Discover of Outcome predictors for MCI condition using MRI

João Miguel da Silva Alves[83624] and Paulo Jorge Alves[84480]

University of Minho, Campus Gualtar, Braga 4710, Portugal

Abstract. This study aimed to explore the feasibility of radiomic features extracted from T1-weighted MRI images to differentiate patients who have Mild Cognitive Impairment (MCI) and will evolve into Alzheimer's disease (AD) from those who won't. Radiomic features were computed from T1 images of patients with AD and MCI, and patients that were healthy. These features were extracted from the brain. Separate classifiers were applied to classify different groups based on a reduced set of most important radiomics features for each classification as determined by the Multilayer Perceptron and XGBoost. The classifiers illustrated an accuracy of 88% for the exams in which the patient had transited and an accuracy of 75% for the exams in which the patient hadn't transited. This study establishes the utility of radiomics to differentiate patients who will evolve into AD using routine T1-weighted images. This may aid in the clinical diagnosis of patients who have MCI, which can have severe consequences in the future.

Keywords: Radiomic Features · T1-weighted MRI · Mild Cognitive Impairment · Alzheimer's disease.

1 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.

Mild cognitive impairment can possibly increase our risk of later developing dementia or other neurological condition. But not always people get worse, they can get better if the diagnosis by the physician is done properly. There is no single cause of mild cognitive impairment, just as there is no single outcome for this disorder. Its symptoms may remain stable for years, or progress to Alzheimer's disease or even improve over time. Changes that come with this condition include abnormal clumps of beta-amyloid protein(plaques) and protein clumps of tau characteristic of Alzheimer's disease (tangles), Lewy bodies, which are microscopic clumps of another protein associated with Parkinson's disease and small strokes or reduced blood flow through brain blood vessels.[8] By using Brain-Imaging, we can see differences and changes that may be associated with MCI such as, shrinkage of the hippocampus, a very important region of the brain which is responsible for memory capacity, enlargement of the brain's fluid-filled

spaces (ventricles) and reduced use of glucose, the sugar that is the primary source of energy for brain's cells. The problem with this condition is how to discover if a certain individual with MCI has a significant probability to develop dementia. To do so we need to find and track down predicators. Although Medical Informatics is a field which we can say it's new, a lot of breakthroughs are appearing and getting more revealing. We live in beautiful times where we can get a diagnose and expect promising results almost immediately.

Medical Imaging is a particular field of Medical Informatics which is a non-invasive method that tries to create a visual representation of images of the human body for medical purposes. There are a lot of types of medical imaging for example the X-ray, Radiology, Magnetic Resonances, Computed tomography scan, Ultrasound, and Nuclear medicine imaging, such as positron emission tomography. This project will be exploring the Neuro-Imaging Magnetic Resonances to discover predictors of images that are provided by magnetic resonances and try to understand whether the consequences will be severe or not in the future [1]. Neuro-Imaging is the type of field which is allowing investigators and scientists to access and learn the brain's structure, function, and power [2]. Magnetic resonances on the brain are evolving and there are countless types and each of them is more appropriated to a certain result. There are the electroencephalography(EEC), position emission tomography(PET), functional magnetic resonance(fMRI) and structural magnetic resonance(sMRI) [3].

Radiomics is a quantitative approach to imaging which provides textural information by mathematical extraction of the spatial distribution of signal intensities and pixel interrelationships [4, 5]. Following the texture feature extraction, machine learning or advanced statistical methods are employed to identify the optimal features. Finally, a clinical classifier model based on the optimal features is constructed, and this can be used to test and diagnose unseen cases. Although frequently employed in the field of oncology for tumor grading and classification, the use of radiomics has recently been extended to neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease [6]. Radiomics based on neuromelanin-sensitive MRI, i.e., spectral pre-saturation.

The present study aims to investigate the feasibility of classifying Alzheimer's disorders, i.e., AD and MCI, based on radiomics features extracted from standard clinical 3D T1- weighted images. These images are routinely acquired as part of a standard clinical scan owing to which they can be used for the differential diagnosis with relative ease. The hypothesis is that radiomic features in the postulated region of interest will demonstrate differences that will consequently support delineating patients with MCI that will evolve to AD and those who won't.

