Identifying Key Brain Biomarkers for Alzheimer's Disease Progression and Stability: A Machine Learning Approach

***Abstract*—**The association of structural brain biomarkers with cognitive impairment in neurodegenerative diseases like Alzheimer's Disease (AD) is a significant research area. Identification of cognitive impairment is crucial to facilitate early intervention, but the conventional methods of diagnosis are certain to miss subtle neural alterations before the onset of the disease. The goal of this paper is to determine the optimal set of brain biomarkers to utilise to make an early diagnosis of Alzheimer's by leveraging the power of state-of-the-art machine learning techniques. From a set of neuroimaging-derived biomarkers and longitudinal cognitive data, we apply feature selection to rank useful predictors of cognitive impairment. The method combines conventional machine learning techniques, such as Random Forest and Support Vector Machines (SVM), with feature selection techniques such as SHAP values to identify linear and non-linear associations of biomarkers and cognitive status. Our results demonstrate that Random Forest models trained on SMOTE-balanced sets performed better in distinguishing patterns of disease progression correctly. SHAP analysis identified novel neuroanatomical characteristics, with models trained on the top 10 features enhancing pMCI classification by over 6 percentage points and sMCI classification by 8.5 percentage points compared to full-feature models. Our results validate some brain areas, including the Left Fusiform Gyrus, Right Middle Frontal Gyrus, and Parahippocampal Gyrus asymmetry, as strong biomarkers for disease course prediction. The output is mapped to conventional biomarker profiles for various levels of cognitive impairment and demonstrates their capability to predict earlier disease. The study highlights the significance of advancing computational methods into the clinic to provide improved diagnostic power and facilitate targeted interventions. With AI-based techniques, the study advances the field of computational neuroscience, providing an effective platform for early diagnosis of Alzheimer's and facilitating personalized medicine strategies.

***Index Terms*—Alzheimer’s Disease, Brain Biomarkers, Machine Learning, Cognitive Decline, Feature Selection, Neuroimaging**

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder, occurring in millions of people worldwide, mainly in elderly people. It causes a progressive decay in cognitive ability, affecting reasoning and quality of life. Early diagnosis of AD is crucial, as it allows timely treatment and the possible retardation of disease progression. Existing diagnostic protocols, however, are centered around the deposition of beta-amyloid plaques in the brain, identified largely by MRI and PET scanning. By the time these biomarkers are evident in significant concentrations, irreversible damage to the neurons is already underway, reducing the effectiveness of treatment regimes. The utilization of beta-amyloid deposition as a diagnostic marker limits the scope for early intervention, as patients are generally diagnosed at a stage when the disease is firmly established and predominantly irreversible. As such, better early detection methods are needed, which will focus on identifying markers of disease progression rather than terminal pathology.

To overcome and address this particular limitation, the present study is framed to find key MRI biomarkers in the brain that are essential in the progression of Alzheimer's disease (AD). Rather than merely indicating the occurrence of beta-amyloid deposits, our multidisciplinary approach is framed to explore longitudinal changes in brain structure, which will allow the detection of cognitive decline at an advanced stage at an early time point. With the use of advanced machine learning (ML) techniques, we aim to identify biomarkers that can distinguish between individuals who develop Alzheimer's disease (AD) and individuals with mild cognitive impairment (MCI) who are cognitively stable or even demonstrate improvement in cognitive functioning. Such a novel approach offers a proactive form of diagnosis, which will allow the implementation of early intervention approaches that have the potential to slow down or diminish the progression of the disease. Through the recognition of structural changes in the brain associated with cognitive decline, this study aims to improve the chances of early diagnosis and open the door to the development of specific therapeutic interventions that are specifically targeted.

In this study, we apply a machine learning method with a feature selection approach to discover and evaluate the most informative AD progression brain MRI biomarkers. We work with two data sets: (1) longitudinal changes in brain biomarkers throughout cognitive impairment across time and (2) baseline measurements of biomarkers from a large cohort of participants with varied cognitive status. We initially train classification models to identify the most important 10 brain biomarkers for each patient class according to longitudinal changes. The identified biomarkers are then used to fine-tune predictive models classifying people into groups such as Alzheimer's disease (AD), cognitively normal (CN), progressive mild cognitive impairment (pMCI), and stable mild cognitive impairment (sMCI). By systematic comparison of the performance of the various models concerning numerous subsets of biomarkers, we are now in a better position to compare the performance and efficacy of our selected features in effectively classifying the disease progression. Based on this set plan, our research is dedicated to the development of better methods that are aimed at early diagnosis of Alzheimer's disease, thereby contributing towards better clinical decision-making plans through better and more accurate methods.

.

2. Literature Review

Alzheimer's disease is a frequent and widespread neurodegenerative disorder that mainly affects individuals in their advanced age, i.e., elderly individuals. Alzheimer's disease induces profound modifications in various cognitive processes and generally occurs only after the onset of severe and incapacitating symptoms. In this particular stage of the disease, it is worth noting that a whole and guaranteed treatment is not currently available; however, the present treatments are dedicated to attempting to postpone or delay the progression of the disease using drug management strategies. In the last few years, because of the continuous advances of technology, both machine learning (ML) and deep learning (DL) strategies have emerged as hopeful prospects with a great ability to resolve complex medical issues and problems.

Goenka and Tiwari [1] used a blend of state-of-the-art techniques from deep learning, machine learning, and explainability techniques in their study. They were trying to detect beta-amyloid deposits in brain MRI scans accurately using different datasets on which they were working. On a different front, Wu et al. [12] proposed a new 3D deep learning model that is specifically fMRI dataset focused. The method provides much more detailed data than the conventional 2D images, thus improving the process of diagnosing Alzheimer's disease. Brain regions affected by the disease and their detection are indeed crucial, but it is to be noted that early-stage diagnosis is the key challenge. Early diagnosis is needed to facilitate early and effective intervention measures in patients.

A basic question in early detection is to identify the most suitable biomarkers. Barber [2] explained several medical tests to determine the biomarkers that can represent Alzheimer's disease at an early stage. With the advancement of technology, ML techniques have been increasingly used to optimize the selection of biomarkers. Kavitha et al. [4] compared the performance of various ML models to identify the most appropriate method for early detection. Subsequently, El-Sappagh et al. [5] incorporated explainability methods to better comprehend how ML models predict the risk of Alzheimer's, providing human-interpretable explanations for model predictions. Sayantan et al. [6] conducted a systematic review of models for Alzheimer's disease classification and identified the four most accurate models, which are also employed in this work.

Along with classifying at-risk patients, it is also important to determine specific attributes that play the most important role in improving prediction accuracy. Eke et al. [3] explored the connection between blood biomarkers and Alzheimer's disease using an extensive analysis of past patient data. Monitoring the progression of Alzheimer's disease has also gained greater academic attention. Shigemizu et al. [10] proposed a method to monitor disease progression from mild cognitive impairment (MCI) to Alzheimer's disease. Jung et al. [9] developed a deep recurrent neural network (RNN) model for the analysis of longitudinal data, emphasizing the necessity of monitoring disease progression over time, rather than using static data alone.

The choice of the right dataset forms a critical element in achieving high predictive accuracy for the development of Alzheimer's disease (AD). If the chosen data is not well correlated with the disease, the model's performance will be affected. Past studies have demonstrated the effectiveness of heterogeneous data modalities in predicting the onset of AD. For example, Altay et al. [15] used MRI images to detect structural brain changes and their correlation with cognitive decline, whereas Eke et al. [3] examined blood plasma biomarkers and their correlation with disease onset. These studies suggest the importance of choosing adequate biomarkers to produce effective predictive results.

