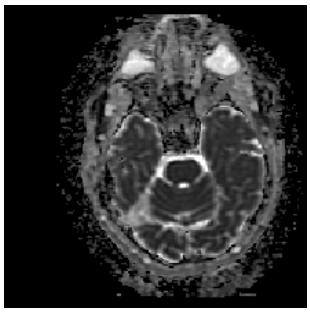


Figure 17: Example of a Misclassified Image. Ground Truth is EMCI with AD but Prediction Showed EMCI Without AD (Albeit at Low Confidence).

the TensorFlow GPU libraries.

### 3.5 Limitations and Future Work

There are some limitations in our study in regards to the ADNI dataset. First, the different ADNI studies (ADNI1, ADNI2 etc.) used scanners that possessed different field strengths, so combining these studies could have affected our results. Additionally, since many ADNI subjects are still alive there is no guarantee that individuals that have yet to

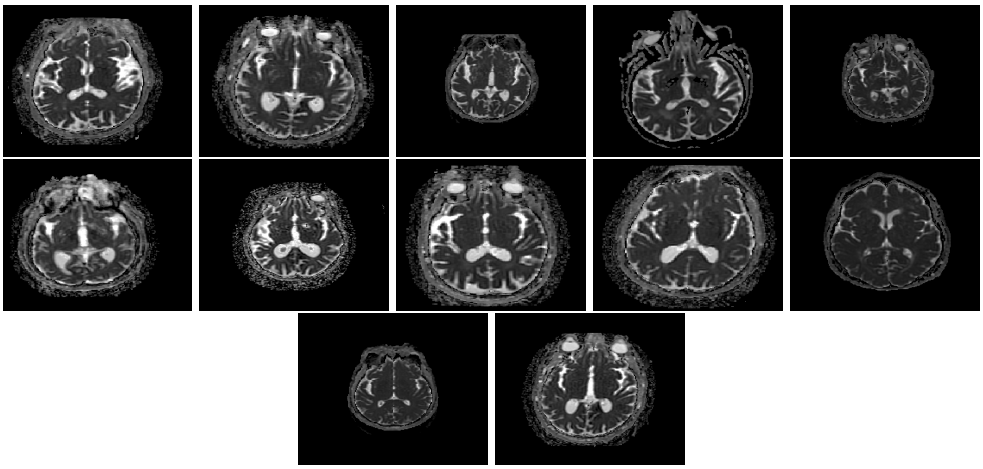
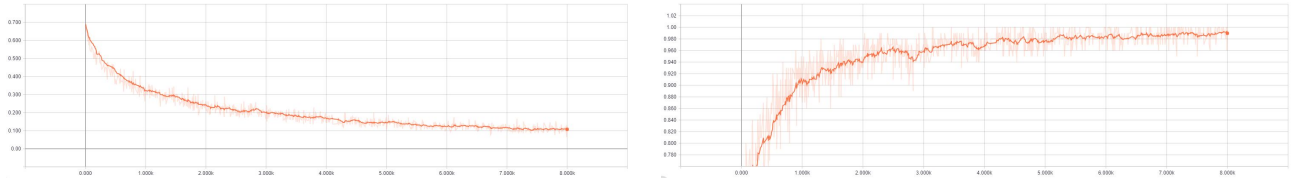


Figure 18: Accurately Predicted Results for EMCI Subjects with Eventual AD (Top) and EMCI Subjects Who Would not Develop AD (Bottom).



(a) Loss (b) Accuracy

Figure 19: Tensorboard Training Metrics: Step-wise Loss and Accuracy

Table 12: Comparative Evaluation for MCI-to-AD Prediction

| Approach | **Modal.** | **Data** | **Model** | **MCI-to-AD Pred.** | **Year** |
|---|---|---|---|---|---|
| Velazquez et al. (our Model 1) | DTI | ADNI | RNN NASNet | **96.4%** | 5 |
| Spasonv et al. (2019) [31] | sMRI | ADNI | 3D CNN | 92.5% | 3 |
| Grassi et al. (2019) [34] | Cognitive measures | ADNI | Ensemble based ML | 88% (ROC) | 3 |
| Li et al. (2019) [33] | PET + clinical data | ADNI | CNN | 81.3% (ROC) | 3 |
| Moscoso et al. (2019) [46] | MRI | ADNI | ML | 84% (AUC) | 5 |
| Khvostikov et al. (2018) [35] | sMRI + DTI | ADNI | 3D CNN | 80% | 5 |

be diagnosed with Alzheimer's won't be diagnosed with the disease in the future. For our work, we sought to reduce this bias by using subjects that were long-time participants in ADNI. We do not believe this would impact our accuracy significantly given our sample size.

Another limitation is that the ADNI subjects are not a real representation of the general population given the different inclusion criteria per study. Also, the demographics of our smaller sample of 90 subjects does not align to the existing population demographics.

While our accuracy was verified by multiple instances of splitting our data, we did not test against the broader population outside of the ADNI participants. Inclusion of other datasets into our model could potentially help with even more variations seen in the DTI scan inputs. Despite these limitations, our work shows that there are detectable tissue differences between the water molecule diffusion of EMCI subjects who would go on to develop AD, and those that would not. Furthermore it shows the effectiveness of using DTI imaging instead of the traditional sMRI approach as a means for early AD detection.

Table 13: Alzheimer's Classification Methods Compared

| Approach | **Modalities** | **Data (size)** | **Model** | **AD/NC** | **AD/MCI** | **MCI/NC** |
|---|---|---|---|---|---|---|
| Velazquez et al. (our Model 1) | DTI | ADNI (405) | RNN NASNet | **99.9%** | **98.1%** | **95.2%** |
| Ahmed et al. (2015) [47] | sMRI | ADNI (509) | Visual Feature/SVM | 83.8% | 69.5% | 62.1% |
| Ebadi et al. (2017) [48] | DTI | custom (34) | | 80% | 83.3% | 70% |
| Lee et al. (2015) [49] | DTI | LONI (141) | SVM | 97.7% | 97.7% | - |
| Lei et al. (2016) [50] | sMRI + PET | ADNI (398) | | 96.9% | - | 86.6% |
| Ahmed et al. (2015) [26] | sMRI + DTI | ADNI (203) | | 90.2% | 76.6% | 79.4% |
| Payan and Montana (2015) [51] | sMRI | ADNI (2265) | 3D CNN | 99.3% | 100% | 94.2% |
| Billones et al. (2016) [52] | sMRI | ADNI (900) | CNN VGGNet | 98.3% | 93.9% | 91.7% |

For our future work we would like to see the differences in accuracy when using multi-modality. Some of the related work shows the additions of sMRI to the DTI modality [26], but it is unclear as to whether that modality addition would improve our performance. As ADNI provides a notable amount of clinician notes, we'd like to incorporate that data into our model to see if it can substantiate the prediction confidence. Additionally, we would like to explore contrasting a patient's prescribed medication list with their progression towards AD. In combination with DTI imaging, this would allow us to build a time-series analysis based on their scan dates.

Performing an analysis on ADNI's Significant Memory Concern (SMC) subjects would also be a category that we'd like to target. SMC subjects are distinguished by having a self-reported significant memory concern that is quantified with the Cognitive Change Index and by having a Clinical Dementia Rating of zero [37]. These subjects do not have progressive memory impairment concerns and thus have not been diagnosed with MCI. If early detection at this level is possible, it would prove very important as many AD clinical drug trials have difficulty due to damage from AD already being done by the time of their participation [53].

From a clinical perspective, when patients present with cognitive and memory decline, current limitations in diagnostics do not allow for a clear differentiation between sub-types of dementia. This commonly leads to misdiagnosis and administration of incorrect medications which can lead to harmful adverse events. Being able to determine the dementia sub-type as early as possible could drastically help with pharmacological management and end-of-life planning for patients and families [24]. Exploring the white

matter differences between these sub-types will be a topic of our future research.

## 3.6 Conclusion

In this chapter, we created a deep learning model capable of predicting Alzheimer's Disease from the EMCI stage based on DTI imaging. As 32% of EMCI patients will develop AD within 5 years, early detection at this point (5-7 years prior to AD) is key to providing an accurate and cost-effective treatment plan. This detection could also aid in clinical trial enrollment as they could potentially assess the prognosis of subjects at a much earlier phase. Our model achieved state-of-the-art performance with accuracy higher than 96% and demonstrated that water molecule diffusion differences can be used to distinguish between patients at-risk for AD. While these results display the effectiveness of DTI as an input, it will be interesting to add additional modalities as well as predictions for other dementia subsets in our future research.

# CHAPTER 4

# MULTIMODALITY ENSEMBLE MODEL

This chapter centers on individualized EMCI (the earliest MCI subset) to AD conversion prediction on multimodal data such as diffusion tensor imaging (DTI) scans and electronic health records (EHR) for their patients using the combination of both a balanced random forest model alongside a convolutional neural network (CNN) model. Our random forest model leverages EHR's patient biometric and neuropsychiatric test score features, while our CNN model uses the patient's diffusion tensor imaging (DTI) scans for conversion prediction.

In our previous chapters, we focused on the prediction aspect of this problem within a single modality, such as electronic health records (EHR) or medical images. This led to a machine learning model (i.e., random forest) that focused on EHR patient clinical data and a Convolutional Neural Network (CNN) model that performed predictions based on Diffusion Tensor Imaging (DTI) scans. While these models performed well, each had limitations and was not focused on being explainable. It became clear that combining these models into an ensemble multi-modality model with the added feature of explainability would be ideal for EMCI conversion prediction.

Additionally, the explainability of a model's predictions has been challenging to determine or is sometimes an afterthought. This has led to many high-performing prediction models that do not provide a clear rationale to healthcare providers. With explainable

models, clinicians can be more confident in their diagnoses when leveraging a clinical decision-making tool. For our multi-modal work, explainability was a key objective.

Therefore, our work focused on developing a multi-modality ensemble model for AD conversion prediction that could explain the rationale behind its predictions. The first piece leverages a random forest, a supervised learning algorithm that is efficient with classification problems [54]. This would focus on interpreting patient clinical features while the other side of the ensemble, the CNN model, would handle a patient's Diffusion Tensor Imaging (DTI) scans.

Diffusion tensor imaging (DTI) is a form of magnetic resonance imaging (MRI) that detects how water moves along the brain's white matter tracts. This water molecule diffusion difference can then be contrasted to show the variation between scans. Our work centered on apparent diffusion coefficient (ADC) DTI scans. ADC measures the magnitude, within the tissue, of water molecule diffusion.

As our classes are originally imbalanced, we provide determinations on how to best balance our data as well as which augmentation forms are most appropriate. In addition, a method for dynamically choosing the ideal weight of each model within the ensemble for any given prediction is also provided. Finally, complete ensemble explainability of both the visual and clinical feature inputs is provided, and analysis is performed. The main contributions in this chapter are (1) building an ensemble model against an imbalanced data set; (2) determining the ideal weighting of that ensemble per patient prediction; (3) explaining model prediction rationale for both visual and clinical features; (4) determining the conversion prediction accuracy of our model. We believe that this work will

provide an understandable tool that can be used to predict patient AD conversion from a prodromal stage. In addition, this work provides both global and local explainability methods for ensemble models.

## 4.1 Related Work

### 4.1.1 MCI-to-AD Conversion Prediction

As the AD conversion problem matures, multiple studies now evaluate based on different mixes of modalities. [55] used a combination of graph theory and machine learning to predict the conversion of MCI subjects to AD based on sMRI/fMRI data. Their work explored multiple feature selection methods (e.g., random subset feature selection algorithm, minimal redundancy maximal relevance, and sparse linear regression) and achieved an accuracy of 84.71%. They also explored the relationship between AD conversion and high-sensitivity brain regions to find that both structural and functional areas were relevant as predictors.

An evaluation between unimodal and multimodal models for AD conversion was performed by [56]. In their work, MRI-derived biomarkers in combination with neuropsychological measures were used to determine early AD warning signs from an MCI population. They achieved an AUC of 95.7% with their multimodality data trained through a support vector machine (SVM).

Lin et al. [57] fused four modalities (MRI, positron emission tomography, cerebrospinal fluid biomarkers, and gene data) which were then individually graded using their Extreme Learning Machine (ELM) model. Their scope focused on conversion prediction

within three years as they achieved an accuracy of 84.7%. In addition, their findings demonstrated a minimum 10% increase in accuracy from using multiple modalities rather than when only a single modality was used.

Focusing on a reduced set of sociodemographic, characteristics, clinical information, and neuropsychological test scores, [14] developed a new machine-learning algorithm for three-year AD conversion prediction. Their work aimed to leverage data that did not derive from expensive, invasive, or otherwise difficult procedures such as lumbar puncture, genetic testing, or neuroimaging techniques. With these restrictions, they could still obtain an AUC of 88% through an SVM.

Huang et al. [58] proposed a predictive nomogram that combined AB concentration, image features, and clinical factors to predict MCI-to-AD conversion. Analysis was also performed on how features were associated with one another and the significance of each feature. To better understand the patterns of AD conversion, they focused on examining the associations at both the micro and macro levels.

Varatharajah et al. [59] focused on which markers would be most relevant for AD conversion models. Using a mix of clinical data, MRI, and FDG-PET, they could isolate large shares of variance in the pathophysiology (amyloid, tau) variables. Their work also revealed the relevance of CR1 (complacent receptor 1) as an individual predictor of AD conversion. As a result of their work, they achieved an AUC of 93% via an SVM.

Rana et al. [60] created MudNet, a CNN model which performed both MCI-to-AD conversion prediction and time-to-AD conversion. They could group patients into high-risk and low-risk categories based on whether they were predicted to convert within

24 months. Their model used a mixture of volumetric MRI and clinical data, which also consisted of neuropsychological tests (RAVLT, ADAS-11, ADAS-13, ADASQ4, MMSE). With these inputs, they achieved an accuracy of 69.8% for conversion predictions and 66.9% for risk classification.

### 4.1.2 Explainability

Explainability for AI models has seen increased demand over recent years. Historically, models were seen as black boxes, but now multiple explainability methods can be used to provide the rationale behind a model's behavior. For the medical domain, this is highly relevant as it allows physicians to understand the process that a decision support system uses to arrive at its recommendation. In one study, Viton et al. [61] focused on using heatmaps to visually explain a CNN model's predictions on in-hospital mortality. Their multivariate time series approach allowed for critical points to be identified and the most influential variables. The visual aid can help justify the model's decisions with this detailed explainability. While the purpose of their work was explainability, they were still able to achieve an AUC of .8207, predicting in-hospital mortality risk.

Maweu et al. [62] also worked to provide an explainable framework for their CNN model. In this case, they targeted ECG signals (one-dimensional time-series data). An interesting aspect of their approach is that they leveraged 1D-CNN models rather than the standard 2D-CNN ones. This allowed them to display descriptive statistics, feature visualization, detection, and mapping for each module of their proposed framework. With this knowledge, they could further explore the relationships between their features and

how they might contribute to misclassification. This idea of identifying the rationale behind misclassification was a focus of our work and will be analyzed within our results.

### 4.1.3 Ensemble Classification

Ensemble classifiers have proven to be highly efficient in recent years. The majority of these ensembles are typically similar algorithms which are then stacked. This usually consists of stacked ML algorithms (RF, SVM, XGBoost, etc.) or a chain of deep learning models. Some ensembles represent the combination of an ML model and a deep learning model, which is the approach we took for this work. We seek to add to this domain by including an explainability layer by combining a random forest classifier with a CNN model.

Mostafiz et al. [63] performed Covid-19 detection via chest x-rays by combining a random forest with a CNN model. After scanning, the initial x-ray is then enhanced and segmented before the key features are extracted. The random forest is then used for detection once the key features have been passed. Their work achieved an accuracy of 98.5% when both sides of their ensemble were engaged. However, when only one side was leveraged, their accuracy dropped to 84%, showing the advantage of ensemble classification.

Another study by [64] combined multiple machine learning algorithms (SVM, Random Forest, MLP, etc.) on top of a CNN, which served to draw out comparisons between the different methods. Their goal was exoplanet detection, and they were able to achieve 99.62% accuracy with their Ensemble-CNN model.

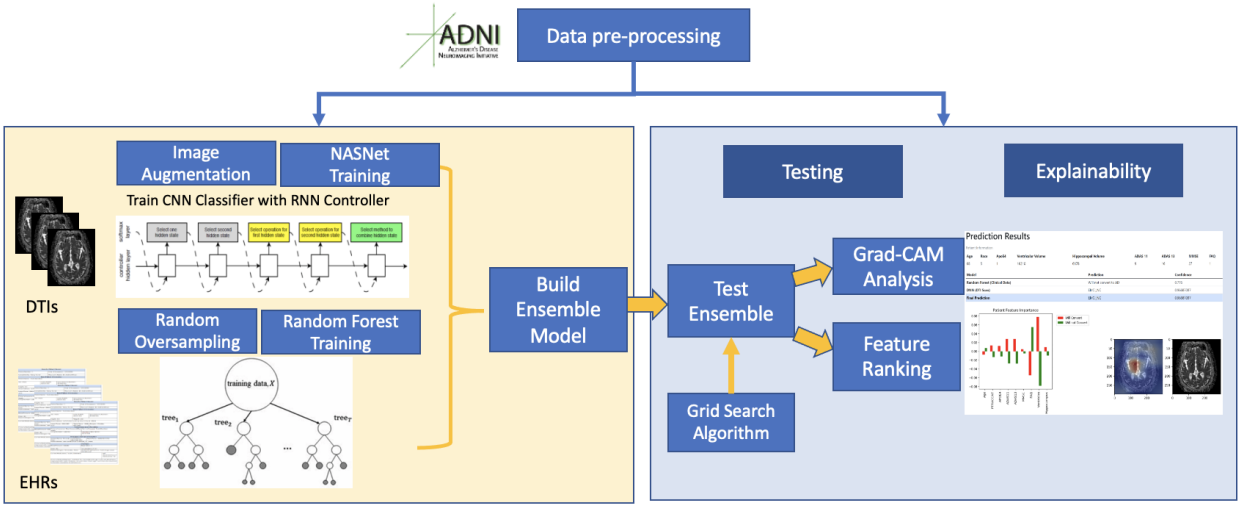



Figure 20: Ensemble Model Workflow

## 4.2 Methods

### 4.2.1 Data Collection

All data used for this chapter were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and included patients from their ADNI-1, ADNI-2, and ADNI-GO studies [12]. "ADNI is a global research study that actively supports the investigation and development of treatments that slow or stop the progression of AD" [12]. ADNI aims to track AD progression using biomarkers and clinical measures to assess the brain over each stage of the disease.

The selection criteria for our work focused specifically on the EMCI subset with patients that had follow-up exams for more than a year. EMCI patients represent the stage typically 5-7 years before a potential AD diagnosis. The Wechsler Memory Scale Logical Memory II test determines this earlier subset compared to the more general MCI

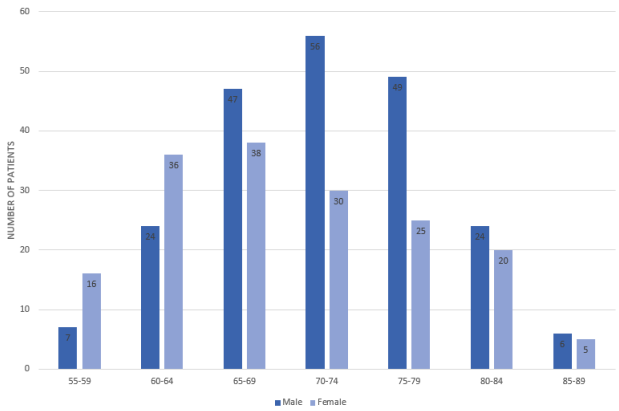



Figure 21: Subjects' Age and Gender Distribution

stage. For our classification problem, the EMCI patients were divided into two classes (EMCI_C, EMCI_NC). EMCI_C represents patients that would eventually convert to an AD diagnosis, whereas EMCI_NC represents patients that would not convert. This distinction was provided by the Clinical Dementia Rating ADNI variable of the patient's last exam diagnosis.

For the clinical feature model, 1806 exam visits were used pre-augmentation. 1608 belonged to the EMCI_NC class, while 198 were from the EMCI_C conversion class. For the DTI model, 405 DTI images were gathered, which, after our pre-processing methods, represented a singular central slice of each scan. These were then grouped into 90 unique EMCI patients, where 16 would convert to AD (EMCI_C) and 74 would not (EMCI_NC). In total, our study consisted of 383 EMCI patients (shown in Figure 21), 49 of these within the EMCI_C class and 335 within the EMCI_NC class. Stratified by age, our largest demographic was ages 70-74, followed by 65-69. Our training/test split for this work was 75% (288 patients) to 25% (95 patients).

Table 14: Clinical Feature Characteristics

| | | EMCI_C | | EMCI_NC | |
|---|---|---|---|---|---|
| | **Subject#** | 49 | | 335 | |
| **ADNI Feature** | **Description** | Mean | SD | Mean | SD |
| DX | Diagnosis | – | – | – | – |
| *Demographic information* | | | | | |
| PTRACCAT | Patient Race | – | – | – | – |
| AGE | Patient Age | 73.5 | 6.47 | 71.1 | 7.49 |
| *Genetic Biomarkers* | | | | | |
| APOE4 | The number of e4 alleles of APOE | 0.9 | 0.71 | 0.4 | 0.46 |
| *Physical Biomarkers* | | | | | |
| Hippocampus | Hippocampal volume | 6875.2 | 947.45 | 7334.1 | 910.20 |
| Ventricles | Ventricular volume | 39282.7 | 21031.66 | 34504.6 | 21394.49 |
| *Neuropsychological scales* | | | | | |
| ADAS13 | 13-item AD Assessment Scale | 15.8 | 6.02 | 13.3 | 5.41 |
| ADAS11 | 11-item AD Assessment Scale | 9.7 | 4.12 | 8.5 | 3.29 |
| FAQ | Functional Activities Questionnaire | 4.1 | 4.38 | 1.82 | 2.50 |
| MMSE | Mini-Mental State Examination | 28.1 | 1.58 | 28.3 | 1.71 |

*EMCI_C* the converter group, *EMCI_NC* the stable group

### 4.2.2 Clinical features selection

For the random forest component of our ensemble model, nine ADNI features were chosen, as seen in Table 14. These features contained physical biomarkers (ventricular and hippocampal volume), genetic biomarkers (APOE4), neuropsychological scale scores (FAQ, MMSE, ADAS13, ADAS11), and demographic variables (age, race). Initially, starting with over 90 features, we could eliminate many variables with a combination of SHAP analysis and Gini importance until an ideal fit had been obtained.

### 4.2.3 Ensemble Classification Model

We assemble an ensemble model that combines Random Forest clinical feature prediction alongside a Convolutional Neural Network (CNN) that performs predictions

based on diffusion tensor imaging (DTI) scans to take advantage of our multimodality data. This allows each model's limitations to be mitigated by engaging the other model for its prediction confidence.

Random Forest, our first classifier, uses a method that constructs a multitude of decision trees which then outputs the majority vote as the prediction. As subsets of features are randomly selected for each decision tree, this provides enhanced tolerance for overfitting. For our work, this classifier can either output EMCI_C (conversion class) or EMCI_NC (stable class). Each decision tree, made up of a random assortment of our nine clinical features, gets to cast a vote. Overall prediction confidence can be determined by observing how many trees voted for the majority class. As we can assess each node's importance in a given tree, we can evaluate each feature's importance for both the model and individual predictions. This allows us a measure of explainability for the clinical feature aspect of our overall ensemble model. Other classifiers were evaluated per Table 15, but the Random Forest algorithm provided the best performance. Additional specifics on this random forest model can be seen in our prior work [65].

Our second classifier consists of a Convolutional Neural Network (CNN) with a NASNet architecture [40] as its backbone. CNN models have previously demonstrated accuracy with MRI scans as they are built to process pixel data [21]. These capture spatial and temporal dependencies within an image, making them ideal for image classification. We had initially built our model with the Inception v3 architecture [42] but found better performance with NASNet. [40] integrate reinforcement learning with a controller RNN to construct a cell or layer for the NASNet network, which delivers cutting-edge ImageNet

Table 15: Clinical Data Classifier Performance Comparison

| Model/Feature | Accuracy | Precision | Recall | F1 Score | AUC | p-value |
|---|---|---|---|---|---|---|
| **Random Forest** | | | | | | |
| 6-Features | 0.892 | 0.907 | 0.980 | 0.942 | 0.88 | 0.91 |
| 9-Features | **0.936** | **0.952** | 0.978 | **0.965** | **0.96** | **0.71** |
| 13-Features | 0.916 | 0.916 | **0.998** | 0.955 | 0.93 | 0.82 |
| **Support Vector** | | | | | | |
| 6-Features | 0.900 | 0.900 | 1 | 0.948 | 0.52 | - |
| 9-Features | 0.900 | 0.900 | 1 | 0.948 | 0.54 | - |
| 13-Features | 0.900 | 0.900 | 1 | 0.948 | **0.55** | - |
| **Logistic Regression** | | | | | | |
| 6-Features | 0.894 | 0.902 | 0.990 | 0.944 | **0.76** | - |
| 9-Features | 0.892 | 0.903 | 0.985 | 0.942 | 0.75 | - |
| 13-Features | **0.896** | **0.904** | **0.990** | **0.945** | 0.75 | - |
| **XGBoost** | | | | | | |
| 6-Features | 0.898 | 0.904 | **0.993** | 0.946 | 0.87 | - |
| 9-Features | **0.920** | **0.930** | 0.985 | **0.957** | **0.89** | - |
| 13-Features | 0.907 | 0.921 | 0.980 | 0.950 | 0.88 | - |

accuracy. Our CNN model combines a NASNet architecture with an RNN controller to recursively search for the best structure as it trains. Creating a network with NASNet makes the search strategy significantly more successful for PNASNet [43].

As shown in Figure 22, these blocks consist of both standard and reduction cells. Normal cells represent convolutional cells that return a feature map of the same dimension. In contrast, reduction cells produce similarly but with the height and width reduced by a factor of two [66]. These are the only structures that the RNN controller subsequently searches. As seen in Table 16, other architectures were evaluated, but NASNet was the leading performer. As a result, this became our ideal architecture despite its computationally intensive approach. As NASNet was trained on ImageNet's 1.2 million images, our

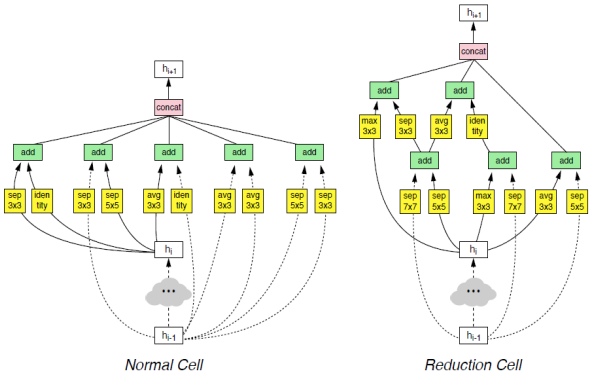



Figure 22: NASNet Architecture

ADNI data was used to retrain the final classification layer using TensorFlow.

We then combine these classifiers to form our ensemble model. This allows us to intake either clinical data, DTI scans, or both to accurately predict AD conversion while mitigating each classifier's weaknesses. A grid search algorithm is then performed to exhaustively determine the ideal weight that each classifier should carry within the ensemble. This optimization (Table 18) resulted in a .55 CNN vs. .45 RF weighting as the ideal balance for AD conversion prediction.

