Deep Learning-Based Classification and Diagnosis of Alzheimer's & Dementia Using Multi-scale Feature Extraction from Baseline MRI Scans

Manel Femmam[1,*], Smain Femmam[2], Mohamed E. Fareh[3], Omar A. Senni[1], and Abdelnour Ferhani[1]

[1] Laboratory of Science for mathematics, computer science and engineering applications, Computer Science Department, Institute of science, University Center of Barika, Barika, Algeria
[2] Computer & Networks Department, Faculty of Sciences, Haute-Alsace University, Mulhouse, France
[3] Computer Science Department, Biskra University, Biskra, Algeria
Email: manel.femmam@cu-barika.dz (M.F.); smain.femmam@uha.fr (S.F.); fareh.me@gmail.com (M.E.F.); senniomarabderrahmane@gmail.com (O.A.S.); abdelnourferhani@gmail.com (A.F.)
*Corresponding author

***Abstract*—Diagnosing neurodegenerative diseases such as Alzheimer's represents a significant challenge in medicine, primarily based on the assessment of symptoms by healthcare professionals. Early detection and appropriate management are crucial to improve patients' quality of life. While medical expertise is essential for identifying early signs, there is a need for automated tools to assist physicians in diagnosis. In this context, our goal is to explore the capability of different new technical computer approaches pre-trained models of CNN (Res-Net, VGG-16, Mobile-Net) with transfer learning for pathological brain image classification (Alzheimer's Disease Detection ADD) using Magnetic Resonance Imaging (MRI) images. Our project specifically aims to classify and automatically detecting four classes (Mild Dementia, Moderate Dementia, Non-Demented, Very Mild Dementia). We rely on deep learning, particularly Convolutional Neural Networks (CNNs), which have demonstrated effectiveness, especially in the medical field. We utilized a specific CNN model, which yielded satisfactory results, confirming the performance of our models. This best of these models could be deployed in clinical settings for early testing and identifying patients at risk of developing Alzheimer's disease.**

***Keywords*—deep learning, Alzheimer's disease, brain MRI, convolutional neural network, transfer-learning, medical image classification**

I. Introduction

Alzheimer's disease is a progressive neurological disorder leading to brain atrophy and death, constituting the most common cause of dementia, and characterized by a persistent decline in cognitive abilities. Early symptoms include recent memory loss, forgetting conversations, and difficulties in performing daily tasks. In our study, we will utilize Magnetic Resonance Imaging (MRI) to assess the degeneration of brain cells associated with the disease. This technique employs radio waves and magnetic fields to generate a detailed representation of the brain, enabling visualization of areas affected by Alzheimer's disease. We will also explore the four stages (Mild Dementia (MD), Moderate Dementia (MOD), Non-Demented (ND), Very mild Dementia (VMD)) of Alzheimer's disease and discuss how MRI images can aid in diagnosing the disease and evaluating its progression. Recent research has leveraged artificial intelligence to develop software capable of detecting the disease by identifying specific brain alterations and using deep learning algorithms based on artificial neural networks (Res-Net, Vgg-16 and Mobile-net). Thus, in this paper we are interested to explore the capability of deep-learning architecture for Alzheimer disease based on MRI images. We aim to improve previous obtained results in some existing research, which have been experimented on the same dataset.

The structure of this paper is as follows: Section II covers the literature review, providing an overview of fundamental concepts such as deep learning, artificial neurons, and convolutional neural networks. Section III details the dataset used in this study. Section IV discusses the results obtained and outlines the architecture of the model. Finally, the paper concludes in Section V.

II. Literature Review

Deep learning (DL) is recognized as a groundbreaking branch of machine learning (ML) research, aimed at advancing ML toward its foundational goal of achieving artificial intelligence. DL architectures typically consist of multiple layers of abstraction and representation, which

Manuscript received June 4, 2024; revised June 22, 2024; accepted August 22, 2024; published December 16, 2024.

facilitate the interpretation of textual, auditory, and visual data [1]. The primary distinction between deep learning and traditional machine learning lies in feature processing: deep learning automatically learns data representations, while traditional machine learning generally requires manual feature extraction. Deep learning has consistently demonstrated superior performance in data classification tasks, validating its accuracy and utility across various domains, including healthcare [2]. Since 2013, publications on deep learning have increased significantly. The exploration of new neural network architectures accelerated during this period, driving the adoption of deeper models, particularly in medical image processing [3]. The importance of deep learning for Alzheimer's disease (AD) detection became evident, with a rapid rise in related publications around 2017, as shown in Fig. 1.

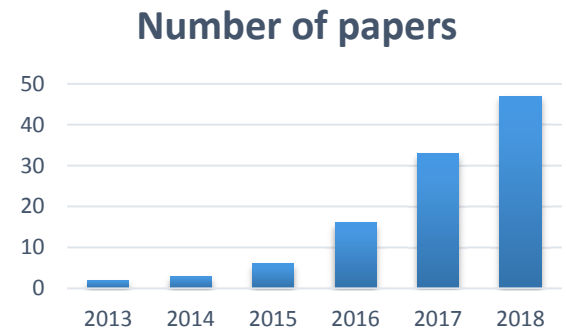

Fig. 1. Increase in deep learning publications on Alzheimer's disease detection using neuroimaging modalities [3].

Following deep learning's success in classifying 2D natural images, there has been a growing number of studies aimed at leveraging these techniques for medical imaging applications [4–5]. Convolutional Neural Networks (CNNs or Conv-Nets) can uncover hidden representations, identifying relationships between different regions of an image, synthesizing overall insights, and effectively detecting disease-related pathologies [6]; CNNs eliminate the necessity for manual feature extraction by automatically deriving features and assigning weights and biases to describe different aspects of an image. This process helps in distinguishing between various elements. Technically, each input image is processed through two main blocks of layers: the first block comprises convolutional layers that generate the features, while the second block, which is dedicated to classification, includes fully connected layers similar in structure to those in a Multi-Layer Perceptron (MLP). Fig. 2 illustrates the layout of the layers in a CNN [7–8].

