Classification of Alzheimer's disease stages from magnetic resonance images using deep learning

Alejandro Mora-Rubio [Corresp., 1] , **Mario Alejandro Bravo-Ortíz** [1] , **Sebastián Quiñones Arredondo** [1] , **Jose Manuel Saborit Torres** [2] , **Gonzalo A Ruz** [3, 4, 5] , **Reinel Tabares-Soto** [1, 5, 6]

[1] Department of Electronics and Automation, Universidad Autonóma de Manizales, Manizales, Caldas, Colombia

[2] Unidad Mixta de Imagen Biomédica FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitario y Biomédica de la Comunidad Valenciana, Valencia, Spain

[3] Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile

[4] Data Observatory Foundation, Santiago, Chile

[5] Facultad de Ingeniería Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile

[6] Department of Systems and Informatics, Universidad de Caldas, Manizales, Caldas, Colombia

Corresponding Author: Alejandro Mora-Rubio
Email address: alejandro.morar@autonoma.edu.co

Alzheimer's Disease (AD) is a progressive type of dementia characterized by loss of memory and other cognitive abilities, including speech. Since AD is a progressive disease, detection in the early stages is essential for the appropriate care of the patient throughout its development, going from asymptomatic to a stage known as Mild Cognitive Impairment (MCI), and then progressing to dementia and severe dementia; is worth mentioning that everyone suffers from cognitive impairment to some degree as we age, but the relevant task here is to identify which people are most likely to develop AD. Along with cognitive tests, evaluation of the brain morphology is the primary tool for AD diagnosis, where atrophy and loss of volume of the frontotemporal lobe are common features in patients who suffer from the disease. Regarding medical imaging techniques, Magnetic Resonance Imaging (MRI) scans are one of the methods used by specialists to assess brain morphology. Recently, with the rise of Deep Learning (DL) and its successful implementation in medical imaging applications, it is of growing interest in the research community to develop computer-aided diagnosis systems that can help physicians to detect this disease, especially in the early stages where macroscopic changes are not so easily identified. This paper presents a DL-based approach to classifying MRI scans in the different stages of AD, using a curated set of images from Alzheimer's Disease Neuroimaging Initiative and Open Access Series of Imaging Studies databases. Our methodology involves image pre-processing using FreeSurfer, spatial data-augmentation operations, such as rotation, flip, and random zoom during training, and state-of-the-art 3D Convolutional Neural Networks such as EfficientNet, DenseNet, and a custom siamese

network, as well as the relatively new approach of Vision Transformer architecture. With this approach, the best detection percentage among all four architectures was around 89% for AD vs Control, 80% for Late MCI vs Control, 66% for MCI vs Control, and 67% for Early MCI vs Control.

Classification of Alzheimer's Disease stages from Magnetic Resonance Images using Deep Learning

Alejandro Mora-Rubio[1], Mario Alejandro Bravo-Ortiz[1], Sebastián Quiñones-Arredondo[1], Jose Manuel Saborit-Torres[2], Gonzalo A. Ruz[3,4,5], Reinel Tabares-Soto[1,3,6], and for the Alzheimer's Disease Neuroimaging Initiative[*]

[1]Department of Electronics and Automation, Universidad Autónoma de Manizales, Manizales 170001, Colombia
[2]Unidad Mixta de Imagen Biomédica FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitario y Biomédica de la Comunidad Valenciana, Valencia 46020, Spain
[3]Universidad Adolfo Ibáñez, Facultad de Ingeniería y Ciencias, Santiago, 7941169, Chile
[4]Center of Applied Ecology and Sustainability (CAPES), Santiago, 8331150, Chile
[5]Data Observatory Foundation, Santiago, 7941169, Chile
[6]Universidad de Caldas, 27985, Departamento de Sistemas e Informática, Manizales, Colombia

Corresponding author:
Alejandro Mora-Rubio[1]

