Classification of Structural MRI for Early Alzheimer's Diagnosis

Jeanne-Marie Linker
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC 28213

Aniqa Arif
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC 28213

Claire Ardern
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC 28213

Deepthi Chinthanippu
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC 28213

Abstract

We propose and compare various machine learning models to diagnose changes in brain structure as noted on Magnetic Resonance Images for Alzheimer's Disease patients. Additionally, we proposed to train on additional publicly available datasets, such as from the Alzheimer's Disease Neuroimaging Initiative (ADNI), to improve the accuracy of our model.

1 Introduction

1.1 Problem Statement

The diagnosis of Alzheimer's disease, a neurodegenerative illness, from MRI scans can be a difficult and time-consuming process for medical practitioners. Over time, several variations of convolutional neural networks (CNNs) have been proposed to address the difficulties of accurately diagnosing dementia from MRI scans. This project is a continuation of the search for a machine learning architecture that can accurately and efficiently diagnose the stages of Alzheimer's disease from MRI scans.

In the completion of this project, three different model architectures were implemented and tested to evaluate the accuracy of classification between four stages of Alzheimer's disease: non-demented, very mildly demented, mildly demented, and moderately demented. These three models include a simple CNN, the pretrained ResNet-50 model, and the U-Net model with a ResNet-50 backbone. The last two of the three models allowed for the use of transfer learning, which was expected to improve classification accuracy. The performance of each of these models are compared, to assess precision, recall, and accuracy. By examining the outcomes, we hope to identify a model that can improve the process of diagnosing dementia from MRI scans.

1.2 Motivation

Millions of people worldwide are affected by Alzheimer's disease. The disease not only causes memory loss but also kills more people than breast and prostate cancer combined. For medical professionals, the only confirmed way to diagnose Alzheimer's is through a brain biopsy, an invasive and potentially dangerous process. Any aid that imaging can give a doctor in diagnosis will be impactful. Technological advances with machine learning models could allow for more prompt and effective treatment and increase diagnostic accuracy.

1.3 Challenges

Data acquisition is a common problem in machine learning tasks.Though we were able to locate a suitable dataset via Kaggle, we decided to augment the dataset with additional images in order to improve the performance of our models. Our team applied for and received admittance to the Alzheimer's Illness Neuroimaging Drive (ADNI) chronicle of clinical images. However, with this new data, finding, sorting, downloading, and converting DICOM images to PNG format proved to be a time-consuming and complicated task. In addition, these MRI images included the skull and facial features of each patient, which was not the case for the original dataset. It became clear that segmentation was necessary to remove the brain from the skull in each of the MRI images to ensure performance and protect patient privacy. However, we had limited success with the segmentation process, despite the fact that there are existing programs designed specifically for skull stripping in MRI images. In order to devote more time to the completion of the machine learning models, the preprocessing of the ADNI data was not completed and this additional data was not used.

The second challenge that was encountered in the process of completing the project was the class imbalance of the dataset. In comparison with the other classes, very few samples fell into the moderate demented class. For this reason, our model was not as effective in differentiating between the classes as we had hoped. In an attempt to mitigate this issue, we performed basic data augmentation techniques such as image mirroring and image rotation to provide additional data in each of the classes. However, as we began with a limited number of images in the moderate demented class, the augmented data did not provide a significant improvement to the class imbalance.

The final challenge we faced in the completion of this project was communication and collaboration. With our team consisting of five busy graduate students, it was difficult to effectively communicate with each of the team members in a timely manner. This became especially prevalent as we approached the project deadline and some members of the team became unresponsive to communication. Despite this, the remaining team members were able to successfully complete the project and create an effective model that has the potential to make important contributions to the field of medical image analysis.

1.4 Summary of Approach

As there are several ways to tackle the problem of image classification in the world of machine learning, our team decided to pursue various approaches. First, a simple CNN for classification was designed with the expectation of a low performance to use as a baseline. Secondly, the U-Net model with a ResNet-50 backbone was implemented with the expectation of the greatest performance due to the use of transfer learning. Lastly, the pre-trained ResNet-50 model was implemented to test the efficacy of classification on MRI images. By implementing all three of these models, our team was able to compare the resulting accuracies to identify the best model for the task of Alzheimer's diagnosis.

We will examine the significance of various metrics like precision and recall as well as the effect of class imbalance. The fundamentals of neural networks, with an emphasis on CNN, the significance of image classification in the healthcare sector, various machine learning classification algorithms, and deep learning are all part of our strategy. Additionally, we will investigate how to achieve a bias-variance trade-off to avoid overfitting and underfitting, how to effectively preprocess data, perform feature engineering and ranking, and how to improve machine learning models with the help of hyperparameters.

2 Background and Related Research

2.1 MRI Imaging in The Diagnosis and Prognosis of AD

Alzheimer's Disease (AD) is a neurodegenerative disorder that is marked by a decline in cognitive ability, and is characterized by a pattern of atrophy that begins in the medial temporal lobe (Scahill et al., 2002). While sMRI imaging has been used for many decades to exclude other potential causes of cognitive decline, its use as a diagnostic tool for AD is becoming more widespread due to advances both in imaging technology and in the application of machine learning techniques to the analysis of such images (Johnson et al., 2012).