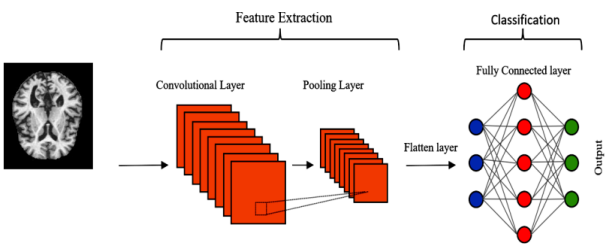

Fig. 2. Standard architecture of a CNN.

Deep learning models have proven effective in analyzing various types of medical images, including structural MRI, Functional MRI (fMRI), Positron Emission Tomography (PET), and Diffusion Tensor Imaging (DTI). Fig. 3 illustrates the prevalence of these neuroimaging modalities in single-modality studies.

Based on our literature review, MRI stands out as the leading diagnostic modality for detecting Alzheimer's disease, which has guided our focus on MRI scans in this paper. In contrast to large-scale image classification datasets with millions of images, neuroimaging datasets usually have only hundreds of images. This limited amount of training data can lead to overfitting issues [9]. In practical terms, it's typical to start with established pre-trained CNNs tailored to a specific domain task, then refine them for new tasks by adjusting only their last layers during retraining [10, 11].

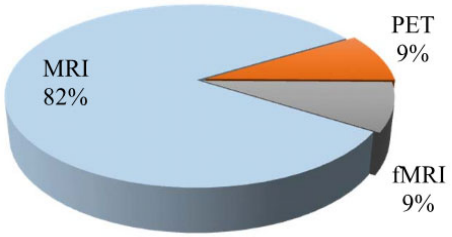

Fig. 3. Prevalence of neuroimaging modalities in single-modality studies [3].

This stems from the fact that the lower layers of CNNs possess more generic features, applicable across a wide range of tasks, and can be transferred from one application domain to another. This concept, known as "transfer learning," has been a powerful tool for training large networks without overfitting. Studies have indicated that transfer learning, even across different domains, yields quicker results and better performance compared to training from scratch [12, 13]. A first transfer learning approach for AD detection using deep learning can be found in the work of Helaly *et al.* [14] introduced a method aimed at the early detection of Alzheimer's disease through a cohesive approach. Their study leveraged Convolutional Neural Networks (CNN) to classify AD, employing two techniques for prediction. The first method utilizes basic CNN architectures to process 2D and 3D structural brain scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, employing both 2D and 3D convolutions. The second method incorporates transfer learning, leveraging pre-trained models like VGG-19 to enhance medical image classification tasks. The results demonstrate that the CNN architectures employed in the initial method exhibit the following traits: they possess simple structures that are well-suited for reducing computational complexity and memory requirements, thereby mitigating overfitting, and ensuring manageable processing times. Additionally, they yield highly encouraging accuracies of 93.61% and 95.17 % for the classification of multi-class AD stages in both 2D and 3D scenarios. Furthermore, fine-tuning of the pre-trained VGG19 model led to an accuracy of 97 % in multi-class AD stage classification. Another work proposed in [15]

developed a model that employs optimized parameters to predict Alzheimer's disease. Classifiers such as Gradient Boosting, Support Vector Machine, Decision Tree, and Voting were utilized to detect AD. The study's findings demonstrate outstanding performance, achieving an average validation accuracy of 83%. This accuracy score during testing surpasses the results of previous endeavors by a significant margin.

Nawaz *et al.* [16] developed a model based on deep features utilizing a pre-trained AlexNet model. They achieved this by transferring the initial layers from the pre-trained AlexNet model and extracting deep features from the Convolutional Neural Network (CNN). To classify the extracted deep features, they employed widely used machine learning algorithms including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF). Evaluation results of the proposed approach indicate that the deep feature-based model surpassed both handcrafted and deep learning methods, achieving an accuracy of 99.21%. Furthermore, the proposed model outperformed existing state-of-the-art methods.

Aderghal *et al.* [17] introduced another approach to transfer learning, wherein they trained three 2D CNNs, each comprising two convolutional layers, on just three slices located at the center of the hippocampal region within certain MRI scans. Given a limited number of DTI images, rather than starting from scratch, they opted for transfer learning from models previously trained on MRI images to the target DTI dataset. Ultimately, they merged all networks and reached a final decision using a majority voting strategy.

Payan *et al.* [18] utilized a combination of a sparse autoencoder and 3D convolutional neural networks in their study. They developed an algorithm designed to identify the disease status of an individual based on a brain Magnetic Resonance Image (MRI) scan. Notably, the incorporation of 3D convolutions represented a significant advancement, outperforming traditional 2D convolutions. Although the convolutional layer was pre-trained using an autoencoder, fine-tuning was not conducted. However, it is anticipated that performance could be enhanced through fine-tuning [19]. In general, to utilize renowned architectures for transfer learning, the issue typically requires simplification from a 3D volume to 2D image slices. Architectures based on slices are often built assuming that 2D image slices can encapsulate specific brain properties. However, rapid training might lead to the loss of spatial relations among slices. As stated in [2], transfer learning models can attain comparable accuracy to a 3D-CNN model trained from the ground up [20, 21].

