An Explainable Alzheimer's Disease Prediction Using EfficientNet-B7 Convolutional Neural Network Architecture

Chapter · June 2023

DOI: 10.1007/978-981-19-8032-9_53

2 authors:

Sobhana Jahan
Bangladesh University of Professionals

M. Shamim Kaiser
Jahangirnagar University

An Explainable Alzheimer's Disease Prediction Using EfficientNet-B7 Convolutional Neural Network Architecture

Sobhana Jahan[1[0000−0003−4882−4327]] and M Shamim Kaiser[2[0000−0002−4604−5461]]

[1] Information and Communication Technology, Bangladesh University of Professionals, Dhaka 1216, Bangladesh
sobhanajahanlisa@gmail.com

[2] Institute of Information Technology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh
msKaiser@juniv.edu

Abstract. Alzheimer's disease is a neurocognitive disease that results from the brain shrinking and brain tissue dying over time. It gradually erodes memory, thinking skills, and the ability to carry out the most basic tasks. The use of an MRI to evaluate brain atrophy is thought to be a reliable way to diagnose and track the progression of Alzheimer's disease. In such studies, deep learning architecture provides outstanding results. One drawback of deep learning is that it necessitates a large number of datasets to train the model. Another drawback is the black-box nature and due to this nature, doctors, patients, and the general public are doubtful of the model's prediction results. To remove these problems, this study proposes a novel Gradient-weighted Class Activation Mapping (Grad-CAM) based explanation of Alzheimer's disease prediction, using the EfficientNet-B7 convolutional neural network architecture. Here, the data augmentation technique is used to make the small dataset suitable for the model. In this work, we have performed a classification between different stages of Alzheimer's disease which are Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). Using the base dataset and augmented dataset, the achieved accuracy for four-class classification is 73.47% and 91.76%, respectively. The resulting accuracy for AD vs CN, CN vs EMCI, CN vs LMCI, AD vs EMCI, AD vs LMCI, and EMCI vs LMCI using augmented dataset are 97.72%, 97.70%, 98.01%, 95.12%, 96.70%, and 96.40%, respectively.

Keywords: Deep Learning · Transfer Learning · Explainable AI · Data Augmentation · Gradient-weighted Class Activation Mapping.

1 Introduction

Alzheimer's disease (AD) is a neurological disorder in which memory loss and cognitive decline are caused due to the death of brain cells. AD and other dementias are thought to affect at least 50 million people worldwide. As Asia is the world's most populous continent, a report estimates that the number of dementia patients in this region will reach nearly 71 million by 2050 [1]. In Western societies, AD is the most common cause of dementia, and around 5.5 million people in the United States are affected by it [2]. There is currently no cure for Alzheimer's disease, but there are medications available to slow the progression of the disease.

An early Alzheimer's diagnosis increases the chances of being benefitted from treatment. It also creates the opportunity for future care and treatment. For early diagnosis of AD, it is necessary to identify those patients who are at Mild Cognitive Impairment (MCI) stage. Researchers discovered that people with MCI are more likely than those without it to develop AD or related dementia. Schmidtke et al. [3] found that among all MCI patients, 44% had converted to AD after a mean delay of 19 months.

For confirming an Alzheimer's diagnosis or ruling out some other potential causes, medical professionals also use brain scans such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), or Positron Emission Tomography (PET). As MRI is one of the most powerful neuroimaging techniques used in deep learning-based AD prediction models, in this work we have used MRI data collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). From the ADNI cohort, MRI images of a total of four types of participants are collected and those types are Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and AD.

Since 2013, deep learning has gotten a lot of attention in Alzheimer's disease detection research. When compared to general machine learning techniques, deep models are more accurate in detecting Alzheimer's disease [4]. Transfer learning is a previously trained deep learning model for representing features of a new dataset. A pre-trained model is typically trained on a large dataset like ImageNet, and the weights acquired from the trained model can be used with one custom neural network for other applications. So, it is commonly used to save time and resources. EfficientNet is a convolutional neural network that uses a compound coefficient to evenly scale all width, resolution, depth, or dimensions.

