ALZHEIMER'S DETECTION WITH DEEP LEARNING

Muhammad Irfan
Student# 1009715612
m.irfan@mail.utoronto.ca

Hwang Wei Ju
Student# 1007874178
william.ju@mail.utoronto.ca

Hitansh Bhatt
Student# 1007931632
hitansh.bhatt@mail.utoronto.ca

Aryan Ghosh
Student# 1008838821
aryan.ghosh@mail.utoronto.ca

ABSTRACT

This proposal presents Group 27's APS360 Final Project, which focuses on developing a deep-learning model for the early detection of Alzheimer's disease using medical imaging. The project aims to classify different stages of Alzheimer's based on brain scan images, enabling early diagnosis and intervention. This document provides an overview of the project objectives, the proposed neural network architecture, relevant research, and additional key components as outlined in the Project Proposal Handout and Rubric. —-Total Pages: 9

1 INTRODUCTION

Alzheimer's disease is a progressive neurological disorder and the leading cause of dementia, affecting over 55 million people worldwide as of 2020, with numbers expected to double every 20 years. Early and accurate detection is crucial for timely interventions and improving patients' quality of life, but the diagnosis often relies on subjective clinical evaluations, which can result in delayed or missed diagnoses, especially in resource-limited settings.

This project aims to use deep learning techniques to automate Alzheimer's detection from MRI scans. By analyzing brain images, deep learning models can identify subtle biomarkers across different stages of dementia, providing a more objective and precise diagnostic tool. This approach would reduce reliance on manual evaluations and enhance diagnostic accuracy, making it particularly valuable as the global prevalence of Alzheimer's continues to rise.

Deep learning, particularly Convolutional Neural Networks (CNNs), is well-suited for this task due to its success in medical imaging. CNNs can automatically extract complex features from MRI scans, eliminating the need for manual feature engineering. This ability to classify various stages of Alzheimer's with high accuracy offers a scalable, non-invasive, and efficient diagnostic tool that could alleviate the burden on healthcare systems.

2 ILLUSTRATION/FIGURE

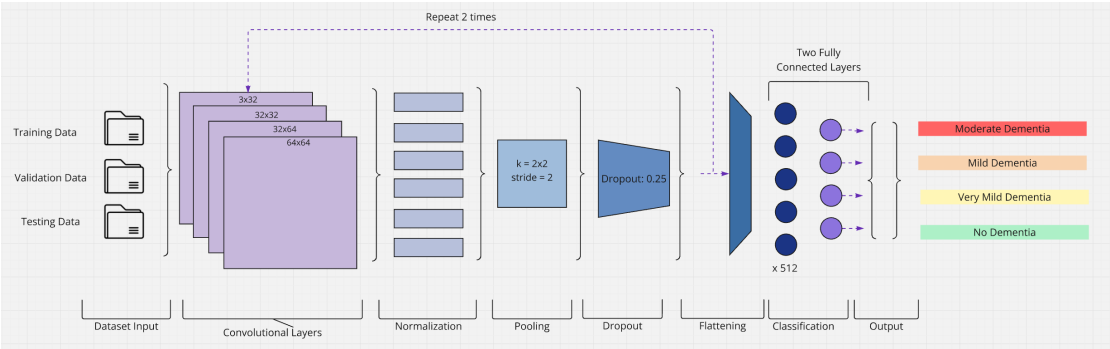

Figure 1: Illustration of Project Workflow and Architecture (Phung & Rhee, 2019).

3 Background & Related Work

Artificial intelligence, especially deep learning, is increasingly applied in biomedical imaging. Below are notable studies that have informed our approach.

3.1 Cardiac MRI Analysis

A CNN achieved 0.988 accuracy in diagnosing cardiovascular disease from cardiac MRIs, demonstrating AI's potential to match expert interpretation (Wang et al., 2024).

