DIAGNOSIS OF ALZHEIMER’S DISEASE USING MACHINE LEARNING

Intelligent Systems

Francisco Carvalho 111000
Group 3

Faculty:
João Miguel da Costa Sousa
Rodrigo Boal Ventura

04.11.2024

PROJECT OVERVIEW

This project aims to leverage machine learning and fuzzy logic models to classify Alzheimer's Disease based on structured patient data.

The goal is to compare different models, such as:

- 01 **Fisrt-order Takagi-Sugeno fuzzy model**
- 02 **Simple neutral network**
- 03 **Multilayer neural network**
- 04 **Support vector machine (SVM)**
- 05 **Decision Tree**

GitHub Repository : https://github.com/FranciscoCarvalho26/Alzeimers-Diagnosis-Project
Dataset DOI : 10.34740/KAGGLE/DSV/8668279

DATA DESCRIPTION

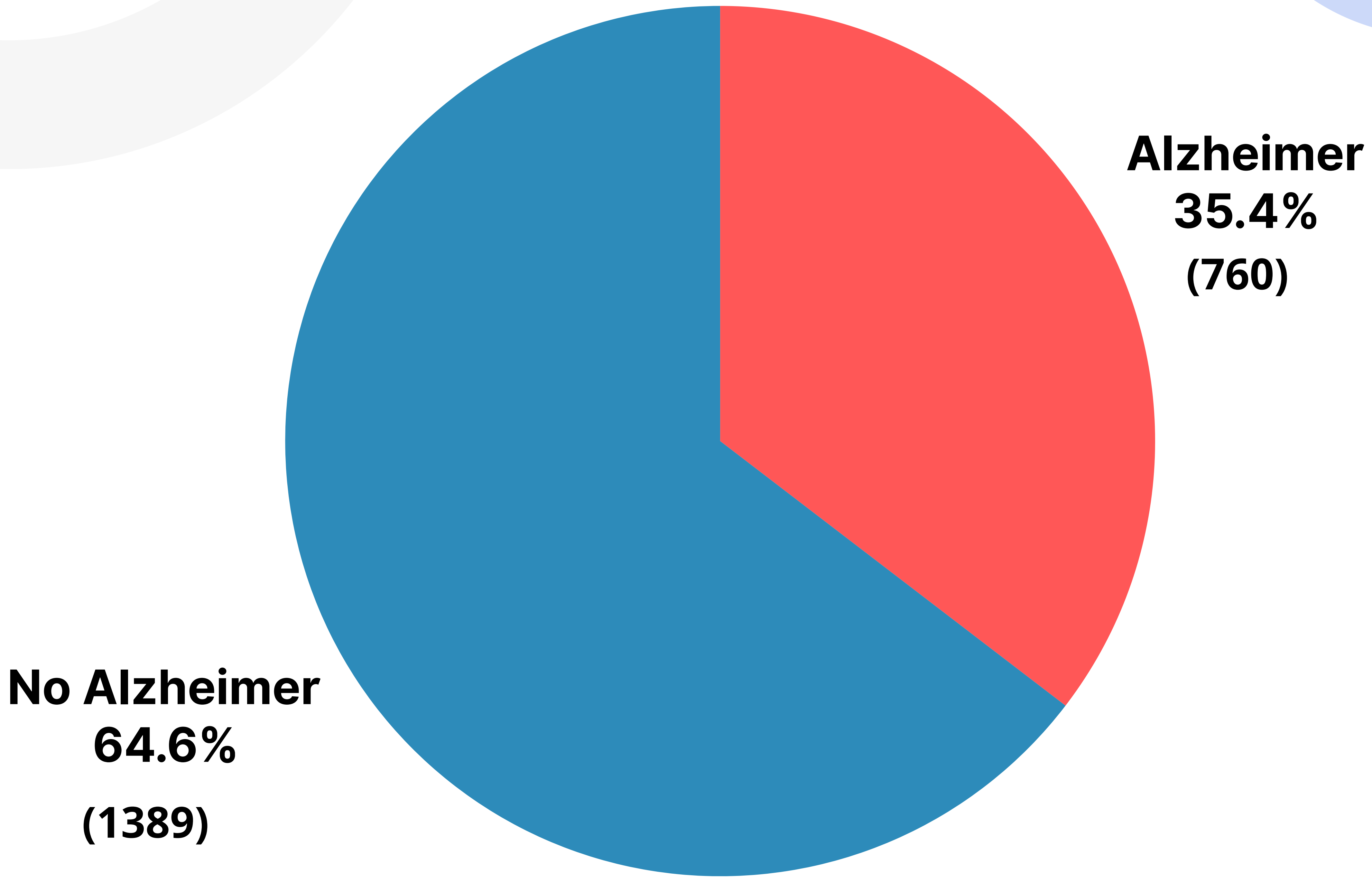
Data Description		
Patient ID	Medical History	Cognitive and Functional Assessments
Demographic Details	Clinical Measurements	Diagnosis Information
Lifestyle Factors	Symptoms	Confidential Information

Cognitive and Functional Assessments:

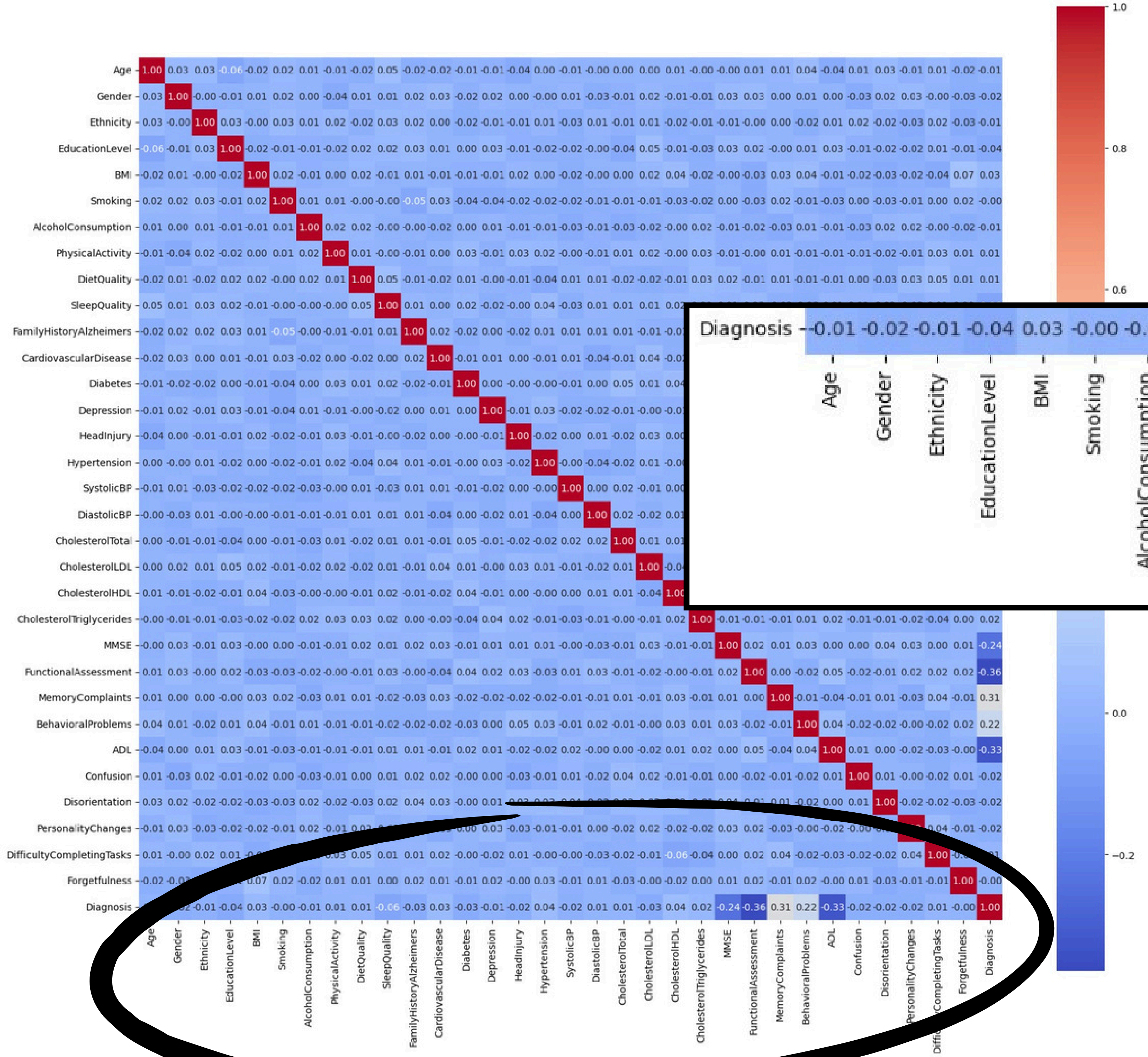
- **MMSE**
- **FunctionalAssessment**
- **MemoryComplaints**
- **BehavioralProblems**
- **ADL**

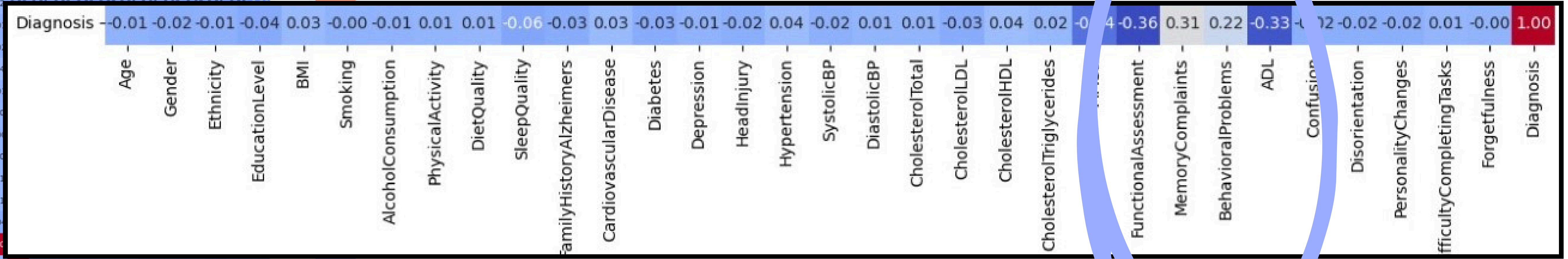
EXPLORATORY DATA ANALYSIS (EDA)

Pie Chart provides a clear visualization of the distribution between patients with and without the Alzheimer's Disease



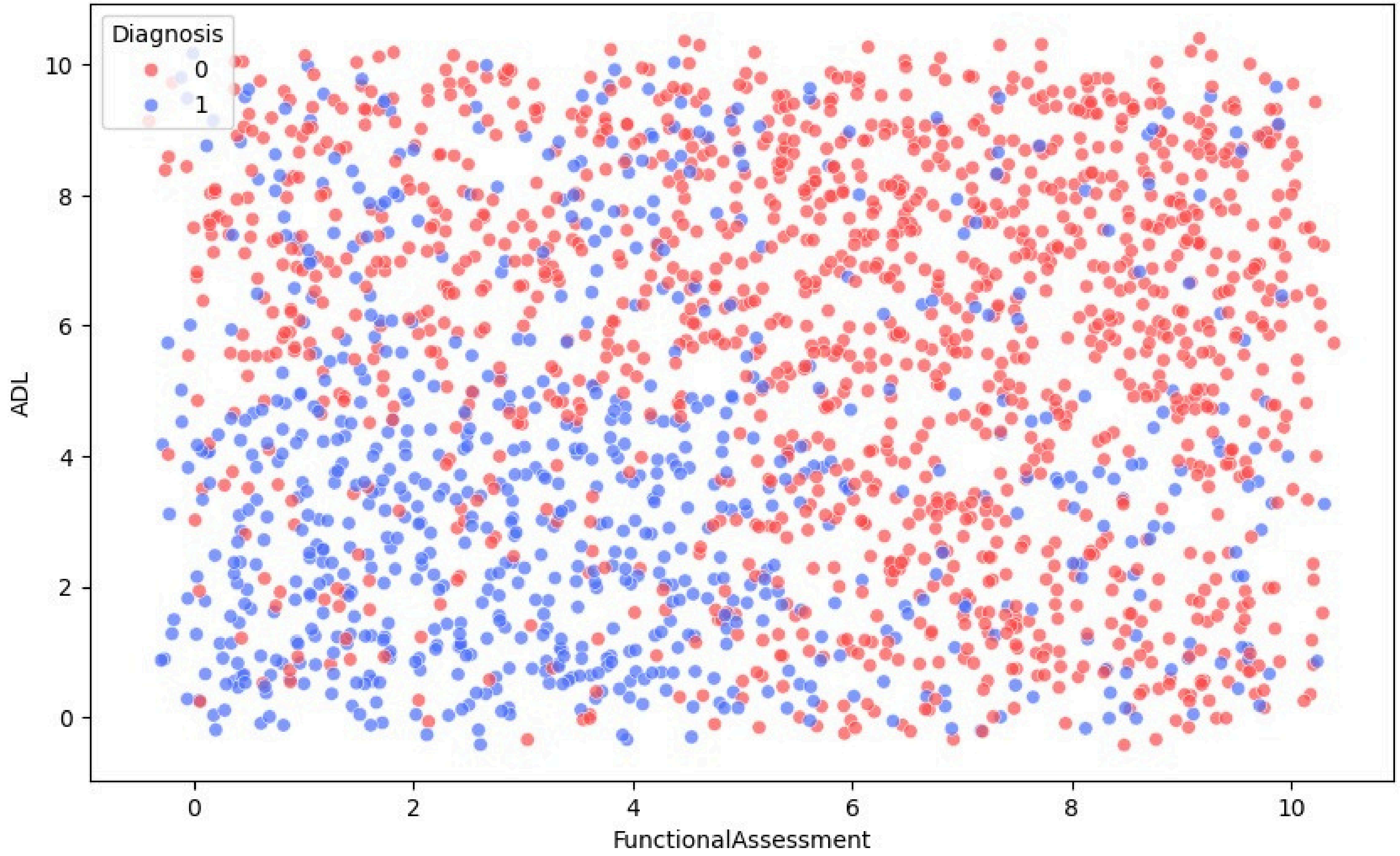
EXPLORATORY DATA ANALYSIS (EDA)





