Medical Imaging Informatics
2020/2021

University of Minho
School of Engineering

Discover of Outcome predictors for MCI condition using MRI

João Alves 83624
Paulo Alves 84480

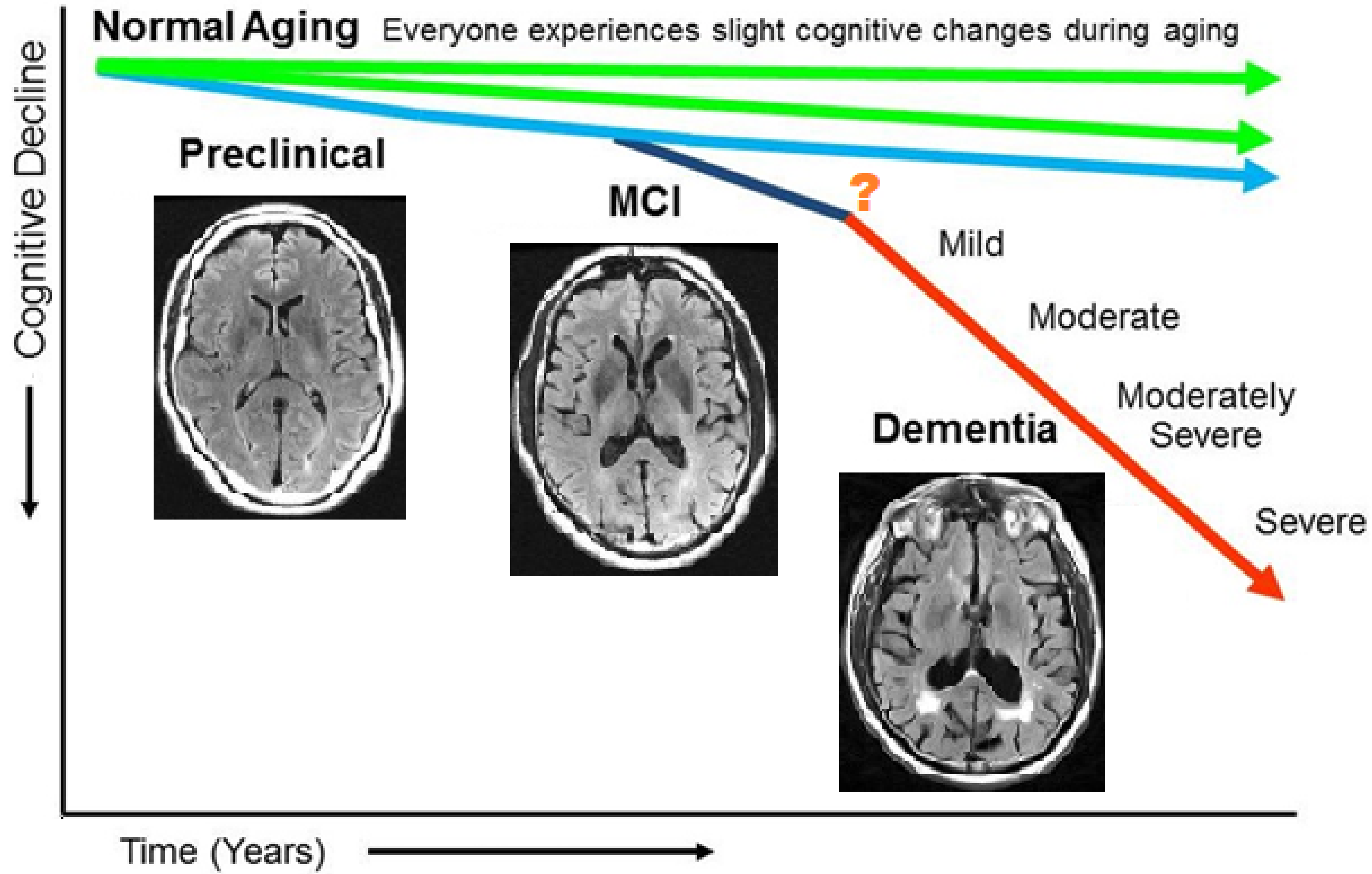
SUMMARY

DISCUSSION POINTS

- Introduction
- Dataset and subject recruitment
- Imaging
- Pre-Processing
- Features Extraction (Radiomics)
- Clinical Data
- Machine Learning models and techniques
- Results
- Discussion

Introduction

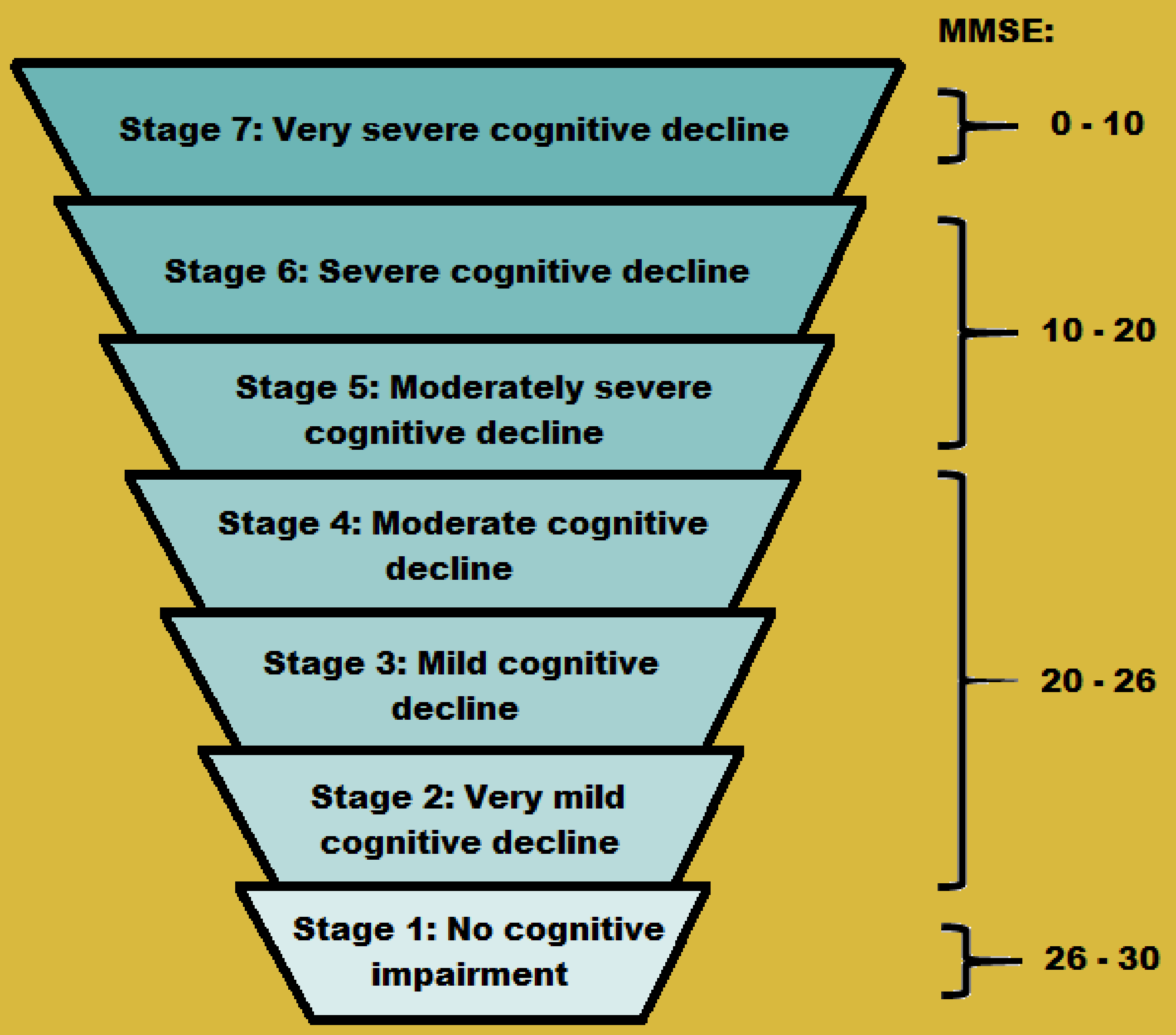
Alzheimer's disease (AD) is the most common form of progressive and irreversible dementia, and accurate diagnosis of AD at its prodromal stage is clinically important.



