A Comparative Study of CNN Models in Alzheimer's Detection

Jiahui Chen
jchen23@stanford.edu

Yvonne Hong
yhong23@stanford.edu

Olivia Weiner
oweiner@stanford.edu

Abstract

This study compares the multiple popular Convolutional Neural Network architectures in their performance in detecting Alzheimer's disease from MRI images and investigates the impact of data augmentation on the model performance in this specific task. We compared the performance of Vanilla CNN, ResNet-101, and DenseNet-121, both with and without data augmentation. Our dataset, sourced from Kaggle, includes 6,000+ MRI images classified into four stages of Alzheimer's. The results indicate that DenseNet-121 without data augmentation achieved the highest performance with an accuracy of 99.53% and an F1 score of 0.995. We attempted data augmentation to prevent overfitting, but it decreased the accuracy of all the models. We discussed the architectures and training methodologies of these models, including the specific data augmentation techniques applied. This study highlights the effectiveness of using CNNs for Alzheimer's detection through MRI imaging. Future work could explore increased dataset size, multimodal data integration, and CNN and conventional machine learning methods hybrid models in the field of Alzheimer's detection.

1. Introduction

Alzheimer is the most common type of dementia around the world, with early detection being critical in delaying the worsening of symptoms and improving patients' quality of life drastically. Research indicates that Alzheimer's may initiate at least two decades before symptomatic presentation, with subtle brain changes occurring. While current therapies aim to slow disease progression, early detection holds promise for enhancing patient quality of life and enabling effective management during the critical stages of decision-making incapacity.

Early research in using deep learning for Alzheimer's detection using MRI images shows promising results [14]. The significance of deep learning in Alzheimer's detection lies in its ability to analyze vast amounts of medical imaging data accurately and efficiently. Deep learning models can identify subtle Alzheimer-related patterns and biomarkers, enabling timely intervention and symptom management. Moreover, deep learning models offer objective and consistent evaluations, reducing variability across healthcare practitioners and institutions.

Research has shown significant improvement of using data augmentation to improve the accuracy of Alzheimer's detection. MRI segmentation techniques, allowing detailed analysis of tissue structures, is starting to gain popularity and a pivotal role in accurate diagnosis for many other diseases as well [17].

1.1. The Problem

The fundamental problem is Alzheimer's detection, namely using MRI imaging to classify the extent to which someone has Alzheimer's dementia: no dementia, very mildly demented, mildly demented, and moderately demented. There have been multiple papers on using CNN in Alzheimer's detection but not across the same datasets. This paper compares the performance of 3 types of CNNs and evaluates their performances.

Another aspect of the project we want to explore is to explore the impact of data augmentation on model accuracies. Researchers have studied numerous deep learning models in Alzheimer's detection but given the recent paper published on the significance of data augmentation in Alzheimer's detection, we want to investigate to what extent data augmentation improves Alzheimer's detection of different model types.

The input to our algorithm is a MRI scans of brain images. We then use different CNN models to output a predicted diagnosis for the Alzheimer stage the patient is in, out of the four classes no dementia, very mildly demented, mildly demented, and moderately demented.

1.2. Overview of Results

Our experiments indicate that DenseNet without data augmentation reaches the best performance in both F1 and accuracy. We have also observed that data augmentation decreases the performance for all the models.

2. Related Work

There have been numerous papers published in the past decades that showcase the efficacy of deep learning and CNN in the medical field, specifically in Alzheimer's detection. The methods used in different CNN papers can be categorized into the following three groups: various CNN approaches, conventional learning approaches, and ensemble learning approaches.

2.1. Conventional Learning Approaches

Support Vector Machines (SVMs) has been a popular method, and an early popular choice, in conventional machine learning for Alzheimer's disease (AD) detection using MRI images. These methods have provided a solid foundation for classification tasks due to their clear principles and relatively good performance.

Suk et al. employed a linear SVM classifier for AD classification, demonstrating its effectiveness in this domain. Suk and Shen (2013)[16], Suk et al. (2015)[4] utilized multi-kernel SVMs to integrate features from multimodal inputs, enhancing the classifier's flexibility and performance. Suk et al. (2014)[15] used a linear SVM in a hierarchical classifier setup, working alongside Deep Boltzmann Machines (DBM) for feature representation. Shi et al. combined stacked deep polynomial networks (DPN) as feature extractors with a linear kernel SVM, highlighting a hybrid approach that leverages deep learning and conventional machine learning [13].

2.2. Ensemble Learning Approaches

Ensemble learning refers to the technique that combines the predictions of multiple base models to produce a single, improved predictive model, and a popular ensemble learning algorithm using traditional machine learning methods is Random Forest. Lebedev et al. (2014)[8] tested Random Forest on the ADNI and AddNeuroMed datasets with MRI images and demographic data. Bi et al. (2020)[1] proposed a Random Forest architecture to handle multimodal data and detect brain abnormalities and pathogenic genes.

