Republic of Tunisia
The Ministry of Higher Education
and Scientific Research

The University of El Manar

The Higher Institute of Computer
Science Ariana

PROJECT REPORT

AlzheimerNet: AI-driven Early Detection of Alzheimer's Disease in MRI Scans

By

Fourat Kandil

Joumen Ben Said

Nermine Dhifala

Nadine Jannen

Supervised by:
Prof. Sahbi Bahrouni

Academic Year: 2024-2025

Contents

List of Figures

Chapter 1

Preliminary Study

1.1 Project Context

Alzheimer's disease is a neurodegenerative disorder that progressively impairs cognitive function, memory, and behavior. It is the most common cause of dementia, posing a significant burden on healthcare systems worldwide. Despite the availability of clinical diagnostic criteria, early detection of Alzheimer's disease remains a considerable challenge due to subtle initial symptoms and overlapping signs with normal aging.

Magnetic resonance imaging (MRI) is a crucial tool in understanding brain structures and identifying abnormalities. Advances in artificial intelligence, particularly in deep learning, provide new opportunities to analyze MRI data effectively. Convolutional neural networks (CNNs) have shown great promise in image classification tasks, including medical imaging, by automating the identification of disease patterns and reducing reliance on manual interpretations. This project leverages CNNs for classifying brain MRIs to enhance early diagnosis of Alzheimer's disease, aiming to address a critical gap in the medical field.

1.2 Problem Statement and Objectives

Alzheimer's disease affects millions of individuals globally, yet its diagnosis is often delayed until significant cognitive decline has occurred. The early stages of the disease, known as mild cognitive impairment (MCI), are particularly difficult to detect as they present symptoms that are often mistaken for normal aging. This delay in diagnosis restricts timely interventions that could improve patient outcomes and slow disease progression.

Traditional diagnostic methods for Alzheimer's, such as cognitive tests and manual MRI analyses, require expert intervention, are time-intensive, and may lack consistency. With the increasing availability of large-scale MRI datasets, there is a pressing need for automated systems that can analyze and classify these images accurately. However, several challenges must be addressed:

Differentiating between normal aging patterns and early-stage Alzheimer's features in MRIs. Detecting advanced stages of Alzheimer's, which may present diverse structural abnormalities. Managing imbalanced datasets, where images of normal MRIs often outnumber those with Alzheimer's-related features. Ensuring model interpretability so healthcare professionals can trust and understand predictions. This project aims to address these challenges by designing a robust machine learning model that leverages CNNs to classify brain MRI images into three categories: normal, early-stage Alzheimer's, and advanced Alzheimer's. Such a system has the potential to augment clinical workflows, improve diagnostic efficiency, and ultimately enhance patient care.[?]

1.3 Evaluation of Existing Technologies and Their Limitations

The application of deep learning, particularly convolutional neural networks (CNNs), in medical imaging has significantly advanced the early detection and classification of Alzheimer's disease (AD) using brain magnetic resonance imaging (MRI). This review examines key studies and methodologies that have contributed to this field.[AAard]

1.3.1 Convolutional Neural Networks in Medical Imaging

CNNs have revolutionized image analysis by automatically learning hierarchical feature representations, making them well-suited for medical image interpretation. Their ability to capture spatial hierarchies in data has led to superior performance in various medical imaging tasks, including disease classification and anatomical segmentation. [AAarb]

1.3.2 Deep Learning for Alzheimer's Disease Detection

Recent studies have demonstrated the efficacy of deep learning models in detecting AD from MRI scans. For instance, a deep learning approach outperformed traditional biomarkers in classifying dementia stages of Alzheimer's disease, highlighting the potential of CNNs in enhancing diagnostic accuracy. [AAard]

Another comprehensive review emphasized the role of deep learning in AD prediction, discussing various architectures and their performance metrics. The study underscored the importance of automated MRI-based models in detecting Alzheimer's disease progression. [AAarc]

1.3.3 Transfer Learning and Pre-trained Models

Transfer learning, which involves fine-tuning pre-trained CNNs on specific medical datasets, has been explored to address challenges posed by limited labeled medical data. A literature review on transfer learning for medical image classification highlighted its effectiveness in improving performance by leveraging knowledge from related tasks. [AAare]

1.3.4 3D Convolutional Neural Networks

The utilization of 3D CNNs has been investigated to capture volumetric information from MRI scans, providing a more comprehensive analysis of brain structures. A review on 3D CNNs in medical image segmentation discussed their potential in enhancing the speed and accuracy of processing and analyzing medical images. [AAara] MDPI

1.3.5 Challenges and Future Directions

Despite the advancements, challenges such as data scarcity, class imbalance, and the need for model interpretability persist. A review on CNNs in medical image analysis discussed these shortcomings and potential solutions, emphasizing the importance of addressing these issues to improve clinical applicability. SPRINGERLINK

In summary, the integration of CNNs in analyzing brain MRI scans has shown promising results in the early detection and classification of Alzheimer's disease. Ongoing research continues to refine these models, aiming to enhance their accuracy, interpretability, and clinical utility.