The study in [22, 23] focuses on developing a diagnostic system using Convolutional Neural Networks (CNN) to distinguish between healthy individuals and Alzheimer's patients. Using 18FDG-PET images of 855 patients (635 healthy individuals and 220 Alzheimer's patients) from the ADNI database, the proposed system achieved an accuracy of 96%, a sensitivity of 96%, and a specificity of 94%, demonstrating its strong performance.

In [24] the authors using transfer learning to build a model for detecting Alzheimer's disease using Magnetic Resonance Imaging (MRI) of the brain. This model classifies images into four stages: Mild Dementia (MD), Moderate Dementia (MOD), Non-Dementia (ND), and Very Mild Dementia (VMD). The pre-trained Alex-Net model was used for this task and achieved excellent performance with an accuracy of 91.70 %. In [25], the authors employed a hybrid ResD approach combining ResNet18 and DenseNet121 for multi-class classification of Alzheimer's disease using an MRI dataset. By integrating information from both pre-trained models [26], they enhanced the classification process. Experimental results demonstrate that the proposed hybrid model surpasses existing techniques. The ResD model achieved a weighted average (macro) accuracy score of 99.61%. A work in [27], the researchers utilized various transfer learning methods based on CNNs for the classification of Alzheimer's disease. They applied different parameters and achieved a remarkable accuracy of 99.05% on the ADNI benchmark dataset. They tested 13 different versions of various pre-trained CNN models using a fine-tuned transfer learning approach in two different domains on the ADNI dataset 94 AD, 138 Mild Cognitive Impairment (MCI), and 146 Normal Controls (NC).

More recently, the largest database of labeled MRI images in the detection of different levels of Alzheimer's was made available on Kaggle [28] containing 6400 images MRI, was addressed in the work of Andrea Loddo, Sara Buttau *et al.* [29], when having implemented a model based on a learning method by ensemble by merging the outputs of three pre-trained models: Alex-Net, ResNet-101, Inception-ResNet-V2 which were adapted to the problem having an accuracy of 97.71% for multi class classification and 96.57% accuracy for binary classification, representing the best result currently obtained on this database.

III. Materials and Methods

In this section, we are interested to build a powerful deep learning model for early detection of Alzheimer Disease (AD) and classification of medical images for different stages of AD. We rely on the CNN model using pre-trained transfer learning models VGG16, ResNet50 and Mobile-Net. Then we compared the results based on the confusion matrix of each model. Four classification metrics were used: mild dementia, very mild dementia, moderate dementia, and non-dementia. To make it more convenient for patients and doctors, the authors developed a web application for remote analysis and verification of AD. It also determines the patient's AD stage based on the AD spectrum.

A. Data Description

The Open Access Series of Imaging Studies (OASIS) [30] MRI dataset, comprising 80,000 brain MRI images, serves as a vital resource for analyzing and identifying early indicators of Alzheimer's disease. The dataset categorizes patients into four distinct classes based on Alzheimer's progression, as determined by metadata

and Clinical Dementia Rating (CDR) values. These classes include demented, very mild demented, mild demented, and non-demented, facilitating the examination and comprehension of various stages of Alzheimer's progression. The primary labeled data sets of four classes (MD, MOD, ND, and VMD) used for training and validation were obtained from OASIS dataset show in Fig. 4. The "non-demented" class represents healthy CN subjects, the "moderate demented" class represents subjects with Alzheimer's Disease (AD), and the two classes "mild demented" and "very mild demented" represent the MCI (Mild Cognitive Impairment) stage. The number of subjects is provided in Table I.

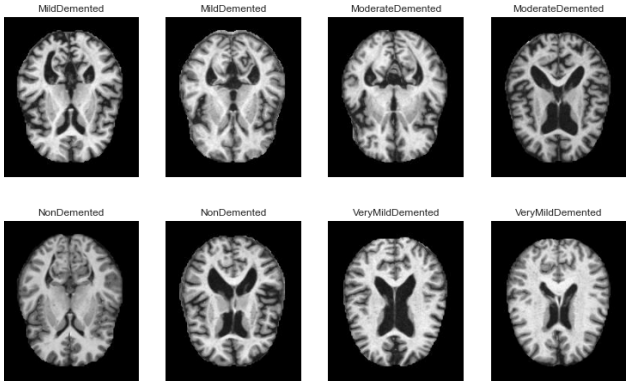

Fig. 4. The sample MRI images from the OASIS dataset.

TABLE I. THE CLASSIFICATION OF THE DATABASE USED

Class Name	Number of images per class
Mild Dementia	5002
Moderate Dementia	488
Non-Demented	67.2 k
Very mild Dementia	13.7 k

B. Data Pre-processing

1) Data splitting

Prior to any processing, it is necessary to partition the images into two sets: a training, validation set, and a test set presented in Table II. to ensure later testing of the model on data it has not previously encountered. This will ensure an accurate evaluation of our model's performance. All classes were divided as follows: 90% for training and 10% for testing, except for moderate dementia, for which we took 20% for testing.

TABLE II. DATA SPLITTING

Class Name	Train &Validation	Test
Mild Dementia	4501	501
Moderate Dementia	390	98
Non-Demented	60499	6723
Very mild Dementia	12352	1373

2) Data balancing

As we observed in the database description earlier, there was a significant difference in the number of data samples in each class. This imbalance negatively affects model performance, as it may lead to misleading outputs (high accuracy for the majority class while performing poorly on minority class samples). Therefore, we need to apply data balancing techniques. Since we are dealing with medical images, it is not advisable to use data augmentation techniques such as rotation and color changes, as they may lead to incorrect diagnoses.

Based on this, we have decided to use oversampling and undersampling techniques. Oversampling involves taking only a subset of data from the majority classes and repeating images in the minority classes until reaching the desired number. While undersampling involves taking only a subset of data from the majority classes.