2.3. Convolutional Neural Networks

CNNs are the current state-of-the-art architecture in various medical field tasks such as detection, classification, and segmentation, including in Alzheimer's. There is a wide variety of architectures used by different scholars in Alzheimer's detection.

LeNet, one of the earliest CNN architectures, has been adapted for AD prediction with reasonable sensitivity and specificity as according to Yang and Liu in 2020 [18]. VGG uses small convolutional filters in a deep network structure, achieving high accuracy in AD detection, especially with transfer learning from ImageNet as accomplished by Jain et al [7]. ResNet addresses the vanishing gradient problem in deep networks with shortcut connections, enabling the training of very deep networks, and has been applied in 3D form for AD and MCI detection with successful outcomes [3]. DenseNet connects all layers directly, mitigating vanishing gradient issues and making full use of features, which is beneficial for small datasets, and has been used in various forms to achieve high accuracy in AD detection [5].

3. Methods

We built 6 individual models with three different architectures and each with and without data augmentation, namely Vanilla CNN, CNN with data augmentation, Vanilla Residual Neural Networks (ResNet), ResNet with data augmentation, Densely Connected Convolutional Networks (DenseNet), and DenseNet with data augmentation.

We will evaluate the accuracy, F1 score and training cost of the models and discuss the strengths and weaknesses of each model in this paper. Below is an overview of how each model architecture works and how data augmentation works in combination with each model architecture.

3.1. Model Architecture

3.1.1 Vanilla CNN

CNNs are specifically designed for image processing tasks and have demonstrated remarkable success in various computer vision applications, including image classification, object detection, and segmentation. In the context of Alzheimer's detection, CNNs can effectively analyze brain MRI images and extract relevant features indicative of Alzheimer's pathology, such as structural abnormalities or subtle changes in brain tissue. CNN is currently the architecture of choice in MRI image classification due to its extraordinary capability to achieve high accuracy. Even very simple training yields promising results, which is one of the key reasons why we are taking vanilla CNN as the baseline model.

The specific architecture of our Vainilla CNN is as follows. We utilize a simple convolutional neural network (CNN) with the following architecture: two convolutional layers, and two linear layers with dropout between them. Each layer is separated with ReLU activation. For this architecture, we reused the code from assignment 2 of CS 231N, with some slight modifications to fit our problem.

3.1.2 ResNet

ResNet is a type of CNN architecture that introduced the concept of residual learning, which helps mitigate the vanishing gradient problem commonly encountered in deep neural networks. ResNet's architecture allows for the training of significantly deeper networks by introducing skip

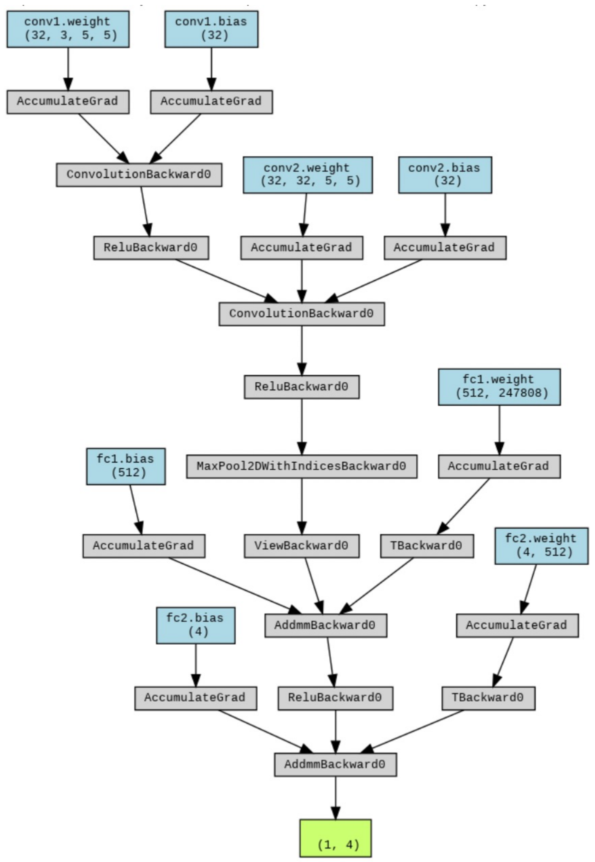

Figure 1. Vanilla CNN Architecture

connections, enabling the network to learn residual mappings. In Alzheimer's detection, ResNet's ability to handle deeper architectures may allow for better representation learning from complex MRI images, potentially leading to improved detection accuracy.

The specific ResNet architecture we have chosen is ResNet-101. ResNet-101 is a deep convolutional neural network designed for image classification, consisting of 101 layers including convolutional layers, batch normalization, ReLU activations, and fully connected layers [3]. The key innovation in ResNet-101 is the use of residual blocks, which include skip connections that help mitigate the vanishing gradient problem by allowing gradients to flow more easily through the network. This architecture enables effective training of very deep networks. It is commonly used for clinical data classification.