### 4.2.4 Data Balancing

Given the nature of our imbalanced data set, with 12.8% of patients belonging to the minority class (EMCI_C), we implement different augmentation methods for our ensemble to have better representation in our training/test data. We perform random over-sampling for our Random Forest classifier to make our two classes equivalent in size. This is done by taking random samples from the EMCI_C class with replacement until the size matches that of the majority class. This provides 2,412 total exam visits for

training rather than the original 1,354 pre-augmentation visits. Our over-sampling method was compared against both under-sampling methods and class weight modifications but continued to perform best.

For the CNN classifier, multiple augmentation methods were performed against our initial 405 EMCI images to increase the overall training size. The most effective augmentation methods were to flip the scans horizontally and to add randomization to an image's brightness. As these scans come in at different brightness levels, augmenting this allowed our model to learn at a far better rate. Variations of cropping or scaling the images did not increase our accuracy. A comparison of our visual augmentation methods sorted by accuracy can be seen in Table 16. Additionally, compared to our augmented data set, our original data set can be observed in Table 17. This table demonstrates our train/test data split and the initial class imbalance.

### 4.2.5 Grid Search Algorithm

We perform a grid search to exhaust possible weight combinations to find the ideal weighting for our ensemble model. First, we define our possible weight values for each model as 0.0 to 1.0 and then iterate through the process in steps of 0.1. After each weight vector is generated, they are normalized to ensure that they sum to one. Once the grid search has been completed, the weights of the highest accuracy run (.55 CNN, .45 RF) are captured and used for the final ensemble model. Other weighting combinations can be observed in both Table 18. As each model running independently is also contained within this table (1, 0 and 0, 1), the advantage of using both the ensemble approach and dynamic

Table 16: Comparative Evaluation with CNN Architectures and Augmentation Methods

| **Architecture** | **Accuracy** | **Model** | **Training Steps (#)** | **Scale Distortion** | **Brightness Distortion** | **Crop Distortion** | **Flipped Images** | **Learning Rate** |
|---|---|---|---|---|---|---|---|---|
| NASNet | **96.4%** | $\text{NASNet}_{M1}$ | 8000 | 0 | 30% | 0 | True | .005 |
| | 89.6% | $\text{NASNet}_{M2}$ | 8000 | 0 | 0 | 0 | True | .005 |
| | 79% | $\text{NASNet}_{M3}$ | 10000 | 30% | 30% | 30% | True | .005 |
| | 77.1% | $\text{NASNet}_{M4}$ | 8000 | 30% | 30% | 30% | False | .005 |
| | 75% | $\text{NASNet}_{M5}$ | 10000 | 50% | 50% | 50% | True | .003 |
| Inception | 94.7% | $\text{Inception}_{M1}$ | 8000 | 0 | 30% | 0 | True | .005 |
| | 73.7% | $\text{Inception}_{M2}$ | 8000 | 50% | 50% | 50% | True | .005 |
| PNASNet | 91.2% | $\text{PNASNet}_{M1}$ | 8000 | 0 | 30% | 0 | True | .005 |
| | 67.7% | $\text{PNASNet}_{M1}$ | 8000 | 50% | 50% | 50% | True | .01 |

Table 17: Data Set by Modality and Class

| | **Clinical Data** | | **DTI** | |
|---|---|---|---|---|
| Data | EMCI_C | EMCI_NC | EMCI_C | EMCI_NC |
| Subject# | 49 Subjects | 335 Subjects | 16 Subjects | 74 Subjects |
| Original Record# | 198 | 1608 | 72 images | 333 images |
| Record# after Over-sampling/ Augmentation | 1608 | 1608 | 576,000 images | 2,664,000 images |
| Training Data | 1206 | 1206 | 432,000 images | 1,998,000 images |
| Testing Data | 402 | 402 | 144,000 images | 666,000 images |

Table 18: Weighted Average Classifier Accuracy Compared

| Iteration | CNN Weight | RF Weight | Accuracy |
|---|---|---|---|
| 1 | .55 | .45 | **98.81%** |
| 2 | .60 | .40 | 97.62% |
| 3 | .40 | .60 | 94.05% |
| 4 | .70 | .30 | 96.43% |
| 5 | .65 | .35 | 96.43% |
| 6 | .50 | .50 | 95.24% |
| 7 | 0 | 1 | 92.86% |
| 8 | 1 | 0 | 96.43% |

weighting can be easily compared.

When individual patients are submitted to our model, each classifier (RF and CNN) generates a prediction and its confidence in that prediction. Our grid search-derived weighting is then factored into this prediction confidence (PC) to determine the overall ensemble prediction. As there can be disagreements between the different modalities, this weighting allows us to slightly prefer the more accurate classifier (CNN).

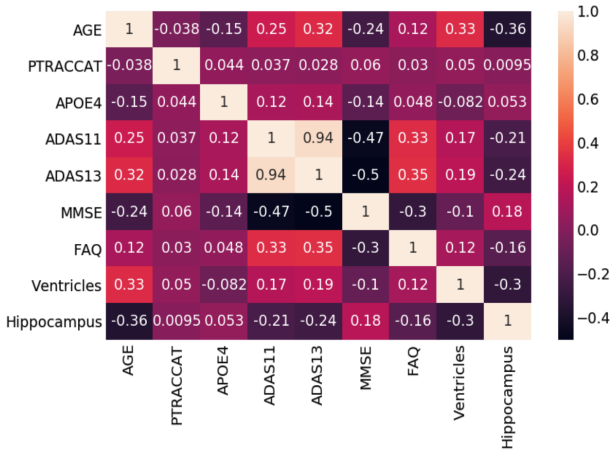



Figure 23: Random Forest Model Correlation Matrix

## 4.3 Results

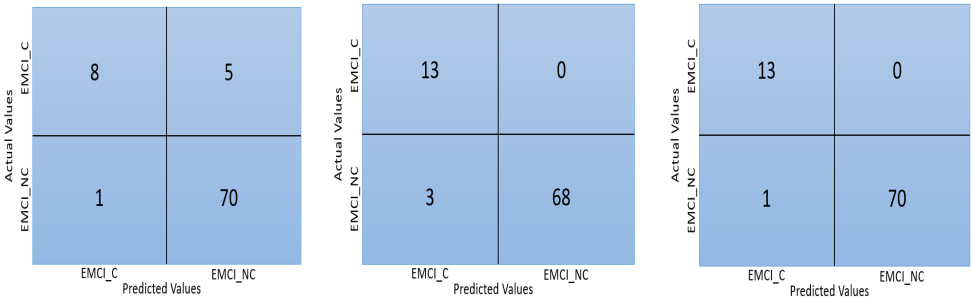
### 4.3.1 Random Forest Feature Characteristics

For our clinical data, the average age of the subjects was 71.4. 55.6% of these were men, and there was a statistically significant age difference between the two groups (P $<$.05). Regarding the genetic and physical biomarkers, APOE4 and hippocampal volume showed substantial differences between the EMCI_C and EMCI_NC classes. Ventricular volume was consistent across both classes. With the neuropsychological scale scores, ADAS13 and FAQ showed significant differences (P $<$.05), whereas ADAS11 and MMSE did not. The correlation matrix in Figure 23 demonstrates the totality of our clinical data feature relationships.

### 4.3.2 Ensemble Model Performance

Our ensemble model workflow can be observed in Figure 20. This demonstrates how the random forest and CNN models work together or independently to output an explainable prediction for our EMCI subjects. Our random forest model is trained with 1000 trees against 2412 exam visits, while our CNN model leverages Tensorflow to retrain the final classification layer of NASNet for DTI analysis. Additionally, we pass a max_depth of 40 with nine max_features as further hyperparameters to the random forest model. We arrive at this tuning by implementing Grid Search to derive the ideal hyperparameters. Next, we optimize our CNN model with the previously discussed augmentation distortions and then train for 8000 steps at a learning rate of .005. During random forest training, 25% of our clinical data are reserved for testing, while the remaining 75% account for the training data. For CNN training, 10% of our images are reserved for testing, 10% for validation, and the remaining 80% for training. These models are then combined to form our ensemble model, after which our Grid Search algorithm is applied to determine the ideal weight distribution. Once the weighting has been applied, our model explainability occurs via the combination of feature ranking and Grad-Cam analysis. This ensures that each outputted prediction has accompanying explainability.

One of the advantages of our ensemble approach is that it allows either modality to be passed absent of the other, and a prediction is still generated. When both modalities (clinical data and DTI scans) are provided, our weighted ensemble model achieves an EMCI-to-AD conversion prediction accuracy of 98.81%. With only clinical data being supplied, our model maintains an accuracy of 92.86%. When only DTI scans are




RF Confusion Matrix CNN Confusion Matrix EM Confusion Matrix

Figure 24: Confusion Matrix: (a) Random Forest (RF) for EHRs, (b) Convolutional Neural Network (CNN) for fMRI, and (c) Ensemble Model (EM)

provided, our model performs at 96.43% accuracy. This flexibility ensures that accurate conversion prediction can be obtained even if one is missing certain features. We also measure the model differences in Figure 25 with the polygon area metric (PAM) proposed by [67]. The individual PAM metrics per model are shown in Table 19. As the model spent significant time being fine-tuned, the experiment was repeated several dozen times per model. After each repeat, performance metrics were assessed to see how to tune the model further. Additionally, cross-validation was performed to ensure different bagging combinations performed well.

The difference between each model's confusion matrix can be seen in Figure 24. While the individual RF model struggled with false negatives, this weakness is removed when transitioning to the ensemble approach. Similarly, while the individual CNN model had three false positives, this was mitigated when predicting as part of the ensemble model.

Table 19: Individual Polygon Area Metrics: Classification Accuracy (CA), Sensitivity (SE), Specificity (SP), Jacard Index (JI), F-Score (F), Area Under Curve (AUC)

| | CA | SE | SP | JI | F | AUC |
|---|---|---|---|---|---|---|
| RF | .929 | .615 | .986 | .571 | .727 | .960 |
| CNN | .964 | 1 | .958 | .813 | .897 | .973 |
| Ensemble | **.988** | **1** | **.986** | **.929** | **.963** | **.992** |

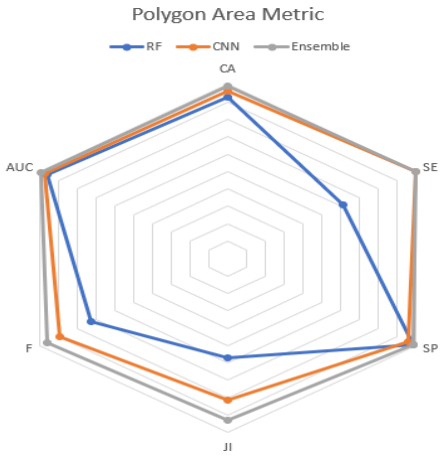
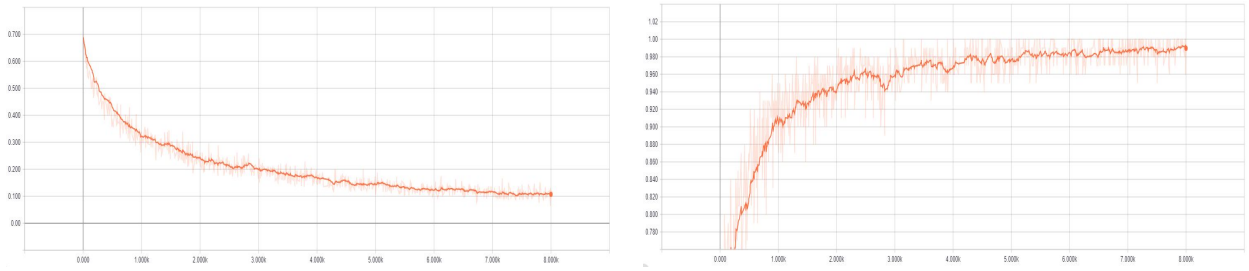



Figure 25: Polygon Area Metric



(a) Loss (b) Accuracy

Figure 26: Tensorboard Training Metrics: Step-wise Loss and Accuracy

### 4.3.3 Ensemble Explainability

A key contribution of this work was to provide accurate conversion prediction and be capable of explaining the rationale behind individual predictions and the overall model. As our ensemble model weighs visual prediction alongside clinical data prediction, it is essential to know the prediction confidence related to each modality. Additionally, understanding the features or pixels that led to the overall decision within each classifier can help instill confidence in a clinical setting. This can be distinguished by providing context around global (model-level) explainability vs. local (individual-level) explainability.

For our clinical global explainability, we perform feature ranking of our nine features based on the following formula:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{4.1}$$

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k\in\ all\ nodes} ni_k} \tag{4.2}$$

$$normfi_i = \frac{fi_i}{\sum_{j\in\ all\ features} fi_j} \tag{4.3}$$

$$RFfi_i = \frac{\sum_{j\in\ all\ trees} norm\ fi_{ij}}{T} \tag{4.4}$$

Initially, in Equation 4.1, we determine the importance of each node per tree (ni). $ni_j$ represents node $j$'s importance, with $C_j$ being a node's impurity value. Additionally,

the weighted samples that reach node j are represented as $w_j$. From this, feature importance (fi) per tree can be calculated as seen in Equation 4.2. This result is then normalized between 0 and 1 (Equation 4.3). This process is then averaged out to the entire forest and divided by the number of trees within the forest per Equation 4.4 [68].

For global explainability, our ensemble model's feature ranking per the above function can be seen in Figure 27. We also demonstrate the permutation importance ranking seen in Figure 28 as well as the Shapley plot (Figure 29). Permutation rankings can reduce high cardinality bias as the features are permuted against a held-out test set. A baseline metric is established for this to occur, which has each feature permuted against itâthe difference between this feature permutation and the baseline metric results in the overall permutation importance. For our primary feature ranking, age, hippocampal volume, and ventricular volume stood out as our model's most important features. With permutation ranking, age and FAQ continued to show strength, with APOE4 gaining in importance compared to its feature ranking. Finally, the Shapley plot shows how strongly each feature contributes to a positive (EMCI_C) versus a negative (EMCI_NC) prediction. The color of each value denotes whether it is high (red) or low (blue) relative to other values for that feature. The combination of these ranking systems aided us in reducing the original feature map to the final model.

For DTI local explainability, we perform Gradient-weighted Class Activation Mapping (Grad-CAM) [69] to generate pixel heat maps. These are then superimposed on the existing image to display the most important regions for the resulting prediction. In this sense, Grad-CAM allows us to understand what our CNN model focuses on by using the

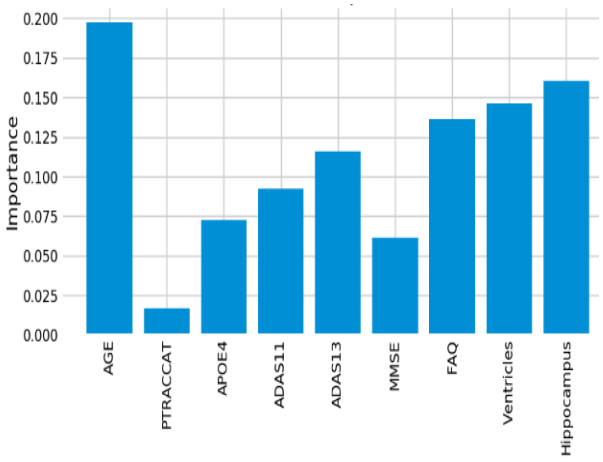



Figure 27: Ensemble Model Feature Importance

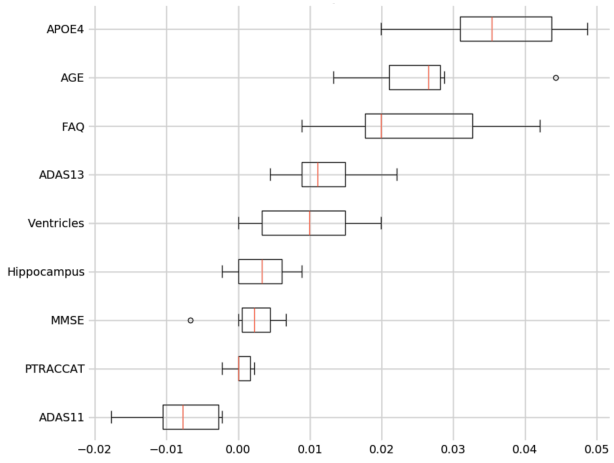



Figure 28: Ensemble Model Permutation Importance

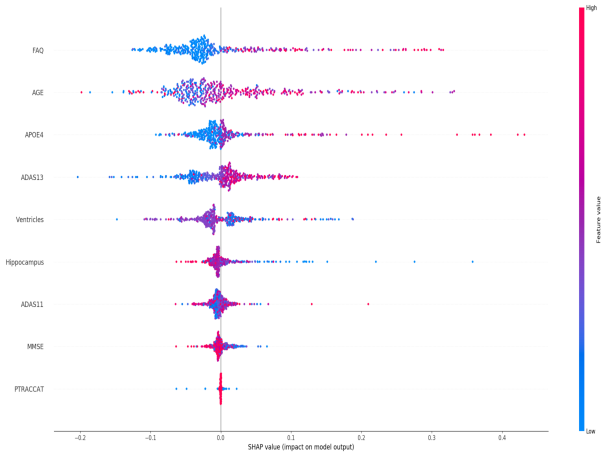



Figure 29: Ensemble Model SHAP Summary

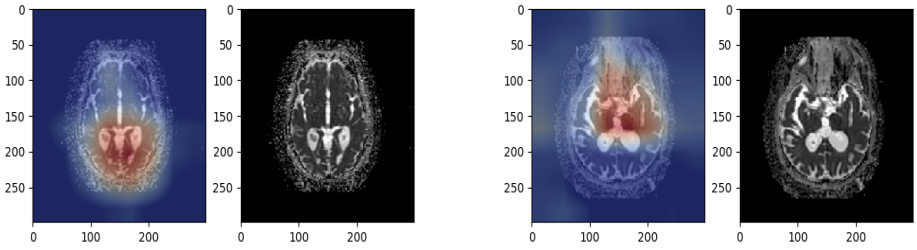



Figure 30: Grad-CAM Explainability: Grad-CAM and DTI Images for EMCI_C Patient (Left) & EMCI_NC Patient (Right).

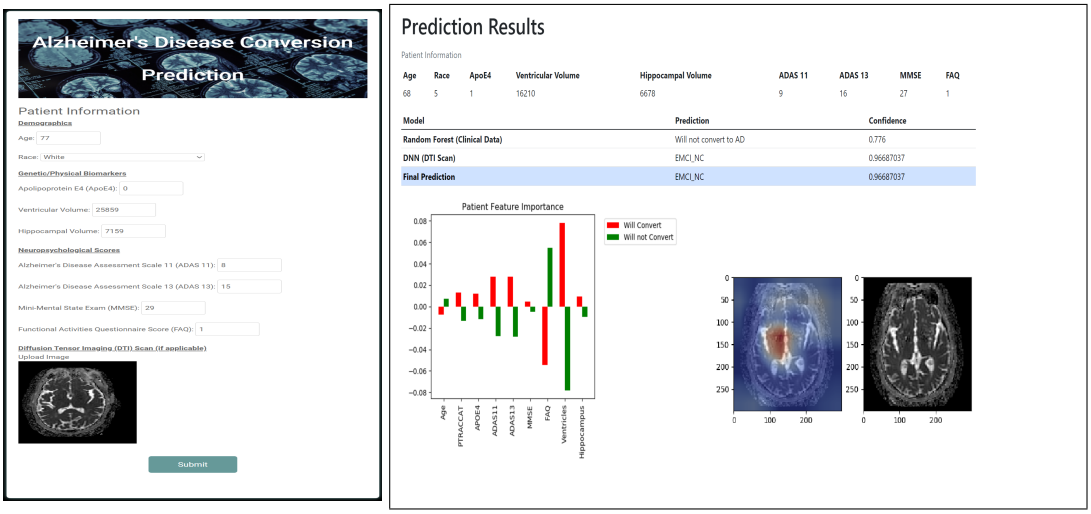



Figure 31: Conversion Prediction Intake (Left) and Results (Right).

gradients that flow into the final convolutional layer. These gradients then use global average pooling to obtain the necessary weights as seen in Equation 4.5. Examples of these heat maps for both an EMCI_C and EMCI_NC patient are shown in Figure 30. For our output, the black and white image represents the initial input before Grad-Cam applies the heatmap. As Grad-Cam assesses which pixels are most relevant, it colors them red at varying intensities to demonstrate that pixel's importance to the prediction. Similarly, darker shades of blue occur when the pixel is deemed not to have a strong contribution to the prediction. Future work will explore aligning these heat maps to segmented brain regions to establish more in-depth global explainability.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{4.5}$$

To assess our ensemble model's local explainability and demonstrate its potential as a clinical decision support tool, we've built a Flask Python application to host our

model and allow for patient intake. Our application allows for patient clinical data to be entered in addition to attaching DTI scans. Partial patient information can also be provided as the application will understand if it has been provided with limited features. For example, only the Random Forest classifier will be engaged if only clinical data is provided. Likewise, the CNN model will serve as the sole predictor if a DTI scan is the only patient data provided. The application also accounts for blanks by substituting the empty field with that feature's mean average. Once the data has been submitted, our ensemble model is engaged, which outputs its prediction and explainability. From our application's output, we can see the importance of the clinical data feature importance alongside the Grad-Cam analysis. We also see the prediction confidence of each independent classifier and the overall ensemble confidence. This informs the user which modality contributed the most to the prediction, highlighting key regions/features of interest. An example of the intake form and a sample prediction can be seen in Figure 31.

Table 20 and Figure 32 detail three unique, individual predictions with local explainability that demonstrate our ensemble model's strength in contrast to a singular model. Within this table, prediction contributions (PC) are also shown. This represents the amount of each clinical feature's contribution to the overall RF prediction. A positive value indicates the contribution towards the ground truth class, whereas a negative value represents the contribution to the incorrect class. As an example, patient 2106 was eventually diagnosed with AD. However, based on their clinical data, our Random Forest component predicted with 52% confidence that they wouldn't convert. In contrast, after assessing the patient's DTI scan, our CNN model predicted that they would convert with

Table 20: Feature Contributions in Cases: Comparing RF, CNN, and Ensemble Predictions

| **Subject 1: 2106** | | | | **Subject 2: 4220** | | | | **Subject 3: 4897** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT | RF | CNN | EN | GT | RF | CNN | EN | GT | RF | CNN | EN |
| C | NC | C | C | NC | NC | C | NC | NC | NC | C | C |
| 100% | 52% | 79% | 65% | 100% | 99% | 58% | 68% | 100% | 89% | 95% | 57% |
| Attribute | Value | | PC | Attribute | Value | | PC | Attribute | Value | | PC |
| Ventricles | 25859 | | .163 | FAQ | 0 | | .033 | FAQ | 7 | | .147 |
| Hippocampus | 7159 | | .07 | ADAS13 | 5 | | .022 | APOE4 | 0 | | -.068 |
| Age | 77 | | .057 | APOE4 | 0 | | .019 | MMSE | 29 | | -.028 |
| Race | White | | .054 | Age | 71 | | .014 | Race | White | | .026 |
| ADAS11 | 8 | | .041 | Hippocampus | 7851 | | .01 | Age | 75 | | -.026 |
| ADAS13 | 15 | | .035 | ADAS11 | 2 | | .008 | Hippocampus | 6676 | | -.023 |
| APOE4 | 0 | | -.015 | MMSE | 30 | | .004 | Ventricles | 34505 | | -.022 |
| MMSE | 29 | | .002 | Ventricles | 23127 | | .002 | ADAS11 | 8 | | -.012 |
| FAQ | 1 | | .001 | Race | White | | -.002 | ADAS13 | 19 | | -.002 |

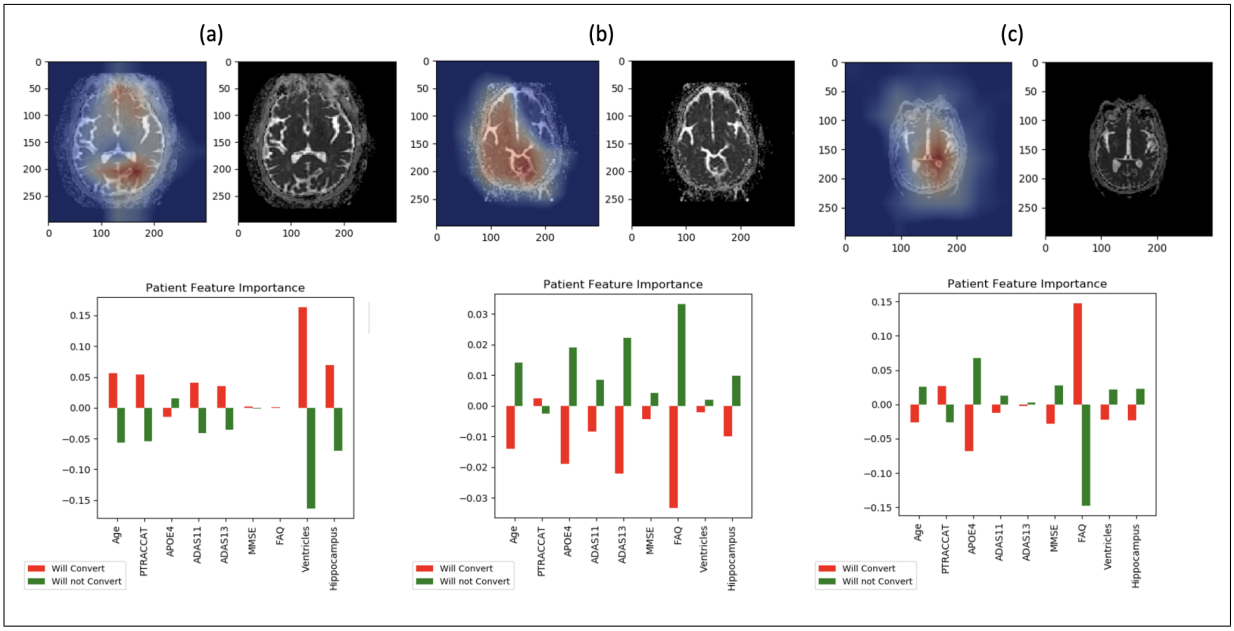



Figure 32: Grad-CAM Explainability (Heat Map Overlay on Left, Intake Image on Right) and Feature Importance Ranking: (a) Patient 2106, (b) Patient 4220, (c) Patient 4897.

79% confidence. With the CNN model being more confident and having more weight in the overall prediction, this resulted in an ensemble confidence of 65% that the patient would convert to AD (EMCI_C). In this case, a singular RF model would have been predicted inaccurately, but with added visual analysis, our ensemble was capable of avoiding the mistake.

Another example from Table 20 can be seen with EMCI_NC patient 4220. However, with this patient, the RF prediction (EMCI_NC, 99% confidence) helped overrule the incorrect CNN prediction (EMCI_C, 58% confidence). The Grad-CAM analysis in Figure 32 shows the difficulty in assessing this patient's DTI scan as the heat map overlaid most of the brain. However, despite the model weighting favoring the visual analysis, the ensemble could still make the correct prediction with a final confidence level of 68%.

Patient 4897 represents an instance where our ensemble model provided an incorrect prediction (EMCI_C, 57% confidence). Despite the clinical data pointing toward an EMCI_NC classification, the DTI prediction confidence won out (95% CNN vs. 89% RF). From Figure 32, we see that the visual model focused heavily on the ventricles, whereas the RF feature ranking placed ventricular volume as 7th in predictive power for this specific individual.