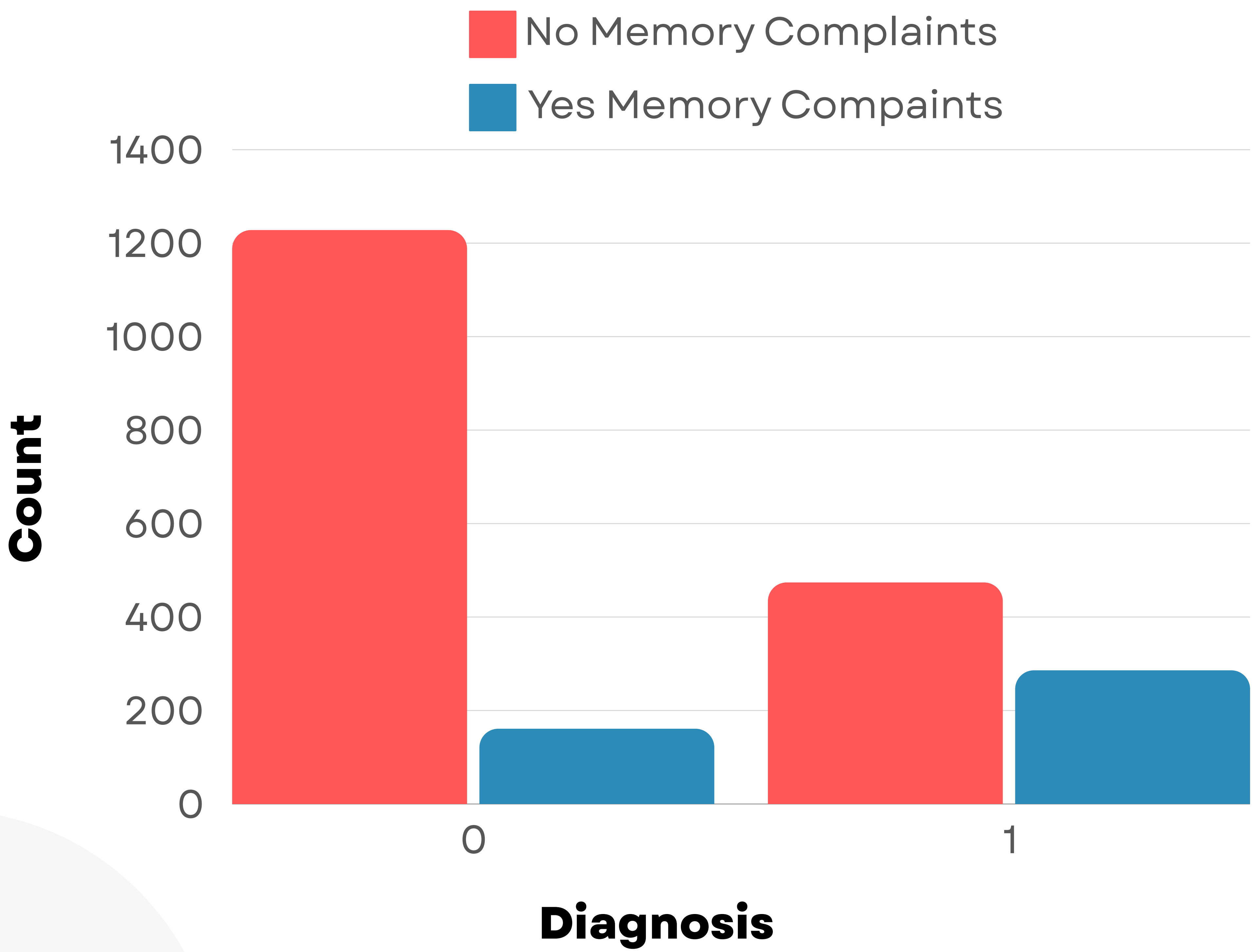
EXPLORATORY DATA ANALYSIS (EDA)

This graphic reveals a noticeable separation between the two groups based on “FunctionalAssessment ” and “ ADL ”



This visual representation strengthens the earlier finding from the correlation analysis , where these variables were moderately correlated with the “ Diagnosis ”