2 Methodology and Materials

2.1 Dataset and subject recruitment

The dataset, that was used to do this project, was obtained in the "The Alzheimer's Disease Neuroimaging Initiative" (ADNI) platform. ADNI unites

researchers with study data as they work to define the progression of Alzheimer's disease (AD). ADNI researchers collect, validate and utilize data, including MRI and PET images, genetics, cognitive tests and blood biomarkers as predictors of the disease.

The dataset was extracted through ADNI advanced search. In the end, 568 exams with CN or MCI or AD were obtained. The metric used to see if the patient had MCI or AD or if the patient didn't have problems was the Mini Mental State Examination (MMSE).

The Mini-Mental State Exam (MMSE) is a widely used test of cognitive function among the elderly, it includes tests of orientation, attention, memory, language and visual-spatial skills. The maximum MMSE score is 30 points. A score of 26 to 30 means that the patient doesn't have any type of dementia. A score of 21 to 25 suggests mild dementia, 13 to 20 suggests moderate dementia, and less than 12 indicates severe dementia.

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*. Then, due to the fact that some patients' images were acquired using a different technique than the ones mentioned before, those exams needed to be eliminated. Finally, 386 exams were used. The 386 exams were from 22 different patients. It is noteworthy to say that despite there are only 22 patients in this study, the number of exams, specially the number of exams that had transitions from CN to AD, CN to MCI and MCI to AD were high. The importance of this study was to know features to predict transitions not to predict if the patient had AD or MCI, therefore the final number of exams is the number which we need to take into consideration.

2.2 Imaging

The images were acquired on *T1-weighted MRI* using different machines and manufactures [7].

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.

2.3 Pre-Processing

Following acquisition, all the images were on NifTI format, thus any format conversion was necessary. Then, processing pipeline was applied to the T1 structural images, **figure 1**. The pre-processing steps involved were removal of non-brain tissue and bias correction. The brain extracted non-uniform intensity corrected image (nu.mgz) was obtained by using the FSL command, Brain Extraction Tool (BET). The usage of this command was done by a script developed to extract the masks for all image volumes.

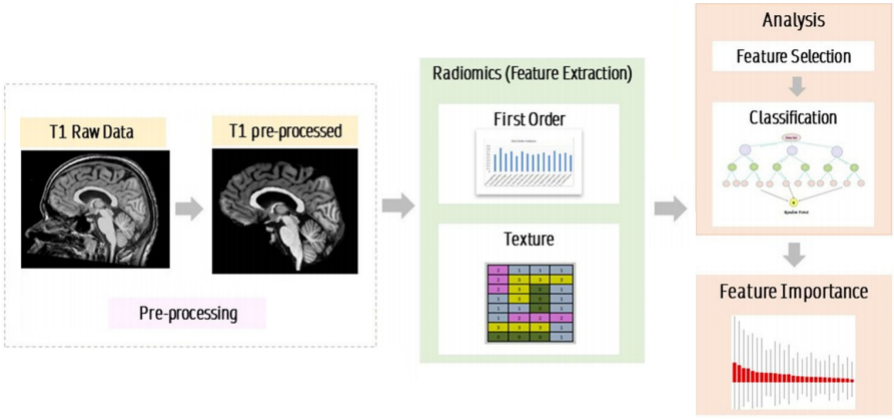


Fig. 1: Pipeline for radiomics analysis and feature extraction (adapted from [8])

2.4 Features Extraction (Radiomics)

All radiomic feature classes, with the exception of shape can be calculated on either the original image and/or a derived image, obtained by applying one of several filters. The shape descriptors are independent of gray value, and are extracted from the label mask. If enabled, they are calculated separately of enabled input image types, and listed in the result as if calculated on the original image. There are three different classes when it comes to radiomic features: First Order Features, Shape Features and Textural Features.

First-order statistics describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics.

Shape Features (2D/3D) includes descriptors of the two/three-dimensional size and shape of the Region of Interest (ROI). These features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask [9].