Email address: alejandro.morar@autonoma.edu.co

ABSTRACT

Alzheimer's Disease (AD) is a progressive type of dementia characterized by loss of memory and other cognitive abilities, including speech. Since AD is a progressive disease, detection in the early stages is essential for the appropriate care of the patient throughout its development, going from asymptomatic to a stage known as Mild Cognitive Impairment (MCI), and then progressing to dementia and severe dementia; is worth mentioning that everyone suffers from cognitive impairment to some degree as we age, but the relevant task here is to identify which people are most likely to develop AD. Along with cognitive tests, evaluation of the brain morphology is the primary tool for AD diagnosis, where atrophy and loss of volume of the frontotemporal lobe are common features in patients who suffer from the disease. Regarding medical imaging techniques, Magnetic Resonance Imaging (MRI) scans are one of the methods used by specialists to assess brain morphology. Recently, with the rise of Deep Learning (DL) and its successful implementation in medical imaging applications, it is of growing interest in the research community to develop computer-aided diagnosis systems that can help physicians to detect this disease, especially in the early stages where macroscopic changes are not so easily identified. This paper presents a DL-based approach to classifying MRI scans in the different stages of AD, using a curated set of images from Alzheimer's Disease Neuroimaging Initiative and Open Access Series of Imaging Studies databases. Our methodology involves image pre-processing using FreeSurfer, spatial data-augmentation operations, such as rotation, flip, and random zoom during training, and state-of-the-art 3D Convolutional Neural Networks such as EfficientNet, DenseNet, and a custom siamese network, as well as the relatively new approach of Vision Transformer architecture. With this approach, the best detection percentage among all four architectures was around 89% for AD vs Control, 80% for Late MCI vs Control, 66% for MCI vs Control, and 67% for Early MCI vs Control.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu

INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurological disorder that primarily affects memory and cognition. Other symptoms, such as language disorders, hallucinations, and in some cases, seizures or parkinsonian features, may also be present (Kumar et al., 2022). One theory regarding the causes of AD links the β-amyloid peptide and mutations in its precursor proteins to weakened brain structures, which in turn impact various cognitive processes in the patient's brain. This theory is supported and assessed through two key markers: amyloid plaques and neurofibrillary tangles (Ballard et al., 2011). Macroscopically, changes in brain morphology play a crucial role in the diagnosis of AD, with frontotemporal lobe atrophy being a common feature. This loss of brain volume in the region results in diminished function of the amygdala and hippocampus, which are responsible for memory processes. Medical imaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans, are utilized by specialists to evaluate the morphological changes in the brain caused by AD (van Oostveen and de Lange, 2021). In addition to brain scans, using cognitive tests, like the "Mini-Mental State Examination" (MMSE), that consider specific physical, psychological, and social skills is essential for a reliable diagnosis of the disease.

Regarding AD progression, as the disease advances the symptoms change, and a series of consecutive stages can be identified, following the convention used by Alzheimer's Disease Neuroimaging Initiative (ADNI, https://adni.loni.usc.edu). A brief description of each stage is presented below.

- Early Mild Cognitive Impairment (EMCI): the affected person starts experiencing episodes of memory loss with words or the location of household items, nevertheless he can function independently and participate in social activities.

- Mild Cognitive Impairment (MCI): the affected people are susceptible to forgetting recent occurrences, becoming disoriented in their homes, and having difficulties with communication. This is often the longest stage of AD, lasting up to 4 years.

- Late Mild Cognitive Impairment (LMCI): at this stage of the disease, patients may need help with daily tasks, facing increasing difficulty communicating and controlling their movements. Their memory and cognitive skills worsen, and changes in behavior and personality may occur.

- Alzheimer's Disease (AD): as the disease progresses, the affected person requires increasing levels of attention and aid with daily tasks. This stage is characterized by growing unawareness of time and space, problems recognizing family and close friends, difficulty walking, and behavioral disturbances that may even lead to aggression.

Furthermore, with the rise of Deep Learning (DL) and its successful implementation in medical imaging applications, researchers have been working on automatic systems able to learn from MRI and PET scans to support the diagnosis of AD (LaMontagne et al., 2019; Petersen et al., 2010; Zeng et al., 2018). This kind of decision support system, created with artificial intelligence at its core, has been used successfully in different situations in the healthcare field improving diagnostic performance, reducing medical errors, and providing a better quality of service in developing countries (Currie and Rohren, 2021; Porumb et al., 2020). In general, DL techniques are developed to find patterns in the data and extract the relevant features for optimal classification (Sharma et al., 2016; Chan et al., 2020; Bravo-Ortiz et al., 2021), making them suitable for tasks with large volumes of labeled data based on human experiences, such as classifying the stages of AD using brain imaging (Liu et al., 2014; Chitradevi and Prabha, 2020; Zeng et al., 2021). In particular, for AD diagnosis, most of the research papers consider a 3D Convolutional Neural Network (CNN) approach to take advantage of the multidimensional features in brain scans (Payan and Montana, 2015; Huang et al., 2019). Alternatively, the work of Cheng and Liu (2017) proposes a 2D CNN approach, combined with recurrent neural networks to link the features of the three-dimensional scans. In the development of such systems, image pre-processing is key to achieving the best possible results, and in that regard, the main tool is FreeSurfer, which allows operations like skull suppression, bias field correction, anatomical registration, segmentation, reconstruction, and parcellation of the cortical surface (Fischl, 2012).