In our implementation using PyTorch TorchVision [11] [9] and other existing code libraries [10] [6], we customized the ResNet-101 model by adjusting the final fully connected layer to match the number of classes in our classification task. The model was trained using the Adam optimizer and cross-entropy loss. During training, the model with the highest validation accuracy is saved as the best model to prevent overfitting and using a sub-optimal model for final testing. Data preprocessing and loading were handled using PyTorch's DataLoader, ensuring efficient batch processing.

3.1.3 DenseNet

DenseNet has a dense connectivity pattern, where each layer receives input from all preceding layers and passes its output to all subsequent layers [5]. This design ensures efficient feature reuse and a good gradient flow. DenseNet's ability to capture detailed and diverse features makes it particularly well-suited for medical imaging tasks. For Alzheimer's detection using MRI images, DenseNet is an excellent choice because its architecture allows for parameter efficiency, reducing the risk of overfitting even with smaller datasets, which is often the case in medical imaging, especially true in the case of our chosen dataset. Studies have shown that DenseNet performs exceptionally well in medical imaging applications, making it a robust and reliable model for detecting Alzheimer's disease [9].

Our chosen DenseNet architecture is DenseNet 121, which includes 121 layers. We chose DenseNet 121 instead of other DenseNet variants because it strikes an optimal balance between depth, computational efficiency, and performance. While deeper networks like DenseNet-169, DenseNet-201, or DenseNet-264 offer potentially higher accuracy, they also come with increased computational costs and a higher risk of overfitting, especially when dealing with smaller medical datasets. Similarly to our ResNet implementation, our DenseNet model is customized based on the Densenet 121 architecture from the PyTorch model library and trained using Adam optimizer and cross-entropy loss[11].

3.2. Data Augmentation

To enhance the robustness and generalization of the convolutional neural network, we implemented data augmentation to enlarge our training dataset. Specifically, we augmented the data through random rotations of up to 90 degrees, random horizontal and vertical flips with a probability of 0.5 each and the application of Gaussian blur with a kernel size of 5 with a probability of 0.3. Additionally, a random resized crop operation was performed, with the scale of the crop ranging between 80% and 100% of the original image size, ensuring the output size remained consistent at 176x176 pixels.

3.3. Cross Entropy Loss

The cross-entropy loss for a single data point is given by:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

where C is the number of classes, y_i is the true label (one-hot encoded), $\hat{y}_i$ is the predicted probability for class i. We used cross entropy loss in the gradient descent of all our models.

4. Dataset

We are using a dataset from Kaggle for Alzheimer's images with 4 classes [2].

The images are split into train, validation, and test sets. The train set contains 5120 images, with 50 moderately demented, 716 mildly demented, 1803 very mildly demented, and 2551 non-demented. The validation set contains 640 images, with 7 moderately demented, 96 mildly demented, 204 very mildly demented, and 333 non-demented. The test set contains 640 images, with 7 moderately demented, 84 mildly demented, 233 very mildly demented, and 316 non-demented.

We chose this dataset because it is well-documented and maintained. It has a sufficiently large dataset size. It has 4 classes, which is very essential since a key aspect in Alzheimer's detection is early diagnosis as early diagnosis has direct correlations with better treatment options for the patients.

We applied standard score normalization to all data. We normalized pixels across the mean and the variance of pixel values across all images. For 3 of the 6 models, we applied data augmentation, and specifically we applied random rotation, random horizontal flip, random vertical flip, and random resized crop. For image data, we directly utilized the pixel values as features after preprocessing and augmentation.

Here are some examples from our dataset:

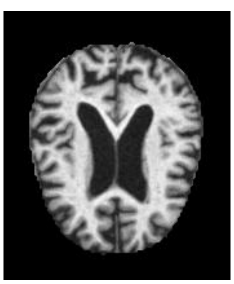

Figure 2. Moderated Demented MRI Brain Image

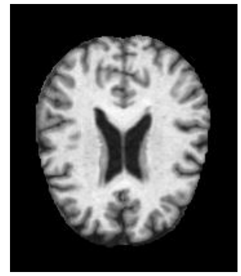

Figure 3. Non Demented MRI Brain Image

5. Experiments

For all six models, we performed hyper-parameter tuning on the learning rate since the model convergence can be sensitive on the learning rate. We initially used Adam optimizer for all the models because it combines momentum and Root Mean Square Propagation, which allows for faster convergence. It is also easy to use out-of-box and does not require much parameter tuning. While training on the vanilla CNN model with data augmentation, we realized that model training's loss remains stagnant after significant epochs. Upon switching to SGD optimizer and some parameter tuning, we were able to train the model to convergence. We used the batch size of 64 because it provides a good balance of training speed and model performance.