For AD patients, early intervention is critical for preserving level of function and slowing the disease (Rasmussen & Langerman, 2019). Since there is a gap between the time when AD can be detected via imaging techniques and the onset of clinical symptoms, early detection via the non-invasive methods like sMRI is vital. In fact, Burton et al. (2009) note that MRI-based assessment of temporal lobe atrophy has a positive predictive value of 80-85% for AD.

2.2 Use of Machine Learning in Structural Magnetic Resonance Imaging (sMRI) Classification for AD

Since the mid 2010s, various studies have been conducted to determine the utility of Machine Learning algorithms for the classification of sMRI images. A summary table is provided (Table 1). Most published studies on this topic to date have performed binary or pairwise classification on relatively small datasets. Amazingly, even with such small datasets and despite the fundamental differences between models, all of the studies cited below have demonstrated strong performance in the task of AD classification from sMRIs.

Authors (year)	ML Model	Classification Type	Subjects	Accuracy
Moradi et al. (2014)	SVM/RF	Binary	200 AD / 231 HC	81.72%
Dybra et al. (2015)	SVM	Binary	78 AD / 72 HC	85%
Yang, Rangarajan, & Ranka (2018)	3D CNN	Binary	47 AD / 56 HC	79.4%
Kruthika, Rajeswari, Maheshappa (2019)	AE, 3D CNN	Pairwise Classification	345 AD / 991 MCI / 605 HC	93 - 94.6%
Wang et al. (2018)	3D CNN	Pairwise Classification	221 AD / 297 MCI / 315 HC	93.6 - 98.83%
Battineni et al. (2020)	NB, RF, SVM, Gradient Boosting	Binary	78 AD / 72 HC	98%
Tufail, Ma, & Zhang(2020)	2D CNN with transfer learning	Binary	90 AD / 90 HC	99.45%
AlSaeed & Omar (2022)	ResNet50 + CNN, SVM, RF	Binary	314 AD / 427 NC	80% - 99%

Figure 1: Table 1. Note the MCI stands for mild cognitive impairment, AD for Alzheimer's Disease, and HC for health control.

Early ML analysis of MRI images for the diagnosis of AD relied heavily on Support Vector Machine (SVM) and Random Forest (RF) models. Moradi et al. (2014) used low density separation to construct an MRI biomarker, and then integrated data about age and cognitive measures to differentiate between patients with stable MCI and those who were likely to "convert" to AD. Dybra et al. (2015) didn't use the MRI images directly for their analysis, but instead used measurements taken from the images to use as input for their SVM. These models were fairly successful (80 - 85% accuracy), but were eclipsed by later, more complex models. That said, Battineni et al. (2020) achieved 98% classification accuracy using RF and gradient boosting with patient data that included numerical features pulled from MRIs.

With more widespread use of Convolutional Neural Networks (CNNs) for image analysis in the early 2010s, CNNs became popular models for AD diagnosis from brain imaging data. In 2018, Yang et al. utilized a 3-dimensional Convolutional Neural Network (CNN) to analyze 3-dimensional MRI

images.The 3D CNN was not as accurate as earlier SVM models, and the authors pointed to our limited ability to explain what the CNN is perceiving in terms of visualization. Also in 2018, Wang et al. used an ensemble method on 3D MRIs, whereby a convolutional layer was followed by a dense block (A), downsampling, another dense block (B), a fully connected layer, and then a softmax layer. This was done with 5 DenseNets with varying hyperparameters. When the results of those DenseNets were fused via a probability-based model, the accuracy on the testing data set was 98.83%. The main downside of this method is that it is computationally expensive.

Tufail and Zang (2020) used transfer learning from Inception 3 and Xception with a custom CNN to perform a binary classification of MRI images as either healthy or having AD with a high level of success. They stated definitively that intensity rescaling is necessary for CNN training - a fact that we took into account when building our own models. The authors also noted that classification is most difficult between classes that are near neighbors (mild AD vs. healthy, for example). They state their intention to expand their work into multiclass classification.

Another article of interest is the analysis of AlSaeed & Omar (2022) using transfer learning from ResNet50 applied to softmax, SVM, and RF layers for a binary classification task. The most accurate combination was ResNet50+softmax at 99%.

These are exciting results that bode well for the future of AD diagnosis via the combination of MRI data with machine learning - specifically with models that include transfer learning and CNNs. In our study, we extend the work of AlSaeed & Omar to multiclass classification (HC, very mild dementia, mild dementia, moderate dementia) on a very large dataset (6500+ images).

3 Methods

3.1 Datasets

The primary dataset for this study was downloaded from Kaggle (Kumar & Shastri, 2022). The documentation on the origin of the dataset is incomplete. According to the provided information, the data was collected from the following public repositories:

- The Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Alzheimers.net
- A 2021 publication by Kamal, et al.
- Data.gov
- A 2020 publication by Bae et al.
- The European Prevention of Alzheimer's Dementia (EPAD)

All images in the dataset have been resized to 128 x 128 pixels. The images are labeled into four classes:

- Non Demented (3200 images)
- Very Mild Dementia (2240 images)
- Mild Dementia (896 images)
- Moderate Dementia (64 images)