In Table III, we show the number of data points in each class after balancing (note that oversampling was not applied to the test data).

TABLE III. NUMBER OF DATA AFTER BALANCING

Class Name	Train &Validation	Test
Mild Dementia	4500	450
Moderate Dementia	4500	98
Non-Demented	4500	450
Very mild Dementia	4500	450

3) Training, validation and testing datasets

At this stage, we create two datasets: one for training and validation, and the other for testing (see Table IV). Both datasets contain the classes, each with its own images. The images were resized to 224×224 pixels to meet the requirements of pre-trained models such as VGG-16, ResNet-50, and MobileNet. Additionally, we enhance the model's efficiency and resource usage by setting the batch size to 32. Subsequently, we divide the first dataset into 90% for training and 10% for validation, ensuring a robust evaluation of the model.

TABLE IV. TRAINING, VALIDATION AND TESTING DATASETS

Class Name	Train	Validation	Test
Mild Dementia	4050	450	450
Moderate Dementia	4050	450	98
Non-Demented	4050	450	450
Very mild Dementia	4050	450	450

IV. RESULT AND DISCUSSION

As aforementioned, the dataset used in our work was taken from the OASIS study. This section aims to present obtained results of our proposed model trained on the mentioned dataset, Fig. 5 illustrates the steps of implementation. The performance of the model is discussed through the evaluation of three metrics: Accuracy, Precision, and F1score. The three metrics are computed using four sets: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) by confusion matrix (see Fig. 6). As a reminder, a true positive occurs when the model accurately identifies the positive class (or conversely, a true negative when it accurately identifies the negative class), while a false positive arises when the model mistakenly identifies the

positive class (or conversely, a false negative when it mistakenly identifies the negative class).

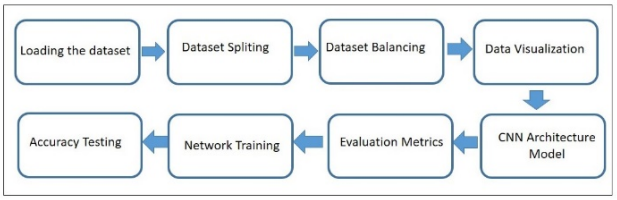

Fig. 5. Steps of implementation.

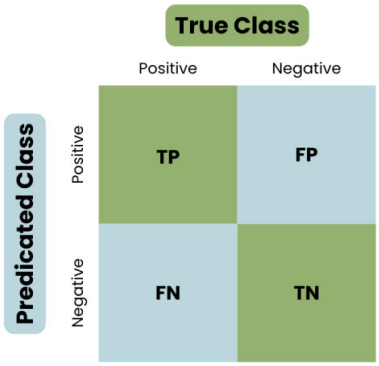

Fig. 6. The basic structure of confusion matrix.

In addition to these three metrics, we assessed the cross-entropy loss (the loss function) between the true labels and predicted labels during the training phase. The equations for the four metrics (accuracy, precision, F1 score, and loss function) are provided below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Loss = -\sum_{i=1}^{n} t_i \times \log(p_i) \quad (3)$$

$$F1-Score = 2\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Such that n is the number of classes, t_i is the truth label, and p_i is the Softmax probability for the *i-th* class. The results are detailed in the following three sections: evaluation of the training phase, evaluation of the test phase, and the confusion matrix.

A. Results of the Test Phase

After explaining the criteria that we will rely on to evaluate the model, we will now begin presenting the results achieved.

1) Accuracy

The following Table V shows the accuracy achieved by each model in training and testing.

TABLE V. MODELS ACCURACY

Model	Training Accuracy	Test Accuracy
Rest-Net	0.9914	0.9758
VGG-16	0.9904	0.9696
Mobile-Net	0.9548	0.9302

As shown in the table above, the testing accuracy was excellent, exceeding 90% across all models. This demonstrates that the three models effectively generalize the knowledge gained during training to unseen data. Following the presentation of the final accuracy for each model, we now analyze the accuracy development curves during the training phase. Figs 7–9 illustrate the accuracy curves for ResNet, VGG-16, and MobileNet, respectively.

Initially, the validation accuracy increases steadily, but at a certain point, fluctuations occur with alternating rises and falls. To address this, we adjusted the models' settings to save the weights corresponding to the lowest validation loss. This adjustment ensures consistently high validation accuracy.

As we notice, the accuracy of the validation data initially begins to rise normally, but at some point, the values begin to fluctuate between rising and falling. To deal with this, we have adjusted the models' settings so that they save the weights in which the loss during validation is at the lowest value. This will also ensure obtaining high accuracy in validation data.

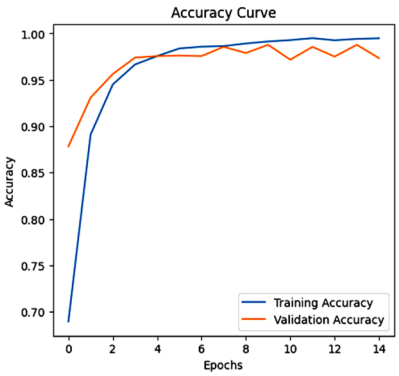

Fig. 7. Accuracy for Res-Net.

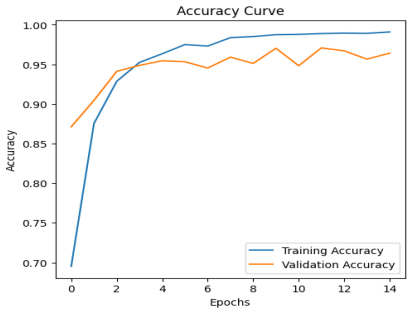

Fig. 8. Accuracy for VGG-16.