Biomarker choice is another significant component of diagnosing Alzheimer's disease. According to El-Sappagh et al. [5], three types of biomarkers can be used to diagnose AD: genetic, cerebrospinal fluid (CSF), and other cognitive biomarkers. Their work confirmed that multimodal diagnosis based on all three types of biomarkers has the highest predictive value because each class of biomarkers carries different information about the progress of the disease. Additionally, some studies have highlighted the requirement for multimodal data fusion. Zhang et al. [9] investigated fusion of imaging, genetic, and clinical data and confirmed substantial improvement in classification performance when these disparate data sources are fused. Feature selection is also important for enhancing early diagnosis by picking the most important features, thus improving model interpretability and accuracy. Wang et al. [7] compared various feature selection methods, such as recursive feature elimination (RFE), mutual information, and SHAP values. The findings indicated that feature selection based on SHAP was the most effective in picking important class-specific biomarkers, resulting in more stable classification models. Following these findings, the current study employs a SHAP-based approach for class-wise feature selection to provide better model interpretability and performance.

3. DATASET DESCRIPTION

The data employed in the study is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a longitudinal research study that intends to clarify the progression of Alzheimer's disease (AD) through neuroimaging methods, clinical assessment, and biomarkers. Specifically, the metadata gathered from MRI scans within this data were first preprocessed and extracted in research work by Ledig et al. [16], who were interested in structural brain imaging of AD and mild cognitive impairment. Their work developed a collaborative morphometry database with the aim of motivating further research in this area and thus providing critical information regarding the structural brain modifications in AD.

The dataset contains 5,074 records, each characterizing a unique MRI scan obtained on an individual at some point in time. The data set has 315 variables: patient identifiers (e.g., RID, VISCODE), demographic variables (e.g., AGE.bl, AGE.scan), cognitive test scores (e.g., MMSE), and diagnostic variables (DX.bl, DX.scan). The data set also contains a set of volumetric measures derived from MRI, asymmetry measures, and structural biomarkers characterizing various regions of the brain. These measures are important indices of neurodegenerative processes and are very informative about patterns of brain atrophy during Alzheimer's disease.

The study employs MRI biomarkers to investigate longitudinal changes in cerebral structure. The longitudinal nature of the database allows for following the course of neurodegeneration over different intervals. The critical features are volumetric estimates of cerebral structures, asymmetry estimates that reflect inter-hemispheric variations, and ventricular estimates that reflect atrophy. By employing this rich metadata, the study seeks to enhance the understanding of the course of Alzheimer's disease and aid in the development of predictive models for its course. The full dataset can be accessed at:<https://gin.g-node.org/ledigchr/MALPEM_ADNI_data/src/master>.

4. Proposed Methodology

This research aims to identify the most important Alzheimer's disease (AD) progression biomarkers in the brain. Understanding the most appropriate biomarkers related to disease progression may allow enhancement of predictive models for early diagnosis, which increases clinical relevance. Toward that end, we employ a two-stage analysis strategy that utilizes two distinct datasets. The first dataset provides longitudinal variability in brain biomarker measurement, allowing us to examine patterns of progression over time. From this first dataset, we extract a subset of the most predictive biomarkers by evaluating feature importance. The second dataset provides baseline biomarker levels in a larger participant group, which we utilize to evaluate the classification utility of the identified features in distinguishing among different diagnostic classes. By comparing models trained from a smaller subset of important biomarkers to those trained on the full set of available features, we aim to determine whether smaller and more meaningful feature sets enhance classification performance and allow better understanding of AD progression.

The **Longitudinal Brain Biomarker Changes dataset** includes cross-sectional brain biomarker values recorded at different points in time and, therefore, allows for research on longitudinal change in cerebral structure. This dataset includes a series of quantitative variables describing different brain regions and corresponding biomarker values, such as cortical thickness measures, hippocampal volume measures, and other imaging-based measures of structure and function. These biomarkers are very helpful in understanding the temporal changes in different brain regions in Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) patients. The main variable in this dataset classifies participants into three categories based on cognitive progression status: (1) Positive Progress, including individuals who progressed from MCI to AD over time, reflecting disease progression; (2) No Progress, including individuals who did not change their cognitive status without deterioration or improvement; and (3) Negative Progress, including individuals who showed unexpected cognitive improvement, progressing from MCI to cognitively normal (CN) status or from AD to CN. This classification allows for careful examination and comparison of the patterns of biomarker changes in different disease courses.

To achieve the most informative biomarkers with regards to disease progression, a supervised classification model is learned from the longitudinal dataset. Various machine learning models such as Random Forest, Logistic Regression, Decision Trees and Support Vector Machines are utilized to achieve robust classification performance. The model is learned to classify the three progression classes, and feature importance scores are obtained to figure out the most important biomarkers responsible for each class's classification. Specifically, we rank the biomarkers in terms of separating progression states and select the 10 most important biomarkers of each class. The selected biomarkers are the features most associated with either worsening cognitive decline (Positive Progress), or cognitive stability (No Progress). The output of this process provides a condensed set of biomarkers hypothesized to be of critical relevance for predicting Alzheimer's disease progression.

The **Baseline Biomarkers and Diagnosis datase**t comprises single-time-point MRI-based biomarker measures from 1,069 participants. As compared to the initial dataset, the Baseline Biomarkers and Diagnosis dataset gives us a snapshot of the brain biomarker values at a single time point and how they compare with diagnostic groups. The dataset comprises only the top 10 biomarkers that were derived from the longitudinal dataset, thus enabling us to determine whether these chosen features remain predictive of discriminating between diagnostic groups. The target variable for this dataset classifies participants into four diagnostic groups: (1) Alzheimer's Disease (AD), which classifies individuals who are diagnosed with AD at assessment; (2) Cognitively Normal (CN), which classifies individuals who are not cognitively impaired and are healthy controls; (3) Progressive Mild Cognitive Impairment (pMCI), which classifies individuals who are diagnosed with MCI and eventually develop AD, the most critical group for early intervention strategies; and (4) Stable Mild Cognitive Impairment (sMCI), which classifies individuals with MCI who do not deteriorate over time. Through the utilization of this dataset, we will determine whether the most predictive biomarkers that were derived from the longitudinal dataset are capable of distinguishing between these diagnostic groups correctly.

To evaluate the usefulness of the identified biomarkers, classification models are trained on three feature sets: (1) the top ten most important biomarkers found under the Positive Progress category, and which are believed to be capable of predicting disease progression well; (2) the top ten leading biomarkers found under the No Progress category, and which are believed to detect features that represent cognitive stability; and (3) the entire set of 300 biomarkers, which represent the entire feature space. A Random Forest Classifier is built to classify these datasets, whose performance is determined using measures such as accuracy, F1-score, precision, and a confusion matrix, reporting the number of true positives across each class. If the pMCI classification accuracy is better on using the top 10 features from the Positive Progress model compared to using the full 300-biomarker model, it would indicate that the identified biomarkers are efficient in capturing disease progression patterns well. Similarly, if there is also an improvement in classification accuracy when using the top 10 biomarkers from the No Progress model for the sMCI category, it indicates that the selected features are indeed significant in detecting cognitive stability. This process allows us to check if a smaller subset of key biomarkers can maximize predictive performance and lead to an efficient diagnostic process.

By finding the most predictive biomarkers and evaluating their predictive value in baseline datasets, this study hopes to provide new evidence regarding the biological pathways of Alzheimer's disease development. The findings of this study have the potential to help clinical decision-making through improved early detection methods by using a refined set of biomarkers, allowing for the early implementation of interventions and on a targeted group of individuals who are likely to develop Alzheimer's disease. The identification of a small number of highly predictive biomarkers can also lead to the development of cost-effective and accessible diagnostic tools, eventually allowing for better patient outcomes and better precision medicine approaches for Alzheimer's disease.

