ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Comparison of Convolutional Neural Networks Approaches Applied to the Diagnosis of Alzheimer's Disease

Luiza Scapinello Aquino da Silva[1] Leonardo Alexandre de Geus[2]
Viviana Cocco Mariani[3] and Leandro dos Santos Coelho[1] *

1 - Graduate Program in Electrical Engineering (PPGEE)
Federal University of Parana (UFPR), Curitiba, PR, Brazil

2 - Undergraduate Program in Mechatronics Engineering
Pontifical Catholic University of Parana, Curitiba, PR, Brazil

3 - Graduate Program in Mechanical Engineering (PGMec)
Federal University of Parana (UFPR), Curitiba, PR, Brazil

Abstract. Alzheimer's disease (AD), a neurodegenerative disorder, progressively impairs memory and cognitive functions. Magnetic resonance imaging (MRI) is used as AD diagnosis and progress monitoring method. Convolutional Neural Network (CNN) is a data-driven deep learning model containing layers transforming data input using convolution filters. The goal of this paper is to present an analysis of the CNN architectures for classifying AD diagnoses using functional brain MRI scans acquired by the experimental dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Results show CNNs variants such as InceptionV3 and InceptionResNetV2 as powerful computational tools for developing predictive neuroimaging biomarkers in AD diagnosis applications, with accuracy above 70%.

1 Introduction

Alzheimer's disease (AD) is an irreversible, progressive, degenerative brain condition that causes neurons to die. At first, it leads to memory loss and decreases thinking skills, ultimately damaging the patient's ability to fulfill simple tasks [1]. There is no universal test for Alzheimer's diagnosis, meaning doctors use a combination of techniques, for instance, the patient's family history, and cognitive and psychological tests. Nevertheless, since these methods are vulnerable to human errors, they may lead to erroneous conclusions and postpone the treatment [2].

The functional Magnetic Resonance Imaging (fMRI) is a technique that allows for the knowledge of the brain's activities through the change of blood

*da Silva would like to thank Coordination of Higher Education Personnel Improvement (CAPES) for its financial support. The authors Mariani and Coelho thank the National Council of Scientific and Technologic Development of Brazil – CNPq (Grants: 314389/2023-7-PQ, 313169/2023-3-PQ, 407453/2023-7-Universal, and 442176/2023-6-Peci).

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

flow, permitting the perception of which part of the brain is responsible for the diverse functions of the body [3]. Additionally, fMRI and magnetic resonance imaging (MRI), non-invasive and less susceptible to human error, are the most used methods for diagnosing AD.

Since brain function networks are comparatively constant among different healthy patients, biomarkers of neural connectivity can be used to predict diseases such as AD [4]. In 2013, Suk and Shen [5] developed a classifier based on a support vector machine and a stacked auto-encoder network to extract low- to mid-level features from images to classify AD stages. The accuracy of the AD/NC (Normal Controls) classification was 95.9% using MRI and Positron Emission Tomography images.

Payan and Montana [6] obtained an accuracy rate of 95.39% in classifying the patient's stage using an auto-encoder with a 3D convolutional neural network (CNN). In 2016, Sarraf and Tofighi [7] classified, with an accuracy testing of 96.85%, the fMRI data of AD subjects from NC using a CNN and the LeNet-5 (Yann LeCun network) architecture. Sarraf and Tofighi [8] also proposed and implemented pipelines, which resulted in an accuracy rate of 99.9% for fMRI pipelines. In 2018, Liu et al. [9] proposed a classification outline based on CNN and Bidirectional Gated Recurrent Units together to capture the features of 3D PET (Positron Emission Tomography) images for AD diagnosis.

In this study, an in-depth analysis of seven CNN architectures, including Visual Geometry Group with 16 and 19 layers (VGG16 and VGG19, respectively), Residual neural network with 50 layers (ResNet50), Inception architecture with residual connections (InceptionV3 and InceptionResNetV2), extreme version of Inception (Xception), and MobileNet are proposed for neuroimaging recognition for the diagnosis of AD. The CNN architectures were trained using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to classify brain images as healthy or indicating signs of dementia.

The remainder of this study is organized with a case study called ADNI-1, discussed in Section 2. Section 3 describes CNN architectures and fundaments. In Section 4, the results and analysis of the data classification obtained are presented. Finally, the conclusions and future research implications are outlined in Section 5.

2 Case Study: ADNI-1 and ADNI Screening

The dataset ADNI-1 has neuroimaging data, biological markers, and clinical and neuropsychological evaluations to detect the progression of mild cognitive impairment (MCI) and early AD. The neuroimaging data are obtained from fMRI and MRI, presenting information describing neural tissues' shape, size, and integrity. It is possible to use the ADNI database to relate the patterns of MRI and fMRI images to the AD diagnosis [10] to assist the patient, leading to a possible improvement in life quality.