1.4 Methodology of Work

In our project, we adopt two complementary frameworks—CRISP-DM (Cross-Industry Standard Process for Data Mining) and Kanban—to structure both the technical workflow and project management. CRISP-DM guides our approach to data science tasks, while Kanban helps us manage and track progress in an organized, flexible manner. Together, these methodologies provide a systematic approach to ensure the successful completion of the project.

1.4.1 CRISP-DM Framework

CRISP-DM is a widely used, iterative process that provides a structured approach to data mining and machine learning tasks. It consists of six phases, each crucial for ensuring the effectiveness of our project.

At the Business Understanding stage, we focus on defining the goals of the project. Our primary objective is to create a convolutional neural network (CNN) that can accurately classify MRI images into three categories: normal, early-stage Alzheimer's, and advanced Alzheimer's. This model will not only help in detecting Alzheimer's disease but also assist in identifying early signs that might go unnoticed by clinicians.

Next, in the Data Understanding phase, we explore and familiarize ourselves with the dataset. We download MRI images from public dataset adni-extracted-axial and inspect their characteristics, such as image quality, distribution, and potential biases. This phase helps us identify the challenges we may face with imbalanced classes or noisy data and allows us to plan the necessary preprocessing steps.

During the Data Preparation phase, we preprocess the data to make it suitable for model training. We resize images to a consistent size, normalize pixel values to standardize the intensity range, and apply data augmentation techniques to artificially increase the dataset's size and variability. This step is essential to improve model robustness and generalization.

The Modeling phase involves designing and training our CNN models. We explore different architectures, such as ResNet and U-Net, to determine the best-performing model for Alzheimer's classification. To improve accuracy and reduce overfitting, we experiment with transfer learning (fine-tuning pre-trained models) and apply regularization techniques like dropout and batch normalization.

Once the models are trained, we enter the Evaluation phase, where we assess their performance using metrics like accuracy, precision, recall, and AUC-ROC. Additionally, we use Grad-CAM or similar techniques to visualize which areas of the MRI images the model is focusing on, which helps us interpret the results and validate that the model is making clinically relevant decisions.

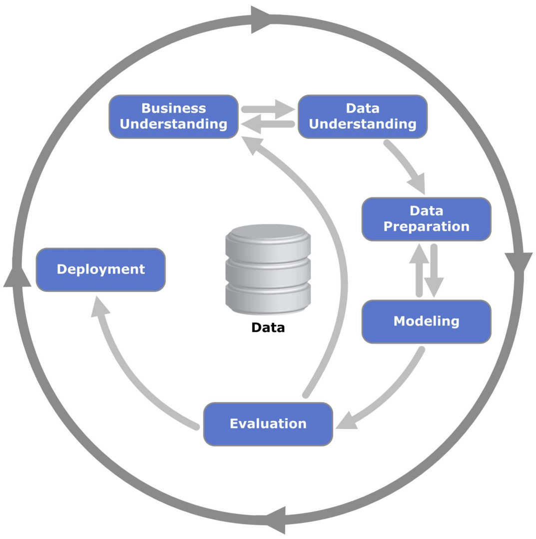

Figure 1.1: CRISP-DM Process Diagram

1.4.2 Kanban Methodology

For managing the tasks and progress of our project, we use the Kanban methodology, which helps us maintain clarity and organization throughout the project lifecycle. Kanban is a visual project management tool that organizes tasks into categories such as "To Do," "In Progress," and "Done." This allows us to focus on one task at a time and avoid overloading ourselves with multiple simultaneous tasks.

In the To Do column, we list all the tasks that need to be completed, such as downloading datasets, researching suitable CNN architectures, or implementing data augmentation strategies. These tasks serve as the foundation for the project, ensuring we have everything in place before we begin training our models.

Once tasks are ready to be worked on, they are moved into the In Progress column. Here, we focus on executing specific tasks like preparing the data, building the CNN models, or testing different architectures. This column helps us stay organized and ensures that each task is actively worked on.

Finally, tasks that are completed are moved into the Done column. For example, once data preprocessing is finished, or once model evaluation metrics are calculated, these tasks are marked as complete. The Kanban board provides a clear visual representation of the project's current state, making it easier to track progress and identify any potential bottlenecks.

Kanban promotes flexibility and continuous progress, allowing us to adapt to changes as new insights or challenges arise during the project. By regularly updating our board and limiting the number of tasks in progress at any given time, we ensure that our work is manageable, efficient, and aligned with our project goals.