5.1. Evaluation Methods

We will evaluate the models with F1 scores as the northstar metric, confusion matrix and examples of common errors. We used the scikit-learn's metrics library to perform the evaluations [12].

5.1.1 F1 Scores

We're choosing the F1 score as our northstar metric because it provides a balanced assessment for the model's performance, evaluating the impact of both false positives and false negatives. It is particularly useful when there is an imbalance between the classes in the dataset, which is the case in our dataset, since most scans in the dataset, as well as in the real world, would be for patients without dementia. The number of scans of patients will serious dementia are significantly smaller.

Since we are conducting multi-class classification, we would compute the F1 score for each class individually and then average them. Our dataset has four classes (no dementia, very mildly demented, mildly demented, and moderately demented), for class i:

$$F1\ Score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i}$$

where Precision is the ratio of true positives to the total predicted positives for class i, and Recall is the ratio of true positives to the total actual positives for class i.

The overall F1 score for the model can be calculated as the weighted average of the F1 scores for each class, where the weights are proportional to the number of instances of each class in the dataset.

5.1.2 Confusion Matrix

We're choosing to evaluate the models with a confusion matrix because a confusion matrix provides a detailed breakdown of the model's predictions, showing the number of true positives, true negatives, false positives, and false negatives for each class. It is critical for doctors to understand the strengths and weaknesses of each model when evaluating the model output, and therefore the confusion matrix would be a helpful tool.

5.2. Results and Discussion

The overall comparison of model performance is shown in Table 1. Our best performing model is DenseNet without data augmentation with an accuracy of 99.53% and an F1 score of 0.995.

5.2.1 Vanilla CNN

A graph of the training and validatoin accuracy during the training can be viewed below in Figure 4. The model achieved 98.28% accuracy on our test set, with an F1 score of 0.983.

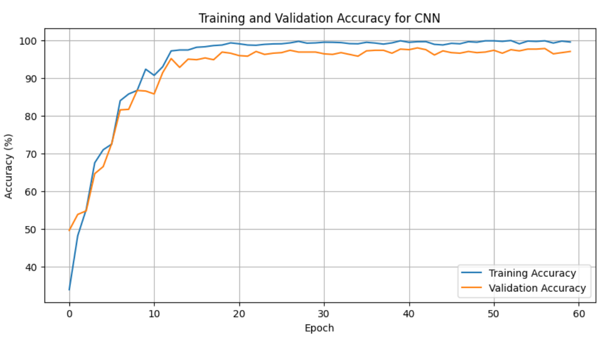

Figure 4. Training and validation accuracy

The confusion matrix of Vanilla CNN is shown in Figure 5. We can observe that the model has the tendency to label non-demented or mild-demented as very-mildly-demented. These errors are more common likely because it is the transition stage between mildly demented and non-demented, and the boundary of the distinction is not as clear to our model.

5.2.2 Vanilla CNN with Data Augmentation

A graph of the training and validatoin accuracy during the training can be viewed below in Figure 6. It achieved an accuracy of 97.81% on our test set and an F1 score of 0.978.

We see a 0.47% decrease in accuracy after data augmentation. This decrease is small in magnitude but nevertheless the same pattern appears across all our model architectures. We believe the following reasons could contribute to this

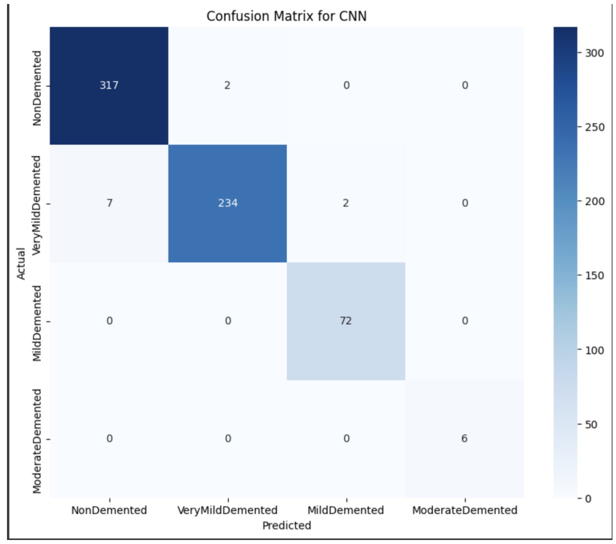

Figure 5. Vanilla CNN Confusion Matrix

pattern: 1) Increase in data complexity: the data augmentation techniques we employ increase the complexity of the data, which can make the learning process harder for the models. Further research is needed to see if the augmented models generalize better in another dataset. 2) Overfitting reduction: data augmentation usually helps to reduce overfitting by making the model more robust to variations of the training data. The small drop in the accuracy could indicate that the model is now less overfitted to the training data and better at generalizing. 3) Wrong data augmentation techniques: data augmentations, when implemented wrong, could introduce distortions that the models find difficult to learn from. This could be the case for our data augmentation techniques. We experimented with a couple of different techniques and all led to a decrease in prediction accuracy. This could just mean we haven't found the right data augmentation techniques for our data.