The ADNI database consists of exams from 55 to 95-year-old patients, with the majority of them aged approximately 75 years old, who were classified into

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

three different groups: individuals diagnosed with AD, people with a progression of MCI, and healthy examples. Four stages of cumulative data capture were made. The number of patients and the time they were involved may be observed as described in Table 1. There are 244 subjects randomly selected from the ADNI-1 and ADNI Screening programs, 124 women and 120 men. Among them, 102 patients were classified with AD, and 142 were labeled as healthy subjects.

Table 1: ADNI program

	ADNI-1	ADNI-GO	ADNI-2	ADNI-3
Start date	Oct/2014	Sept/2009	Sept/2011	Sept/2016
Healthy	200	200	350	483
MCI	400	600	1000	1000
AD patients	200	200	350	350

All fMRI neuroimage files were transferred through the ADNI, with each patient having 100 images. Next, the images were conjoined to a type representing a brain in two dimensions using $dcm2niix$. Further format conversion was also needed since the trained CNNs used two-dimensional images at network entrances, so the $med2imag$ library was used to separate a slice from the patients' brains. The images were then resized and normalized to 256x256 pixels. Finally, a file type was generated containing various pieces of information regarding the instances, such as age, gender, the name of the images of the subject, and the target label (class) for the task of classification.

3 CNN Architectures

The CNNs are biologically inspired by Hubel and Wiesel's [11] early work on the cat's visual cortex. A standard feed-forward CNN consists of layers described as: convolutional, pooling, and fully connected. The convolutional layer consists of a set of kernels (filters) that are convolved over the height and width of the input image in the case of two-dimensional images. In this layer, each neuron is locally connected to parts of the neurons in the previous layers. Based on this process, the network learns filters that activate when they detect a visual feature, which produces a separate two-dimensional activation map [12].

After several convolutional and pooling layers and a normalizing layer, the high-level reasoning in the CNN is done via fully connected layers, which take all the neurons from the previous layer and connect them to every single one of its neurons [13], capturing correlations between different features previously produced. Moreover, a possible last layer of a CNN is an output layer, and for cases of classification tasks, the softmax operator is commonly used.

The main CNN architectures include the VGG16 and VGG19 networks [14], ResNet50 [15], InceptionV3 and its InceptionResNetV2 variant [13], Xception [16], and MobileNet [17]. Transfer Learning was used to simplify the training of neural networks. The method uses the knowledge of a pre-trained neural network through a more complete database, increasing the generalization in the classification task of another neural network, to reduce the training time

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

and computational power required. Then the transfer of the weights trained in the ImageNet image bank was loaded, which contains about 150,000 images in a thousand classes and is used as a benchmark for the DL algorithms. For parameter hypertuning the approach used was Bayesian optimization.

All the evaluated CNN models were trained using Elastic Compute Cloud machines. The k-fold cross-validation method was used for the training and the creation of the results. The k-fold cross-validation consists of evaluating and comparing learning algorithms that divide all the data into $k > 1 \in \mathbb{Z}$ sets. The process occurs when the data are randomly selected and divided into k groups; one group is selected for testing and the others for training. The more optimized value for k depends on the amount of data for training, and a poor choice may result in an unsatisfactory representation of the skills of the model. Performance indicators such as accuracy, precision, recall, F1 score, and Matthew's correlation coefficient (MCC) are applied to evaluate CNN. These indicators are given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

where (TP) True Positives is the number of correct predictions for the positive class, (TN) True Negatives is the number of correct predictions for the negative class, (FP) False Positives is the number of incorrect predictions for the positive class, and (FN) False Negatives is the number of incorrect predictions for the negative class.

4 Results and Discussion

All images were split randomly from the ADNI1 database: Screening (SC), among the 224 selected patients, there were 168 placed in the training group and about 58 in the test group. The adopted ratio for dividing data over the test and training set was 75%. After the training process, the same results were obtained using the statistical metrics. The results are presented in Table 2. The values shown are the averages of the cross-validation of 10 folds, while the standard deviation was inserted next to each value to give more consistency to the results. The set of tests is evaluated with many metrics: accuracy, precision, recall, F1 score, and MCC. The best results in Table 2 are presented in bold.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Table 2: Results of the classification performance metrics