Here in this research work, the EfficientNet-B7 model is used for training and preprocessing the input data. As the EfficientNet-B7 is a Convolutional Neural Network (CNN) model, it requires a huge amount of data for providing a good result. This work performed data augmentation to remove the problem of data size. In data analysis, data augmentation refers to techniques for increasing the amount of data by trying to add slightly altered copies of existing data or creating new synthetic data from existing data.

Due to the black-box nature of deep learning models, general people are unable to know the reasons behind the model's prediction. That is why doctors and patients do not believe in the deep learning models' prediction. To overcome this issue, this work has proposed the Gradient-weighted Class Activation Mapping (Grad-CAM) technique to visualize the decision-making parts of the MRI image.

The main contributions of this work are:

1. We have proposed a data augmentation technique for increasing the size of the training and testing dataset.
2. We have presented the EfficientNet-B7 model to classify the stages of AD.
3. We have used Grad-CAM to convert the black-box decisions into explainable ones so that anyone can understand the reasons for any prediction.

The rest of the paper is structured as follows: Section 2 includes related works, Section 3 contains the proposed methodology, Section 4 comes up with performance comparisons, and Section 5 includes the concluding remarks.

2 Related Works

Nanni et al. [5] presented a comparison study between ensemble SVM and ensemble transfer learning methods for predicting Alzheimer's in the early stage using MRI. It was observed that the ensemble SVM model provides better AUC than ensemble transfer learning for AD vs CN and CN vs MCIc prediction, which is 93.2% and 90.6%, respectively. For MCIc vs MCInc prediction ensemble transfer learning was better performing than machine learning and that value is 70.6%. By using a multi-slice and multi-model ensemble approach, Kang et al. [6] proposed an AD prediction model with the help of CNNs. When categorizing MCI vs AD, CN vs MCI, and CN vs AD; this ensemble method obtained accuracy values of 77.19%, 72.36%, and 90.36%, respectively.

Wee et al. [7] used a spectral graph CNN with cortical thickness and geometry to detect MCI and AD. The graph-CNN had shown accuracy rate of AD vs CN at 85.8%, AD vs EMCI at 79.2%, CN vs LMCI 69.3%, EMCI vs LMCI 60.9%, LMCI vs. AD 65.2%, and EMCI vs CN 51.8% for the ADNI-2 dataset. Mehmood et al. [8] separated EMCI, CN, LMCI, and AD using the VGG architecture. The best accuracy was 98.73% for AD vs CN. For LMCI vs EMCI patients, evaluated the accuracy was 83.72%.

Using MRI data, Ji et al. [9] proposed an early diagnosis of Alzheimer's disease using CNN. In that paper, three base ConvNets were used and these three classifiers are eNASNet, eResNet-50, and eMobileNet. Using the ADNI dataset, the eNASNet model had achieved a sensitivity of 80.56%, a specificity of 94.00%, an AUC of 0.96%, and an accuracy of 88.37% for CN vs MCI estimation. Naz et al. [10] used the AlexNet, VGG, Inceptionv3, GoogLeNet, ResNet, DenseNet, and InceptionResNet architectures to detect AD. Using the ADNI dataset, VGG 16 achieved 97.06% and 98.89% accuracy for MCI vs CN and AD vs CN, respectively. VGG19 achieved the highest accuracy of 99.27% for MCI vs AD.

3 Proposed Methodology

3.1 Dataset Acquisition

The Alzheimer Disease Neuroimaging Initiative-ADNI data set is the gold standard for Alzheimer disease detection [11]. ADNI was established in 2003 under the direction of Dr. Michael W. Weiner and is funded through public-private cooperation. The ADNI is a multisite, longitudinal study aimed at developing clinical, genetic, imaging, and biomarkers for the early detection of AD. ADNI, ADNI GO, ADNI 2, and ADNI 3 in these three databases, there are 1800 subjects, both females and males [12].

In this study, a subset of the ADNI dataset is used to train, validate, and test different architectures of CNN classifiers. Axial T2-Star weighted longitudinal MRI data of 85 AD, 85 CN, 76 EMCI, and 76 LMCI is chosen for this work. As a result, our base dataset contains a total of 1,633 MRI data points.