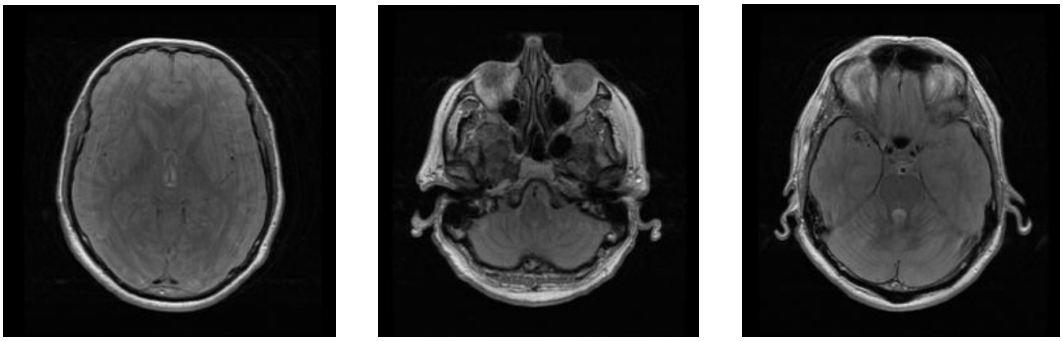
It's not clear what method was used to delineate levels of dementia, though we note that the Alzheimer's Society of the UK describes the levels of dementia as mild, moderate, and severe.

We proposed to add to this dataset by adding images from the ADNI database, which is a large and formally published MRI dataset from the University of Southern California (ADNI, n.d.). This was undertaken for several reasons. First, downloading raw data from a publicly available repository would give us the opportunity to perform for ourselves some of the preprocessing tasks as outlined in the lettered steps above. It would also allow us to demonstrate the effectiveness of our models on 'real world' data. Finally, finding additional examples of MRI scans labeled as 'moderate dementia' would allow us to address the imbalance between the classes of our dataset.

We applied for and were granted permission to access the ADNI database, download the images contained therein, and use the images for research purposes. From the Image and Data Archive (IDA) dashboard, images can be filtered according to modality, weighting, image description, and Mini-Mental State Exam (MMSE) score - a measure of dementia.

3.2 Data Segmentation

The ADNI MRI data consists of full scans of MRI acquisitions, including metadata and anatomical features of the subject scanned. It is important to remove anatomical features from the dataset due to the random and unrelated nature of facial and skull features, that would not help us learn from the brain structure itself. Additionally, facial features that can be regenerated from MRI slices fall under protected health information and need to be removed and is part of the anonymizing step in our pre-processing. We made some significant effort in trying to segment the brain from the skull. At first we used python libraries to threshold the skull from the brain using Otsu's method (Barnes, 2022). Not having success with the separating the component of the brain from the skull entirely, we tried some open source and published toolsets (Van, 2023)(Hoops, et al., 2023), after which we realized that the software (FreeSurfer, n.d.) on which the tools were based was a resource and time intensive task that was beyond the scope of this project.



(a) Slice 1 (b) Slice 2 (c) Slice 3

Figure 2: Example MRI Slices Without Segmentation

3.3 Data Pre-Processing

The Kaggle dataset (Kumar and Shastri, 2022), is a pre-processed dataset. Though we did not find the exact sources of the complete dataset, nor could we find the methods used for pre-processing, from reading the literature (Tufail and Zang, 2020, for example) and examining the images, we inferred that likely pre-processing that was done was: a) anonymizing the images b) removing the skull from the MRI image, c) denoising the image from noise from the acquisition, d) a smoothing filter over the image for better model performance, e) pixel scaling for efficient learning f) resizing the image to a size suitable for the machine learning model.

Determining our steps, we then proceeded to perform them, from which first, we had to read our DICOM standard images into python to manipulate them. We did that using a library pydicom (Mason, D. L., et al., 2023), and extracted our pixels from the metadata. To remove noise from the image, after trying a gaussian denoising step, we realized that our images were already clean enough (probably were cleaned after acquisition). Hence, denoising did not make much difference, we left out that step. Some raw images do need to be denoised, but seeing as this is a published dataset, it's likely that they have already performed that before publishing.

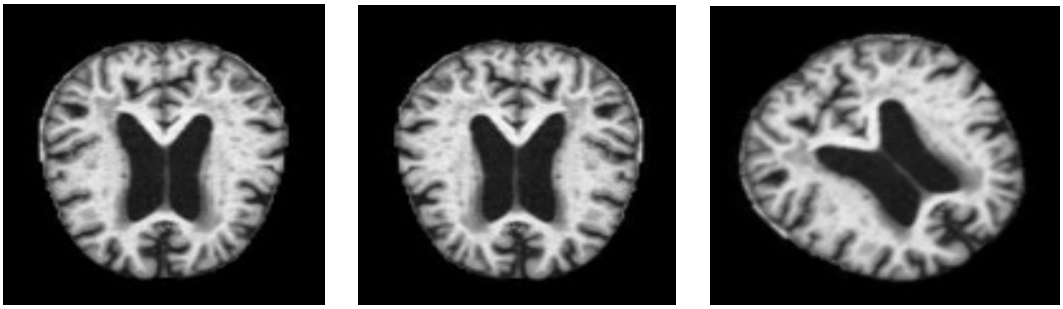
For the ADNI dataset, while selecting slices, we tried to select slices that were in the medial temporal lobe, as the literature suggests that is the region where alzheimer's patients' brain structure seems to change (Khashper & Del Carpio-O'Donovan, 2014), thus with our limited knowledge we chose slices 35-50.

Finally, we scaled our pixels, by simple normalization, changed our color channel to grayscale, resized the images using bilinear interpolation, and reshaped to form a tensor, as well as saved the images as PNG and TIFF files (Abadi, et al., 2016)(Foong, 2022)(Johnson, et al., 2012).