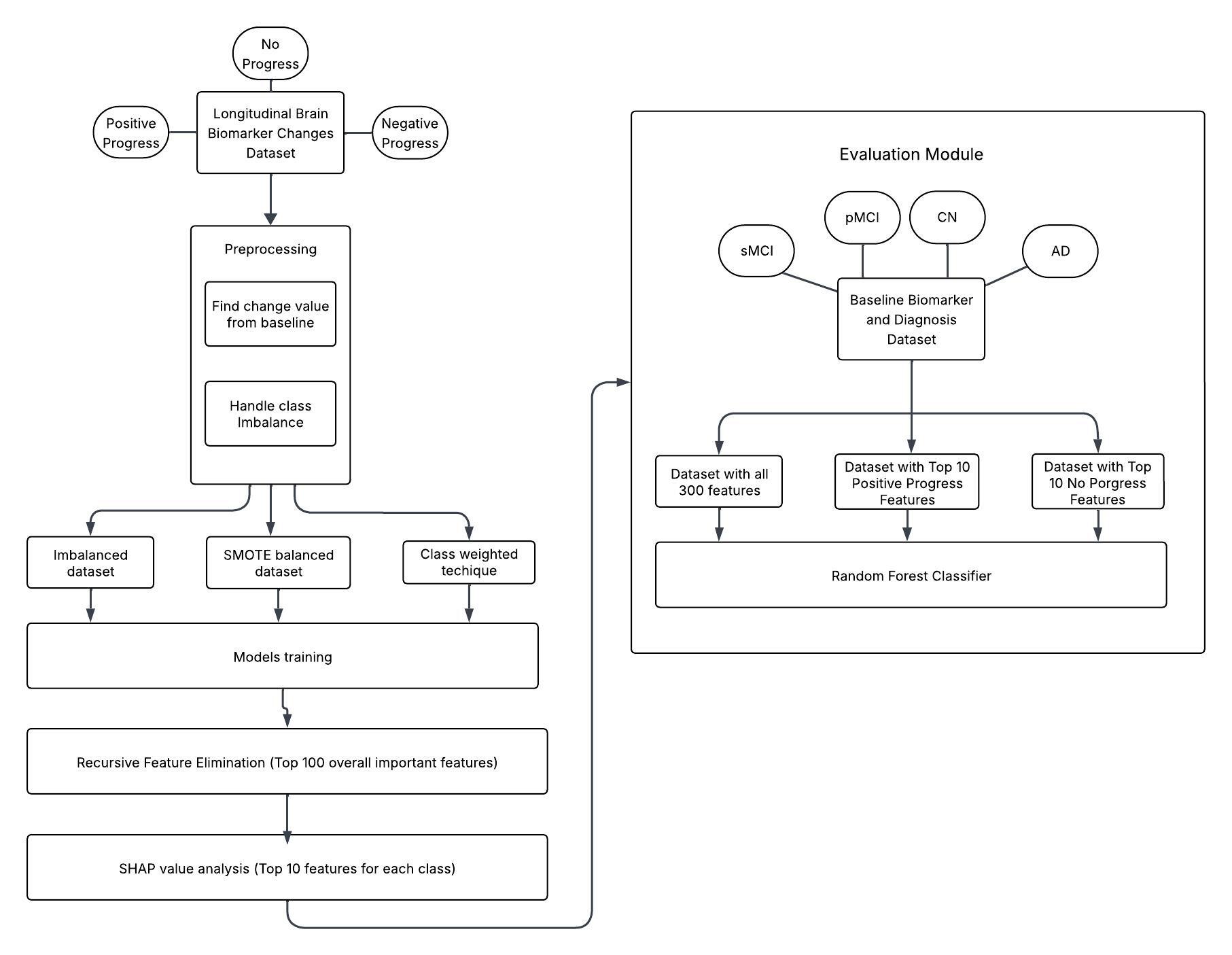


Fig. 1. Overall Proposed Architecture

5. IMPLEMENTATION

5.1. DATA PREPROCESSING

Data preprocessing is a fundamental part of any machine learning activity, particularly for medical data with complex and high-dimensional features. The quality and consistency of input data have a direct impact on the performance and interpretability of predictive models. In this paper, we employ a series of preprocessing methods with the aim of enhancing data credibility, correcting inconsistencies, and ensuring our model picks up on important patterns associated with the development of Alzheimer's disease. The preprocessing process includes data cleaning, feature engineering, and target labeling on the  **Longitudinal Brain Biomarker Changes dataset**. These processes lead to the removal of errors, normalization of values, and the creation of a credible target variable that best describes disease development or stabilization.

5.1.1 DATA CLEANING

Data cleaning is a key process for the elimination of inaccuracies that would compromise the integrity of our analysis. Missing values in biomarker data can result in bias and negatively impact model performance. Thus, the first process in our preprocessing pipeline is the identification and handling of missing values, the identification of anomalies, and the checking for consistency in biomarker measurements at several time points. In this research, the observations with missing (NaN) values and redundant and useless columns were found and removed from the dataset. Since the incidences of such occurrences were few, their removal did not affect the overall data distribution significantly. The data cleansing was performed to maintain only clean, complete, and high-quality records for model building and assessment.

5.1.2. FEATURE ENGINEERING – CHANGE CALCULATION AND NORMALIZATION

Feature engineering is the most important machine learning process that transforms raw data into more informative inputs to facilitate the learning of meaningful patterns by the model. It is particularly important in medical studies, as measurements for biomarkers among subjects can be different due to biological variation, imaging protocol, and other extraneous factors. By using relevant transformations, we can highlight those most meaningful features associated with the onset of the disease and also remove noise, as well as help in making models more interpretable. We utilized two primary feature engineering methods across this research:

Change Calculation: Since Alzheimer's disease is a progressive disorder, it is more informative to monitor longitudinal changes in biomarker levels than to look at cross-sectional measurements. To quantify longitudinal changes, we computed the difference between each biomarker's value at the current time and its baseline value. This transformation makes the model sensitive to changes from the baseline state rather than absolute biomarker values and, therefore, more sensitive to trends in disease progression.

Normalization: The calculated changes in biomarker values differed in both positive and negative directions. To introduce stability in model performance, we used feature scaling methods to normalize these values, scaling them to a common range. This minimizes the impact of extreme variations and allows models to generalize better across patients.

After these steps, the baseline biomarker rows (i.e., the first measurements before any change computed) were removed from the dataset to keep the dataset free of redundant rows.

5.1.3. TARGET LABELING

Precise labeling of the targets is a key component to ensure that the classification model will be accurate in recognizing disease progression patterns. With the consideration that the patients in the data have diverse patterns of cognitive worsening, stability, or improvement, proper labeling based on these patterns have been created. Well-formed target labels assist the model in differentiating the varying stages of progression of Alzheimer's disease and thereby enhance its prediction ability as well as its relevance to clinics.

In this study, the terminal target variable was allocated according to each patient's initial diagnosis and their most recent diagnosis found in the longitudinal dataset. Patients were divided into three different groups to illustrate various disease trajectories. The Positive Progress group consisted of individuals who moved from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD), indicating disease progression. The No Progress group consisted of individuals whose diagnostic status remained consistent over time, indicating cognitive stability. The Negative Progress group has individuals who experienced cognitive improvement, moving from either MCI to Cognitively Normal (CN) or from AD to CN. This classification system gives an overall representation of disease progression.

To understand the labeling process adopted, we have prepared the distribution of transitions in diagnoses as shown in Table-1. The table represents the frequency and categorization of each diagnostic pattern, which allows us to assess the percentage of individuals under each category.