3.2 Data Pre-processing & Data Augmentation

For data pre-processing, we have performed rotation, and resizing operations. After that we have performed image augmentation which is an artificial technique for increasing the size of a training dataset by constructing modified versions of the images. Horizontal flip, 5 units rotation, 0.1 width shift, 0.1 height shift, and 0.1 zooms are used in this work to perform augmentation. After the data augmentation, we have achieved a large dataset of 6,392 data including 1420 CN, 1562 LMCI, 1740 EMCI, and 1645 AD. Here, 85% image data are used as training data, and the rest are used as testing data. The algorithm of the data augmentation technique is mentioned below:

Input: Original image data θ;
Output: Augmented training image set λ;
Define $\theta = \lambda$;
For each image sample $\mathrm{I}^k(fi)$ in θ do
  For offset β_ζ = -υ:υ_0:υ do
    Add β_ζ to the true azimuth $\hat{\beta}$, $\beta' = \hat{\beta} + \beta_\zeta$;
    Generate the beamformed signals;
    Generate the feature ν_ζ^k;
    $\lambda = \lambda$ U ν_ζ^k;
  End
End

Here, the original data is denoted as θ, which is disturbed at the time of the feature extraction to get the augmented features λ. Apparently, for each sample in θ, we have found the augmented feature ν_ζ^k (here, is the sample index in θ) by proposing an offset angle β_ζ to the real azimuthangle $\hat{\beta}$. The augmented signal is calculated by the azimuth angle $\beta' = \hat{\beta} + \beta_\zeta$.

3.3 Classification Model

EfficientNet is one of the most effective CNN models that achieve State-of-the-Art accuracy on both ImageNet and common image classification transfer learning tasks, which was first created by Tan and Le in 2019 [13]. EfficientNet has eight variants (B0 to B7), and the width, depth, and resolution of each variant are hand-picked and demonstrated to produce better results. Considering the performance of all 8 variants, it is found that EfficientNet-B7 is the best performing. So, this work is constructed using EfficientNet-B7. To simplify the model, we divided the total 813 layers into 6 modules and one stem. In this case, stem includes the following actions or layers: Input, Rescaling, Normalization, Zero Padding, Conv2D, Batch Normalization, and Activation. Module 1's layers are: Depthwise Conv2D, Batch Normalization, and Activation. Similarly, the other five modules are constructed, as shown in Fig. 1.

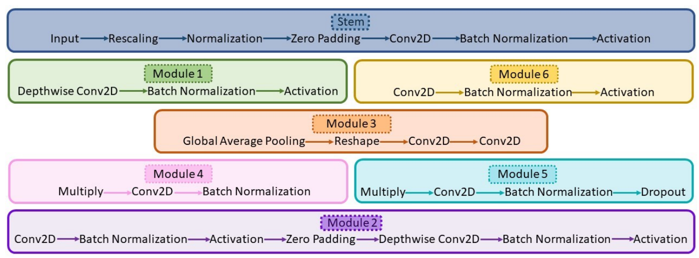

Fig. 1. Modules and stem of EfficientNet-B7 architecture.

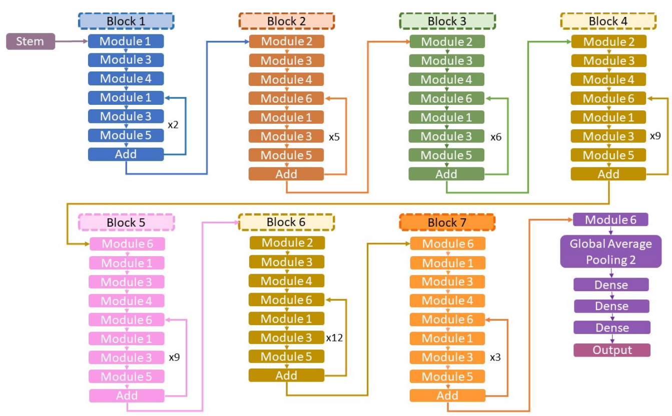

Fig. 2. Complete architecture of EfficientNet-B7 model.