Acknowledging the growing number of patients with AD, and the importance of improving their quality of life, it is essential to develop more effective and efficient tools to diagnose the disease. This paper presents a DL-based approach to the classification of MRI scans in the different stages of AD,

using the EfficientNet, DenseNet, a custom siamese architecture, and a Vision Transformer, the detection percentage among all four architectures was around 89% for AD vs Control, 80% for Late MCI vs Control, 66% for MCI vs Control, and 67% for Early MCI vs Control, and it is organized as follows: a Methods section describing the dataset, CNN architectures, and training set-up; a Results section presenting the experiments and classification results obtained; a Discussion section going over the results and their impact on the field; and finally a Conclusion of this work.

METHODS

Dataset

For this task, open-access imaging databases ADNI (https://adni.loni.usc.edu) and OASIS (LaMontagne et al., 2019) are the most commonly used sources of MRI scans to train and evaluate DL models. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

For the purpose of this paper, a dataset comprising images from both databases was used. There are a total of 2,559 images from 1,126 subjects, where the gender distribution is relatively even with 54% of images from males and the remaining 46% from females; and ages ranging from 55 to 96 years old with a mean of 75 years old. Following the labels used by ADNI, there are a total of five different classes in the dataset: Cognitively Normal or Control (CN), EMCI, MCI, LMCI, and AD; OASIS labels, which are expressed numerically according to the *Clinical Dementia Rating*, were changed to ADNI format: label 0 corresponds to CN, (0,1) to MCI, and [1,3] to AD. Figure 1 shows the number of samples per class in the dataset.

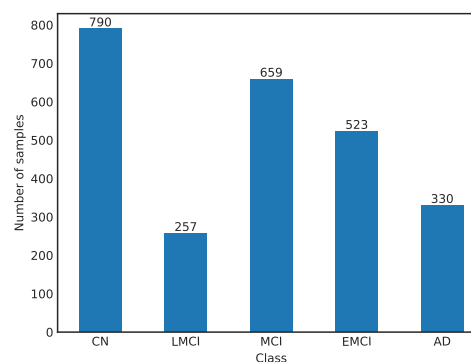



Figure 1. Class distribution in the dataset.

In order to reduce data-induced bias in the model, the images were partitioned into the training, validation, and test sets on a subject basis, obtaining a training set with 683 patients (1569 images), a validation set with 222 patients (513 images), and a test set with 221 patients (477 images), where every session from a patient is contained in a single set. Additionally, a similar age distribution within each partition was also maintained (see Figure 2)

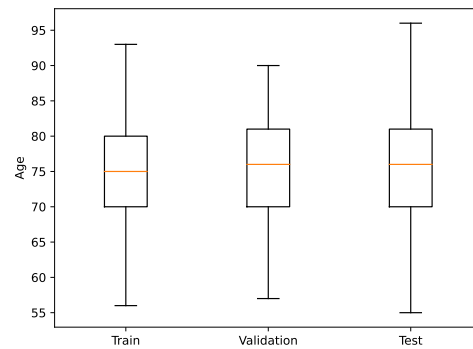



Figure 2. Age distribution in each dataset partition.

Deep Learning architectures

Siamese 3D

In this type of architecture, the data flows through parallel routes at the same time, which are then combined to make a final prediction. In this case, each path is composed of sets of 3D convolutional layers, 3D batch normalization, and average pooling operations, in the end, the feature maps are flattened, concatenated, and then passed to the fully connected layers for the final prediction. It is worth mentioning, that this is a custom architecture, designed for the classification of AD stages using 3D MRI scans (Torres, 2019). In the proposed configuration, this model has a total of 392,822 trainable parameters for the binary classification tasks, and 392,923 for the 3-way classification tasks. Figure 3 shows the model architecture.

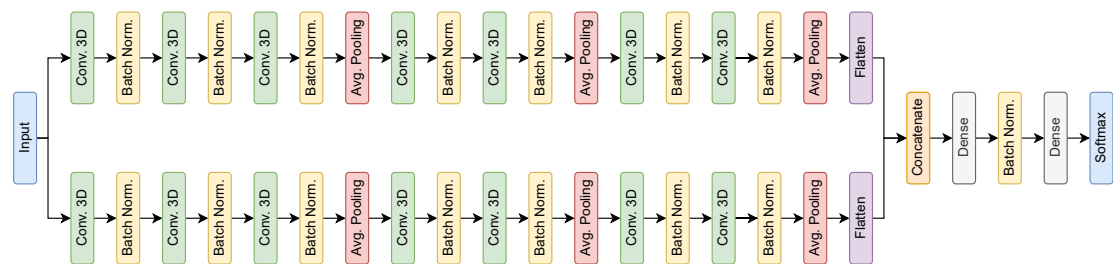



Figure 3. Siamese 3D model architecture.