Dataset and subject recruitment

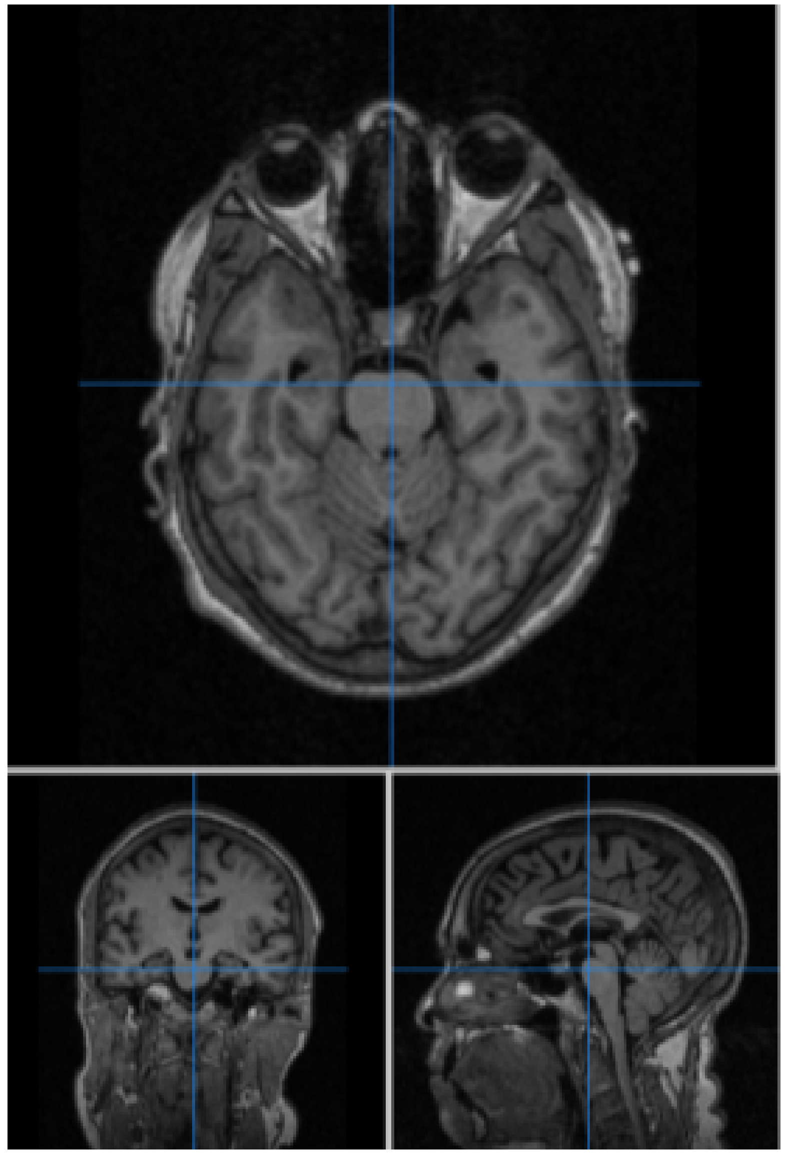
- All exams were extracted from **ADNI**.
- The metric used was **MMSE** (Mini Mental State Examination).
- The exams were obtained 6 months, 12 months, 24 months and 36 months after the first one.
- All the images were acquired using *B1-calibration_Body*, *B1-calibration_Head*, *MPRAGE_br*, *MPRAGE_Repeat*.
- 386 exams were used from 22 different patients.

