Predicting the Risk of Alzheimer's Disease

Team: Samyak, Himanshu, Manikya

Outline

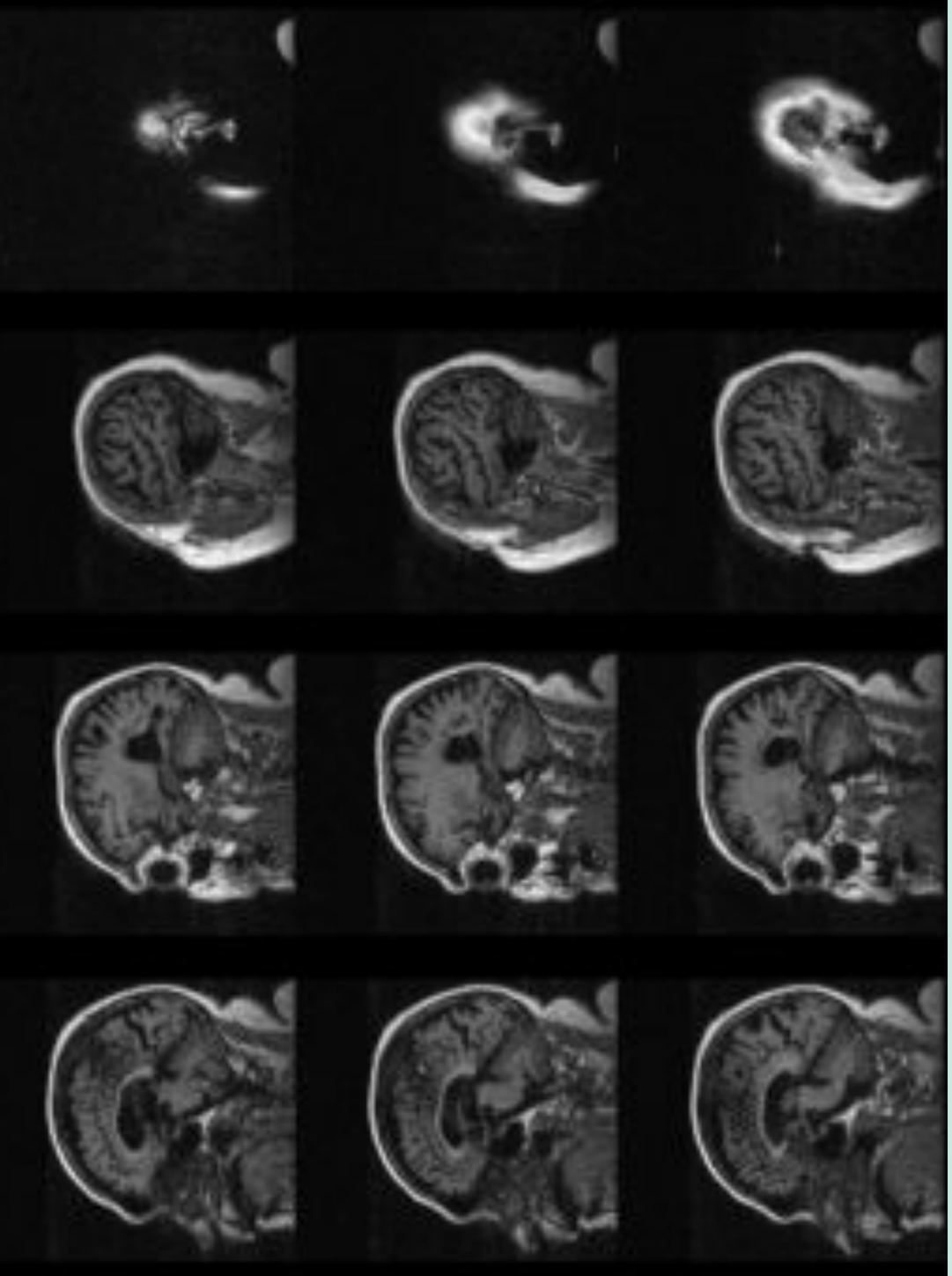
Introduction

Exploratory Data Analysis

Methodology

Results

References



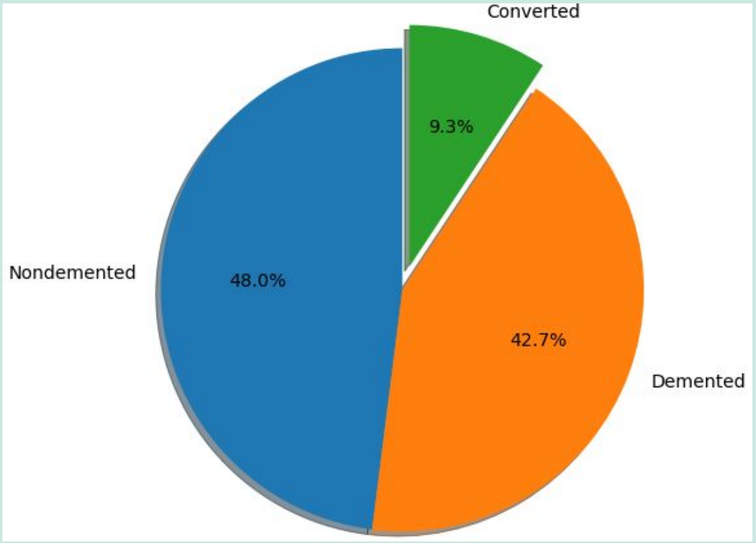
Introduction

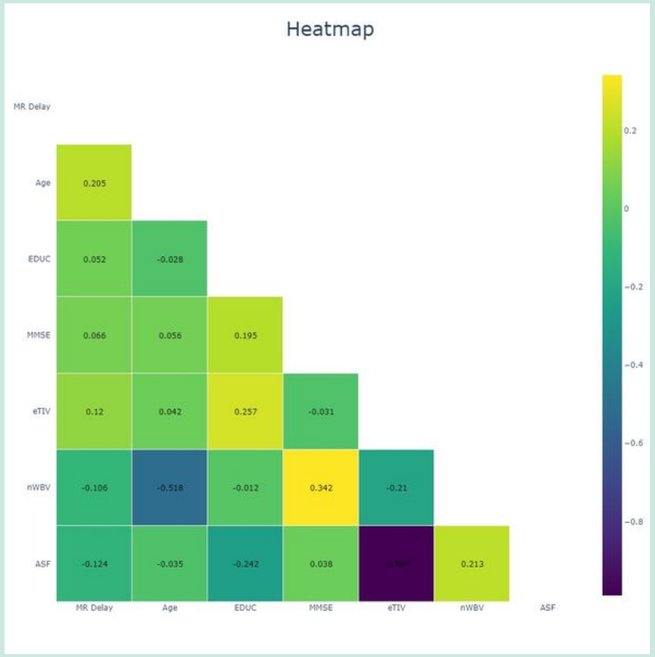