Vision Transformer (ViT)

Vision transformers (ViT) (Dosovitskiy et al., 2020) are a groundbreaking approach to computer vision that has recently gained popularity due to their impressive performance on various tasks. Unlike traditional CNNs, which rely on hand-engineered feature maps, vision transformers use self-attention mechanisms to learn features directly from the input data dynamically. This approach has several advantages, including increased flexibility and scalability and improved performance on object detection, segmentation, and classification tasks. Additionally, ViT (Khan et al., 2022) has shown promise in domains beyond computer vision, including natural language processing and speech recognition. One of the critical innovations of ViT is self-attention, which allows the network to attend to different parts of the input data and dynamically adjust the importance of each feature. ViT (Khan et al., 2022) contrasts traditional CNNs, which rely on fixed, pre-defined feature maps that may only be optimal for some tasks. Another advantage of vision transformers is their ability to learn from large-scale datasets. Pre-training on massive datasets such as ImageNet has been shown to significantly improve the performance of ViT on downstream tasks, demonstrating the importance of unsupervised learning in deep learning models (Dosovitskiy et al., 2020). Despite their impressive performance, ViT is still a relatively new approach to computer vision, and there is ongoing research into how best to design and optimize these models. In the proposed configuration, this model has a total of 1,853,698 trainable parameters.

DenseNet

DenseNet was originally proposed as a general-purpose image classification architecture, where the dense convolutional blocks improved significantly the vanishing gradient problem experienced by very deep neural networks. Each of these dense convolutional blocks refers to a set of convolutional layers that are all connected between them, this means that "the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers" (Huang et al., 2016). By concatenating several of these blocks, the DenseNet architecture and all its variants are created. This network has been successfully used in different biomedical imaging classification problems, such as the detection of COVID-19 in Computed Tomography scans avoiding manual delineation of the lesions (Xiao et al., 2022), the detection of AD from brain scans based on features extracted from the hippocampal region (Cui and Liu, 2019), and the detection of skin cancer by classifying the images of skin lesions (Villa-Pulgarin et al., 2022), which motivate its inclusion in this work. In the proposed configuration, this model has a total of 11,244,674 trainable parameters for the binary classification tasks, and 11,245,699 for the 3-way classification tasks.

EfficientNet

EfficientNet refers to a series of convolutional neural networks that scale the three dimensions of the network at the same time, achieving better classification results with fewer trainable parameters compared to other state-of-the-art architectures. The dimensions affected are depth (number of layers), width (number of channels or filters per layer), and resolution (size of the input images and feature maps). In particular, the B7 variant was used for this paper, which is the biggest one as presented in the original paper (Tan and Le, 2019). As for DenseNet, this architecture has been implemented in various tasks of biomedical image processing such as the detection of diabetic retinopathy from digital fundus images (Jaiswal et al., 2021; Wang et al., 2020), the detection of tuberculosis in chest X-Ray images (Munadi et al., 2020), and aging classification based on skin microstructure characteristics (Moon and Lee, 2022). In the proposed configuration, this model has a total of 68,764,114 trainable parameters for the binary classification tasks, and 68,766,675 for the 3-way classification tasks.

Training procedure

Regarding the model training set-up, the MONAI(MONAI Consortium, 2022) was used, it is based on PyTorch and allows easy implementation of input data pipelines with image intensity scaling to the [0,1] range, image resizing to $91 \times 91 \times 91$ pixels, and spatial data-augmentation operations, such as random 90° rotation, random flip, and random zoom. The latter operations aim to improve the models' generalization ability by adding variability to the images during training. In terms of training hyperparameters, all models were trained for 250 epochs, using a weighted Cross Entropy Loss to account for data imbalance, and the Adadelta optimizer with a learning rate of 1.0, $\rho = 0.95$, and $\varepsilon = 1 \times 10^{-7}$. All the experiments were performed using NVIDIA QUADRO RTX 8000 graphic processor units.

RESULTS

This section presents the experiments performed over the previously described dataset, and the classification results achieved for each one of them. The first experiment consists of the binary classification between the Control cases and AD, which is the benchmark task in AD detection. Table 1 presents the training, validation, and test results achieved for this task, where the ViT architecture performed the best.

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.9844	0.8646	0.7448	0.6776	0.5943
ViT	0.7804	0.9203	***0.8902***	***0.8902***	***0.7401***
DenseNet	0.9656	0.8854	0.7448	0.6322	0.5905
EfficientNet B7	0.9672	0.8750	0.8542	0.8085	0.6835

Table 1. Classification results for 2 classes - Control and Alzheimer's Disease.