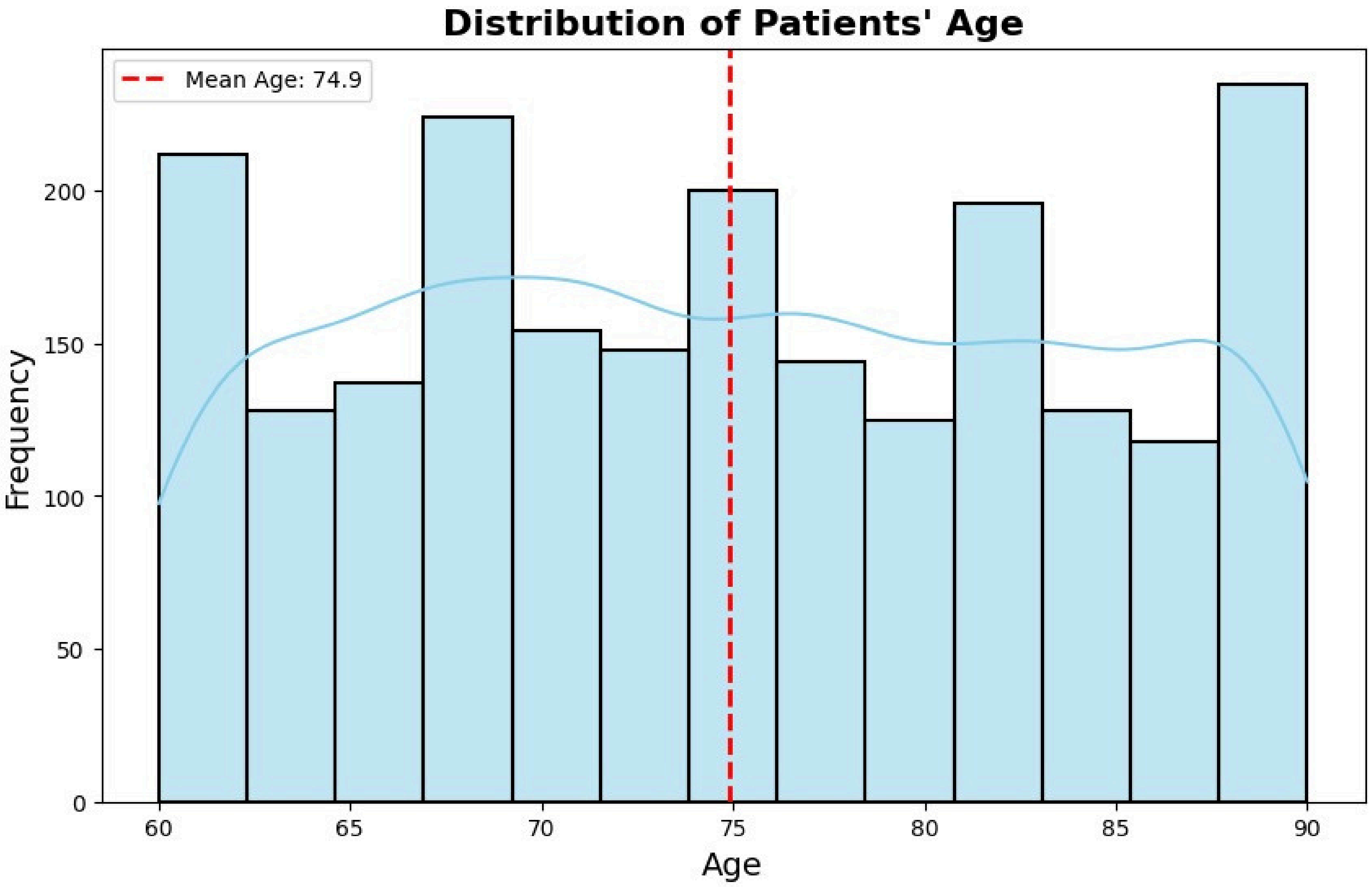
EXPLORATORY DATA ANALYSIS (EDA)



- **Patients without Alzheimer´s =no report memory issues= cognitive health**
- **Patients with Alzheimer´s = memory complaints = association between the disease and cognitive decline**

This pattern reinforces the understanding that Alzheimer's significantly impacts memory function.

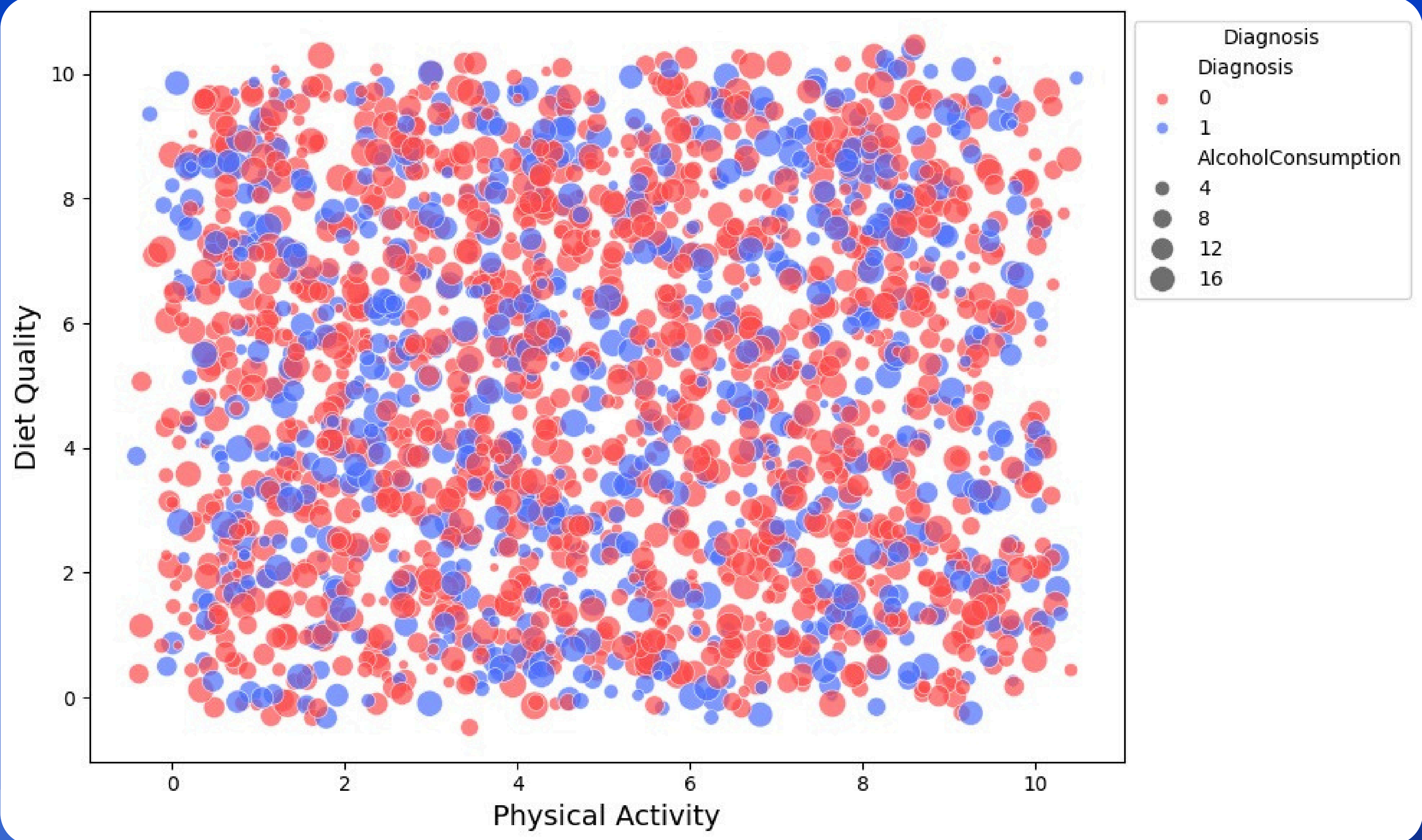
EXPLORATORY DATA ANALYSIS (EDA)

