



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Linear and Quadratic Discriminant Analysis of Alzheimer's and Non-Alzheimer's Patients Based on Lifestyle Factors, Clinical Measurements, and Cognitive and Functional Assessments

In partial fulfillment of the requirements in
STAT 20253: Multivariate Analysis

By

Almer John Sta. Ines

Dencie Mae Saguano

Jerolle Nonato

Joy Ellen Mae Yangyang

Juan Raphael Pimentel

Roldan Libay Jr.

2025



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

TABLE OF CONTENTS

1 CHAPTER I

THE PROBLEM AND ITS SETTING	1
1.1 Introduction	1
1.2 Objectives	3
1.3 Statements of the Problem	3
1.4 Significance of the Study	4
1.5 Scopes and Limitations	5
1.6 Defintion of Terms	6

2 CHAPTER II

REVIEW OF LITERATURE AND STUDIES	7
2.1 Factors Influencing Alzheimer's Disease	7
2.1.1 Lifestyle Factors	7
2.1.2 Clinical Measurements	7
2.1.3 Cognitive and Functional Assessment	8
2.2 Detection of Alzheimer's Disease using Discriminant Models	9
2.2.1 Linear Discriminant Analysis (LDA)	10
2.2.2 Quadratic Discriminant Analysis (QDA)	11
2.3 Feature Selection Techniques in Discriminant Analysis	12
2.3.1 Recursive Feature Elimination (RFE)	12
2.3.2 Lasso Regression	12
2.4 Synthesis	13



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

3 CHAPTER III

METHODOLOGY	14
3.1 Method of Research	14
3.2 Data and Variables	14
3.3 Data Preprocessing	15
3.4 Statistical Treatment of Data	16
3.4.1 Recursive Feature Elimination	16
3.4.2 Lasso Regression	16
3.4.3 Linear Discriminant Analysis	17
3.4.4 Quadratic Discriminant Analysis	17
3.4.5 Model Evaluation	17

4 CHAPTER IV

RESULTS AND DISCUSSION	19
4.1 Exploratory Data Analysis of Alzheimer's Patients and Non-Alzheimer's	19
4.1.1 Assumptions of Discriminant Analysis	19
4.1.2 Lifestyle Factor Distributions of Patients	22
4.1.3 Clinical Measurements of Patients	23
4.1.4 Cognitive and Functional Assessments of Patients	24
4.2 Discriminant Modeling of AD Patients	24
4.2.1 Complete Linear and Quadratic Discriminant Models	24
4.2.2 RFE-Reduced Linear and Quadratic Discriminant Models	27
4.2.3 Lasso Regression-Reduced Linear and Quadratic Discriminant Models	30
4.3 Classification Metrics Comparisons of LDA and QDA Models	33



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

5 CHAPTER V

CONCLUSION AND RECOMMENDATION	34
5.1 Summary of Findings	34
5.2 Recommendations	34



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

CHAPTER I THE PROBLEM AND ITS SETTING

This chapter contains the introduction, statements of the problem, research hypothesis, significance of the study, scope and limitations, and definition of terms.

Introduction

Every three seconds, someone in the world develops dementia, adding to the more than 55 million people already living with the condition as of 2020 (Alzheimers Disease International, n.d.). This number is expected to nearly double every 20 years, reaching 78 million in 2030 and a staggering 139 million by 2050. Alzheimer's disease (AD) the most common cause of dementia, is a progressive brain disorder that gradually affects memory, thinking skills, and the ability to carry out everyday tasks (Alzheimer's Association, 2021). As the prevalence of AD continues to rise, it affects people not only altering their lives but also placing an emotional and financial strain on families and healthcare system. Early detection is crucial as it allows for timely interventions, potential treatments, and better supportfor both patients and caregivers (Dubois et al., 2016). However, diagnosing AD in its early stages remains challenging, as its symptoms often overlap with other cognitive conditions, making it difficult to distinguish from normal aging or mild cognitive impairment (Jack et al., 2018). Understanding the key indicators of AD can help improve early diagnosis and lead to better patient outcomes.

A study by Livingston et al. (2020) suggests that lifestyle factors, medical history, clinical measurements, and cognitive and functional assessments can provide essential insights into the early idenitifaction of Alzheimer's disease. Lifestyle factors,



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

such as Body Mass Index (BMI), smoking status, alcohol consumption, physical activity, diet quality, and sleep quality can impact the risk of developing AD (National Institutes of Health, n.d.). Additionally, certain medical conditions, including hypertension, diabetes, cardiovascular diseases, depression, family history of AD, and head injury have been linked to increased susceptibility to AD (Kivipelto et al., 2018). Clinical measurements, particularly cardiovascular health indicators, have been increasingly recognized as significant risk factors for the disease. Moreover, a study conducted by Leszek et al. (2020) suggests that cardiovascular diseases can contribute to the development and progression of AD by impairing blood flow to the brain and causing inflammation (Leszek et al., 2021). Functional assessments, which evaluate how well a person can manage daily tasks and adjust to cognitive changes, can also be an important tool in detecting early signs of AD. (Sperling et al., 2011)

The effectiveness of classification models in predicting AD is crucial for developing reliable diagnostic tools. LDA and QDA are two widely used statistical techniques for classification problems. LDA assumes equal covariance structures among groups, making it more effective when the assumption holds, while QDA relaxes this assumption, allowing for more flexibility in classification (Hastie et al., 2009)

The study aims to investigate significant differences between Alzheimer's patients and Non-Alzheimer's patients concerning lifestyle factor, clinical measurements, and cognitive and functional assessments. Additionally, it seeks to identify substantial indicators of AD and compare the discriminative power of LDA and QDA in early prediction. By analyzing these aspects, this research contributes to the development of more accurate and efficient classification models for early AD diagnosis.



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Objectives

The general objective of the study is to accurately discriminate Alzheimer's patients to Non-Alzheimer's patients based on their lifestyle factors, clinical measurements, and cognitive and functional assessment.

Specifically, it aims to achieve the following objectives.

1. To identify which clinical indicators can effectively discriminate Alzheimer's patients to Non-Alzheimer's patients.
2. To optimize a discriminant classification model through feature reduction.
3. To compare the discriminative power of LDA and QDA for an early prediction of AD.

Statements of the Problem

Following the mentioned objectives, the study sought to answer the following questions.

1. Are there significant differences between Alzheimer's and Non-Alzheimer's patients in terms of:
 - (a) Lifestyle Factors
 - (b) Clinical Measurements
 - (c) Cognitive and Functional Assessment
2. What are the substantial indicators of AD? in terms of:
 - (a) Lifestyle Factors



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

- (b) Clinical Measurements
 - (c) Cognitive and Functional Assessment
3. How do the discriminative powers of LDA and QDA compare in the early prediction of AD?

Significance of the Study

This research aims to develop a classification model for the early prediction of AD by analyzing key factors such as lifestyle, clinical measures, and cognitive and functional assessments. The findings of this study will be significant to the following sectors:

Healthcare Practitioner and Neurologist. This study will provide a data-drive approach for early AD prediction, aiding in timely intervention and personalized treatment plan. Furthermore, this research will enhance the understanding of the most substantial indicators of AD based on statistical multivariate models.

Researchers and Data Scientist. The study will contribute to the growing field of medical data analysis by exploring the effectiveness of LDA and QDA in disease membership classification. It will serve as a foundation for the future studies integrating machine learning and statistical methods in healthcare analytics which will potentially lead to more advanced diagnostic tools.

Patients and Families. This study will benefit them as early detection tools can help them prepare for disease management and necessary lifestyle adjustments. This will also raise awareness of the key lifestyle and medical factors that may contribute to the risk of AD and encourage them to have preventive measures that may delay the onset of disease.



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Public Health and Policy Makers. This study holds significance to the general public health sector by supporting the development of data-drive healthcare policies focused on early screening and intervention programs. The insights gained from this research can help allocate resources in AD research, prevention and treatment programs.

Scopes and Limitations

This study aims on developing a classification model for early prediction of AD using LDA and QDA. It focuses on analyzing and identifying important factors such as lifestyle habits, clinical measurements, and cognitive and functional assessments to distinguish between Alzheimer's and Non-Alzheimer's patients. The study also compares the discriminative power of LDA and QDA in identifying substantial predictors of the disease and compares their classification performance.

The dataset used in this study consists of 2,149 patient records with unique identifications ranging from 4571 to 6900. It includes demographic details, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and Alzheimer's diagnosis. Since the dataset is synthetically generated and designed from educational purposes, it provides a structured and controlled environment for analysis. However, because it is not based on real patient data, it may not fully capture the variability and complexity of real-world medical dataset.