Textural features are the most complex ones. Their calculations are not simple, however many values can be calculated in different subclasses of textural features like values of Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) features, Gray Level Run Length Matrix (GLRLM) features, Neighbouring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM) features.

To extract radiomic features, it is needed to mention all the filters which have to be applied into the images. The reason behind this is that radiomic features will be extracted not only from the original image, but they will also be extracted from the images originated by the application of the filters. The filters applied were: Original, Wavelet, LoG, Sigma with values of [1.0, 2.0, 3.0, 4.0, 5.0], Square, SquareRoot, Logarithm, Exponential, Gradient, LBP2D and LBP3D.

In the end, 1899 features were extracted. Then, we eliminated all the features that weren't numeric. The remaining 1823 numeric features were the ones which were used to the machine learning models.

2.5 Clinical Data

Figure 2 provides the demographic and clinical information of the dataset under consideration. The dataset had 90 exams in which the patient was considered healthy, in 223 exams the patient had moderate dementia (MCI) and in 73 exams the patient had severe dementia (AD).

Regarding gender, the dataset is quite balanced, except on the healthy exams, as there were 73 exams performed on healthy males and 17 performed on healthy females. In the MCI part, 114 exams were performed on males and 109 performed on females. In the AD subset, 40 exams were performed on males and 33 performed on females.

In the age section, the range is quite similar between the subsets. The minimum age recorded on an exam was 62 years and the maximum was 90 years.

		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

Fig. 2: Clinical data and information about the exams

2.6 Machine Learning models and techniques

Machine learning on medical imaging is getting excellent results overall, mostly because of the performance of some machine learning algorithms such as: XGBoost and Multilayer Perceptron (MLP). After the training of these algorithms, techniques like Shap Values and Feature Importance can help understand what were the most important features to the training of the respective algorithm.

XGBoost XGBoost is an optimized distributed gradient boosting library 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 [10]. XGBoost is an algorithm that has

recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance [10].

MLP A Multilayer Perceptron (MLP) is a deep, artificial neural network. It is composed of more than one perceptron. They are composed of an input layer to receive the signal, an output layer that makes a decision or prediction about the input, and in between those two, an arbitrary number of hidden layers that are the true computational engine of the MLP. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training [11].

Multilayer perceptrons are often applied to supervised learning problems: they train on a set of input-output pairs and learn to model the correlation between those inputs and outputs. Training involves adjusting the parameters, or the weights and biases, of the model in order to minimize error. Backpropagation is used to make those weight and bias adjustments relative to the error, and the error itself can be measured in a variety of ways, including by root mean squared error (RMSE) [11].

SHAP VALUES SHAP (SHapley Additive exPlanations), **figure 3**, are based on Shapley values, a concept coming from game theory. But game theory needs at least two things: a game and some players. It is important to stress that a "game" concerns a single observation. One game: one observation. Indeed, SHAP is about local interpretability of a predictive model. Shapley values are based on the idea that the outcome of each possible combination (or coalition) of players should be considered to determine the importance of a single player [12].

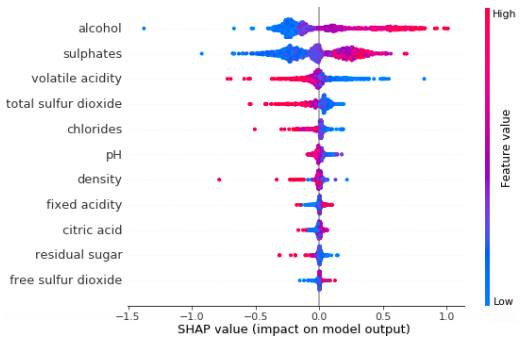


Fig. 3: Example of a graph using Shap Values [13]

The SHAP value is a united approach to explaining the output of any machine learning model, as it has three major benefits. The first one is global interpretability — the collective SHAP values can show how much each predictor

contributes, either positively or negatively, to the target variable. The second benefit is local interpretability — each observation gets its own set of SHAP values. It can explain why a case receives its prediction and the contributions of the predictors. Third, the SHAP values can be calculated for any tree-based model, while other methods use linear regression or logistic regression models as the surrogate models [12].

