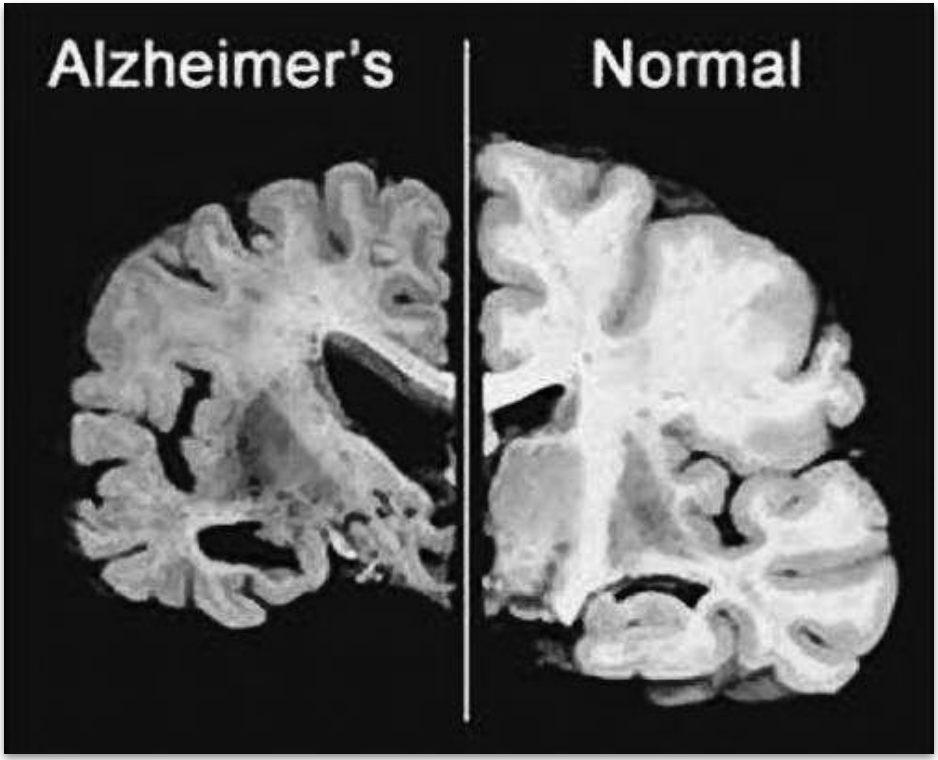
Dementia Prediction



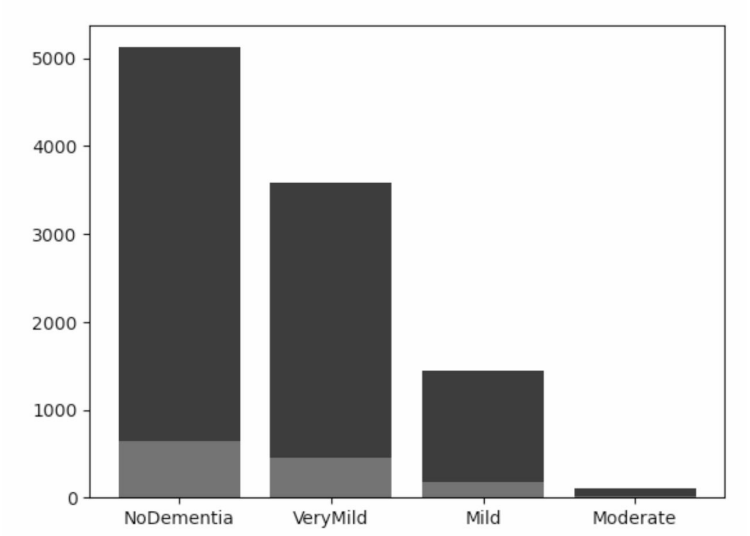
Problem

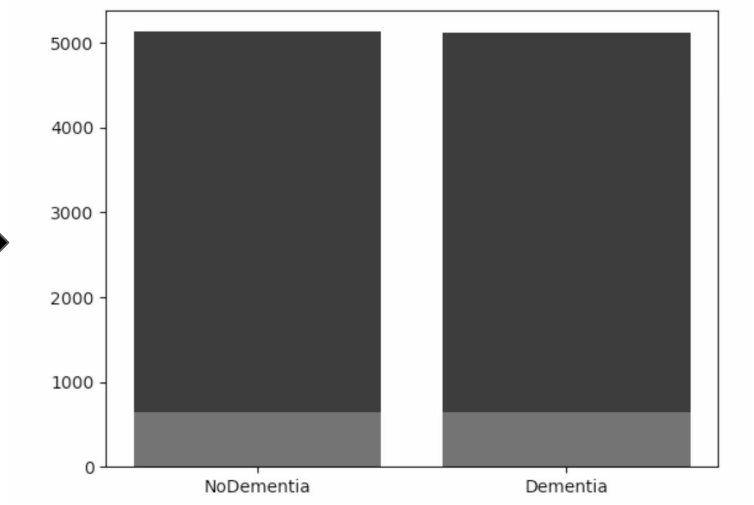
Dementia is a syndrome in which there is deterioration in cognitive function beyond what might be expected from the usual consequences of biological aging.