-Alzheimer’s Disease is the most common disease of dementia, which may involve the decline of memory, communication, and judgment, yet it is hard to diagnose.

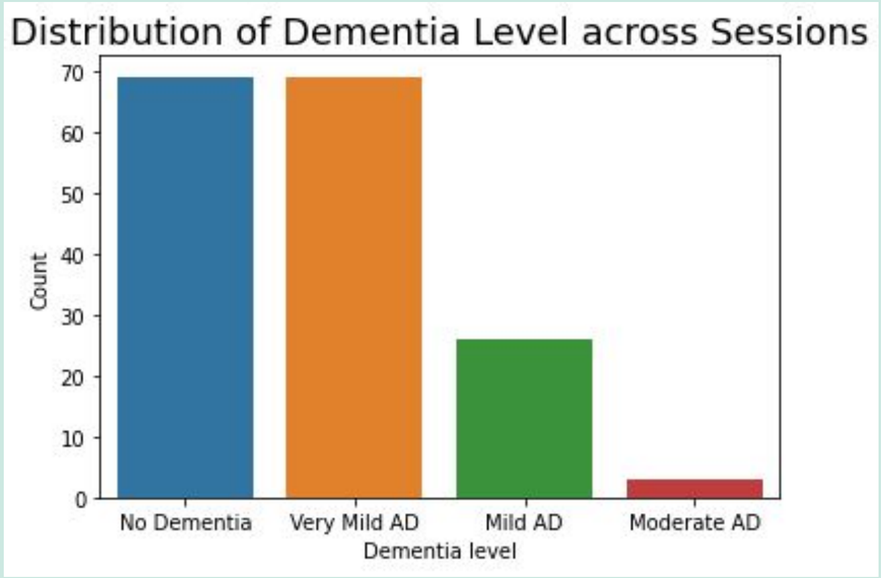
- This project involves predicting the risk of Alzheimer's Disease in a patient given their health parameters and scans.