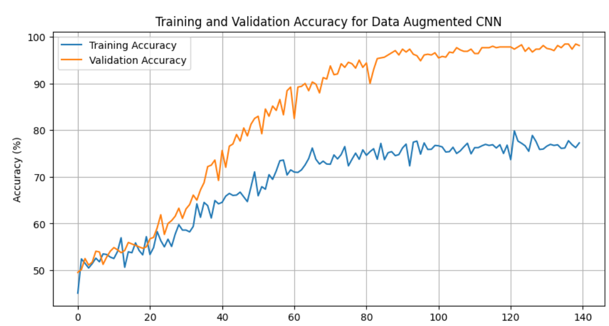

Figure 6. CNN with DA Training and validation accuracy

The confusion matrix of Vanilla CNN with data augmentation is shown in Figure 7. In comparison to the confusion matrix of Vanilla CNN without data augmentation, there

Model Type	Accuracy	F1
Vanilla CNN	98.28%	0.983
Vanilla CNN with DA	97.81%	0.978
ResNet	99.38%	0.994
ResNet with DA	98.12%	0.981
DenseNet	99.53%	0.995
DenseNet with DA	97.97%	0.98

Table 1. A Comparison of Model Performances

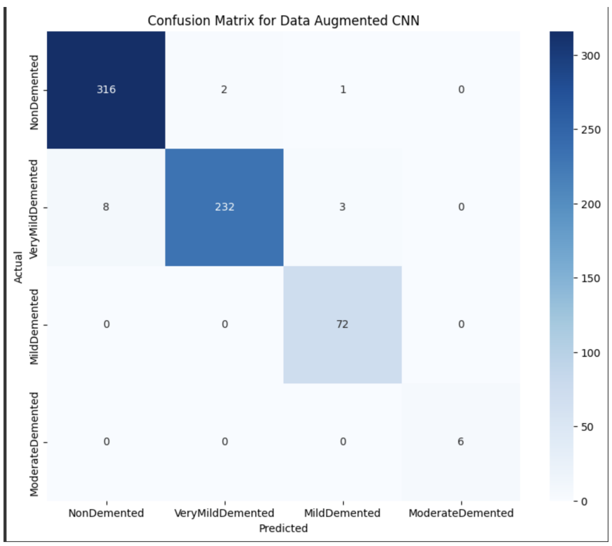

Figure 7. Vanilla CNN with DA Confusion Matrix

is a slight increase in errors for the VeryMildDemented and NonDemented classes after augmentation, while the MildDemented and ModerateDemented classes remain unaffected. There is a drop of 3 correct predictions overall after augmentation. This suggests that the data augmentation did not improve the model's ability to distinguish the subtle differences between the VeryMildDemented and NonDemented classes, which we implemented data augmentation to accomplish. See the previous paragraph for possible reasons.

5.2.3 ResNet

A graph of the training and validation accuracy during the training can be viewed below in Figure 8. It achieved an accuracy of 99.38% on our test set and an F1 score of 0.994. The confusion matrix is as shown in Figure 9. According to the confusion matrix, the model has a slight bias towards non-demented, which is likely due to the larger sample size of the non-demented train data. It also confuses one NonDemented example as a VeryMildDemented example, as well as one VeryMildDemented example as a MildDemented example. Since the overall accuracies are quite high, these are more likely outliers in the training data that lie very close to the decision boundaries of different classes.