By doing this systematic preprocessing, we build an accurate, homogeneous dataset that represents the progression of Alzheimer's disease. The following preprocessing steps form the foundation for creating robust prediction models that can distinguish strongly between different decline and stability patterns. The enhanced target labels not only enhance the accuracy of the classification but also the interpretability of the model, thereby allowing for more accurate and clinically meaningful investigation of the progression of the disease.

| **Baseline Diagnosis** | **Diagnosis during the scan** | **Number of DataPoints** | **Class** |
| --- | --- | --- | --- |
| AD | AD | 316 | No progress |
| AD | MCI | 3 | Negative Progress |
| CN | AD | 20 | Positive Progress |
| CN | CN | 889 | No Progress |
| CN | MCI | 100 | Positive Progress |
| EMCI | AD | 23 | Positive Progress |
| EMCI | CN | 22 | Negative Progress |
| EMCI | MCI | 354 | Positive Progress |
| LMCI | AD | 621 | Positive Progress |
| LMCI | CN | 43 | Negative Progress |
| LMCI | MCI | 1009 | Positive Progress |

TABLE-1 Diagnosis Transition and their Respective Class

| **Class** | **Number of Datapoints** |
| --- | --- |
| Positive Progress | 2127 |
| No Progress | 1205 |
| Negative Progress | 68 |

TABLE-2 Target Frequency Distribution

5.2. HANDLING CLASS IMBALANCE

Class imbalance is a common problem in medical data sets, in which some of the patient groups might be underrepresented compared to others. For classification problems, the existence of an imbalanced data set leads to biased model predictions since the algorithm will favor the majority class and will be unable to extract useful information from the minority classes. This consequently leads to low classification accuracy, particularly for the underrepresented categories. Additionally, an imbalanced data set can also undermine the analysis of feature importance since the model would overfit the majority class, hence undermining the validity of biomarkers that are discovered for disease development. Class imbalance must thus be handled in ensuring the model provides fair performance for all categories and provides clinically useful information.

The data employed in this study has a large class imbalance, as evident in Table-2. The Positive Progress class has 2,127 samples, the No Progress class has 1,205 samples, and the Negative Progress class has significantly fewer samples, with only 68 samples. The imbalance can significantly impact the model in detecting appropriate patterns in cognitive improvement (Negative Progress) and stability (No Progress). As a solution to this problem, we use two popular approaches: the Synthetic Minority Oversampling Technique (SMOTE) and Class Weighting. SMOTE generates synthetic samples to improve minority class representation, whereas Class Weighting tunes the model's learning process by assigning more importance to minority classes. To determine the best approach for our data, we train and evaluate models on three versions of the data: the original imbalanced data, the SMOTE-balanced data, and the Class-Weighted data. The best method in terms of classification accuracy is utilized for further feature importance analysis.

5.2.1. SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) is an oversampling approach specifically designed to handle class imbalance by creating synthetic minority instances. Compared to the traditional oversampling methods, where a copy of the samples is produced from earlier ones, SMOTE generates new samples by interpolating between the available instances. The process involves the selection of an instance from a minority class, the identification of its k-nearest neighbors, and the subsequent creation of a new synthetic instance on the connecting line of the objects. Employing the technique, SMOTE can ensure that the newly created instances reflect the structure and minority class distribution without suffering from the issue of redundant duplication that results in overfitting. This technique has seen widespread use in medical research to improve the effectiveness of classification in data sets that are dominated by highly imbalanced class distributions.

As the Negative Progress class was strongly underrepresented in our data, the Synthetic Minority Over-sampling Technique (SMOTE) was employed to create additional synthetic instances, hence achieving a balanced dataset. Nearest neighbor-based interpolation was utilized to fill up the synthetic samples without compromising on the underlying structure of the data. The utility of SMOTE was checked by comparing the Principal Component Analysis (PCA) plots of the data before and after oversampling (see Fig. 1 and Fig. 2), illustrating the minority class distribution enhancement after applying SMOTE and revealing that the created synthetic points lie close to existing data points. Models were trained using this balanced dataset after the oversampling, and the classification performance was checked to see whether the created samples were beneficial in model generalization.

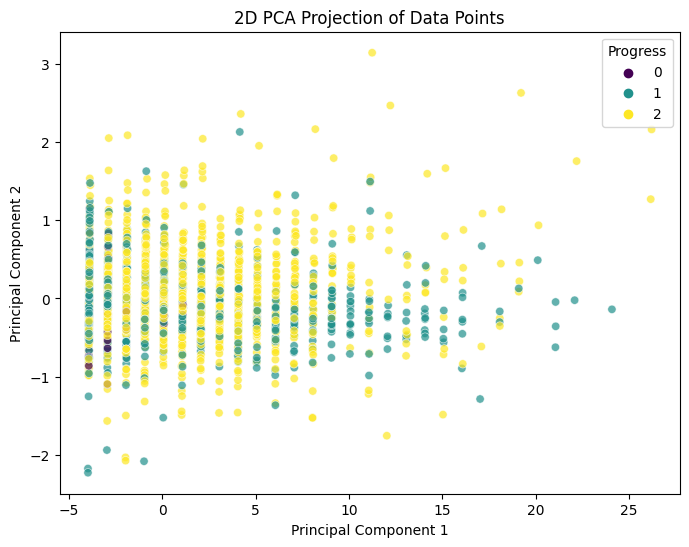


Fig. 2. PCA plot of the Data before oversampling using SMOTE

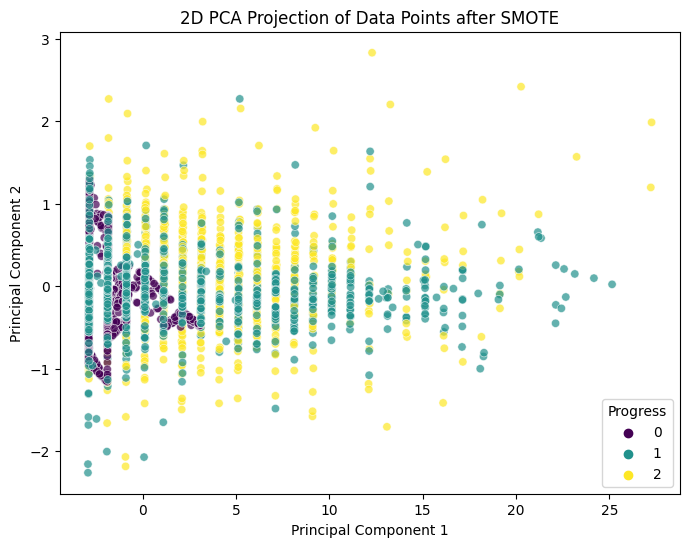


Fig. 3. PCA plot of the Data after oversampling using SMOTE

5.2.2. CLASS WEIGHTING TECHNIQUE

The Class Weighting Technique is another technique used to address class imbalance by varying the loss function during the model training process. In contrast to the oversampling of the minority class to elevate the number of samples, Class Weighting sets higher penalties on misclassifications that occur in underrepresented classes and lower penalties to the majority classes. Through the variation of the learning dynamics of the model, Class Weighting guarantees that minority classes influence the training process more strongly and therefore avoid majority class bias. This technique finds application with medical datasets, where the creation of synthetic data may lead to artificial variations that endanger the interpretability of clinical results.

Class Weighting was employed in the present work due to its ability to maintain the original distribution of the dataset without the introduction of artificial values. Class Weighting does not interfere with the integrity of real patient data like SMOTE, and it is promoted to be used in medical classification problems, as in the research work by Bakırarar, Batuhan & ELHAN, Atilla [8]. Class Weighting Technique was performed by calculation of weight given to a class via the Inverse Class Frequency Method shown in the following formula:

By this strategy, the Negative Progress category was given the highest weight to compel the model to concentrate more on appropriately classifying such underrepresented cases. Weighted loss functions were employed in training the models with the data, and classification accuracy assessment was done to see the impact of this adjustment.

To make a systematic comparison of the two methods, models were trained on three forms of the dataset: the original imbalanced dataset, which was used as the baseline for model performance evaluation; the SMOTE-balanced dataset, where synthetic oversampling methods were used to strengthen the representation of minority classes; and the Class-Weighted dataset, where the model was trained with weighted class weights to balance against imbalance without adding synthetic data. Classification performance on the three datasets was compared to determine which of the methods achieved the best accuracy and generalization for all classes. The method that had the most balanced and stable classification performance was used in the analysis of feature importance to ensure that the generated model best represents patterns of Alzheimer's disease development without class bias.