3.2 AI in Radiology

AI techniques like CNNs and autoencoders are assisting radiologists in identifying abnormalities across imaging modalities, including colonoscopy and pelvic scans (Hosny et al., 2018).

3.3 Lung Cancer Detection

ML algorithms applied to CT scans extracted 440 features, with a Naive Bayes classifier reaching an AUC of 0.72 in tumor detection (Wu et al., 2016).

3.4 Breast and Pulmonary Lesion Diagnosis

The OverFeast CNN was used on CT scans to detect cancerous lesions in breast and lung tissue, achieving an AUC of 0.086 (Cheng et al., 2016).

3.5 Chest Radiograph Interpretation

A deep learning model interpreted emergency chest radiographs with radiologist-level accuracy across a broad age range (Huang et al., 2023).

4 Data Processing

Proper data preparation is a crucial step in ensuring the success of a deep learning model. High-quality, well-preprocessed data can significantly enhance model performance, reduce biases, and prevent overfitting. This section outlines the essential steps taken to clean, preprocess, and balance the dataset to ensure robust and reliable learning. Effective data preparation not only improves accuracy but also ensures that the model generalizes well to unseen data. The following steps outline the data preparation process:

4.1 Dataset Source

The image dataset is sourced from the OASIS website, which provides a collection of Alzheimer's and other brain-related MRI images:

- **Original Source:** OASIS Brains Website (Open Access Series of Imaging Studies (OASIS), 2019)
- **Cleaned and Organized Dataset:** Kaggle: Alzheimer's MRI Images Dataset (Marcus et al., 2007)

For convenience, the Kaggle dataset will be used in this project as it is pre-organized into folders based on labels.

4.2 Data Preparation

1. **Data Management:** After importing, we need to manage the dataset and get the dataset file location into the array. In this case, it would help us to train the model by using this array to access each file in the folder.

(a) Creates a 2D list (array) with 4 empty lists (4 because we have 4 categories) which will store file paths.
(b) Loops through all folders inside a given `folder_path`. Walks through each folder (including subfolders) to find all files.
(c) Stores the full path of each file in one of the 4 lists inside the array.
(d) Increases `count` to move to the next list (but incorrectly, so it may cause an error).

2. **Data Overview:** Check the number of dataset for each category to get the information about the distribution and if either it's balanced or not.

	Folder	Count	Percentage
0	NonDemented	9600	28.25
1	MildDemented	8960	26.37
2	VeryMildDemented	8960	26.37
3	ModerateDemented	6464	19.02

Figure 2: Dataset distribution for each category

3. **Data Augmentation:** To address class imbalance in the Alzheimer's dataset, data augmentation is essential for helping the model learn from all categories more evenly. Although augmentation was already applied by the dataset provider, we highlight some typical techniques suitable for this type of medical imaging:
   (a) **Horizontal Flipping:** Flips the image with a 50% probability (p = 0.5).
   (b) **Brightness & Contrast Adjustment:** Random changes with 20% probability (p = 0.2).
   (c) **Image Rotation:** Randomly rotates within ±20° with a 50% probability.
   (d) **Normalization:** Scales pixel values using ImageNet mean and standard deviation.