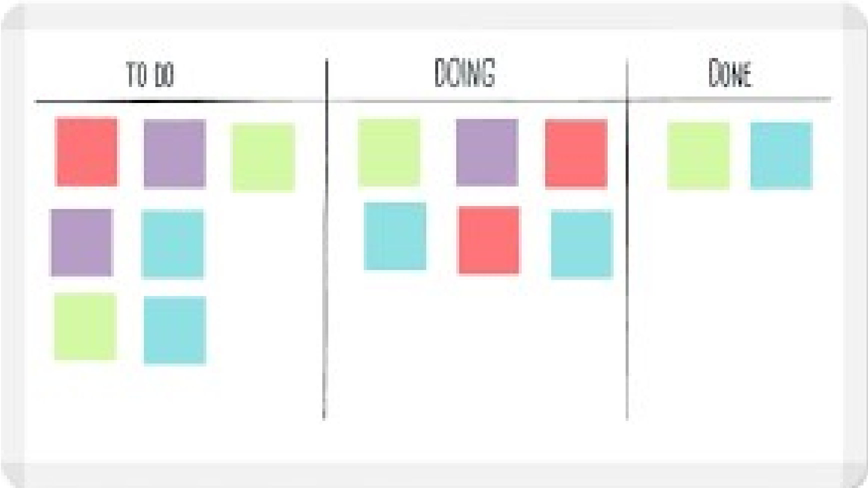

Figure 1.2: Kandban Diagram

Chapter 2

Classification of Alzheimer's Cases Using CNNs

Classification of Alzheimer's Cases Using CNNs

1 Introduction

In this chapter, we will provide a detailed overview of our work. We developed two distinct projects: one utilizing four classes and the other using three classes. Our goal was to analyze the impact of the data structure on the results. Additionally, we tested pretrained models and thoroughly evaluated their performance

2 First Experiment: Classification with four Classes

The dataset used consists of images categorized into four classes: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. The dataset contains 5154 images divided into the following classes:

- **Mild Demented:** 896 images
- **Moderate Demented:** 64 images
- **Non-Demented:** 3200 images
- **Very Mild Demented:** 2240 images

Images were resized to 128x128 pixels for uniformity.

2.0.1 Data Preparation

2.0.2 Preprocessing

The images were normalized by scaling pixel values to the range [0, 1]. Labels were encoded using a one-hot encoding scheme. The dataset was split into training (80%) and testing (20%) subsets.

2.0.3 Data Augmentation

To enhance model generalization, data augmentation techniques were applied, including:

- Rotation (up to 40 degrees)
- Width and height shifting (up to 30%)
- Shearing (up to 30%)
- Zooming (up to 30%)
- Horizontal flipping

2.1 CNN Architecture

The architecture of the custom CNN model is designed to extract meaningful features from the MRI images and perform classification efficiently. Below are the details of the architecture:

1. **Input Layer**: Accepts images of size 128x128x3 (RGB channels).
2. **Convolutional Layer 1**:
    - Filters: 32
    - Kernel size: 3x3
    - Activation: ReLU
3. **MaxPooling Layer 1**:
    - Pool size: 2x2
4. **Convolutional Layer 2**:

- Filters: 64
- Kernel size: 3x3
- Activation: ReLU

5. **MaxPooling Layer 2**:
   - Pool size: 2x2
6. **Convolutional Layer 3**:
   - Filters: 128
   - Kernel size: 3x3
   - Activation: ReLU
7. **MaxPooling Layer 3**:
   - Pool size: 2x2
8. **Convolutional Layer 4**:
   - Filters: 256
   - Kernel size: 3x3
   - Activation: ReLU
9. **MaxPooling Layer 4**:
   - Pool size: 2x2
10. **Flatten Layer**: Converts the 2D feature maps into a 1D feature vector.
11. **Fully Connected Layer 1**:
    - Neurons: 512
    - Activation: ReLU
12. **Dropout Layer**: Dropout rate of 0.5 to reduce overfitting.
13. **Output Layer**:
    - Neurons: 4 (corresponding to the classes)
    - Activation: Softmax

A transfer learning approach was also implemented using a pre-trained VGG16 model with custom dense layers.

2.2 Training

The model was compiled using the Adam optimizer with a learning rate of 0.0001. The categorical cross-entropy loss function was used. Early stopping was employed to prevent overfitting.

3 Evaluation Metrics

Performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC curves.