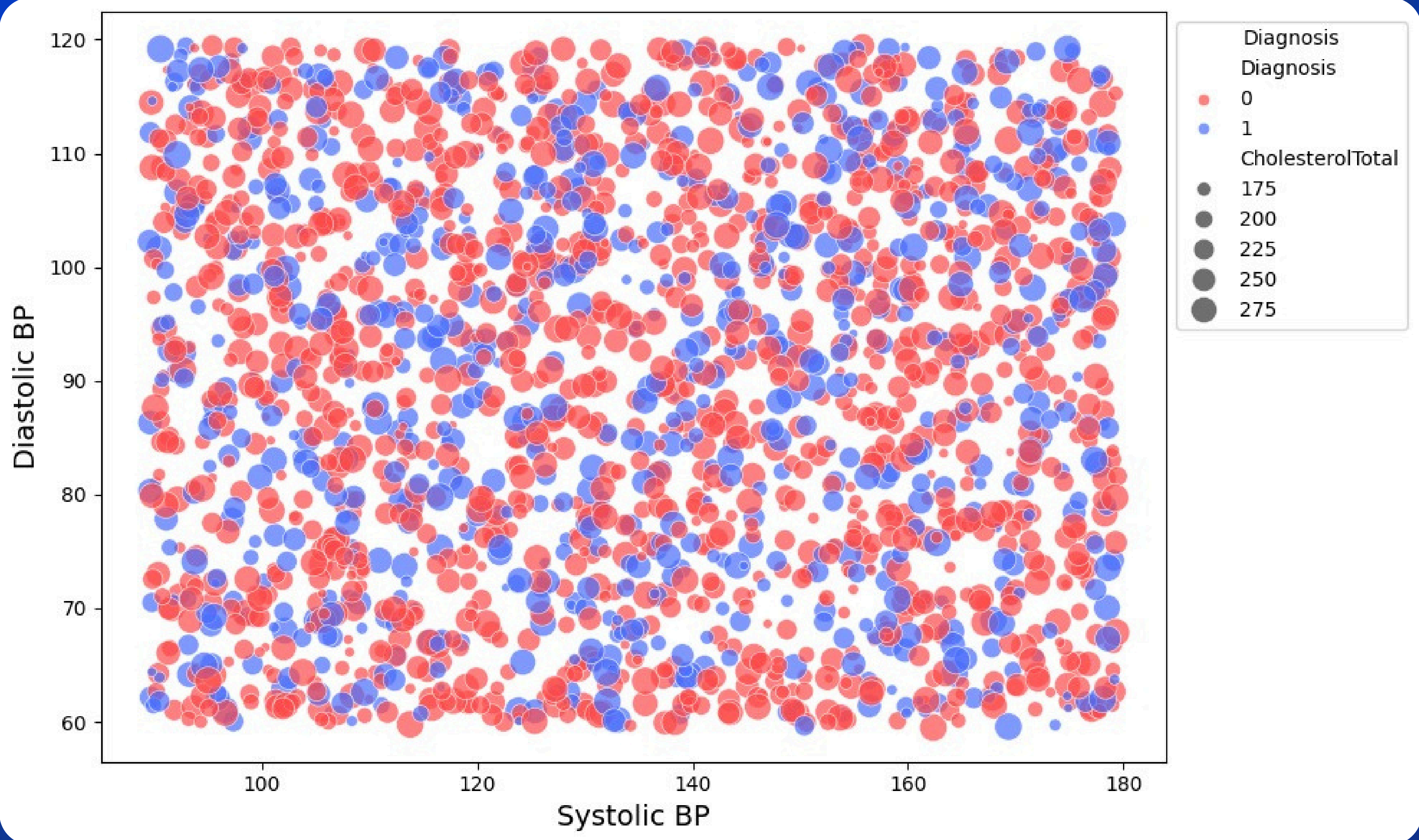


Histogram with a KDE (Kernel Density Estimate)

Note that there are several age groups with slightly higher frequencies, particularly at ages 60, 68 , 75 , 82 and 90

EXPLORATORY DATA ANALYSIS (EDA)





DATA PREPROCESSING

Data Cleaning

- No missing values were identified;
- Each feature´s data type was reviewed;
- Columns “PatientID” and “DoctorInCharge” were removed.

Data Splitting

- Dataset divided into training and testing subsets;
- Random state set to 42 to enable consistent results across experiments.

Data Balacing

- Synthetic Minority Oversampling Technique (SMOTE) was employed to ensure adequate representation of both classes.

Data Standardization

- Standardization of the input variables was performed using MinMaxScaler;
- Technique was applied to the training set using the *fit_transform* method.

MODEL CREATION

Firsst-Order Takagi-Sugeno

Developed using the PyFume library

One-layer network

Implemented with the default settings of the Multi-Layer Perceptron (MLP)

