A Multimodal Deep Learning Model for Early Detection of Alzheimer's Disease

Karina Martinez

December 8, 2023

1 Introduction

Alzheimer's disease is a brain disorder that primarily affects cognitive functions, memory, and behavior. It is the most common cause of dementia, affecting millions of individuals worldwide [3]. The disease typically manifests in older adults, although early-onset Alzheimer's can occur in individuals under the age of 65. Detecting Alzheimer's disease is challenging, however, because most patients experience a sporadic form, characterized by a late onset [1]. Therefore, it is important to devise different methods to identify early markers of the disease.

The aim of this project was to use deep learning models to predict the severity of dementia by descriminating between mild, very mild or no cases, based on MRI images and associated metadata obtained from the study "OASIS-1: Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults [5]."

1.1 Outline of Shared Work

All members of the team made major contributions to this group effort.

- Proposal: Edison and Nina
- Data assessment/acquisition: Nina and Karina
- EDA: Karina and Edison
- Models: Each team member created their own model, then helped to optimize each other's models and debug
- Combining code: Karina
- Streamlit: Edison
- Final Report: All members

2 Residual Network (ResNet) with Metadata

The availability of diverse biomedical data, such as medical imaging, genome sequencing and electronic health records, has paved the way for the emergence of multimodal artificial intelligence in health. Although most applications to date have focused on one data modality, there is a growing trend toward the use of deep learning architectures to integrate multiple data types [2]. Inspired by a paper on feature fusion using neuroimages from brain tumors, I aimed to combine the output of a convolutional neural network with demographic and derived anatomic volumes [6].

Residual Networks were first introduced in 2015 by He and colleagues [4] and have become widely used in computer vision tasks, and particularly image classification. These deep convolutional neural networks effectively address the challenge of training very deep neural networks through the use of

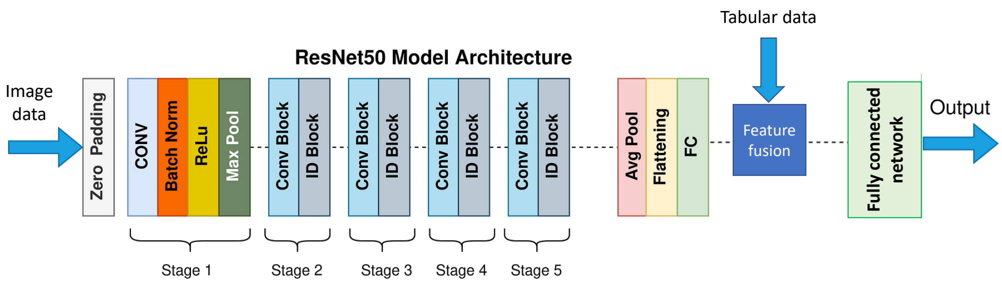

Figure 1: ResNet50 with Metadata architecture. Adapted from

residual blocks, which contain skip connections. By skipping over some layers, residual networks mitigated the vanishing gradient problem that results when many convolutional layers are stacked together.

ResNet50 is comprised of 50 layers, with a modular architecture that allows for stacking and repeating blocks to achieve its depth. Due to its success on various image recognition benchmarks, ResNet50 is often used as a pre-trained model for transfer learning in various computer vision applications. For this reason, a ResNet50 architecture was used as the backbone for image feature extraction.

The ResNet50 classifier was replaced with a concatenation step, where the tabular metadata features were fused with the image features. The combined features were then connected to a 5-layer fully connected network (FCN). Each step of the FCN included a dropout layer, batch normalization and ReLU activation function. The ResNet with Metadata architecture is presented in Figure 1.

3 Individual Contribution

3.1 Data assessment/acquisition

The team decided to work on a medical problem so I volunteered to help look for imaging data in biomedical databases due to my background in bioinformatics. I suggested we search for a dataset with accompanying metadata so that we could attempt to integrate multiple modalities in our models. I found several datasets on the NIH's Protein Data Commons and Nina found the OASIS Brains dataset on Kaggle, which we ultimately chose to work with. The metadata was not available through Kaggle, but I located the correct study and file directly through the OASIS website.

3.2 EDA

For the EDA, I prepared statistics and plots based on the demographic, clinical and derived data. I suggested that the imbalance of the dataset could be offset by eliminating patients under 60 years old from our models. In addition to mitigating imbalance, this would also ensure that all patients in our dataset were age-matched, as brain size is known to decrease with age.

3.3 Models

From the outset, my goal was to try two models: one CNN with integration of image and tabular data, and a 3D CNN. However, after training the ResNet50 with Metadata model, I switched gears to combining our team's code into one file and debugging. I discovered a serious data leakage issue with our models and focused on fixing that rather than pursuing the 3D CNN architecture.

3.4 Combining code

I generated a data file that contained the image paths and metadata and wrote functions to ensure the combined code file was interoperable. I then integrated each team member's models into my combined file, debugged, and helped to integrate and test the tabular data and attention with the other models. I created images and final models for the Streamlit presentation.