3.1 Results

3.1.1 Classification Metrics

	precision	recall	f1-score	support
MildDemented	0.92	0.74	0.82	128
ModerateDemented	0.00	0.00	0.00	11
NonDemented	0.84	0.98	0.91	535
VeryMildDemented	0.87	0.74	0.80	351
accuracy			0.86	1025
macro avg	0.66	0.61	0.63	1025
weighted avg	0.85	0.86	0.85	1025

3.1.2 Confusion Matrix

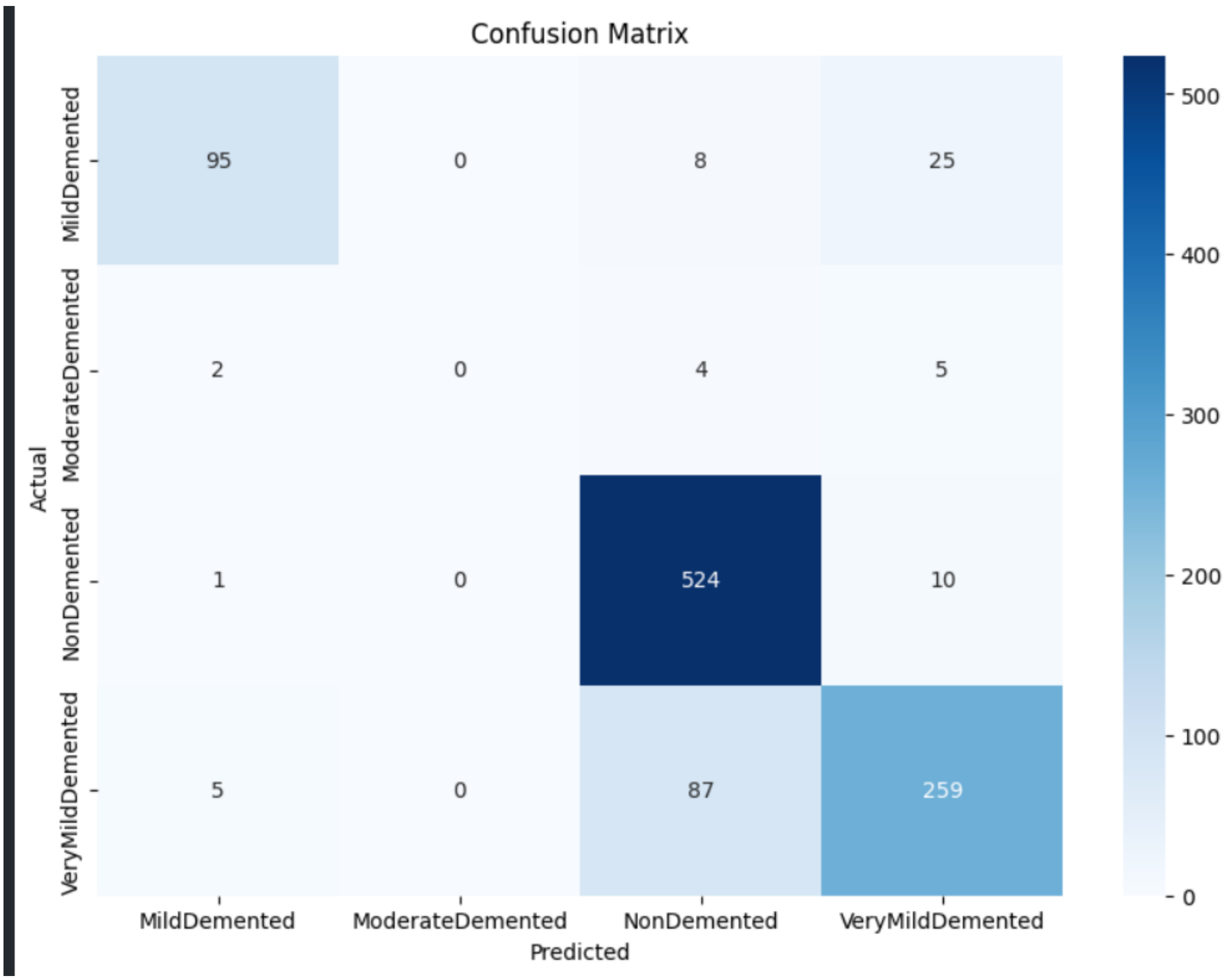

Figure 1: Confusion matrix for test set predictions.