Fig. 2 comes up with the complete architecture of the EfficientNet-B7 model. As it's name suggests, there are total seven blocks in this model. The layers of modules and stems are stated in the Fig. 1. In the first block, module 1 takes input from the stem. The output of module 1 goes to the module 3. Module 4's input is the output of module 3. Module 4's output goes to the module 1. The output of module 1 goes to the module 3 and this module 3's output further goes to the module 5. Add layer takes the output of modules 5. Module 1, 3, 5, and Add stays inside a loop and this loop continues 2 times. After the execution of this loop the final result from block 1 transfers to block 2's module 2. Fig. 2 explains all the modules and their sequence for creating the prediction model. We have constructed the last three layers manually by inserting dense layer. Among those three layers, the first two layers are defined as Dense layers with the unit of '128' and the activation function is 'ReLu'. For four-class classification, the output layer or the last layer contains the unit value '4' along with the 'softmax' activation function. For two-class classification, the last layer contains the unit value '2' along with the 'sigmoid' activation function. For remaining layers, we have used the EfficientNet-B7 model's default layers.

3.4 Explain Ability of the Model

Grad-CAM [14] is a technique for improving the transparency of CNN-based models by visualizing the input territories that are "crucial" for the model's predicted results [15]. Grad-CAM generates a class-specific heatmap based on an input image [16]. Through that heat map, one can easily understand the most dominating (red-colored) and less dominating (blue-colored) portions for any decision.

The main operation of Grad-CAM is described in the following sections. To begin, the oth rectified feature map Po can be denoted as

$$P^o = \frac{1}{Z}\sum_r \sum_s B^o_{r,s} \tag{1}$$

Here, Z is the feature map's number of pixels, r and s are the feature map's row and column indexes, and $B^o_{r,s}$ is the score of the pixel in the rth row and sth column. The class C score (R^C), also recognized as the class C's score before the softmax layer.

$$R^C = \sum_o \beta^C_o P^o \tag{2}$$

Here β^C_o is the oth feature map's weight The first activation map ($W^C_{Grad-CAM}$) produced by Grad-CAM is

$$W^C_{Grad-CAM} = RELU(\sum_o \beta^C_o P^o) \tag{3}$$

The gradient method can be used to analyze the weights of a pixel. The procedure can be represented as follows:

$$\beta^C_o = \frac{\delta R^C}{\delta P^o} = \frac{\delta R^C}{\delta B^o_{r,s}} \frac{\delta B^o_{r,s}}{\delta P^o} = \frac{\delta R^C}{\delta B^o_{r,s}}.Z \tag{4}$$

The average of weights of every pixel in the feature map is the weight of the oth feature map. The oth feature map's weight is

$$\frac{1}{Z}\sum_r \sum_s \beta^C_o = \frac{1}{Z}\sum_r \sum_s \frac{\delta R^C}{\delta B^o_{r,s}}.Z \tag{5}$$

Equation 5 can be simplified because the weights of every pixel in the oth feature map has to be the same.

$$\frac{1}{Z}.Z.\beta_o^C = \frac{1}{Z}\sum_r \sum_s \frac{\delta R^C}{\delta B_{r,s}^o}.Z = \sum_r \sum_s \frac{\delta R^C}{\delta B_{r,s}^o} \tag{6}$$

Using Grad-CAM, the normalized localization map (attention map or heat map) can be defined as

$$S = \frac{1}{Z}\sum_r \sum_s \sum_o \beta_o^C B_{r,s}^o \tag{7}$$

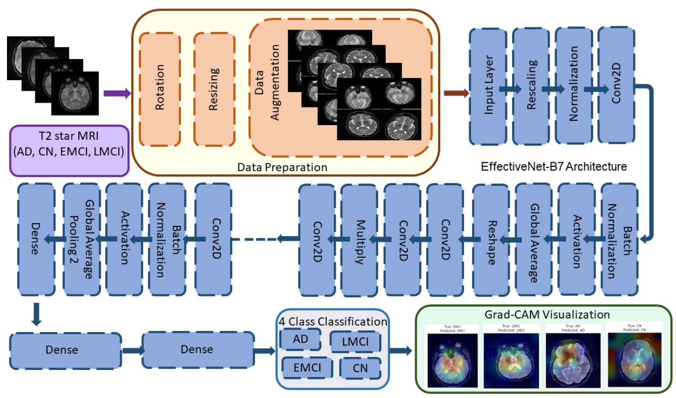

Fig. 3. Proposed model architecture for Alzheimer's prediction and visualization.