Overall, with our EMCI_C subset, 38% of the ensemble predictions had disagreements between the RF and CNN model but resulted in a correct ensemble prediction. For EMCI_NC, 4.3% of the predictions encountered disagreements between modalities. Given these findings, we see that the ensemble benefits conversion class prediction more significantly than the EMCI_NC class.

Table 21 shows that our proposed model outperforms recently published multimodality models for AD conversion prediction. A defining difference is our usage of DTI over traditional sMRI and our ensemble classification in place of a single classifier. While many competing authors leverage multiple modalities, they typically limit their studies to a single classifier rather than an ensemble approach. Additionally, our model can predict from 5 to 7 years out due to focusing on EMCI rather than the more general MCI data set.

## 4.4 Limitations and Future Work

A limitation of our study is that all patients were derived from the ADNI data set. For our work, this was acceptable. However, a clinical setting implementation of our model could benefit from additional data sets to account for further feature variation.

Table 21: MCI-to-AD Conversion Prediction Model Comparison

| **Approach** | **Modalities** | **Data** (subject size) | **Model** | **MCI-to-AD Pred.** | | **Year** |
|---|---|---|---|---|---|---|
| | | | | ACC | AUC | |
| Proposed Model (ours) | Clinical data/DTI | ADNI (383) | Ensemble (RF/CNN) | **98.81%** | **99.2%** | 5 |
| CNN Model (2019) (ours) | DTI | ADNI (383) | CNN | 96.43% | 98.1% | 5 |
| RF Model (2021) (ours) | Clinical data | ADNI (383) | RF | 92.86% | 96% | 5 |
| Zhang (2021) | sMRI/rs-fMRI | ADNI (108) | SVM | 84.71% | 88.8% | 3 |
| Minhas (2021) | Clinical data/MRI | ADNI (85) | SVM | 81% | 95.7% | 1 |
| Pan et al. (2020) | MRI | ADNI (509) | Ensemble (CNN/EL) | 62% | 59% | 3 |
| Lin et al. (2020) | Clinical data/MRI/FDG-PET | ADNI (617) | ELM | 84.7% | 88.8% | 3 |
| Rana et al. (2020) | Clinical data/MRI | ADNI (559) | CNN | 69.8% | 83% | 5 |
| Grassi et al. (2019) | Clinical data | ADNI (550) | SVM | – | 88% | 3 |
| Huang et al. (2019) | Clinical data/MRI | ADNI (290) | SVM | 80% | 84.6% | 5 |
| Varatharajah et al. (2019) | Clinical data/MRI/FDG-PET | ADNI (135) | SVM | 93% | 93% | 3 |

In addition, these different ADNI studies (ADNI1, ADNIGO, etc.) also leveraged scanners with varying field strengths, which could have affected our results. This will be a consideration for our future work as we aim for more robust modeling.

Another limitation is that gradient-based saliency techniques have shown some unreliability regarding medical imaging [70]. For future work, alternate mappings will be explored and evaluated against our Grad-Cam maps.

Future work will explore performing time-series analysis via an ensemble model in addition to binary classification. This would allow patient progression trajectories to be determined rather than distinguishing between EMCI_C and EMCI_NC. In addition, the generated heatmaps from our CNN model would also be aligned to segmented brain areas to explore potential findings from that relationship.

## 4.5 Conclusion

An ensemble model for EMCI to AD conversion probability within five years is proposed. Either DTI scans, clinical data, or both can be used for this reason. First, our balanced random forest assesses the clinical data input before our CNN evaluates the DTI scan. Each modality generates separate prediction confidence, which is then factored into our ideal model weight (45% RF, 55% CNN). With this approach, our model achieves an accuracy of 98.8% on EMCI to AD conversion prediction within five years. In this study, we observed that DTI scans are better at AD conversion prediction (96.43%) than clinical data alone (92.86%). We also demonstrated ensemble explainability by employing clinical data feature ranking and Grad-CAM analysis for DTI heat map generation. This

allows for greater confidence and understanding of the prediction rationale when framing our model as a decision-support tool.

# CHAPTER 5

# CHARTING AD PROGRESSION: TIME-TO-EVENT PREDICTIVE MODELS AND NOVEL CATEGORIZATION

In this chapter, we propose a Random Forest Regression Model augmented with SMOGN to predict the time to conversion from MCI to AD in months, along with an explainable feature-ranking and progression category layer. Additionally, we include a new Convolutional Neural Network (CNN) that classifies the diffusion tensor imaging (DTI) scans from our prior work [71] into this study's novel progression categories. By categorizing patients based on their predicted progression rate towards AD and providing feature-importance assessment, our model offers clinical decision-making support and can impact treatment plans for patients with MCI. In this chapter, we'll explore our methodology, showcase our results, and discuss the broader implications of our findings.

Although the median survival time after an AD diagnosis is 4.8 years, some individuals live longer. As discussed in prior chapters, 32% of the MCI population will eventually convert to an Alzheimer's disease diagnosis [1]. Therefore, it is crucial not only to predict which individuals will convert to AD but also to the timeline of their conversion.

However, Alzheimer's Disease progression can be non-linear, making it challenging to determine the effectiveness of treatment plans and the aggressiveness of the disease's decline [72]. Additionally, it is unclear whether there are common patterns of clinical feature progression in AD. To address these challenges, we propose a novel framework

that provides MCI-to-AD conversion prediction by month, identifies patient progression slopes, performs novel patient grouping based on progression non-linearity, and provides feature ranking across longitudinal data.

Our framework combines a random forest regression model with SMOGN (Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise) [73] to work in tandem with our prior CNN/RF ensemble classification model [71] detailed in Chapter 4. The prior model determined which MCI patients would eventually convert to Alzheimer's disease based on a combination of clinical data and diffusion tensor images (DTI). The regression model, from this work, uses patient clinical data to predict how many months a patient has until an AD diagnosis, providing further explainability to the conversion patients identified in the ensemble classifier. By performing these estimates across multiple exam visits, we can determine the non-linearity of a patient's progression, allowing the model to form novel groupings of MCI-to-AD patients based on the aggressiveness of their predicted decline.

The primary contribution of this study is a proposed framework that can accomplish the following tasks: 1) providing MCI-to-AD conversion prediction by month, 2) identifying patient progression slopes, 3) performing novel patient grouping based on progression non-linearity, and 4) providing feature ranking across longitudinal data. Our framework represents a significant advancement in both explainability and progression modeling for Alzheimer's disease.

## 5.1 Related Work

For longitudinal data disease progression analysis there have been significant findings in recent years. Ramamoorthy et al. [74] attempted to determine if there were patterns within amyotrophic lateral sclerosis (ALS) disease progression. They were successful in clustering patients with similar progression patterns by leveraging a mixture of Gaussian techniques. They also showed that progression patterns can often be non-linear with stages of rapid decline observed between stable periods.

Lars Lau Raket [72] proposed a model that predicted the disease month from a summary of patient biomarkers that they aggregated as ADAS-cog. Additionally, their model was able to make predictions on disease stage, rate of decline, and cognitive deviation from the mean. This allowed for the interpretation of different factors on how they could potentially impact cognitive function. The impact of proposed treatment plans could also be considered given the knowledge of a patient's predicted rate of decline.

El-Sappagh et al. [75] used an LSTM model to classify a patient as cognitively normal, MCI, or AD, and to predict the time until conversion for those classified as MCI. Their classification stage showed an accuracy of 93.87%, while their regression stage demonstrated a mean absolute error of 0.1375. However, one limitation of their study was the lack of explainability for their model's decisions.

El-Sappagh et al. [76] developed an ensemble learning framework to predict AD progression for up to 2.5 years in the future. Their focus was on accuracy and diversity metrics as they fine-tuned their model until both were balanced and provided optimal output. Notably, their classifier achieved high accuracy without including neuroimaging

data, which could potentially lower the cost compared to MRI-based approaches.

David Loeffler [77] performed a review of the literature exploring how different variables can impact the development of AD, categorizing them as modifiable (vascular risk factors, malnutrition, etc.), non-modifiable (age, family history, etc.), and clinical (neuropsychiatric symptoms, extrapyramidal signs, baseline cognitive level). He found that six factors consistently displayed a positive association with AD: malnutrition, genetic variants, altered gene regulation, baseline cognitive level, neuropsychiatric symptoms, and extrapyramidal signs. However, he did not assess how these factors might impact a patient's rate of decline, leaving that for future explainability work.

Our approach builds upon existing methodologies but introduces key innovations. Previous studies have primarily focused on either clinical data or imaging analysis independently. Our research bridges this gap by integrating both modalities. Our combined methodology not only enhances the accuracy of progression prediction but also provides a multi-faceted view of the disease's impact.

## 5.2 Methods

We have established methods to track the progression from Early Mild Cognitive Impairment (EMCI) to Alzheimer's disease. Our approach includes two categories: Early MCI to Conversion (EMCI_C) and Early MCI Non-Converter (EMCI_NC). In this process, we are integrating multiple dimensions, such as Cognitive Decline (CD), Neuroimaging (NI), and Biochemical Markers (BM). This section provides a comprehensive overview of our dataset, our methodological approach, and rationale for our selection of

Table 22: EMCI Data Set for Machine Learning

| | EMCI_C | EMCI_NC |
|---|---|---|
| Subject# | 49 | 335334 |
| Visit# | 198 | 1608 |
| Record# after Oversampling | 1608 | 1608 |
| Training Data | 1206 | 1206 |
| Testing Data | 402 | 402 |

clinical features.

### 5.2.1 Alzheimer's Disease Neuroimaging Initiative Data

All data used for this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and included patients from their ADNI-1, ADNI-2, and ADNI-GO studies [12].

For this study, all Early Mild Cognitive Impairment (EMCI) subjects who had exam visits for over a year were eligible. This EMCI subset represents patients that are 5-7 years prior to a possible AD diagnosis. The Wechsler Memory Scale Logical Memory II test identifies this subset and we further distinguish them into two groups based on whether they convert to AD (EMCI_C) or fail to convert (EMCI_NC). For our original conversion classifier [71], 1806 exam visits were originally used for training prior to further augmentation. For our month of conversion regression model, we focus only on the EMCI_C group. This group is represented by 198 exam visits (45 patients) prior to SMOGN augmentation.

#### 5.2.1.1 Early MCI with Conversion (EMCI_C)

In the EMCI_C (Early MCI with Conversion) subclass, patients exhibit a definitive progression from Early Mild Cognitive Impairment to Alzheimer's disease. This progression is characterized by a cognitive decline coupled with observable changes in daily functional abilities, social interactions, and possibly mood and behavior. The underlying cause of this progression may be sped-up neurodegenerative processes, such as synaptic dysfunction, neuronal loss, and the buildup of amyloid plaques and neurofibrillary tau tangles. Disruption of neurochemical pathways and inflammation may also play significant roles. In addition to genetic predispositions like the APOE4 allele, other factors such as cardiovascular health, metabolic conditions like diabetes, psychosocial factors, and environmental influences can significantly impact the progression rate.

#### 5.2.1.2 Early MCI Non-Converter (EMCI_NC)

The Early MCI Non-Converter (EMCI_NC) group consists of individuals who present with EMCI symptoms but do not progress to Alzheimer's disease during standard observation periods. Their cognitive impairment might remain stable or even show signs of improvement. The stability in this group could be attributed to factors such as higher cognitive reserve due to lifelong learning and intellectual engagement, robust social networks, physical exercise, a healthy diet, and possibly innate genetic resilience. Investigating the EMCI_NC group can offer valuable insights into preventive strategies and potential therapeutic targets for Alzheimer's disease.

#### 5.2.1.3 Clinical Features Selection

Our random forest regression model was trained on a combination of neuropsychological scale scores (ADAS13, ADAS11, FAQ, MMSE), demographic variables (age, race), genetic biomarkers (APOE4), and physical biomarkers (hippocampal and ventricular volume). These match the features used in our prior EMCI-to-AD conversion prediction model [65].

#### 5.2.1.4 Diffusion Tensor Imaging Data

For the secondary progression category classification, our Convolutional Neural Network (CNN) model was trained on the diffusion tensor imaging (DTI) scans from our prior work [71]. Diffusion Tensor Imaging (DTI) is a type of magnetic resonance imaging (MRI) technology that is specifically focused on measuring the directional movement of water molecules within tissue. This is particularly useful in the brain, where DTI can be used to map neural pathways and understand the integrity of white matter tracts. As a result, microstructural changes in brain tissue can be tracked over time, which is valuable for both clinical and research purposes. These DTI scans were drawn from both EMCI_C and EMCI_NC patients. The EMCI_C scans were specifically from patients that were included in our regression model time series prediction.

#### 5.2.1.5 Mathematical Model for Alzheimer's Progression

The mathematical model for the progression of Alzheimer's Disease in Early Mild Cognitive Impairment with Conversion encompasses several dimensions, namely Cognitive Decline (CD), Neuroimaging (NI), and Biochemical Markers (BM).

- **Cognitive Decline (CD)** is evaluated through comprehensive neuropsychological testing, encompassing a range of cognitive domains.
- **Neuroimaging (NI)** employs advanced techniques such as volumetric MRI for brain atrophy, PET scans for amyloid and tau pathology, and functional MRI for neural network activity.
- **Biochemical Markers (BM)** include a range of biomarkers in the CSF, including amyloid-beta and tau proteins, alongside emerging markers like neuroinflammatory indicators.

The rate of progression $R(t)$ is modeled by:

$$R(t) = f(CD(t), NI(t), BM(t))$$

#### 5.2.1.6 Formal Definitions for Progression Categories in EMCI_C

- **Rapid Progression** Rapid Progression is characterized by a swift and significant decline in cognitive functions, evidenced by rapid brain atrophy on neuroimaging and elevated levels of pathological biomarkers. Clinically, this progression leads to Alzheimer's Disease in less than 12 months. The rate of cognitive decline $R(t)$

is significantly above a high threshold, necessitating immediate and intensive interventions.

  - Accelerated cognitive decline, significant impairments in daily activities.
  - Elevated neurodegenerative biomarkers, rapid neuroimaging changes.
  - $T_{AD} < 12$ months, $R(t) > R_{rapid}$.

- **Moderate Progression** Moderate Progression is defined by a steady decline in cognitive abilities, with moderate changes in neuroimaging and biomarker levels. This progression results in Alzheimer's Disease typically between 12 and 47 months. The rate of decline $R(t)$ lies between the thresholds of Rapid and Gradual Progression. Management strategies include a combination of cognitive therapies, lifestyle adjustments, and pharmacological treatments.

  - Steady cognitive decline, difficulties in complex tasks.
  - Moderate neuroimaging and biomarker changes.
  - $12 \leq T_{AD} < 47$ months, $R_{moderate} \leq R(t) \leq R_{rapid}$.

- **Gradual Progression** Gradual Progression involves a slow decline in cognitive functions, with less severe changes in neuroimaging and biomarkers. The transition to Alzheimer's disease typically occurs over a period exceeding 47 months. The rate of cognitive decline $R(t)$ falls below the moderate threshold. Clinical management focuses on lifestyle interventions, cognitive stimulation, and regular monitoring.

  - Very slow cognitive decline, subtle changes initially.

- Slight changes in neuroimaging and biomarkers.
- $T_{AD} \geq 47$ months, $R(t) < R_{moderate}$.

These formal definitions enhance our understanding of the progression categories within the EMCI_C class of Alzheimer's Disease. They highlight the importance of personalized and timely interventions, based on the specific progression rate and characteristics of each patient. This approach underlines the need for continuous research and development of advanced diagnostic tools for accurate categorization and effective management of Alzheimer's disease progression.

### 5.2.2 Multi-Modal AD Progression Predictive Model

A sophisticated, multi-modal model has been developed, integrating clinical data, neuroimaging (DTI), and advanced machine learning techniques, to accurately predict Alzheimer's disease progression.

This multi-modal predictive model, leveraging the synergy of clinical data, neuroimaging, and advanced machine learning techniques, offers a nuanced and precise prediction of Alzheimer's Disease progression. The use of t-SNE, SHAP, and Grad-CAM for validation underscores the model's robustness and explainability, particularly in categorizing and understanding specialized progression paths within the EMCI_C cohort.

### 5.2.3 Normalization and Categorization Methodology

We employ a detailed normalization and categorization methodology to align and categorize patient data effectively for predictive modeling.

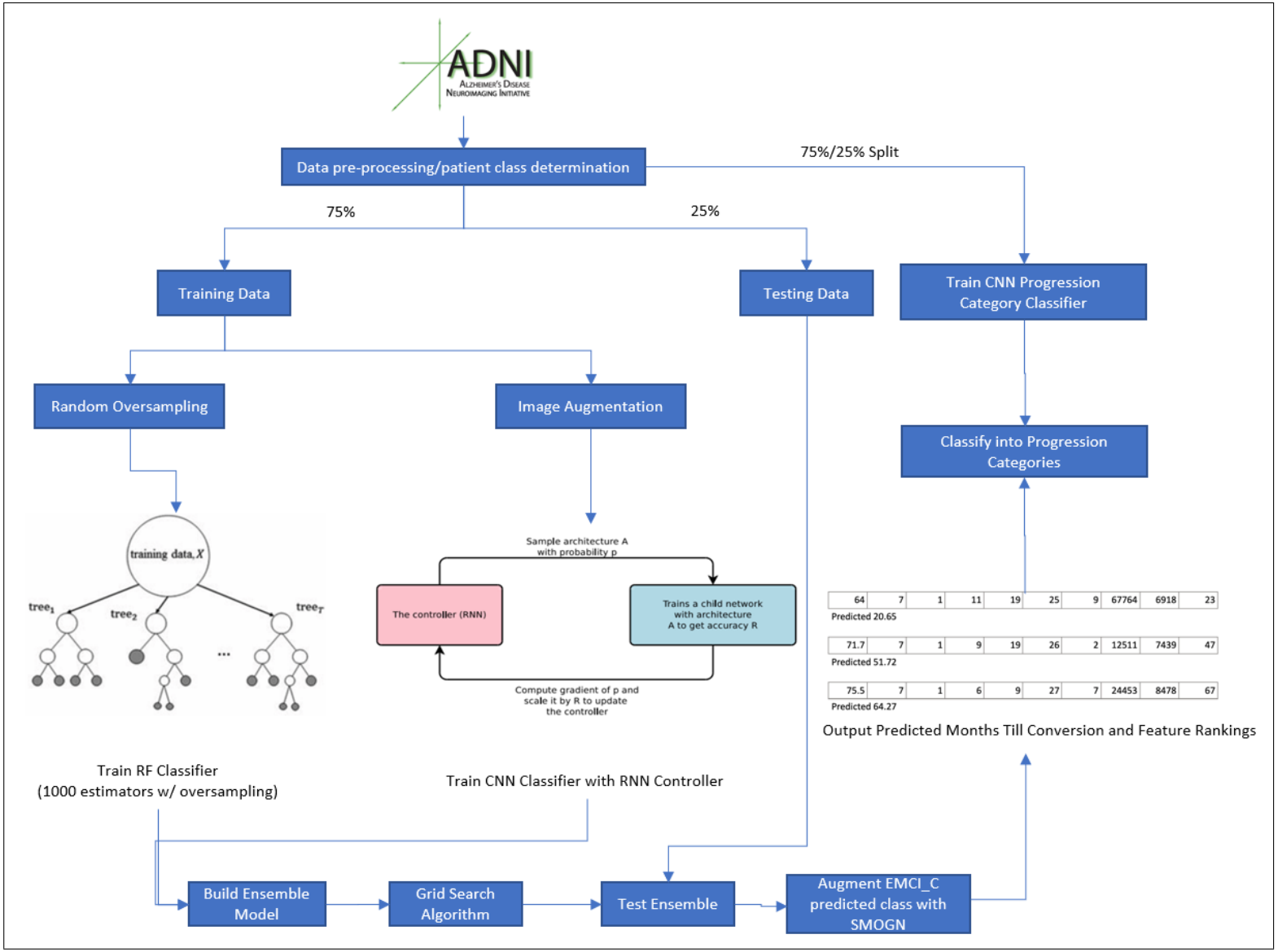



Figure 33: Ensemble Classification and Regression Model Workflow

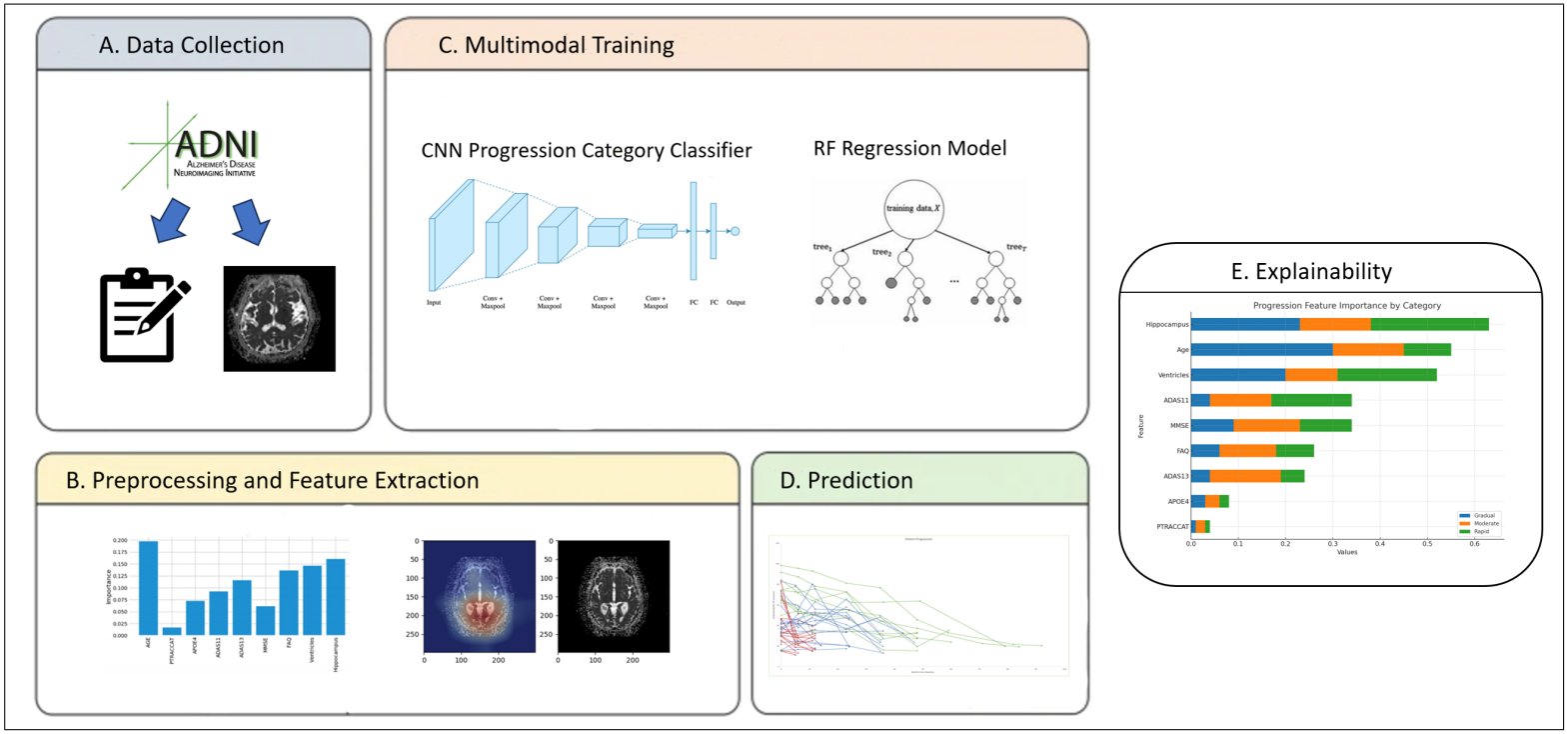



Figure 34: Progression Model Workflow

**Algorithm 1** Multi-Modal Alzheimer's Disease Progression Predictive Model

**Require:** EMCI patient data (Clinical, DTI)
**Ensure:** Comprehensive predictive models for AD progression
**Step 1: Multi-Modal Clinical Data Analysis**
**for** each patient in EMCI group **do**
*Analyze clinical records and DTI data:* Assess clinical progression and neuroimaging markers over time.
**end for**
**Step 2: Progression Slope Analysis**
**for** each patient in EMCI group **do**
*Determine progression trajectory:* Quantify the rate of progression using clinical and DTI data.
**end for**
**Step 3: K-Means Clustering for Specialized Categories**
*Identify distinct progression clusters:* Apply k-means clustering to categorize patients into specialized progression groups based on EMCI_C.
**Step 4: Validation Using Advanced Imaging Techniques**
**for** each cluster **do**
*Employ Grad-CAM for neuroimaging validation:* Correlate clinical progression with specific brain regions affected in each category.
**end for**
**Step 5: CNN-Based Predictive Models for Each Category**
*Develop tailored CNN models:* Construct separate CNN models for each progression category, integrating clinical and DTI data.
**Step 6: Comprehensive Model Explainability**
*Implement t-SNE, SHAP, and Grad-CAM:* Utilize t-SNE for data visualization, SHAP for clinical data interpretation, and Grad-CAM for neuroimaging analysis to validate and explain each specialized progression category.

$$\text{Normalized Time} = \text{Month-wise alignment of patient data} \tag{5.1}$$

$$\text{Category} = \begin{cases} \text{Rapid Progression,} & \text{if } T_{AD} < 12 \text{ months, } R(t) > R_{rapid} \\ \text{Moderate Progression,} & \text{if } 12 \leq T_{AD} < 47 \text{ months, } R_{moderate} \leq R(t) \leq R_{rapid} \\ \text{Gradual Progression,} & \text{if } T_{AD} \geq 47 \text{ months, } R(t) < R_{moderate} \end{cases} \tag{5.2}$$

### 5.2.4 Feature Selection

In our study on Alzheimer's Disease (AD) progression, we employed grid search optimization to fine-tune the parameters of k-means and hierarchical clustering algorithms. These algorithms were applied to a dataset informed by nine crucial biomarker parameters, carefully selected based on findings from our previous studies [65]. These biomarkers were chosen to determine the conversion from early dimensions to AD prediction more accurately and achieve state-of-the-art detection accuracy.

In the Random Forest algorithm, the significance of features is derived from the variation in prediction error across the ensemble of decision trees. This methodology is detailed in [9]. Initially, the importance of each node in a tree is assessed, as formulated in Equation 5.3. In this equation, $ni_j$ denotes the importance of node $j$, $w_j$ represents the weight of the samples reaching node $j$, and $C_j$ indicates the impurity measure at the node. Following this, the importance of features within each individual tree is computed as per Equation 5.4, and subsequently normalized to lie within a range of 0 to 1, according to

Equation 5.5. The final step involves averaging these normalized values across all the trees in the forest and then dividing by the total number of trees, as presented in Equation 5.6 [19].

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{5.3}$$

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k\in\ all\ nodes} ni_k} \tag{5.4}$$

$$normfi_i = \frac{fi_i}{\sum_{j\in\ all\ features} fi_j} \tag{5.5}$$

$$RFfi_i = \frac{\sum_{j\in\ all\ trees} norm\ fi_{ij}}{T} \tag{5.6}$$

The importance scores seen in Table 23 quantitatively reflect the predictive power of various biomarkers within two distinct models. The 'Classification Importance Score' gauges each biomarker's capacity to predict the likelihood of conversion from Mild Cognitive Impairment (MCI) to AD, thus serving as a critical indicator for identifying patients at high risk of developing AD. On the contrary, the 'Progression Importance Score' assesses the biomarker's predictive strength in determining the time until conversion, providing an estimation of how many months it may take for a patient with MCI to convert to AD, based on our regression model.