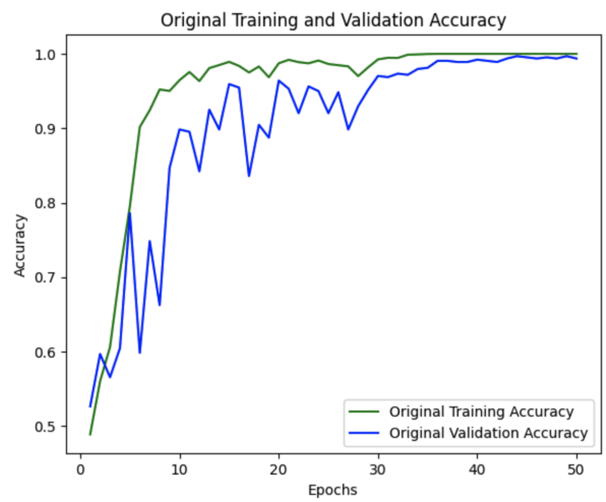

Figure 8. ResNet Training and validation accuracy

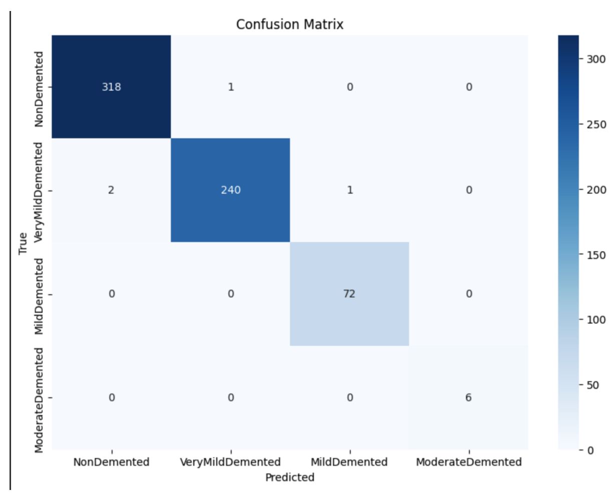

Figure 9. ResNet Confusion Matrix

5.2.4 ResNet with Data Augmentation

A graph of the training and validatoin accuracy during the training can be viewed below in Figure 10. It achieved an accuracy of 98.12% on our test set and an F1 score of 0.981. The confusion matrix is as shown in Figure 11.

In comparison to ResNet without data augmentation, the accuracy and F1 scores of ResNet with data augmentation decreased, which can be due to reasons outlined in the Vanilla CNN with Data Augmentation section. From Figure 10, we can see the steady increase and eventual plateau of the training accuracy, which suggests the model is effectively learning from the augmented data. On the other hand, the validation accuracy is generally higher than the training accuracy, which is somewhat unusual and can suggest that the model is learning to generalize beyond the training data. At the same time, there are high fluctuations in the validation accuracies, meaning the training has a high sensitivity to validation data variability. This could mean we need to further tune the augmentation parameters and add additional regulations.

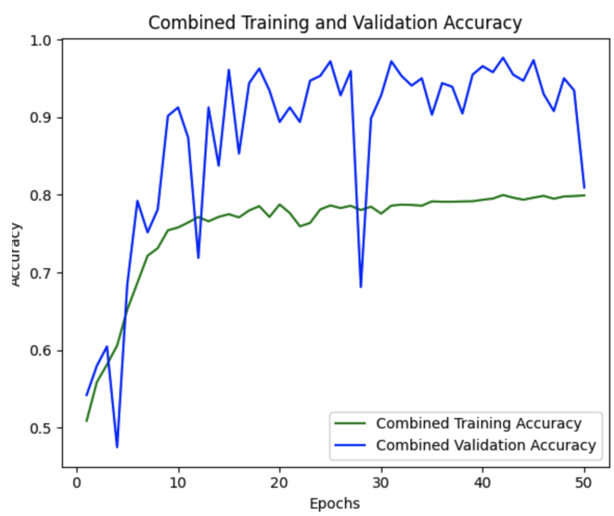

Figure 10. ResNet with DA Training and validation accuracy

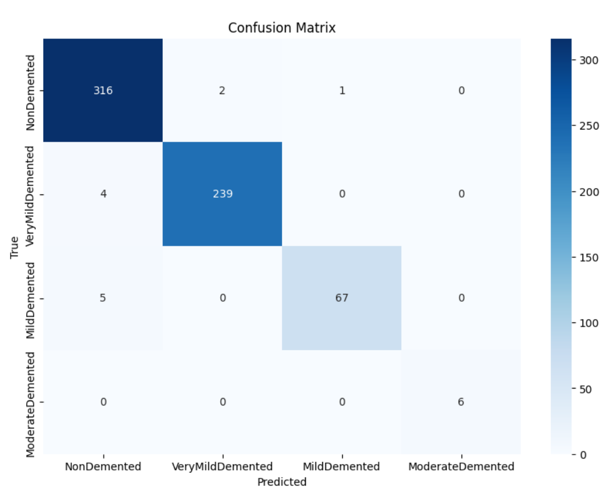

Figure 11. ResNet with DA Confusion Matrix

5.2.5 DenseNet

Based on the comparison of model performances chart, DenseNet without data augmentation reached the highest accuracy at 99.53% and the highest F1 score of 0.995. Therefore, DenseNet is our state-of-the-art model.

A graph of the training and validatoin accuracy during the training can be viewed below in Figure 12. The confusion matrix is as shown in Figure 13.

Other models tend to struggle to optimize training and validation accuracy but DenseNet reached an almost perfect train and validation accuracy. We believe a key aspect in DenseNet's extraordinary performance is the relative small size of our dataset. By connecting each layer to every other layer, DenseNet requires fewer parameters compared to traditional deep networks, which helps in preventing overfitting. This is crucial when working with relatively small datasets.