3.5 Proposed Explainable Alzheimer's Disease Prediction Using EfficientNet-B7 Architecture

Fig. 3 presents the model architecture of this research work. Here, T2 star MRI images of 4 types (CN, AD, EMCI, and LMCI) of participants are used. Initially for this task, 1,633 MRI images are used. For preparing the dataset, we have performed image rotation, resizing, and data augmentation. To make sufficient dataset for the EfficientNet-B7 model, we have performed augmentation. After augmentation, the number of data has become 6,392 which is overall satisfactory for the EfficientNet-B7 CNN model. Then, we have performed the classification operation using the EfficientNet-B7 pre-trained CNN model and found the classification result. For visualizing the decision-making portions of the MRI images we have used the Grad-CAM algorithm. The last convolutional layer of the EfficientNet-B7 named "*topconv*" is passed to the Grad-CAM algorithm and the final output is shown using the heat map.

4 Performance Analysis

4.1 Performance Metrics

A confusion matrix is an overview of classification problem's prediction results. From confusion matrix we can calculate the Accuracy, Precision, Recall, and F1-score using True positive (TP), True negative (TN), False Positive (FP), and False Negative (FN) values. The rules of Accuracy, Precision, Recall, and F1-score are written below:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

4.2 Data Augmentation vs Without Augmentation

For analyzing the effect of the augmented dataset, we have performed 4 class classifications on the augmented dataset and the base dataset. Without augmentation, the overall accuracy value is 73% and with the augmented dataset this accuracy is 91.76%. Fig. 4 comes up with the normalized confusion matrix where it is clear that the performance using the augmented dataset is far better than the base dataset. Using the augmented dataset the highest Precision value achieved for EMCI and that is 95%. We found the highest Recall value of 94% for AD and LMCI class.

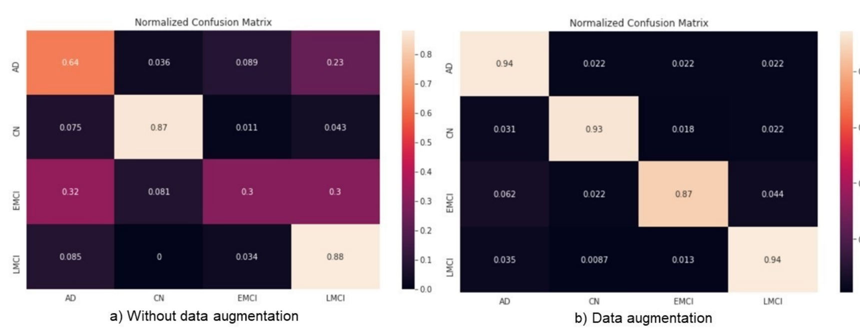

Fig. 4. Confusion matrix of the model using augmented dataset and non-augmented dataset.

4.3 Comparison Between Transfer Learning Models

There are several outstanding existing transfer learning models which are usually used in AD vs CN vs EMCI vs LMCI classification. Table 1 shows that the accuracy of EfficientNet-B7 is higher than other models.

Table 1. Accuracy comparison among various transfer learning models for AD vs CN vs EMCI vs LMCI classification.

Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
InceptionResNet-V2	74.45%	MobileNet-V2	87.49%	ResNet-34	83.73%
ResNet-50	84.78%	ResNet-101	90.71%	ResNet-152	89.89%
VGG-16	88.74%	VGG-19	90.09%	DenseNet-121	84.05%
DenseNet-201	88.63%	Xception	75.91%	NASNetMobile	70.59%
Inception-V3	67.67%	EfficientNet-B0	88.63%	EfficientNet-B1	91.35%
EfficientNet-B2	91.66%	EfficientNet-B3	90.51%	EfficientNet-B4	90.30%
EfficientNet-B5	88.11%	EfficientNet-B6	90.82%	EfficientNet-B7	**91.76%**

Table 2. Accuracy comparison among similar works on six binary classifications.