For example, the Age feature displays a higher Classification Importance Score (.19) than Progression Importance Score (.07), suggesting that while age is a significant

factor in identifying which MCI patients may convert to AD, it provides less predictive value regarding the timeframe of conversion. In contrast, hippocampal volume demonstrates a greater Progression Importance Score (.22) than Classification Importance Score (.17), underscoring its predictive utility in estimating the timeline for conversion rather than just the likelihood.

This nuanced differentiation between the two importance scores is critical for developing targeted clinical strategies. The Classification Importance Score informs clinicians about which patients require close monitoring for potential conversion to AD, while the Progression Importance Score helps in crafting timelines for intervention, potentially allowing for preemptive measures to slow or alter the disease's trajectory.

Figure 35 presents a correlation heatmap that shows the relationships between our selected features. Each square in the heatmap corresponds to the Pearson correlation coefficient between pairs of features, providing a visual summary of how each parameter may be related to another within the context of AD progression.

Notably, a positive correlation between Alzheimer's Disease Assessment Scale (ADAS) scores (ADAS11 and ADAS13) and Ventricular Volume suggests that as the cognitive impairment severity increases, so does the volume of the brain's ventricles. This correlation reinforces the clinical observation that ventricular enlargement is often associated with the progression of AD. The ADAS scores also exhibit a positive correlation with each other, as expected, due to their mutual aim of assessing cognitive dysfunction.

Conversely, we observe a negative correlation between age and hippocampal volume. Similarly, the MMSE score shows a negative correlation with the ADAS scores,

Table 23: Biomarker Parameters with Importance Scores

| **Parameter** | **Description** | **Classification Importance Score** | **Progression Importance Score** |
|---|---|---|---|
| Alzheimer's Disease Assessment Scale (ADAS13) | A cognitive assessment scale designed to measure the severity of cognitive impairment in AD patients. | .12 | .18 |
| Alzheimer's Disease Assessment Scale (ADAS11) | Another cognitive assessment scale used to evaluate cognitive dysfunction in AD. | .09 | .06 |
| Functional Activities Questionnaire (FAQ) | A tool used to assess a person's ability to perform daily activities independently. | .14 | .15 |
| Mini-Mental State Examination (MMSE) | A widely used cognitive screening tool to assess cognitive impairment. | .06 | .09 |
| Age | The age of the individuals in the dataset, can be an important demographic factor in AD progression. | .19 | .07 |
| Race | Demographic information about the individuals, can provide insights into potential disparities in AD risk. | .02 | .02 |
| Apolipoprotein E4 (APOE4) | A genetic marker associated with an increased risk of developing AD. | .07 | .04 |
| Hippocampal Volume | The volume of the hippocampus, a brain region crucial for memory and often affected in AD. | .17 | .22 |
| Ventricular Volume | The volume of the brain's ventricles, which can change with disease progression and may be indicative of AD. | .14 | .17 |

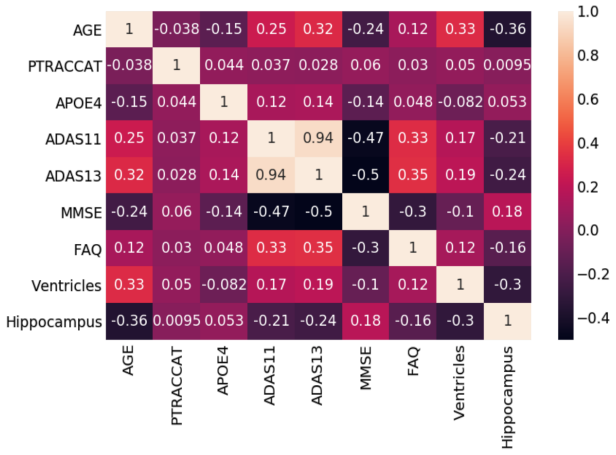



Figure 35: Random Forest Model Correlation Matrix

indicating that as cognitive impairment worsens (higher ADAS scores), the cognitive function measured by the MMSE decreases.

The heatmap also highlights the relatively lower correlation coefficients of APOE4 and PTRACCAT with other features, indicating that these may provide unique information in the AD prediction model that is not captured by other parameters.

These observed correlations inform our feature selection process by identifying features that not only offer independent predictive power, but also those that may act in tandem to enhance the model's performance. The heatmap aids in avoiding redundancy within our predictive model and ensures a robust feature set that is both comprehensive and representative of the complex biological interactions in AD progression.

The integration of sophisticated validation techniques, such as Silhouette Width Analysis and t-distributed Stochastic Neighbor Embedding (t-SNE), has refined our feature selection process to align closely with the distinct needs of patient groups within

our dataset. By leveraging these techniques and interpreting the calculated Importance Scores, we have pinpointed a subset of biomarkers that show the highest predictive alignment with the progression categories, optimizing our feature selection for robust, future-focused predictive modeling. The precise selection of these features, informed by their respective importance scores, significantly enhances our model's ability to support early and personalized treatment plans for patients at various stages.

### 5.2.5 Clustering Techniques with Grid Search Optimization

This section explores the application of grid search optimization in clustering techniques, specifically focusing on K-Means and Hierarchical clustering. By employing grid search, we aim to fine-tune the parameters for these methods, enhancing the efficacy of our clustering models.

#### 5.2.5.1 K-Means with Grid Search Optimization

K-means clustering was optimized using a grid search. The objective function is defined as:

$$S = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{5.7}$$

where $k$ is the number of clusters, $S_i$ is each cluster, $x$ is each data point, and $\mu_i$ is the centroid of each cluster.

#### 5.2.5.2 Hierarchical Clustering with Grid Search Optimization

Hierarchical Clustering with Grid Search Optimization: For HC, we employed grid search over different linkage methods. Hierarchical Clustering (HC) was optimized

**Algorithm 2** K-Means Clustering with Grid Search

**Require:** Dataset $D$, Range of $k$ values
**Ensure:** Optimal clusters $C$
Initialize best score to infinity
**for** each $k$ in the range **do**
    Initialize $k$ centroids randomly
    **repeat**
        Assign each point in $D$ to the nearest centroid
        Update each centroid to the mean of its assigned points
    **until** convergence
    Calculate silhouette score for current $k$
    **if** score is better than best score **then**
        Update best score and $C$
    **end if**
**end forreturn** $C$

using grid search over different linkage methods.

#### 5.2.5.3 Validation with t-SNE and Silhouette Width Analysis

**t-Distributed Stochastic Neighbor Embedding (t-SNE)** The t-SNE algorithm minimizes the KullbackâLeibler divergence between two distributions. t-SNE was used for dimensionality reduction and visualization:

$$\text{Minimize } KL(P \parallel Q) \tag{5.8}$$

**Silhouette Width Analysis** The silhouette width calculation for each data point is given by:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5.9}$$

Silhouette width for each point $i$ is given by:

**Algorithm 3** Hierarchical Clustering with Grid Search

**Require:** Dataset $D$, Set of linkage methods $L$
**Ensure:** Optimal dendrogram $T$
Initialize best score to infinity
**for** each linkage method $l$ in $L$ **do**
    Initialize each point in $D$ as a separate cluster
    **while** number of clusters $> 1$ **do**
        Find the closest pair of clusters
        Merge the closest pair
        Update the distance matrix
    **end while**
    Construct dendrogram $T_l$ for $l$
    Calculate silhouette score for $T_l$
    **if** score is better than best score **then**
        Update best score and $T$
    **end if**
**end forreturn** $T$

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Grid search optimization in both k-means and HC allowed for a more precise determination of the clustering parameters, thereby enhancing the accuracy and relevance of our findings in the context of AD progression.

### 5.2.6 Progression Group Categorization

In our study, we utilized an expanded EMCI Dataset, detailed in Table 24, to enhance predictive modeling for Alzheimer's disease progression. This dataset distinguishes between EMCI patients with and without conversion, providing an extensive overview of subject demographics, visit numbers, and cognitive scores. The dataset's comprehensive nature, underscored by careful oversampling and split into training and testing groups,

Table 24: Expanded EMCI Dataset for Predictive Modeling

| **Parameter** | **EMCI_C** | **EMCI_NC** |
|---|---|---|
| Total Subjects | 49 | 334 |
| Total Visits | 198 | 1608 |
| Records after Oversampling | 1608 | 1608 |
| Training Dataset | 1206 | 1206 |
| Testing Dataset | 402 | 402 |
| Average Age | 72 years | 70 years |
| Gender Distribution | 40% Female, 60% Male | 45% Female, 55% Male |
| Average Cognitive Score | 23 | 25 |
| Rapid Progression Visits | 14 | - |
| Moderate Progression Visits | 75 | - |
| Gradual Progression Visits | 36 | - |

Table 25: DTI Data for EMCI Progression Category Classification

| **Patient Group** | **Description** | #Image |
|---|---|---|
| EMCI_C | Rapid Progression | 18 |
| | Moderate Progression | 13 |
| | Gradual Progression | 15 |

facilitates an in-depth analysis of progression patterns.

Table 25 presents DTI data employed for EMCI progression category classification. It categorizes patients based on progression rates and offers a comparative analysis with a substantial collection of scans. This approach underlines the utilization of advanced neuroimaging techniques in discerning progression trajectories in Alzheimer's patients, demonstrating the potential of DTI in clinical applications. These datasets collectively form the backbone of our study, laying the foundation for nuanced insights into Alzheimer's progression.

The advanced normalization and categorization methodology, combined with detailed data analysis, forms a robust framework for understanding and predicting the progression of Alzheimer's disease.

### 5.2.7 Random Forest Regression Model

For regression problems, random forests create numerous decision trees while training and then output the mean prediction of the individual trees. This averaging not only improves the overall model accuracy but also reduces overfitting [2]. For our workflow, the random forest regressor is making predictions of how many months a patient has until AD diagnosis. This is performed specifically on EMCI_C patients that have been previously identified by our ensemble model. Each decision tree casts a vote which is then aggregated into a final Months_Till_AD prediction. One advantage of random forest is that it allows us to assess both local (individual prediction) and global (model) explainability as we can extrapolate feature importance. When measuring longitudinal data, this provides us a method in which we can determine how feature importance might change between patient exam visits, and how that might impact the overall rate of decline.

#### 5.2.7.1 Regression Model Data Augmentation

Data augmentation is an important aspect of machine learning, especially when working with smaller clinical datasets. Small datasets can lead to inaccurate predictions which can have serious consequences in the clinical setting. To address this issue, we applied the SMOGN (Synthetic Minority Oversampling Technique with Gaussian Noise)

algorithm to balance our clinical dataset. SMOGN is a widely used method for imbalanced regression that has been shown to be effective in a variety of applications [78]. It works by generating new synthetic samples that are obtained by either the SmoteR or Gaussian Noise techniques. The chosen technique depends on the distance between the cases supplied. If they are close, SMOGN will leverage the SmoteR method, otherwise, it will introduce Gaussian Noise. The algorithm also includes a noise filtering step, which helps to remove the noise from the synthetic examples that are not useful for improving the regressor's performance. To apply SMOGN to our clinical dataset, we first split the data into a training set and a testing set. We then used the SMOGN algorithm to oversample important ranges, that might otherwise have been ignored, in the training set to generate synthetic examples. We added these synthetic examples to the training set and retrained our model using the balanced training set.

### 5.2.8 CNN Classification Model

For our progression category classification problem, we leverage a Convolutional Neural Network (CNN) with a NASNET architecture to classify Diffusion Tensor Imaging (DTI) scans into categories based on the expected rate of progression to AD from the MCI stage. NASNET was chosen as the architecture given its high performance with complex image analysis. We distinguish our progression categories as gradual, moderate, or rapid as they reflect the anticipated speed of progression towards an AD diagnosis. We define these three distinct categories based on the timespan from the initial condition to their AD diagnosis as well as the severity of the initial stage:

We have developed a Convolutional Neural Network (CNN) classification system with four distinct categories to enhance the diagnosis and monitoring of Alzheimer's Disease progression. The categories are as follows:

1. **EMCI_C (Early MCI with Conversion)**: This category is further divided into three subcategories based on the rate of progression to Alzheimer's Disease (AD):

   (a) **Rapid Progression**: Patients showing significant cognitive decline and conversion to AD within a period of less than twelve months.

   (b) **Moderate Progression**: Patients whose transition from Mild Cognitive Impairment (MCI) to AD occurs within 12 to 47 months, accompanied by moderate deterioration in cognitive functions.

   (c) **Gradual Progression**: Patients experiencing a slower decline, with the transition to AD occurring over a period exceeding 47 months.

2. **EMCI_NC (Early MCI Non-Converter)**: This category includes patients who present with EMCI symptoms but do not progress to Alzheimer's disease within the standard observation periods.

By aligning patients into these progression categories, we enhance our understanding of the disease trajectory. The convolutional layers specifically seek to detect key image features or patterns that correlate to each of these progressions. This is supplementary to our clinical data classification as it adds a layer of visual biomarker analysis.

**CNN Model Data Augmentation.** Augmentation was also key to our CNN model given

Table 26: CNN Augmentation Methods Compared

| Variant | **Accuracy** | **Scale Dist.** | **Bright. Dist.** | **Crop Dist.** | **Flipped Img.** |
|---|---|---|---|---|---|
| Variant 1 | **86.5%** | 0 | 30 | 0 | True |
| Variant 2 | 85.1% | 0 | 0 | 0 | True |
| Variant 3 | 83.8% | 0 | 30 | 0 | False |
| Variant 4 | 82.6% | 0 | 0 | 0 | False |
| Variant 5 | 78.2% | 30 | 0 | 0 | False |
| Variant 6 | 76.5% | 0 | 0 | 30 | False |

the small dataset. Limited data can be prone to overfitting and reduce a model's generalizability. To mitigate this, we performed image-specific augmentation techniques that enhanced the diversity and robustness of our model. The techniques we focused on were horizontal flipping, brightness adjustment, and minor scaling variations. These variations were chosen based on our comparative analysis seen in Table 26. The augmented dataset was formed by applying these transformations to the original DTI images which allowed us to expand our dataset without compromising the integrity of the medical details. This augmented dataset not only helps us to prevent overfitting, but also increases how well our model can generalize to new data as evidenced by the enhanced post-augmentation performance.

## 5.3 Results

### 5.3.1 Regression Model Performance

The workflow of our progression model is depicted in Figure 34. This shows how our random forest regressor alongside our CNN classifier takes EMCI_C predictions from

our previous ensemble model as input and produces the predicted months until AD conversion as output. Our regression model is trained with 10000 estimators against 198 exam visits, split with 75% reserved for training and 25% for testing, which are then augmented by SMOGN. This split was additionally grouped by patient to prevent the same patient existing in both the training and validation datasets. For individual exam prediction, feature ranking explainability is performed after the prediction output. For multi-exam prediction, feature ranking is performed, and a progression category is assigned based on the non-linearity of the patient's progression.

The overall workflow of our framework is shown in Figure 33. This demonstrates how a patient progresses through conversion prediction into progression analysis.

In Figure 37, both features are plotted against the number of months that the patient has been in the study (Months Since Baseline), as well as the model's predicted months until AD conversion. This allows us to see the trend of each feature over time and how different progression categories, such as Rapid (red), Moderate (blue), and Gradual (green), are affected. These progression categories are determined by a combination of the patient's conversion timeline and K-means clustering.

The K-means algorithm was chosen due to its effectiveness in handling vast datasets and generating distinct, non-overlapping clusters. This quality proved crucial for our dataset, facilitating the identification of clear AD progression patterns. A comparison with hierarchical clustering revealed K-means to be superior in terms of consistency and interpretability, supported by silhouette scores as shown in Table 27.

An in-depth subset iteration test was conducted to ascertain the most significant

Table 27: Comparative Analysis of K-Means and Hierarchical Clustering

| **Aspect** | **K-means Clustering** | **Hierarchical Clustering** |
|---|---|---|
| Silhouette Score | 0.651 | 0.636 |
| Method | Centroid-based | Connectivity-based |
| Formation | Centroid proximity clusters | Distance linkage hierarchy |
| Complexity | Lower | Higher |
| Flexibility | Fixed cluster number | Variable clusters |
| Interpretability | Straightforward | Detailed dendrogram |
| Efficiency | Faster for large data | Slower, more intensive |
| Use Cases | Uniform cluster sizes | Diverse natural groupings |

features for clustering. This test measured the silhouette score for various feature combinations, resulting in the identification of Race, APOE4, MMSE, and Ventricular Volume as the feature subset yielding the highest silhouette score of 0.655. This high score reflects well-separated clusters, suggesting that these specific features capture the essence of AD progression most effectively.

Figure 36 displays a t-SNE visualization of the clustering performed with the best feature subset. The visualization reveals a clear separation between the clusters, with each cluster suggesting different rates of progression of AD.

As depicted, the data points coalesce into clusters that are differentiated with a gradient color scheme, indicating the cluster to which each data point belongs. The gradation also subtly reflects the degree of membership, providing insight into the certainty of the clustering. We observe that the points form a series of discrete, yet contiguous, clusters stretching across the t-SNE feature space. This pattern suggests that while each group is distinct, there is a progression or gradation from one group to the next, which may mirror the progressive nature of Alzheimer's Disease.

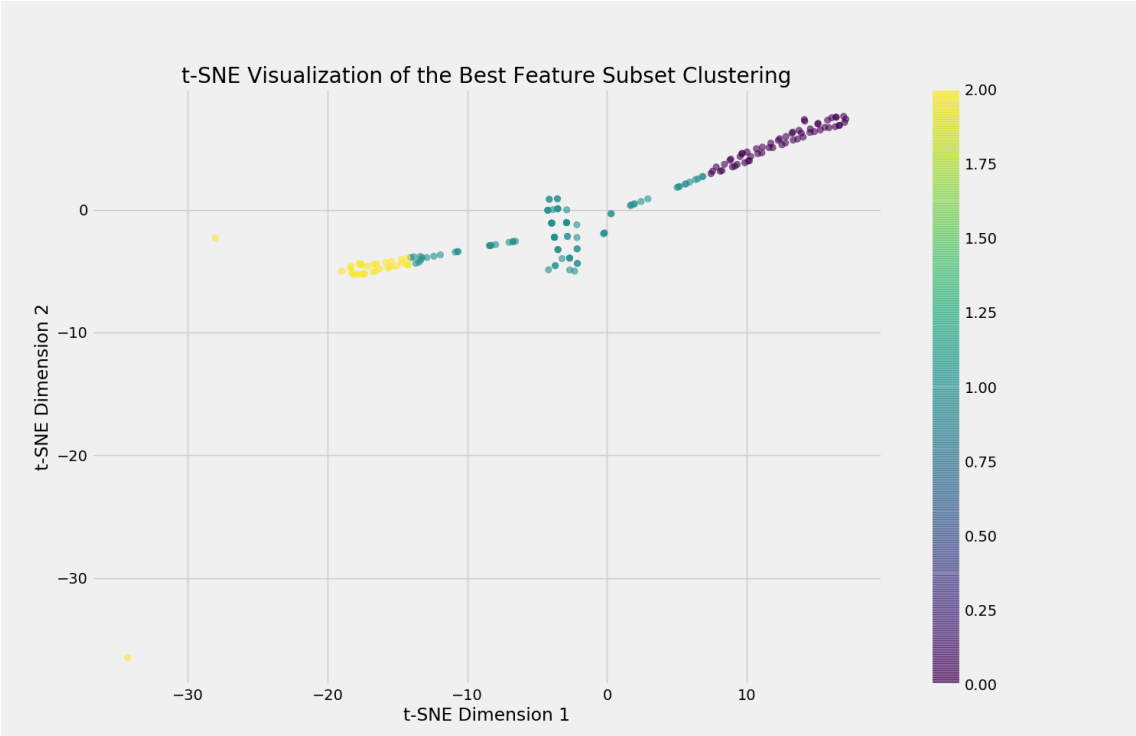



Figure 36: T-SNE Clusters of Patient Data by Key Features and AD Progression Category

The separation of clusters along the t-SNE dimensions affirms the discriminative power of the selected features. The tightness of the clusters, particularly noted in the cluster associated with the rapid progression category (as denoted by the darker color in the upper right quadrant), suggests a high degree of homogeneity within this group. In contrast, the gradual progression category, represented by a lighter color, is characterized by a broader spread across the t-SNE space, implying a more heterogeneous mix of features within this group.

The middle region where clusters merge represents patients with a moderate progression rate, displaying an intermediate feature profile between the rapid and gradual

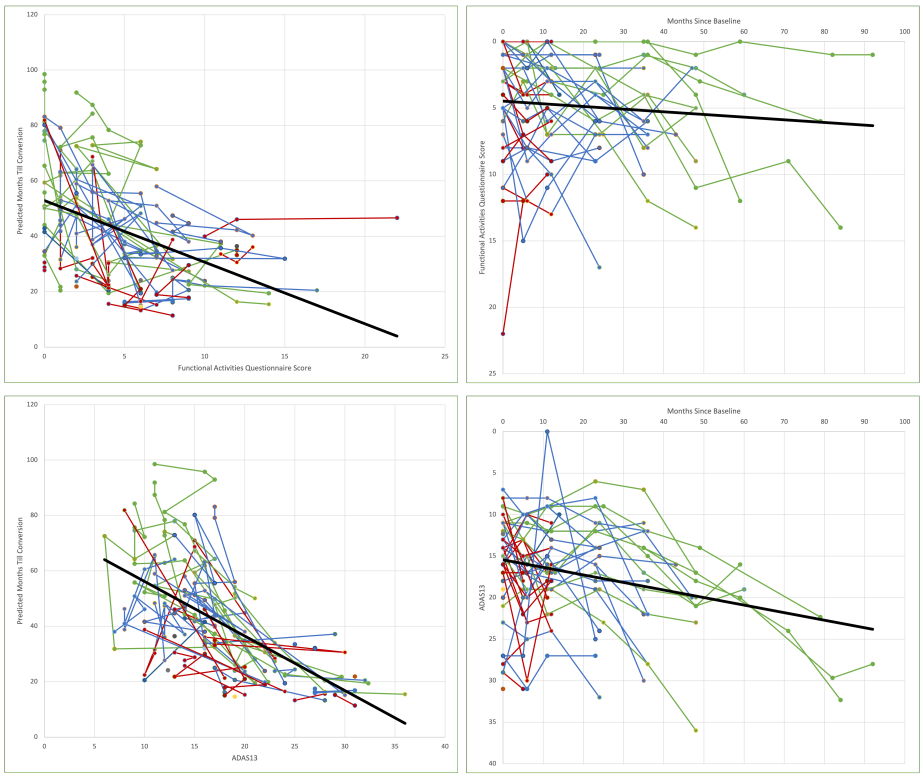



Figure 37: FAQ Progression (top) & ADAS13 Progression (bottom)

groups. This visual segregation aligns with the clinical understanding of Alzheimer's Disease, where patients exhibit a spectrum of symptoms and progression rates, influenced by a complex interplay of genetic, cognitive, and neuroanatomical factors.

In summary, the t-SNE visualization corroborates the efficacy of our feature selection process, emphasizing the value of these features in capturing varying rate of progressions within AD. The discernible cluster patterns not only validate the analytical approach but also underscore the potential for these biomarkers to contribute to a refined stratification of AD patients.

For the month conversion prediction, we conducted several experiments to validate our model, as outlined in Table 28. Our best-performing configuration was a random

Table 28: Regression Model RMSE and Accuracy Compared

| Model | Augmentation | RMSE | Accuracy |
|---|---|---|---|
| Random Forest | SMOGN | 12.61 | **78.34%** |
| Random Forest | Quantile Transformer | 21.64 | 42.75% |
| XGBoost | None | 20.69 | 34.96% |
| Random Forest | None | 19.28 | 37.74% |
| SVC | None | N/A | 11.53% |

Table 29: Sample Individual Months Till Conversion Predictions: PC: Progression Category, VT: Ventricles, HP: Hippocampus, GT: Ground Truth, PM: Predicted Months

| PC | Rapid | Moderate | | Gradual |
|---|---|---|---|---|
| **Age** | 68 | 64 | 72 | 76 |
| **APOE4** | 2 | 1 | 1 | 1 |
| **ADAS11** | 16 | 11 | 8 | 6 |
| **ADAS13** | 22 | 19 | 19 | 9 |
| **MMSE** | 23 | 25 | 26 | 27 |
| **FAQ** | 10 | 9 | 2 | 7 |
| **VT** | 57676 | 67764 | 12511 | 24453 |
| **HP** | 7528 | 6918 | 7439 | 8478 |
| **GT** | 11 | 23 | 47 | 67 |
| **PM** | 22 | 20 | 51 | 64 |

forest regressor with SMOGN, which achieved an accuracy of 78.34% and an RMSE of 12.61. To ensure that our model was not overfitting, we also performed cross-validation.

Table 29 displays the predicted months until conversion for individual patients based on various input features. Although useful on their own, these predictions allow our model to further classify each patient into an associated progression category when evaluated across multiple exam visits.

We also evaluated the predictive accuracy of our model across our different AD progression categories, as detailed in Table 30. This table delineates the mean values of key features and the corresponding RMSE for each category: Gradual, Moderate, and Rapid.

The Gradual category, with an average age of 72.40 years and mean feature values such as an APOE4 frequency of 0.78, ADAS11 score of 8.33, and Hippocampal volume of 6993.93, shows a lower RMSE of 6.58. This suggests that the model is particularly adept at predicting disease progression in this group, where changes might be more subtle and gradual.

In the Moderate category, the patients are slightly older on average (73.57 years), and display different biomarker profiles, such as a higher ADAS13 score (18.13) and a lower average Hippocampal volume (6582.79). The RMSE here is 14.23, indicating a moderate level of predictive challenge, which could be attributed to a more varied progression pattern within this group.

The Rapid category represents the most aggressive disease progression with the youngest average age (70.69 years), the highest APOE4 frequency (1.00), and the largest Ventricular volume (43272.17). Despite these challenging characteristics, the model achieves an RMSE of 10.46, reflecting reasonable prediction accuracy under conditions of rapid and unpredictable disease progression.

The heatmap in Figure 38 illustrates the confusion matrix of this model. Rows represent actual categories, while columns represent predictions. The color intensity indicates prediction counts, with darker shades signifying higher numbers. The diagonal cells show correct predictions, with 17 accurate 'Rapid' and 48 'Gradual' predictions. Notably, the model misclassified 5 'Rapid' instances as 'Moderate', highlighting a potential area for improvement. This visualization offers a succinct overview of the model's performance, emphasizing its strengths in predicting 'Moderate' and 'Gradual' stages and the

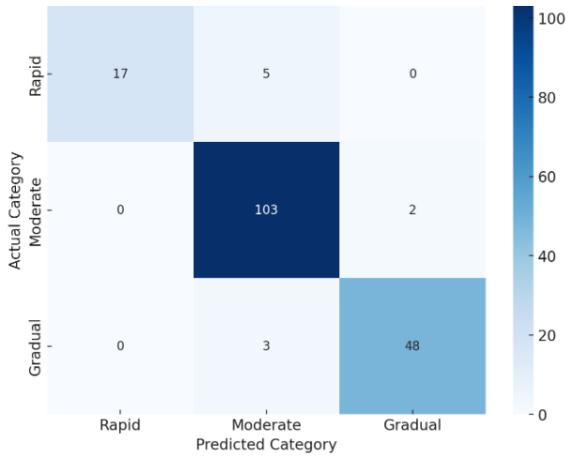



Figure 38: Random Forest Regressor Confusion Matrix

need to enhance 'Rapid' stage identification.