Figure 4a shows the Receiver Operator Curve for the CN vs. AD task using the predictions from the EfficientNet model.

The second experiment involves the binary classification between Control cases against LMCI. Table 2 presents the results for this experiment.

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.7656	0.8021	0.7578	0.5156	0.5156
ViT	0.7656	0.9070	*0.8056*	0.6870	0.6530
DenseNet	0.9984	0.8385	0.7031	*0.6991*	*0.6633*
EfficientNet B7	0.7375	0.8177	0.7734	0.5833	0.5833

Table 2. Classification results for 2 classes - Control and Late Mild Cognitive Impairment

Figure 4b shows the Receiver Operator Curve for the CN vs. LMCI task using the predictions from the DenseNet model.

Similarly, the third experiment consists of the classification of Control cases against MCI. Table 3 presents the results for this experiment.

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.8966	0.7305	0.6250	0.6017	0.6017
ViT	0.8506	0.7103	0.6432	0.5970	0.6420
DenseNet	0.9976	0.7188	*0.6641*	0.6449	*0.6633*
EfficientNet B7	0.9976	0.7383	*0.6641*	*0.6530*	0.6591

Table 3. Classification results for 2 classes - Control and Mild Cognitive Impairment

Figure 4c shows the Receiver Operator Curve for the CN vs. MCI task using the predictions from the DenseNet model.

The fourth experiment consists of the classification of Control and EMCI, which is significantly more challenging since the physical effect of the disease is not as aggressive as for AD, but ultimately it is more relevant for physicians since it would help them identify potential AD patients in the early stages. Table 4 presents the results for this experiment.

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.7331	0.7344	*0.6719*	*0.6749*	*0.6511*
ViT	0.7932	0.7020	0.6320	0.6720	0.6311
DenseNet	0.9596	0.7383	0.5938	0.6094	0.5913
EfficientNet B7	0.5807	0.6523	0.5042	0.4915	0.4640

Table 4. Classification results for 2 classes - Control and Early Mild Cognitive Impairment.

Figure 4d shows the Receiver Operator Curve for the CN vs. EMCI task using the predictions from the Siamese 3D model.

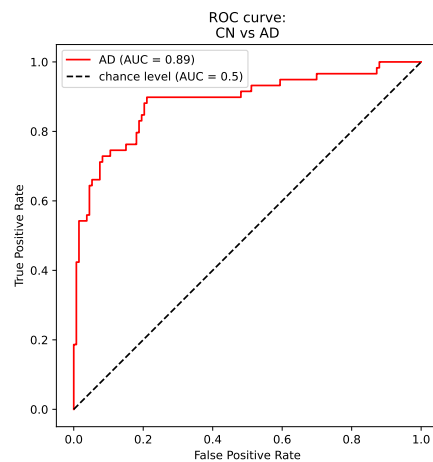
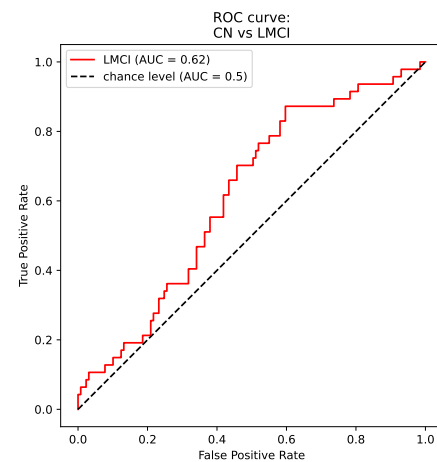
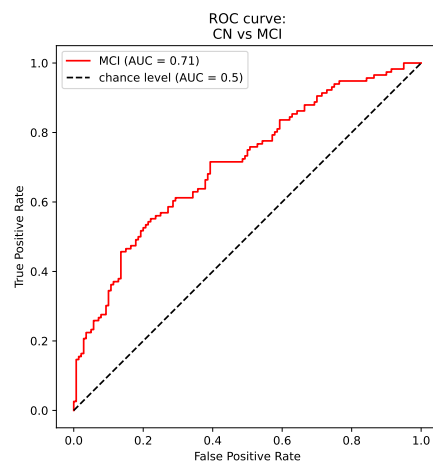
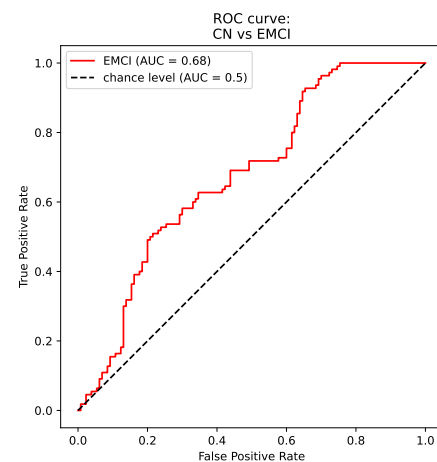
The fifth experiment aimed to classify the early and late stages of the disease progression; the early stages were the Control and EMCI cases, and MCI, LMCI, and AD were considered as late stages. The main goal was to take advantage of the complete dataset aiming to improve classification results. Table 5 presents the results for the fourth experiment.