Two-layer network

Configured with two hidden layers of 100 neurons each

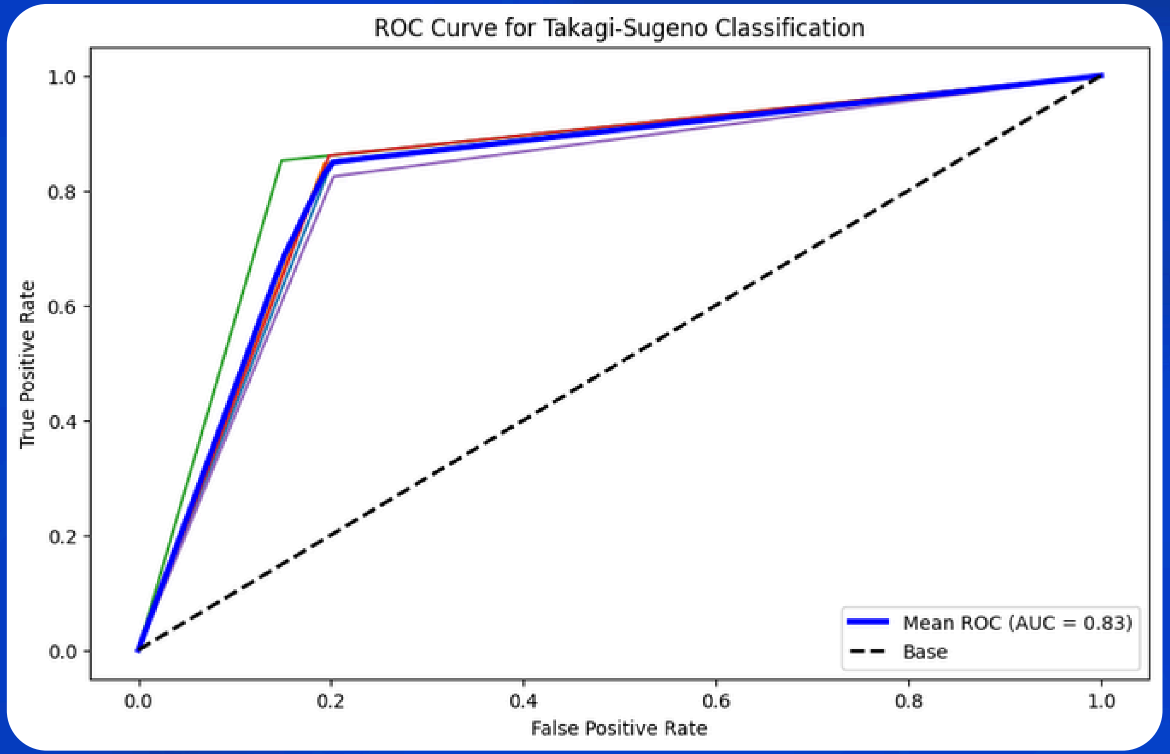
SVM model

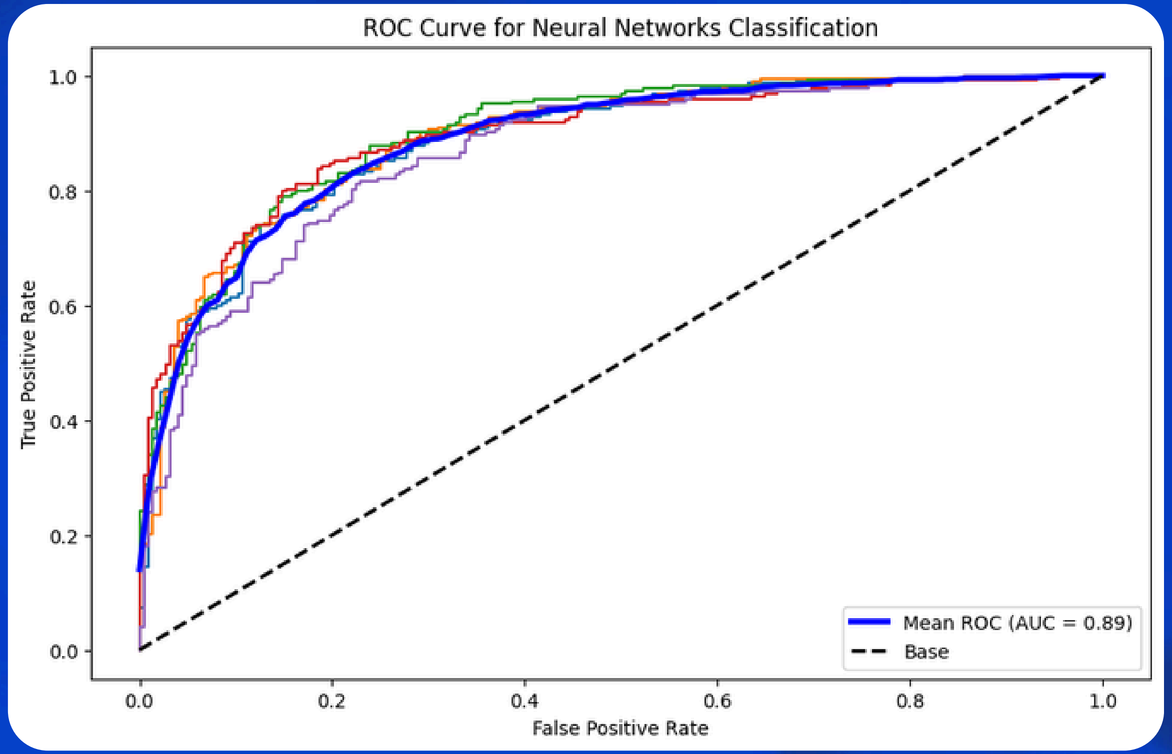
Established with the 'probability' parameter enabled to allow for probability estimation