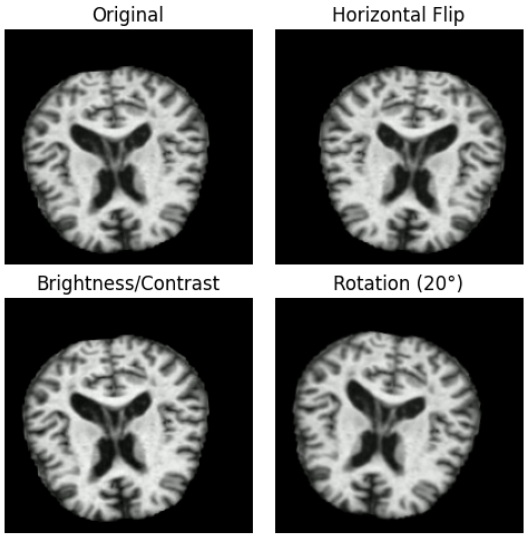

Figure 3: Augmentation of Alzheimer's dataset

4. **One-Hot Encoding:** The dataset is divided into four categories: *Non-Demented*, *Mild Dementia*, *Moderate Dementia*, and *Very Mild Dementia*. These string labels will be converted into integer labels using one-hot encoding for compatibility with the deep learning model.
   - **Categories:** {Non-Demented, Mild Dementia, Moderate Dementia, Very Mild Dementia}
   - **Integer Encoding:** {0, 1, 2, 3}

5. **Data Splitting:** The dataset will be split into training, validation, and testing sets. Specifically, 70% of the data will be used for training, 15% for validation, and 15% for testing. This ensures that the model is trained effectively while also having separate datasets for tuning hyperparameters and evaluating final performance.

5 ARCHITECTURE

For Alzheimer's detection using medical imaging, we initially developed a custom convolutional neural network (CNN) named **AlzheimerNet**. However, after further experimentation, we adopted **ResNet152** with transfer learning, which provided superior performance, achieving an accuracy of **90.83%**.

5.1 CUSTOM CNN: ALZHEIMERNET

AlzheimerNet was designed as a lightweight CNN for efficient feature extraction and classification. The architecture consisted of:

- **Two convolutional blocks**, each containing two convolutional layers (kernel size = 2), batch normalization, max-pooling (kernel size = 2), and dropout (0.25 for convolutional layers, 0.5 for fully connected layers).
- **Fully connected layers**: A dense layer with 512 neurons and ReLU activation, followed by dropout (0.5) and a final output layer with softmax activation for classification into four categories.

While AlzheimerNet was computationally efficient, its performance was limited by the relatively shallow depth of the network, leading us to explore deeper architectures.

5.2 TRANSFER LEARNING WITH RESNET152

Given the complexity of MRI scan patterns, we transitioned to using **ResNet152**, a deep residual network pre-trained on ImageNet. This allowed us to leverage well-established hierarchical feature representations. Our approach included:

- **Freezing initial layers**: Retaining the lower convolutional layers to preserve fundamental feature extraction capabilities.
- **Unfreezing deeper layers**: Specifically, layers 2, 3, and 4 were fine-tuned to learn Alzheimer's-specific patterns.
- **Custom classifier head**: The original fully connected layer was replaced with a multi-layer structure:
  - Linear (2048 neurons) with ReLU activation
  - Dropout (0.5) for regularization
  - Linear (4096 neurons) with ReLU activation
  - Dropout (0.5)
  - Linear (1024 neurons) with ReLU activation
  - Final softmax layer for classification into four categories

5.3 FEATURE EXTRACTION AND FINE-TUNING

The lower convolutional layers of ResNet152 extracted fundamental image features such as edges and textures. Fine-tuning was applied to layers 2, 3, and 4 to allow specialization in detecting patterns relevant to Alzheimer's diagnosis.

5.4 TRAINING STRATEGY AND HYPERPARAMETERS

The model was trained with the following hyperparameters, optimized for performance:

- **Learning rate**: 5.199×10^{-5}
- **Batch size**: 16
- **Epochs**: 50

This configuration yielded our best-performing model, demonstrating that ResNet152's depth and pre-trained knowledge significantly improved classification accuracy over our custom CNN approach.

6 BASELINE MODEL

The baseline model chosen for comparison is a Random Forest classifier due to its simplicity, interpretability, and widespread use as a reference in classification tasks. It provides a minimum performance benchmark, allowing us to assess whether more complex models, such as neural networks, are necessary. In this case, we use Random Forest because it offers a balance between accuracy and efficiency while being less prone to overfitting compared to more complex models.