- Currently more than 55 million people live with dementia worldwide, and there are nearly 10 million new cases every year.
- Dementia is currently the seventh leading cause of death among all diseases and one of the major causes of disability and dependency among older people globally.
- Dementia has physical, psychological, social and economic impacts, not only for people living with dementia, but also for their carers, families and society at large.

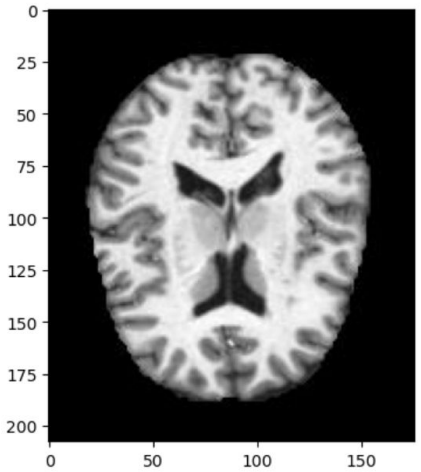
— Data Distribution —

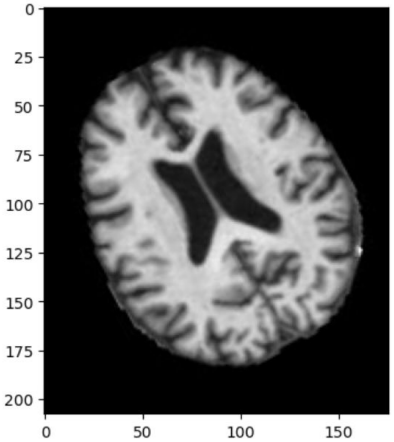
For the first iteration of the project, a binary classification was created with dementia of any kind on the one side and a “normal” brain on the other. This perhaps led to a decrease in performance, so the next step of the project is to retain the categories while weighting due to the imbalanced data set.

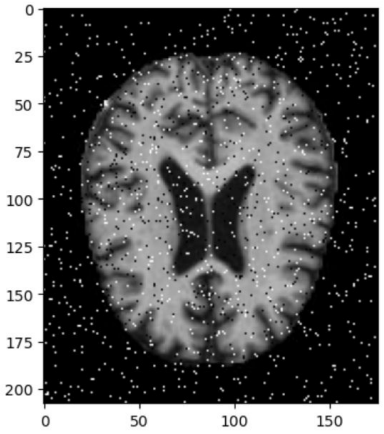




— Image Augmentation —







In order to avoid overfitting, image augmentation was manually completed.
Some of the methods used include:

- horizontal flip
- vertical flip
- image noise
- rotation

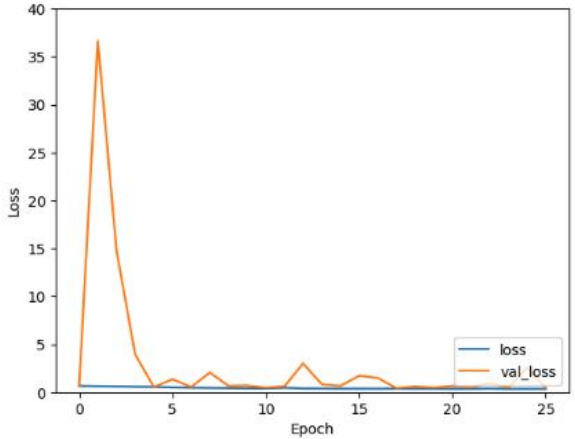
Model Comparison

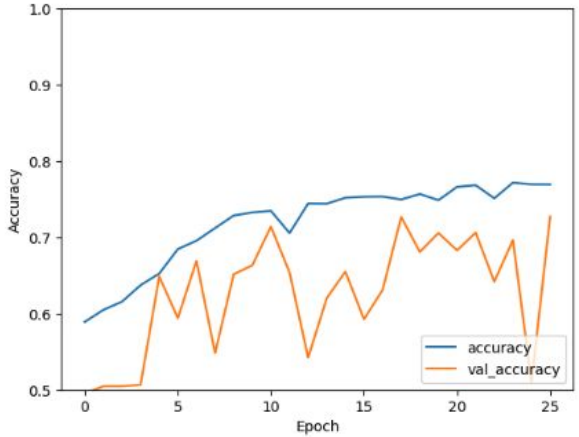
Model	Accuracy	Selected
VGG19	.72	no
Resnet50	.74	no
CustomCNN	.78	yes