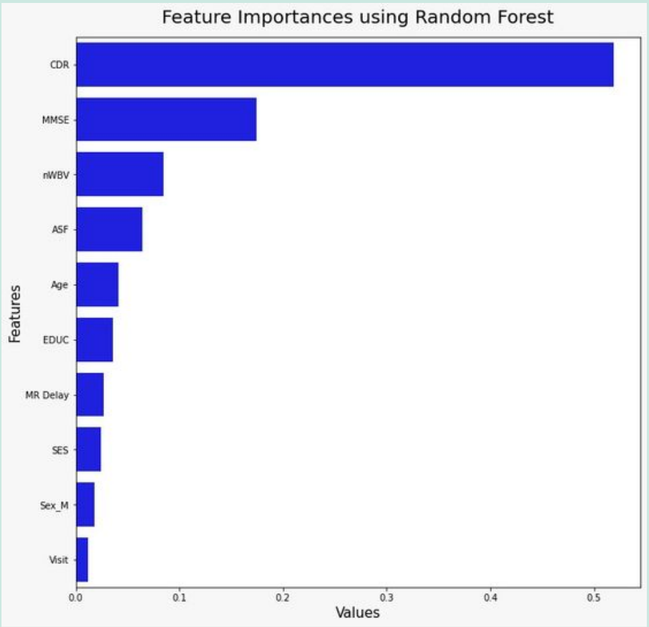
Exploratory Data Analysis

- We use the Oasis 2 dataset (https://www.oasis-brains.org/) which contains a longitudinal collection of 150 subjects aged 60 to 96.
- We make use of the demographical data provided and the CDR scores of the patients.
- We also use the T1-weighted MRI scans to make predictions about the risk of Alzheimer's Disease.