These findings underscore the diverse predictive performance of the model across the progression spectrum of AD. The variance in RMSE across categories highlights the complex interplay of age, genetic factors (like APOE4), and neuroimaging biomarkers (such as Hippocampus and Ventricles volume) in forecasting disease progression. This analysis not only demonstrates the model's capabilities and limitations but also illuminates the potential for tailored prognostic strategies based on specific patient profiles.

Table 31 demonstrates a comparison between existing time-to-event prediction models. While the events predicted are distinct, similar modalities are used. A benefit of our proposed model is that our scope of time extends to 5-7 years, whereas many of the comparative models work within a one-year span. Within Alzheimer's disease conversion prediction timing specifically, we are not aware of any competing models that assess this from an MCI stage.

Table 30: Mean Feature Values with Standard Deviation and RMSE by Progression Category: PC: Progression Category, VT: Ventricles, HP: Hippocampus

| | **Progression Category** | | | |
|---|---|---|---|---|
| **PC** | **Gradual** | **Moderate** | **Rapid** | **EMCI_NC** |
| **Age** | 72.40(6.54) | 73.57(5.73) | 70.69(5.63) | 70.53(7.29) |
| **APOE4** | 0.78(0.81) | 0.80(0.71) | 1.00(0.62) | 0.43(0.59) |
| **ADAS11** | 8.33(3.13) | 11.20(3.93) | 13.80(6.20) | 7.97(4.19) |
| **ADAS13** | 13.49(4.50) | 18.13(5.88) | 20.03(7.32) | 12.31(6.15) |
| **MMSE** | 28.12(1.72) | 27.03(1.94) | 26.50(2.09) | 28.28(1.75) |
| **FAQ** | 2.20(2.09) | 5.53(4.14) | 7.73(3.88) | 1.79(3.11) |
| **VT** | 32512.13(17457.28) | 38163.84(20530.91) | 43272.17(20953.00) | 34779.82(19564.64) |
| **HP** | 6993.93(1170.74) | 6582.79(1058.94) | 6438.56(942.24) | 7300.77(1009.95) |
| **RMSE** | 6.58 | 14.23 | 10.46 | - |

| **Approach** | **Condition** | **Model** | **Accuracy** | **RMSE** |
|---|---|---|---|---|
| Proposed Model (ours) | Alzheimer's | RF | **78.34%** | **12.61** |
| Bossa et al. (2023) | Alzheimer's | Bayesian | - | 16.1 |
| Wijeratne et al. (2023) | Alzheimer's | TEBM | - | 21.73 |
| Rajkomar et al. (2018) | Hospital Readmission | RNN | 75% | - |
| Weng et al. (2017) | Cardiovascular Events | RF | 74.5% | - |
| Avati et al. (2018) | Palliative Care | DNN | 69% | - |
| Wang et al. (2022) | Cancer | 3D-Resnet | 69% | - |

Table 31: Time to Event (Monthly Prediction) Model Comparison

### 5.3.2 CNN Model Performance

Our CNN progression category classification model was trained over 18 epochs, with an early stopping parameter that prevents overfitting. This also serves to optimize our overall training time. Our Variant 1 model was trained with a learning rate of .005 and a batch size of 64. Many variations of this were tested but these proved to be the optimal parameters. Our use of NASNET as the backing architecture played a significant role

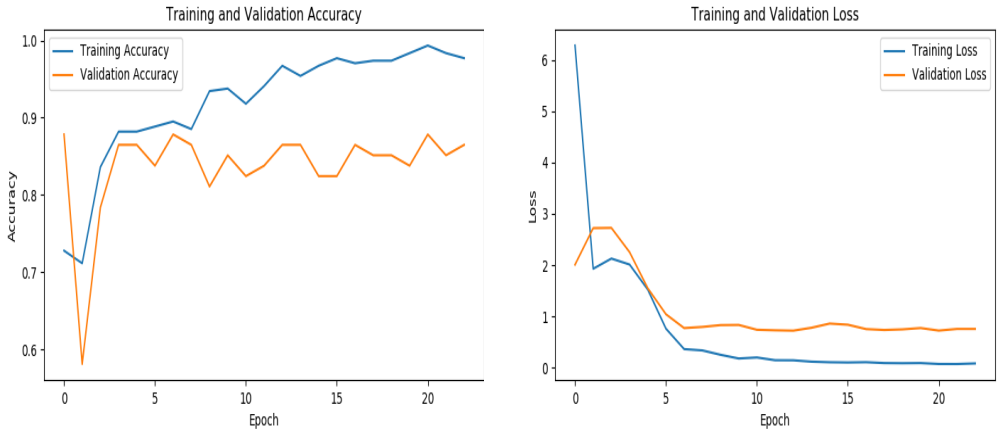



Figure 39: CNN Accuracy & Loss Graphs

in achieving high performance, given its capabilities with image classification. DTI images from 80 patients were used for training across 4 classes (Gradual, Moderate, Rapid, EMCI_NC). Prior to training, these images were augmented with brightness distortion as well as horizontal flipping with 20% of the dataset reserved for validation testing. In addition to progression category classification, we perform GradCAM analysis to further explain our patient's trajectory from a visual perspective.

Our model demonstrated significant efficacy in classifying DTI scans into the defined progression categories by achieving an accuracy of 86.49%. Our accuracy and loss curves can be seen in Figure 39. This bolsters the output provided by our regression model by providing a visual justification. When converting our clinical data regression model output to our progression categories, we achieve an accuracy of 87.64%.

A comparative analysis of these two models is presented in Table 32 alongside related progression categorization models. The slightly higher accuracy of the regression model indicates its potential in effectively categorizing progression stages, despite being

Table 32: Comparative Evaluation for AD Progression Category Classification

| **Approach** | **Model** | **Modality** | **Accuracy (%)** |
|---|---|---|---|
| Proposed RF Model (ours) | RF Regressor | Clinical Data | 87.64 |
| Proposed CNN Model (ours) | CNN | DTI | 86.49 |
| Hashemifar et al. (2023) [79] | 3D CNN | MRI | 82.34 |
| Van Der Haar et al. (2023) [80] | SVM | Clinical Data | 89.35 |
| Pan et al. (2022) [81] | 3D CNN | MRI | 79.01 |
| El-Sappagh et al. (2022) [82] | LSTM | Clinical Data | 82.82 |

primarily designed for Months Till Conversion prediction. This underscores the model's adaptability and reinforces the utility of combining clinical data insights with advanced machine learning techniques.

We have expanded the approaches from three significant studies in AD progression analysis to create a nuanced and comprehensive model. Building upon the DeepAD model [79], which integrates clinical, genomic, demographic, and MRI data using advanced techniques for broad analysis, we have introduced our RF Regression model with SMOGN, combined with our CNN model for DTI scan analysis. This enhancement allows for a more detailed categorization of AD progression rates, offering greater diagnostic precision and valuable insights for treatment planning.

We also contrast our approach with the Manifold Learning study [80], which employs manifold learning techniques such as t-SNE, UMAP, and sparse denoising autoencoders on ADNI data. While this study provides insights into potential subcategories within AD progression, our research emphasizes specific, measurable progression rates, integrating advanced imaging techniques with machine learning models for a more actionable clinical understanding.

Additionally, we evolve the methodologies presented in the "Deep Learning for

Brain MRI" study [81], which uses an Ensemble of 3D CNNs for MRI analysis, developing a P-score biomarker for assessing neurodegeneration. Our model extends these insights by creating a comprehensive multimodal framework that captures both neuroimaging nuances and clinical progression patterns.

The juxtaposition of our clinical data approach against our CNN model in categorizing progression categories underlines the diverse methodologies available in medical image analysis. It also highlights the importance of tailoring the model architecture and parameters to the specificities of the dataset and the task at hand. Our integrated approach, combining visual justification through CNN with clinical data insights, offers a more comprehensive understanding of AD progression.

### 5.3.3 Ensemble Progression Category Classification Performance

To enhance the predictive accuracy of the progression category classification, we employed an ensemble approach combining our CNN classifier and our RF regressor. The primary objective was to leverage the distinct strengths of these two models in a complementary fashion. To this end, a grid search was conducted to determine the optimal weighting of each model within the ensemble. The results of this search are summarized in Table 33.

The grid search systematically varied the weights assigned to the CNN and RF models, ranging from 0 to 1 in increments of 0.1. Each combination was evaluated for its classification accuracy, with the findings indicating a nuanced interplay between the weights and the overall ensemble performance.

A notable observation from the results is the peak accuracy of 89.4%, achieved with a weight of 0.4 for the CNN and 0.6 for the RF model. This specific combination outperforms the individual standalone models, underscoring the efficacy of the ensemble approach. The enhancement in performance can be attributed to the synergy between the CNN's image-based analytical prowess and the RF's robustness in handling structured clinical data. The weighted ensemble evidently capitalizes on the strengths of each model while mitigating their individual weaknesses.

Interestingly, as the weight shifts progressively towards the CNN (beyond the 0.4 mark), a decline in accuracy is observed. This trend highlights the RF model's slightly superior standalone performance and its critical role in the ensemble's overall effectiveness. Conversely, an ensemble dominated by the RF model (weights lower than 0.4 for the CNN) fails to harness the CNN's valuable insights fully, leading to suboptimal performance.

These findings underscore the importance of balance in model weighting within an ensemble framework. The optimal weight distribution not only maximizes predictive accuracy but also ensures a harmonious integration of diverse analytical methodologies. The ensemble model, with its finely tuned weighting, presents a robust solution to the progression category classification challenge, demonstrating the potential of hybrid machine learning approaches in complex analytical tasks.

Table 33: Grid Search Results for Ensemble Weighting

| Weight_CNN | Weight_RF | Accuracy |
|---|---|---|
| 0.0 | 1.0 | 87.6% |
| 0.1 | 0.9 | 87.9% |
| 0.2 | 0.8 | 88.1% |
| 0.3 | 0.7 | 88.4% |
| 0.4 | 0.6 | **89.4%** |
| 0.5 | 0.5 | 88.5% |
| 0.6 | 0.4 | 87.5% |
| 0.7 | 0.3 | 87.4% |
| 0.8 | 0.2 | 87.2% |
| 0.9 | 0.1 | 87.1% |
| 1.0 | 0.0 | 86.5% |

### 5.3.4 Explainability

Our study's main contribution is the ability to provide explanations for our model's predictions, including the identification of features that may contribute to a patient's rapid decline. We accomplish this through the output of feature rankings at both the global (model) and individual levels, as well as the use of SHAP values to explain each feature's contribution to the overall prediction. These explanations can be seen in Figure 40.

#### 5.3.4.1 Global and Individual Feature Analysis

At a global level, the hippocampal volume and the ADAS13 test results contribute the most towards our months till conversion prediction. Ventricular volume and FAQ also show high importance whereas APOE4 and age are less meaningful in comparison to the previous classification work. For individual predictions this ranking varies, however FAQ, ventricular volume, and hippocampal volume are consistently represented towards the top.

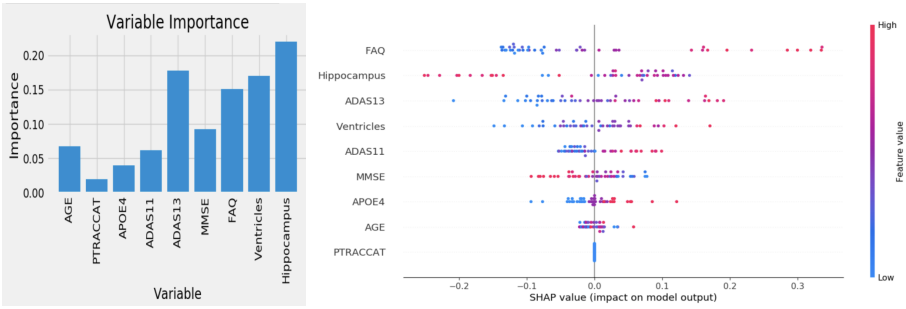



Figure 40: Feature Importance and SHAP Summary

#### 5.3.4.2 Progression Category Analysis

Breaking out the feature importance analysis by our progression categories reveals additional insights. Figure 41 demonstrates this feature importance breakout. In the Gradual category, Age and Hippocampal volume emerge as the most influential factors, suggesting that older age and hippocampal atrophy are strong predictors for patients who will progress to AD more slowly (over 47 months). The significance of hippocampal volume is consistent with clinical understanding, as hippocampal atrophy is a well-known marker of AD progression.

For the Moderate progression category, where patients are predicted to convert to AD between 12 and 47 months, ADAS13 shows the highest feature importance, followed closely by hippocampal volume and age. This underscores cognitive function, as measured by ADAS13, along with age and hippocampal volume as key indicators of a moderate progression timeline.

In the Rapid category, hippocampal volume is the feature with the highest importance, indicating a strong relationship between rapid disease progression (less than 12 months until conversion) and hippocampal atrophy. This is followed by ventricular

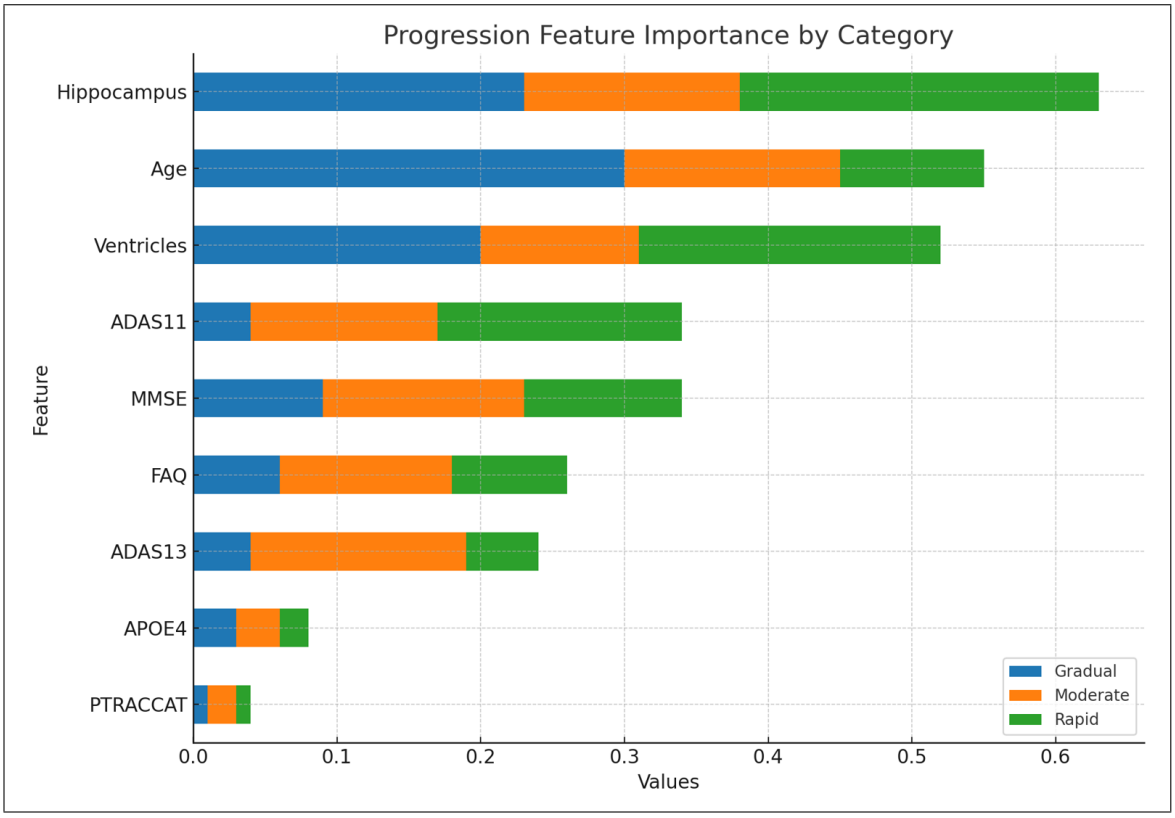



Figure 41: Progression Feature Importance by Category

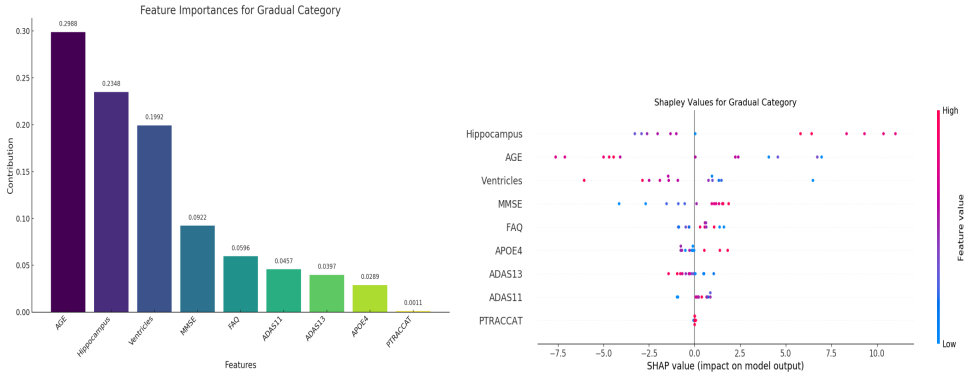



Figure 42: Feature Importances and SHAP Values for the Gradual Progression Cadtegory, Highlighting the Dominance of Age and Hippocampal Volume.

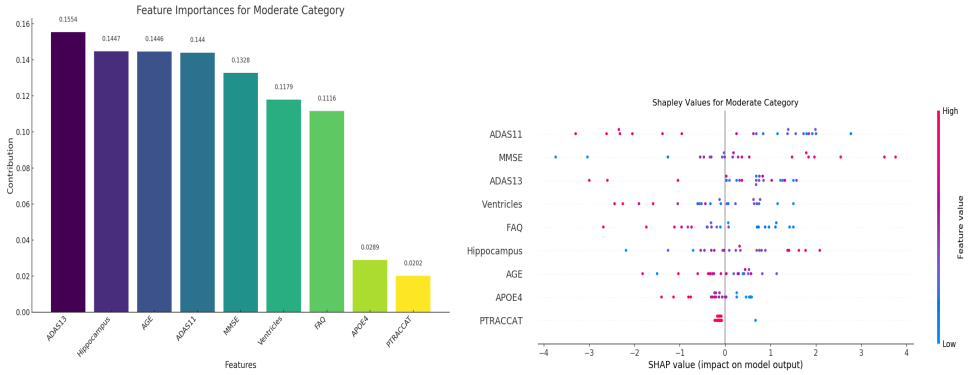



Figure 43: Feature Importances and SHAP Values for the Moderate Progression Category, with ADAS13, Age, and Hippocampal Volume as the Most Informative Predictors.

volume, which may reflect the increased brain atrophy and ventricular enlargement associated with more aggressive AD.

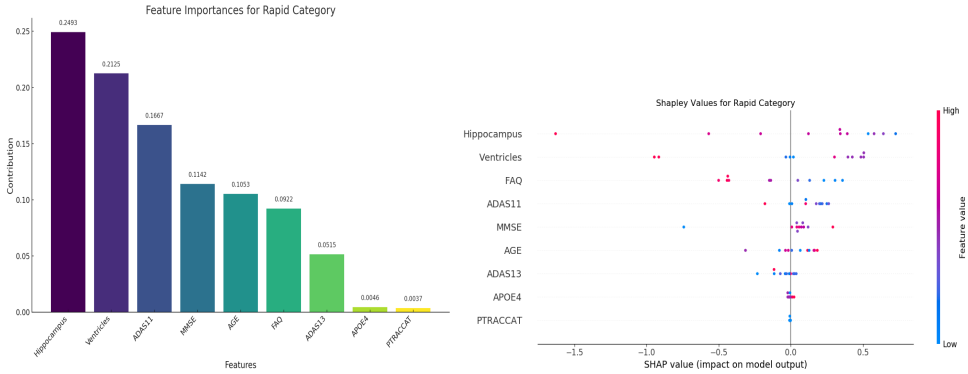



Figure 44: Feature Importances and SHAP Values for the Rapid Progression Category, Emphasizing the Importance of Hippocampal and Ventricular Volume in Forecasting Swift Disease Progression.

Across all categories, hippocampal volume consistently appears as a critical feature, underscoring its importance in AD progression. However, the variation in feature rankings across categories may indicate different pathological or clinical patterns that characterize the rate of progression.

Table 34: Feature Importance in AD Categories: A Quantitative Comparison

| Feature | Gradual | Moderate | Rapid | Sum |
|---|---|---|---|---|
| Age | **0.30** | **0.15** | 0.10 | 0.55 |
| Hippocampus | 0.23 | **0.15** | **0.25** | **0.63** |
| Ventricles | 0.20 | 0.11 | 0.21 | 0.52 |
| ADAS11 | 0.04 | 0.13 | 0.17 | 0.34 |
| ADAS13 | 0.04 | **0.15** | 0.05 | 0.24 |
| MMSE | 0.09 | 0.14 | 0.11 | 0.34 |
| FAQ | 0.06 | 0.12 | 0.08 | 0.26 |
| APOE4 | 0.03 | 0.03 | 0.02 | 0.08 |
| PTRACCAT | 0.01 | 0.02 | 0.01 | 0.04 |

#### 5.3.4.3 Visualizations and Interpretations

Table 34 provides a quantitative complement to the visual feature importance plots, enabling a direct, numerical comparison across the progression categories. Age and hippocampal volume, as quantitatively underscored in the table, are the predominant features influencing the Gradual progression category. This aligns with the visual interpretations that indicated these features as critical predictors for patients with a slower transition to AD. Interestingly, while the table confirms ADAS13 as a significant indicator for Moderate progression, it is ADAS11 that gains prominence in the Rapid category. This shift may reflect the differential sensitivity of these cognitive scales to the stages of AD progression.

Additionally, we present a heatmap visualization in Figure 45. This heatmap employs a color gradient to represent the importance scores of each feature within the categories of Gradual, Moderate, and Rapid progression.

As depicted in the heatmap, warmer colors signify higher importance scores, indicating a stronger predictive value for the respective AD progression category. Conversely, cooler colors correspond to lower importance scores, suggesting a lesser predictive value. The numerical values within each cell of the heatmap provide exact scores consistent with Table 34.

Noteworthy observations from the heatmap include the pronounced importance of age in the Gradual progression category, as indicated by the deep red coloration, with a score of 0.30. In contrast, hippocampal volume shows a high importance score of 0.25 in the Rapid category, marked by a similarly intense color.

The consistent presence of hippocampal volume as a key feature across all categories, as highlighted in both the plots and the table, reiterates its central role in the disease's progression. Conversely, PTRACCAT's minimal importance across the board raises questions about its utility and may suggest the need for alternative measures that can capture the nuances of AD progression more effectively. The variability in feature importances suggests that different pathological mechanisms may be at play as the disease progresses at different rates, warranting a tailored approach to the prediction and management of AD.

Figure 48 illustrates the predicted months until conversion to AD from MCI, plotted against the elapsed months since baseline for each patient within the study. The patients are categorized into three distinct progression trajectories: rapid (red), moderate (blue), and gradual (green). This visualization provides a compelling overview of the heterogeneity in AD progression rates and validates the efficacy of the predictive modeling

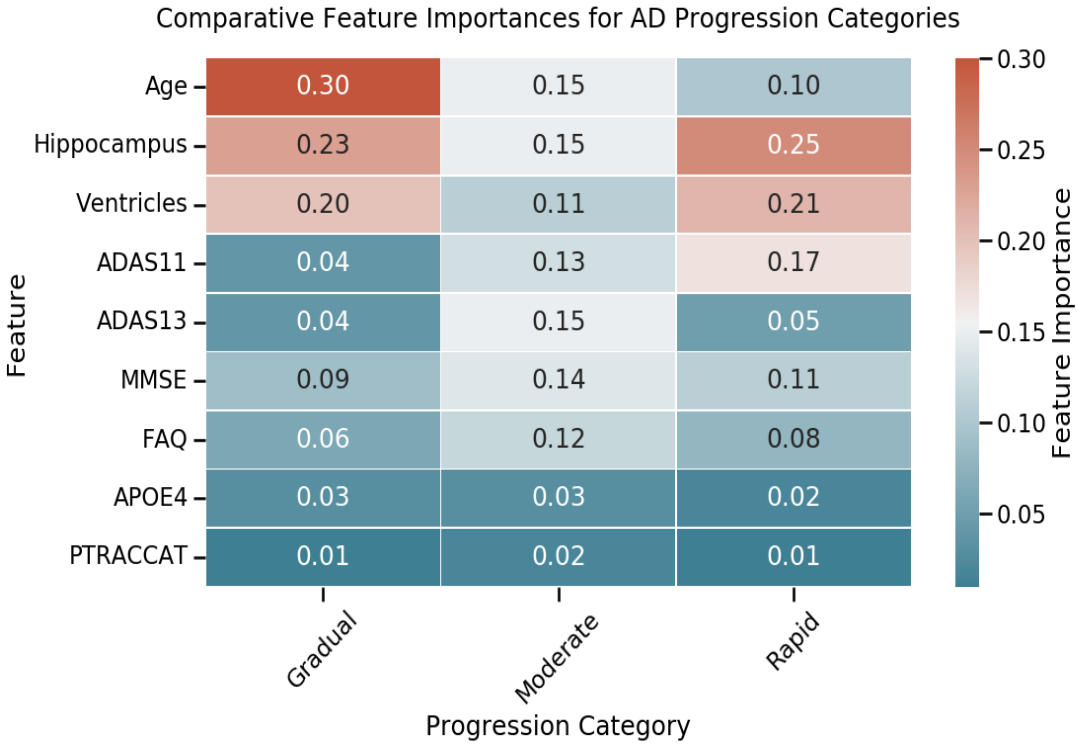



Figure 45: Feature Importance by Progression Categories Heatmap

approach used in this study.

The rapid progression group, depicted in red, is characterized by steep trajectories, indicating a swift decline within a shorter timeframe. In contrast, the moderate (blue) and gradual (green) progression groups exhibit less steep trajectories, reflecting a slower transition to AD. The plot reveals a clear demarcation between the progression rates, underscoring the potential of our model to differentiate between varying rates of disease advancement effectively.

A key observation from this plot is the non-linear progression slopes, which align closely with the known non-linear course of AD. This non-linearity is particularly crucial as it underpins the complex nature of AD progression, where each patient's journey is unique and not strictly time-bound. The implications of these findings are significant for clinical practice, as they suggest that treatment and monitoring strategies should be highly personalized, taking into account the patient's predicted progression rate.

Moreover, the plot serves as a validation tool for our predictive model, demonstrating its capacity to not only forecast the time to AD conversion with a high degree of accuracy but also to categorize patients into clinically relevant progression categories. Such categorization could be instrumental in tailoring intervention strategies, potentially improving patient outcomes by allowing for earlier and more targeted therapeutic approaches.

For this work, a t-SNE visualization was employed to explore the intrinsic structure of the patient data in a two-dimensional space (Figure 46). This non-linear dimensionality reduction technique revealed distinct clusters that correspond to our AD progression categories. In the t-SNE plot, each point represents an individual patient, color-coded according to the progression category: green for Gradual, yellow for Moderate, and purple for Rapid progression. Notably, patients with a Moderate progression pattern tend to cluster centrally, while Rapid and Gradual progressions are somewhat dispersed. This clustering pattern may suggest underlying differences in the progression pathways of AD, with Moderate progression showing the most distinguishable grouping in the reduced space.