(a) Control and Alzheimer's Disease.

(b) Control and Late Mild Cognitive Impairment.

(c) Control and Mild Cognitive Impairment.

(d) Control and Early Mild Cognitive Impairment.

Figure 4. ROC curves

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.8542	0.6875	0.6473	0.6292	0.6292
ViT	0.7920	0.6990	0.6540	0.6560	0.6520
DenseNet	1.0000	0.7051	0.6540	0.6505	0.6505
EfficientNet B7	0.7760	0.7246	*0.6652*	*0.6550*	*0.6550*

Table 5. Classification results for 2 classes - Early and late stages

The sixth and final experiment consists of a multi-class classification problem involving Control, MCI (including early and late MCI), and AD cases. Table 6 presents the results of this experiment.

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Test Sensitivity	Test Specificity
Siamese 3D	0.6250	0.5645	0.5625	0.3585	0.6664
ViT	0.6120	0.5780	0.5510	*0.4260*	0.6650
DenseNet	0.5964	0.5820	0.5446	0.4199	*0.6861*
EfficientNet B7	0.5443	0.5703	*0.5893*	0.4163	0.6765

Table 6. Classification results for 3 classes - Control, Mild Cognitive Impairment, and Alzheimer's Disease

DISCUSSION

This paper presents a DL-based approach to the classification of MRI scans in the different stages of AD, using a curated set of images from both ADNI and OASIS databases. There are studies from different countries around the world that suggest an increasing prevalence of dementia in older patients, and how AD can be considered one of the leading causes of it (Rasmussen and Langerman, 2019), it is important for the scientific and healthcare community to develop and improve current diagnostic techniques, in particular during its early stages. Since there is yet to be a treatment that can revert the physiological effects of severe dementia, the early diagnosis of the disease is imperative for appropriate care of the patient, and a realistic projection of the following years and stages. In this regard, brain imaging and CAD systems are of high interest to researchers, where the high capacity of DL algorithms to extract relevant features from images can be leveraged to detect the subtle changes present in the early stages of the disease.

Our methodology encompasses image pre-processing using FreeSurfer, spatial data augmentation techniques such as rotation, flip, and random zoom during training, and state-of-the-art 3D CNNs, including EfficientNet, DenseNet, and a custom siamese network, as well as the relatively new approach of Vision Transformer architecture. Using this approach, the best detection percentages among all three architectures were approximately 89% for AD vs. Control, 80% for Late MCI vs. Control, 66% for MCI vs. Control, and 67% for Early MCI vs. Control. It is noteworthy that all three models involved in the experiments achieve competitive results and outperform one another in specific cases, with the Siamese 3D model's performance on the CN vs. Early MCI task and the ViT model's performance on the CN vs. AD task being particularly remarkable. Regarding these results, it is essential to acknowledge the difficulty in comparing them with similar research papers due to the numerous factors affecting model performance. If these factors are not adequately reported, reproducing the results and directly comparing them with ours becomes nearly impossible. Consequently, the dataset and selected partition are crucial for obtaining reliable results that accurately represent how the model would perform in real-life scenarios. As reported in medical tasks, a significant portion of predictive models presented in the scientific literature offer little to no value to the healthcare community due to the presence of bias in the models, which affects the accuracy of predictions (Roberts et al., 2021; Wynants et al., 2020). Bearing this in mind, the code and references to the data used during the experimentation stage of our work are available in a GitHub repository (https://github.com/MoraRubio/alzheimer-stages-dl), and this paper describes, to the best of our ability, the materials and methods used throughout the process.

In particular, compared to the referenced works on classification using 3D CNNs, The work by Huang et al. (2019) also includes FDG-PET images, which offer relevant information for the diagnosis, but at a

246 greater cost for the patient or healthcare system, resulting in less viable option for developing countries
247 with lower resources, as well as, fewer images for model training and evaluation. On the other hand, the
248 work by Payan and Montana (2015), uses an autoencoder for feature extraction instead of feeding the
249 images directly to the CNN, which adds extra complexity for model development and deployment. In this
250 work, only MRI images are considered, since they are more common and accessible than PET images,
251 and the images are directly fed to the CNN architectures. As mentioned above, a direct comparison of the
252 detection percentages is not easily performed.

