

Alzheimer

A tool for Doctors

Author: Raúl García Gómez



Content Synopsis

OVERVIEW OF KEY IDEAS

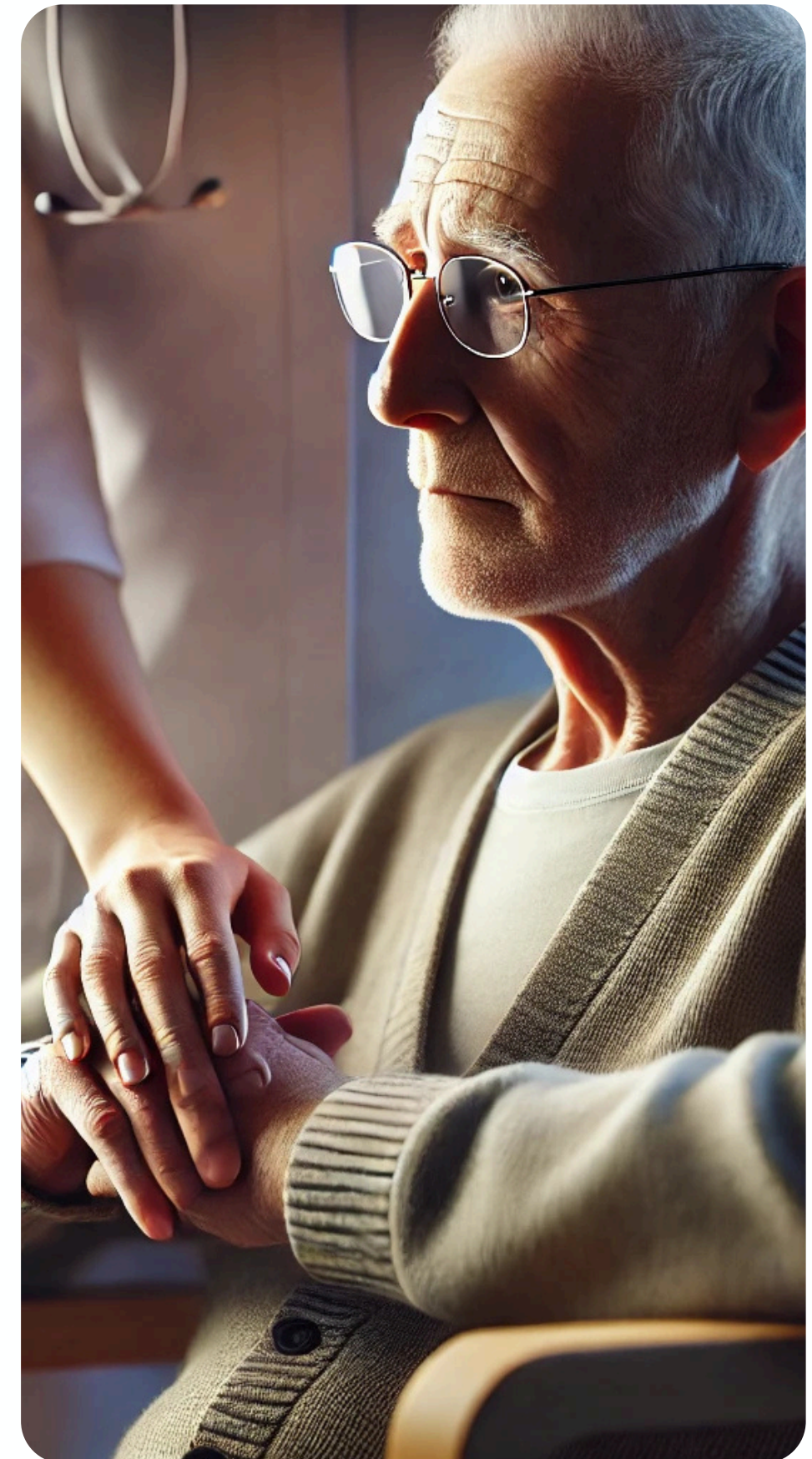
Context and goal of the project

Data Analysis and Features

Model Construction - Recall

Model Comparison and Winners

Final Notes



Alzheimer & Dementia

DEMENTIA IS NOT EQUAL TO ALZHEIMER

There are over 55 million people worldwide with dementia, according to WHO. 50-60 % of them are affected by Alzheimer's disease, a type of dementia. This project focuses on Alzheimer.

There is no cure for any of them, but there are treatments which slow down the progression of the condition.



Goal of the project

The Goal

The goal of the project is to help doctors and healthcare workers on their **diagnosis of Alzheimer** on patients. The models are just a tool to help the doctor, who take **full responsibility** on the final diagnosis.

Two Models

To ensure the diagnosis is evidence-based, two models have been trained with Machine Learning for this occasion:

- A first model, which based on preliminary data of the patient can predict **whether they present signs of Alzheimer or not**.
- The second model predicts both whether there are signs of Alzheimer **from their MRI scan**, and which is the development of the disease (provided there is Alzheimer).