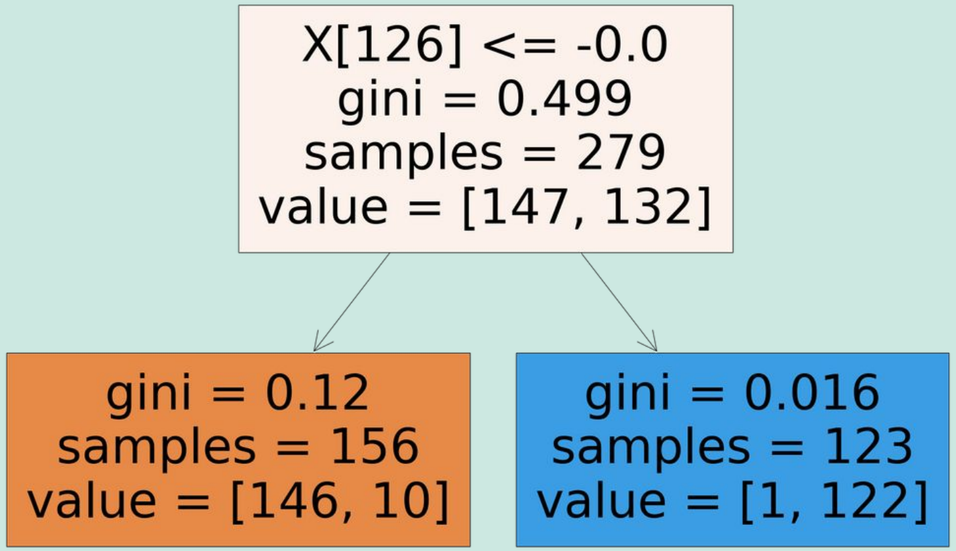






Methodology

Tree based Ensemble Models →



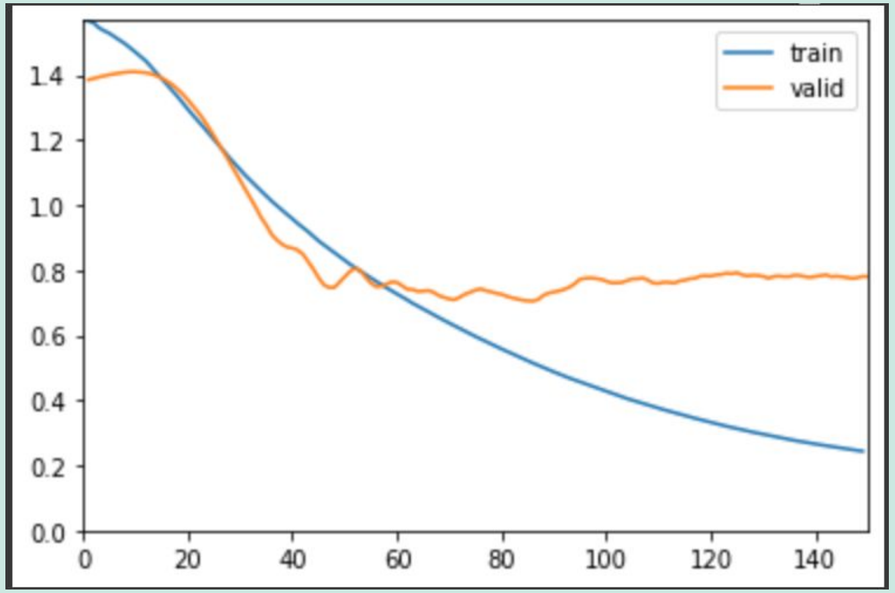
Different models ranging from linear to ensemble-based were used to predict if the given patient had dementia or not

We used a combination of feature-engineered data along with different types of hyperparameter tuning.

Finally, the best model was selected based on F1 score and training time

Methodology

Neural Network and Embeddings →



Loss vs Epochs*

We used neural networks as well for predicting the CDR score for Alzheimer's Disease and were able to get ~81 % accuracy for this task.

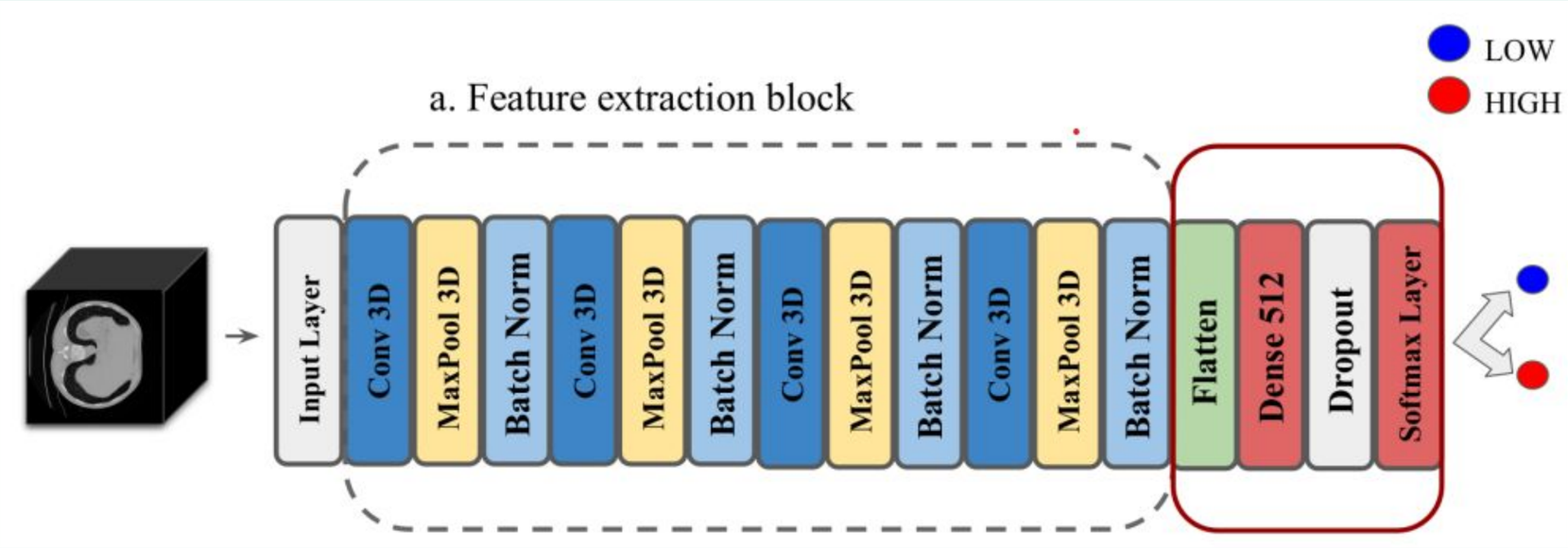
The categorical variables were represented using Embeddings, which gave much better results than other encoding approaches like One-hot encoding.