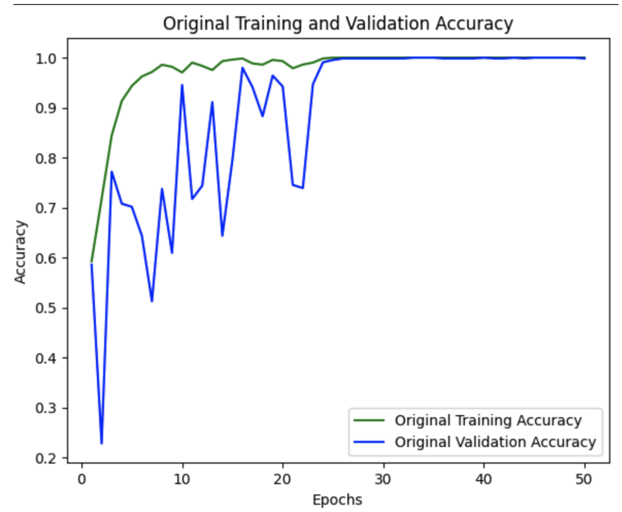

Figure 12. DenseNet Training and validation accuracy

5.2.6 DenseNet with Data Augmentation

A graph of the training and validatoin accuracy during the training can be viewed below in Figure 14. It achieved an accuracy of 97.97% on our test set and an F1 score of 0.98. The confusion matrix is as shown in Figure 15.

Similar to Resnet with data augmentation, the curve for training accuracy is quite smooth and converges to an optimum in the end, while the curve for validation accuracy has high fluctuations and generally stays above the training accuracy after the initial epochs. This can be explained by the additional variation in training data introduced by data augmentation. One interesting difference between the two is that ResNet's training accuracy converges to around 0.8, and DenseNet's training accuracy converges to beyond 0.9.

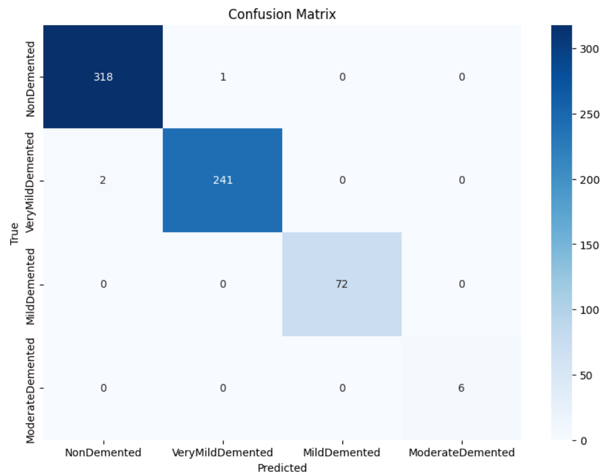

Figure 13. DenseNet Confusion Matrix

This suggests that ResNet is slightly underfitting the training data or it is well-regulated to avoid overfitting. In the context of data augmentation, this suggests that DenseNet is more effective in learning the variations introduced by data augmentation, which ultimately manifests in higher accuracy on the test set.

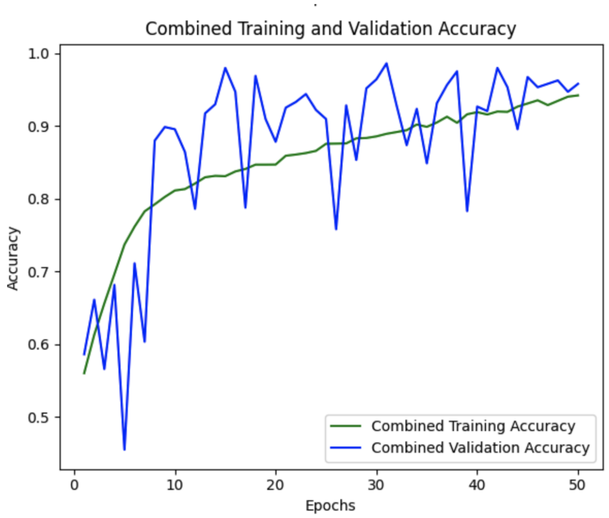

Figure 14. DenseNet with DA Training and validation accuracy

6. Conclusions

This paper compared the performance of various CNNs in detecting Alzheimer's disease through MRI images, additionally focusing on the impact of data augmentation on model accuracy. We evaluated three primary architectures: Vanilla CNN, ResNet-101, and DenseNet-121, both with and without data augmentation. Our findings indicate that DenseNet-121 without data augmentation achieved the highest performance with an accuracy of 99.53% and an F1 score of 0.995. This superior performance is likely due to DenseNet's dense connectivity pattern, which allows for efficient feature reuse and good gradient flow, making it particularly effective for small datasets such as the one used in our paper.