6.0.1 IMPLEMENTATION DETAILS

1. **Input Format**
   - Images are resized to $128 \times 128 \times 3$ for consistency.
   - Flattened into one-dimensional vectors.
2. **Preprocessing**
   - Pixel values are normalized to $[0, 1]$ for better convergence.
   - No additional feature extraction is applied.
3. **Model Architecture and Regularization**
   - Random forest with softmax activation for multi-class classification.
   - L_2 regularization ($\lambda = 1.0$) to prevent overfitting.
4. **Training and Evaluation**
   - Trained using cross-entropy loss and optimized via LBFGS solver (100 iterations).
   - Performance evaluated using a confusion matrix and standard metrics: accuracy, precision, recall, and F1-score.

6.1 RESULTS:

The baseline model demonstrates strong overall performance, achieving balanced scores for accuracy (74.25%), recall (74.25%), and precision (73.95%). However, since the primary objective of this project is to detect Alzheimer's disease, recall is the most critical metric to minimize false negatives.

7 QUANTITATIVE RESULTS

Our model demonstrates high accuracy in Alzheimer's stage classification, with a configuration of learning rate: 7.74×10^{-5}, batch size: 32, and epochs: 10. These results maintain strong clinical relevance, particularly in minimizing false negatives—a crucial factor in medical diagnosis.

7.1 TRAINING DYNAMICS

The model showed fast and stable convergence within just 10 epochs:

- **Training Error**: 20.75% → **0.90%**
- **Validation Error**: 8.83% → **3.50%** (60.4% relative reduction)
- **Training Loss**: 0.7134 → **0.3730** (47.7% reduction)
- **Validation Loss**: 0.5098 → **0.4399** (13.7% reduction)

7.2 Final Model Performance

Table 1: Optimized Model Metrics

Metric	Value	Clinical Interpretation
Accuracy	**96.50%**	High correct classification rate
Precision	**96.71%**	Very few false positives
Recall	**96.50%**	Critical: Only 3.5% of cases missed
F1-Score	**96.48%**	Strong balance between precision and recall

7.3 Class-Wise Performance

Table 2: Stage-Specific Recall

Stage	Recall	Key Challenge
Moderate Demented	100.0%	Limited but well-separated class
Non-Demented	88.7%	Misclassified as early dementia
Very Mild Demented	98.6%	Borderline with Mild Dementia
Mild Demented	98.7%	Some confusion with Very Mild

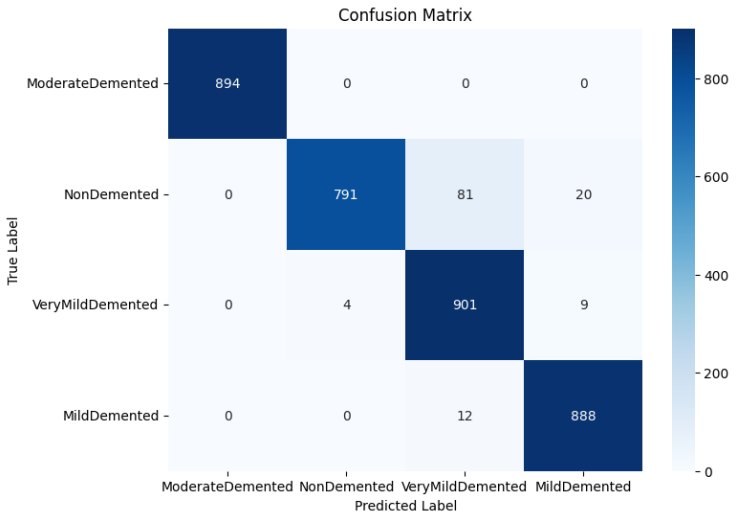

Figure 4: Final Model's Confusion Matrix Result

7.4 Comparative Analysis

- **Accuracy:** Improved from 74.25% (Random Forest) to 96.5%, a gain of **22.25%**.
- **Recall:** Increased from 74.25% to 96.50%, a **30.0% relative improvement** in minimizing false negatives.
- **Precision:** Improved from 73.95% to 96.71%, a **22.76% absolute gain**.
- **F1-Score:** Enhanced from approximately 74.1% (baseline average) to 96.48%, indicating a significant gain in overall balance and robustness.