Network	Accuracy	Precision	Recall	F1 Score	MCC
VGG16	0.942 ± 0.02	0.924 ± 0.05	0.992 ± 0.04	0.922 ± 0.03	0.877 ± 0.05
VGG19	0.918 ± 0.04	0.896 ± 0.06	0.886 ± 0.07	0.889 ± 0.06	0.827 ± 0.09
ResNet50	0.799 ± 0.04	0.796 ± 0.08	0.623 ± 0.08	0.696 ± 0.07	0.562 ± 0.09
InceptionV3	**0.989 ± 0.01**	0.986 ± 0.02	**0.984 ± 0.01**	**0.985 ± 0.01**	**0.977 ± 0.02**
InceptionResNetV2	0.983 ± 0.01	**0.988 ± 0.02**	0.966 ± 0.02	0.976 ± 0.01	0.963 ± 0.02
Xception	0.890 ± 0.02	0.876 ± 0.04	0.823 ± 0.04	0.847 ± 0.03	0.764 ± 0.05
MobileNet	0.953 ± 0.02	0.939 ± 0.04	0.935 ± 0.03	0.937 ± 0.02	0.901 ± 0.04

The results using InceptionV3 had the best accuracy, with high mean values and low standard deviation. InceptionResNetV2 and VGG19 presented a promising performance in terms of the accuracy measure. MobileNet presented a median accuracy near 95%. This result was close to that of VGG16, which had a median close to 94%. The classification in this study had a high computational cost because the architectures had 143 million parameters to be adjusted.

The Friedman test [18] is a non-parametric statistical test that detects significant treatment differences across multiple test subjects. After calculating the Friedman test for the results in Table 2 the achieved p-value was 0.00011. Since this value is less than 0.05, it is possible to reject the null hypothesis that the metric mean is the same for all seven models. Showing sufficient evidence to conclude that the type of model used leads to statistically significant differences in each metric.

5 Conclusion and Future Research

Several CNN architectures were compared using statistical metrics to classify AD through fMRI images. The results achieved the expected objectives, as the high accuracy values found after the training of the CNN models proved the task of classifying Alzheimer's and non-Alzheimer's cases using fMRI images as a database to be feasible. All seven architectures produced good performances regarding the task, with the algorithms that presented the best results in terms of accuracy being the InceptionV3 and InceptionResNetV2 networks, with 98.93% and 98.30% mean accuracy, respectively.

In terms of future work, further analysis will be done using additional datasets for higher comparison, such as the Capsule networks [19] and NASNet (Neural Architecture Search Network) [20], which have previously shown promise to be better than the Inception networks.

References

[1] David S Knopman, Helene Amieva, Ronald C Petersen, Gäel Chételat, David M Holtzman, Bradley T Hyman, Ralph A Nixon, and David T Jones. Alzheimer disease. *Nature reviews Disease primers*, 7(1):33, 2021.

[2] Rudy J Castellani, Raj K Rolston, and Mark A Smith. Alzheimer disease. *Disease-a-month: DM*, 56(9):484, 2010.

[3] Hassaan Tohid. Advancements in the field of neuroscience. *International Journal of Neurology Brain Disorders*, 3(1), 2016.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

[4] Seung-Jun Kim, Vince D Calhoun, and Tülay Adalı. Flexible large-scale fmri analysis: A survey. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA*, pages 6319–6323, 2017.

[5] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for ad/mci classification. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI: 16th International Conference, Nagoya, Japan, Proceedings, Part II 16*, pages 583–590. Springer, 2013.

[6] Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.

[7] Saman Sarraf, Danielle D DeSouza, John Anderson, Ghassem Tofighi, and Alzheimer's Disease Neuroimaging Initiative. Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri. *BioRxiv*, page 070441, 2016.

[8] Saman Sarraf and Ghassem Tofighi. Deep learning-based pipeline to recognize alzheimer's disease using fmri data. In *Future Technologies Conference (FTC), San Francisco, CA, USA*, pages 816–820. IEEE, 2016.

[9] Manhua Liu, Danni Cheng, Weiwu Yan, and Alzheimer's Disease Neuroimaging Initiative. Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. *Frontiers in Neuroinformatics*, 12:35, 2018.

[10] Feng Li, Loc Tran, Kim-Han Thung, Shuiwang Ji, Dinggang Shen, and Jiang Li. A robust deep model for improved classification of ad/mci patients. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1610–1616, 2015.

[11] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106, 1962.

[12] Enrique Garcia-Ceja, Md Zia Uddin, and Jim Torresen. Classification of recurrence plots' distance matrices with a convolutional neural network for activity recognition. *Procedia Computer Science*, 130:157–163, 2018.

[13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA*, pages 2818–2826, 2016.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA*, pages 1251–1258, 2017.

[17] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[18] Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, 1(4):221–228, 1996.

[19] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 30, 2017.

[20] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA*, pages 8697–8710, 2018.