Despite its strengths, the study has certain limitations. Since the dataset is synthetic and not sourced from actual medical records, the generalizability of the findings to real-world clinical settings may be restricted. Additionally, LDA and QDA rely on certain statistical assumptions, like normality and homoscedasticity for LDA, which



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

may effect their predictive accuracy in more complex datasets. This study also does not compare LDA and QDA with other machine learning models, limiting the analysis to just these two multivariate techniques.

Defintion of Terms



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

CHAPTER II REVIEW OF LITERATURE AND STUDIES

Factors Influencing Alzheimer's Disease

Lifestyle Factors

Lifestyle behaviors play a vital role in maintaining cognitive health and reducing the risk of Alzheimer's disease. Engaging in regular physical activity, for instance, has been linked to a lower likelihood of cognitive decline. Dominguez et al., 2021 emphasized that aerobic exercises not only boost brain function but can also help delay the onset of dementia. Similarly, following a healthy diet, particularly one rich in fruits, vegetables, and whole grains, like the Mediterranean diet, has been shown to enhance cognitive performance and decrease the risk of Alzheimer's. Maintaining an active social life also contributes significantly, as frequent social interactions and participation in activities foster cognitive resilience and lower the chances of developing dementia. (Dominguez et al., 2021)

Sleep quality and duration are emerging as important factors in Alzheimer's risk. Poor sleep habits, such as not getting enough rest or experiencing disrupted sleep, have been linked to a buildup of amyloid-beta in the brain, a hallmark of Alzheimer's. Dominguez et al., 2021 highlights that good-quality sleep is essential for brain health. It helps clear out metabolic waste products and supports memory consolidation, making it a critical part of maintaining cognitive function.

Clinical Measurements

Recent advancements in biomarker research have greatly improved the ability



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

to classify and detect Alzheimer's disease at an early stage. Blood-based biomarkers, such as specific microRNAs (miRNAs), are showing promise in identifying individuals at risk even before symptoms appear (JAMA Network, 2019). Genetics also play a key role, with the APOE4 allele being a well-known risk factor. People who carry this variant have a higher chance of developing Alzheimer's, and integrating genetic information into classification models can enhance accuracy in predicting the disease (JAMA Network, 2019).

Health conditions like diabetes, heart disease, and high blood pressure have been linked to a higher risk of developing Alzheimer's disease. These issues can worsen the impact of vascular problems on brain function and may even interact with Alzheimer-related changes in the brain. Taking steps to manage these conditions through healthy lifestyle choices and medical treatments, play a key role in lowering the overall risk of dementia (PMC, 2021).

Cognitive and Functional Assessment

Cognitive tests play a crucial role in identifying and classifying Alzheimer's disease. Standard assessments like the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) help measure cognitive function in a structured way. The results from these tests provide valuable insights that can improve classification models, making it easier to distinguish between healthy individuals, those with mild cognitive impairment, and those with Alzheimer's disease (PMC, 2021). Functional assessment, on the other hand, focuses on understanding how well a person can handle everyday tasks and adapt to changes in their thinking abilities. Tools like the Alzheimer's Disease Cooperative Study-Activities of Daily Living In-



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

ventory (ADCS-ADL) and the Functional Activities Questionnaire (FAQ) are often used to evaluate these abilities (Custodio et al., 2022). Supported by the study of Cummings2017 which stated that these assessments go beyond cognitive tests, offering a more complete picture of how Alzheimer's disease affects a person's overall functioning and quality of life.

Detection of Alzheimer's Disease using Discriminant Models

Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are commonly used classification techniques in medical datasets, particularly for AD. According to Jain et al., 2022, LDA stands out for its simplicity, especially when data is linearly separable, while QDA handles more complex, non-linear patterns effectively but at a higher computational cost. This makes the choice between the two methods dependent on the nature of the dataset.

Recent research emphasizes the strengths of both approaches. LDA is praised for its ease of interpretation and ability to identify key predictors, offering reliable and computationally efficient models. In contrast, QDA, with its ability to handle non-linearities, excels in sensitivity, especially in detecting early cognitive decline, but may sacrifice some interpretability (Arbabshirani et al., 2017). Studies show that LDA often has higher specificity, whereas QDA is more sensitive in distinguishing between healthy individuals and those with Alzheimer's (Nguyen and Lee, 2020).

Model interpretability and computational complexity remain critical considerations. LDAs straightforward structure allows for a better understanding of variable relationships, making it a practical choice in clinical settings. On the other hand, QDA's flexibility can come at the cost of higher computational demands and reduced



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

clarity in variable significance (Wang and Zhao, 2018).

Linear Discriminant Analysis (LDA)

LDA was used as a classification method to predict Alzheimer's disease patients based on clinical biomarkers, neuroimaging data, and metabolic profiles. This technique has been utilized in several studies to enhance diagnostic accuracy and distinguish between Alzheimer's disease, mild cognitive impairment (MCI), and healthy controls. Data records mainly consisted of medical imaging data, metabolic biomarkers, and cognitive assessment scores, which were processed through feature selection and dimensionality reduction techniques to enhance classification performance.

LDA assumes that different classes (AD and non-AD groups) share the same covariance structure and constructs linear boundaries to maximize class separability. In the study by (Le et al., 2020), an adapted LDA approach was implemented to handle high-dimensional medical datasets, ensuring better feature selection and classification accuracy. Similarly, Salas-Gonzalez et al., 2010 utilized LDA in combination with factor analysis to select the most relevant features from 18F-FDG PET images, optimizing classification between Alzheimer's and control groups.

The study by Yilmaz et al., 2021 applied artificial intelligence and machine learning techniques alongside discriminant analysis to identify biomarkers associated with Alzheimer's progression. Feature selection was crucial in reducing redundancy and enhancing predictive capability. Maroco et al., 2011 conducted a comparative analysis between LDA, logistic regression, neural networks, support vector machines, classification trees, and random forests. They assessed model performance using key evaluation metrics, demonstrating that LDA achieved high accuracy in distinguishing



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

AD patients from controls when optimized with feature selection.

Quadratic Discriminant Analysis (QDA)

QDA, unlike LDA, relaxes the assumption of shared covariance structures and allows each class to have its own covariance matrix, making it suitable for datasets where feature distributions exhibit non-linearity. This flexibility enables QDA to better handle complex biomarker distributions in Alzheimer's disease classification.

QDA has demonstrated strong classification performance in Alzheimer's research. Studies such as Zhang et al., 2017 and Pereira and Ferreira, 2020 explored the application of QDA in neuroimaging-based Alzheimer's diagnosis, showing that its ability to model class-specific covariance matrices led to improved classification precision and recall. Additionally, the work of Maroco et al., 2011 indicated that QDA often achieved higher F1-scores in datasets with more complex patterns compared to LDA.

Furthermore, studies have shown that QDA performs well in scenarios where biomarker distributions exhibit high variability. For instance, Lee and Kim, 2015 demonstrated that QDA outperformed LDA in identifying Alzheimer's subtypes based on MRI-derived biomarkers, particularly in cases with overlapping clinical features. Similarly, Garcia-Rodriguez and Gonzalez-Escamilla, 2016 employed QDA to analyze cerebrospinal fluid biomarkers, showing significant improvements in classifying AD and MCI patients. The higher Area Under the Curve (AUC) values reported in these studies suggest that QDA provides superior discrimination between AD and non-AD cases when biomarker variability is high.



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Feature Selection Techniques in Discriminant Analysis

Recursive Feature Elimination (RFE)

RFE is a backward feature elimination technique that iteratively removes the least significant features based on their impact on model accuracy. This study implements RFE to systematically refine feature subsets, ensuring that only the most relevant biomarkers contribute to classification. The process starts by training an initial model on all features, ranking their importance, and recursively eliminating the least informative ones until an optimal subset remains.

Previous studies, such as Balakrishnan et al., 2020, combined RFE with artificial neural networks to identify critical features for Alzheimer's disease classification. Similarly, Alshamlan et al., 2018 demonstrated that RFE facilitates biomarker gene selection, improving classification performance. In this study, RFE is applied to clinical biomarkers, cognitive scores, and neuroimaging markers to eliminate redundant features, refine the input space, and mitigate overfitting, thereby enhancing the performance of LDA and QDA.

Lasso Regression

Lasso regression, a form of regularization, employs an L1 penalty constraint to shrink the coefficients of less relevant features to zero, effectively selecting only the most significant predictors. This study applies Lasso regression to ensure that the discriminant functions of LDA and QDA are derived from the most influential biomarkers while minimizing noise and improving model generalizability.

Gu et al., 2020 evaluated Lasso regression as a feature selection method for



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Alzheimer's diagnosis, emphasizing its capacity to identify high-impact features while preserving interpretability. Additionally, Spooner et al., 2020 demonstrated Lasso's efficacy in biomarker discovery, enhancing model robustness in distinguishing between disease stages. By incorporating Lasso regression, this study aims to improve classification accuracy by retaining only the most essential biomarkers.