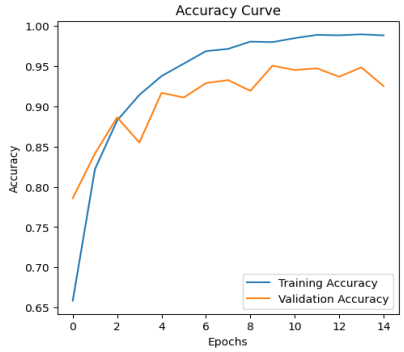

Fig. 9. Accuracy for Mobile-Net.

2) Precision, recall and F1-score

We relied on these three criteria to evaluate the performance in each class. The following tables (Tables VI–VIII) show the results achieved by each model in each class.

TABLE VI. REST-NET EVALUATION METRICS

Class Name	Precision	Recall	F1-score
Mild Dementia	0.98	1	0.99
Moderate Dementia	1	1	1
Non-Demented	0.95	0.97	0.96
Very mild Dementia	0.98	0.94	0.96

TABLE VII. VGG-16 EVALUATION METRICS

Class Name	Precision	Recall	F1-score
Mild Dementia	0.98	0.99	0.98
Moderate Dementia	1	0.98	0.99
Non-Demented	0.97	0.93	0.95
Very mild Dementia	0.95	0.97	0.96

TABLE VIII. MOBILE-NET EVALUATION METRICS

Class Name	Precision	Recall	F1-score
Mild-Dementia	0.98	0.98	0.98
Moderate Dementia	0.98	0.98	0.98
Non-Demented	0.87	0.92	0.89
Very mild Dementia	0.92	0.86	0.89

On the Res-Net model, classes Mild dementia and Moderate Dementia achieved a very high rating, as the full percentage was achieved in all measures for the second. For the first one, the results were between 0.98 and 1.

As for the classes Non Demented and Very Mild Dementia, their evaluation was somewhat lower than the other two classes. However, the decline was not high, as percentages above 92% were achieved in all classes.

In VGG-16 the result was somewhat similar to Res-Net, where classes Mild dementia and Moderate dementia also achieved a high rating, with scores on all metrics greater than 0.98. However, they are slightly lower than the evaluations of the Res-Net model.

As for Non Demented and Very Mild Dementia, there was a slight decrease in the level of results compared to the previous two, so all measures were between 93% and 97%.

Like the other two models, Mobile-Net scores well on all metrics for the mild and moderate dementia classes, with all ratings being 0.98. However, it witnessed a noticeable decline in the remaining two classes, as Precision's rating in non demented class reached 0.87, meaning that the model predicted a large amount of data within this class even though it belongs to another class, and Recall's rating in very mild dementia reached 0.86, where it predicted the model has 14% of the data for this class in another class.

3) Loss

The loss function during the training phase for the Res Net, Vgg-16 and Mobile-Net (see Figs. 10–12) respectively.

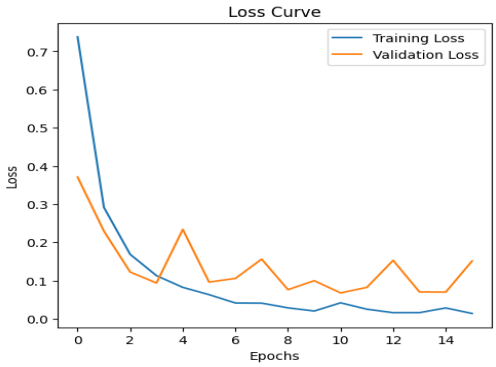

Fig. 10. Loss for Res-Net.

4) Confusion matrix

These matrixes show the numbers of data that were correctly classified into the target class and those that were correctly classified outside of it, as well as the number of data that were incorrectly classified into the target class and that were incorrectly classified outside of it. As the images show, across all models, many data were predicted correctly, but there were classes that significantly outperformed others on Mobile-Net.

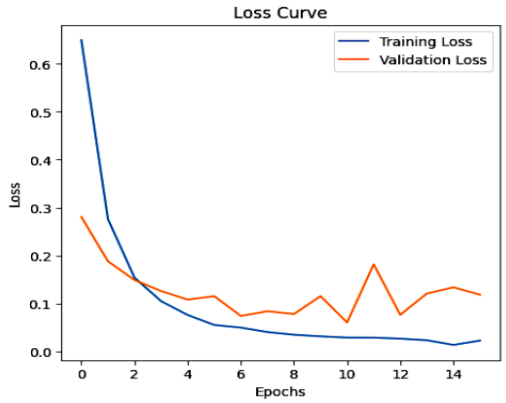

Fig. 11. Loss for Vgg-16.

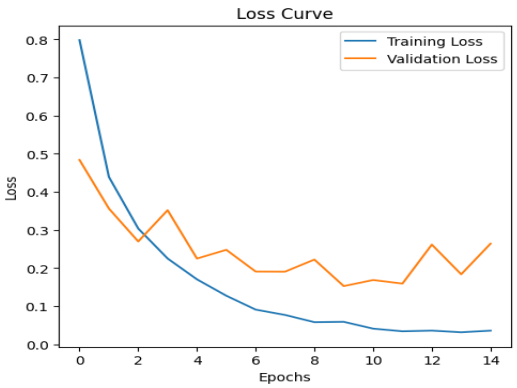

Fig. 12. Loss for Mobile-Net.

As for Res-Net and VGG-16, although it also had superior classes, but even the less accurate classes had most of the data predicted correctly, specifically on Res-Net. Figs. 13–15 illustrate the confusion matrix, which reflects the model's performance.