Binary Classification	Korolev et al. [20] (ResNet)	Shakeri et al. [21] (Multi Layer Perceptron)	Our Proposed Work (Data Augmentation+ EfficientNet-B7 +Grad-CAM)
AD vs CN	80%	84%	**97.72%**
CN vs EMCI	56%	56%	**97.70%**
CN vs LMCI	61%	59%	**98.01%**
AD vs EMCI	63%	81%	**95.12%**
AD vs LMCI	59%	67%	**96.70%**
EMCI vs LMCI	52%	63%	**96.40%**

4.4 Accuracy Comparison With Other Recent Works

From Table 2 we have found that for all six types of binary classification our proposed model achieved highest accuracy scores than others. For CN vs LMCI classification we have found 98.01% accuracy, which is better than other binary classifications. Here, all the authors have used the same dataset collected from the ADNI repository.

Table 3 consists the comparison between various recently published work using the AD and CN MRI data collected from the ADNI repository. From this table it is clear that the accuracy of our proposed work is higher than others and that value is 97.72%.

Table 3. Accuracy comparison among various recently published similar works for AD vs CN classification.

Authors	Model	AD vs CN
Helaly et al. [17]	2D-M2IC	97.11%
Wee et al. [18]	graph-CNN	81.0%
Khvostikov et al. [19]	3D Inception	93.3%
Dan et al. [20]	CNN-EL	84%
Zhang et al. [21]	TRRA+EfficientNet-B1	93%
Our Proposed Work	Augmentation+ EfficientNet-B7+Grad-CAM	**97.72%**

4.5 Result Visualization

As we have already stated that, we have used Grad-CAM technique to visualize the reason behind any specific classification result. Fig. 5 shows the Grad-CAM visualization of this model. The heatmap represents the level of dominance or attention for any classification. Here, the red colored areas are highly involved for the decision making.

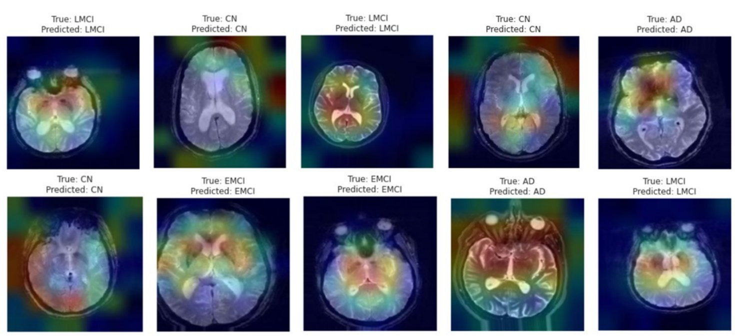

Fig. 5. Explaination of prediction result using Grad-CAM.

5 Conclusion

In this study, we have proposed a Grad-CAM-based EfficientNet-B7 model for Alzheimer's prediction, where the data augmentation technique is used to increase the number of MRI data as required for deep learning. With the help of this approach, we can easily provide predictions on CN, AD, EMCI, and LMCI stages. Our proposed approach removes the black-box nature of the deep learning-based prediction model and also outperforms many recently published works in this field. For AD vs CN vs EMCI vs LMCI classification, our proposed approach achieved 91.76% accuracy. The accuracies achieved from the augmented dataset for AD vs CN, CN vs EMCI, CN vs LMCI, AD vs EMCI, AD vs LMCI, and EMCI vs LMCI are 97.72%, 97.70%, 98.01%, 95.12%, 96.70%, and 96.40%, respectively. In the future, we will try to implement a deep learning model for early-stage AD prediction. For improving the model's performance we are likely to implement an ensemble CNN model.

Acknowledgement

This work was supported by the Master's Level Fellowship fund of the ICT Division of the Government of the People's Republic of Bangladesh in 2020 - 2021 (3rd Round).