Synthesis



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

CHAPTER III

METHODOLOGY

Method of Research

The researchers employed a comparative quantitative research design, focusing on the evaluation and comparison of discriminant analysis models for classifying Alzheimer's patients. The study used the publicly available Alzheimer's Disease dataset from Kaggle which was divided into two randomized subsets of 80% for the training data and 20% for testing data, ensuring that distinct portions are used for each iteration of 10-fold cross validation for all complete and reduced linear and quadratic discriminant models in the study.

The reduced models of LDA and QDA will be derived using Recursive Feature Elimination (RFE) and Lasso Regression, producing two versions of reduced model for LDA and QDA. This will identify the most significant predictors of Alzheimer's patients classification, ensuring that most relevant features are emphasized. These models will be compared and evaluated using classification performances metrics, including, accuracy, specificity, sensitivity, F1-Score, and No Information Rate.

Data and Variables

The dataset used in this study contains a comprehensive health information for 2,149 patients, including their demographic details, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and Alzheimer's disease diagnoses. Each variables are summarized in table 1.



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Table 1

Variable Summary of Kharoua R.E.'s Alzheimer's Disease Dataset

Variable Name	Description	Unit of Measure
Lifestyle Factors		
BMI	Body Mass Index	kilogram per meter squared
AlcoholConsumption	Weekly Alcohol consumption in units	in units
PhysicalActivity	Weekly physical activity	in hours
DietQuality	Diet quality assessment	score from 0 to 10
SleepQuality	Sleep quality assessment	score from 4 to 10
Clinical Measurements		
SystolicBP	Systolic Blood Pressure	mmHg
DiastolicBP	Diastolic Blood Pressure	mmHg
CholesterolTotal	Total Cholesterol Levels	mg/dL
CholesterolLDL	Low-density Lipoprotein Cholesterol Levels	mg/dL
CholesterolHDL	High-density Lipoprotein Cholesterol Levels	mg/dL
CholesterolTriglycerides	Triglycerides Levels	mg/dL
Cognitive and Functional Assessments		
MMSE	Mini-Mental State Examination Score	Lower scores indicate cognitive impairment
FunctionalAssessment	Functional Assessment Score	Lower scores indicate greater impairment
ADL	Activities of Daily Living Score	Lower scores indicate greater impairment
Diagnosis Information		
Diagnosis	Status of Alzheimer's Disease	0 means No, 1 mean yes

Data Preprocessing

Before analysis, the data underwent preprocessing to ensure accuracy and reliability. Outliers are identified and assessed for their impact, and if there exist extreme values, their influence on the analysis was carefully evaluated before deciding whether to retain, transform, or remove them.

To meet the assumptions of discriminant analysis, three key aspects are examined: normality of predictor variables, homogeneity of variance - covariance matrices, and multicollinearity among independent variables.



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Since the data is generated synthetically, the researchers assumed the distributions to be approximately normal. The homogeneity of variance-covariance matrices was assessed using Box's M test, which tests whether the covariance matrices are equal across the groups. Lastly, Multicollinearity is evaluated using the Variance Inflation Factor (VIF) not exceeding the threshold value of 10. Variables with high VIF values is considered to improve model stability.

Statistical Treatment of Data

Recursive Feature Elimination

RFE is employed to iteratively remove the least informative feature based on model performance. To determine the optimal number of features for the reduced model of LDA and QDA, the control was set using the random forest function with cross-validation over 10 iterations. Root Mean Squared Error (RMSE) across all feature size from 1 to 15— based from the set controls, is assessed to determine the number of features to consider in the reduced model. The optimal variables to be included is derived from highest overall coefficients yielded from the same results of RFE in the caret package in R.

Lasso Regression

Lasso regression is applied to introduce sparsity by shrinking coefficients of less important features toward zero, effectively selecting only the most relevant features. Using the glmnet package in R, an initial Generalized Linear Regression Model is trained using the binomial family and alpha of 1. The best lambda to consider is assessed using cross-validation of GLM to determine the minimum value of lambda



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

that will penalize least important features in the dataset. The features with the highest coefficients is chosed to be included in the reduced model.

Linear Discriminant Analysis

LDA is applied to classify individuals as diseased or non-diseased based on the selected features. It identifies a linear combination of features that maximizes the separation between two classes. The classification process involves estimating the mean vector for each class and computing a pooled covariance matrix, which is used to determine decision boundaries. Prior probabilities for each class were either assumed to be equal or estimated based on class distributions in the dataset.

Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) was also implemented to account for potential non-linear relationships between the selected biomarkers and Alzheimer's disease outcomes. Unlike LDA, QDA does not assume equal variance-covariance matrices across groups, allowing for greater modeling flexibility. This approach estimates separate covariance matrices for each class, allowing for quadratic decision surfaces. The classification process followed a similar structure to LDA but accounted for differences in group-specific covariance structures. This additional flexibility will enable QDA to capture more complex relationships between biomarkers and Alzheimer's disease status.

Model Evaluation

The performance of LDA and QDA models is assessed using various evaluation



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

metrics, including accuracies across 10-folds of cross validation, specificity, sensitivity, F1-score, and No Information Rate (NIR). These metrics provided a comprehensive assessment of the models' classification ability, ensuring both overall performance and the balance between sensitivity and specificity were considered.

A comparative analysis was conducted between LDA and QDA to determine which model performed better based on these evaluation metrics. The model with higher classification accuracy, a better balance between sensitivity and specificity, and a non-significant NIR were considered more effective for predicting Alzheimer's disease outcomes.

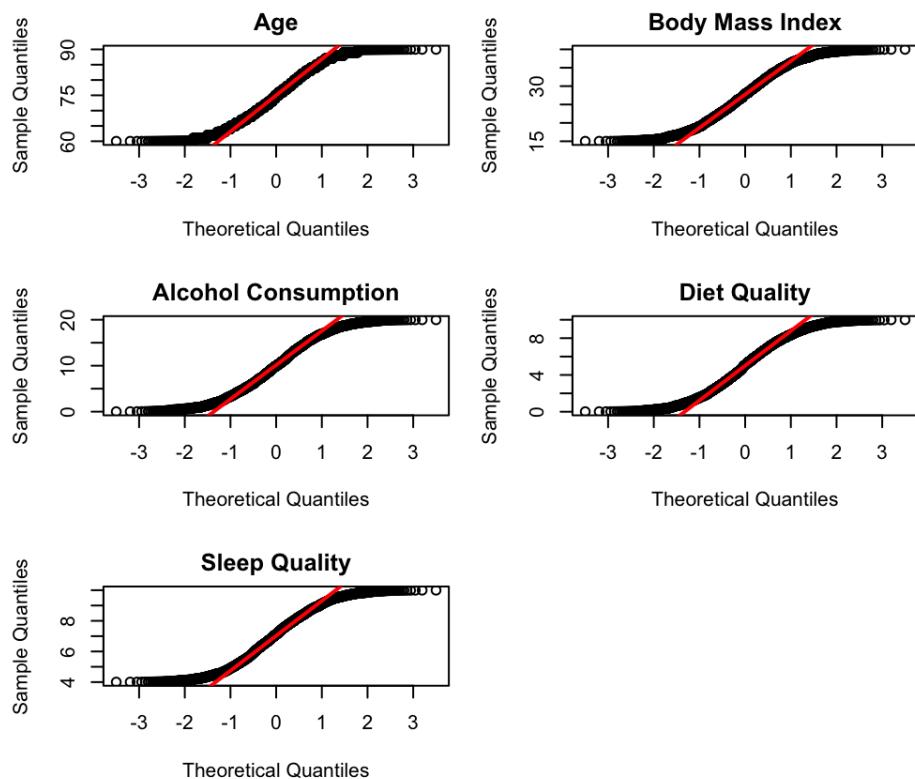
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

CHAPTER IV RESULTS AND DISCUSSION

Exploratory Data Analysis of Alzheimer's Patients and Non-Alzheimer's Assumptions of Discriminant Analysis