ADNI
ALZHEIMER'S DISEASE
NEUROIMAGING INITIATIVE



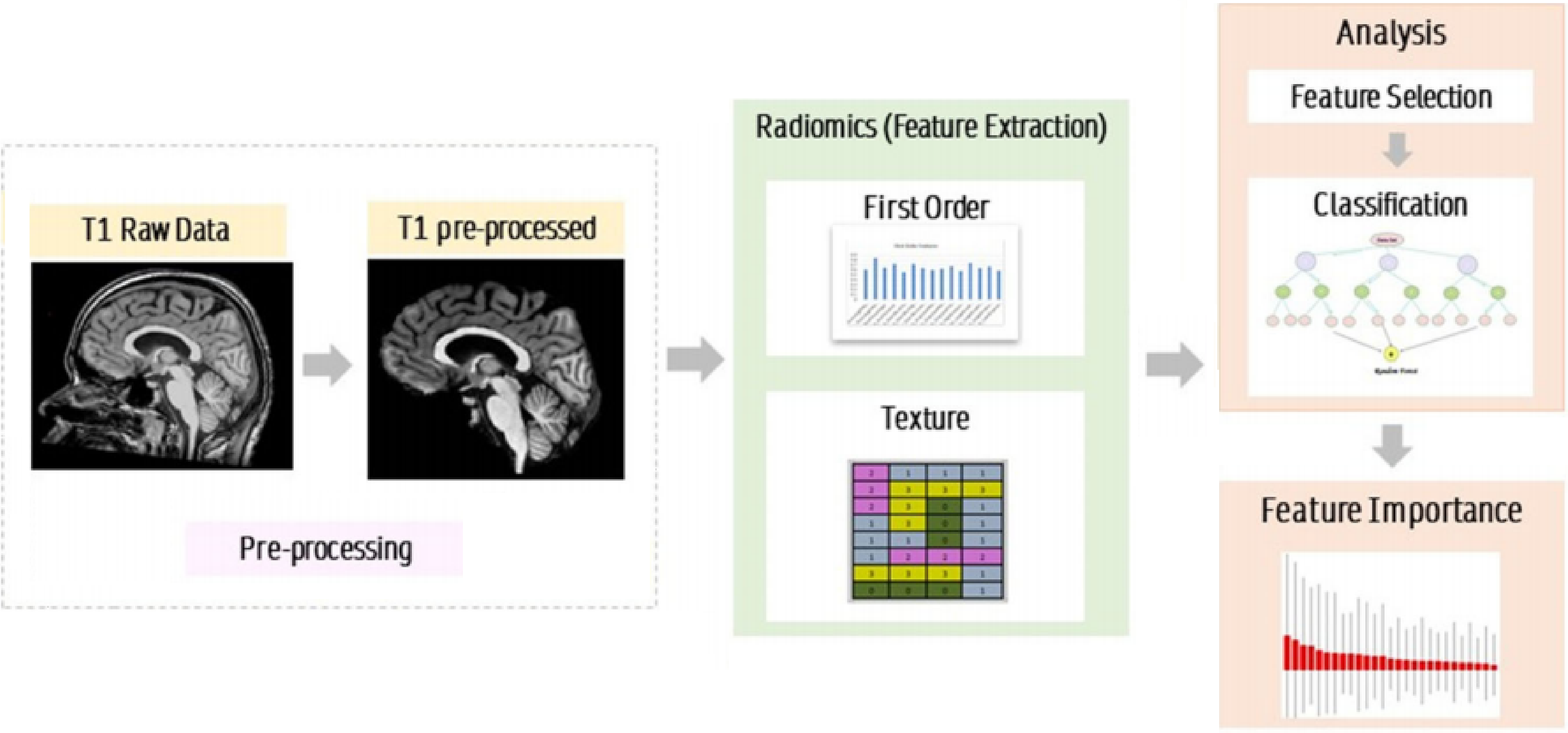
Imaging

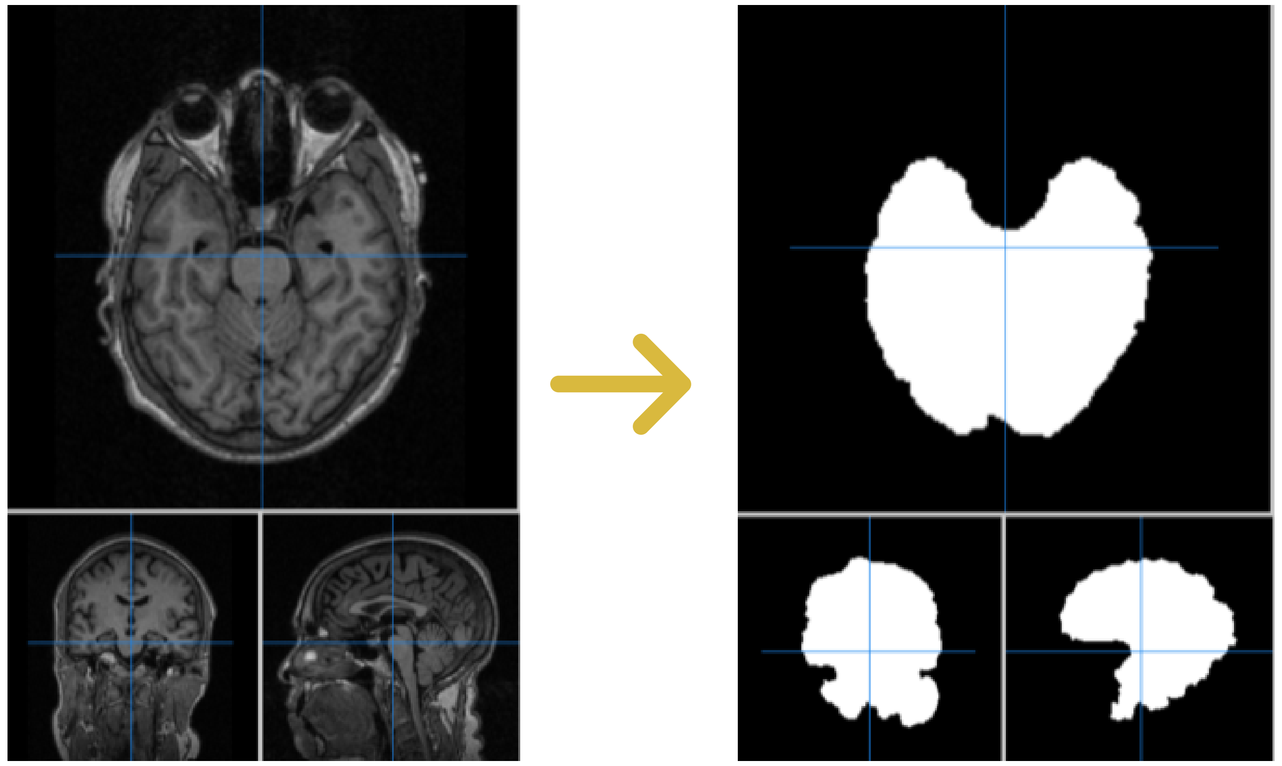
The images were acquired on T1-weighted MRI using different machines and manufactures. High-resolution images were acquired with a voxel size = 1 × 1 × 1 mm, slice thickness = 1 mm, and they were taken in sagittal, axial and coronal slices.



Pre-Processing

The brain extracted non-uniform intensity corrected image (nu.mgz) was obtained by using the FSL command, Brain Extraction Tool (BET).



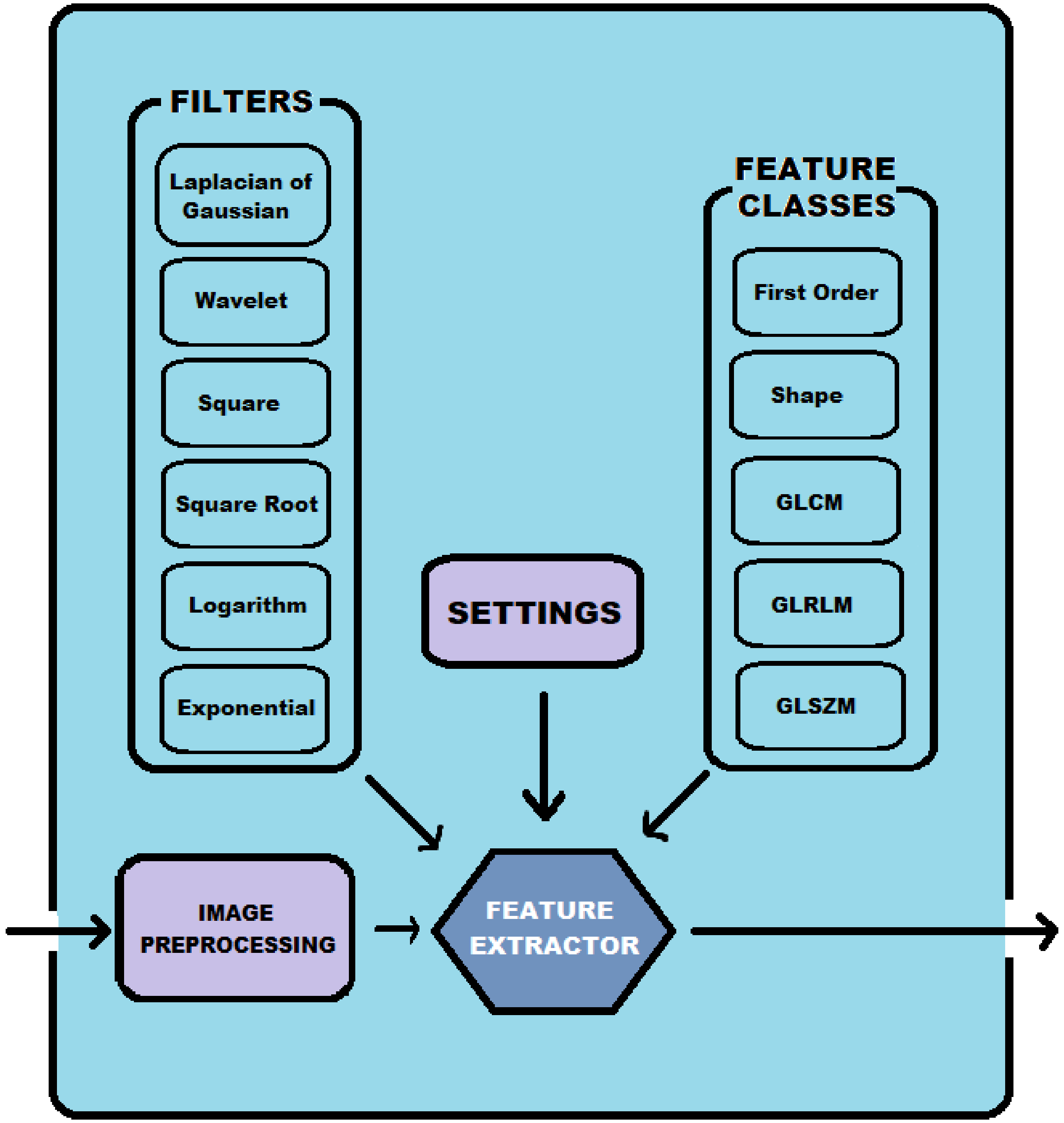


Feature Extraction (Radiomics)

- All subclasses of radiomics were extracted.