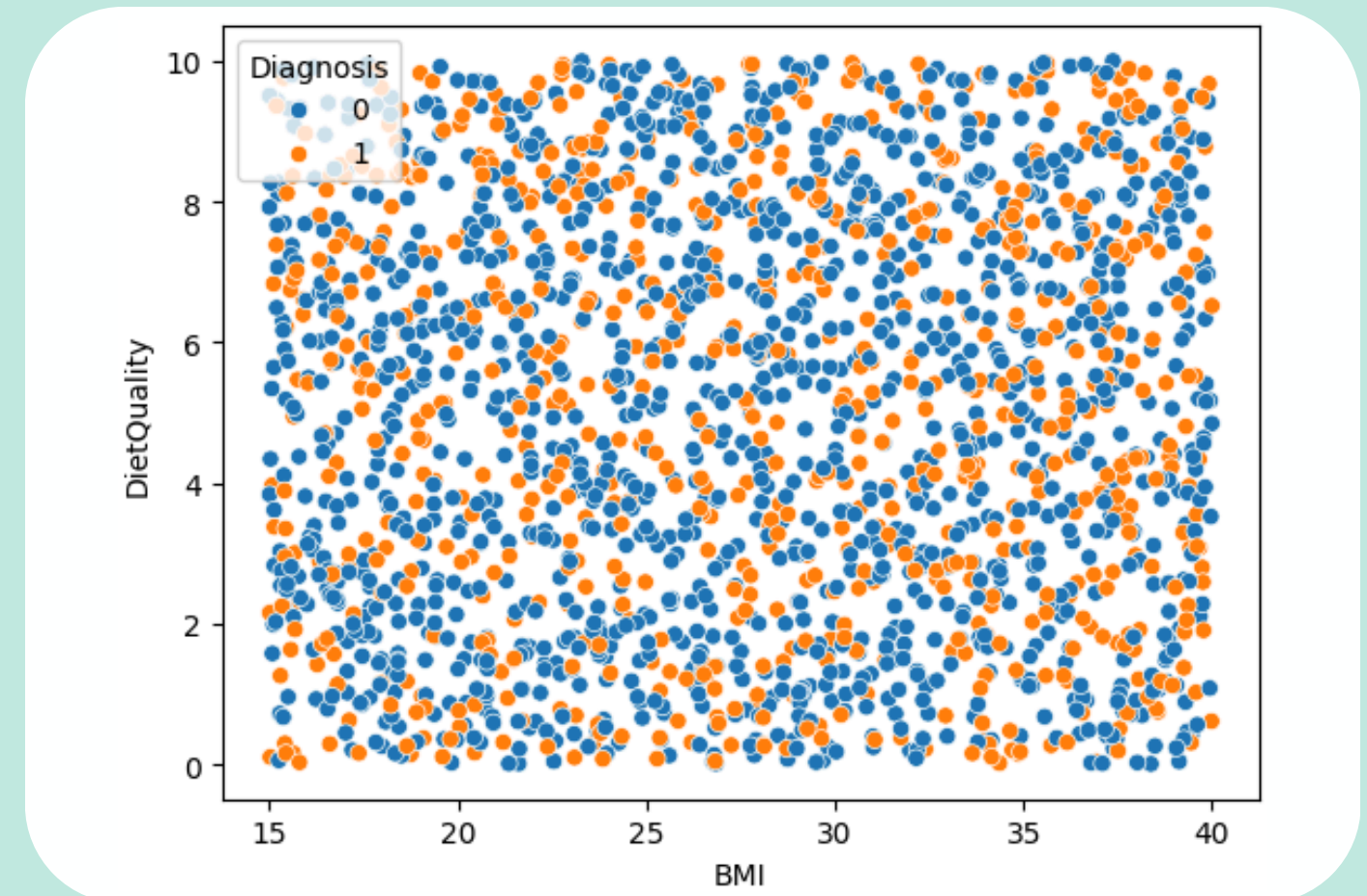
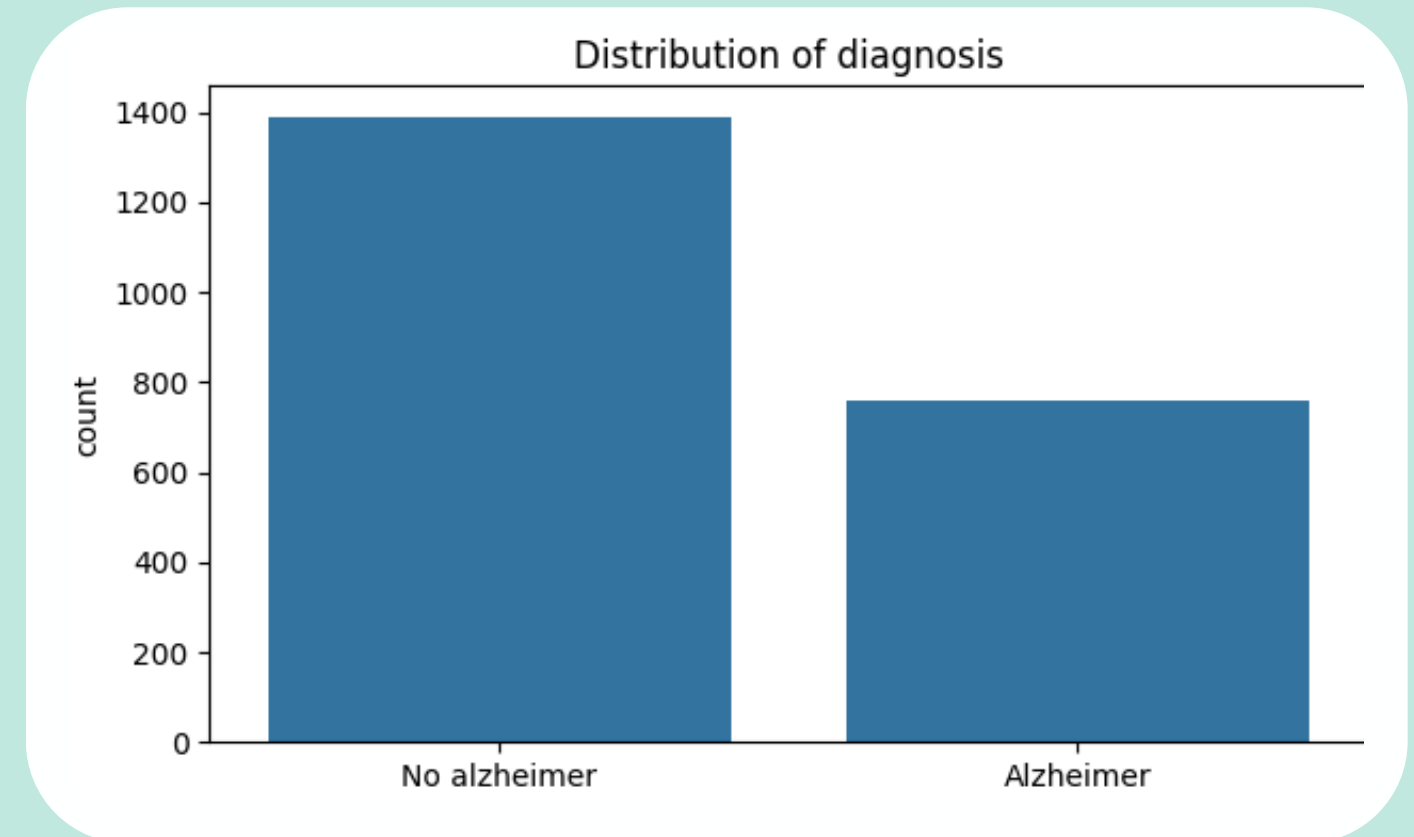
Main metrics