### 5.3. MODEL TRAINING

To set up a comprehensive paradigm of classification for the identification of the most important biomarkers associated with the development of Alzheimer's disease, we conducted a series of training sessions on various machine learning (ML) models against the already prepared datasets. From a study by Sayantan et al. [6], which compared the performance of various classifiers utilizing the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, we took four ML models that showed high performance in similar research. These models selected for this study were Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, and Random Forest Classifier. All these models were trained separately against three different datasets: the original imbalanced dataset, the SMOTE-balanced dataset, and the Class-Weighted dataset. Comparative performance of these models was checked to find the best methodological approach to achieve the maximum classification accuracy and generalizability for all classes.

### 5.3.1. LOGISTIC REGRESSION

Logistic Regression (LR) is a popular statistical classification method that predicts the likelihood of an instance being in a given class using the sigmoid function. The model is especially useful for binary and multi-class classification and can be utilized for distinguishing between many progression pathways of Alzheimer's disease. In the present research, L2 regularization (or Ridge Regression) was employed to fight overfitting and stabilize the model.

### 5.3.2. SUPPORT VECTOR MACHINE (SVM)

The Support Vector Machine (SVM) classifier is a robust supervised learning algorithm that designs a hyperplane to separate data points of two classes. It attempts to maximize the margin between the two closest data points (support vectors) of each class so that it can generalize best. To handle non-linearly separable data, the radial basis function (RBF) kernel was employed in this study so that SVM can map input features into a higher-dimensional space where linear separation is possible. SVM has been a prevalent algorithm in medical studies due to its capability of handling high-dimensional and small-sample-size data, and hence it is a suitable method for Alzheimer's progression classification.

### 5.3.3. DECISION TREE CLASSIFIER

Decision Tree (DT) Classifiers work in a tree model to make decisions by recursively partitioning the data based on feature values. At the node, the algorithm chooses the feature yielding the maximum information gain, computed based on the Gini impurity or entropy criterion. Tree generation is a recursive partitioning strategy where instances are divided into subsets based on conditions to reduce the classification uncertainty. Decision Trees are used because of their interpretability and the capability to process non-linear feature relations.

### 5.3.4. RANDOM FOREST CLASSIFIER

The Random Forest (RF) Classifier is an ensemble learning algorithm that builds an ensemble of multiple Decision Trees and combines their predictions to improve classification accuracy. The individual tree in the ensemble is trained on a bootstrap sample of the dataset using bootstrap sampling (also referred to as Bagging), and further, it also considers a random subset of features at each split, which avoids overfitting and improves generalization. The final classification decision is made based on a majority voting scheme, in which the class with the most frequent predictions among the individual trees is chosen. Random Forest is specifically well-suited to handle high-dimensional data, and hence it is a good fit for this research project.

### 5.3.5. EVALUATION

The performance of the trained models was evaluated by using 10-fold cross-validation (10-fold CV) as the primary evaluation tool. Cross-validation is a resampling technique where the given dataset is split into k subsets of the dataset, and the model is trained on (k-1) subsets of the dataset and tested on one subset in one round. The process is repeated 10 times, where in each iteration a different subset is taken as the test set. The final classification performance is calculated across all the folds to obtain a good approximation of the generalizability of the model.

The 10-fold cross-validation method was chosen since it can test models against multiple partitions of data, thereby reducing the possibility of overfitting and improving the robustness of performance measurements. Since each fold is used as a test set only once, this method ensures that the model is exposed to multiple validation samples, thereby improving its robustness. Additionally, studies by Eke et al. [3] have demonstrated that 10-fold cross-validation offers an excellent trade-off between performance reliability and computational efficiency, thereby being one of the most widely used validation methods in medical data analysis. Through the application of this evaluation method, we investigate which of the trained classifiers consistently show stable performance on all Alzheimer's progression classifications, thereby allowing for unbiased measurement of model performance.

### 5.4. FEATURE IMPORTANCE ANALYSIS

Identification of the most influential biomarkers is required to explain the mechanisms of Alzheimer's disease development and build predictive models to be applied to support early diagnosis. Feature importance analysis enables us to identify the most influential biomarkers that have the greatest impact on disease classification, providing a reference point for future studies of Alzheimer's disease development. For this purpose, a two-stage feature selection procedure was implemented: (1) Recursive Feature Elimination (RFE) to select the top 100 most influential features accounting for overall model performance and (2) SHapley Additive exPlanations (SHAP) analysis to select the top 10 features for each class concerning their contribution to classification results. The step-wise strategy preserves model interpretability and dimensionality reduction with improved computational efficiency without sacrificing essential information.

#### Phase 1: Recursive Feature Elimination (RFE) for Dimensionality Reduction

The first step of feature importance evaluation utilises Recursive Feature Elimination (RFE), a wrapper feature selection technique that sequentially removes less important features to enhance the accuracy of the model. RFE works by fitting a machine learning algorithm—a Random Forest classifier—over the data and finding feature importance scores. The process uses an iterative approach in which the model is trained using all the available features, the feature considered the least important is removed, and the loop is executed until the most important features remain. The importance of features is determined through the model's inherent ranking scores, i.e., Gini importance used in decision trees.

Mathematically, RFE chooses features with optimal selection by minimizing model error by backward elimination, leaving behind the features that have a significant impact on the classification result. This minimizes dimensionality from the original 300 biomarkers into the most relevant 100 features, improving model efficiency and interpretability while retaining the most predictive variables. In addition, reducing the features makes it computationally feasible to produce SHAP values, allowing more precise analysis in the next phase.

#### Phase 2: SHAP Analysis for Class-Specific Feature Importance

After obtaining the top 100 features from RFE, the next step is to identify the most impactful features per individual class from SHapley Additive exPlanations (SHAP) values. SHAP is a game-theoretic approach to approximat the effect of every feature on a model's predictions by distributing prediction differences to all possible feature subsets. In contrast to the traditional feature importance approaches, SHAP provides global and local interpretability, and therefore, we are able to identify features that are most impactful for specific class categories rather than the overall model performance. Through SHAP, we identify the top 10 most impactful features for each class (Positive Progress, No Progress, and Negative Progress) by identifying features with the highest SHAP values. This step allows us to identify biomarkers that are most associated with specific disease courses, so we are able to build predictive models aimed at them.

5.5. EVALUATION OF SELECTED FEATURES

The selected feature evaluation is a crucial part of this research, and the selected biomarkers are ensured to be accurate and effective in their portrayal of Alzheimer's disease progression. While the feature significance analysis provides a ranked list of the most important biomarkers, they are to be evaluated for prediction in an independent classification model. This evaluation ensures that a reduced subset of the features is able to attain equivalent or better classification accuracy than the full feature set, thus enhancing interpretability and reducing computational complexity. As long as the selected biomarkers are able to provide a good representation of disease stability and progression, they can serve as a clinically relevant and useful set of predictors for the monitoring of Alzheimer's disease.

We train a Random Forest classifier model on the Baseline Biomarkers and Diagnosis dataset, which classifies subjects into four classes: Alzheimer's Disease (AD), Cognitively Normal (CN), Progressive Mild Cognitive Impairment (pMCI), and Stable Mild Cognitive Impairment (sMCI) to evaluate the predictive ability of the selected features. To compare the effect of feature selection systematically, the dataset is preprocessed in three ways for comparison. First, a model is trained on the full feature set, consisting of the full 300 original biomarkers as input features, as a baseline for comparison of performance. Second, a model is trained on the top 10 biomarkers selected for the Positive Progress class, which are most likely to be most important for predicting disease progression. Third, a third model is trained on the top 10 biomarkers selected for the No Progress class, taking into account features that correspond to cognitive stability. By assessing classification performance across these heterogeneous feature sets, we hope to assess whether a reduced, clinically interpretable subset of biomarkers can have comparable or superior predictive performance in the prediction of disease progression and stability.