References

1. Dementia in the Asia Pacific Region. page 64. [Online]. Available: https://www.alzint.org/u/Dementia-Asia-Pacific-2014.pdf. [Accessed: 05- Nov-2021].
2. Richard Mayeux and Yaakov Stern. Epidemiology of Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(8):a006239, August 2012.
3. Klaus Schmidtke and Sonja Hermeneit. High rate of conversion to Alzheimer's disease in a cohort of amnestic MCI patients. *International Psychogeriatrics*, 20(1):96–108, February 2008.
4. Mr Amir Ebrahimighahnavieh, Suhuai Luo, and Raymond Chiong. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine*, 187:105242, April 2020.
5. Loris Nanni, Matteo Interlenghi, Sheryl Brahnam, Christian Salvatore, Sergio Papa, Raffaello Nemni, and Isabella Castiglioni. Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease. *Frontiers in Neurology*, 11:576194, November 2020.
6. Wenjie Kang, Lan Lin, Baiwen Zhang, Xiaoqi Shen, and Shuicai Wu. Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis. *Computers in Biology and Medicine*, 136:104678, September 2021.
7. Chong-Yaw Wee, Chaoqiang Liu, Annie Lee, Joann S. Poh, Hui Ji, and Anqi Qiu. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *NeuroImage: Clinical*, 23:101929, January 2019.
8. Atif Mehmood, Shuyuan Yang, Zhixi Feng, Min Wang, AL Smadi Ahmad, Rizwan Khan, Muazzam Maqsood, and Muhammad Yaqub. A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. *Neuroscience*, 460:43–52, April 2021.
9. Huanhuan Ji, Zhenbing Liu, Wei Qi Yan, and Reinhard Klette. Early Diagnosis of Alzheimer's Disease Using Deep Learning. In *Proceedings of the 2nd International Conference on Control and Computer Vision*, ICCCV 2019, pages 87–91, New York, NY, USA, June 2019. Association for Computing Machinery.
10. Saeeda Naz, Abida Ashraf, and Ahmad Zaib. Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Systems*, 28(1):85–94, February 2022.

11. Christopher Weber, Maria Carrillo, William Jagust, Clifford Jack, Leslie Shaw, John Trojanowski, Andrew Saykin, Laurel Beckett, Cyrille Sur, Naren Rao, Patricio Mendez, Sandra Black, Kuncheng Li, Takeshi Iwatsubo, Chiung-Chih Chang, Ana Sosa, Christopher Rowe, John Morris, and Michael Weiner. The Worldwide Alzheimer's Disease Neuroimaging Initiative: ADNI-3 updates and global perspectives. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 7, December 2021.
12. Clifford R. Jack Jr., Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J. Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M. Dale, Joel P. Felmlee, Jeffrey L. Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A. Ward, Gregory J. Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.
13. Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946 [cs, stat]*, September 2020. arXiv: 1905.11946.
14. Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2):336–359, February 2020. arXiv: 1610.02391.
15. Kira Vinogradova, Alexandr Dibrov, and Gene Myers. Towards Interpretable Semantic Segmentation via Gradient-weighted Class Activation Mapping. *arXiv:2002.11434 [cs, eess]*, February 2020. arXiv: 2002.11434.
16. Grad-CAM: Gradient-weighted Class Activation Mapping.
17. Hadeer A. Helaly, Mahmoud Badawy, and Amira Y. Haikal. Deep Learning Approach for Early Detection of Alzheimer's Disease. *Cognitive Computation*, November 2021.
18. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations - ScienceDirect.
19. Alexander Khvostikov, Karim Aderghal, Andrey Krylov, Gwenaelle Catheline, and Jenny Benois-Pineau. *3D Inception-based CNN with sMRI and MD-DTI data fusion for Alzheimer's Disease diagnostics.* July 2018.
20. Frontiers | Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning | Neuroscience.
21. Fan Zhang, Bo Pan, Pengfei Shao, Peng Liu, Shuwei Shen, Peng Yao, and Ronald X. Xu. An explainable two-dimensional single model deep learning approach for Alzheimer's disease diagnosis and brain atrophy localization. *arXiv:2107.13200 [cs, eess]*, July 2021. arXiv: 2107.13200.