Both models resolve classification issues. Although accuracy will be shown as a common way of comparison, the main goal is to **maximize recall**.

Descriptive Analysis

Preliminary model

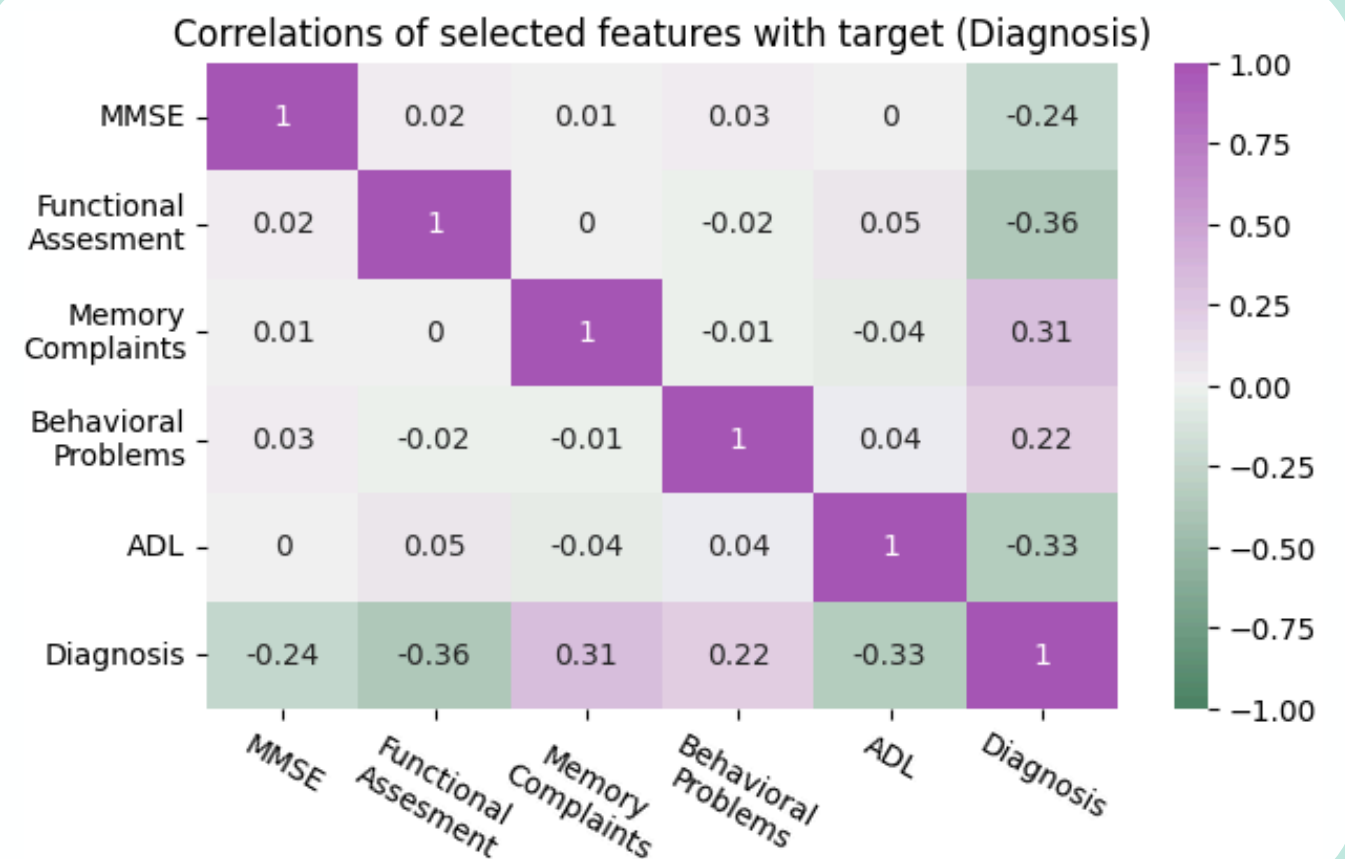
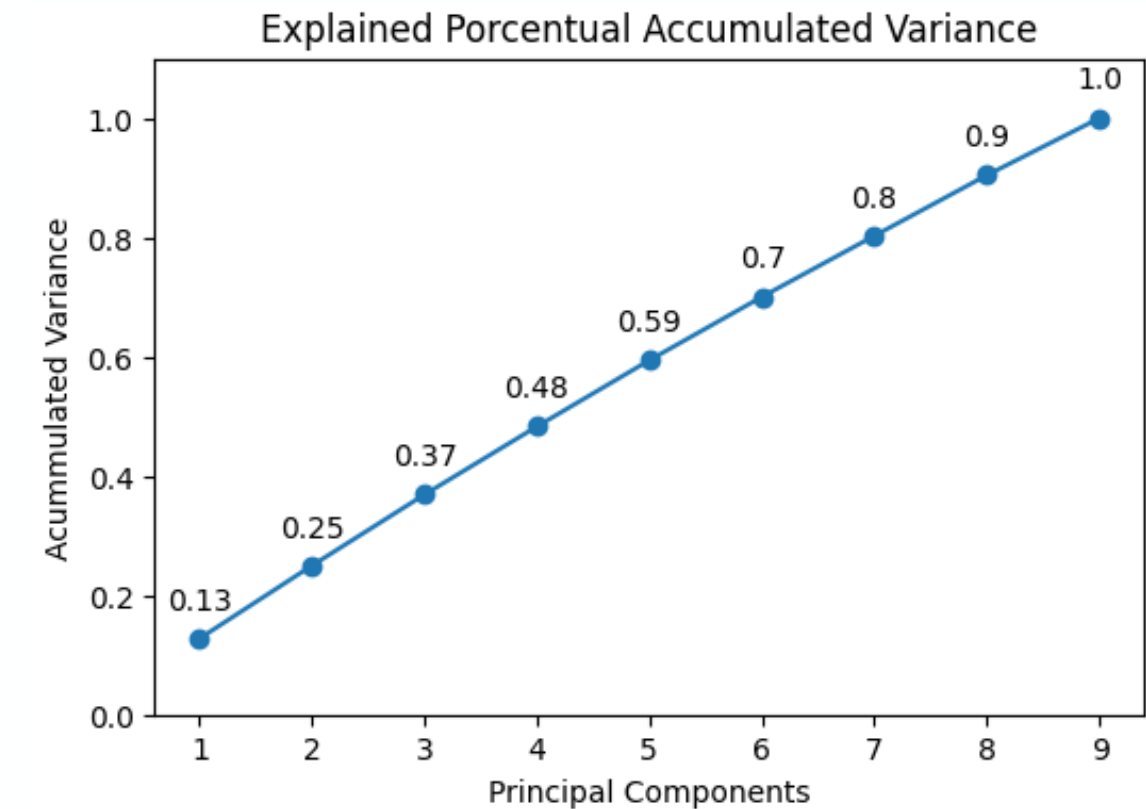
- 2149 rows, containing 31 numerical columns with related information.
- Little data cleaning to be done.
- Several techniques were tested to maximize recall (feature engineering, feature reduction via PCA), but none topped the original results.
- This is probably due to the fact that the dataset is synthetic.