The performance of the selected features is evaluated on the basis of classification accuracy between the pMCI and sMCI classes. If higher classification accuracy among pMCI cases for the Positive Progress feature set is noted compared to the full feature set, it validates that such biomarkers best capture disease progress. Likewise, if the No Progress feature set model enhances sMCI cases' classification accuracy, then it validates that these biomarkers are reliable predictors for cognitive stability. By comparing performance between such models, we establish the diagnostic utility and prediction power of selected biomarkers to validate that these are reliable predictors for Alzheimer's disease progress and stability.

6. RESULTS AND DISCUSSION

The detailed evaluation of the classification models developed to analyse the progression of Alzheimer's disease progression, with focus on model performance and feature importance analysis are discussed below. To offer accurate and unbiased predictions, several methods of handling class imbalance were explored, including Synthetic Minority Oversampling Technique (SMOTE) and class weight balancing to find the technique best suited for the Longitudinal Brain Biomarker Changes dataset. The classification models were compared using their mean 10-fold cross-validation scores to find the best method that is best for detecting disease progression patterns. Feature importance analysis was also conducted to find the most important biomarkers that influence disease progression. The results of these analyses offer critical information on the best classification model with the best class imbalance handling method, as well as the most important biomarkers for forecasting Alzheimer's disease progression and stablity. An evaluation of the selected features was also conducted to validate their performance in improving classification accuracy, ensuring that the extracted features offer meaningful biological significance in disease progression and stability knowledge.

6.1. PROGRESSION CLASSIFICATION MODEL ANALYSIS WITH VARIOUS CLASS IMBALANCE HANDLING TECHNIQUES

To find the optimal classification model for disease progression prediction, training was conducted on three datasets: the original unbalanced dataset, the SMOTE-balanced dataset, and the class weight-balanced dataset. The mean 10-fold cross-validation (CV) scores of each classification model on these datasets are reported in Table-3, Table-4, and Table-5, respectively. The aim of this study was to identify the model configuration that yielded the highest classification accuracy, which would be used to select the top 10 important biomarkers for each class using SHAP values.

| **Model** | **With all 300 features** | **With top 100 features** |
| --- | --- | --- |
| Random Forest | 0.6334 | 0.6564 |
| Logistic Regression | 0.5818 | 0.6071 |
| SVM | 0.6231 | 0.6236 |
| Decision Tree | 0.5980 | 0.5832 |

TABLE-3 Mean 10-fold cross-validation score of the initial unbalanced dataset

As can be observed in Table-3, giving the overall performance for the original unbalanced data set, the Random Forest model had the best performance with a mean cross-validation score of 0.6334 when the model used all 300 features and 0.6564 when using the 100 top features. The mean 10 fold cross-validation scores of the Logistic Regression and SVM models were lower, whereas the Decision Tree model gave the poorest effective performance.

| **Model** | **With all 300 features** | **With top 100 features** |
| --- | --- | --- |
| Random Forest | 0.8360 | 0.8513 |
| Logistic Regression | 0.6770 | 0.6508 |
| SVM | 0.5182 | 0.5176 |
| Decision Tree | 0.7658 | 0.7704 |

TABLE-4 Mean 10-fold cross-validation score of the dataset balanced with SMOTE

Table-4 demonstrates the performance after using SMOTE to balance the data. Here, the Random Forest model performed best, with a mean cross-validation equal to 0.8360 when all 300 features were used and 0.8513 when the top 100 features were employed. The Decision Tree model also improved; however, the SVM performed relatively low rates of performance.

| **Model** | **With all 300 features** | **With top 100 features** |
| --- | --- | --- |
| Random Forest | 0.6310 | 0.6384 |
| Logistic Regression | 0.4698 | 0.4521 |
| SVM | 0.3984 | 0.3976 |
| Decision Tree | 0.5812 | 0.6006 |

TABLE-5 Mean 10-fold cross-validation score of the dataset balanced with Class Weights

On the other hand, Table-5 shows the results regarding class weighting technique. The overall performance of most of the models was considerably worse compared to SMOTE. The Random Forest model had an average cross-validation measure of 0.6310 when using all 300 features and 0.6384 when using the top 100 features. This confirms class weighting was worse compared to SMOTE in handling class imbalance in this specific dataset.

Based on the above analysis, we conclude that SMOTE was the best method for addressing class imbalance in the dataset because it resulted in the highest classification accuracy for all models. Additionally, of the models experimented with, Random Forest performed better consistently than others, and its performance was significantly improved when the model was trained on the SMOTE-balanced dataset. Moreover, using the top 100 features resulted in a slight increase in classification accuracy, which implies that dimension reduction through feature selection improves model performance. Thus, the Random Forest model trained on SMOTE-balanced data and the top 100 features was chosen as the best setup for future feature importance analysis using SHAP values.

6.2. SHAP VALUE ANALYSIS FOR TOP 10 FEATURES OF POSITIVE AND NO PROGRESS CLASSES

Shapley Additive exPlanations (SHAP) values form a solid foundation for calculating individual feature contribution to model prediction by class. In our study, this method is especially valuable since it enables us to define the neuroanatomical areas and cognitive markers most strongly linked to disease course in Alzheimer's. By identifying features with the greatest effect on model output for the Positive Progress and No Progress classes, we are more able to elucidate the underlying biological drivers of disease course. Not only does this analysis facilitate model interpretability but also creates clinically relevant results that ultimately might be used to inform early intervention and treatment planning.

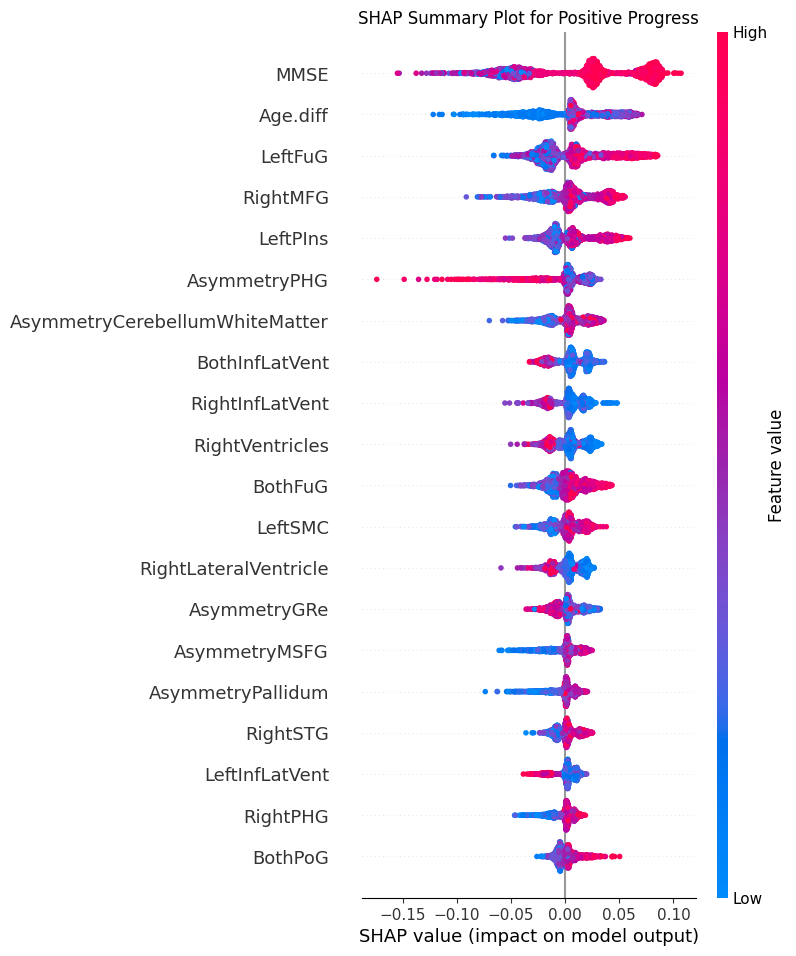


Fig.4. Top 10 features contributing to the class “Positive Progress”

The summary plot for Positive Progress class (Fig. 4) shows several features that have a significant impact on model predictions. The MMSE score has the largest effect, with lower values (blue) having a high probability of positive disease progression. Of the neuroanatomical features, Left Fusiform Gyrus (LeftFuG) and Right Middle Frontal Gyrus (RightMFG) have extremely strong effects, with larger feature values (red) having a tendency to confirm the predictions of disease progression. Moreover, the asymmetry of the Parahippocampal Gyrus (AsymmetryPHG) is also a strong predictor, suggesting that hemispheric asymmetry in this region can predict disease progression.