Figure 1

Normality Assumption of Lifestyle Factors

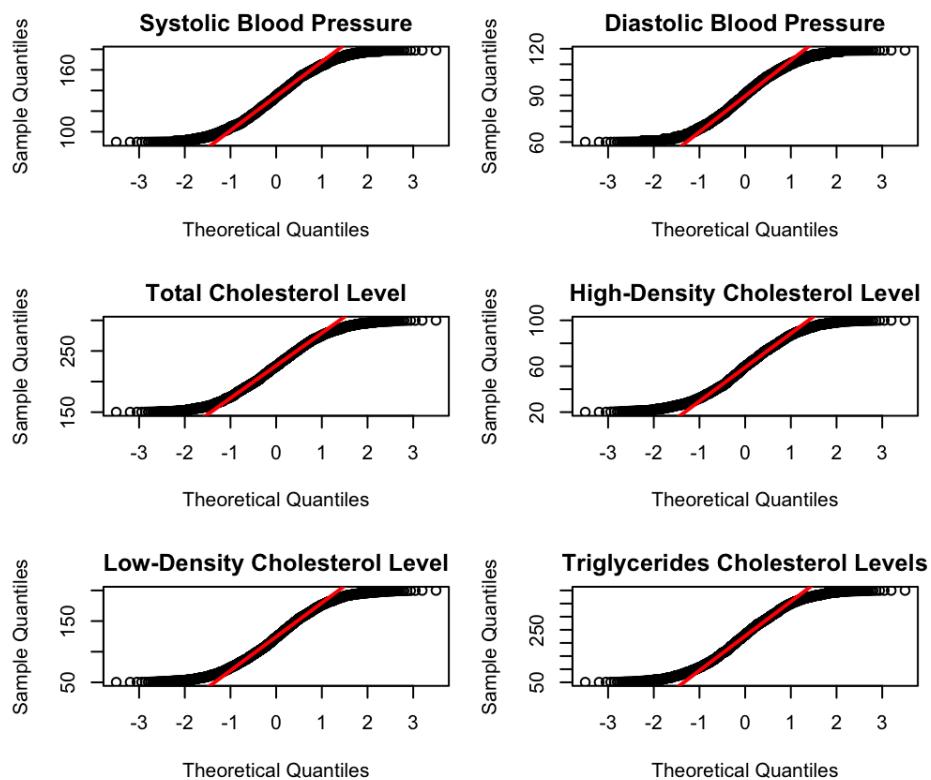


Insert interpretation

Figure 2

Normality Assumption of Clinical Measurements

POLYTECHNIC UNIVERSITY OF THE PHILIPPINES



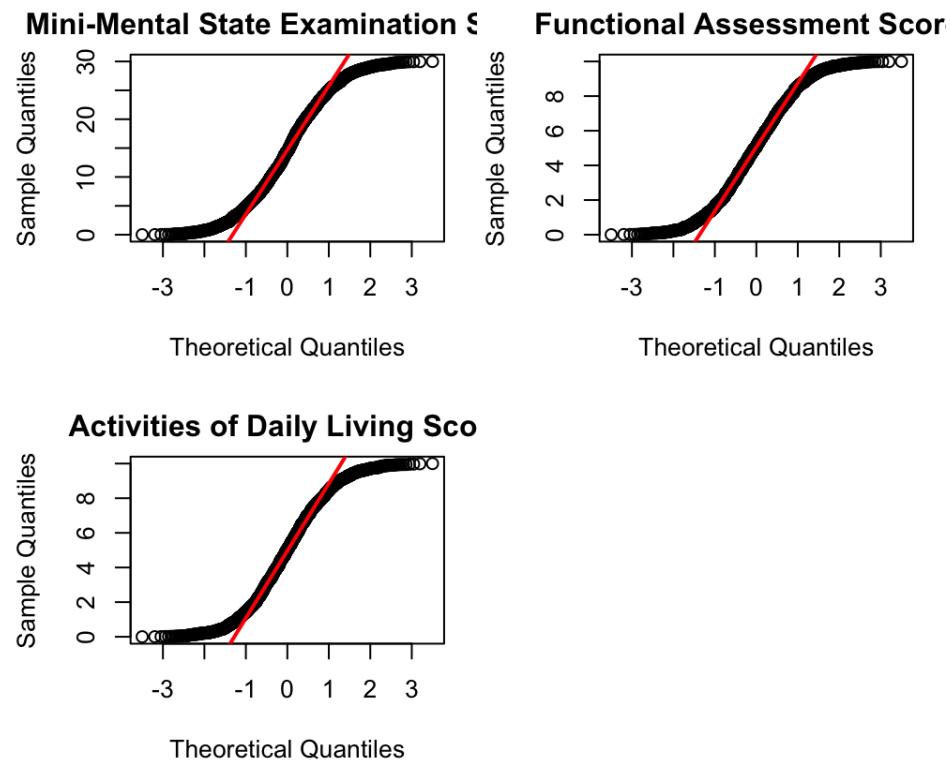
Insert interpretation

Figure 3

Normality Assumption of Cognitive and Functional Assessments



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES



Insert interpretation

Table 2

Homogeneity of Covariance Matrices

X^2	df	$p\text{-value}$
205.57	120	< 0.01

Performing Discriminant Analysis, specifically LDA, requires the groups of interest in the data to have homoscedasticity. However, for QDA, it does not assume homoscedasticity since each class in the dataset will be assigned an estimated individual covariance matrix.

As shown in table 2, in checking the assumption for the Alzheimer's disease,

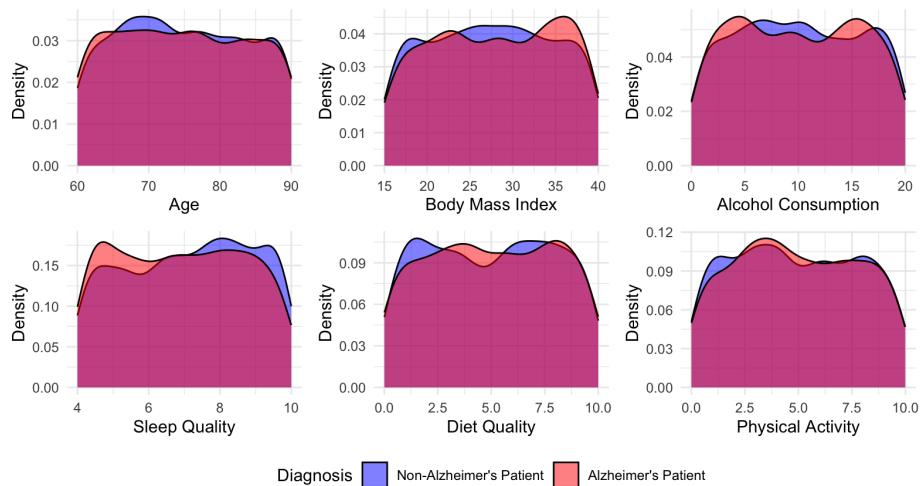
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

setting the diagnosis information as the grouping variable, the Box's M-test for homoscedasticity was utilized, given the multivariate normality has been assumed. The resulting p-value from this test was less than 0.01, computed from $X^2 = 205.57$ and $df = 120$. This means that the groups of AD and Non-AD do not have equal covariances. As a violation of this assumption will negatively impact the performance of LDA, QDA may prove to be more reliable in cases such as this, wherein there are unequal covariances between the two classes.

Lifestyle Factor Distributions of Patients

Figure 4

Lifestyle Factors Density of Patients by Diagnosis



The figure above presents the density plots for Age, BMI, Alcohol Consumption, Sleep Quality, Diet Quality, and Physical Activity, comparing Alzheimer's patients (blue density plot) and Non-Alzheimer's patients (red density plot), with overlapping areas appearing purple.

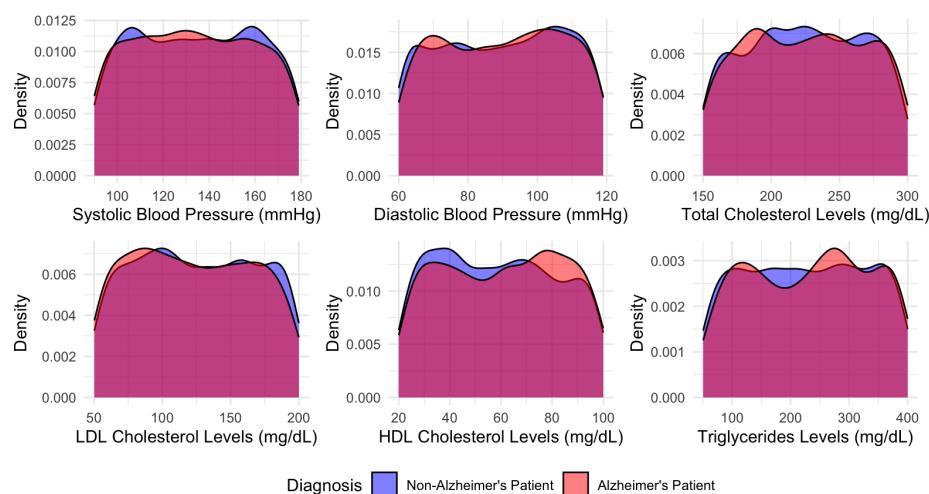
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Through visual inspection, it reveals that the age distribution of both groups are fairly uniform, following a similar trend with minimal variation. Similar patterns are observed across other variables, with slight differences but no substantial disparities. However, Sleep Quality, deviates among the other, as it exhibits a more noticeable distinction, where Alzheimer's patients show higher density levels compared to Non-Alzheimer's patients.