These embeddings are also useful as features for other algorithms.

* Although the graph makes it look like the model is overfitting, we use callbacks to make sure the best model is used at the end.

Methodology

3D CNN Model on T1-weighted MRI Scans

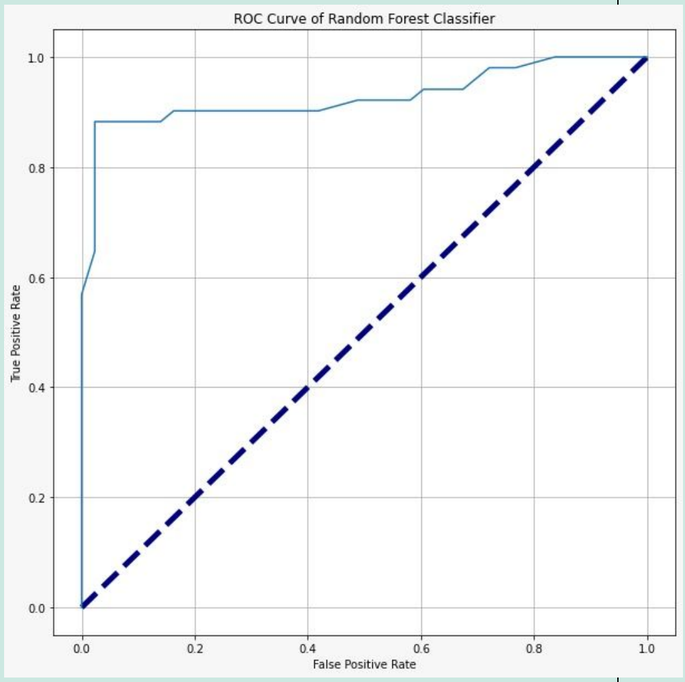


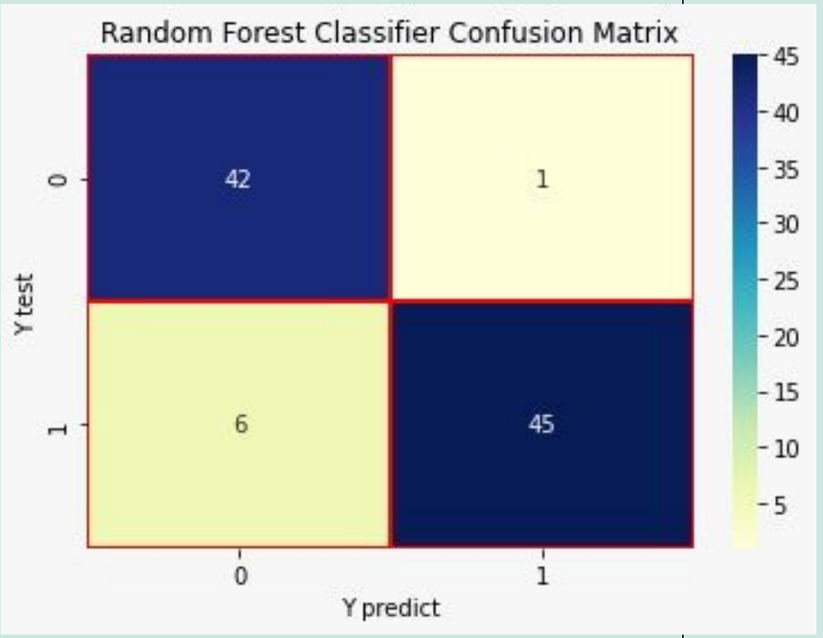
- We used 3D CNN model of the architecture described in this paper
- After some hyperparameter tuning, the best validation accuracy we could get was about 60%