3.4 Data Augmentation

We decided to use straightforward data augmentation methods to expand the size of our initial dataset because of the restrictions imposed by our attempts to include additional data. The images were rotated, flipped, and shifted at random, resulting in multiple versions of each image and increasing the diversity of the dataset.

Even though we were unable to add additional data because of segmentation, the data augmentation methods that were applied to the initial dataset helped the model increase accuracy by 2% after balancing our dataset with more moderate class images which were added using augmentation techniques . Our report features the significance of cautiously considering the achievability of proposed approaches and changing them in like manner, given the assets accessible.



(a) Original Image (b) Mirrored Image (c) Rotated Image

Figure 3: Example Data Augmentation

3.5 Models

We designed three models for the classification of the images in our dataset.

3.5.1 Model 1: Simple CNN

Our first model is a simple Convolutional Neural Network (CNN). The model consists of a intensity rescaling layer (used to convert pixel intensities to a 0 to 1 scale, as recommended by Tufail and Zang (2020)), convolutional and pooling layers, two dropout layers, a flattening layer, and then three dense layers. The figure below provides a graphical representation of the model.

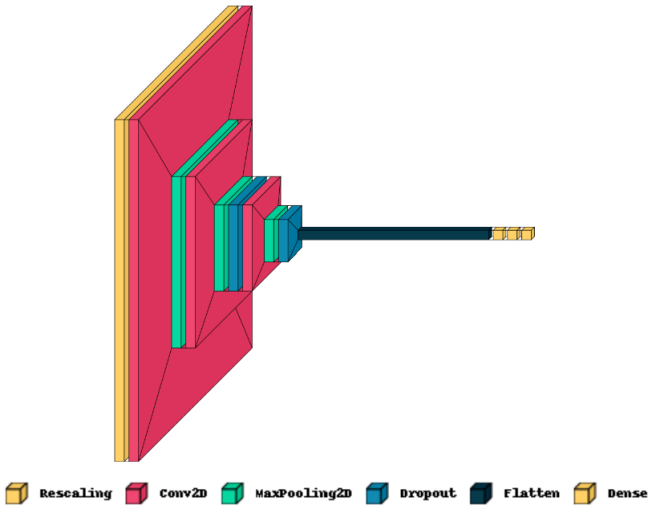


Figure 4: Simple CNN Model Architecture

Convolutional layers and the first two dense layers utilize a ReLU activation function. The final dense layer is the softmax layer.

Though the details of this model (number of layers, number of units per layer) differs from previous models designed for this task, the only substantive difference is that the model was designed for multi-class classification.

3.5.2 Model 2: ResNet-50

ResNet-50 is a popular deep learning model used for image recognition and classification tasks. It is a part of the Residual Network (ResNet) family, which was created to make training deep neural networks easier and more effective. The main idea behind ResNet is the use of residual connections, also known as skip connections. These connections help the network learn more effectively by allowing it to focus on learning new features from the input data. ResNet-50 has 50 layers in total, including different types of layers like convolutional layers, activation layers, and fully connected layers.

One of the reasons ResNet-50 is so popular is because it's often used as a pre-trained model for transfer learning. This means that the model is first trained on a large dataset like ImageNet, and then fine-tuned for specific tasks using smaller datasets. This approach saves time and helps the model perform well even with less training data.

In short, ResNet-50 is a powerful deep learning model that is great for many image recognition and classification tasks. Its unique feature, the residual connection, helps it learn better and makes training deep networks more efficient.

3.5.3 Model 3: U-Net with ResNet-50

U-Net is an architecture for a fully convolutional neural network that specializes in image segmentation, otherwise known as semantic segmentation. Image segmentation is a process of predicting where a specific object is in a picture and creating a mask showing exactly where the object is and its dimensions. This process is a variation of classification – every pixel gets assigned a class that it belongs to, which creates the mask. As this method of classification is slightly different from the standard classification model and, in our case, is used in combination with transfer learning with the addition of ResNet-50, we hypothesized that this model may be more successful than the other models included in this project.

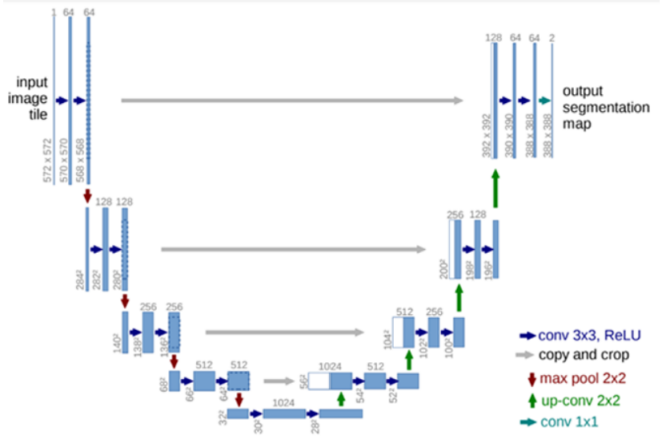


Figure 5: U-Net Basic Model Architecture

The basic U-Net architecture seen in the above figure can be divided into two main parts: the encoder and the decoder. The encoder, which makes up the left half of the U shape, uses convolutions and max pooling at each level of the architecture to compress the input and learn more complex relationships within the image data. Each level of the encoder consists of two 3x3 convolutional layers, each followed by a ReLU activation. The transition between levels is done with a 2x2 max pooling layer

with a stride of 2 for down-sampling. By the time the data reaches the bottom of the U shape, the model has a fairly good understanding of the image.