Descriptive Analysis

Preliminary model

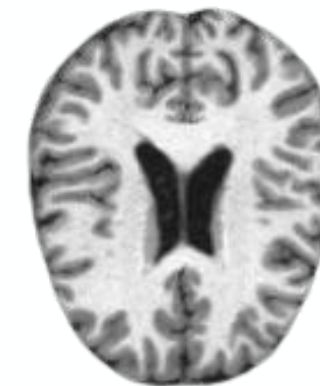
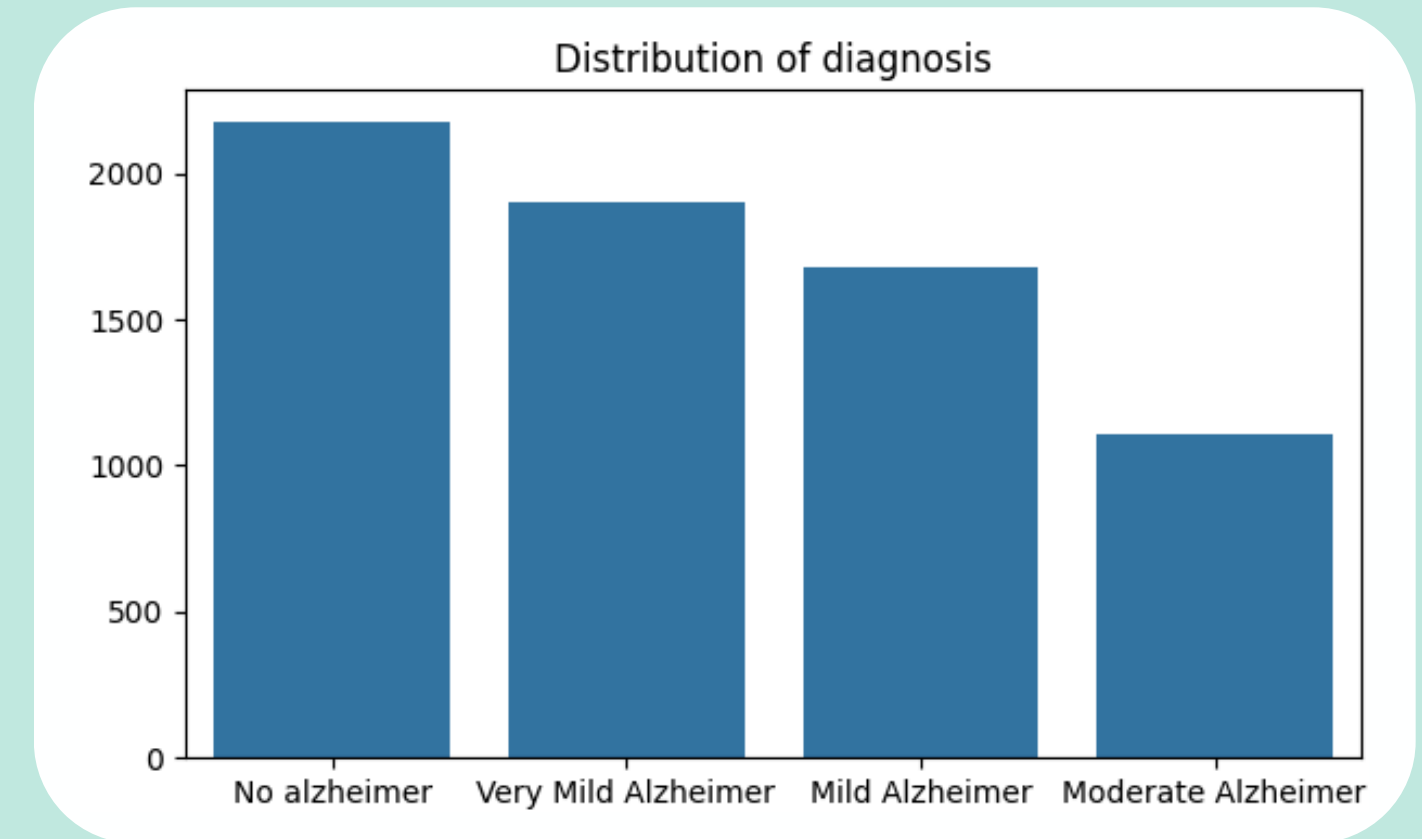
- 2149 rows, containing 31 numerical columns with related information.
- Little data cleaning to be done.
- Several techniques were tested to maximize recall (feature engineering, feature reduction via PCA), but none topped the original results.
- This is probably due to the fact that the dataset is synthetic.