- **Filters:**
Original, Wavelet, LoG, Sigma with values of [1.0, 2.0, 3.0, 4.0, 5.0], Square, SquareRoot, Logarithm, Exponential, Gradient, LBP2D,LBP3D.

- Initially, there were 1899 features. Elimination of all the features that weren't numeric. In the end, **1823 numeric features were extracted.**



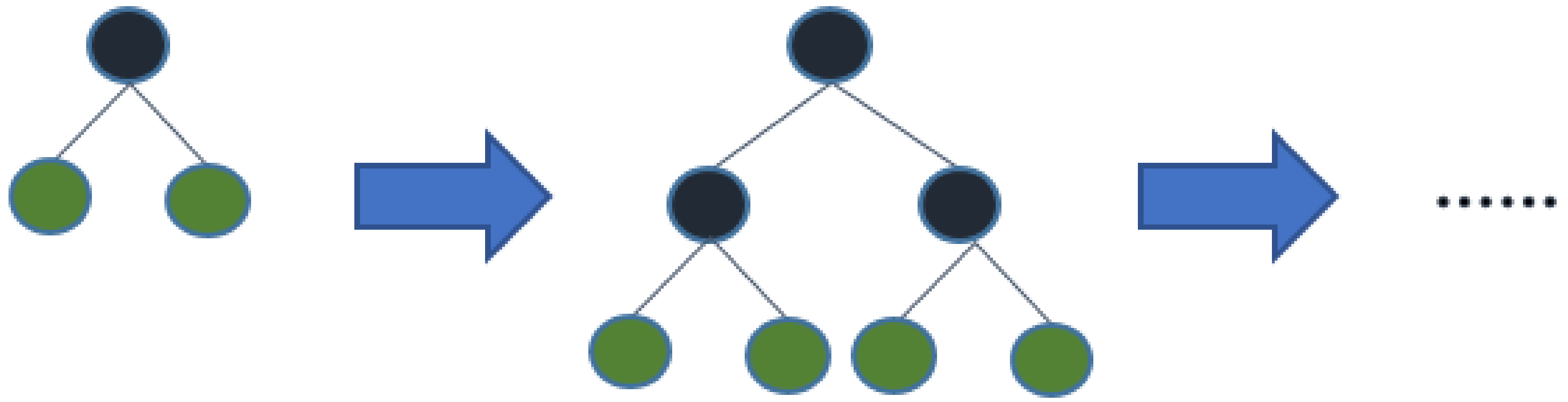
Clinical Data

		CN (normal)	MCI	AD
MMSE Score		26-30	21-26	10-21
Total Count		90	223	73
Count by Gender	Male	73	114	40
	Female	17	109	33
Age Range		62.0 – 87.8	62.3 – 89.7	73.6 – 90.3

Clinical data and information about the exams

Machine Learning Models

XGBoost

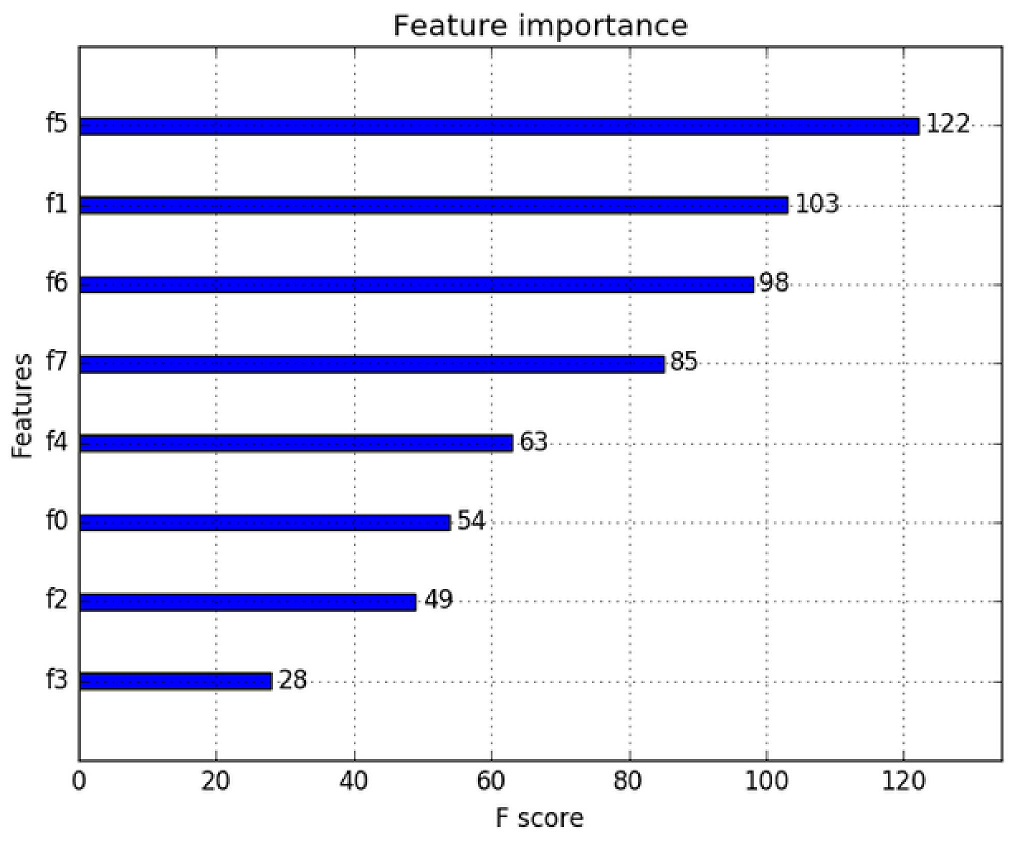


- Designed to be highly efficient, flexible and portable;
- It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way;
- It is an implementation of gradient boosted decision trees designed for speed and performance.