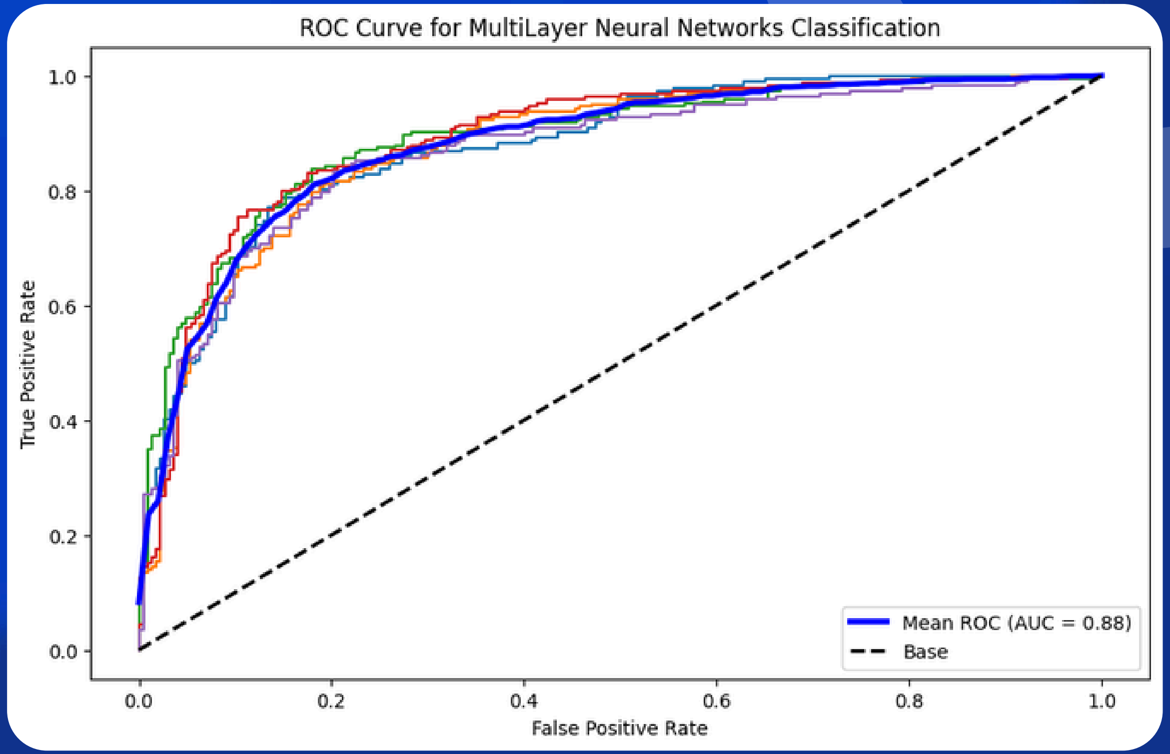
Decision Tree

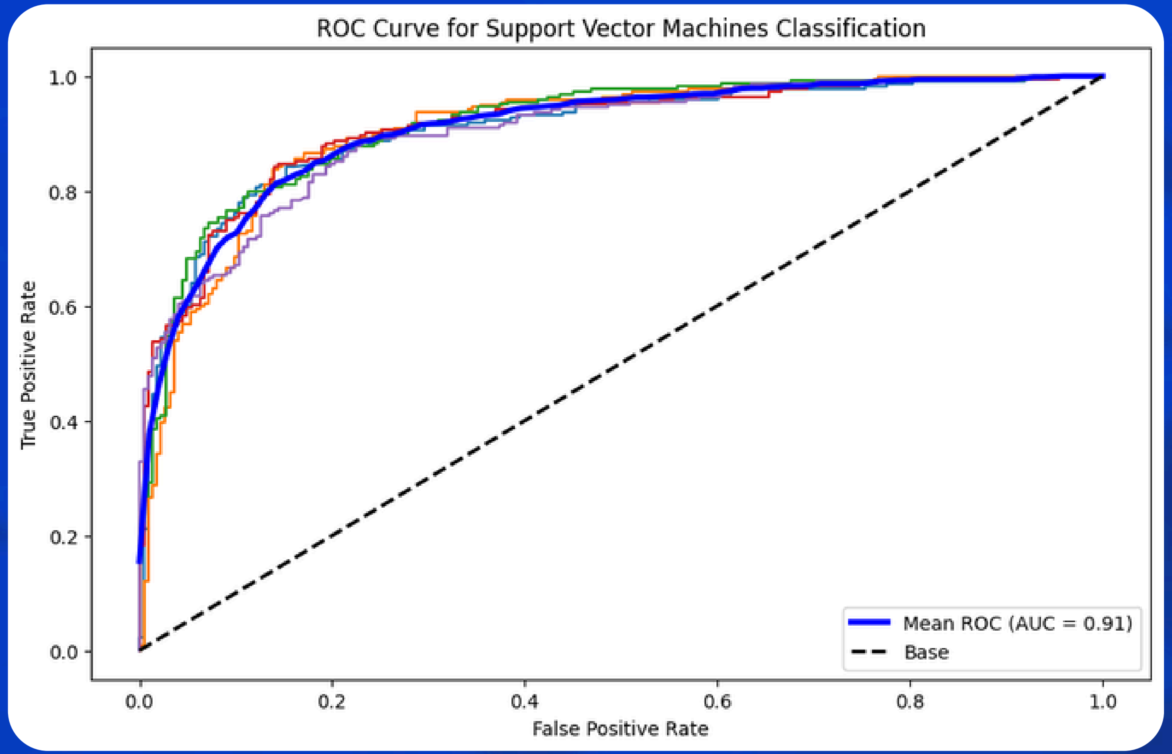
Implemented with standard settings for straightforward interpretability.

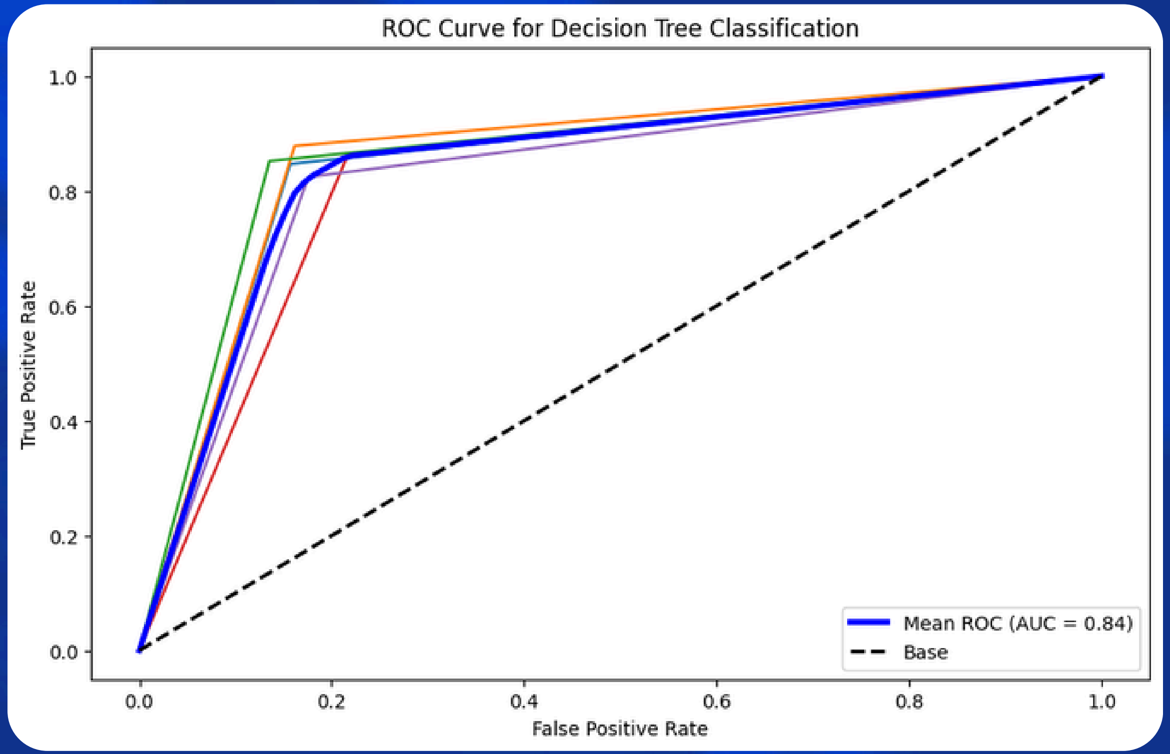
K-FOLD CROSS-VALIDATION











HYPERPARAMETERS TUNING AND FEATURE SELECTION

Models	Hyperparameters	Results
First-Oder Tagaki-Sugeno	Number of Clusters	5
	Distance Metric	‘euclidean’
	Max Clustering Iterations	100
	Membership Function Shape	'gauss’
	Global Fit	False