For our CNN progression category model, we perform Grad-CAM (Gradient-weighted Class Activation Mapping) analysis to better interpret the model's decision-making process. For a comprehensive analysis, we have included detailed case studies of Patients 2216, 4765, and 2106. These cases were selected due to their unique progression patterns, providing critical insights into the variability of AD. Figure 48 (a) presents a visualization of Grad-CAM for Patient 2216 within the rapid progression category. The sequence of images represents the evolution of DTI scan features, captured from their initial exam visit (leftmost image), an intermediate exam (middle image), and finally their final exam visit (rightmost image). These heatmaps highlight the brain regions that our model identified as critical in predicting disease progression, with the heatmap intensity reflecting our model's focus during a particular exam visit. This specific sequence indicates a significant change of intensity which aligns with the expected rapid progression

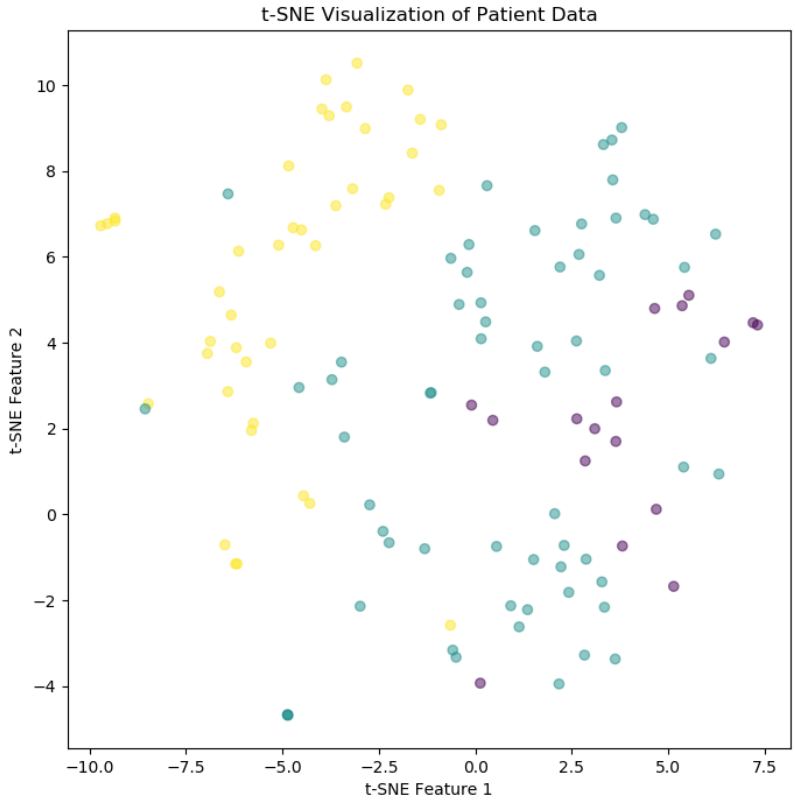
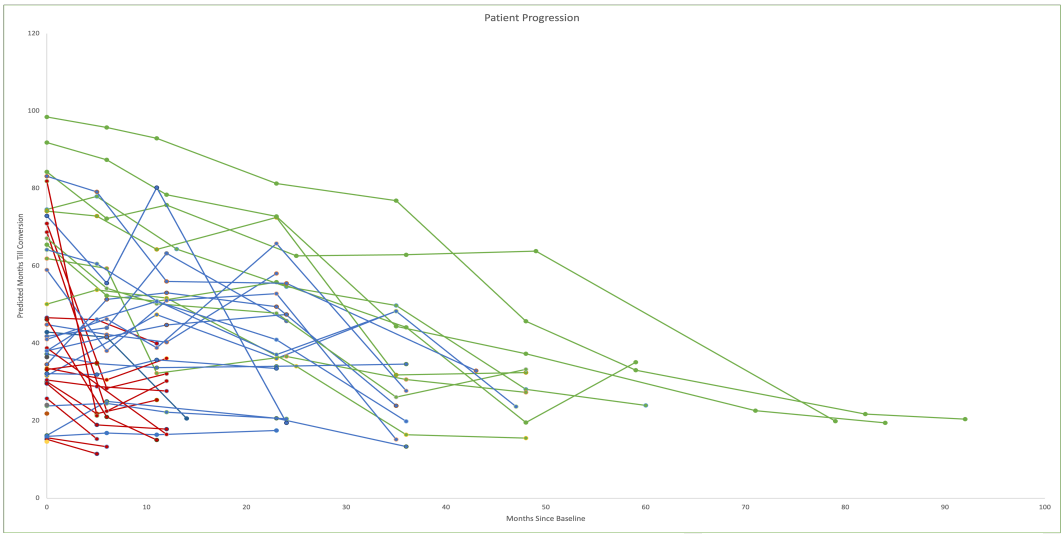
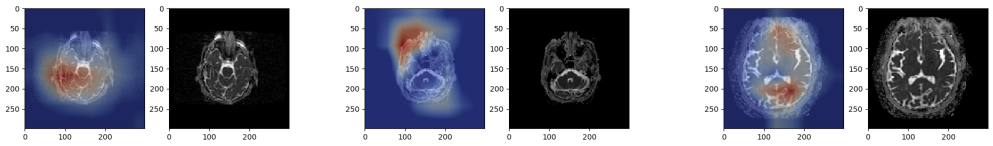


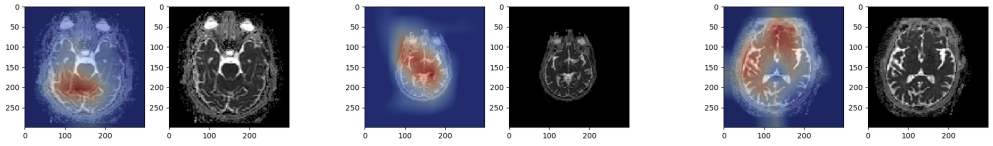

Figure 46: T-SNE Patient Data: 2D Embedding with Progression Categories (Green: Gradual, Yellow: Moderate, Purple: Rapid)

of this category. For the initial exam, the heatmap is primarily centralized at the core brain regions. Progressing to the intermediate exam, we see that the heatmap expands both in scope and intensity. This expansion traverses the surrounding areas, along with providing a more vivid coloration, indicating an escalation in the disease activity. Finally, the latest exam exhibits the brightest areas indicative of significant disease progression. This heatmap is significantly more widespread and intensified reflecting the imminent conversion period to AD.

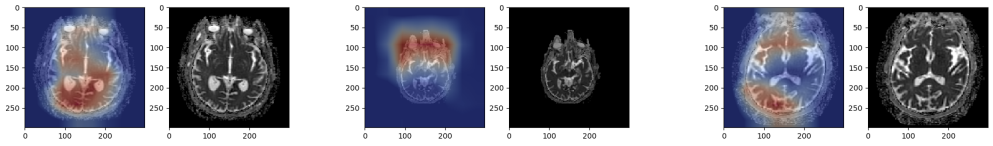


Initial Stage: (a) Rapid (Red) (b) Moderate (Blue) (c) Gradual (Green)



Intermediate Stage: (a) Rapid (Red) (b) Moderate (Blue) (c) Gradual (Green)



Final Stage: (a) Rapid (Red) (b) Moderate (Blue) (c) Gradual (Green)

Figure 47: Grad-CAM Rapid Progression Category Example for Initial, Intermediate, and Final stages for (a) Rapid Progress: Patient 2216 (b) Moderate Progress: Patient 4765 (c) Gradual Progress: Patient 2106

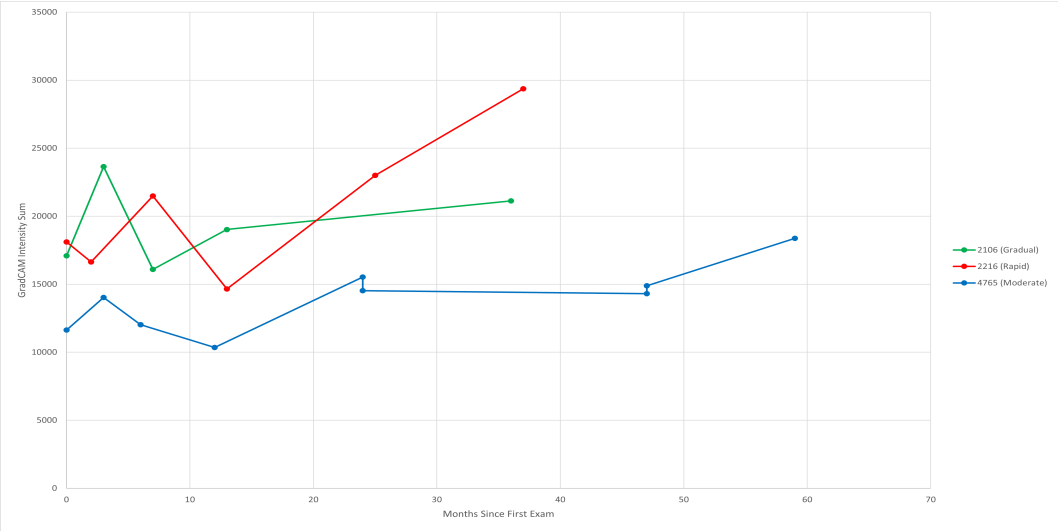



Figure 48: Grad-CAM Intensity by Sample Patients

Figure 48(b) illustrates the Grad-CAM for Patient 4765 in the moderate progression category. The intensity changes are more moderate over time, reflecting an intermediate disease trajectory. This visualization is less aggressive than our rapid example, however, it contains more variability than our gradual progression category.

Figure 48(c) shows the Grad-CAM for Patient 2106 in the gradual progression category. The intensity regions change subtly over time, demonstrating a more gradual development towards AD-specific features.

The sequence of these images demonstrates a clear trend: an escalation in both the spatial distribution and intensity of the highlighted areas as the disease progresses. The progression from sparser, cooler regions to widespread, warmer areas aligns with our hypothesized AD progression, thus offering a compelling visual narrative of a patient's AD journey.

The Grad-CAM Intensity chart presented in Figure 49 quantitatively depicts the

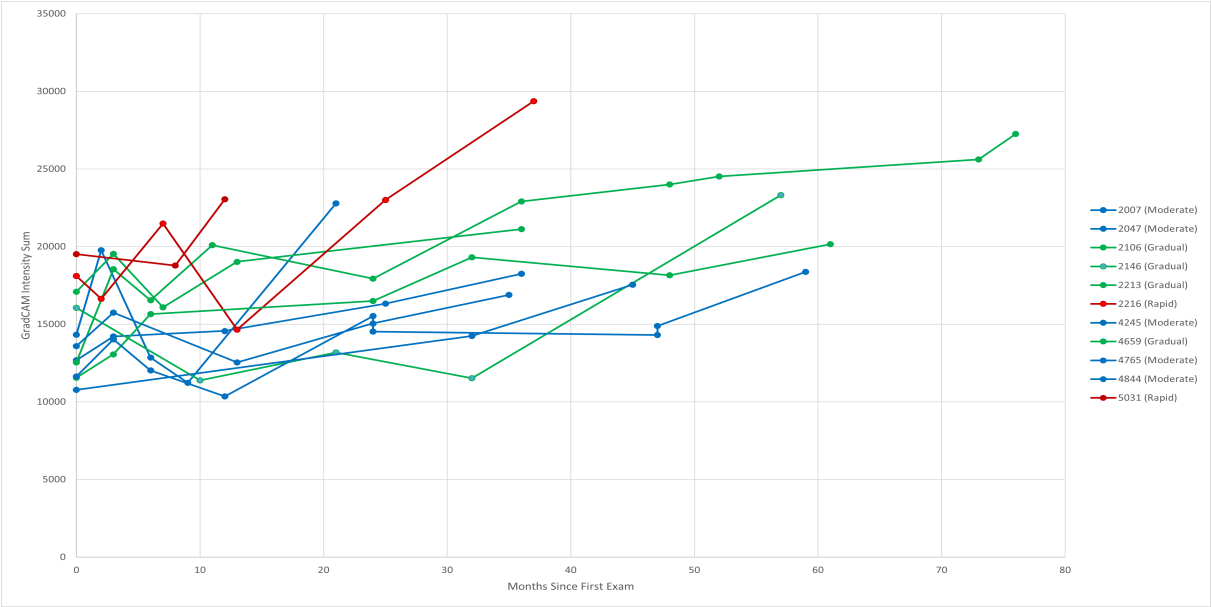



Figure 49: Progression of GradCAM Intensity Sum Over Time

progression of disease-associated features in patients over time, as identified by our CNN model. This chart plots the sum of Grad-CAM intensities for our sample patients at various times since their initial exam. This provides a numerical representation of our model's focus per visit. Each line corresponds to a different patient and aligns with the Grad-CAM analysis seen in Figure 48.

For patients within our progression categories, the chart shows an increasing trend in Grad-CAM intensities. This suggests that our model consistently recognizes stronger AD-related features as time progresses. This trend correlates with our previously observed patterns, where intensity became more pronounced within later exam date images. By comparing these quantitative trends with the qualitative Grad-CAM examples, we are able to validate our model's ability to discern progression rates accurately over time.

Through the combination of our patient progression slope and Grad-CAM analysis, alongside the individual feature ranking, we are able to obtain a more holistic view of each conversion patient. This will allow clinicians to cater their treatment plans based on the aggressiveness of a patient's progression category.

## 5.4 Conclusion

In this study, we have developed a comprehensive analytical framework for AD progression, utilizing both clinical and imaging data. We proposed a regression model that not only predicts the time until AD conversion for MCI patients, but also classifies these patients into rate of progression categories. This regression model, enhanced with SMOGN augmentation, led to an accuracy of 78.34% and an EMSE of 12.61. This provides valuable explainability through our feature importance ranking and novel progression categories.

In addition to our clinical data approach, we introduced a CNN model to further classify patients with DTI scans into progression categories. This augmented model demonstrated an accuracy of 86.49%, illustrating it's ability to learn patterns from the DTI scans.

This dual approach, combining time-to-conversion predictions with progression categorization, represents a significant advancement in AD research, offering a more nuanced understanding of the disease's progression. The integration of Grad-CAM analysis with our model's predictions has enabled us to validate the model's accuracy and provide a more holistic view of each patient's progression. This comprehensive approach is not

only academically significant but also holds immense clinical value. It equips clinicians with a powerful tool to tailor treatment plans according to the aggressiveness of a patient's progression, potentially improving patient outcomes in AD management.

One limitation of our study is that all patients were derived from the ADNI dataset, which may limit the generalizability of our findings. Future work will consider the inclusion of additional datasets to account for further feature variation and enhance the robustness of our model for clinical implementation.

Overall, our study stands at the forefront of AD research, offering novel insights into disease progression through advanced predictive modeling. The combination of multimodal monthly predictive modeling alongside progression categorization represents a substantial contribution to the field of personalized healthcare.

# CHAPTER 6

# CLINICAL DECISION SUPPORT APPLICATION

As we explore the Clinical Decision Support Application, the groundwork laid in the preceding chapters becomes both the foundation and the impetus for innovation. This chapter is dedicated to not just describing a tool but to delineating a new approach to patient care. Here, the abstract meets the practical, and the potential of predictive modeling is funneled into a format that aims to be intuitive, insightful, and integral to the management of Alzheimer's disease. Through this application, we seek to not only predict but to empower, offering a new dimension to decision-making in the complex landscape of neurodegenerative disease management.

Studies on visual attention of patients with Alzheimer's disease and Dementia is a promising way for keeping track of the individual patient's image recognition ability over time. This research seeks to expand upon the current applications of combining the Android operating system with TensorFlow by providing a visual question answering platform for image analysis. This application, Cognitive Visual Recognition Tracker (CVRT), provides an entry point by which the user can ask questions concerning any image of their choosing, and then receive cumulative metrics over time to better assess any diminishing cognitive ability (i.e. Alzheimer's patients). In this chapter, recurrent neural networks as well as semantic analysis are leveraged to provide an interactive VQA experience. One of the main objectives of CVRT is for physicians to be able to determine trends from patient

data that could either be applicable to the individual patient, or to many patients if an aggregate is formed from many individual datasets. On an individual level, these metrics would provide a way for the physician to monitor daily cognitive capability, whereas on a grander scale, these joint datasets could be used to provide better overall treatment for the disease with the future inclusion of predictive analytics. The final contribution is an interactive metrics platform by which other users can assess the primary user's cognitive capacity based on features of their questioning, and to then provide them with accurate trending or possible remediation plans based on their condition.

Patients with Alzheimer's disease and dementia have difficulties in performing tasks of everyday life such as the ability to think, remember, and reason due to loss of cognitive functioning [83]. The number of Alzheimer's disease patients is predicted to almost double every twenty years [84]. Currently, there are approximately 5.4 million Americans with Alzheimer's disease [84]. Studies on visual attention of patients with Alzheimer's disease and dementia is a promising way to keep track of the individual patient's image recognition ability.

There have been significant advances in big data analytics and mobile computing. Recently, mobile applications have been found to be helpful in diagnosis and treatment of diseases due to their accuracy and ability to assist medical practitioners. For example, the iWander app allows for caregivers to cost effectively monitor Alzheimer's disease or dementia patients remotely [85].

While there is some work done in Visual Question Answering [86], very few exist that have applicable results for the medical field. With this in mind, this research seeks

to expand upon the current applications of combining the Android operating system with TensorFlow by providing a visual question answering platform for image analysis. Additionally, we also wanted to add QA diagnostics into our model to further expand the practicality. Cognitive Visual Recognition Tracker (CVRT), was designed to build on the existing Visual Question Answering space, but to produce practical, medically relevant data as a result. This application is significant because it provides a unique way to track the progression of Alzheimer's disease based on image recognition in a more data-driven way than was previously possible. With the combination of intelligent semantic analysis alongside image classification, new methods are now able to be explored in the medical domain, and with CVRT, an attempt is made to build a powerful data model alongside the added flexibility of a mobile UI. By using the COCO-QA dataset mentioned in [87], alongside the TensorFlow and Android framework, CVRT is able to serve the following features to the end user:

- User may choose any picture to be displayed in the application with a question-answer focused interface.
- Users can ask questions about those images in order to discover their identities/characteristics. (i.e., What is happening in this picture? Where is this? What item is on the table?)
- Users can ask non-image related questions regarding medical definitions (e.g. What is Donepezil? What is Galantamine?)
- Users can initiate a preliminary medical diagnosis by describing their symptoms (e.g. I am having memory loss. I feel disoriented.)

- Ability to track metrics for the individual patient's image recognition ability over time.
- Short-term monitoring that can detect delirium and advise for emergent medical treatment.
- Ability to aggregate multiple patients' data to allow for better predictive analytics.

A common scenario for this application would be a family wishing to track the cognitive status of a loved one with dementia or Alzheimer's disease. This individual would be the primary user as they would be the one launching the application to ask questions. The application will allow them to skip as many pictures as they would like, or they can stay focused on one particular image until their curiosities are satisfied. The application will provide the primary user with answers to their questions while simultaneously logging results to a database. This database would then allow for family members to view metrics on the individual's cognitive visual recognition health.

Apart from the deep learning image analysis aspects of CVRT, the medical strength of the application is intrinsically tied to these metrics. For example, by measuring the rate of an individual's cognitive decline, one can determine the type of dementia (vascular vs. Alzheimer's) as well as the aggressiveness of the disease. With the inclusion of semantic analysis, the vocabulary patterns of the individual can also be assessed to measure the progression of aphasia. With this proposed implementation, these metrics function on an individual basis, however the possibility exists for future medical studies on dementia to be performed that leverage this model's aggregate data.

## 6.1 Related Work

### 6.1.1 VQA Methods

In regards to the Visual Question Answering space, similar work has been done related to the questions being asked of an image with answers generated in response. These demos exist online and serve as a form of introduction to the domain itself, however, they do not apply to the medical field.

The work done by [88] attempted to solve the VQA problem by leveraging the Visual Genome dataset, which connects structured image concepts to language. This work has a nice advantage in that the Visual Genome dataset allows for human-generated structure annotations on a per image basis in the form of scene graphs. These scene graphs are based on the nodes of visual elements and while it does handle some question types accurately, their study found that it struggles on "what" questions as those tend to be the most granular in nature. Other papers, such as [89] tackle the problem by leveraging only a CNN model and a bag-of-words approach when predicting their images.

The most ideal approach to this problem from a non-medical standpoint in terms of accuracy and question flexibility was presented in [90]. Their work is very new and they propose a conversational AI bot which can have meaningful dialog with a user about any custom image of the user's choosing. This was an inspiration for the initial back-and-forth image interaction provided by CVRT.

The fact-based VQA method from [91] with its supporting facts functionality, is something that I would be interested to implement in the future work of CVRT. This "common-sense" framework can be used to create a more conversational atmosphere for

the user experience.

### 6.1.2 Methods Closest to CVRT Model

The VQA approaches from [92], [87], [93] are the closest to the one used in CVRT as they rely on the combination of an LSTM model alongside a CNN one. Similar to CVRT, a VGG-16 model [94] is used with an MSCOCO dataset. Where CVRT differs from these implementations is primarily in its mobile platform as well as its medical focus.

One method that did have a relatable medical focus was presented in [86]. They mention having open-ended questions/answers in order to possibly aid the visually impaired, and is one of the few recent examples of a medically applicable VQA solution.

Finally, the work done by [95] was a useful study to realize that humans often have very different attention regions compared to popular VQA models. This was used on a cautionary basis when fine-tuning the CVRT model.

### 6.1.3 Visual Cognition Studies within Alzheimer's Disease

Part of the idea of the metrics component of CVRT derived from the work done in [96]. This study of repetition-lag effects was the basis for the idea of tracking the questions per image frequency and the date grouping. The work done in [97] and [98] highlighted the idea that visual dysfunction can serve as an early indicator of Alzheimer's disease. This served as the medical rationale for the predictive analytics offered by CVRT. With this work, a strong positive correlation was found between visual recognition ability and Alzheimer's disease progression, which formed the starting point for CVRT.

Finally, breakdown characteristics of Alzheimer's disease was studied in [99] and [100], which focused on the specific types of information that is typically lost at the greatest margin. This served as the basis for the CVRT metrics centered on question type and object vs. person cognitive visual recognition performance.

## 6.2 Proposed Work

### 6.2.1 CVRT's Question Answering Model

One primary component of the work is related to providing a preliminary diagnosis based on the users reported symptoms. The CVRT system (Figure 53) is composed of three question answering models: (1) diagnostic question answering (GQA) (2) definition-type question answering (DQA) (3) visual question answering (VQA).

#### 6.2.1.1 Diagnostics Question Answering Model

For the Diagnostics Question Answering (GQA) model of CVRT, it is noticed that despite the many visual features in the chosen image, CVRT identifies that the patient is asking a non-image related question and provides the appropriate follow-up question in response (Figure 50). The GQA model was designed with Infermedica [101] that aims to checker alongside natural language processing to extrapolate the key items out of the users speech.

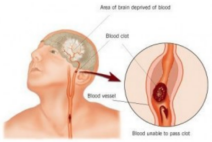
#### 6.2.1.2 Definition-Type Question Answering Model

Regarding the Definition Question Answering (DQA) model of CVRT, one can see that despite prefacing the query with What, CVRT remains capable of distinguishing

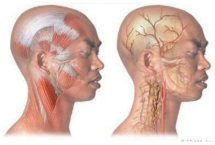
(a) Memory Loss
*"I am experiencing memory loss."*
*Response: "Do you feel disoriented?"*
*Prediction: "Dementia"*

(b) Headache
*"I have a headache."*
*Response: Is your vision impaired?*
Prediction: *"Tension headache"*

Figure 50: Diagnostics Question Answering Examples



(a) Degenerative Vascular Disorder
Question: *"What is vascular dementia?"*
Response: *"a degenerative vascular disorder affecting the brain; it is caused by the blockage of the blood supply to the brain"*
*Prediction: "Degenerative Vascular Disorder"*



(b) Tension Headache?
Question: *"What is a tension headache?"*
Response: *"a headache located at the back of the head; usually caused by body tension."*
Prediction: *"Tension Headache"*

Figure 51: Definition-Type Question Answering Examples

the definition-type query from a visual query (Figure 51). The DQA model was designed with Wolfram Alpha [102] that aims to provide medical definitions for items that the diagnostic feature returns.

#### 6.2.1.3 Visual Question Answering Model

For the visual question answering (VQA) model in CVRT, we can observe that despite the clutter of the image, the visual model is able to correctly assess that there are only two dogs in the image (Figure 52). It is also able to distinguish between the frisbee that is in the dogs mouth, and the one that is lying on the grass.

*"What does the dog have in its mouth?"*
Prediction: *frisbee*
*How many dogs are there?*
Prediction: *two*

Figure 52: Visual Question Answering Examples

Table 35: Additional Result Examples

| Query | Response | Type |
|---|---|---|
| My hand is shaking. | Do you have an elevated body temperature or fever? | GQA |
| I am having neck pain. | Do you have hypertension? | GQA |
| What is bipolar disorder? | A major affective disorder marked by severe mood swings. | DQA |
| What is an antidepressant? | Any of a class of drugs used to treat depression | DQA |
| Where is this? | Hallway | VQA |
| What color is the man's coat? | White | VQA |

### 6.2.2 Proposed System

For the TensorFlow model side, this work was primarily done using recurrent neural networks in combination with semantic analysis. By leveraging the COCO-QA dataset in conjunction with the DAQUAR QA dataset, each custom image can be attributed a proper QA set from which the Android UI can interact with through verbal invocation. The second part of the work is concerned with metrics generation from the SQLite database created from the Android code.

The main analytic tool that will be used for the metrics component of this application will be done with the Google Charting API. This API allows for our stored database information (imageID, question, answer) to be rendered as user-friendly charts that will be displayed via an Android WebView. Bar graphs and line charts will be featured heavily to measure the frequency of the patient's questions in relation to the image that they decided to focus upon. Over time, this can be leveraged to determine trends that could possibly predict the patient's rate of cognitive visual recognition impairment. In addition to identifying trends, these types of graphs will help recognize outliers which could signify the sudden emergence of delirium.

### 6.2.3 Deep Learning Models

For CVRT, the image recognition is being done via the TensorFlow model for neural VQA that is composed of both visual representation and text representation (as shown in Figure 53). As for the CVRT model, using a modified TensorFlow implementation of the VIS+LSTM model seen in [87], the LSTM sentence model forms the basis of the text processing schema. The LSTM outputs are then used to create the answers by being processed through the Softmax layer. In regards to visual embeddings, these are done by extracting features from the FC7 RELU layer within the VGG-16 model itself.

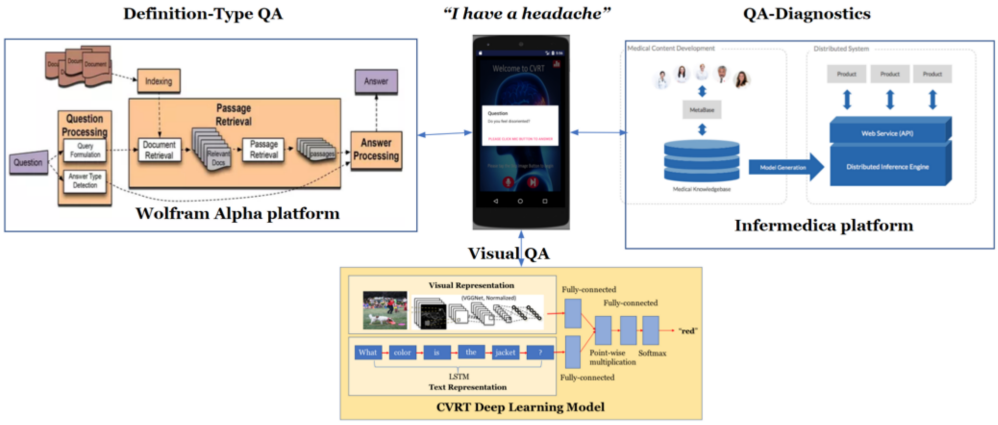



Figure 53: The CVRT Deep Learning Model Schema

The best CNN models for image classification are GoogLeNet [103] and ResNet [104]. However, we used VGG Net [94] that showed good performance on ILSVRC-2012. VGG-16 is also relatively small, simple as it has only three fully-connected layers. The VGG-16 has been reduced to 7.5% of its original size that is critical for real-time image processing. Considering the benefits of the VGG-16 model, it is suitable for our purpose.