- These results demonstrate the clear advantage of deep learning approaches in capturing subtle structural differences in brain scans, significantly outperforming traditional methods like Random Forests.

8 Qualitative Results

To supplement our quantitative analysis, we also examined the model's predictions on real MRI scans across Alzheimer's stages. Figure 5 displays representative examples for each class.

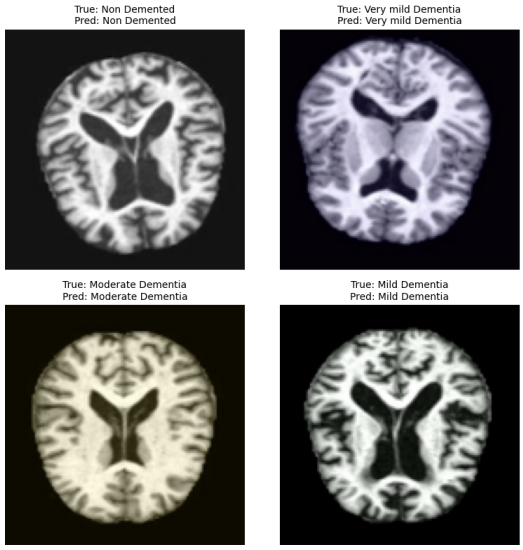

Figure 5: Model predictions on sample MRI scans from each class.

Key observations:

- **Clear Classifications:** The model performs extremely well on Moderate Dementia cases, showing near-perfect recall. Structural differences in these cases are more pronounced, leading to fewer misclassifications.
- **Non-Demented Challenges:** Interestingly, Non-Demented cases were the most frequently misclassified, typically as early-stage dementia. This suggests the model sometimes mistakes healthy brain scans for very subtle signs of degeneration—possibly due to overlapping features or natural brain variation.
- **Borderline Stages:** While confusion still occurs between Very Mild and Mild Dementia, the model distinguishes them reasonably well. However, these stages remain difficult due to their visual similarity, even for human experts.
- **Error Patterns:** Most misclassifications involve either Non-Demented being predicted as Very Mild, or minor confusion between Very Mild and Mild Dementia. These patterns reflect the inherent difficulty in distinguishing adjacent stages of cognitive decline.

9 Evaluate Model on New Data

[H]

10 Discussion

In this project, I explored various models to classify brain MRI images for deementia detection. Among them, my best-performing model was the ResNet512, which significantly outperformed the baseline model.