Summary of Findings

Each model became progressively better with each adjustment. Ultimately, a custom trained model performed better overall than transfer classification models such as VGG19 and Resnet50. Starting with a CNN trained on brain scans led me to believe that a transfer learning model would perform better. However, the subtleties of different kinds of brains (dementia, mild dementia, no dementia) is perhaps lost on any classifier not specifically trained with the brain as the focus of its training.

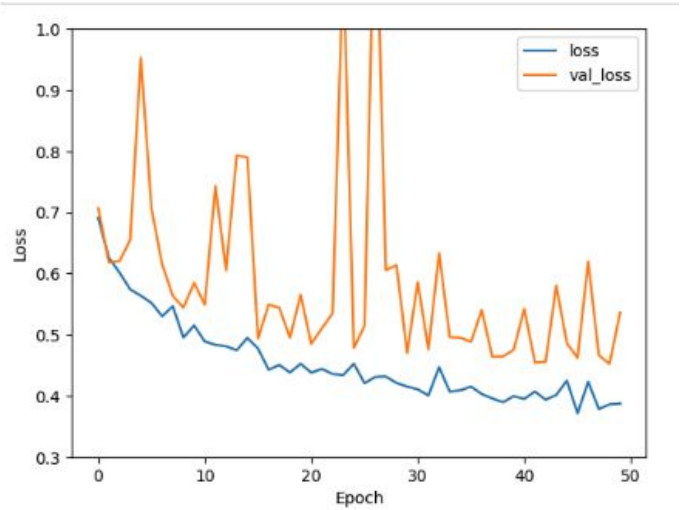
Resnet 50

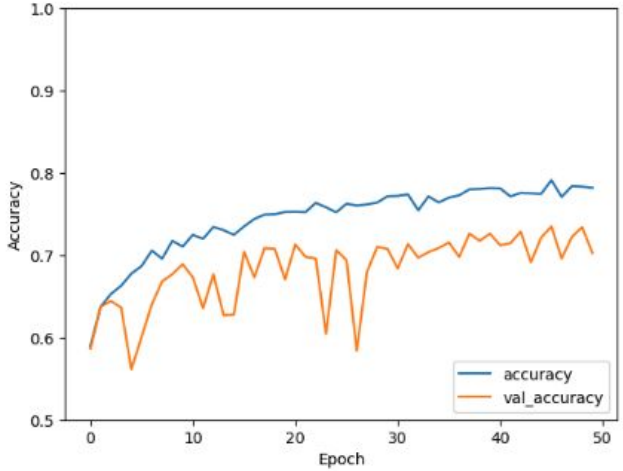




- Resnet training had a highly irregular validation loss that spiked at 36, which means that during the first few epochs the model was entirely incapable of classifying the input data. Perhaps a different loss function would perform better, or perhaps this is the result of using binary accuracy as opposed to accuracy as a metric for the model.