Descriptive Analysis

MRI model

- 40,000 MRI scans of healthy brains, and different degrees of Alzheimer.
- Many images where coloured, so a grey filter is applied.



Healthy Brain



Brain with
moderate dementia

Model selection

Preliminary model

Three models have been taken into account for the preliminary model:

BASELINE

- Used as a baseline metric to compare further models.
- **XGBoost.**

MODEL 2

- Deeper feature engineering, focused on variable aggregation.
- PCA was tried, but not implemented.
- Hyper-parametrization, comparing a total of 1112 candidates.
- **Gradient Boosting.**

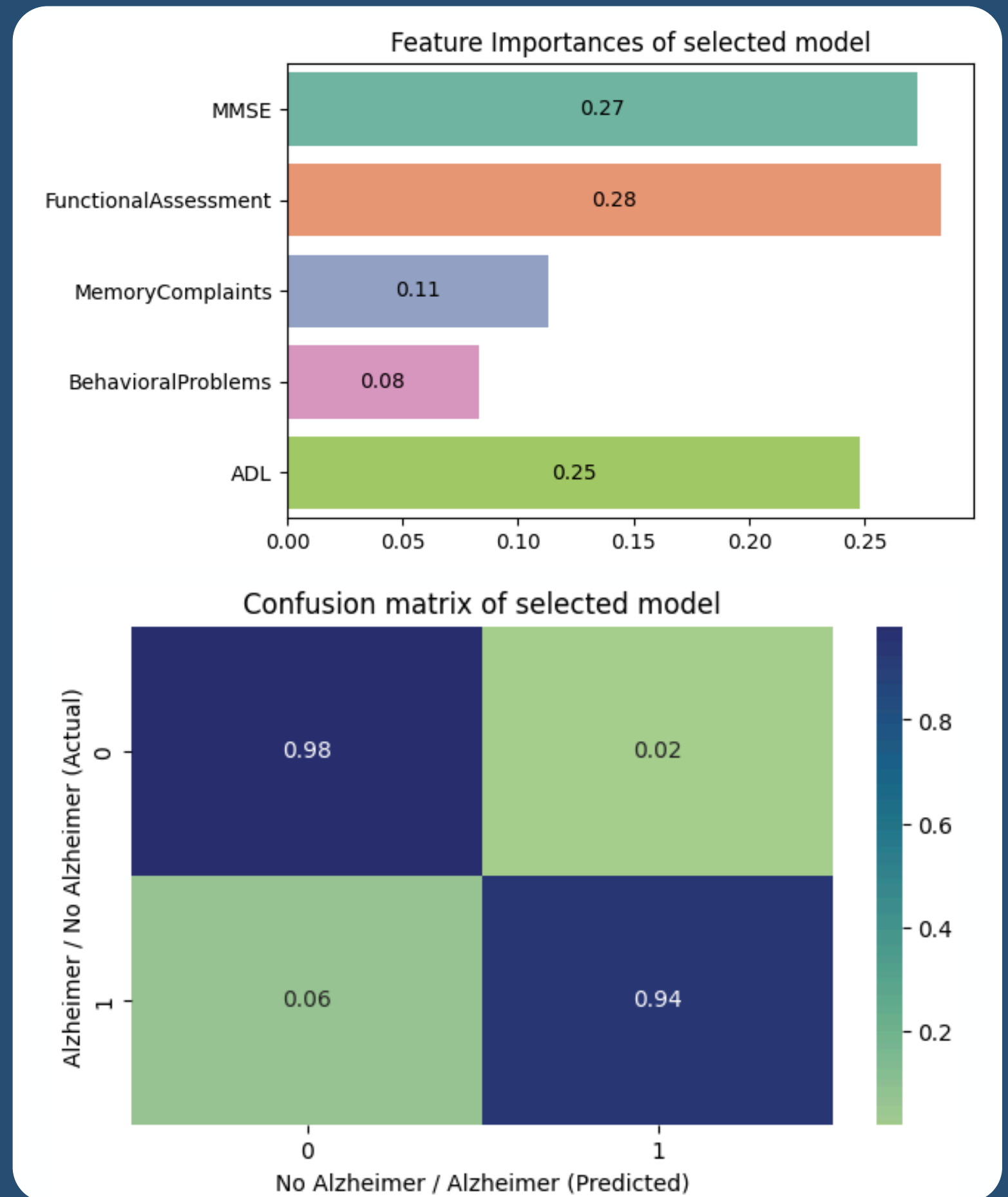