Clinical Measurements of Patients

Figure 5

Clinical Measurement Densities of Patients by Diagnosis



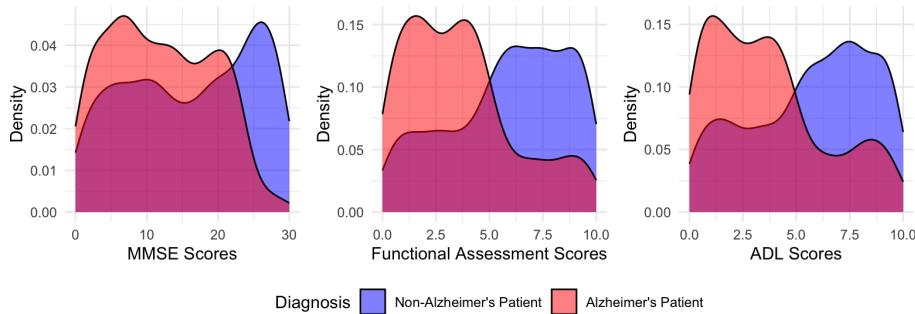
The figure above displays the density distributions of clinical markers measured from both groups. Across all identifiers, both groups follow a similar trend with little to none variations.

POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Cognitive and Functional Assessments of Patients

Figure 6

Cognitive and Functional Assessment Score Densities of Patients by Diagnosis



In terms of Cognitive and Functional Assessments, as illustrated in Figure 6, there are significant disparities across the variables. Specifically, MMSE scores show that Alzheimer's patients tend to have lower scores compared to Non-Alzheimer's patients, suggesting that most AD patients experience moderate to severe cognitive impairment. Comparably, Non-AD patients have higher functional assessment scores and ADL scores than AD patients. These marked differences in densities across variables may imply a significant indicator of AD.

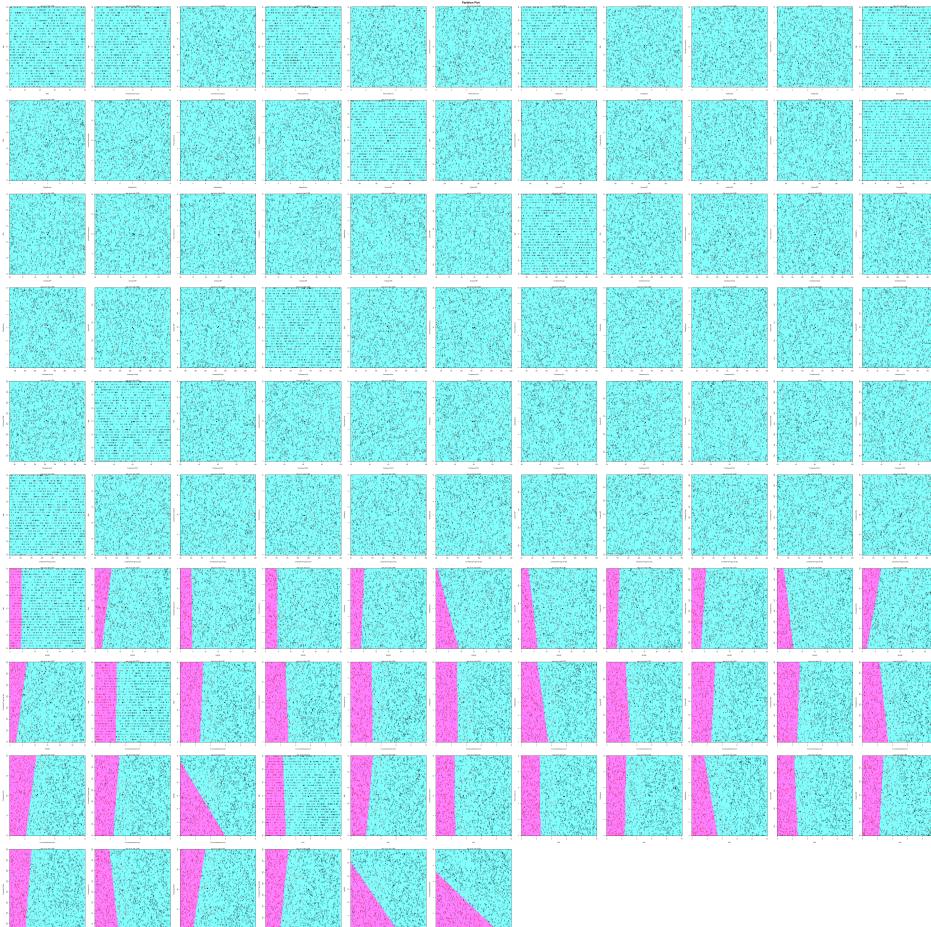
Discriminant Modeling of AD Patients

Complete Linear and Quadratic Discriminant Models

Figure 7

Complete Linear Class Discrimination of AD and Non-AD Patients

POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

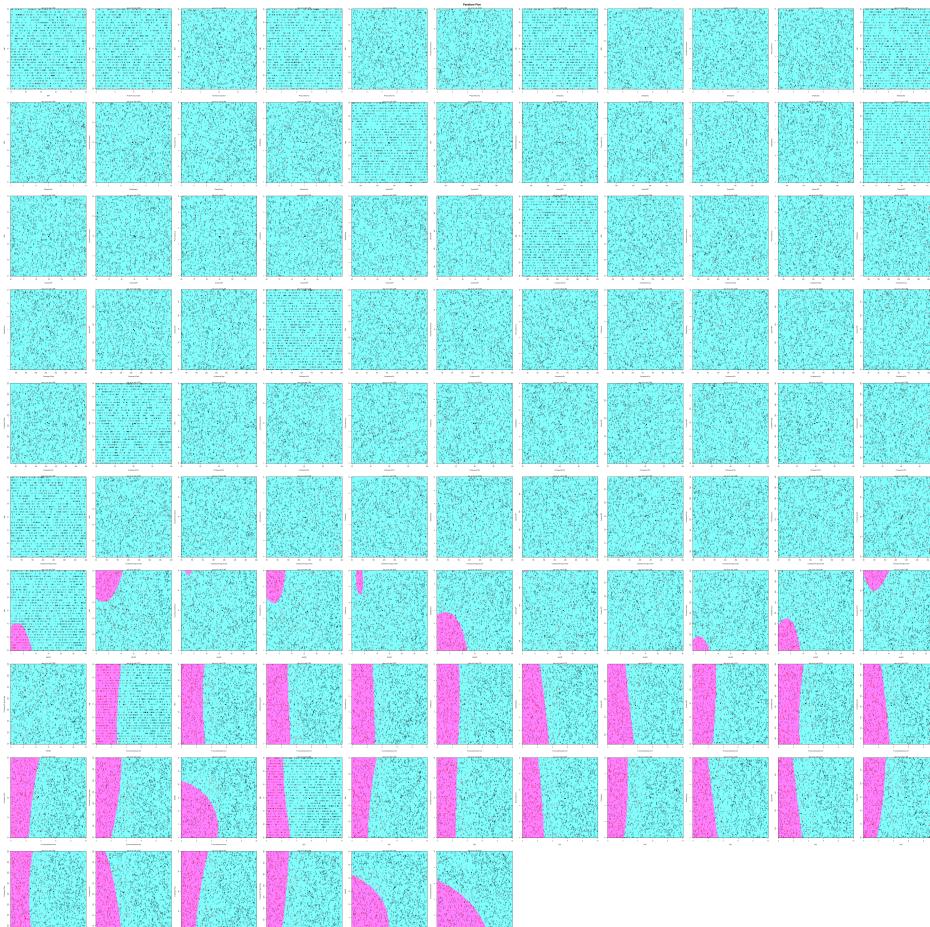


The figure displays that among the 105 possible partition plots for LDA, only 39 pairs show clear separation between AD and Non-AD patients. Notably, all of these 39 plots include at least one of the following variables: MMSE, Functional Assessment, or ADL. This suggests that these three variables may hold the strongest discriminative power in distinguishing the two groups, as they consistently contribute to visible class separation. In contrast, the remaining variables, when analyzed in pairs without MMSE, Functional Assessment, and ADL, do not provide distinct visual partitions, implying that these lack stronger linear discriminative ability on their own.