Machine Learning Techniques

Feature Importance

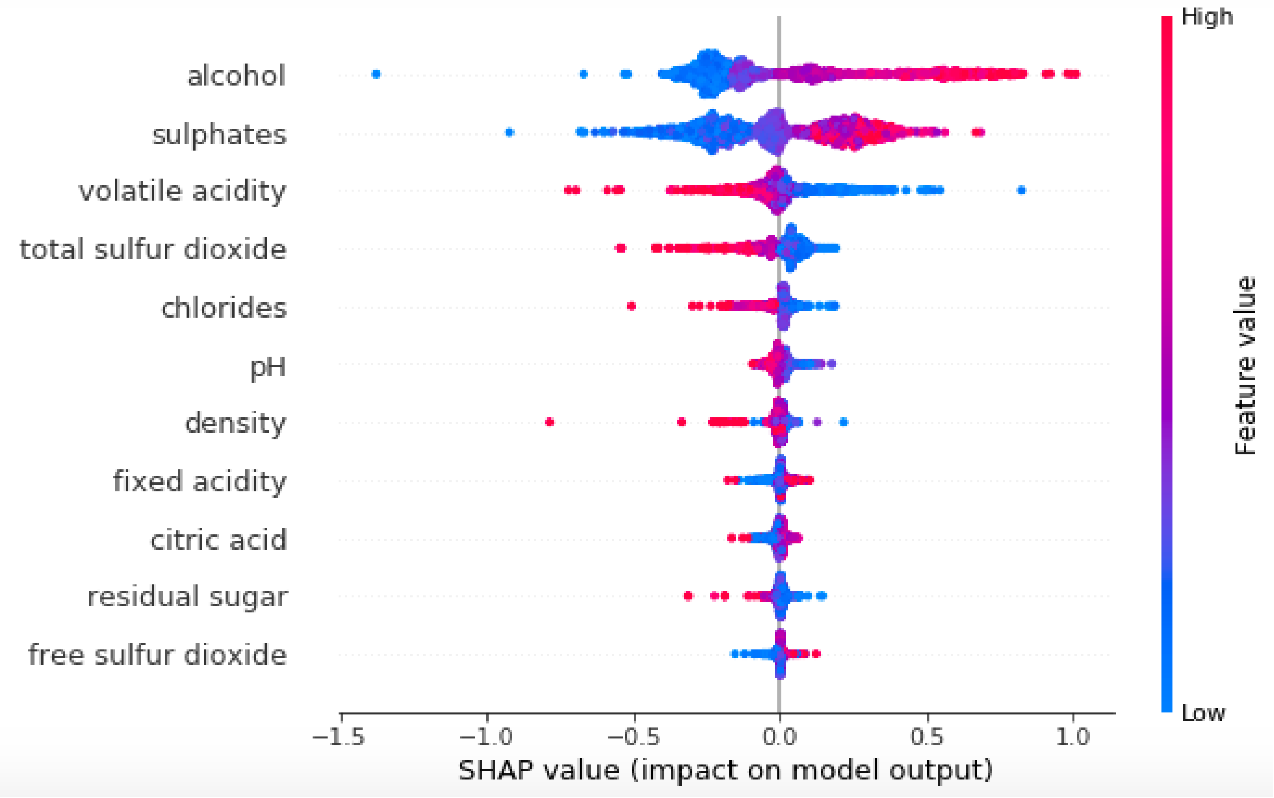
- Can be used to understand which are the most important features to the specific model;
- Assigns a score to input features based on how useful they are at predicting a target variable;
- Provides insight into the data and the model;
- Dimensionality reduction and feature selection that can improve the efficiency and effectiveness of a predictive model.



Machine Learning Techniques

SHAP Values (SHapley Additive exPlanations)

- Based on game theory;
- Looks at the contribution of a model's feature for each patient (local interpretability);
- Global importance also possible by averaging across patients;
- Contribution is based on shapley value – this is the total impact of a feature when considering all combinations of other features.

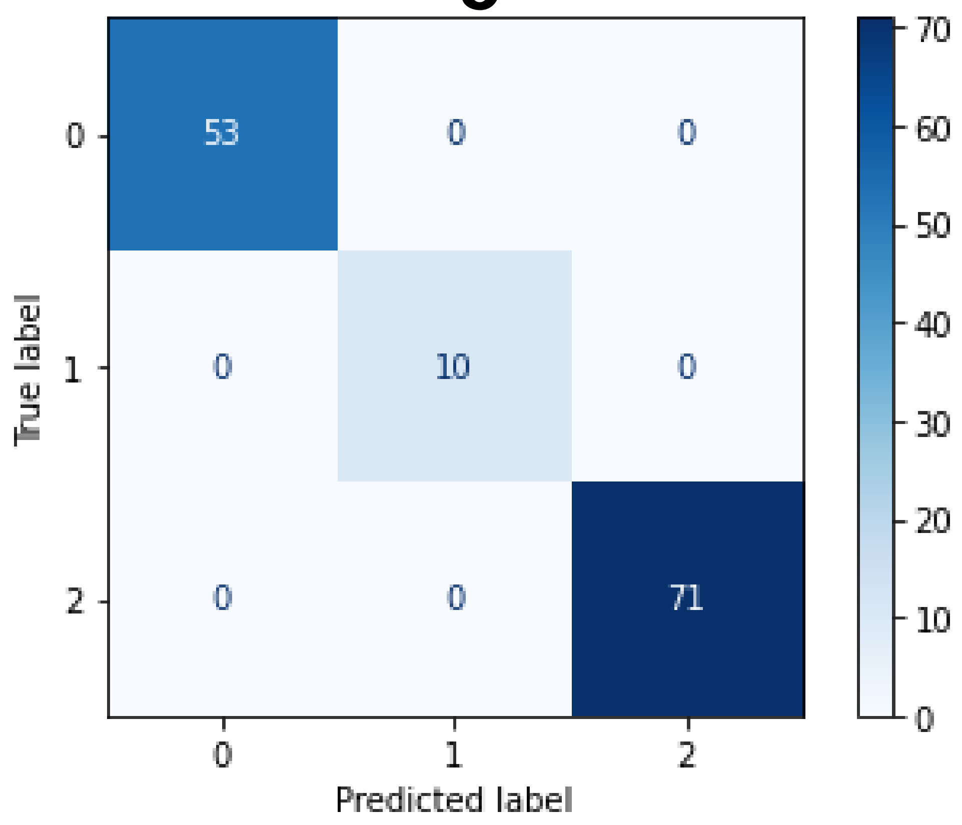


Results

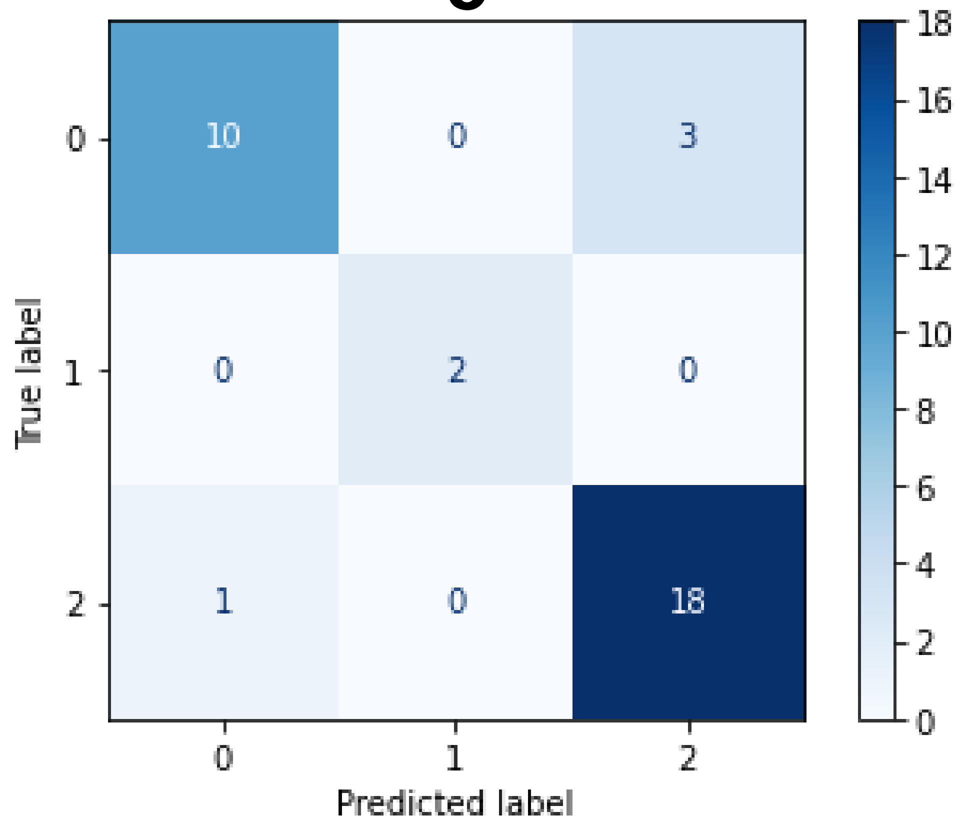
Transitions

- The dataset had **168 exams** in which the patient had transited from a condition to another;
- It was split into **80% for training** and **20% for testing**;
- **3 classes**: Class 0: CN to MCI; Class 1: CN to AD; Class 2: MCI to AD.