Model	Hyperparameters	Results
One-Layer Neural Network	Learning Rate	0.001
	Activation Function	'tanh’
	Hidden Layer Size	(50,)

HYPERPARAMETERS TUNING AND FEATURE SELECTION

Models	Hyperparameters	Results
Two-Layer Neural Network	Learning Rate	0.001
	Activation Function	'tanh’
	Hidden Layer Sizes	(100, 100)

Model	Hyperparameters	Results
Support Vector Machines	Regularization Parameter	10
	Kernel Type	'rbf’

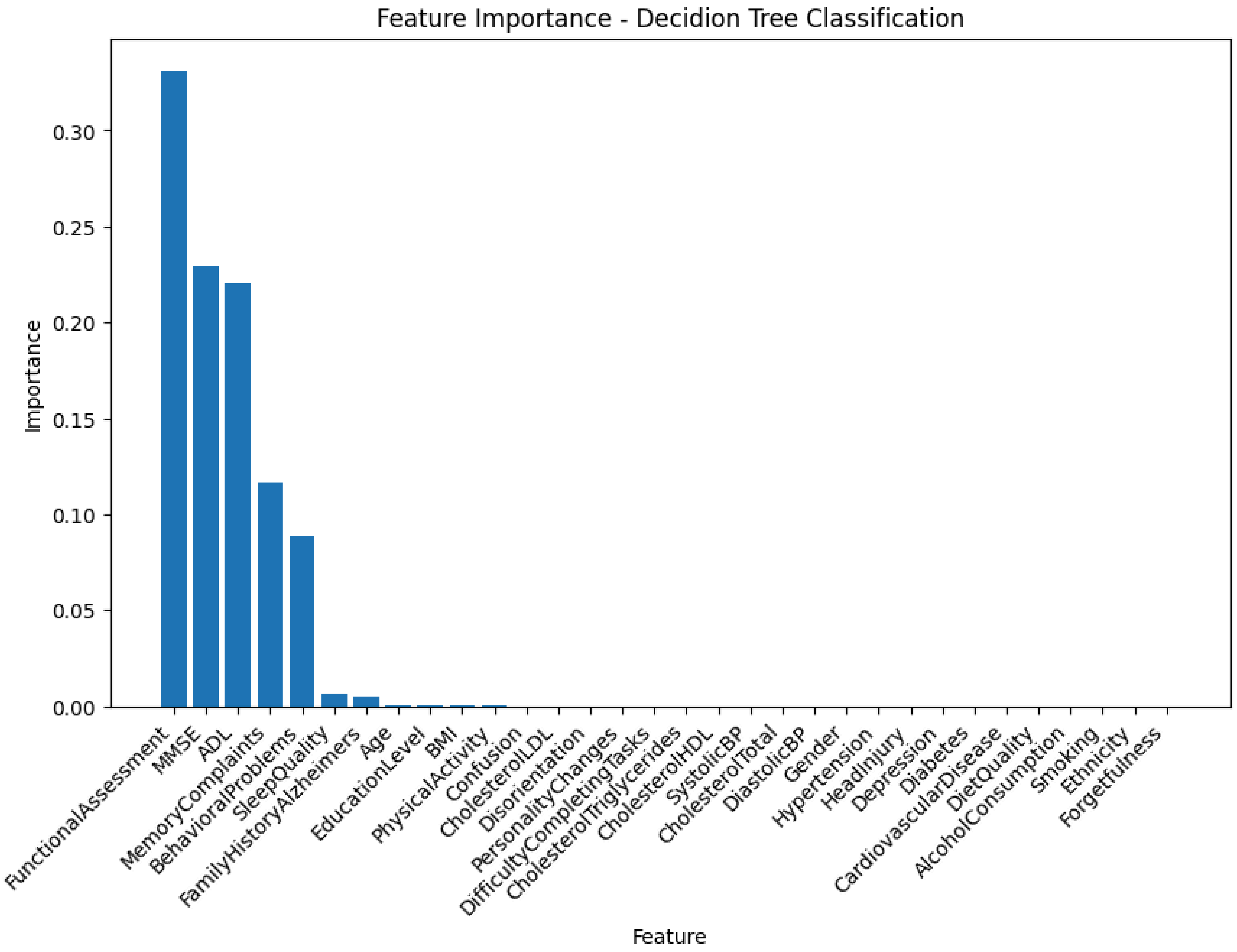
Model	Hyperparameters	Results
Decision Tree	Maximum Depth	None
	Minimum Samples Split	0.1

TEST AND COMPARE

Predictions for each model were generated and stored in corresponding variables. The performance metrics provide an objective measure of the model's effectiveness in a specific task, enabling informed choices during development.

	Metric				
		Accuracy	Precision	Recall	F1 Score
Model	First-Order Takagi-Sugeno	0.79	0.68	0.78	0.73
	One-Layer Neural Network	0.77	0.66	0.76	0.70
	Two-Layer Neural Network	0.80	0.69	0.77	0.73
	Support Vector Machines	0.78	0.69	0.69	0.69
	Decision Tree	0.93	0.93	0.88	0.90

FEATURE IMPORTANCE



01 Decision Tree model is evaluated using the *feature_importances_ attribute*

02 The importance values are extracted into a *DataFrame*

03 Key contributors to the model's predictive accuracy are cognitive and functional assessments

04 “**FunctionalAssessment**” is the most important feature, followed by **“MMSE”**, “**ADL**”, and “**MemoryComplaints**”

CONCLUSION

Among the models evaluated, the Decision Tree consistently outperformed others across all metrics, likely due to its adaptability to binary features within the dataset.

Exploring additional feature selection methods, such as SHAP, could provide more nuanced insights into non-linear and interaction effects between features.

Expanding the model comparison to include ensemble methods like Random Forests or Gradient Boosting could provide additional insights into predictive accuracy.

Moreover, incorporating additional data sources, such as imaging data or genetic markers, , could add layers of depth to the model’s diagnostic capabilities

THANK YOU

FOR YOUR ATTENTION