Figure 8

POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Complete Quadratic Class Discrimination of AD and Non-AD Patients



On the other hand, the figure above shows the 105 possible partition plots for QDA, and almost similar to LDA, only 38 pairs were able to provide distinct separation between diagnosed AD and Non-AD. Specifically, the pairing of MMSE and Diastolic Blood Pressure did not show any partition in QDA, as compared to LDA. Through visual inspection, the pairs that include MMSE could only partition smaller and limited regions. This could mean that its effect on distinguishing AD and Non-AD patients may not be consistent across all values through QDA. Instead, its discriminative ability might be more pronounced in specific subgroups or at extreme values, which in this case



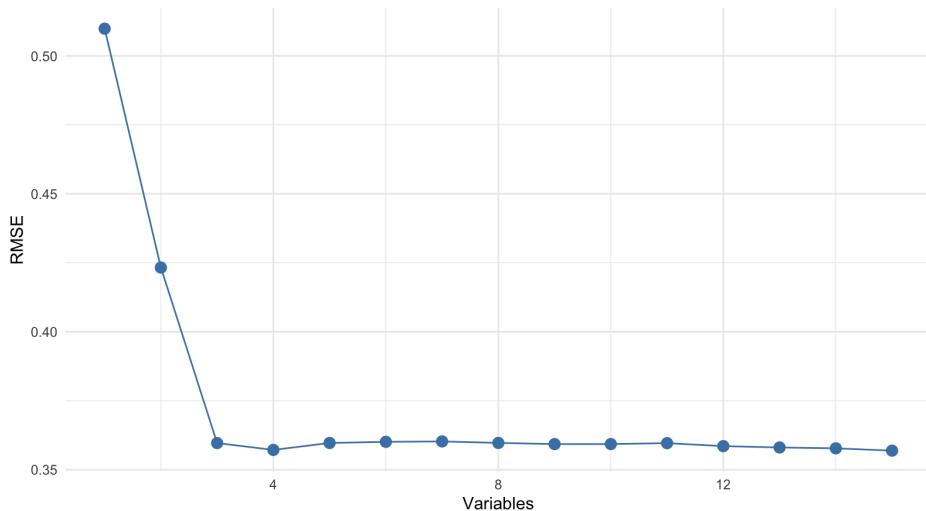
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

are low MMSE scores, rather than across the entire scale range. This simply means that while its relevant in distinguishing, MMSE might not be a standalone predictor in nonlinear classification method.

RFE-Reduced Linear and Quadratic Discriminant Models

Figure 9

Root Mean Squared Error Comparison of Increasing Number of Features



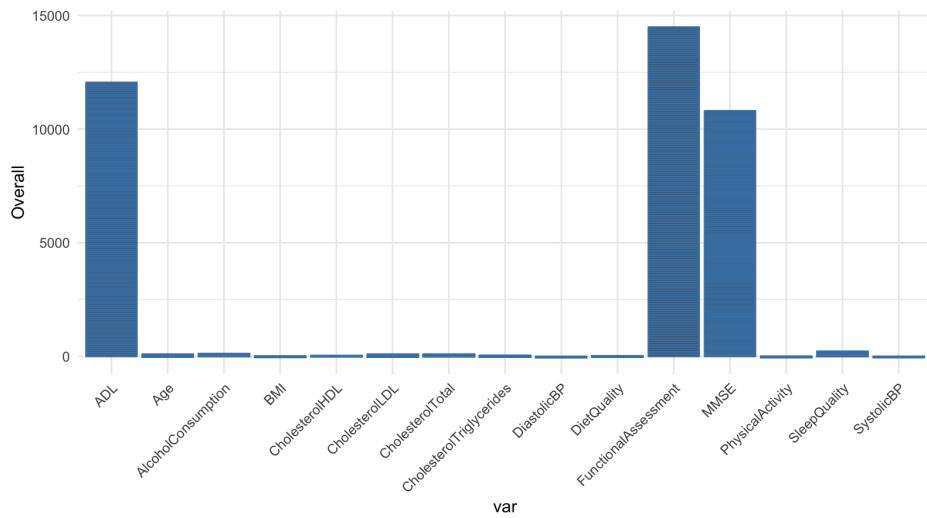
To determine the optimal number of features for discriminating between AD and Non-AD patients, as shown in figure 9, the error metrics of each model is evaluated as the number of features increases. A significant decline in error is observed after testing the model with three features. Beyond this, the model's error stabilized, with values consistently close to 0.35. Thus, the reduced model for both LDA and QDA should only have at least three features as predictors to achieve lower error measures and stability.

Figure 10



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Recursive Feature Comparison by Importance

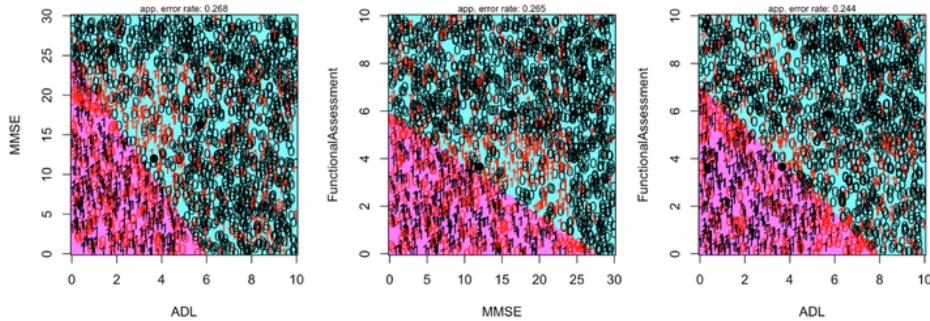


Using the same caret algorithm, variable importance is determined using RFE. In figure 10, MMSE, Functional Assessments, and ADL play a significant role in differentiating AD and Non-AD patients, which is consistent with the findings from the density distributions of cognitive and functional assessment and from the complete model of the previous discussions. Aside from these three, other variables have no substantial contribution to the distinction of AD and Non-AD patients as evident on the graph.

Figure 11

RFE-Reduced Linear Class Discrimination Model of AD and Non-AD Patients

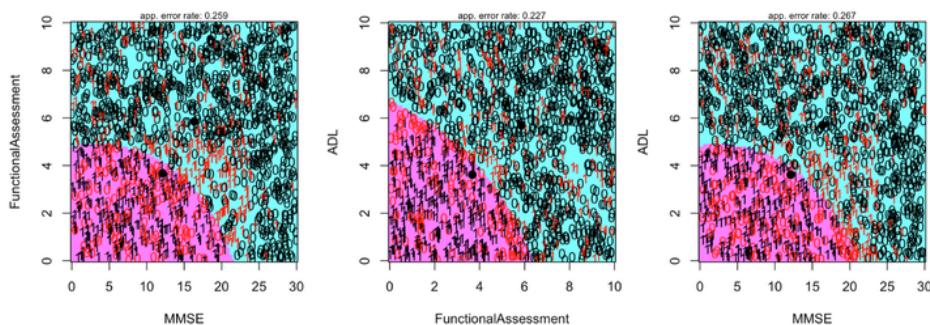
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES



Consequently, using the reduced number of features and the variables with the highest importance in figures 9 and 10, respectively, figure 11 displays the linear discrimination of AD patients and Non-AD patients. Compared to the full model, all partition pairs in the RFE-reduced linear model captured the distinct separation between the two classes. However, the combination of pairings was reduced to three.

Figure 12

RFE-Reduced Quadratic Class Discrimination Model of AD and Non-AD Patients



Similar trends can be observed in the quadratic discrimination of AD patients and Non-AD patients. All partition pairs in the reduced quadratic model captured the nonlinear distinct separation between the two classes. However, consistently, the combination of pairings was reduced to three.

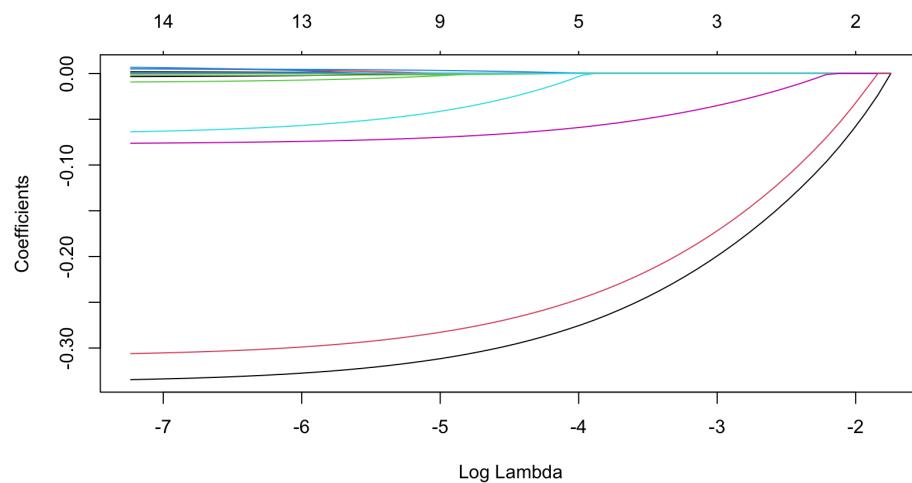


POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Lasso Regression-Reduced Linear and Quadratic Discriminant Models