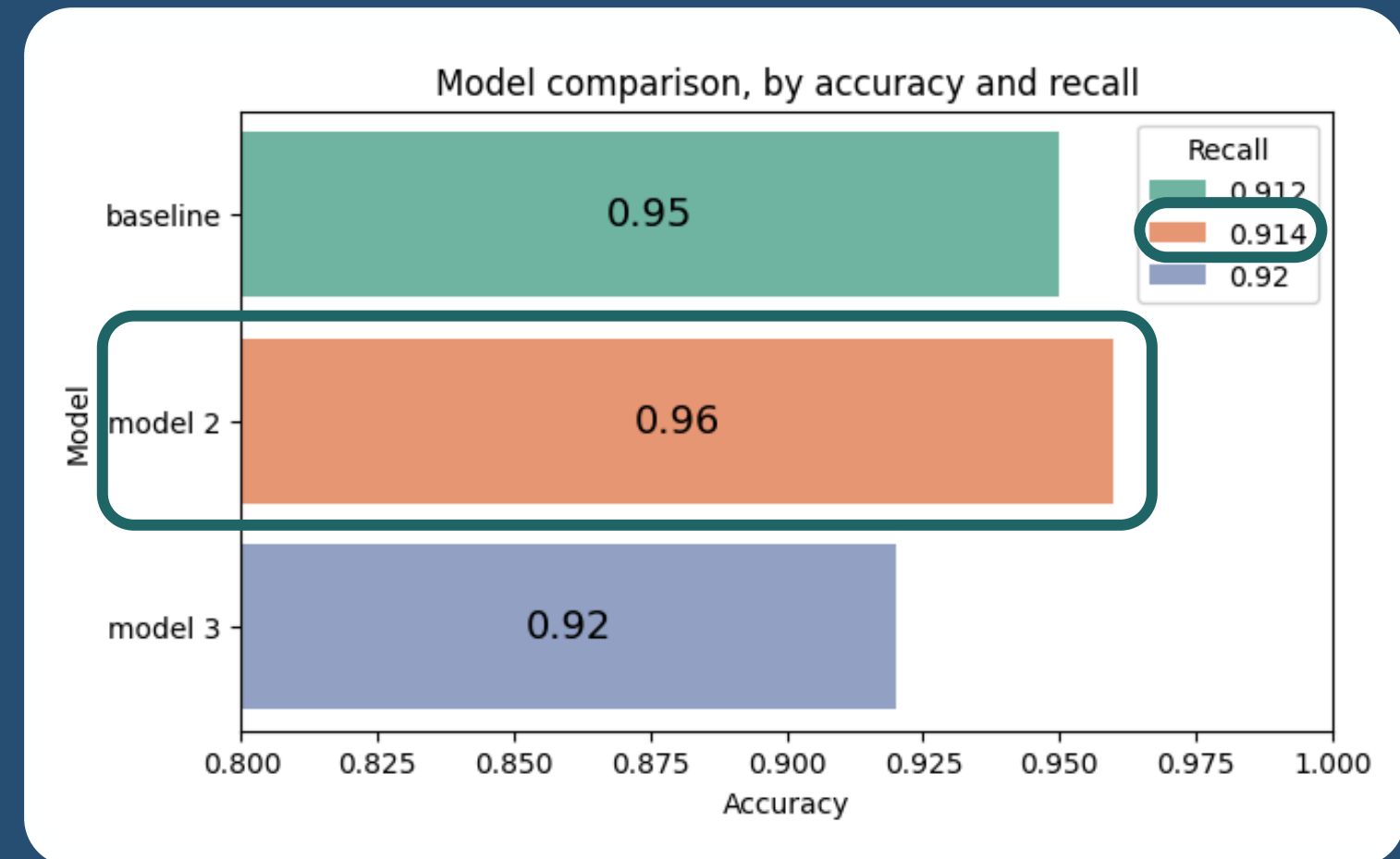
MODEL 3

- Same variables as in Model 2, but with a **Neural Network.**
- **NN(5x300x100x1)**

PRELIMINARY MODEL

Model 2 is the winner

2ND HIGHEST RECALL, AND EXPLAINABLE



Model selection

Image model

Five models have been taken into account for the image model:

BASELINE

- Used as a baseline metric to compare further models.
- Images reduced to 32x32 pixels.
- Adam optimizer.
- **CNN** (32x64x32x4)

MODEL 2

- Rearranged datasets to ensure robustness.

MODEL 3

- Uses DataGenerator.

MODEL 4

- New CNN layer.
- Layers are also deeper.
- Images reduced to 64x64 pixels.
- Optimizer: **RMSProp**.
- **CNN** (64x256x128x64x4).