Then, the data must be passed to the decoder to return to its original state. The decoder uses essentially the same architecture as the encoder, but in reverse. As the decoder takes an input with condensed information at the bottom of the U shape, the data must be decompressed to return the image back to its original size. This is the main difference between the encoder and decoder – the model is up-sampling here. So, instead of max pooling at each level, there are 2x2 transposed convolutions to transition between levels. In doing this, the model should have an output image that is a similar size to the input image.

When using this model in combination with ResNet-50, the encoder and decoder structures are replaced by the architecture of the ResNet-50 model, which is discussed in the ResNet-50 model section of this report. This is where transfer learning is used. As the ResNet-50 model has already been trained on a large dataset and has been proven useful for image classification, minimal additional training is needed to make the model effective for classifying MRI images. This shortened training process is done by freezing the majority of layers in the model and only performing training on the final layers of the model that are used specifically to classify the images. In other words, since the model is pre-trained, only the weights of the final layers in the model are being adjusted to ensure that the model can understand and differentiate between the specifics of this particular dataset containing MRI images.

3.6 Difference from Existing Methods

In our study, we extend the work of AlSaeed & Omar to multiclass classification (HC, very mild dementia, mild dementia, moderate dementia) on a very large dataset (6500+ images). The main aspect of our three models that sets our work apart from existing methods is the use of multi-class classification.

4 Experiments

We used tensorflow image preprocessing and the splitfolders library to randomize our data, split it into a training set (70%), validation set (10%) and testing set (20%), and segment it into batches of size 64. We tried a 80:10:10 split at first, but 20% testing data seemed more reasonable so as to get a more accurate estimate of the performance of our data. We encoded our labels in integer format so as to be able to calculate sparse categorical accuracy after training. Additionally, we specified our input image size as 128x128, though we later discovered that Resnet50 performs best with 224x224 pixel size images, which is something we can incorporate into future work.

4.1 Discussion of Results

4.1.1 Model 1: Simple CNN

We trained the simple CNN model for 25 epochs using a sparse categorical crossentropy loss function with the Adam optimizer with accuracy as the metric.

The graphs in the figure below track the accuracy and loss for the training and validation data sets.

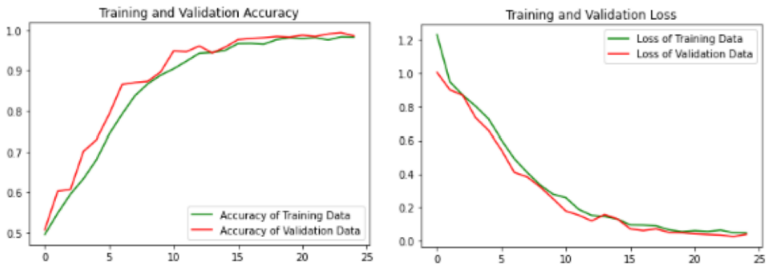


Figure 6: Simple CNN Accuracy and Loss Graphs

This model performed well, with high accuracy for both the training and validation sets. Note that the validation accuracy mirrors the training accuracy well, so we do not see signs of overfitting. Likewise, the loss decreases over time in a consistent manner, and does so for both training and validation data.

Once trained, we tested the model's accuracy with the testing data, and achieved an accuracy of 97.9%.

An examination of the confusion matrix comparing model predictions with ground truth shows that where images were misclassified, they are most likely to be confused with a neighboring class (very mild dementia misclassified as mild dementia, for example).

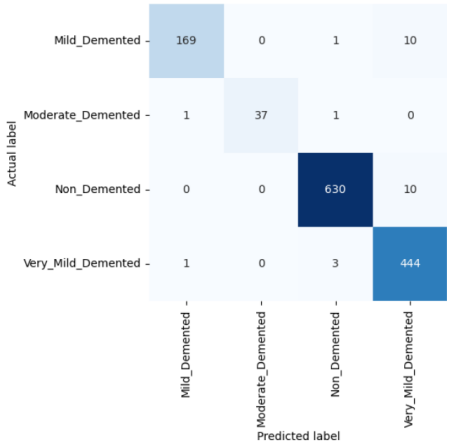


Figure 7: Simple CNN Confusion Matrix

An example of the model output can be seen in the figure below.