Results & Conclusion

Top 4 models

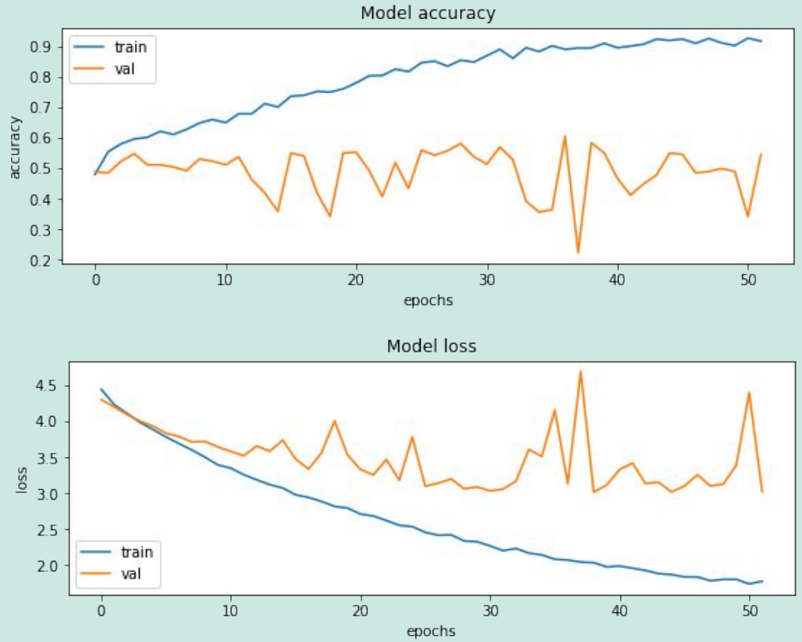
	Model Name	Feature Scaling	Accuracy	Recall	Precision	F1	training time (ms)
78	Random Forest Classifier	none	0.925532	0.882353	0.978261	0.927835	135.991096
80	Random Forest Classifier	Min-Max Scaler	0.925532	0.882353	0.978261	0.927835	163.813591
82	Random Forest Classifier	Standard Scaler	0.925532	0.882353	0.978261	0.927835	140.283108
105	Boosting Classifier	Polynomial	0.925532	0.882353	0.978261	0.927835	274.342537





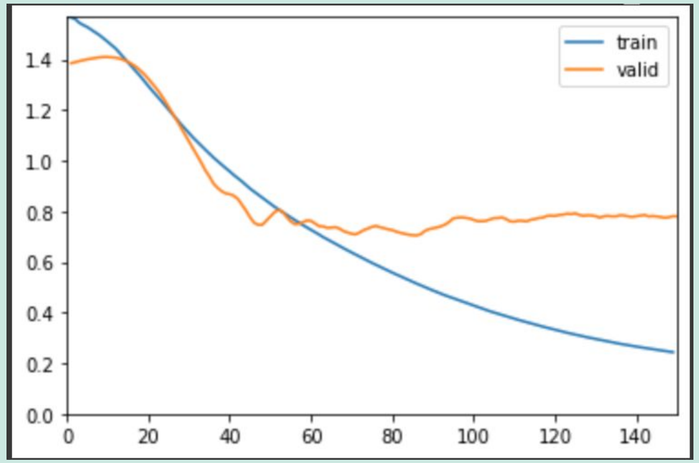
Results & Conclusion

3D Convolutional Network



- The 3D CNN network did not perform well - Validation Accuracy ~ 60%
- The MRI scans available in Oasis dataset was very less and neural networks generally require lot more data
- Also, the classes in the model were very imbalanced (very few moderate AD samples and very high nondemented samples)

Neural Networks & Embeddings



- Accuracy with this approach was reasonable - Validation accuracy ~ 81%

* Although the graph makes it look like the model is overfitting, we use callbacks to make sure the best model is used at the end.