Training Matrix



Testing Matrix



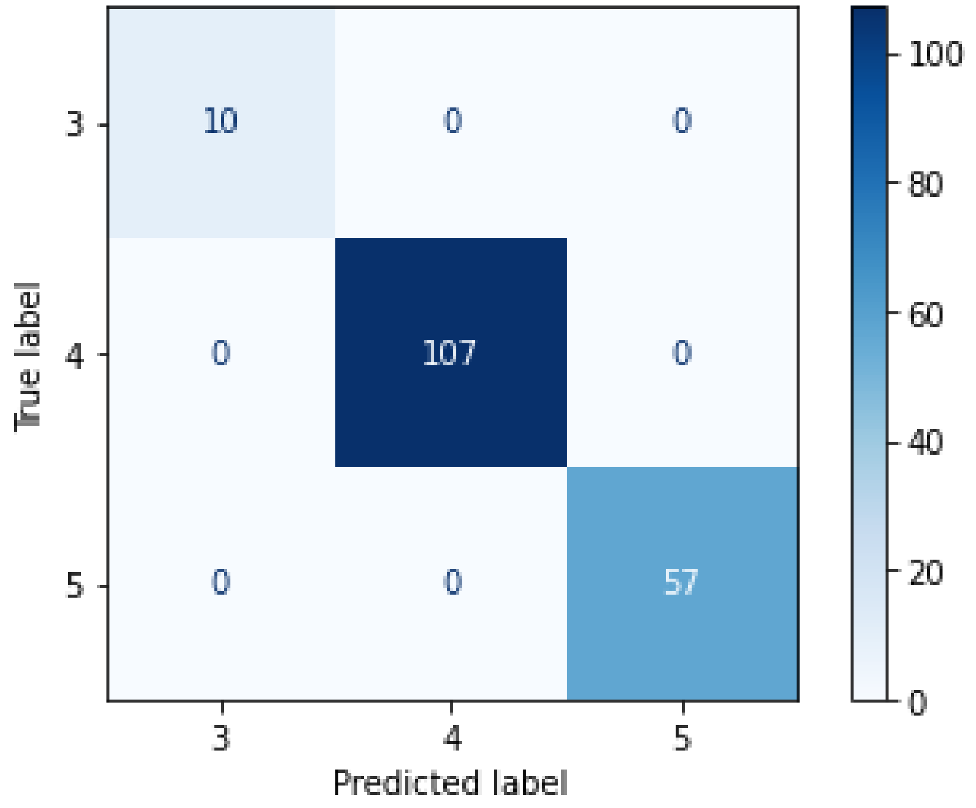
Accuracy = 88.24%

Results

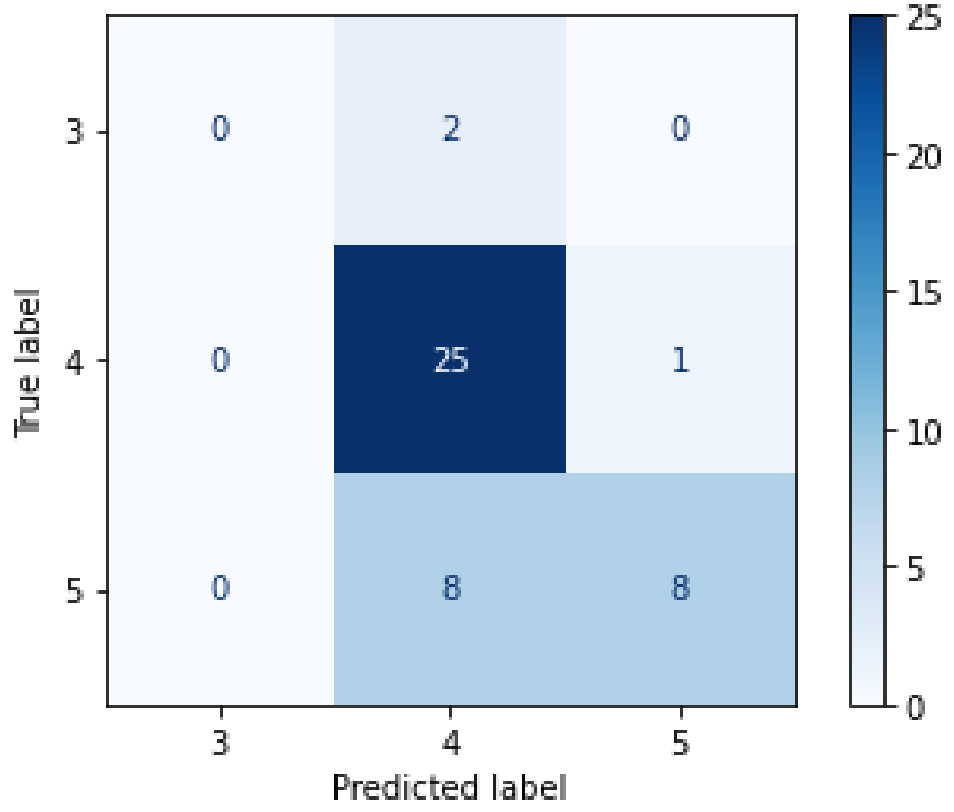
No Transitions

- The dataset had **218 exams** in which the patient hadn't transited from a condition to another.
- The dataset was split into **80% for training** and **20% for testing**.
- **3 classes:** Class 3: CN to CN; Class 4: MCI to MCI; Class 5: AD to AD.