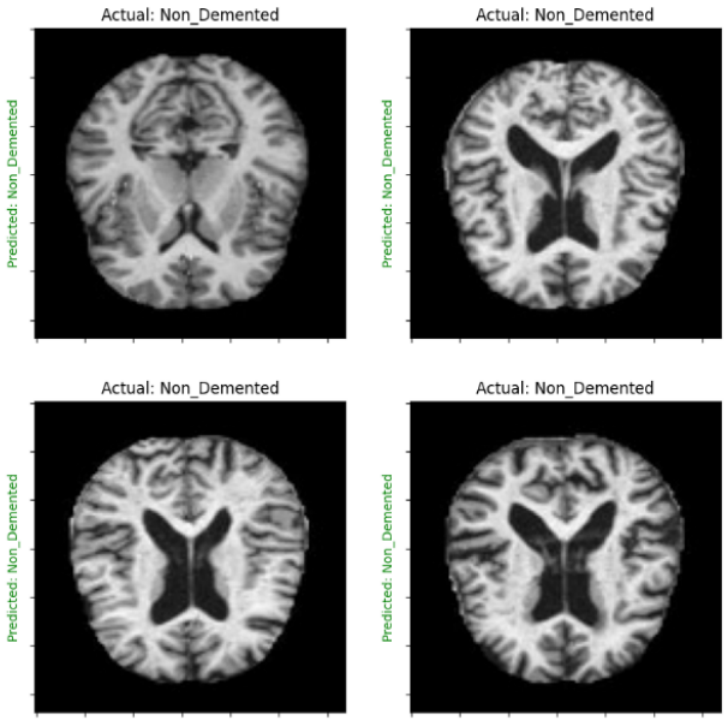


Figure 8: Simple CNN Example Output

4.1.2 Model 2:ResNet-50

We trained the ResNet-50 model for 25 epochs using a sparse categorical crossentropy loss function with the Adam optimizer with accuracy as the metric. A learning rate of 0.0001 was used for training. The learning rate was kept small so that we may retain as much information from the pre-trained model parameters as we can.

The graphs in the figure below track the accuracy and loss for the training and validation data sets.

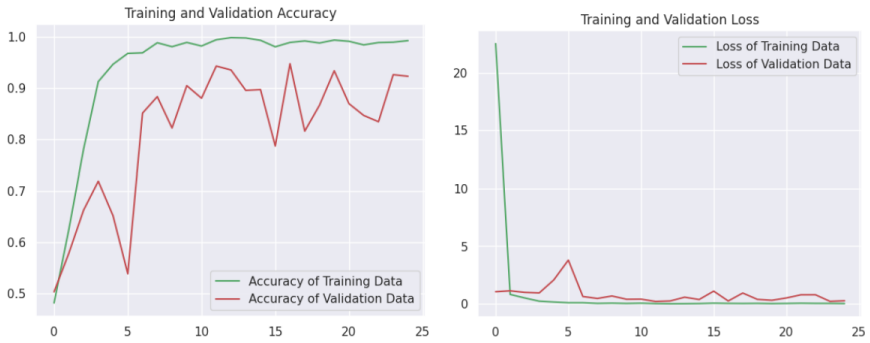


Figure 9: ResNet-50 Accuracy and Loss Graphs

This model performed well, with relatively high accuracy for both the training and validation sets. Note that the validation accuracy curve seems to fluctuate constantly, showing signs of overfitting. Likewise, the loss decreases over time in a consistent manner, and does so for both training and validation data, though there seems to be some fluctuation.

Once trained, we tested the model's accuracy with the testing data, and achieved an accuracy of 95.9%.

An examination of the confusion matrix comparing model predictions with ground truth shows that where images were misclassified, they are most likely to be confused with a neighboring class (very mild dementia misclassified as mild dementia, for example).

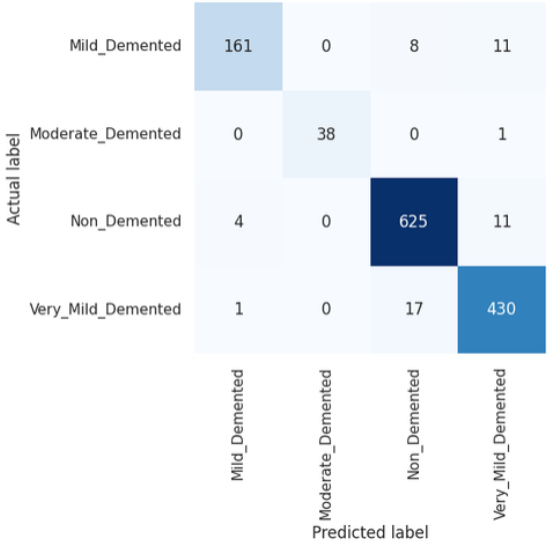


Figure 10: ResNet-50 Confusion Matrix

An example of the model output can be seen in the figure below.

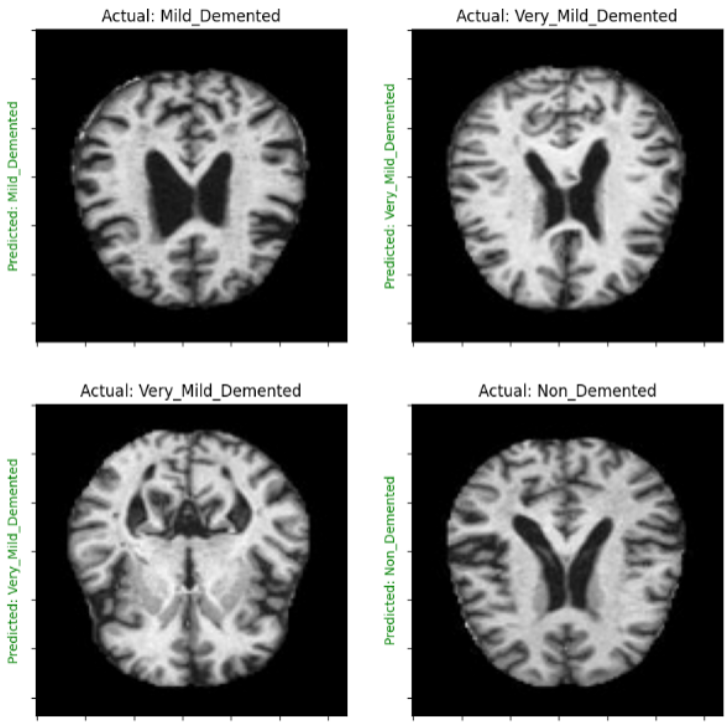


Figure 11: ResNet-50 Example Output

4.1.3 Model 3: U-Net with ResNet-50

We trained the U-Net with the ResNet-50 backbone model for 25 epochs using a sparse categorical crossentropy loss function with the Adam optimizer with accuracy as the metric. A learning rate of 0.0001 was used for training. The learning rate was kept small so that we may retain as much information from the pre-trained model parameters as we can.