FEATURE IMPORTANCE Feature importance can be used in machine learning models to understand which were the most important features to the specific model.

This technique assigns a score to input features based on how useful they are at predicting a target variable.

Feature importance scores play an important role in a predictive modelling project, including providing insight into the data and into the model, and the basis for dimensionality reduction and feature selection that can improve the efficiency and effectiveness of a predictive model on the problem. The scores can also be used in a range of situations in a predictive modelling problem, such as: a better understanding about the data and the model and reducing the number of input features [14].

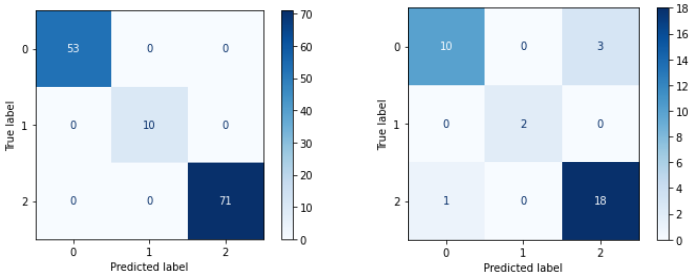
3 Results

Two different situations were used to develop the machine learning models. The first one was to only pick up the exams where the patient had transited from CN to MCI or CN to AD or MCI to AD. The other situation was to only choose the exams where the patient hadn't transited from a condition to another, thus the patient remained in the same condition.

3.1 Transitions

In this approach, the dataset had 168 exams in which the patient had transited from a condition to another. Two different models were used. The first one was Multilayer Perceptron. The dataset was divided into 80% for training and 20% for testing. The final accuracy value was 88.24%.

However, due to the fact that the implementation of this model was quite simple, with just one hidden layer, thus it was also implemented a XGBoost model that was far more robust. In this model, the dataset was split into 80% for training and the remaining for testing. In the end, in the label variable (y), there were 3 classes: 0 matched the class of CN to MCI, 1 matched the class of CN to AD and the last one matched the class of MCI to AD. In this model, it was possible to see a more credible confusion matrix than the ones obtained with the Multilayer Perceptron. Thus, in this paper it is only presented the confusion matrices (training and testing) of the XGBoost model, **figure 4**. The final accuracy obtained was 88.24%.



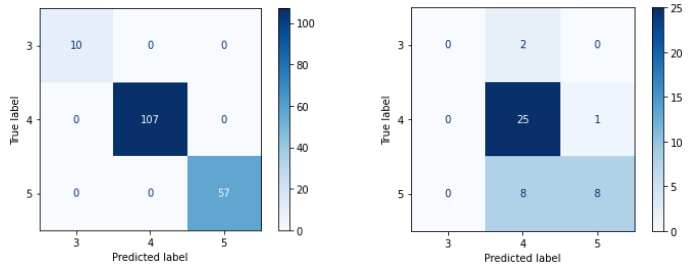
(a) Training (b) Testing

Fig. 4: Correlation Matrices

3.2 No-Transitions

In this approach, the dataset had 218 exams in which the patient hadn't transited from a condition to another. It was used two different models. The first one was Multilayer Perceptron. The dataset was divided into 80% for training and 20% for testing. The final accuracy value was 79.55%.

However, as in the first transitions approach, due to the fact that the implementation of this model was quite simple, with just one hidden layer, thus it was also implemented a XGBoost model that was far more robust. In this model, the dataset was split into 80% for training and the remaining for testing. In the end, in the label variable (y), there were 3 classes: 3 matched the class which the patient remained normal, 4 matched the class which the patient remained with MCI and the last one matched the class which the patient remained with AD. In this model, it was possible to see a more credible confusion matrix than the ones obtained with the Multilayer Perceptron. Thus, in this paper it is only presented the confusion matrices (training and testing) of the XGBoost model, **figure 5**. The final accuracy obtained was 75%.