3.1.3 ROC Curves

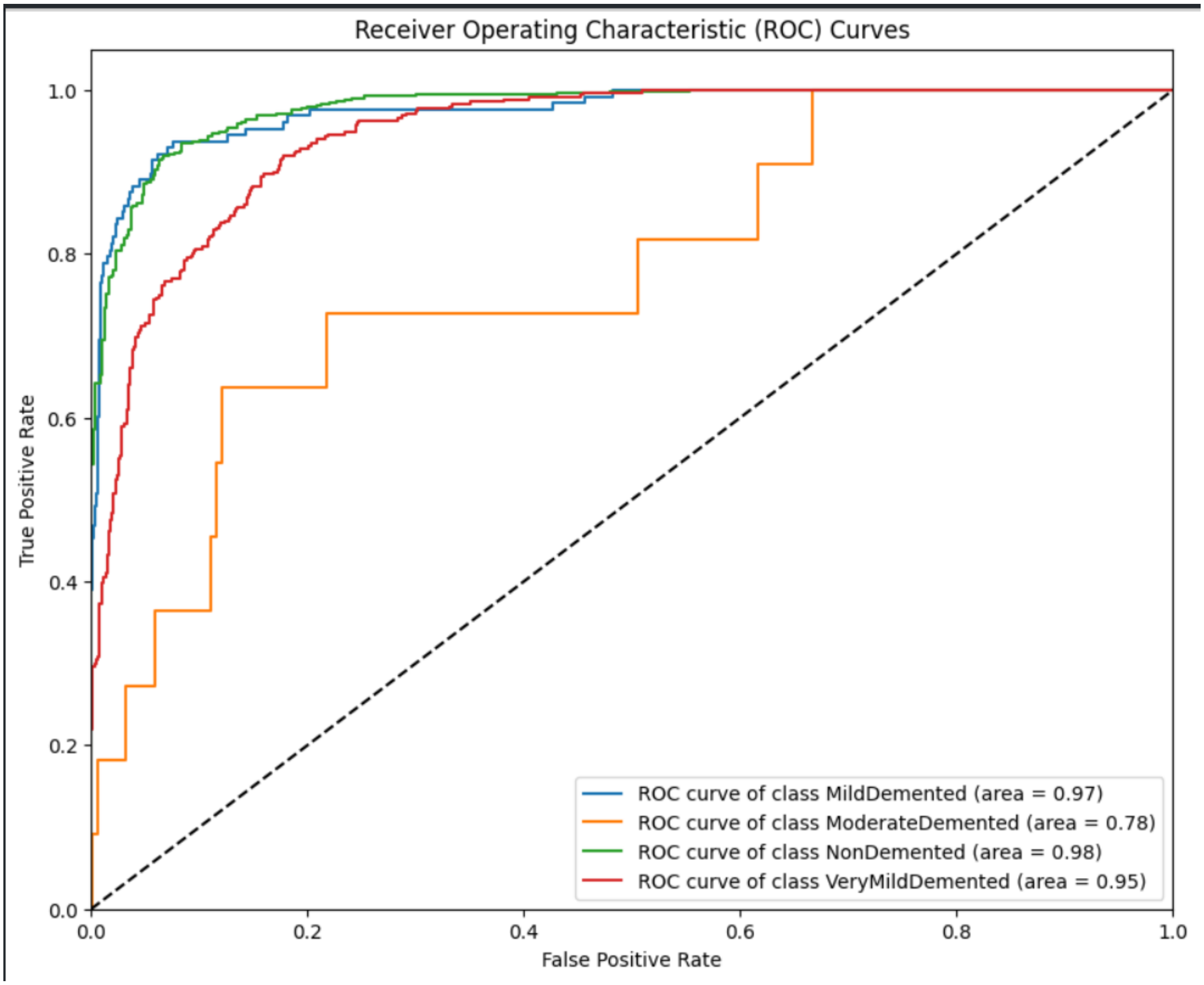

Figure 2: ROC curves for each class.

3.2 Conclusion

The CNN model achieved an overall accuracy of 86% on the test set. While the model performed well for the Mild, Non-Demented, and Very Mild classes, it struggled with the Moderate Demented class due to the limited number of samples. Future work could focus on obtaining a more balanced dataset and exploring other architectures.

4 Second Experiment: Classification with Three Classes

4.1 Introduction

In this second experiment, we used a dataset containing three classes: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Cognitively Normal (CN). The goal was to examine the impact of reducing the number of classes on model performance and to compare these results with those obtained in the first project.

4.2 Model Architecture

The model used for this experiment is a sequential architecture consisting of:

- Two 2D convolutional layers with 200 and 100 filters, respectively, and ReLU activation.
- Two max-pooling layers with a size of 3×3.
- A flattening layer to transform outputs into 1D vectors.
- Two fully connected layers with 100 and 50 units activated by ReLU, respectively.
- An output layer with three units and softmax activation for multi-class classification.

The model was compiled with the Adam optimizer, the `SparseCategoricalCrossentropy` loss function, and accuracy as a metric.

4.3 Results

Model performance was evaluated on validation and test sets. The results are as follows:

4.3.1 Learning Curve

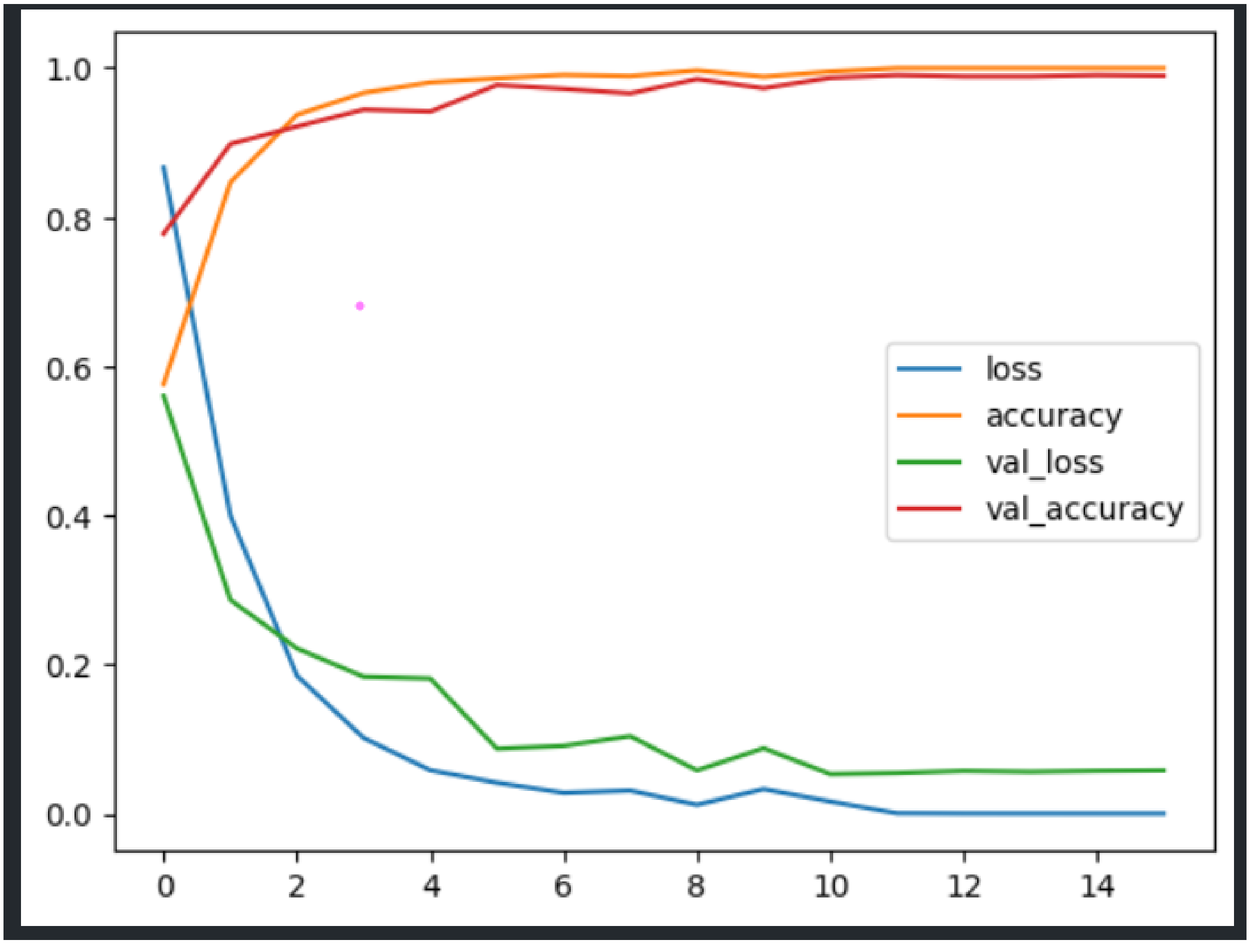

Figure 3: Learning Curve.

4.3.2 Validation Set

	precision	recall	f1-score	support
0	0.99	1.00	0.99	383
1	0.99	0.99	0.99	383
2	0.99	0.98	0.99	383
accuracy			0.99	1149

macro avg	0.99	0.99	0.99	1149
weighted avg	0.99	0.99	0.99	1149

4.3.3 Test Set

	precision	recall	f1-score	support
0	0.98	0.99	0.98	450
1	0.99	0.98	0.98	450
2	0.98	0.97	0.98	451
accuracy			0.98	1351
macro avg	0.98	0.98	0.98	1351
weighted avg	0.98	0.98	0.98	1351