Figure 13

Lambda Regularization by Feature Coefficients

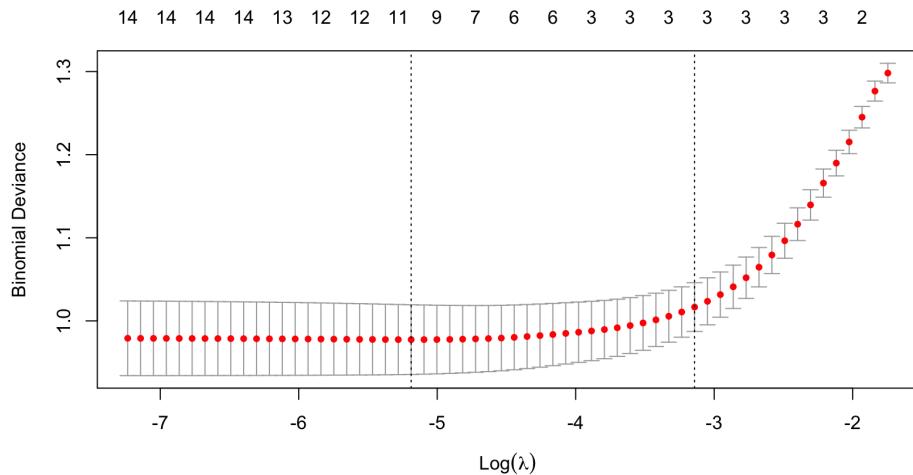


The logarithm of lambda controls the regularization strength of the coefficients for feature reduction. As the value of lambda increases, more coefficients shrink towards zero which indicates that more features are being penalized by the model. Based on the graph, along lambda values of -5 to -3, the model achieves the best trade-off between coefficient shrinkage and performance.

Figure 14

Binomial Deviance and Logarithmic Lambda

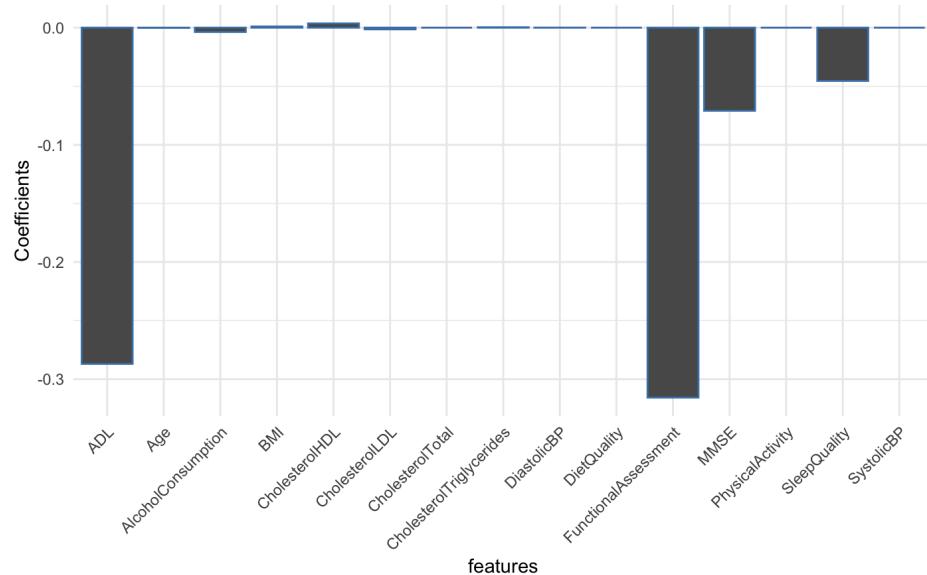
POLYTECHNIC UNIVERSITY OF THE PHILIPPINES



Upon cross-validated, figure 9 displays the error rate of the lambda values as it increases. The red dot are the mean coefficients and vertical lines are its error rates. As the lambda value increases the error rate is also increasing, shrinking the coefficients down to zero.

Figure 15

Feature Importance based on the Optimal Lambda

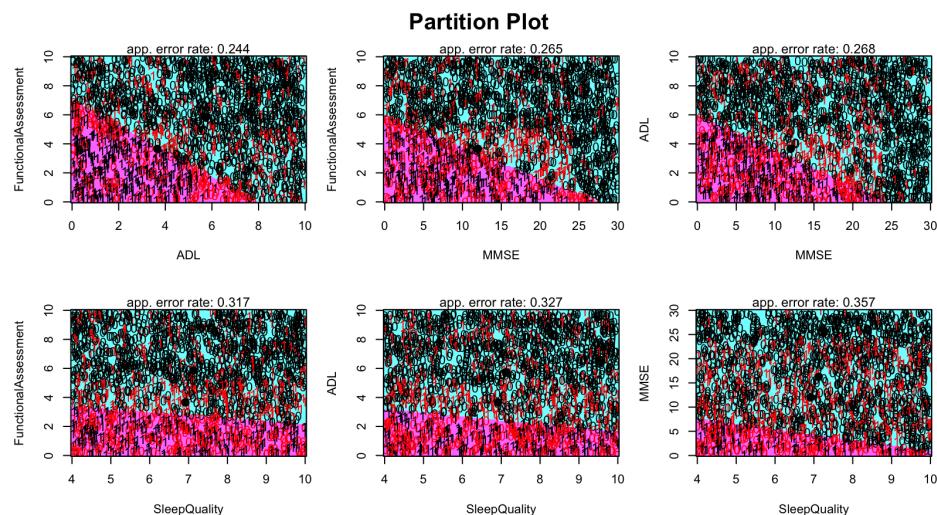


POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Thus, based on the optimal lambda from the lasso regression, ADL, Functional Assessments, MMSE, and Sleep Quality were the top four coefficients compared among the features who almost shrunk to zero. It implies that these four variables contribute the most in discriminating AD patients and Non-AD patients.

Figure 16

Lasso-Reduced Linear Class Discrimination of AD and Non-AD Patients

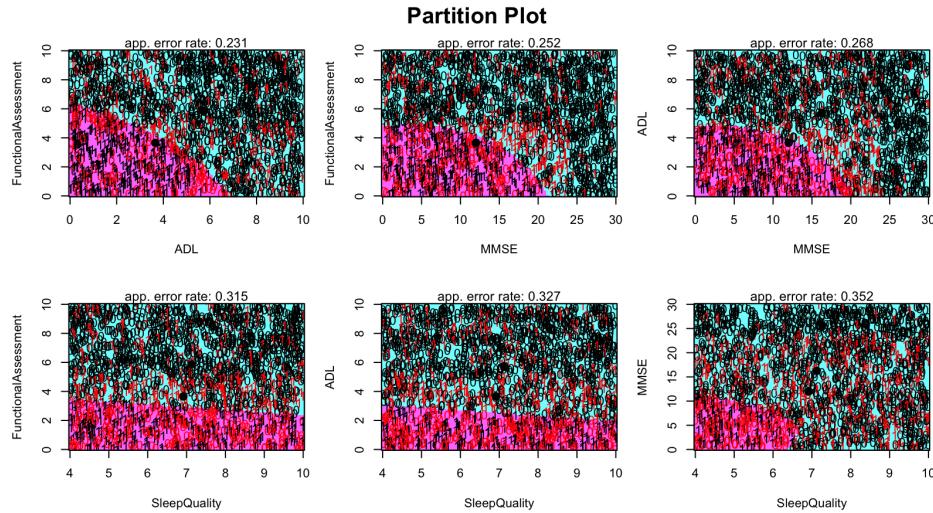


Tama na please.