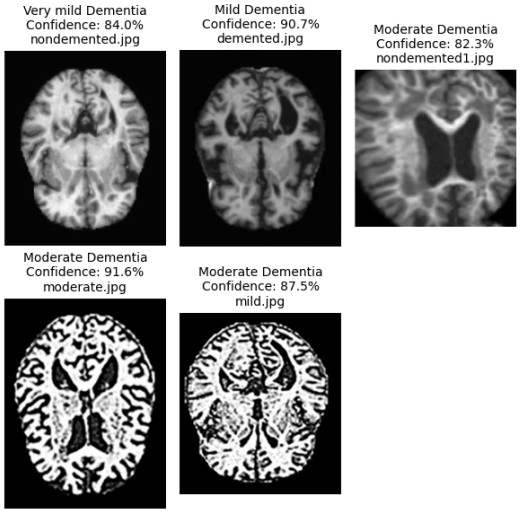

Figure 6: Never seen data

- **Best Transfer Learning Model:** Our transfer learning model (ResNet152) achieved the highest recall (96.5 %) and generalization performance. Its ability to extract complex, hierarchical features helped it better distinguish between subtle differences across the classes. Fine-tuning the deeper layers and using data augmentation also contributed to reducing overfitting and improving validation accuracy.
- **Comparison with Baseline:** Our initial baseline has accuracy and recall 74.25 %. It means our model shows great improvement with the increase about 21.25 %.
- **Interesting or Surprising Observations:** One surprising observation was that the model sometimes misclassified images in the Very Mild Dementia class, likely due to the visual similarity to the Moderate Dementia class. This highlights the difficulty of the task and suggests that either more training data or domain-specific preprocessing might be helpful.
- **Model Behavior and Learning:** I noticed that even small changes in learning rate or optimizer made a noticeable impact, which underscores the sensitivity of transfer learning. Also, using pre-trained models helped drastically cut down training time and improved convergence speed.
- **What I Learned:** This project taught me that model architecture and data quality both play critical roles in performance. It also reinforced how powerful transfer learning can be, especially in medical or low-data domains. I gained a deeper understanding of how fine-tuning and hyperparameter tuning affect deep learning pipelines.

11 Ethical Considerations

The use of deep learning for Alzheimer's detection presents several ethical challenges. A major concern is dataset bias—publicly available data may lack diversity in age, gender, ethnicity, or socio-economic status, leading to uneven diagnostic performance and reinforcing health disparities.

Privacy is another key issue, as MRI scans contain sensitive information. Strict data protection measures, such as compliance with PIPEDA (Office of the Privacy Commissioner of Canada, 2021), are essential. In this project, we use anonymized data to reduce privacy risks, but broader deployment requires robust governance.

Model explainability also matters. Deep learning models are often "black boxes," which can undermine trust and accountability in clinical settings. These tools should support—not replace—medical professionals, with clear mechanisms for interpreting predictions.

Finally, the impact of false positives and false negatives must be carefully managed, as both carry serious consequences. Rigorous validation and transparency about model limitations are crucial for ethical deployment.

12 Link To Google Colab

https://colab.research.google.com/drive/1dK2eVPZ8EtbMdCNKKIENec7bjbIFtX2t?usp=sharing

REFERENCES

J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen. Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans, 2016. URL https://doi.org/10.1038/srep24454. Accessed: 2025-02-01.

A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and Hugo. Artificial intelligence in radiology, 2018. URL https://doi.org/10.1038/s41568-018-0016-5. Accessed: 2025-02-01.

J. Huang, L. Neill, M. Wittbrodt, D. Melnick, M. Klug, M. Thompson, J. Bailitz, T. Loftus, S. Malik, A. Phull, V. Weston, J. A. Heller, and M. Etemadi. Generative artificial intelligence for chest radiograph interpretation in the emergency department, 2023. URL https://doi.org/10.1001/jamanetworkopen.2023.36100. Accessed: 2025-02-01.

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. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007. doi: 10.1162/jocn.2007.19.9.1498.

Office of the Privacy Commissioner of Canada. The personal information protection and electronic documents act (pipeda), 2021. URL https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/. Accessed: 2025-02-01.

Open Access Series of Imaging Studies (OASIS). Open access series of imaging studies (oasis), 2019. URL https://doi.org/10.1162/jocn.2007.19.9.1498. Accessed: 2025-02-01.

Phung and Rhee. A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. pp. 3, 2019.

Y.-R. J. Wang, K. Yang, Y. Wen, P. Wang, Y. Hu, Y. Lai, Y. Wang, K. Zhao, S. Tang, A. Zhang, H. Zhan, M. Lu, X. Chen, S. Yang, Z. Dong, Y. Wang, H. Liu, L. Zhao, L. Huang, and Y. Li. Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging, 2024. URL https://doi.org/10.1038/s41591-024-02971-2. Accessed: 2025-02-01.

W. Wu, C. Parmar, P. Grossmann, J. Quackenbush, Philippe Lambin, Johan Bussink, R. Mak, and Hugo. Exploratory study to identify radiomics classifiers for lung cancer histology, 2016. URL https://doi.org/10.3389/fonc.2016.00071. Accessed: 2025-02-01.