4.3.4 Confusion Matrix

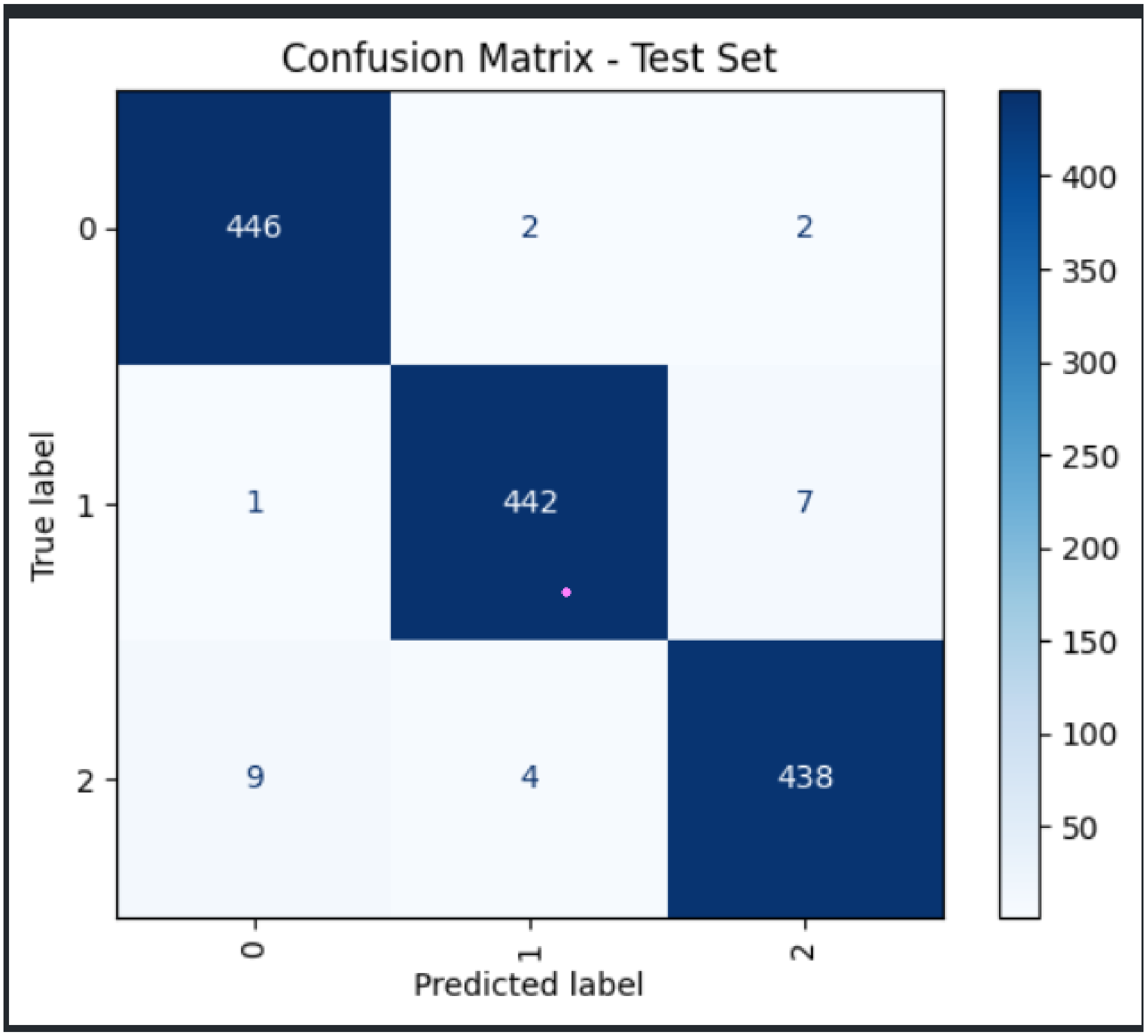

Figure 4: Confusion matrix for test set predictions.

4.4 Discussion

The results show that the model achieved an overall accuracy of 99% on the validation set and 98% on the test set. These performances are slightly higher than those obtained with four classes in the first project. Reducing the number of classes appears to have simplified the classification task, leading to slightly improved metrics.

5 Comparing the two examples

Comparing the two experiments, we can conclude that:

- Reducing the number of classes led to a marginal improvement in overall performance.
- Both models demonstrated robust capability in distinguishing classes, with high F1 scores across all categories.
- Overall accuracy increased from 97% to 98% on the test set when reducing from four to three classes.

These results highlight the importance of class definitions in classification tasks and demonstrate that simplifying the task can lead to measurable performance gains.

6 Interface

we developper a user interface with streamlit that allow users to pass IRM images to classify the images.

Figure 5: Interface With Streamlit.

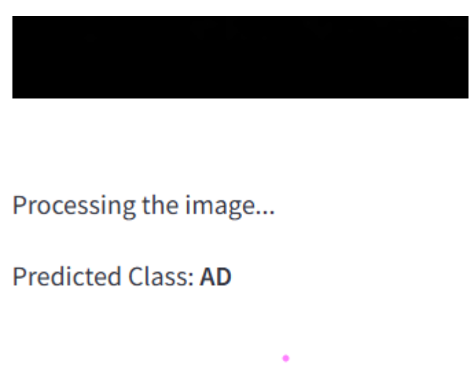

Figure 6: Interface ouput.

7 Comparison of Pretrained Models

7.1 Description

In this section, we evaluated several pretrained models for image classification, including Residual Net, Efficient Net, and Mobile Net. These models, initially trained on large datasets such as ImageNet, were fine-tuned for our specific task using transfer learning. The objective was to compare their performances with those of the custom models developed in the previous projects.

7.2 Models and Results

7.2.1 1. Residual Net

Residual Net achieved the best performance among the three models, demonstrating strong convergence and robust accuracy.