## 6.3 Implementation and Evaluation

The primary data source when the application launches will be the Android device's gallery or sdcard photos. Ideal images to select would be common locations that the patient should be familiar with to hopefully gauge if they've retained their familiarity with those locations. If not, how much has it declined? Once the application has loaded, the relevant information will come via the Google Voice Speech Recognizer as the patient asks questions concerning the image. This will be stored in a variable that is then passed to the rest of the application's workflow.

Table 36: Overview of Key Services and APIs Utilized in the Application Development

| **Service Name** | **Description** |
|---|---|
| Heroku | PaaS that allows for web app delivery |
| SQLite | Database local to Android device |
| Android | Primary OS for mobile platforms |
| Google Charting API | Charting engine for data rendering |
| TensorFlow | Machine learning library |
| Google Speech Recognition API | Provides Speech Recognition |

The primary input for CVRT (Cognitive Visual Recognition Tracker) are the questions that are asked via voice recognition by the patient. After this speech is recognized and assigned to a variable, it is then passed to the Python Server alongside the image details to perform the image analysis. This combination of inputs is then used to generate an appropriate answer to the user's question as the output. Another output that can be expected is the write to the database of the image ID, question, and answer. These will then output to a WebView for statistical viewing via the Android device.

### 6.3.1 Software Architecture

Once an image has been selected from the user's phone, it imports as a bitmap and is then converted into a byte array. After this conversion, it is then compressed before being passed to the python server. This is done as a background asynchronous task so that it does not interfere with the processes occurring on the main thread. The python predict.py script asks for only two parameters (image and question) to generate a proper answer back to the Android device.

For the CVRT Android client, once the image, question, and answer are stored in the database, each chart references the data points by using a combination of JavaScript and HTML to pull the appropriate metrics from the SQLite database. This is slightly different depending on the amount of dependent or independent variables used in the chart, however, they are essentially always referencing variables in the JavaScript portion of the CVRT Android model.

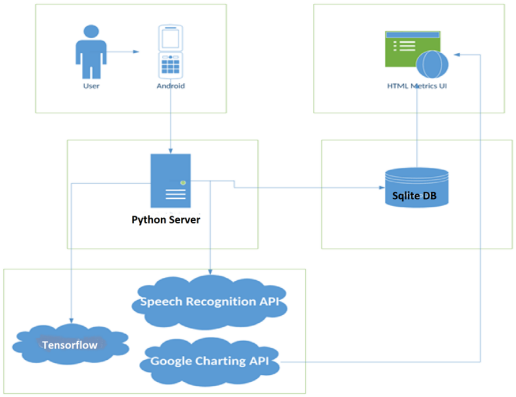



Figure 54: CVRT Software Architecture

### 6.3.2 Dataset

The COCO-QA dataset elements that were used for CVRT are as follows:

- Training dataset: 78,736 question-answer pairs
- Test dataset: 38,948 question-answer pairs
- Image count: 123,287 images

These are further broken down into four distinct types of questions: object, number, color, and location. Also, it is important to note that answers for this dataset are all one-word.

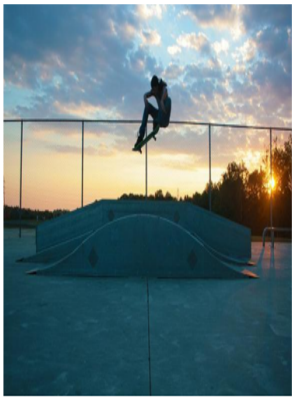
### 6.3.3 Learning Results

The result examples presented below demonstrates the model's reaction to several types of questioning. These differences in question type determines the method the model needs to take for it to properly generate an answer. We can observe that despite the clutter of the top-right image, the model is able to correctly assess that there are only two dogs in the image. It is also able to distinguish between the frisbee that is in the dog's mouth, and the one that is lying on the grass.

Table 37 demonstrates the model accuracy as it relates to the question-answer type. As observed below, the model does well with object type questions as the CNN component handles image classification well. Color questions also seem to perform well despite how cluttered or busy an image might be with objects.

**What does the dog have in its mouth?**
Prediction: frisbee

**How many dogs are there?**
Prediction: two



**What sport is this?**
Prediction: skateboarding

**What color is the strawberry?**
Prediction: red

**Where is this?**
Prediction: hospital

**What is in the hand?**
Prediction: medication

Figure 55: Result Examples of Visual Question Answer

Table 37: COCO-QA Accuracy Per Type

| Object | Number | Color | Location |
|---|---|---|---|
| .6008 | .4534 | .5122 | .4893 |

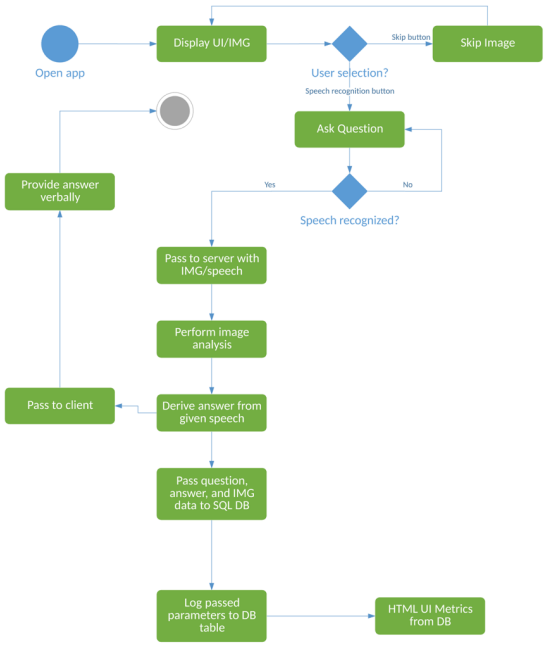



Figure 56: CVRT Workflow

## 6.4 Workflow

The activity diagram shown in Figure 56 gives a good example of the typical workflow that a user can be expected to experience while using CVRT. The speech recognition button triggers the first of our API usages and prepares the intent for the 'Ask Question' node. The question variable is then assigned based on whether the 'Speech Recognized' node can interpret the user's speech or not.

#### 6.4.0.1 Medical Knowledge Engines

One primary component of the CVRT system is related to providing a preliminary diagnosis based on the users reported symptoms. This uses a combination of the Infermedica API symptom checker [101] alongside natural language processing to extrapolate

the key items out of the users speech. This works in tandem with a highly customized Wolfram Alpha API [102] integration that can provide medical definitions for items that the diagnostic feature returns. 1) Medical Knowledge Engines: The two medical knowledge engines that CVRT uses to augment its answer accuracy are the Infermedica platform [101] and the Wolfram Alpha computational knowledge engine [102]. The Infermedica platform derives their data from a combination of their team of physicians using respected literature as well as their AI model based on machine learning. Because of this continual learning, this platform gains accuracy daily. For the Infermedica engine [101], once a question has been given by the user it scans the content using natural language processing to see if any symptoms or other medical keywords are present in the users query. If any are identified then the Infermedica API will return a symptom ID for each and CVRT will then concatenate them into an evidence array. This array is then passed to the diagnostics API which will return both a follow-up question and the likeliest diagnosis at the time. These follow-up questions are used to gather further symptoms from the user by prompting them with relevant yes/no queries to their originally proposed symptoms (e.g. Are you feeling disoriented?). Based on the users response, additional symptoms will be added or negated into the original evidence array. These follow-up questions will continue until the returned diagnosis achieves a probability higher than 70%. Regarding the Wolfram Alpha computational knowledge engine [102], if the users query is identified as asking a definition-type question (e.g. What is Alzheimer's?), then the Wolfram Full Results API will return the answer, typically a medical definition, into an alert dialog. This dialog box differs from the other systems by waiting for user confirmation as more time might be

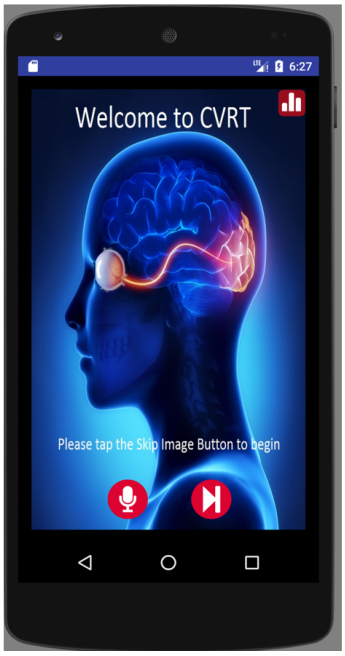



Figure 57: Initial User Interface

necessary to fully read the definition.

#### 6.4.0.2 Visual Question Answering Engine

The implementation of CVRT begins with the user opening the application on their Android device. This presents them with a simple UI that has three options (as shown in Figure 57):

The button on the right acts as a 'Skip' option which prompts the user to select a new image from their phone's gallery in case they are not interested in questioning the existing image. This will allow for increased flexibility as the user can choose any image that is on their phone. This functionality is exemplified by the screenshot of the selection screen as shown in Figure 58.

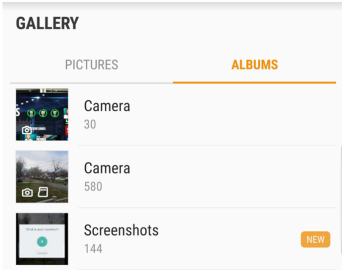



Figure 58: Gallery Selection

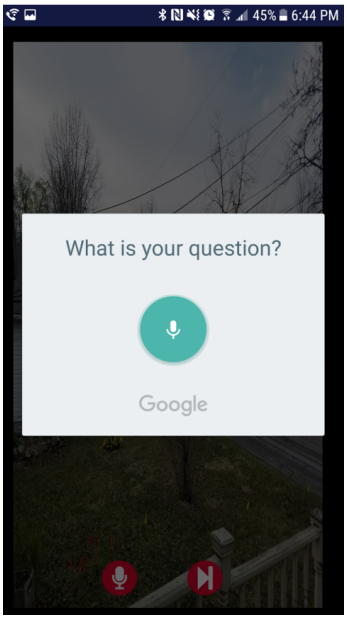



Figure 59: Ask Question UI

The button on the left side of the UI acts as an 'Ask Question' option and allows the user to verbally dictate a question that they have concerning the image. By clicking this option, the following dialog will display as shown in Figure 59.

If no speech is recognized, then CVRT will prompt the user to speak again. Once speech has been detected, the speech is converted to a string and then stored within a variable. After this speech is recognized, it is then passed to three separate systems. The


Figure 60: Example of Returned Answer

Infermedica API [101] is trusted to identify if the users intention is that of a diagnostic nature (e.g. I feel confused). The Wolfram Alpha API [102] is responsible for identifying whether a definition-type question is being asked. If it identifies that a definition is the users intention then it will return the appropriate result and override the returned responses from the other systems. The Python server containing TensorFlow is responsible for the image analysis. This combination of returned results from each system is then used to generate an appropriate answer to the users question as the output. The Android application will then create an asynchronous task to leverage the TensorFlow model for image analysis, and then provide an answer based on its findings. This answer will be verbally returned to the user on the main UI where they can choose to continue asking questions, thus repeating the process, until they are satisfied.

An example of this implementation is as follows. After choosing a picture in the user's phone gallery of a woman on a bicycle, they then tapped on the "Ask a Question" button. This triggered the "What is your question?" speech prompt, which is then answered with "Tell me about this picture." After passing this to the TensorFlow model, the application provides the user with a verbal answer. Also, included with the verbal response is a toast message with the same results in case the user is deaf or hard of hearing. An example of these returned results can be seen in Figure 60. A simple question and answer can be extended as a dialogue to get the user's attention as follows:

**Q:** Who is she?

**A:** Your grandma.

**Q:** What is she doing?

**A:** She is riding a bicycle.

**Q:** What color jacket is she wearing?

**A:** Green.

...

After the answer has been returned to the user, the image ID, question, and answer will be stored into a database table. This can then be leveraged by clicking on the metrics button located in the top-right of the UI once database results have been entered. This button launches a WebView HTML page that takes the database information, calls the

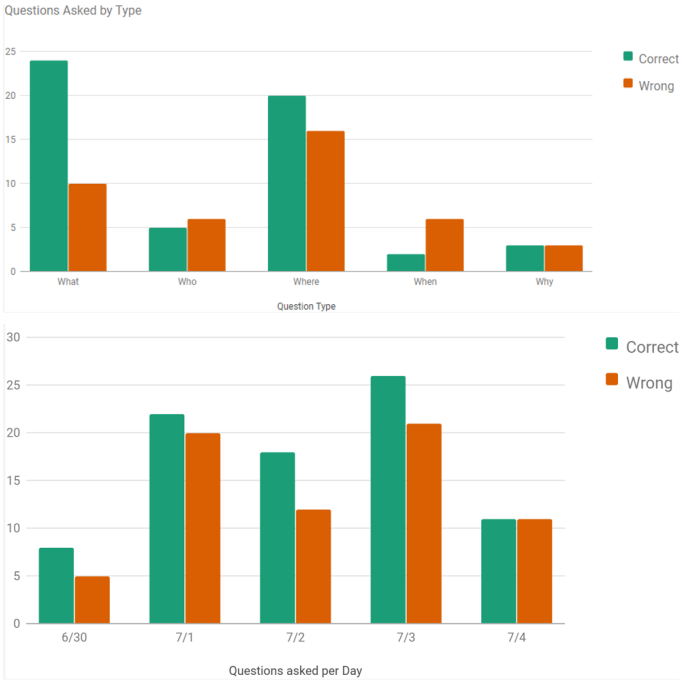



Figure 61: Metrics UI

Google Charting API, and then displays the report metrics as shown in Figure 61. This metrics interface will be viewable to the patient as well as any family members/doctors that would be interested in the individual's cognitive progress/decline.

The possible metrics capable of being displayed can be added in a flexible nature as the Google Charting API does the actual chart creation and only requires certain parameters to be passed into it. The way these additional graphs are added is simply by creating a query against the local SQLite db. As long as the query returns the result set preferred, the rest of the code will handle the conversion through javascript/HTML into the eventual WebView. This provides a way that once ideal medical metrics are identified for the individual or healthcare providers' use case, they can easily be added to the

application in a format that is specific to their needs.

### 6.4.1 Medical Evaluation

The criteria that was used to measure the medical relevance of CVRT correlates with the three factors identified in [105]. They identify these factors as:

1. Feasibility in relation to organizational and systems readiness.

2. Acceptability of the solution.

3. Usability for different users.

When evaluating the feasibility of CVRT the largest barrier is in identifying the best metrics to follow. If the scope is increased to aggregating this data across multiple devices in order to perform further studies, then mutual agreement will be necessary in order to track the same data. If healthcare providers (HCPs) can't agree on the best way to measure data then the application's scope remains solely for the individual end-user. As for system readiness, the application would only require a download from the Google Play store in order to be ready to use. If additional metrics are requested beyond the initial ones provided then that would increase the implementation timeframe. This is something that will be designed in a modular style for future work in order to ease this process and give the control to the user themselves, rather than relying on the developer.

For acceptability evaluation, the flexibility of the metrics UI gives an added bonus as HCPs can customize the tracked datapoints to their specific preference. They could also expose multiple users to different preferences if they wanted to critique the effectiveness

of one solution versus the other. In the case of hospitals or medical research institutions, this would allow for greater control of a given study as they wouldn't have the disagreement issue mentioned in the feasibility analysis given their more organized structure. If ideal metrics were identified and proven to be relevant on a smaller scale, then the willingness of adopting this new technology on a larger scale would greatly increase [105]. On an individual basis, acceptability would be determined by their ability to form a habit to use CVRT almost daily (if necessary). To gather relevant data, semi-frequent user interaction is necessary so that trends can be established. This would be similar to how weight loss applications currently function. Essentially, if the user is not consistent with their application usage, then trending data would not be meaningful.

Finally, the usability evaluation is a critical factor when dealing with a solution that requires consistent usage. Given that only three options are presented to the user for interaction, the UI for CVRT was designed in a manner to facilitate ease of use. Other than the metrics UI, the entire application functions from the main screen rather than sending the user down multiple android Action pages. This simplicity was intentional in order to make the target demographic as comfortable as possible when using the application. The more complicated functionality of interpreting the metric results or determining which ones to deem as relevant are largely meant for the HCP, and as a result, the end-user may never have a need to view that component. Because of this, the metrics option exists only in the top-right of the UI and is not central to the user-driven workflow.

Launching the application is also intended to be simple for the patient as no login is required at this stage so password retention is not necessary. This could be added

in future work should there be a need for multiple users per device, however, it would need to leverage biometric authentication so that the individual's memory recall is never a requirement for entry.

### 6.4.2 Comparative Evaluation

For the first iteration of CVRT work, the Clarifai API was used in order to generate the top captions related to the passed image. While this did have better overall accuracy than the TensorFlow model in regards to basic image captioning, it was limited in regards to its inability of handling dynamic questions and providing appropriate answers. While the Clarifai API model only took a few seconds to call the API and also did not require a local webserver, its lack of VQA essentially relegated it to a foundation for the eventual classification path that CVRT would take. After transitioning to a TensorFlow model, CVRT experienced lessened accuracy at around 50%, however, this was in relation to dynamic questioning, as opposed to generic image captioning. This TensorFlow implementation also required the python prediction script to be referenced over HTTP via Android. This required a local server to be setup that was previously not the case when dealing with the Clarifai API. However, similar functionality existed from the Android side as only one additional parameter (question) was necessary to be passed to the python server in comparison to Clarifai.

## 6.5 Discussion and Limitation

In the current state of CVRT, some of the unanswered questions are more related to the medical side. The cognitive visual recognition tracking outlined above has very

little research done to show its impact on a patient's long-term/short-term mental health [106]. While these metrics are interesting from a familial perspective to see the rate of a person's cognitive decline, it is unproved as to whether these results would be useful to a medical professional given the large quantity of existing medical procedures. As for the data model, future studies could focus on further semantic analysis to determine the frequency of a patient's vocabulary and how that might correlate with changes in their visual recognition ability.

## 6.6 Conclusion

With CVRT, the QA Diagnostics/VQA problem is presented through a new interface via Android. In conjunction with the Infermedica and Wolfram Alpha knowledge engines, TensorFlow model, and semantic analysis, it is able to take any user-selected image, process its features and the passed question, and then return an appropriate answer verbally, providing for an engaging experience for the user. While this model may not be concrete when answering certain types of questioning, it is able to provide a general answer for most question types and images that are presented to it. With the addition of the metrics platform that allows for truly data-driven medical results, my hope is that further research can be done to best exemplify possible treatment plans to better help anyone with this condition.

# CHAPTER 7

# CONCLUSIONS

In this dissertation, we embarked on an exploratory journey into Alzheimer's disease prediction. The aim was clear: to unveil new pathways for early detection and intervention in Alzheimer's disease, a condition that stands as a formidable challenge in the healthcare landscape.

The motivation for this research was rooted in the stark reality of Alzheimer's disease  a neurodegenerative disorder that not only impairs cognitive function but also places a heavy emotional and financial burden on patients, families, and healthcare systems. This backdrop set the stage for a crucial investigation: Could AI transform the way we predict and ultimately confront AD?

The research questions were shaped by this overarching goal. They were designed to probe the potential of AI in enhancing the accuracy and explainability of Alzheimer's disease progression, using a variety of data sources and analytical techniques. This dissertation thus unfolded as a series of methodical inquiries, each chapter building upon the last, to construct a comprehensive understanding of the current state of Alzheimer's prediction and how it might be revolutionized.

As we traverse through the subsequent sections of this conclusion, we will revisit the pivotal moments of this research. We will synthesize the key findings from each chapter, reflect on the methodologies employed, and consider the broader implications of

this work. This synthesis is not merely a recapitulation of the dissertation's contents but an invitation to contemplate the future of Alzheimer's disease management through the lens of technological innovation.

## 7.1 Chapter Summaries

**Chapter 1: Introduction:** The introduction laid the groundwork for the dissertation, establishing the context and significance of Alzheimer's disease prediction. It set the stage for the exploration of AI techniques, outlining the research objectives and the potential impact of these technologies in revolutionizing Alzheimer's disease diagnosis and management.

**Chapter 2: Feature-based Random Forest Model:** This chapter delved into the development and application of a feature-based Random Forest model for Alzheimer's disease prediction. It highlighted the model's ability to handle diverse feature types and its effectiveness in identifying key features associated with the disease, showcasing the model's potential as a tool for early diagnosis.

**Chapter 3: Diffusion Tensor Imaging (DTI) Deep Neural Network (DNN) Model:** The third chapter focused on the application of a Deep Neural Network (DNN) model for analyzing Diffusion Tensor Imaging (DTI) data. This innovative approach demonstrated how advanced imaging techniques, combined with deep learning, can provide critical insights into the structural changes in the brain associated with Alzheimer's disease.

**Chapter 4: Multimodality Ensemble Model:** In this chapter, the research moved towards the integration of multiple modalities of data through an ensemble model. This

combined our previous two models into a better performing model capable of handling multiple types of input data. This multimodality approach underscored the complexity of Alzheimer's disease and the need for comprehensive models that can synthesize information from various sources to improve prediction accuracy.

**Chapter 5: Charting AD Progression: Time-to-Event Predictive Models and Novel Categorization:** This chapter presented an in-depth analysis of AD progression using longitudinal data. A new Random Forest regression model was created alongside a new CNN classifier model. Working in tandem, these models not only were capable of predicting how many months a patient has until AD conversion, but were also able to categorize the rate of their progression into novel progression categories. The focus on temporal changes and progression patterns offered a dynamic perspective on the disease, highlighting the importance of understanding Alzheimer's over time for effective prediction and intervention.

**Chapter 6: Clinical Decision Support Application:** The final chapter described the development of the Cognitive Visual Recognition Tracker (CVRT) clinical decision support application, a practical tool designed to aid healthcare professionals in diagnosing and managing Alzheimer's disease. This application, grounded in the research findings and models developed in previous chapters, represented a crucial step towards translating academic research into practical, clinical tools.

## 7.2 Key Findings

This research has led to several pivotal findings, each contributing significantly to the field of AD prediction.

**Innovation in Feature-based Random Forest Modeling:** The development and application of the Feature-based Random Forest model marked a significant achievement. This model not only identified key features, but demonstrated an accuracy of 93.6% in AD conversion prediction, showcasing the potential of machine learning to improve diagnostic accuracy.

**Breakthroughs in DTI Data Analysis with DNN:** The exploration and application of a Deep Neural Network (DNN) model for analyzing Diffusion Tensor Imaging (DTI) data represented a major step forward. This approach provided crucial insights into the structural brain changes associated with Alzheimer's, illustrating the power of combining advanced imaging techniques with deep learning algorithms.

**Advancements in Multimodality Ensemble Modeling:** Perhaps one of the most significant contributions of this research was the development of a multimodality ensemble model. By effectively integrating diverse data types, this model offered a more comprehensive and accurate approach (98.81%) to predicting Alzheimer's conversion, paving the way for more nuanced and effective diagnostic techniques.

**Insights from Progression Analysis:** The analysis of progression patterns in Alzheimer's disease added a dynamic dimension to the study. This approach allowed for a deeper understanding of the progression of the disease over time, highlighting the importance of temporal factors in effective prediction and management.

**Impact of the Clinical Decision Support Application:** The development of a clinical decision support application based on these models bridged the gap between advanced research and practical clinical tools. This application stands as a crucial tool for healthcare professionals, translating complex data analysis into actionable insights for patient care.

## 7.3 Research Significance

The research conducted in this dissertation holds significant implications for the field of Alzheimer's disease prediction and the broader context of healthcare technology.

**Advancing Alzheimer's Disease Prediction:** The innovative use of AI models, particularly the feature-based Random Forest model, the DTI DNN model, and together as a multi-modality ensemble model, represents a significant step forward in predicting Alzheimer's disease. These models' ability to process and analyze complex, multi-dimensional data offers a more accurate and comprehensive approach to early detection, which is crucial for timely intervention.

**Contributions to AI in Healthcare:** This research contributes substantially to the evolving domain of AI applications in healthcare. By demonstrating how different AI approaches can be effectively utilized in the context of a complex neurological disorder, this work sets a precedent for future research in the field and highlights the versatility and potential of these technologies in medical diagnostics.

**Impact on Clinical Practices:** The development of a clinical decision support application based on these models bridges the gap between theoretical research and practical application. This tool has the potential to transform clinical practices by providing healthcare

professionals with advanced, data-driven insights for early diagnosis and management of Alzheimer's disease, thereby improving patient outcomes.

**Fostering Interdisciplinary Collaboration:** The research underscores the importance of interdisciplinary approaches in tackling complex health challenges. By integrating techniques from computer science, neurology, and data science, the study serves as an example of how collaborative efforts can lead to more comprehensive and effective solutions in healthcare.

**Setting New Directions for Future Research:** The methodologies and findings of this dissertation open new avenues for further exploration in Alzheimer's disease prediction. They challenge future researchers to continue innovating and expanding the boundaries of artifical intelligence in medical applications, encouraging a continued focus on multimodal and longitudinal data analysis.

In summary, the significance of this dissertation lies in its successful demonstration of how AI techniques can be harnessed to make meaningful advancements in the diagnosis and treatment of Alzheimer's disease. It stands as a beacon of innovation and interdisciplinary collaboration, paving the way for future research and advancements in the field of healthcare technology.

## 7.4 Limitations

While this dissertation has made significant strides in the field of Alzheimer's disease prediction using AI, it is important to acknowledge its limitations. These limitations

offer a realistic perspective on the findings and provide valuable directions for future research.

**Data Scope and Diversity in the Models:** One of the primary limitations involves the scope and diversity of data used in the models. While these models are powerful, their effectiveness is contingent on the ADNI dataset that they were trained on. Future research would benefit from incorporating additional datasets to enhance the models' generalizability.

**Generalizability of the Clinical Decision Support Application:** While the clinical decision support application developed in this study represents a significant practical contribution, its applicability and usability in diverse clinical settings have yet to be fully assessed. Further research and field testing are needed to determine its effectiveness across different healthcare environments.

**Interdisciplinary Collaboration Opportunities:** The study, while interdisciplinary in nature, also points towards opportunities for even broader collaboration. Future research could benefit from a more extensive integration of expertise from fields such as geriatrics, psychology, and genetics, to provide a more holistic approach to Alzheimer's disease prediction and management.

These limitations, while underscoring areas for improvement, also provide a roadmap for future research. They highlight the continuous need for innovation, collaboration, and adaptation in the pursuit of advanced solutions for Alzheimer's disease.

## 7.5 Future Work

The conclusions drawn from this dissertation pave the way for further research in the field of Alzheimer's disease prediction. The following topics are proposed for future investigation, aiming to enhance the predictive capabilities and understanding of the disease's progression.

### 7.5.1 Comorbidity Correlation

**Comorbidity Correlation:** Future research should delve into the comorbidity correlations within the ADNI dataset to evaluate their impact on time-to-event conversion prediction. Understanding how comorbid conditions influence the progression of Alzheimer's disease could significantly refine prediction models.

### 7.5.2 Analysis of Clinical Notes

**Analysis of Clinical Notes:** Investigating the informational content of patients' clinical notes offers a promising direction. Analyzing narrative records could provide deeper insights into the individual's disease trajectory and inform the rate of progression, potentially enhancing predictive accuracy.