The graphs in the figure below track the accuracy and loss for the training and validation data sets.

Figure 12: U-Net with ResNet-50 Accuracy and Loss Graphs

This model performed well, with relatively high accuracy for both the training and validation sets. Note that the validation accuracy curve seems to fluctuate constantly, showing signs of overfitting. Likewise, the loss decreases over time in a consistent manner, and does so for both training and validation data, though there seems to be some fluctuation.

Once trained, we tested the model's accuracy with the testing data, and achieved an accuracy of 94.9%.

An examination of the confusion matrix comparing model predictions with ground truth shows that where images were misclassified, they are most likely to be confused with a neighboring class (very mild dementia misclassified as mild dementia, for example).

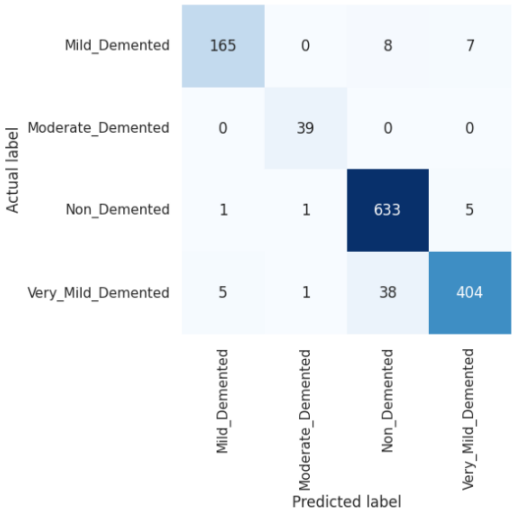


Figure 13: U-Net with ResNet-50 Confusion Matrix

An example of the model output can be seen in the figure below. Correct predictions are marked in green text while incorrect predictions are marked in red text.

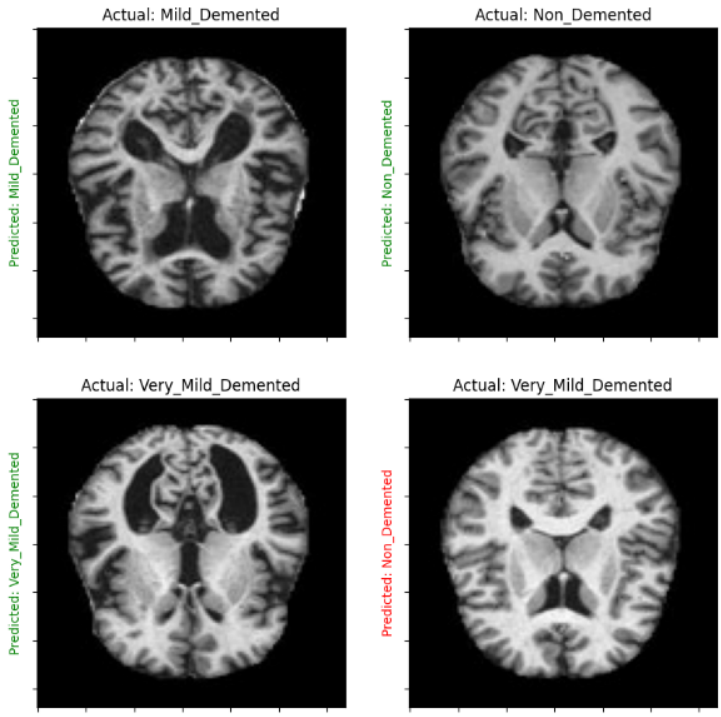


Figure 14: U-Net with ResNet-50 Example Output

5 Reflections

For our initial presentation, Dr. Lee commented that we should include dates for our citations in our presentation slides. This was fixed for the final presentation.

For our midterm report, Dr. Lee commented that our claims lacked uniqueness or novelty. He suggested that we focus on the challenges raised by moving from the binary or pairwise classifications

seen in earlier research to the multiclass classification that we proposed. The facet of this transition that proved the most problematic is the fact that our dataset, though quite large compared to the datasets used in other analyses, lacked balance between the classes.

Our initial idea for rectifying this imbalance was to introduce additional data from the ADNI database. We were unable to utilize the data that we found, as we were unable to properly segment the images to remove the skull and facial features. Instead, we focused on adding data by augmenting the images in our initial data set.

We still have an imbalance between classes, which might cause us to lose accuracy in our model.

6 Conclusions

This project successfully helped us synthesize material that we'd learned throughout the semester. It serves as a nice way to tie everything together. With a coding portion, a presentation portion, and a written report, it was an immense amount of work.

Though we are disappointed that we were unable to complete the preprocessing of the ADNI data in order to add it to our model, we are pleased with the results that we achieved with our three models. There's something to be said about the value of wrestling with a problem, even when you are ultimately unsuccessful. We learned about data acquisition, converting DICOM files to PNGs, denoising, and even segmentation. It's certainly more complicated than we'd initially thought.