The most important features found for the Positive Progress class correspond to known neurobiological patterns in Alzheimer's disease progression. MMSE scores, as a direct indicator of cognitive capacity, not surprisingly exhibit high correlation with disease progression when scores are low. Activation of the Left Fusiform Gyrus corroborates earlier findings associating this area with visual recognition impairment in early Alzheimer's. Likewise, variation in the Right Middle Frontal Gyrus corresponds to executive function impairment typical of disease progression. The salience of asymmetry measures, especially in the Parahippocampal Gyrus, corroborates mounting evidence that asymmetric degeneration is a primary marker of Alzheimer's pathology. The Left Lateral Ventricle enlargement, another highest feature, corresponds to the ventricular enlargement that is typically seen as brain atrophy increases.

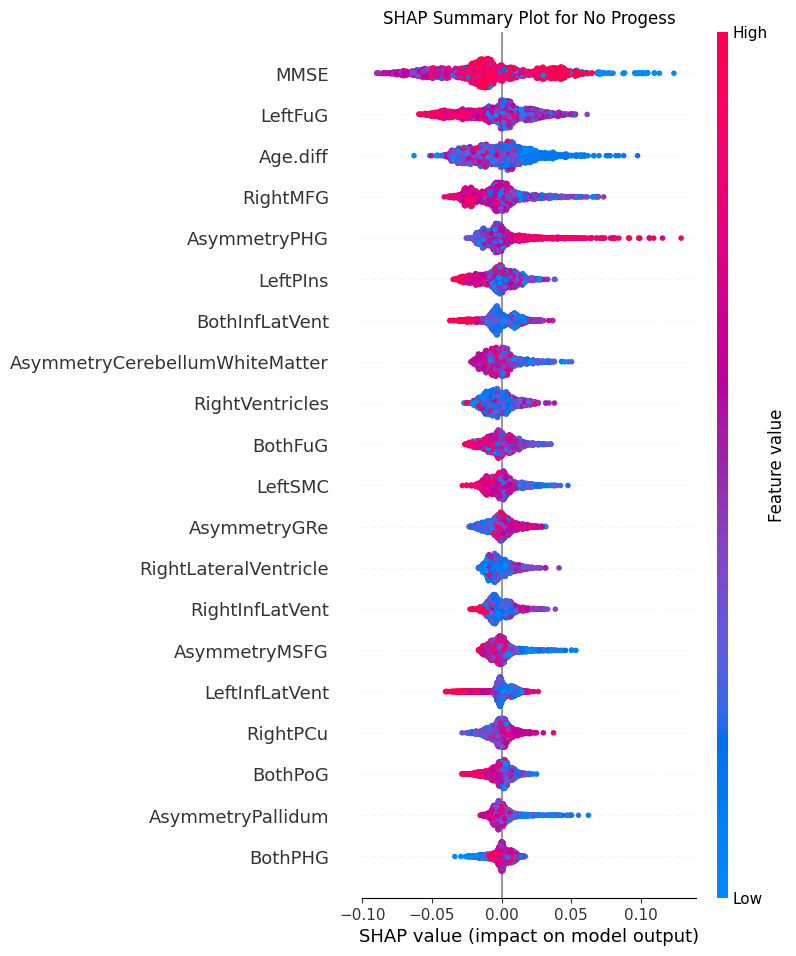


Fig. 5. Top 10 features contributing to the class “No Progress”

For the No Progress class (Fig. 5), the MMSE once more is the top predictor, but with a negative pattern wherein higher values (red) are good predictors of stability of disease. The Left Fusiform Gyrus (LeftFuG) has a strong pattern wherein lower values are good predictors of stability of disease. Age difference has a complex pattern, but generally lower values are good predictors of No Progress. Right Middle Frontal Gyrus (RightMFG) and Left Pins have moderate effects with some patterns of contribution depending on their values. Some asymmetry measures, such as AsymmetryPHG and AsymmetryCerebellumWhiteMatter, are among the top features, which implies that balanced preservation of these structures is crucial for stability of disease.

The features recognized for the No Progress category produce interesting data about factors concerning stability of the disease. The positive impact of high MMSE scores is in line with clinical findings suggesting that preservation of cognitive function is connected with retardation of disease progress. The use of features concerning ventricles (RightVentricles, RightLateralVentricle) suggests that preserved ventricular volumes might represent the absence of atrophy characteristic of progressive stages of the disease. The occurrence of multiple bilateral features (BothFuG, BothInflatVent, BothPoG, BothPHG) among primary predictors of stability suggests the possible use of preserved bilateral symmetry in the preservation of cognitive functions. The impact of Left SMC and Asymmetry features also suggests that the preservation of specific regional patterns and equal hemispheric relations might characterize stable stages of the disease in Alzheimer's.

6.3. EVALUATION OF FEATURES SELECTED

To verify the clinical relevance and predictive validity of the features obtained by SHAP value analysis, we conducted a validation experiment on the Baseline Biomarkers and Diagnosis dataset. The goal of this validation approach is to determine if the chosen features indeed capture future patterns of disease progression from baseline data alone. We trained three different classification models: (1) a baseline model that used all 300 available features, (2) a focused model that used only the top 10 features of Positive Progress class, and (3) a focused model that used only the top 10 features of No Progress class. All the models were trained to predict four diagnostic classes (AD, sMCI, pMCI, and CN) and were tested with per-class classification accuracy.

| **Class** | **Per-class Classification Accuracy** | | |
| --- | --- | --- | --- |
| **With all 300 features** | **Top 10 features for “Positive Progress” class** | **Top 10 features for “No Progress” class** |
| AD | 0.8415 | 0.8361 | 0.7027 |
| sMCI | 0.6790 | 0.7586 | 0.7640 |
| pMCI | 0.7527 | 0.8182 | 0.7375 |
| CN | 0.8676 | 0.7841 | 0.4321 |

TABLE 6 Per-class Classification Accuracy

The results reported in Table-6 give us a glimpse into the success of our discrimination power-based feature selection algorithm. Of particular interest is the model that was trained on the top 10 features for Positive Progress, which achieved a much higher classification accuracy for progressive MCI (pMCI) patients, at 81.82%, compared to the model that used all 300 features, with the accuracy being 75.27%. This more than 6 percentage point improvement indicates that the selected features capture significant patterns uniquely representative of the progression of the disease. While the accuracy levels for Alzheimer's Disease (AD) cases were comparatively close (83.61% vs. 84.15%), the accuracy for Cognitively Normal (CN) subjects did fall slightly (78.41% vs. 86.76%). However, the enhanced ability to correctly classify pMCI cases is of particular clinical significance. The confusion matrices attest to this, as seen in Fig. 7, showing a significantly higher number of true positives for pMCI compared to the all-features model shown in Fig. 6, and hence strongly indicating that our selected features capture the neurobiological markers of disease progression well.

In the same vein, we also trained a model with the top 10 features for No Progress, which performed better in the identification of stable MCI (sMCI) cases, with an accuracy rate of 76.40%, against 67.90% for the all-features model. This is an improvement of approximately 8.5 percentage points in the stability classification per se. Although this model performed worse in AD and CN classification, the improved capability to detect stable MCI cases with accuracy has most critical implications for treatment intensity and frequency of patient monitoring decisions within the clinic. The comparison of confusion matrices from Fig. 6 and Fig. 8 also supports this trend, with the No Progress features model having a significantly higher number of true positives for sMCI than the full model. This finding strongly validates our approach to feature selection in the identification of the brain MRI biomarkers most pertinent to disease stability, thereby enabling more accurate predictions with significantly fewer variables.