253 Acknowledging the limited size of the involved dataset, and the number of DL methods; future work
254 should be aimed at exploring different network architectures, such as Convolutional Vision Transformers,
255 to enhance the detection accuracy of the different stages. Initially, this will focus on binary classification
256 to differentiate intermediate stages from Control and AD patients. However, the ultimate goal is to identify
257 those MCI patients who are most likely to develop AD, particularly in the early stages, where prediction
258 is crucial for healthcare professionals in providing optimal patient care. Moreover, assembling a larger
259 image dataset while adhering to ethical principles (Tabares-Soto et al., 2022) and aiming to minimize
260 data-induced bias in the models can positively impact the detection capabilities and, more importantly, the
261 generalization ability of the models, which is essential for developing viable CAD systems. Furthermore,
262 it is vital to conduct in-depth analysis of the results using visualization or interpretation techniques, such
263 as activation maps, occlusion sensitivity, or gradient-based heat maps. These methods could enable
264 comparison with physical markers assessed by medical specialists or even identify new regions of interest
265 for AD detection.

CONCLUSIONS

267 Early diagnosis of AD is key to providing adequate care to the patient, but existing detection techniques
268 are not definitive enough to provide a certain diagnosis. Evaluation of brain morphology and its changes
269 through time is one of the tools used by specialists to diagnose the disease, however, changes in the early
270 stages are difficult to identify in plain sight, and the use of CAD systems is of great interest for this task.
271 The results presented in this paper regarding the classification of different stages of AD using brain MRI
272 indicate that DL models do have the ability to correctly identify control/healthy patients from the ones
273 suffering from dementia, nevertheless, there is still a lot of work to be done in order to create reliable
274 models and test them appropriately in real-life scenarios.

REFERENCES

276 Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., and Jones, E. (2011). Alzheimer's disease.
277   *Lancet*, 377(9770):1019–1031.
278 Bravo-Ortiz, M. A., Arteaga-Arteaga, H. B., Tabares-Soto, R., Padilla-Buritica, J. I., and Orozco-Arias, S.
279   (2021). Cervical cancer classification using convolutional neural networks, transfer learning and data
280   augmentation. *Revista EIA*, 18(35):100–111.
281 Chan, H.-P., Samala, R. K., Hadjiiski, L. M., and Zhou, C. (2020). Deep learning in medical image
282   analysis. *Deep Learning in Medical Image Analysis*, pages 3–21.
283 Cheng, D. and Liu, M. (2017). Combining convolutional and recurrent neural networks for alzheimer's
284   disease diagnosis using pet images. In *2017 IEEE International Conference on Imaging Systems and
285   Techniques (IST)*, pages 1–5. IEEE.
286 Chitradevi, D. and Prabha, S. (2020). Analysis of brain sub regions using optimization techniques and
287   deep learning method in alzheimer disease. *Applied Soft Computing*, 86:105857.
288 Cui, R. and Liu, M. (2019). Hippocampus analysis by combination of 3-d densenet and shapes for
289   alzheimer's disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 23:2099–2107.
290 Currie, G. and Rohren, E. (2021). Intelligent imaging in nuclear medicine: the principles of artificial
291   intelligence, machine learning and deep learning. In *Seminars in Nuclear Medicine*, volume 51, pages
292   102–111. Elsevier.
293 Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M.,
294   Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An image is worth 16x16
295   words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
296 Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2):774–781.

297 Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2016). Densely connected convolutional
298   networks. *arXiv*.

299 Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X., and (ADNI), A. D. N. I. (2019). Diagnosis of
300   alzheimer's disease via multi-modality 3d convolutional neural network. *Frontiers in neuroscience*,
301   13:509.

302 Jaiswal, A. K., Tiwari, P., Kumar, S., Al-Rakhami, M. S., Alrashoud, M., and Ghoneim, A. (2021). Deep
303   learning-based smart iot health system for blindness detection using retina images. *IEEE Access*,
304   9:70606–70615.

305 Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision:
306   A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41.

307 Kumar, A., Sidhu, J., Goyal, A., and Tsao, J. W. (2022). Alzheimer disease. *StatPearls*, pages 1–27.

308 LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E.,
309   Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., and Marcus, D. (2019).
310   Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer
311   disease. *MedRxiv*.

312 Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., and Feng, D. (2014). Early diagnosis of alzheimer's disease
313   with deep learning. In *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, pages
314   1015–1018. IEEE.