It was interesting to observe, when working on the literature review, the way machine learning approaches to medical diagnostics have changed through the last 2 decades. The shift from SVMs and RFs to CNNs to CNNs with ensemble or transfer learning shows a clear trend towards the development and use of more and more sophisticated and powerful tools.

All code can be found at the following GitHub repository: https://github.com/aniqa/MRI_and_alzheimers_classification

References

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others. (2016). Tensorflow: A system for large-scale machine learning. OSDI, 16, page 265–283.

[2] AlSaeed, D., & Omar, S. F. (2022). Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning. Sensors (Basel, Switzerland), 22(8), 2911. https://doi.org/10.3390/s22082911

[3] ADNI (n.d.) Alzheimer's disease neuroimaging initiative. Retrieved February 10, 2023, from https://adni.loni.usc.edu/

[4] Bae, J. B., Lee, S., Jung, W. et al. (2020) Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Scientific Reports 10, 22252. https://doi.org/10.1038/s41598-020-79243-9

[5] Barnes, R. (2022). Using Otsu's method for skull-brain segmentation (v1.0.2). Zenodo. https://doi.org/10.5281/zenodo.7382654

[6] Battineni G., Chintalapudi N., Amenta F., Traini E. (2020) A Comprehensive Machine-Learning Model Applied to Magnetic Resonance Imaging (MRI) to Predict Alzheimer's Disease (AD) in Older Subjects. Journal of Clinical Medicine 2020;9:2146. doi: 10.3390/jcm9072146.

[7] Burton E. J., Barber R., Mukaetova-Ladinska E. B., Robson J., Perry R. H., Jaros E., Kalaria R. N., O'Brien J. T. (2009) Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: a prospective study with pathological verification of diagnosis, Brain, Volume 132, Issue 1, January 2009, Pages 195–203, https://doi.org/10.1093/brain/awn298

[8] Dyrba M., Grothe M., Kirste T., Teipel S. J. (2015) Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. Human Brain Mapping. 2015;36:2118–2131. doi: 10.1002/hbm.22759.

[9] Foong, N. W. (2022). Convert images to tensors in Pytorch and tensorflow. Medium. Retrieved May 5, 2023, from https://towardsdatascience.com/convert-images-to-tensors-in-pytorch-and-tensorflow-f0ab01383a03

[10] FreeSurfer (n.d.) FreeSurfer software suite. Retrieved March 1, 2023 from https://surfer.nmr.mgh.harvard.edu/

[11] Hoopes, A., Mora, J. S., Dalca, A. V., Fischl, B., & Hoffmann, M. (2022). SynthStrip: skull-stripping for any brain image. NeuroImage, 260, 119474. https://doi.org/10.1016/j.neuroimage.2022.119474

[12] Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). Brain imaging in Alzheimer disease. Cold Spring Harbor Perspectives in Medicine, 2(4), a006213. https://doi.org/10.1101/cshperspect.a006213

[13] Kamal M. S., Northcote A., Chowdhury L., Dey N., Crespo R. G., Herrera-Viedma E. (2021) Alzheimer's Patient Analysis Using Image and Gene Expression Data and Explainable-AI to Present Associated Genes, IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-7, 2021, Art no. 2513107, doi: 10.1109/TIM.2021.3107056.

[14] Khashper, A., Chankowsky, J., Del Carpio-O'Donovan, R. (2014). Magnetic resonance imaging of the temporal lobe: normal anatomy and diseases. Canadian Association of Radiologists journal = Journal l'Association canadienne des radiologistes, 65(2), 148–157. https://doi.org/10.1016/j.carj.2013.05.001

[15] Kruthika K., Rajeswari, Maheshappa H. (2019) CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. Informatics in Medicine Unlocked. 2019;16:100227. doi: 10.1016/j.imu.2019.100227.

[16] Kumar, S., Shastri, S. (2022). Alzheimer MRI Preprocessed Dataset. Retrieved February 10, 2023 from https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset

[17] Mason, D. L., et al. (2023) pydicom: An open source DICOM library, https://github.com/pydicom/pydicom [Online; accessed 2023-04].

[18] Moradi E., Pepe A., Gaser C., Huttunen H., Tohka J. (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage. 2015;104:398–412. doi: 10.1016/j.neuroimage.2014.10.002.

[19] Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N., & Fox, N. C. (2002). Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. Proceedings of the National Academy of Sciences of the United States of America, 99(7), 4703–4707. https://doi.org/10.1073/pnas.052587399

[20] Tufail A., Ma Y.-K., Zhang Q.-N. (2020) Binary Classification of Alzheimer's Disease Using sMRI Imaging Modality and Deep Learning. Journal of Digital Imaging. 2020;33:1073–1090. doi: 10.1007/s10278-019-00265-5.

[21] Van A. (2023) brainextractor [Source code]. https://github.com/vanandrew/brainextractor

[22] Wang H., Shen Y., Wang S., Xiao T., Deng L., Wang X., Zhao X. (2018) Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. Neurocomputing. 2019;333:145–156. doi: 10.1016/j.neucom.2018.12.018.

[23] Yang C., Rangarajan A., Ranka S. (2018) Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification; Proceedings of the Annual Symposium Proceedings, AMIA Symposium 2018; San Francisco, CA, USA. 3–7 November 2018; p. 1571.