While testing all of our models on the combined code, I realized that we'd all achieved 100% accuracy on the test data. While we had accepted that the problem must be easier to solve than we'd initially thought, these results seemed improbable. It was at this point, at the very end of our project, that we realized the oversight.

All of our initial models were trained on train test splits that contained separate image slices from the same scans. Because we had not stratified the patients into different splits, there was essentially test data in the training set. I then updated our preprocessing codes to account for this oversight.

4 Results

The ResNet50 with metadata model was trained with MRI slice images and three tabular features, eTIV, ASF and nWBV, for 20 epochs with batch size 64. The loss function was Focal Loss ($a = 1$, $y = 2$) and the optimizer was Stochastic Gradient Descent with $momentum = 0.9$ and $learningrate = 0.001$. My initial models used an Adam optimizer, but to avoid overfitting, I switched to SGD with momentum.

The results presented in Figure 2 show that the validation loss was actually lower than the training loss, while the training accuracy was higher than that of the validation. Possible reasons for this are that the validation set is easier than the training set, the validation set is too small, or it wasn't properly sampled. Because each of the MRI images was treated as a separate observation, the same patient metadata was presented multiple times as slices from the same MRI scan were presented to the model. The pleateau of the train loss, and the low accuracy of both sets could also be due to sub-optimal concatenation of the features from the image and tabular data.

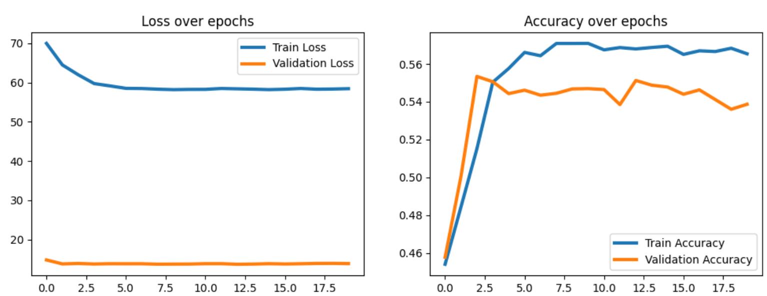

Figure 2: Loss and Accuracy on training set and validation set over 20 epochs

Metric	Value
Accuracy	0.58
F1 Score	0.35
COH Score	0.09

Table 1: Performance Metrics of the ResNet with Metadata model

The accuracy score of the ResNet plus metadata model was the highest of the three models, while the F1 and Cohen's Kappa scores fell in between the other models (Fig 1). The confusion matrix indicates that this model performed best on the majority class, non-demented, but underperformed on the minority classes compared to the other models (Fig 3). The high number of false negatives in the minority groups indicates that this model was biased toward the majority group, and would not be helpful as an early prediction tool for dementia.

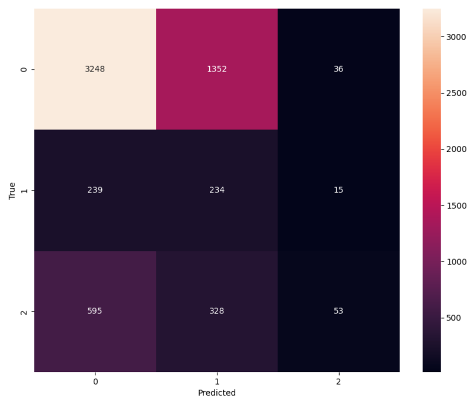

Figure 3: ResNet with Metadata Confusion Matrix. 0 refers to 'No dementia', 1 refers to 'Very mild dementia' and 2 refers to 'Mild dementia'.

5 Code Percentage

I used an example from the website https://rosenfelder.ai/multi-input-neural-network-pytorch/ to implement the feature fusion in my model, but I copied very little code from the internet. I estimate that 5% of the code I submitted was copied, and the other 95% was written by myself or based on the exam templates from this course.

6 Summary and Conclusion

In conclusion, our project developed three different models to predict the severity of Alzheimer's disease based on MRI images. Although the the performance of the models was not optimal, each member of the team learned how to implement a new deep learning architecture. The metrics for my multimodal ResNet modal were similar to the gMLP and CNN with Attention models of my teammates, indicating that none of these models was particularly effective at identifying the patterns of dementia in MRI images. In the future, it would be interesting to implement a 3D CNN, however a dataset with many more patients would be ideal.

References

[1] Alzheimer's disease. *Nature Reviews Disease Primers*, 1, Oct. 2015.

[2] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol. Multimodal biomedical ai. *Nature Medicine*, 28(9):1773–1784, Sept. 2022.

[3] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones. Alzheimer's disease. *Lancet*, 377(9770):1019–1031, Mar. 2011.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[5] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 19(9):1498–1507, Sept. 2007.

[6] D. Nie, J. Lu, H. Zhang, E. Adeli, J. Wang, Z. Yu, L. Liu, Q. Wang, J. Wu, and D. Shen. Multi-channel 3d deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Scientific Reports*, 9(1), Jan. 2019.