315 MONAI Consortium (2022). Monai: Medical open network for ai.

316 Moon, C. I. and Lee, O. (2022). Skin microstructure segmentation and aging classification using cnn-based
317   models. *IEEE Access*, 10:4948–4956.

318 Munadi, K., Muchtar, K., Maulina, N., and Pradhan, B. (2020). Image enhancement for tuberculosis
319   detection using deep learning. *IEEE Access*, 8:217897–217907.

320 Payan, A. and Montana, G. (2015). Predicting alzheimer's disease: a neuroimaging study with 3d
321   convolutional neural networks. *arXiv preprint arXiv:1502.02506*.

322 Petersen, R. C., Aisen, P., Beckett, L. A., Donohue, M., Gamst, A., Harvey, D. J., Jack, C., Jagust,
323   W., Shaw, L., Toga, A., Trojanowski, J., and Weiner, M. (2010). Alzheimer's disease neuroimaging
324   initiative (adni): clinical characterization. *Neurology*, 74(3):201–209.

325 Porumb, M., Stranges, S., Pescapè, A., and Pecchia, L. (2020). Precision medicine and artificial
326   intelligence: a pilot study on deep learning for hypoglycemic events detection based on ecg. *Scientific
327   reports*, 10(1):1–16.

328 Rasmussen, J. and Langerman, H. (2019). Alzheimer's disease – why we need early diagnosis. *Degenera-
329   tive Neurological and Neuromuscular Disease*.

330 Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C.,
331   McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., and
332   Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect
333   and prognosticate for covid-19 using chest radiographs and ct scans. *Nature Machine Intelligence*.

334 Sharma, S., Sharma, V., and Sharma, A. (2016). Performance based evaluation of various machine learning
335   classification techniques for chronic kidney disease diagnosis. *arXiv preprint arXiv:1606.09581*.

336 Tabares-Soto, R., Bernal-Salcedo, J., García-Arias, Z. N., Ortega-Bolaños, R., Hermosilla, M. P., Arteaga-
337   Arteaga, H. B., and Ruz, G. A. (2022). Analysis of ethical development for public policies in the
338   acquisition of ai-based systems. *IGI Global*, pages 184–212.

339 Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks.
340   *arXiv*.

341 Torres, J. M. S. (2019). Análisis poblacional de patología de alzheimer a partir de imágenes de resonancia
342   magnética.

343 van Oostveen, W. M. and de Lange, E. C. (2021). Imaging techniques in alzheimer's disease: A review
344   of applications in early diagnosis and longitudinal monitoring. *International Journal of Molecular
345   Sciences*, 22:1–34.

346 Villa-Pulgarin, J., Ruales-Torres, A., Arias-Garzón, D., Bravo-Ortiz, M., Arteaga-Arteaga, H., Mora-
347   Rubio, A., Alzate-Grisales, J., Mercado-Ruiz, E., Hassaballah, M., Orozco-Arias, S., Cardona-Morales,
348   O., and Tabares-Soto, R. (2022). Optimized convolutional neural network models for skin lesion
349   classification. *Computers, Materials and Continua*, 70.

350 Wang, J., Yang, L., Huo, Z., He, W., and Luo, J. (2020). Multi-label classification of fundus images with
351   efficientnet. *IEEE Access*, 8:212499–212508.

Wynants, L., Calster, B. V., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M., Damen, J. A., Debray, T. P., Vos, M. D., Dhiman, P., Haller, M. C., Harhay, M. O., Henckaerts, L., Kreuzberger, N., Lohmann, A., Luijken, K., Ma, J., Navarro, C. L. A., Reitsma, J. B., Sergeant, J. C., Shi, C., Skoetz, N., Smits, L. J., Snell, K. I., Sperrin, M., Spijker, R., Steyerberg, E. W., Takada, T., Kuijk, S. M. V., Royen, F. S. V., Wallisch, C., Hooft, L., Moons, K. G., and Smeden, M. V. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *British Medical Journal*.

Xiao, B., Yang, Z., Qiu, X., Xiao, J., Wang, G., Zeng, W., Li, W., Nian, Y., and Chen, W. (2022). Pam-densenet: A deep convolutional neural network for computer-aided covid-19 diagnosis. *IEEE Transactions on Cybernetics*, 52:12163–12174.

Zeng, N., Li, H., and Peng, Y. (2021). A new deep belief network-based multi-task learning for diagnosis of alzheimer's disease. *Neural Computing and Applications*, pages 1–12.

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H., and Li, Y. (2018). A new switching-delayed-pso-based optimized svm algorithm for diagnosis of alzheimer's disease. *Neurocomputing*, 320:195–202.