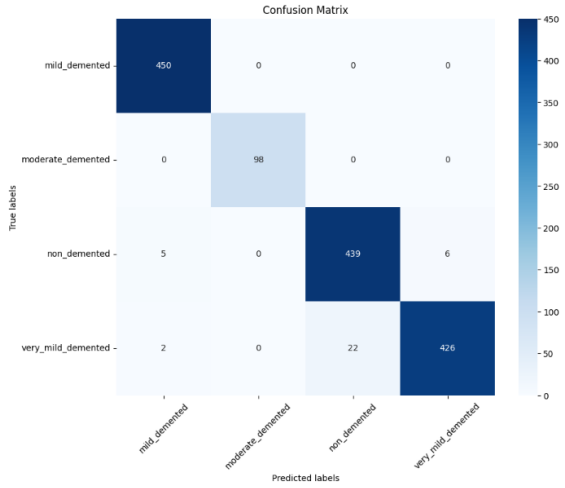

Fig. 13. Res-Net Confusion matrix.

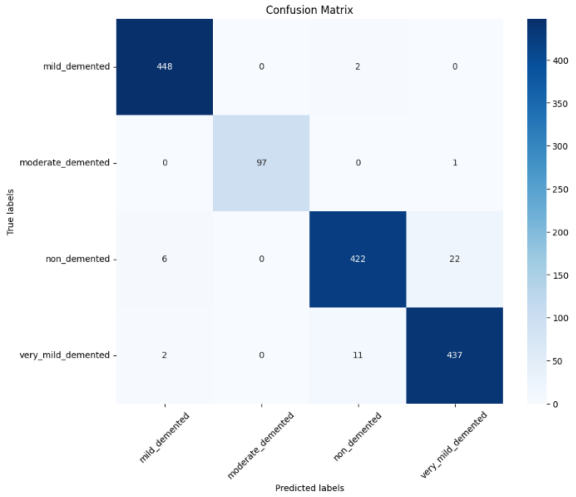

Fig. 14. Vgg-16 confusion matrix.

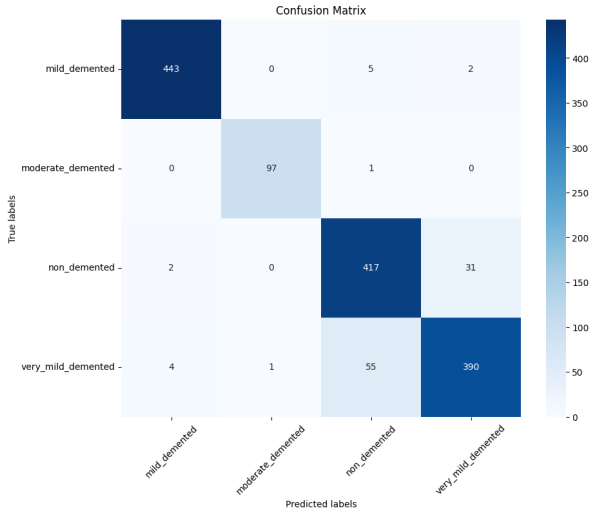

Fig. 15. Mobile-Net confusion matrix.

5) Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) score

Res-Net demonstrates exceptional performance with ROC AUC scores of 1.0000 for the mild-demented and moderate-demented classes, and high scores of 0.9980 and 0.9988 for the non-demented and very-mild-demented classes respectively (see Table IX). The micro-average ROC AUC of 0.9994 indicates strong overall discriminative ability across all classes, as illustrated in Fig. 16 (ResNet-50 AUC-ROC curve). VGG-16 also performs well, achieving ROC AUC scores of 1.0000 for the moderate-demented class. However, it shows slightly lower scores compared to ResNet-50 for the non-demented and very-mild-demented classes (0.9906 and 0.9946 respectively), as depicted in Fig. 17 (VGG-16 AUC-ROC curve). MobileNet exhibits moderate performance with ROC AUC scores of 0.9983, 1.0000, 0.9807, and 0.9840 for the mild-demented, moderate-demented, non-demented, and very-mild-demented classes respectively. The micro-average ROC AUC of 0.9933, shown in Fig. 18 (MobileNet AUC-ROC curve), indicates decent discriminative ability but slightly lower compared to ResNet-50 and VGG-16.

TABLE IX. AUC-ROC SCORE

Class Name	Res-Net50	VGG-16	Mobile-Net
Mild Dementia	1	0.9998	0.9983
Moderate Dementia	1	1	1
Non-Demented	0.9980	0.9906	0.9807
Very mild Dementia	0.9988	0.9946	0.9840
Micro average	0.9994	0.9971	0.9933

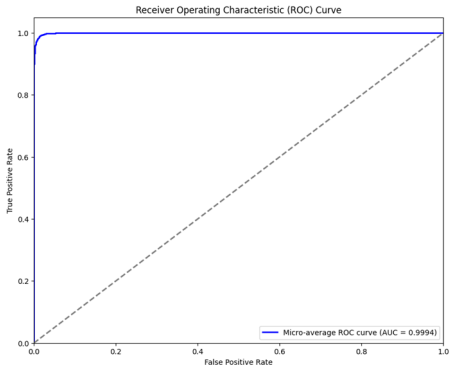

Fig. 16. Res-Net50 AUC-ROC curve.

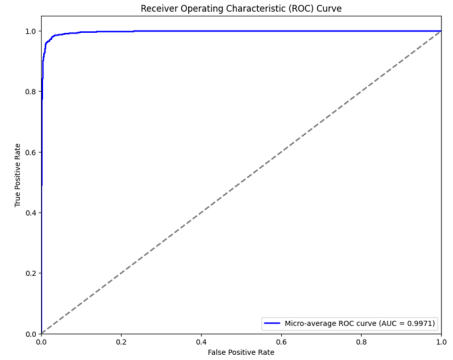

Fig. 17. VGG-16 AUC-ROC curve.