Training Matrix



Testing Matrix

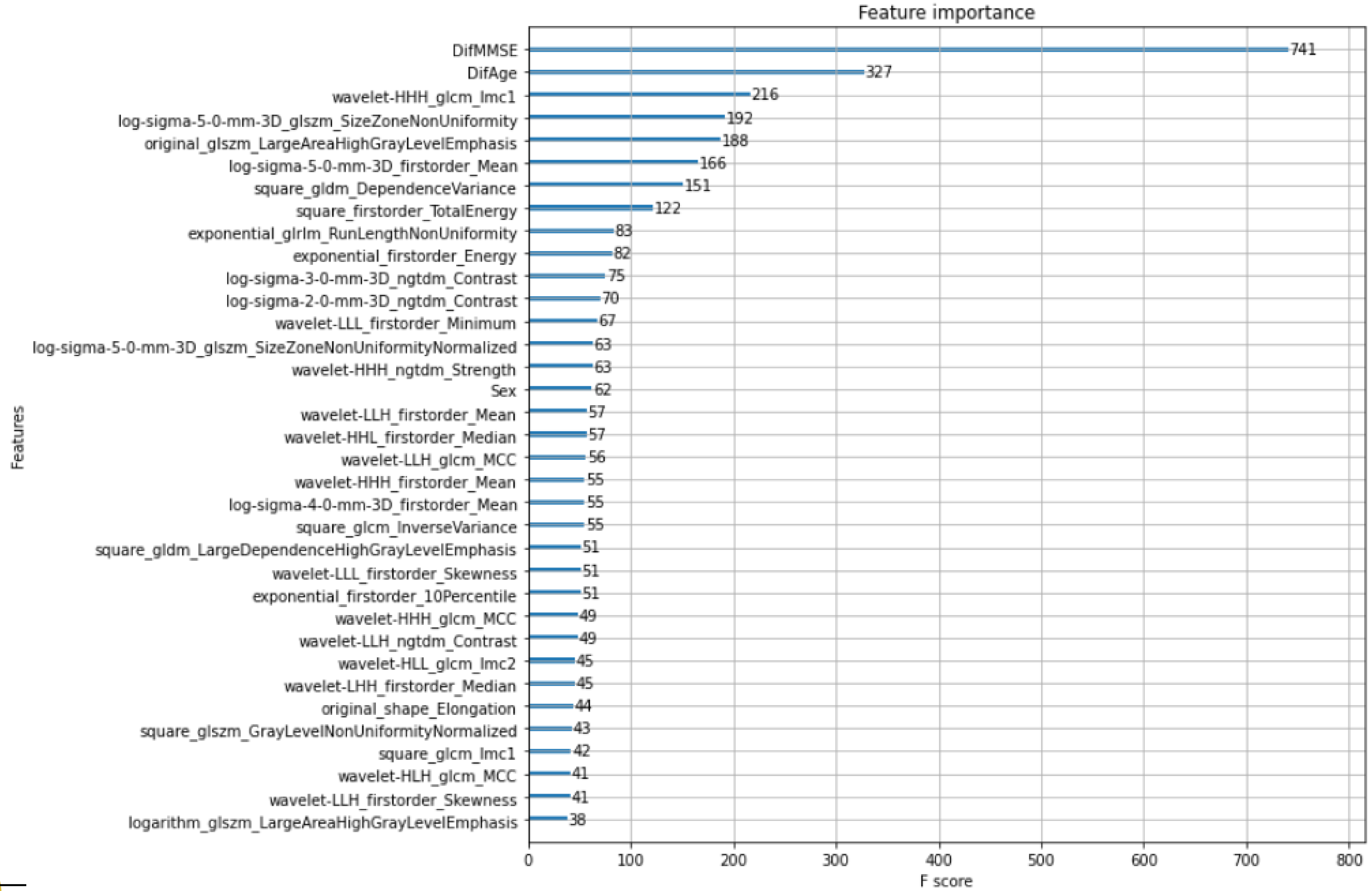


Accuracy = 75%

Discussion

Feature Importance

The most relevant features in descending order of F score.



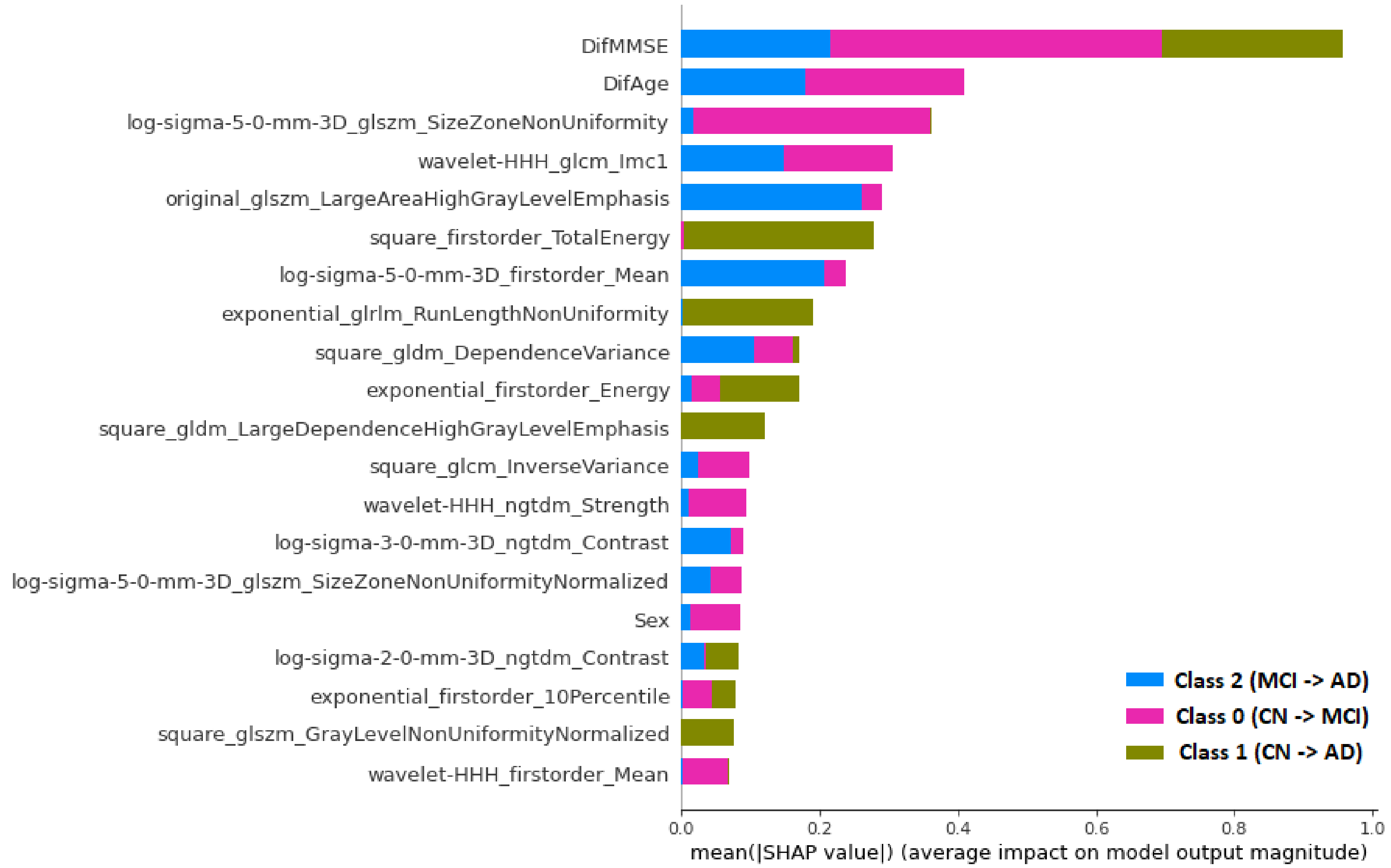
- 35 features displayed, as the minimum score of a specific feature was 38.
- All the features that had a worse f score than 38 were dismissed, because their score quickly stagnated.