(a) Training (b) Testing

Fig. 5: Correlation Matrices

4 Discussion

The discovery of outcome predictors for MCI condition can be done using post-modelling techniques such as Shap Values and Feature Importance. In this section, it will be discussed what were the most important features of them all to the XGBoost model for training and testing the data. It will also be compared if the most relevant features in the situation where the patient evolve from one condition to another were also present or not in the situations where he didn't transit at all.

4.1 Feature Importance

The graph in **figure 6** shows the most important features which were mostly used by the XGBoost model. The graph exhibits the most relevant features in descending order of F score. In the graph, there are 35 features displayed, as the minimum F 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. By the analysis of the graph, it is clear that the difference of MMSE (DifMMSE) and the difference of age (DifAge) in the exams that had transitions were the most important features. Regarding radiomic features, the most relevant ones were: *wavelet-HHH_glcm_lcm1, log-sigma-5-0-mm-3D_ glszm_SizeZoneNonUniformity, original_glszm_LargeAreaHighGrayLevelEmphasis,log-sigma-5-0-mm-3D_firstorder_Mean.*

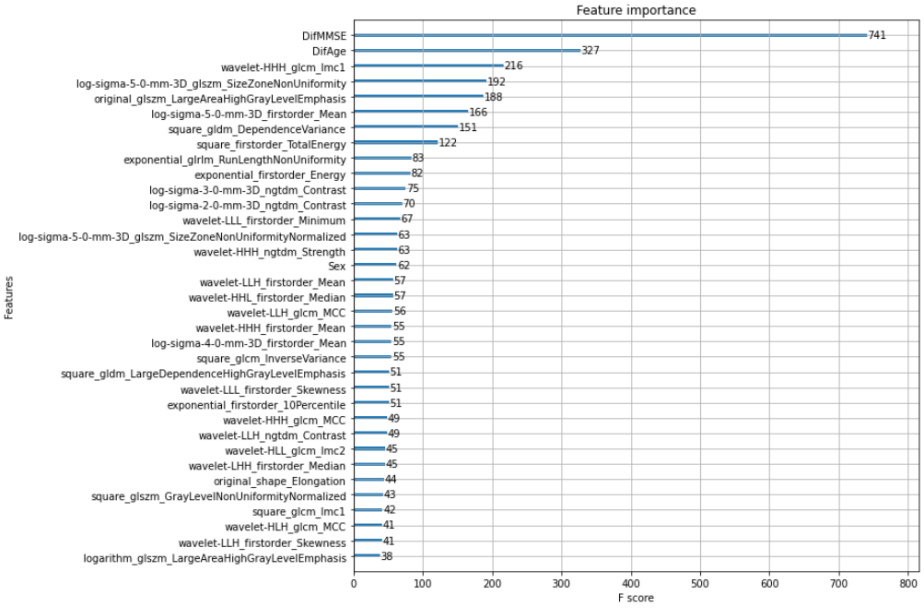


Fig. 6: Feature Importance

4.2 Predictors using Shap Values

The clinical diagnosis of Alzheimer's disorders is often complicated in the early stages of illness owing to similarities in clinical features. Both **figures 7** and **8**, give information regarding what were the most important features for all the classes.