Figure 17

Lasso-Reduced Quadratic Class Discrimination of AD and Non-AD Patients

POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

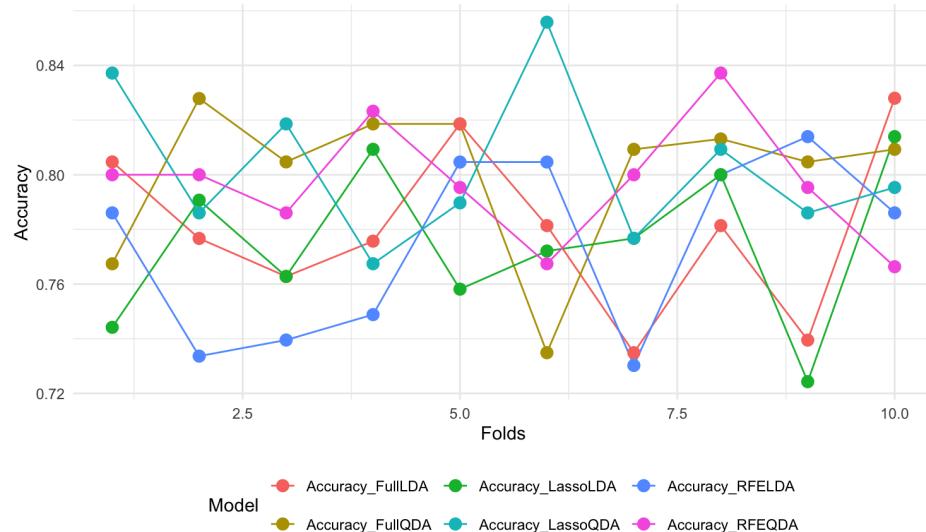


Babasahin mo ba to pj

Classification Metrics Comparisons of LDA and QDA Models

Figure 18

Model Accuracies Across 10-Folds



Amazing



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

CHAPTER V CONCLUSION AND RECOMMENDATION

Summary of Findings

Recommendations



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

References

- Alshamlan, H., Omar, S., & Aljurayyad, R. (2018). Identifying effective feature selection methods for alzheimer's disease biomarker gene detection using machine learning [Accessed: 2025-01-30]. *International Journal of Advanced Computer Science and Applications*, 9(4), 168–174. <https://thesai.org/Publications/ViewPaper?Volume=9&Issue=4&Code=ijacsa&SerialNo=23>
- Alzheimer's Association. (2021). 2021 alzheimer's disease facts and figures. *Alzheimer's and Dementia*, 17(3), 327–406. <https://doi.org/10.1002/alz.12328>
- Alzheimers Disease International. (n.d.). Dementia statistics [Accessed: 2025-01-30]. <https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/>
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *Frontiers in Neurology*, 8, 315. <https://doi.org/10.3389/fneur.2017.00315>
- Balakrishnan, D., Mariappan, U., Prudhvi, B. S., Saikumar, P., Mahesh, B. M., & Mani Sai, P. (2020). Alzheimer disease prediction using recursive feature elimination and artificial neural network [Accessed: 2025-01-30]. *International Journal of Advanced Science and Technology*, 29(4), 746–755. https://www.researchgate.net/publication/374803499_Alzheimer_Disease_Prediction_Using_Recursive_Feature_Elimination_and_Artificial_Neural_Network
- Custodio, N., Montesinos, R., Chambergo-Michilot, D., Herrera-Perez, E., Pintado-Caipa, M., Seminario, W., Cuenca, J., Mesía, L., Failoc-Rojas, V., & Diaz, M. (2022). A functional assessment tool to distinguish controls from alzheimer's



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

- disease in lima, peru [Accessed: 2025-01-31]. *ResearchGate*. https://www.researchgate.net/publication/361057550_A_Functional_Assessment_Tool_to_Distinguish_Controls_From_Alzheimer's_Disease_in_Lima_Peru
- Dominguez, L. J., Barbagallo, M., Munoz-Garcia, M., & Godos, J. (2021). Nutrition, physical activity, and other lifestyle factors in the prevention of cognitive decline and dementia. *Nutrients*, 13(11), 4080. <https://doi.org/10.3390/nu13114080>
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genton, R., ... Jack, C. R. (2016). Preclinical alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's and Dementia*, 12(3), 292–323. <https://doi.org/10.1016/j.jalz.2016.02.002>
- Garcia-Rodriguez, A., & Gonzalez-Escamilla, G. (2016). Qda for cerebrospinal fluid biomarker analysis in alzheimer's disease and mild cognitive impairment [Accessed: 2025-01-30]. *Alzheimer's and Dementia*, 12(7), 748–756. <https://doi.org/10.1016/j.jalz.2016.01.015>
- Gu, F., Ma, S., Wang, X., Zhao, J., Yu, Y., & Song, X. (2020). Evaluation of feature selection for alzheimer's disease diagnosis [Accessed: 2025-01-30]. *Frontiers in Neuroscience*, 14, 8. <https://www.frontiersin.org/articles/10.3389/fnins.2020.00008/full>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd) [Accessed: 2025-01-30]. Springer Series in Statistics. [https://www.researchgate.net/publication/319770542_The_Elements_of_Statistical_Learning:_Data_Mining,_Inference,_and_Prediction_\(2nd_Edition\)](https://www.researchgate.net/publication/319770542_The_Elements_of_Statistical_Learning:_Data_Mining,_Inference,_and_Prediction_(2nd_Edition))



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

- Elements_of_Statistical_Learning_Data_Mining_Inference_and_Prediction_Second_Edition_Springer_Series_in_Statistics
- Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, J., Molinuevo, J. L., Montine, T. J., Phelps, C. H., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., & Sperling, R. (2018). Nia-aa research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*, 14(4), 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
- Jain, A., Gupta, S., & Kumar, R. (2022). Comparative study of lda and qda in health-care data classification. *Journal of Healthcare Engineering*, 2022, 1–10. <https://doi.org/10.1155/2022/7829795>
- JAMA Network. (2019). Genetic risk factors and biomarkers in alzheimer's disease detection. *JAMA Network*.
- Kivipelto, M., Mangialasche, F., & Ngandu, T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and alzheimer disease. *Nature Reviews Neurology*, 14(11), 653–666. <https://doi.org/10.1038/s41582-018-0070-3>
- Le, K. T., Chaux, C., Richard, F. J. P., & Guedj, E. (2020). An adapted linear discriminant analysis with variable selection for the classification in high-dimension, and an application to medical data [Accessed: 2025-01-30]. *Computational Statistics and Data Analysis*, 152, 107031. <https://www.sciencedirect.com/science/article/abs/pii/S0167947320300304>
- Lee, J., & Kim, H. (2015). Mri biomarker identification using quadratic discriminant analysis in alzheimer's subtypes [Accessed: 2025-01-30]. *NeuroImage: Clinical*, 8, 1–10. <https://doi.org/10.1016/j.nicl.2015.04.008>



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Leszek, J., Mikhaylenko, E. V., Belousov, D. M., Koutsouraki, E., Szczechowiak, K., Kobusiak-Prokopowicz, M., Mysiak, A., Diniz, B. S., Somasundaram, S. G., Kirkland, C. E., & Aliev, G. (2021). The links between cardiovascular diseases and alzheimer's disease. *Current Neuropharmacology*, 19(2), 117–132. <https://doi.org/10.2174/1570159X18666201008143743>

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity, and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees, and random forests [Accessed: 2025-01-30]. *BMC Research Notes*, 4, 299. <https://bmcresearchnotes.biomedcentral.com/articles/10.1186/1756-0500-4-299>

National Institutes of Health. (n.d.). Combination of healthy lifestyle traits may substantially reduce alzheimer's [Accessed: 2025-01-30]. <https://www.nih.gov/news-events/news-releases/combination-healthy-lifestyle-traits-may-substantially-reduce-alzheimers>

Nguyen, H. V., & Lee, J. (2020). Advances in discriminant analysis for alzheimer's diagnosis. *Machine Learning in Healthcare Applications*, 25(1), 78–89.

Pereira, F., & Ferreira, J. (2020). Covariance-based qda for improved alzheimer's diagnosis using mri [Accessed: 2025-01-30]. *Frontiers in Aging Neuroscience*, 12, 238. <https://doi.org/10.3389/fnagi.2020.00238>

PMC. (2021). The role of comorbidities in alzheimer's disease progression. *PubMed Central*.

Salas-Gonzalez, D., Górriz, J. M., Ramírez, J., Illán, I. A., López, M., Segovia, F., Chaves, R., Padilla, P., & Puntonet, C. G. (2010). Feature selection using factor



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

- analysis for alzheimer's diagnosis using 18f-fdg pet images [Accessed: 2025-01-30]. *Medical Physics*, 37(11), 6084–6095. <https://pubmed.ncbi.nlm.nih.gov/21158318/>
- Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, W., Morrison-Bogorad, M., ... Phelps, C. H. (2011). Toward defining the pre-clinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's and Dementia*, 7(3), 280–292. <https://doi.org/10.1016/j.jalz.2011.03.003>
- Spooner, A., Mohammadi, G., Sachdev, P. S., Brodaty, H., & Sowmya, A. (2020). Ensemble feature selection with data-driven thresholding for alzheimer's disease biomarker discovery [Accessed: 2025-01-30]. *Frontiers in Neuroscience*, 14, 524. <https://www.frontiersin.org/articles/10.3389/fnins.2020.00524/full>
- Wang, L., & Zhao, K. (2018). Computational complexity and model interpretability in ad prediction using lda and qda. *Journal of Bioinformatics and Computational Biology*, 16(1), 75–89.
- Yilmaz, A., Ustun, I., & Graham, S. F. (2021). A community-based study identifying metabolic biomarkers of mild cognitive impairment and alzheimer's disease using artificial intelligence and machine learning [Accessed: 2025-01-30]. *Journal of Alzheimer's Disease*, 79(1), 123–132. <https://content.iospress.com/articles/journal-of-alzheimers-disease/jad200714>



POLYTECHNIC UNIVERSITY OF THE PHILIPPINES

Zhang, D., Wang, Y., & Liu, Y. (2017). Neuroimaging-based quadratic discriminant analysis for alzheimer's disease prediction [Accessed: 2025-01-30]. *Journal of Neuroscience Methods*, 281, 9–16. <https://doi.org/10.1016/j.jneumeth.2017.02.007>