Training and Validation Performance:

- **Epoch 1:** Train Loss: 1.0809, Validation Loss: 1.0115, Accuracy: 51.41%
- **Epoch 5:** Train Loss: 0.8981, Validation Loss: 1.1198, Accuracy: 43.84%
- **Epoch 10:** Train Loss: 0.6518, Validation Loss: 0.8229, Accuracy: 64.21%

Test Results:

- Correct Predictions: 662 out of 1,031
- **Accuracy:** 64.21%

7.2.2 2. Efficient Net

Efficient Net showed moderate performance with stable but lower accuracy compared to Residual Net.

Training and Validation Performance:

- **Epoch 1:** Train Loss: 1.1364, Validation Loss: 0.9980, Accuracy: 54.41%
- **Epoch 5:** Train Loss: 1.0331, Validation Loss: 1.5481, Accuracy: 39.67%

- **Epoch 10:** Train Loss: 1.0295, Validation Loss: 1.1158, Accuracy: 51.50%

Test Results:

- Correct Predictions: 530 out of 1,031
- **Accuracy:** 51.50%

7.2.3 3. Mobile Net

Mobile Net demonstrated intermediate performance between Residual Net and Efficient Net, with accuracy improving over epochs but falling short of Residual Net.

Training and Validation Performance:

- **Epoch 1:** Train Loss: 1.1493, Validation Loss: 1.0310, Accuracy: 53.25%
- **Epoch 5:** Train Loss: 1.0011, Validation Loss: 0.9529, Accuracy: 55.58%
- **Epoch 10:** Train Loss: 0.9622, Validation Loss: 0.8914, Accuracy: 56.74%

Test Results:

- Correct Predictions: 585 out of 1,031
- **Accuracy:** 56.74%

7.3 Comparative Analysis

Metric	Residual Net	Efficient Net	Mobile Net
Validation Loss	0.8229	1.1158	0.8914
Accuracy (%)	64.21	51.50	56.74

Table 1: Performance comparison of models.

Residual Net outperformed the other models in terms of accuracy and validation loss, making it the most effective model for this task. Efficient Net struggled with stability, particularly with higher validation loss. Mobile Net, while consistent, showed slightly lower accuracy compared to Residual Net.

7.4 Conclusion

Residual Net is the most suitable model for classifying Alzheimer's diagnoses in this study. Its higher accuracy and lower validation loss indicate better generalization and performance on the test dataset.

General Conclusion

This study explored the classification of Alzheimer's disease using convolutional neural networks (CNNs) and pretrained models, focusing on the impact of dataset structure and complexity on performance. Two primary experiments were conducted : one involving classification with four classes ("Mild Demented," "Moderate Demented," "Non-Demented," and "Very Mild Demented") and another with three classes ("Alzheimer's Disease," "Mild Cognitive Impairment," and "Cognitively Normal"). The four-class model achieved 86 PERCENT accuracy, struggling with the "Moderate Demented" class due to limited samples, while the three-class model reached 98 PERECENT accuracy, highlighting the benefits of reduced complexity. Among pretrained models, Residual Net outperformed Efficient Net and Mobile Net with 64.21 PERECENT accuracy. Future work should focus on enhancing datasets by incorporating additional imaging modalities such as PET scans and expanding the diversity of the sample population to include underrepresented demographics. Advanced architectures like Vision Transformers and graph neural networks should be explored for improved feature extraction and contextual understanding. Explainability techniques, such as Grad-CAM and SHAP, should be employed to gain deeper insights into model decisions. Validation efforts must include benchmarking on external datasets from diverse clinical settings and real-world testing in collaboration with healthcare professionals. Integration of longitudinal data and the development of predictive tools to assess the progression from mild cognitive impairment to Alzheimer's disease are essential. Furthermore, leveraging federated learning frameworks can ensure patient privacy while enabling robust model training across decentralized datasets. These efforts aim to create accurate, interpretable, and deployable solutions for improving Alzheimer's disease diagnosis and management.

Bibliography

[AAara] Firstname Author1 and Firstname Author2. A comprehensive review on the application of 3d convolutional neural networks in medical image analysis. *MDPI*, Volume:Pages, Year.

[AAarb] Firstname Author1 and Firstname Author2. Convolutional neural networks in medical image understanding. *PMC*, Volume:Pages, Year.

[AAarc] Firstname Author1 and Firstname Author2. Deep learning for alzheimer's disease prediction: A comprehensive review. *PMC*, Volume:Pages, Year.

[AAard] Firstname Author1 and Firstname Author2. A deep learning mri approach outperforms other biomarkers of alzheimer's disease. *Alzheimer's Research*, Volume(Issue):Pages, Year.

[AAare] Firstname Author1 and Firstname Author2. Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, Volume:Pages, Year.