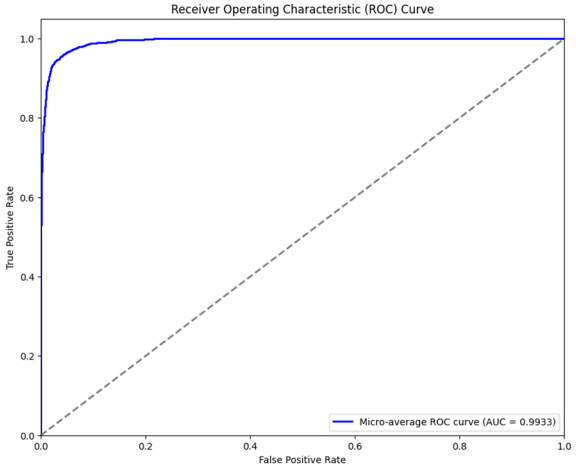

Fig. 18. Mobile-Net AUC-ROC curve.

B. Model Description

1) Model architecture

Based on our evaluations, the ResNet model showed the highest accuracy and was therefore chosen as our final model. We started the model building process by loading the ResNet50 model while preserving its pre-trained weights by freezing all layers. This approach aims to reduce training time and ensure that the features learned by the model on the ImageNet dataset remain intact and are not affected by any modification during training.

However, since freezing layers in pre-trained models limits their ability to learn from new data, additional layers must be added to ensure the model adapts to the new dataset and achieves excellent performance. Based on this, we added new layers on top of the ResNet50 base. Our model starts with two convolutional layers, each containing 32 filters of size (3, 3) with a stride of (1, 1), ReLU activation, and "same" padding. These layers aim to extract additional features from the input images and adapt previously extracted features from the ResNet50 model to the new dataset, followed by normalization to standardize activations and enhance convergence. Next comes the MaxPooling layer, for which we haven't set any specific values, meaning it will use the default dimensions. Therefore, their values will be as follows: kernel size (2, 2), stride (2, 2), and "valid" padding. This layer reduces the dimensionality of the input and effectively captures the essential information.

Afterward, the data is flattened and passed through a fully connected layer. This layer contains 256 nodes and utilizes the ReLU activation function. This layer facilitates the capture of complex patterns within the feature maps created by the previous layers. However, adding extra layers may pose the risk of overfitting; to address this issue, a dropout layer was incorporated with a value of 0.6. This value helped solve the problem of overfitting without affecting the model's accuracy.

Finally, we add an output layer that uses softmax activation to generate probability distributions for the classes in our classification task. Fig. 19 shows the model architecture.

2) Training setting

After presenting the model architecture, we will be presenting the training settings. The Adam algorithm was used as an optimizer to update the model weights during training with a learning rate of 0.001. We also chose the loss function sparse categorical cross entropy, which is considered appropriate for classification problems. Additionally, we used accuracy as a measure to monitor the model's performance during training.

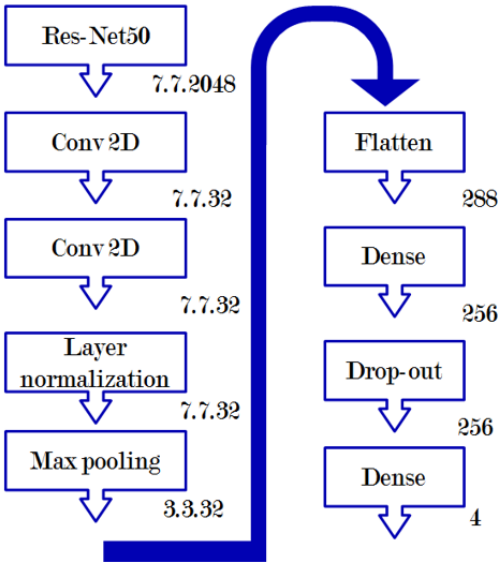

Fig. 19. Proposed model.

Regarding the number of epochs, we chose 20 but included early stopping using the Early Stopping callback. If there was no decrease in the loss for 5 epochs, training would stop early. We also requested to restore the weights at the point where the validation loss was at its lowest value.

C. Comparison with Other Related Works

To conclude this work, Table X. compares our proposition to some related works, where we have chosen accuracy as comparison criteria. In our study, we achieved an accuracy of 98% for the multi-class classification of Alzheimer's disease using the OASIS dataset. This result is comparable and even slightly superior to some approaches using similar techniques. Furthermore, we are the only ones among these studies to utilize the OASIS dataset, demonstrating the robustness and generalization of our model.

TABLE X. COMPARISON OF OUR RESULT TO OTHER WORKS

Work	Accuracy	Number of classes	Dataset
Our work (Multi class classification)	98 %	4	OASIS
CNN model (Binary Classification) [23]	96 %	2	ADNI
CNN+ Transfer Learning [24]	91.70%	4	Kaggle
ResD (Resnet18 + Densenet121) [25]	99.61%	4	ADNI
CNN + Transfer Learning [27]	99,05%	4	ADNI
AlexNet, ResNet101, InceptionResNet-V2 [29]	97.71% 96.57%	4 2	Kaggle

V. Conclusion

In recent years, researchers have become increasingly interested in deep learning and its models for the classification of medical images and the segmentation of brain images, to assist in the diagnosis of certain diseases, such as those affecting the brain.

The aim of our work was to achieve the classification of Alzheimer's disease from brain MRI images using Transfer-learning, specifically convolutional neural networks. Convolutional neural networks are currently the most effective models for image classification compared to other machine learning models.