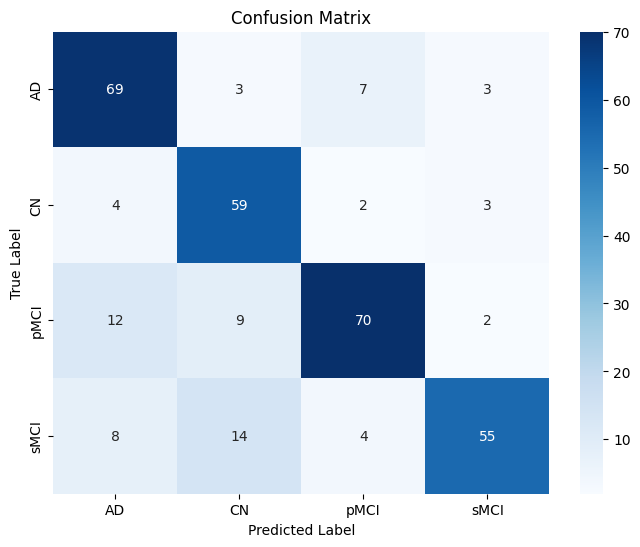


Fig. 6. Confusion matrix of model trained with all 300 features

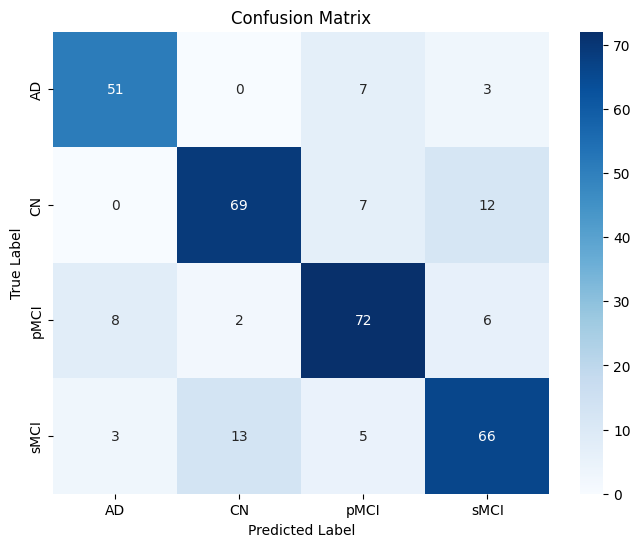


Fig.7. Confusion matrix of Positive Progress feature model

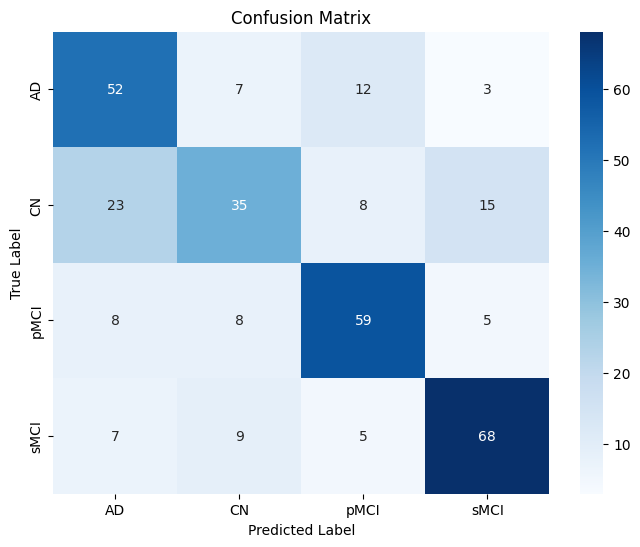


Fig. 8. Confusion matrix of No Progress feature Model

7. CONCLUSION

In this paper, we effectively developed and validated a machine learning model that predicts Alzheimer's disease progression based on neuroimaging biomarkers and cognitive test scores. Our meticulous analysis verifies that Random Forest models trained on SMOTE-balanced data and dimensionality reduction are superior in distinguishing between disease progression patterns. By examining SHAP values, we revealed characteristic sets of brain MRI biomarkers uniquely linked to disease progression and stability. MMSE score, Left Fusiform Gyrus, Right Middle Frontal Gyrus, and Parahippocampal Gyrus asymmetry were identified as robust predictors of disease progression, whereas ventricle-related features and bilateral symmetry markers robustly indicated disease stability. The validation experiment with the top 10 features per class confirmed their clinical significance, with Positive Progress features boosting pMCI classification accuracy by more than 6 percentage points (81.82% vs. 75.27%) and No Progress features enhancing sMCI classification by about 8.5 percentage points (76.40% vs. 67.90%) compared with the use of all 300 features. These results not only verify the efficacy of our feature selection strategy but also offer useful insights into the neurobiological underpinnings of Alzheimer's disease course, potentially informing targeted intervention strategies and personalized patient monitoring protocols in clinical settings.

8. REFERENCES:

[1] Goenka, N., Tiwari, S. Deep learning for Alzheimer prediction using brain biomarkers. *Artif Intell Rev* **54**, 4827–4871 (2021).

[2] Barber, Robert C. "Biomarkers for Early Detection of Alzheimer Disease" *Journal of Osteopathic Medicine*, vol. 110, no. s98, 2010, pp. 10-15.

[3] C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson and E. Ifeachor, "Early Detection of Alzheimer's Disease with Blood Plasma Proteins Using Support Vector Machines," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 218-226, Jan. 2021,

[4] Kavitha C, Mani V, Srividhya SR, Khalaf OI and Tavera Romero CA (2022) Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. *Front. Public Health* 10:853294.

[5] El-Sappagh, S., Alonso, J.M., Islam, S.M.R. *et al.* A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer’s disease. *Sci Rep* **11**, 2660 (2021).

[6] Sayantan Kumar, Inez Oh, Suzanne Schindler, Albert M Lai, Philip R O Payne, Aditi Gupta, Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review, *JAMIA Open*, Volume 4, Issue 3, July 2021.

[7] Wang, H., Liang, Q., Hancock, J.T. *et al.* Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *J Big Data* **11**, 44 (2024).

[8] Bakırarar, Batuhan & ELHAN, Atilla. (2023). Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research. Turkiye Klinikleri Journal of Biostatistics. 15. 19-29. 10.5336/biostatic.2022-93961.

[9] Wonsik Jung, Eunji Jun, Heung-Il Suk, Deep recurrent model for individualized prediction of Alzheimer’s disease progression, NeuroImage, Volume 237,2021.

[10] Shigemizu, D., Akiyama, S., Higaki, S. *et al.* Prognosis prediction model for conversion from mild cognitive impairment to Alzheimer’s disease created by integrative analysis of multi-omics data. *Alz Res Therapy* **12**, 145 (2020).

[11] Katabathula, S., Wang, Q. & Xu, R. Predict Alzheimer’s disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations. *Alz Res Therapy* **13**, 104 (2021).

[12] Wu, Jianfeng, et al. "Improved Prediction of Beta-Amyloid and Tau Burden Using Hippocampal Surface Multivariate Morphometry Statistics and Sparse Coding." *arXiv preprint arXiv:2211.05235* (2022).

[13] Dong, Mengjin, et al. "DeepAtrophy: Teaching a Neural Network to Differentiate Progressive Changes from Noise on Longitudinal MRI in Alzheimer's Disease." *arXiv preprint arXiv:2010.12948* (2020).

[14] Feng, X., Provenzano, F.A., Small, S.A. *et al.* A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer’s disease. *Alz Res Therapy* **14**, 45 (2022).

[15] Altay, Fatih, et al. "Preclinical stage Alzheimer's disease detection using magnetic resonance image scans." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 17. 2021.

[16] Ledig, C., Schuh, A., Guerrero, R. *et al.* Structural brain imaging in Alzheimer’s disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Sci Rep* **8**, 11258 (2018).