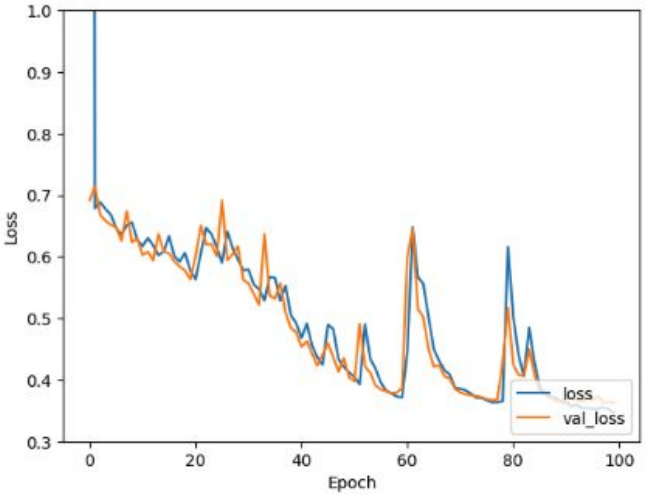
VGG19

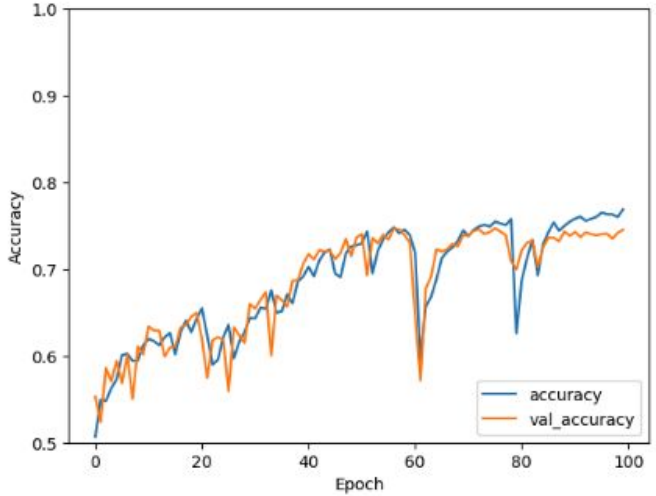




- VGG training was particularly slow (50 epochs took 3 days to complete, whereas the 40 epochs of resent took less than 24 hours). Additionally, the accuracy achieved was less than that of resnet. There was no reason to investigate this model any further given the performance and training time.

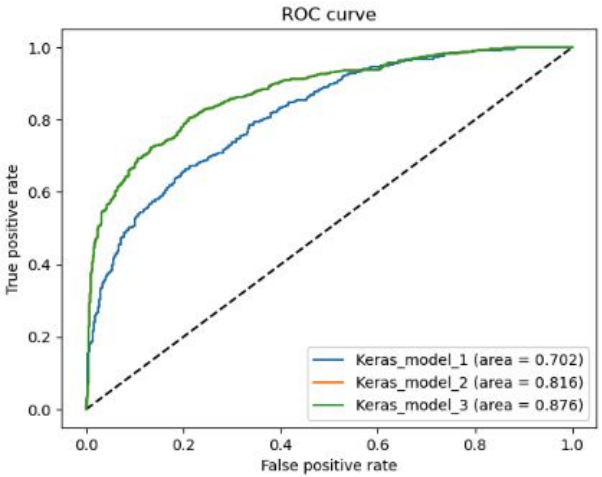
CNN

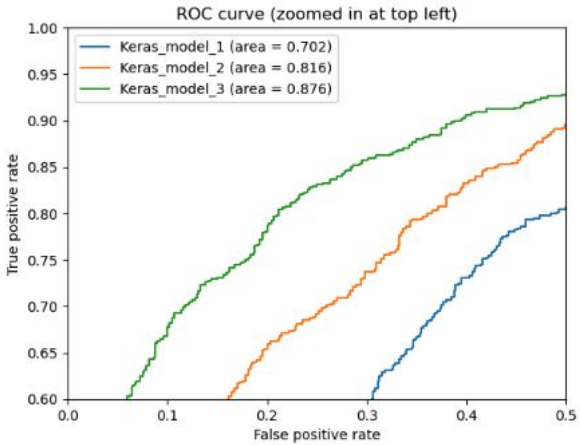




- The above is the loss and accuracy of the best performing CNN model. Arguably, there is a separation around epoch 85 between accuracy and validation accuracy, but further tests with early stopping could be run to determine how relevant the apparent separation is at that time. Put simply, there's a chance that the CNN begins to overfit at that time, but the conclusion to that question could only be determined with further questioning.

CNN Method





The following adjustments were made between models 1 and 2:

- Layers from (32, 64, 63) to (32, 64, 128, 128)
- Dropout of 0.5 after the final convolutional layer
- Created a batch size of 256
- Learning rate from .001 (standard) to .0003
- Epochs from 10 to 50

The following adjustments were made between models 2 and 3:

- Epochs from 50 to 100

The reasons only Epochs were adjusted is that the model seemed to be performing considerably better and didn't overfit.

Conclusion & Further Research

Conclusion:

After the final adjustments, accuracy went from .60, to .73, to .78 and AUC went from .70, to .81, to .87 between the three models. Conceivably, based on the increase in performance and the apparent lack of overfitting, further Epochs could be introduced. However, the additional 50 epochs only introduced 5 points in both accuracy and AUC.

The accuracy of the model increased by .18 between three versions, and arguably there is room to cross the .80 threshold that every model in this test failed.

Further Research:

- Introduce more layers on the custom cnn, with early stopping
- Refine layer architecture
- Apply similar model to a non-binary classification, which can underscore the possibility of early detection