In our study, we classified four stages of Alzheimer's disease MD, MOD, ND, and VMD using three architectures. The first architecture represents the pre-trained Res-Net, the second architecture represents VGG-16 and the third is Mobile-Net. We implemented all three architectures by Colab. According to the results obtained, the ResNet model performs better for classification than VGG-16 and MobileNet. In terms of future work, our approach can be improved to achieve better results by following these recommendations:

- Using a different dataset than the one we used.
- Utilizing other CNN models such as AlexNet and GoogleNet.
- Employing other deep learning models like RNN.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Femmam M. conceived the idea of the paper, conducted the research, and wrote the paper; Femmam S. developed the methodology; Mohamed E. Fareh performed the formal analysis and examined the data; Omar A. Senni and Abdelnour Ferhani developed the software; all authors approved the final version.

References

[1] L. Deng *et al.*, "Deep learning: Methods and applications," *Found. Trends Signal Process*, vol. 7, no. 3–4, pp. 197–387, Jun. 2014.

[2] Y. B. I. Goodfellow and A. Courville, *Deep Learning*, MIT press, 2016.

[3] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[4] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[5] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no 7553, p. 436–444, 2015.

[7] M. Nennouche and A. M. Atchi, "Feature fusion for classification of different levels of Alzheimer's disease dementia," *Dissertation, Dept. Electron. Eng., École Polytechnique*, Algeria, 2022. https://repository.enp.edu.dz/jspui/bitstream/123456789/10541/1/NENNOUCHE.Mohamed_ATCHI.Abdelmalek.pdf (in French)

[8] A. Alsaleh and C. Perkgoz, "A space and time efficient convolutional neural network for age group estimation from facial images," *PeerJ Computer Science*, vol. 9, e1395, 2023.

[9] A. Ebrahimi-Ghahnavieh, S. Luo, and R. Chiong, "Transfer learning for Alzheimer's disease detection on MRI images," in *Proc. 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, Bali, Indonesia, 2019, pp. 133–138.

[10] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv Print, arXiv: 1704.06857, 2017.

[11] D. Su, H. Zhang, H. Chen, J. Yi, P.-Y. Chen, and Y. Gao, "Is robustness the cost of accuracy? A comprehensive study on the robustness of 18 deep image classification models," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 631–648.

[12] A. Valliani and A. Soni, "Deep residual nets for improved Alzheimer's diagnosis," *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017, pp. 615–615.

[13] V. Wegmayr and D. Haziza, "Alzheimer classification with MR images: Exploration of CNN performance factors," *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL)*, 2018, pp. 1–7.

[14] H. A. Helaly, M. Badawy, A. Y. Haikal, "Deep learning approach for early detection of Alzheimer's disease," *Cogn. Comput*., vol. 14, p. 1711, 2021.

[15] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. T. Romero, "Early-stage Alzheimer's disease prediction using machine learning models," *Front. Public Health*, 2022.

[16] H. Nawaz, M. Maqsood, and S. Afzal *et al.*, "A deep feature-based real-time system for Alzheimer disease stage detection," *Multimedia Tools and Applications*, vol. 80, pp. 35789–35807, 2021.

[17] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning," *Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 345–350.

[18] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D deep learning on medical images: A review," *Sensors*, vol. 20, no. 18, p. 5097, 2020.

[19] K Jarrett, K. Kavukcuoglu, M. A. Ranzato and Y. LeCun, "What is the best multi-stage architecture for object recognition?" *IEEE Int. Conf. Comput. Vis.*, pp. 2146–2153, 2009.

[20] V. Wegmayr and D. Haziza, "Alzheimer classification with MR images: Exploration of CNN performance factors," *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL)*, 2018, pp. 1–7.

[21] O. Liouane, T. Bakir, S. Femmam, and A. Benabdelali, "On-line sequential ELM based localization process for large scale Wireless Sensors Network," in *Proc. The IEEE 5th International Conference on Control, Automation and Diagnosis (ICCAD'21) Grenoble*, 2021. doi:10.1109/ICCAD52417.2021.9638725

[22] R. Aschheim, S. Femmam, and M. F. Zerarka, "New 'graphiton' model: A computational discrete space, self-encoded as a trivalent graph," *Computer and Information Science Journal*, vol. 5, no. 1, pp. 1–12, 2012. doi: 10.5539/cis.v5n1p2

[23] M. Hamdi, S. Bourouis, K. Rastislav, and F. Mohmed, "Evaluation of neuro images for the diagnosis of Alzheimer's disease using deep learning neural network," *Frontiers in Public Health,* 2022.

[24] T. M. Ghazal and G. Issa, "Alzheimer disease detection empowered with transfer learning," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5005–5019, 2022.

[25] M. Odusami, R. Maskeliunas, and S. Misra *et al.*, "ResD hybrid model based on Resnet18 and Densenet121 for early alzheimer disease classification," *Intelligent Systems Design and Applications*, pp. 296–305, 2022.

[26] I. Aouissaoui, T. Bakir, A. Sakly, and S. Femmam, "Improved one-dimensional piecewise chaotic maps for information security," *Journal of Communications*, vol. 17, no. 1, pp. 11–16, 2022. doi: 10.12720/jcm.17.1.11-16

[27] A. Abida, N. Saeeda, and P. Mukesh *et al.*, "Deep transfer learning for Alzheimer neurological disorder detection," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30117–30142, 2021.

[28] Alzheimer's dataset (4 class of images)|kaggle. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-class-of-images

[29] A. Loddo, S. Buttau, and C. D. Ruberto, "Deep learning-based pipelines for Alzheimer's disease diagnosis: A comparative study

and a novel deep-ensemble method," *Computers in Biology and Medicine*, vol. 141, 105032, 2022.
[30] Open access series of imaging studies. [Online]. Available: http://www.oasis-brains.org/