### 7.5.3 Leveraging LLM Models

**Leveraging LLM Models:** Exploring the application of Large Language Models (LLMs) presents an exciting frontier. Comparative studies assessing the performance of LLMs against current methods could reveal new perspectives and techniques for predicting Alzheimer's disease progression.

By exploring these avenues, future studies will likely yield novel insights and methodologies, contributing to the ongoing efforts to combat Alzheimer's disease with advanced predictive analytics.

## 7.6 Reflection on Personal Learning and Development

Undertaking this dissertation has been an enriching journey of personal and professional growth, marked by a deepened understanding and a broadened skill set in the realm of artificial intelligence and its application in healthcare.

**Enhanced Understanding of Artificial Intelligence and Neurology:** Delving into advanced models such as the Feature-Based Random Forest, Random Forest Regressor, DTI DNN, and Multimodality Ensemble Model for AD prediction has significantly expanded my knowledge and expertise. This experience has illuminated the intricate ways in which AI can be applied to complex neurological disorders, offering insights into both the potential and the challenges of this technology in healthcare.

**Skills Development in Advanced Data Analysis:** The process of developing and refining these sophisticated models has honed my skills in data analysis, particularly in handling and interpreting complex datasets. It has taught me the value of precision and rigor in research, as well as the importance of being adaptable in the face of evolving technological landscapes.

**Appreciation for Interdisciplinary Research:** This journey has reinforced my belief in the power of interdisciplinary collaboration. Integrating insights from various fields has not only enriched the research but also highlighted the importance of diverse perspectives

in tackling complex healthcare challenges.

**Personal Growth and Resilience:** Throughout this research process, I have developed resilience in overcoming challenges and setbacks. It has taught me the value of perseverance, critical thinking, and the pursuit of knowledge, shaping my approach to future research and professional endeavors.

In reflection, this dissertation journey has been transformative, shaping my professional identity and preparing me for a future career in data science. The skills, knowledge, and insights gained through this research are invaluable assets that I will carry forward in my ongoing quest for innovation and excellence in the field.

## 7.7 Concluding Thoughts

As this dissertation journey concludes, it not only marks the completion of a significant academic endeavor but also represents a meaningful contribution to the field of AD prediction and the broader domain of healthcare technology.

This research has traversed the complex landscape of AI in healthcare, tackling the formidable challenge of AD prediction and progression analysis with innovative approaches. The development and analysis of the Feature-based Random Forest Model, the DTI DNN Model, the Random Forest Regression Model,and the Multimodality Ensemble Model have shed new light on the capabilities and potential of AI in medical diagnostics. These models, each with their unique strengths, collectively represent a significant advancement in our ability to predict and understand Alzheimer's disease.

The journey, however, extends beyond the technical achievements. It has been a

testament to the strength of interdisciplinary research and collaboration. The integration of knowledge from fields such as computer science, medical bioinformatics, and data analytics has underscored the importance of a holistic approach in addressing complex health challenges.

Moreover, the development of CVRT, the clinical decision support application, highlights the critical transition from theoretical research to practical, real-world applications. This tool embodies the essence of the dissertation â bridging the gap between computational research and clinical practice, making sophisticated predictive models accessible and useful in the context of patient care.

In reflection, this dissertation stands as a source of optimism and advancement in the fight against Alzheimer's disease. It represents the convergence of technological innovation and medical research. While this journey may be concluding, it paves the way for future exploration and advancements. The quest for knowledge, innovation, and the betterment of human health continues, fueled by the foundations laid by this research.

In closing, this dissertation is more than a collection of models and analyses; it is a narrative of discovery, challenge, and the relentless pursuit of a future where the early detection and intervention of Alzheimer's disease is not just a possibility, but a reality.

Bibliography

[1] Disease International. *World Alzheimer Report 2011: The benefits of early diagnosis and intervention*. Tech. rep. Alzheimer's Disease International, 2011. URL: www.alz.co.uk/worldreport2011.

[2] *Random Forests Definition — DeepAI*. URL: https://deepai.org/machine-learning-glossary-and-terms/random-forest.

[3] Alessia Sarica, Antonio Cerasa, and Aldo Quattrone. "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review". In: *Frontiers in Aging Neuroscience* 9.OCT (Oct. 2017), p. 329. ISSN: 1663-4365. DOI: 10.3389/fnagi.2017.00329. URL: http://journal.frontiersin.org/article/10.3389/fnagi.2017.00329/full.

[4] Michael W. Weiner et al. "The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception". In: *Alzheimers and Dementia* 9.5 (Sept. 2013). ISSN: 15525260. DOI: 10.1016/j.jalz.2013.05.1769.

[5] Melanie Ganz et al. "Relevant feature set estimation with a knock-out strategy and random forests". In: *NeuroImage* 122 (Nov. 2015), pp. 131–148. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2015.08.006.

[6] A. V. Lebedev et al. "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness". In: *NeuroImage: Clinical* 6 (Jan. 2014), pp. 115–125. ISSN: 22131582. DOI: 10.1016/j.nicl.2014.08.023.

[7] Han Li et al. "Hierarchical interactions model for predicting mild cognitive impairment (MCI) to Alzheimer's disease (AD) conversion". In: *PLoS ONE* 9.1 (Jan. 2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0082450.

[8] Kexin Huang et al. “A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer’s disease by using a predictive nomogram”. In: *Neuropsychopharmacology* 45.2 (2020), pp. 358–366. DOI: 10.1038/s41386-019-0551-0. URL: https://doi.org/10.1038/s41386-019-0551-0.

[9] P. J. Moore, T. J. Lyons, and J. Gallacher. “Random forest prediction of Alzheimer’s disease using pairwise selection from time series data”. In: *PLOS ONE* 14.2 (Feb. 2019). Ed. by Stephen D Ginsberg, e0211558. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0211558. URL: https://dx.plos.org/10.1371/journal.pone.0211558.

[10] Sijan S Rana et al. “A multi-modal deep learning approach to the early prediction of mild cognitive impairment conversion to Alzheimer’s disease.” In: Institute of Electrical and Electronics Engineers (IEEE), Dec. 2020, pp. 9–18. DOI: 10.1109/bdcat50828.2020.00013.

[11] A. Thushara et al. “Multimodal MRI based classification and prediction of Alzheimer’s disease using random forest ensemble”. In: *Proceedings - 2020 Advanced Computing and Communication Technologies for High Performance Applications, ACCTHPA 2020*. Institute of Electrical and Electronics Engineers Inc., July 2020, pp. 249–256. ISBN: 9781728164533. DOI: 10.1109/ACCTHPA49271.2020.9213211.

[12] Alzheimer’s Therapeutic Research Institute: ADNI Team. *Alzheimer’s Disease Neuroimaging Initiative (ADNI)*. URL: http://adni.loni.usc.edu/study-design/.

[13] Jack Albright. “Forecasting the progression of Alzheimer’s disease using neural networks and a novel preprocessing algorithm”. In: *Alzheimer’s and Dementia: Translational Research and Clinical Interventions* 5 (Jan. 2019), pp. 483–491. ISSN: 23528737. DOI: 10.1016/j.trci.2019.07.001. URL: /pmc/articles/PMC6804703/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6804703/.

[14] Massimiliano Grassi et al. “A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer’s disease using socio-demographic characteristics, clinical information, and neuropsychological measures”. In: *Frontiers in Neurology* 10.JUL (2019). ISSN: 16642295. DOI: 10.3389/fneur.2019.00756. URL: https://pubmed.ncbi.nlm.nih.gov/31379711/.

[15] Mostafa Mehdipour Ghazi et al. *Robust training of recurrent neural networks to handle missing data for disease progression modeling*. Tech. rep. 2018. URL: https://arxiv.org/abs/1808.05500.

[16] *5 SMOTE techniques for oversampling your imbalance data*. URL: https://towardsdatascience.com/5-smote-techniques-for-oversampling-your-imbalance-data-b8155bdbe2b5.

[17] Hui Han, Wen Yuan Wang, and Bing Huan Mao. “Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning”. In: *Lecture Notes in Computer Science*. Vol. 3644. PART I. Springer Verlag, 2005, pp. 878–887. DOI: 10.1007/11538059{_}91.

[18] Haibo He et al. “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”. In: *Proceedings of the International Joint Conference on Neural Networks*. 2008, pp. 1322–1328. ISBN: 9781424418213. DOI: 10.1109/IJCNN.2008.4633969.

[19] Stacey Ronaghan. *The mathematics of decision trees, random forest and feature importance in scikit-learn and spark*. URL: https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3.

[20] Rita Guerreiro and Jose Bras. “The age factor in Alzheimer’s disease”. In: *Genome Medicine* 7.1 (Oct. 2015), pp. 1–3. ISSN: 1756994X. DOI: 10.1186/s13073-015-0232-5. URL: /pmc/articles/PMC4617238/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617238/.

[21] Matthew Velazquez et al. “RNN-based Alzheimer’s disease prediction from prodromal stage using diffusion tensor imaging”. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) RNN-Based* Mci (2019), pp. 1665–1672. DOI: 10.1109/bibm47256.2019.8983391.

[22] Fanny M. Elahi and Bruce L. Miller. “A clinicopathological approach to the diagnosis of dementia”. In: *Nature Reviews Neurology* 13.8 (Aug. 2017), pp. 457–476. ISSN: 17594766. DOI: 10.1038/nrneurol.2017.96. URL: /pmc/articles/PMC5771416/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5771416/.

[23] Samuel Iddi et al. “Predicting the course of Alzheimer’s progression”. In: *Brain Informatics* 6.1 (Dec. 2019). ISSN: 21984026. DOI: 10.1186/s40708-019-0099-0. URL: /pmc/articles/PMC6598897/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6598897/.

[24] CDC. *U.S. burden of Alzheimer’s disease, related dementias to double by 2060*. https://www.cdc.gov/media/releases/2018/p0920-alzheimers-burden-double-2060.html. [Online; accessed 19-August-2019]. 2018.

[25] Kejal Kantarci et al. “White-matter integrity on DTI and the pathologic staging of Alzheimer’s disease”. In: *Neurobiology of Aging* 56 (2017), pp. 172–179.

[26] Olfa Ben Ahmed et al. “Classification of Alzheimer’s disease subjects from MRI using hippocampal visual features”. In: *Multimedia Tools and Applications* 74.4 (2015), pp. 1249–1266.

[27] Jose Soares et al. “A hitchhiker’s guide to diffusion tensor imaging”. In: *Frontiers in Neuroscience* 7 (2013), p. 31.

[28] Tagm Huisman. “Diffusion-weighted and diffusion tensor imaging of the brain, made easy”. In: *Cancer Imaging* 10 Spec no A (2010), S163–71.

[29] Kenichi Oishi et al. “DTI analyses and clinical applications in Alzheimer’s disease”. In: *Journal of Alzheimer’s Disease* 26.s3 (2011), pp. 287–296.

[30] Stephanos Leandrou et al. “Quantitative MRI brain studies in mild cognitive impairment and Alzheimer’s disease: a methodological review”. In: *IEEE Reviews In Biomedical Engineering* 11 (2018), pp. 97–111.

[31] Simeon Spasov et al. “A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer’s disease”. In: *Neuroimage* 189 (2019), pp. 276–287.

[32] Nicola Amoroso et al. “Deep learning reveals Alzheimer’s disease onset in MCI subjects: Results from an international challenge”. In: *Journal of Neuroscience Methods* 302 (2018), pp. 3–9.

[33] Hongming Li et al. “A deep learning model for early prediction of Alzheimer’s disease dementia based on hippocampal magnetic resonance imaging data”. In: *Alzheimer’s & Dementia* (2019).

[34] Massimiliano Grassi et al. “A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimerâs disease using socio-demographic characteristics, clinical information, and neuropsychological measures.” In: *bioRxiv* (2019), p. 564716.

[35] Alexander Khvostikov et al. *3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies*. Tech. rep. URL: https://ida.loni.usc.edu.

[36] Clifford R Jack Jr et al. “The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods”. In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.4 (2008), pp. 685–691.

[37] ADNI. *ADNI — Study Design - Alzheimer’s Disease Neuroimaging Initiative*. http://adni.loni.usc.edu/study-design/. [Online; accessed 19-August-2019]. 2019.

[38] Harvard Health Letter. *Staving off dementia when you have mild cognitive impairment*. https://www.health.harvard.edu/staying-healthy/staving-off-dementia-

when-you-have-mild-cognitive-impairment. [Online; accessed 19-August-2019]. 2018.

[39] Alex Ward et al. “Rate of conversion from prodromal Alzheimer’s disease to Alzheimer’s dementia: a systematic review of the literature”. In: *Dementia and Geriatric Cognitive Disorders Extra* 3.1 (Sept. 2013), pp. 320–332. ISSN: 1664-5464. DOI: 10.1159/000354370. URL: https://pubmed.ncbi.nlm.nih.gov/24174927/.

[40] Barret Zoph et al. “Learning transferable architectures for scalable image recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8697–8710.

[41] AutoML. *AutoML for large scale image classification and object detection*. https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html. [Online; accessed 19-August-2019]. 2019.

[42] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2818–2826.

[43] Chenxi Liu et al. “Progressive neural architecture search”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 19–34.

[44] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 248–255.

[45] Hoo-Chang Shin et al. “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1285–1298.

[46] Alexis Moscoso et al. “Prediction of Alzheimer’s disease dementia with MRI beyond the short-term: Implications for the design of predictive models”. In: *NeuroImage: Clinical* 23 (2019), p. 101837.

[47] Olfa Ben Ahmed et al. “Alzheimer’s disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex”. In: *Computerized Medical Imaging and Graphics* 44 (2015), pp. 13–25.

[48] Ashkan Ebadi et al. “Ensemble classification of Alzheimer’s disease and mild cognitive impairment based on complex graph measures from diffusion tensor images”. In: *Frontiers in Neuroscience* 11 (2017), p. 56.

[49] Wook Lee, Byungkyu Park, and Kyungsook Han. “Svm-based classification of diffusion tensor imaging data for diagnosing Alzheimer’s disease and mild cognitive impairment”. In: *International Conference on Intelligent Computing*. Springer. 2015, pp. 489–499.

[50] Baiying Lei et al. “Discriminative learning for Alzheimer’s disease diagnosis via canonical correlation analysis and multimodal fusion”. In: *Frontiers in Aging Neuroscience* 8 (2016), p. 77.

[51] Adrien Payan and Giovanni Montana. “Predicting Alzheimer’s disease: a neuroimaging study with 3D convolutional neural networks”. In: *arXiv preprint arXiv:1502.02506* (2015).

[52] Ciprian D Billones et al. “Demnet: A convolutional neural network for the detection of Alzheimer’s Disease and mild cognitive impairment”. In: *2016 IEEE Region 10 Conference (TENCON)*. IEEE. 2016, pp. 3724–3727.

[53] New York Times. *An Alzheimer’s drug trial gave me hope, and then it ended*. https://www.nytimes.com/2019/03/22/well/mind/alzheimers-drug-trial-study-biogen-dementia-treatment-cure.html. [Online; accessed 19-August-2019]. 2019.

[54] Yanjun Qi. “Random forest for bioinformatics”. In: *Ensemble Machine Learning*. Springer, 2012, pp. 307–323.

[55] Tingting Zhang et al. “Predicting MCI to AD Conversation using integrated sMRI and rs-fMRI: machine learning and graph theory approach”. In: *Frontiers in Aging Neuroscience* 13 (July 2021), p. 429. ISSN: 16634365. DOI: 10.3389/FNAGI.2021.688926/BIBTEX.

[56] Sidra Minhas et al. “Early MCI-to-AD conversion prediction using future value forecasting of multimodal features”. In: *Computational Intelligence and Neuroscience* 2021 (2021). ISSN: 1687-5273. DOI: 10.1155/2021/6628036. URL: https://pubmed.ncbi.nlm.nih.gov/34608385/.

[57] Weiming Lin et al. “Predicting Alzheimer’s Disease conversion from mild cognitive impairment using an extreme learning machine-based grading method with multimodal data”. In: *Frontiers in Aging Neuroscience* 0 (Apr. 2020), p. 77. ISSN: 1663-4365. DOI: 10.3389/FNAGI.2020.00077.

[58] Kexin Huang et al. “A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer’s disease by using a predictive nomogram”. In: *Neuropsychopharmacology* 45.2 (2020), pp. 358–366.

[59] Yogatheesan Varatharajah et al. “Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics”. In: *Scientific Reports* 9.1 (2019), pp. 1–15.

[60] Sijan S Rana et al. “A multi-modal deep learning approach to the early prediction of mild cognitive impairment conversion to Alzheimer’s disease”. In: *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*. IEEE. 2020, pp. 9–18.

[61] Fabien Viton et al. “Heatmaps for visual explainability of CNN-based predictions for multivariate time series with application to healthcare”. In: *2020 IEEE International Conference on Healthcare Informatics, ICHI 2020* (Nov. 2020). DOI: 10.1109/ICHI48887.2020.9374393.

[62] Barbara Mukami Maweu et al. “CEFEs: A CNN explainable framework for ECG signals.” In: *Artificial Intelligence in Medicine* 115 (May 2021), p. 102059. ISSN: 0933-3657. DOI: 10.1016/J.ARTMED.2021.102059.

[63] Rafid Mostafiz et al. “Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features”. In: *Journal of King Saud University - Computer and Information Sciences* (Dec. 2020). ISSN: 1319-1578. DOI: 10.1016/J.JKSUCI.2020.12.010.

[64] Ishaani Priyadarshini and Vikram Puri. “A convolutional neural network (CNN) based ensemble model for exoplanet detection”. In: *Earth Science Informatics 2021 14:2* 14.2 (Feb. 2021), pp. 735–747. ISSN: 1865-0481. DOI: 10.1007/S12145-021-00579-5. URL: https://link.springer.com/article/10.1007/s12145-021-00579-5.

[65] Matthew Velazquez and Yugyung Lee. “Random forest model for feature-based Alzheimer’s disease conversion prediction from early mild cognitive impairment subjects”. In: *PLOS ONE* 16.4 (Apr. 2021). Ed. by Stephen D. Ginsberg, e0244773. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0244773. URL: https://dx.plos.org/10.1371/journal.pone.0244773.

[66] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 618–626.

[67] Onder Aydemir. “A new performance evaluation metric for classifiers: polygon area metric.” In: *Journal of Classification* 38.1 (Apr. 2020), pp. 16–26. ISSN: 14321343. DOI: 10.1007/S00357-020-09362-5. URL: https://dl.acm.org/doi/10.1007/s00357-020-09362-5.

[68] William La Cava et al. “Interpretation of machine learning predictions for patient outcomes in electronic health records”. In: *AMIA Annual Symposium Proceedings*. Vol. 2019. American Medical Informatics Association. 2019, p. 572.

[69] Ramprasaath R Selvaraju et al. “Grad-CAM: Why did you say that?” In: *ArXiv Preprint ArXiv:1611.07450* (2016).

[70] Nishanth Arun et al. “Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging”. In: *Radiology: Artificial Intelligence* 3.6 (Oct. 2021). ISSN: 26386100. DOI: 10.1148/RYAI.2021200267. URL: https://pubs.rsna.org/doi/10.1148/ryai.2021200267.

[71] Matthew Velazquez and Yugyung Lee. “Multimodal ensemble model for Alzheimer’s disease conversion prediction from early mild cognitive impairment subjects”. In: *Computers in Biology and Medicine* 151.Pt A (Dec. 2022). ISSN: 1879-0534. DOI: 10.1016/J.COMPBIOMED.2022.106201. URL: https://pubmed.ncbi.nlm.nih.gov/36370583/.

[72] Lars Lau Raket. “Statistical disease progression modeling in Alzheimer’s disease”. In: *Frontiers in Big Data* 3 (Aug. 2020), p. 24. ISSN: 2624909X. DOI: 10.3389/FDATA.2020.00024/BIBTEX.

[73] Paula Branco et al. “SMOGN: a pre-processing approach for imbalanced regression”. In: *Proceedings of Machine Learning Research* 74 (2017), pp. 36–50.

[74] Divya Ramamoorthy et al. “Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data”. In: *Nature Computational Science 2022 2:9* 2.9 (Sept. 2022), pp. 605–616. ISSN: 2662-8457. DOI: 10.1038/s43588-022-00299-w. URL: https://www.nature.com/articles/s43588-022-00299-w.

[75] Shaker El-Sappagh et al. “Two-stage deep learning model for Alzheimer’s disease detection and prediction of the mild cognitive impairment time”. In: *Neural Computing and Applications* 34 (2022), pp. 14487–14509. ISSN: 14333058. DOI: 10.1007/s00521-022-07263-9. URL: https://link.springer.com/article/10.1007/s00521-022-07263-9.

[76] Shaker El-Sappagh et al. “Automatic detection of Alzheimer’s disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers”. In: *Neurocomputing* 512 (Nov. 2022), pp. 203–224. ISSN: 0925-2312. DOI: 10.1016/J.NEUCOM.2022.09.009.

[77] David A. Loeffler. “Modifiable, non-modifiable, and clinical factors associated with progression of Alzheimer’s disease”. In: *Journal of Alzheimer’s Disease* 80.1 (Jan. 2021), pp. 1–27. ISSN: 1387-2877. DOI: 10.3233/JAD-201182.

[78] Radmila Janković et al. “Machine learning models for ecological footprint prediction based on energy parameters”. In: *Neural Computing and Applications* 33.12 (June 2021), pp. 7073–7087. ISSN: 14333058. DOI: 10.1007/S00521-020-05476-4/FIGURES/7. URL: https://link.springer.com/article/10.1007/s00521-020-05476-4.

[79] Somaye Hashemifar, Claudia Iriondo, and Mohsen Hejrati. “DeepAD: a robust deep learning model of Alzheimer’s disease progression for real-world clinical applications”. In: (). URL: https://arxiv.org/abs/2203.09096.

[80] Dustin van der Haar et al. “An Alzheimer’s disease category progression subgrouping analysis using manifold learning on ADNI”. In: *Scientific Reports 2023 13:1* 13.1 (Dec. 2023), pp. 1–14. ISSN: 2045-2322. DOI: 10.1038/s41598-023-37569-0. URL: https://www.nature.com/articles/s41598-023-37569-0.

[81] Dan Pan et al. “Deep learning for brain MRI confirms patterned pathological progression in Alzheimer’s disease”. In: *Advanced Science* 10.6 (Dec. 2023). ISSN: 21983844. DOI: 10.1002/ADVS.202204717.

[82] Shaker El-Sappagh et al. “Two-stage deep learning model for Alzheimer’s disease detection and prediction of the mild cognitive impairment time”. In: *Neural Computing and Applications* 34 (2022), pp. 14487–14509. ISSN: 14333058. DOI: 10.1007/s00521-022-07263-9. URL: https://link.springer.com/article/10.1007/s00521-022-07263-9.

[83] Alzheimer's Disease Education and Referral Center. "Alzheimer's disease fact sheet". In: *NIH Publication No. 08-6423* (2010). URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf.

[84] Alzheimer's Association. "Alzheimer's disease facts and figures". In: *Alzheimer's and Dementia* (2016). URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf.

[85] Frank Sposaro, Justin Danielson, and Gary Tyson. "iWander: an android application for dementia patients". In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE. 2010.

[86] Stanislaw Antol et al. "VQA: visual question answering". In: 2015. URL: https://arxiv.org/abs/1505.00468.

[87] Mengye Ren, Ryan Kiros, and Richard Zemel. "Exploring Models and Data for Image Question Answering". In: *ArXiv* (May 2015). URL: https://arxiv.org/abs/1505.02074.

[88] Qi Wu et al. "Visual question answering: a survey of methods and datasets". In: *Computer Vision and Image Understanding* (2016). URL: https://www.sciencedirect.com/science/article/abs/pii/S1077314217300772.

[89] Bolei Zhou et al. "Simple baseline for visual question answering". In: *ArXiv* (Dec. 2015). URL: https://arxiv.org/abs/1512.02167.

[90] Abhishek Das et al. "Visual dialog". In: 2017. URL: https://arxiv.org/abs/1611.08669.

[91] Peng Wang et al. "FVQA: fact-based visual question answering". In: Dec. 2016. URL: https://arxiv.org/abs/2303.10699.

[92] Issey Masuda Mora. "Open-ended visual question-answering". In: (2016). URL: https://arxiv.org/pdf/1610.02692.pdf.

[93] Paarth Neekhara. *Visual question answering in tensorFlow*. Nov. 2016. URL: https://github.com/paarthneekhara/neural-vqa-tensorflow.

[94] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *Computer Vision and Pattern Recognition* abs/1409.1556 (2014). URL: https://arxiv.org/abs/1409.1556.

[95] Abhishek Das et al. “Human attention in visual question answering: do humans and deep networks look at the same regions”. In: *Computer Vision and Pattern Recognition* (2016). URL: https://arxiv.org/abs/1606.03556.

[96] Maria Viggiano et al. “Visual recognition memory in Alzheimer’s disease: repetition-lag fffects”. In: *Experimental Aging Research* (2008).

[97] M Didic et al. “Impaired visual recognition memory predicts Alzheimer’s disease in amnestic mild cognitive impairment”. In: *Dementia and Geriatric Cognitive Disorders* (May 2013).

[98] Alice Cronin-Golomb, Suzanne Corkin, and John Growdon. “Visual dysfunction predicts cognitive deficits in Alzheimer’s disease”. In: *Optometry and Vision Science* 72.3 (Mar. 1995), pp. 168–176.

[99] Sven Joubert et al. “The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer’s disease”. In: *Neuropsychologia* 48.4 (Mar. 2010), pp. 978–988.

[100] Janet Duchek, David Balota, and Veronica Thessing. “Inhibition of visual and conceptual information during reading in healthy aging and Alzheimer’s disease”. In: *A Journal on Normal and Dysfunctional Development* 5.3 (Aug. 2010), pp. 169–181.

[101] Infermedica. *Artificial Intelligence for Healthcare Companies Infermedica*. http://www.infermedica.com/. 2017.

[102] *Wolfram Alpha APIs: Computational Knowledge Integration*. https://products.wolframalpha.com/api/.

[103] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[104] Kaiming He et al. “Deep residual learning for image recognition”. In: *Computer Vision and Pattern Recognition* (2016).

[105] Matthew Nancy et al. “Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review”. In: *JMIR Mhealth Uhealth* (Apr. 2016).

[106] S. Kevin Zhou, Hayit Greenspan, and Dinggang Shen. *Deep Learning for Medical Image Analysis*. London, United Kingdom: Elsevier Academic Press, 2017.

# VITA

Matthew Velazquez was born in Stuart, Florida, before moving to Kansas City, Kansas. He completed his undergraduate studies at Ottawa University in 2010 and further pursued an MBA from Texas A&M Commerce, which he concluded in 2013. In 2016, he took a new step in his academic journey by enrolling in the Ph.D. program at the University of Missouri-Kansas City. In the domain of Computer Science and Medical Bioinformatics, and under the guidance of Dr. Yugyung Lee, Matthew has been studying early prediction AI models for Alzheimer's disease.

During his time at UMKC, Matthew has also been associated with TIAA. He has learned and grown professionally in his roles, initially as a Senior Developer and more recently as a Senior Development Manager. This industry exposure has provided him with practical insights alongside his academic pursuits.

Matthew has had the opportunity to share some of his research work through six publications during his Ph.D. program. As he nears the end of this academic chapter, Matthew intends to continue his work at TIAA as a Senior Development Manager, hoping to apply and expand on the knowledge he has gained.