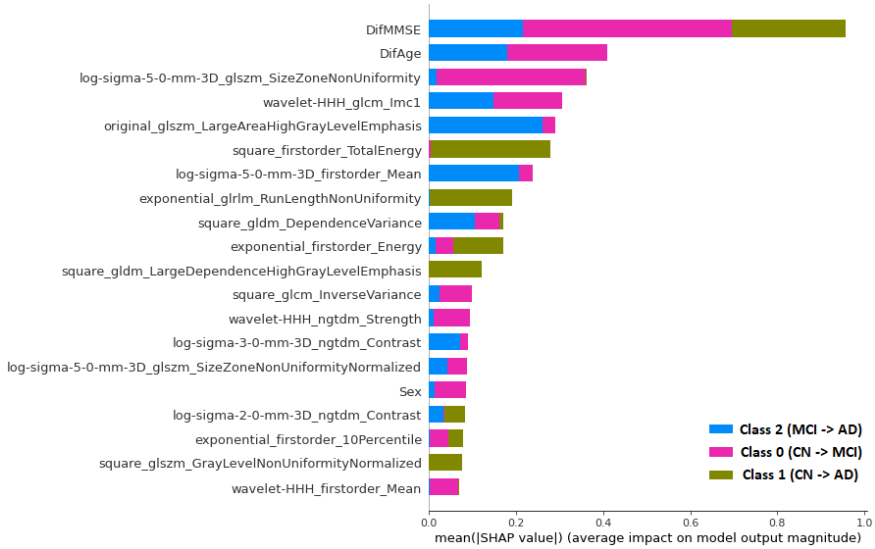


Fig. 7: Shap Values for the transitions

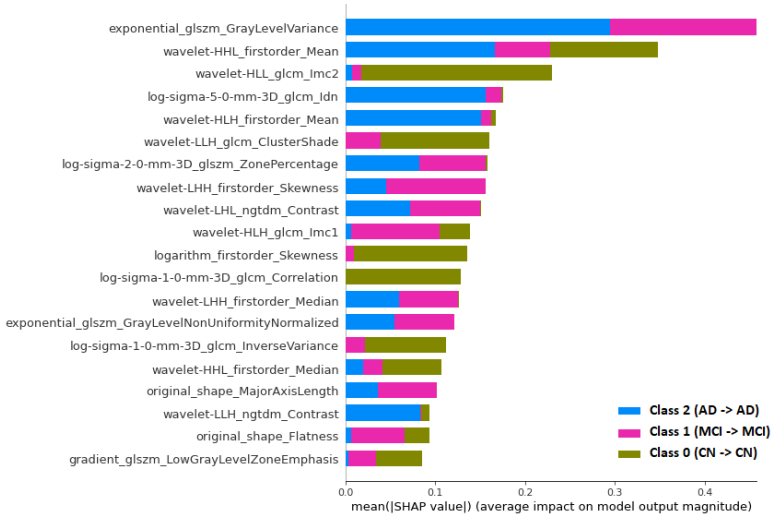


Fig. 8: Shap Values for the no transitions

To discover what were the important features in the transition from CN to MCI (class 0 of the **figure 7**) and from CN to AD (class 1 of the **figure 7**), it was needed to understand what were the important ones to patients that didn't transit from CN to something (class 0 of the **figure 8**). By analysing both situations, it is clear that all the features that were important in the transitions from CN to MCI or AD weren't important to the no transition situation. Thus, features which can be considered predictors to discover if a patient that is normal will evolve to MCI are:

- *DifMMSE*;
- *DifAge*;
- *log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformity*;
- *wavelet-HHH_glcm_lcm1*;
- *original_glszm_LargeAreaHighGrayLevelEmphasis;*
- *log-sigma-5-0-mm-3D_firstorder_Mean*;
- *square_gldm_DependenceVariance*;
- *exponential_firstorder_Energy*;
- *square_glcm_InverseVariance*;
- *wavelet-HHH_ngtdm_Strength;*
- *log-sigma-3-0-mm-3D_ngtdm_Contrast;*
- *log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized;*
- *Sex;*
- *exponential_firstorder_10Percentile;*
- *wavelet-HHH_firstorder_Mean.*

Features that can be considered predictors to discover if a patient that is normal will evolve to AD are:

- *DifMMSE;*
- *square_firstorder_TotalEnergy;*
- *exponential_glrlm_RunLenghtNonUniformity;*
- *exponential_firstorder_Energy;*
- *square_gldm_LargeDependenceHighGrayLevelEmphasis;*
- *log-sigma-2-0-mm-3D_ngtdm_Contrast;*
- *exponential_firstorder_10Percentile;*
- *square_glszm_GrayLevelNonUniformityNormalized.*

Then, to find outcome predictors to see if a patient with MCI will evolve to AD, it was needed to see what were the important features in the transition from MCI to AD (class 2 of **figure 7**) and what were the important ones in the situation where a patient with MCI didn't evolve to AD (class 1 of **figure 8**). Again, by analysing both situations, it is understandable that all the features that were important in the transition from MCI to AD weren't important to the no transition situation. Therefore, features which can be considered predictors to discover if a patient that has MCI will evolve to AD are:

- *DifMMSE;*
- *DifAge*;
- *wavelet-HHH_glcm_lcm1;*
- *original_glszm_LargeAreaHighGrayLevelEmphasis*;
- *log-sigma-5-0-mm-3D_firstorder_Mean;*
- *square_gldm_DependenceVariance, square_glcm_InverseVariance*;
- *log-sigma-3-0-mm-3D_ngtdm_Contrast;*
- *log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized*;
- *log-sigma-2-0-mm-3D_ngtdm_Contrast.*

All this features can help in the decision making. Even with new images or image volumes, these predictors will certainly help doctors to understand if a patient who is normal will develop MCI or AD or if a patient who has MCI will evolve into AD in the future.

5 Conclusion

The present study establishes the utility of discovering predictors for MCI condition based on radiomic features extracted from T1-weighted images. All routine clinical scans include T1-weighted images, and the results obtained from this study expand the possibility of the role of machine learning–based classification and prediction techniques in clinical practice. Future studies could aim to well characterize these textural features that 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.

The relation between radiomic features and machine learning succeeds in the discovery of outcome predictors in medical practice. This information will help doctors and clinical staff on the decision making and consequently, it will help patients' lives.

References

1. University of Illinois at Chicago. "What is Medical Informatics?," Accessed Jun. 6, 2021. [Online]. Available: https://healthinformatics.uic.edu/blog/what-is-medical-informatics/
2. NPS MedicineWise. "Imaging explained," Accessed Jun. 3, 2021. [Online]. Available: https://www.nps.org.au/consumers/imaging-explainedwhat-are-my-imaging-choices
3. Michael Brammer, "The role of neuroimaging in diagnosis and personalized medicine-current position and likely future directions," Dec. 11, 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181933/
4. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, et al., "Radiomics: the bridge between medical imaging and personalized medicine," Dec. 14, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28975929/

5. Janita E. van Timmeren, Davide Cester, Stephanie Tanadini-Lang, Hatem Alkadhi, Bettina Baessler, "Radiomics in medical imaging—"how-to" guide and critical reflection," Dec. 11, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7423816/
6. Christian Salvatore, Isabella Castiglioni, Antonio Cerasa, "Radiomics approach in the neurodegenerative brain," Aug. 19, 2019. [Online]. Available: https://doi.org/10.1007/s40520-019-01299-z
7. ADNI. "Alzheimer's Disease Neuroimaging Initiative," Accessed Jun. 1, 2021. [Online]. Available: http://adni.loni.usc.edu/
8. Priyanka Tupe-Waghmare, Archith Rajan, et al., "Radiomics on routine T1-weighted MRI can delineate Parkinson's disease from multiple system atrophy and progressive supranuclear palsy," May 4, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33945022/
9. Pyradiomics community. "Welcome to pyradiomics documentation!," Accessed Jun. 7, 2021. [Online]. Available: https://pyradiomics.readthedocs.io/en/latest/
10. Xgboost developers. "XGBoost Documentation," Accessed Jun. 5, 2021. [Online]. Available: https://xgboost.readthedocs.io/en/latest/
11. "Multilayer perceptron," *Wikipedia*. Accessed Jun. 7, 2021. [Online]. [Online]. Available: https://en.wikipedia.org/wiki/Multilayer_perceptron
12. Mustafa Hassan Fuuast, et al. "SHAP Values," Accessed Jun. 7, 2021. [Online]. Available: https://www.kaggle.com/dansbecker/shap-values
13. Dr. Dataman. "Explain Your Model with the SHAP Values," Accessed Jun. 5, 2021. [Online]. Available: https://towardsdatascience.com/explain-your-model-with-the-shap-values-bc36aac4de3d
14. Jason Brownlee. "How to Calculate Feature Importance With Python," Accessed Jun. 5, 2021. [Online]. Available: https://machinelearningmastery.com/calculate-feature-importance-with-python/