Discussion

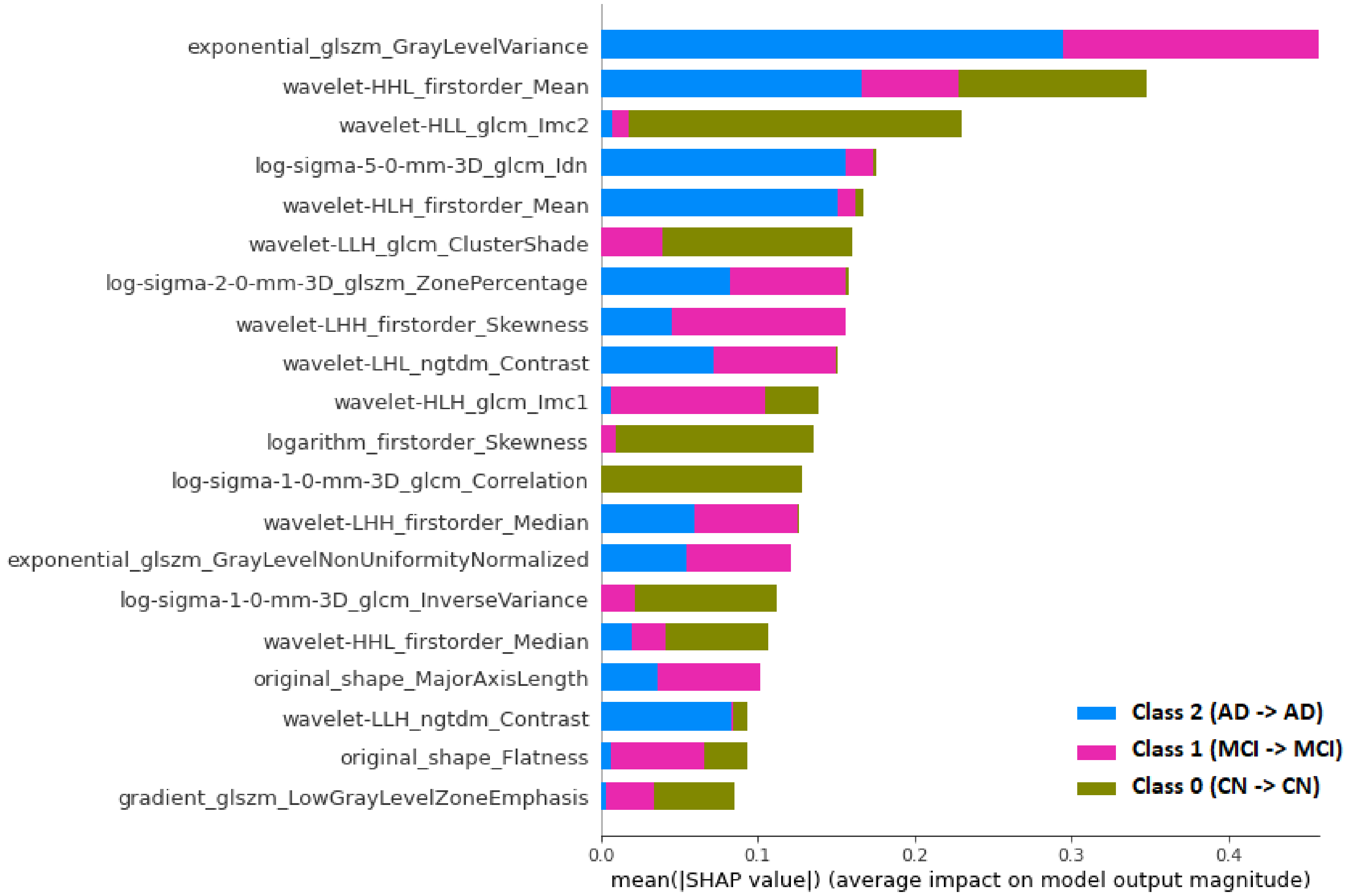
Shap Values

What were the most important features for all the classes?

TRANSITIONS



NO TRANSITIONS

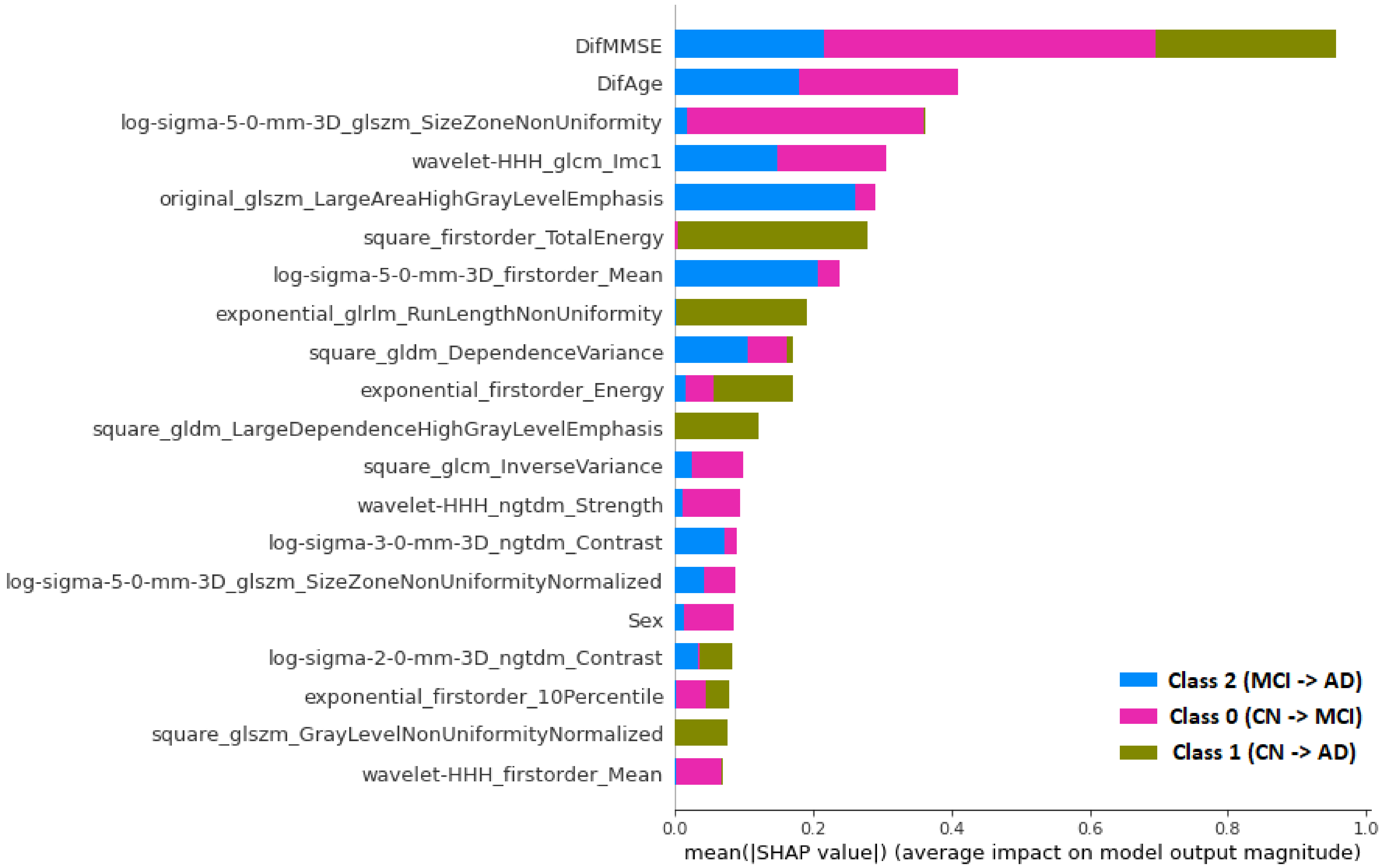


To discover what were the important features in the transition from CN to MCI and from CN to AD, it was needed to understand what were the important ones to patients that didn't transit from CN to something. The same logic was applied to discover the predictors of MCI to AD.

Discussion

Shap Values

What were the most important features for all the classes?



Conclusion

- The present study establishes the utility of discovering **predictors** for MCI condition based on **radiomics features** extracted from **T1- weighted images**.

- Future studies could aim to examine higher order local and global spatial relationships between **pixels** and relate them to underlying **pathological** microstructural changes.

- **Shap values** and **feature importance** were crucial to the development and discussion of the final results to achieve the main purpose of this project.

- This information will help **doctors** and **clinical staff** on the decision making and consequently, it will help **patients' lives**.

Thank you!

João Alves (joaomsalves26@gmail.com)
Paulo Alves (paulojorgealves18@gmail.com)