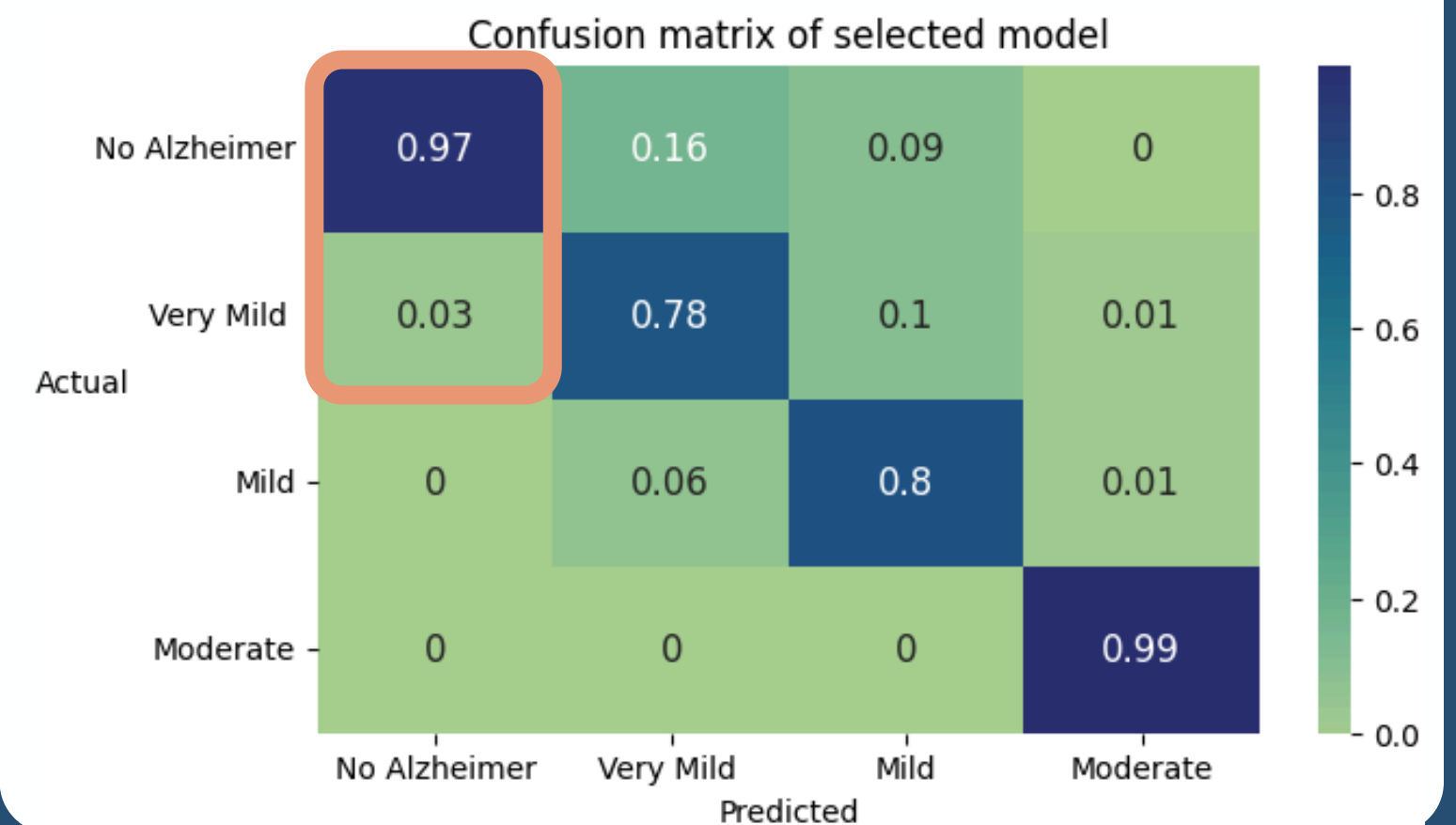
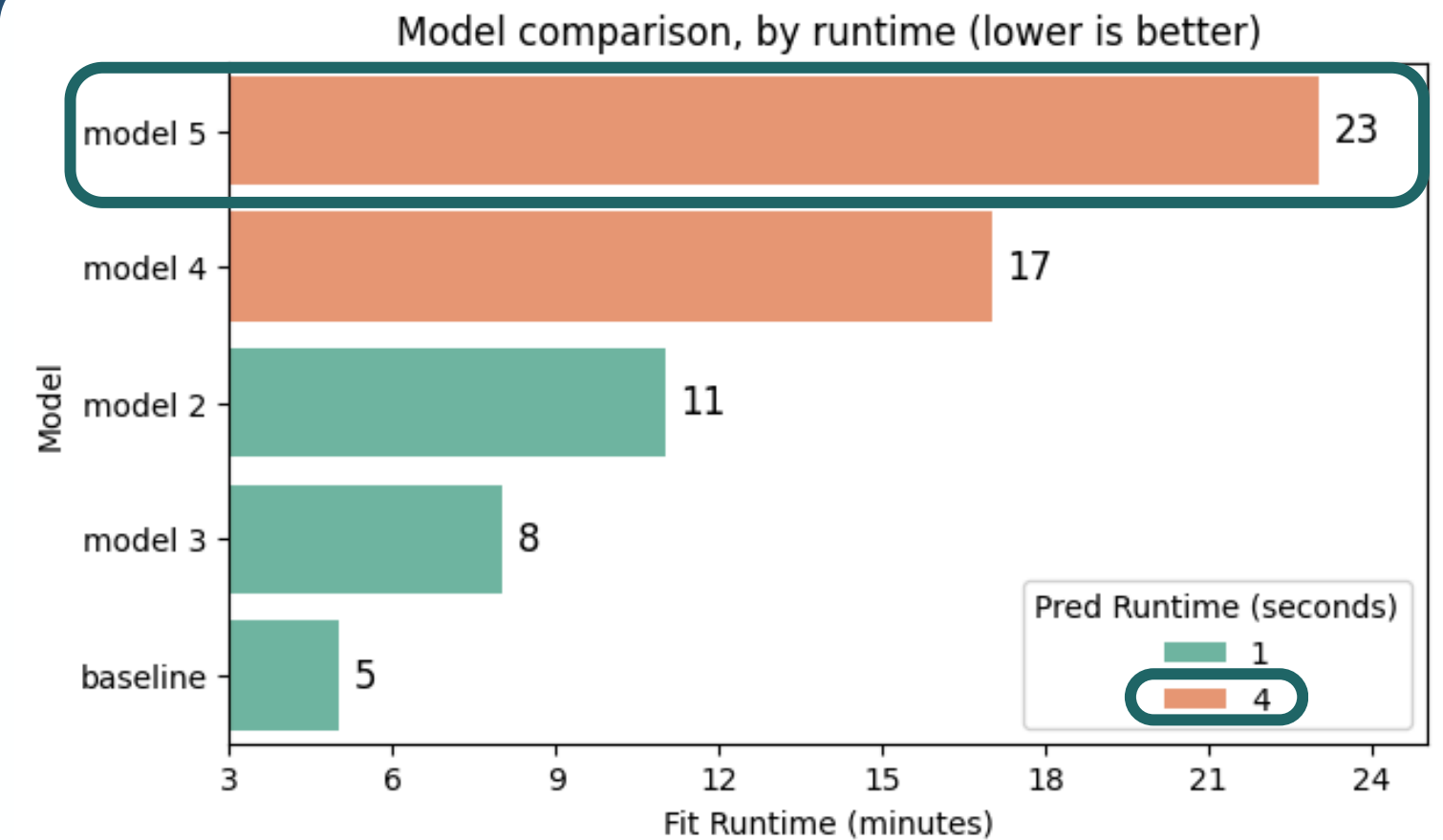
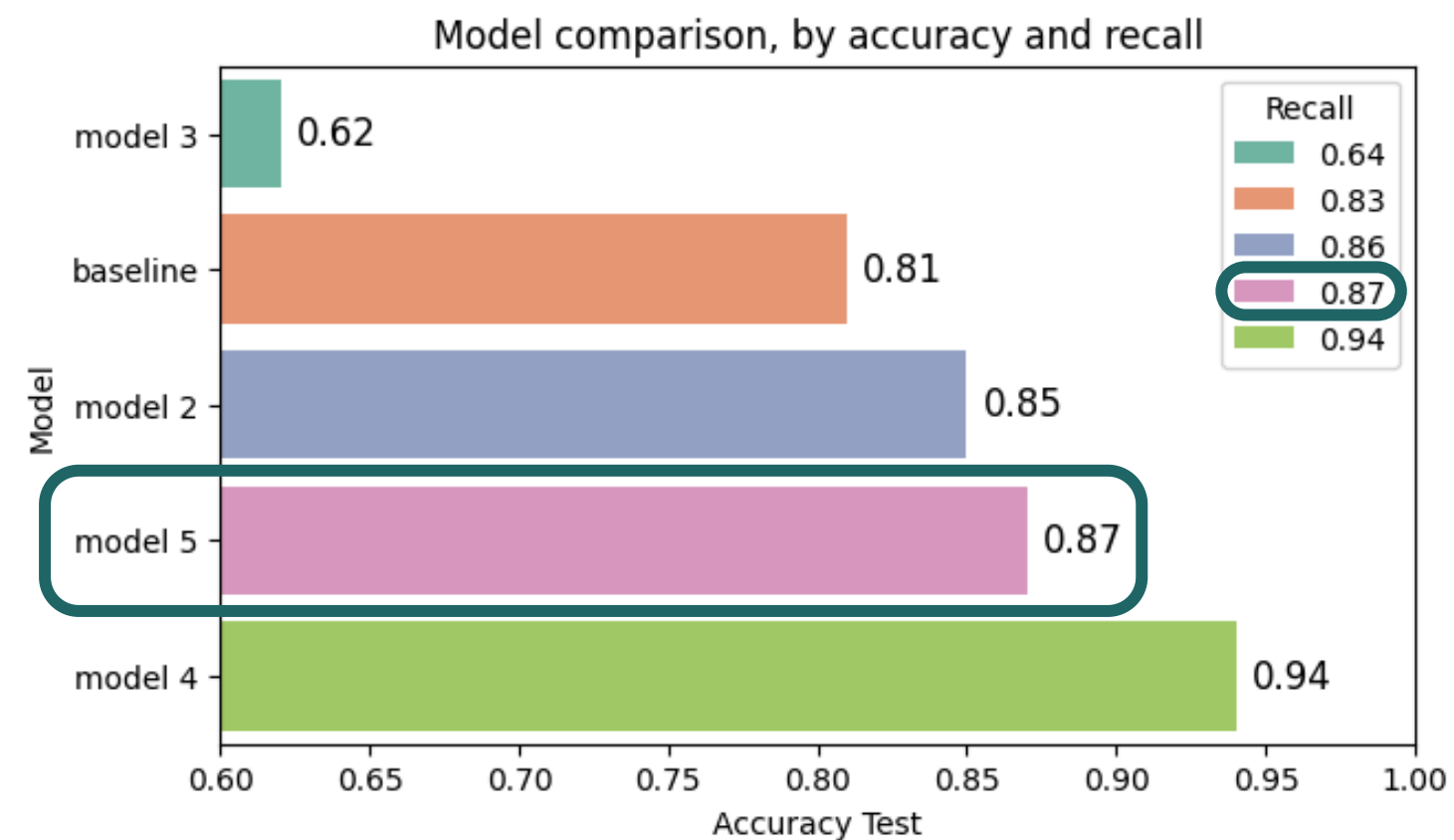
MODEL 5

- Fewer and less deep layers, to prevent overfitting.
- Extra layer of dropout.
- **CNN** (64x128x64x4).

PRELIMINARY MODEL

Model 5 is selected

BEST GENERALIZATION, GOOD RECALL WHERE MATTERS



Final Notes

Beware of the results

Class model was trained with a synthetic dataset, and shouldn't be considered as the only tool for diagnosis

Better model for MRI scans

There's probably room for improvement, and further models which equilibrate recall and overfitting may be done.

MRI Scans from more sources

Model behaves surprisingly good with images outside the dataset, but more robustness can be achieved.

Image data generator

Image Data Generator gave suspiciously bad results, and further investigation would be welcome.

Thank
you

Raúl García Gómez

raul.garciagomez@hotmail.com

linkedin.com/in/raulggomez/

github.com/raugargom