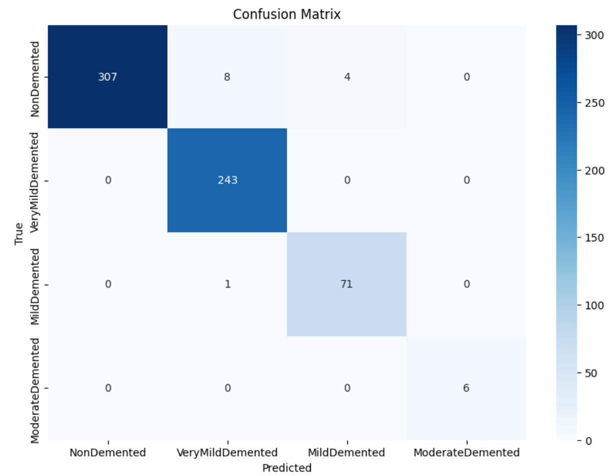

Figure 15. DenseNet with DA Confusion Matrix

While data augmentation generally aims to improve model robustness and accuracy by artificially increasing the diversity of the training data, it decreases the accuracy of all our models in predicting the test dataset. This decline is likely due to these models' inherent capacity to learn from the available data without additional manipulations, where the added variability introduced by augmentation could potentially confuse the training process.

For future work, several areas could be explored to further improve the performance of Alzheimer's detection using CNNs. Increasing the dataset size could provide more training data, especially for classes with fewer examples, potentially improving model accuracy. Incorporating multimodal data, such as combining MRI scans with genetic or demographic information, could also improve the models' performance and be an interesting interdisciplinary area of study. Furthermore, exploring more advanced architectures or hybrid models that combine deep learning with traditional machine learning techniques might also have the potential to improve performance in Alzheimer's detection.

7. Contributions & Acknowledgements

Jiahui Chen implemented the data augmentation, and training and evaluation for ResNet and DenseNet. Yvonne Hong conducted the literature review and decided on the algorithms we will implement for this project. She also drafted this project report and created the project presentation. Olivia Weiner implemented the training and evaluation for Vanilla CNN. She also helped prepare the project presentation.

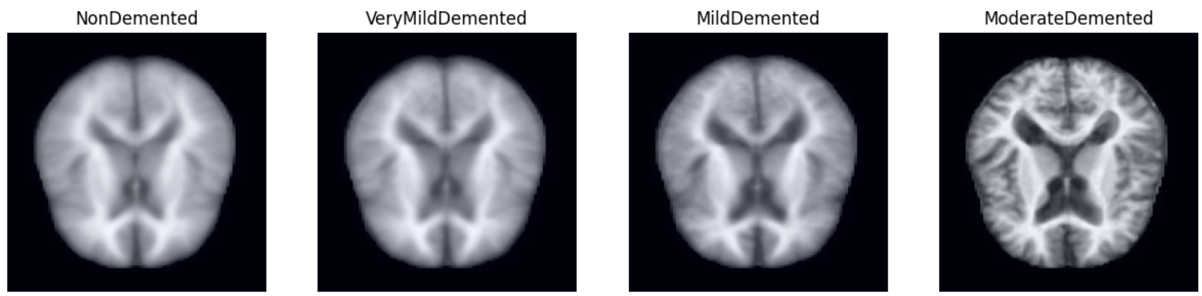

Figure 16. Average Class Images for DenseNet

References

[1] X.-a. Bi, X. Hu, H. Wu, and Y. Wang. Multimodal data analysis of alzheimer's disease based on clustering evolutionary random forest. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2973–2983, 2020.

[2] S. Dubey. Alzheimer's dataset (4 class of images).

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[4] S.-W. L. Heung-Il Suk and D. Shen. Latent feature representation with stacked auto-encoder for ad/mci diagnosis, 2015.

[5] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks, 2018.

[6] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[7] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth. Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research*, 57:147–159, 2019.

[8] A. Lebedev, E. Westman, G. Van Westen, M. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone, and A. Simmons. Random forest ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, 6:115–125, 2014.

[9] T. maintainers and contributors. Torchvision: Pytorch's computer vision library. `https://github.com/pytorch/vision`, 2016.

[10] T. pandas development team. pandas-dev/pandas: Pandas, Feb. 2020.

[11] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 2017.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[13] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, 22(1):173–183, 2018.

[14] S. E. Sorour, A. A. A. El-Mageed, K. M. Albarrak, A. K. Alnaim, A. A. Wafa, and E. El-Shafeiy. Classification of alzheimer's disease using mri data based on deep learning techniques. *Journal of King Saud University - Computer and Information Sciences*, 36(2):101940, 2024.

[15] H.-I. Suk, S.-W. Lee, and D. Shen. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.

[16] H.-I. Suk and D. Shen. Deep learning-based feature representation for ad/mci classification. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 583–590, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[17] N. Yamanakkanavar, J. Y. Choi, and B. Lee. Mri segmentation and classification of human brain using deep learning for diagnosis of alzheimer's disease: A survey. *Sensors (Basel)*, 20(11), 2020.

[18] Z. Yang and Z. Liu. The risk prediction of alzheimer's disease based on the deep learning model of brain 18f-fdg positron emission tomography. *Saudi Journal of Biological Sciences*, 27(2):659–665, 2020.