CNN-Based Categorization of Alzheimer's Disease Progression Phases

Ziqi Li *

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China

* Corresponding author: ziqi_li@bupt.edu.cn

Abstract. This study addresses the challenge of intelligent diagnosis in Alzheimer's Disease by employing machine learning to classify MRI images depicting various disease stages. The author conducted a comparative analysis to assess the efficacy of diverse models for this purpose, yielding several key findings. The primary model utilized in the research is founded on ResNet152V2, encompassing 152 pre-trained layers, 1 GlobalAveragePooling2D layer, 3 dropout layers, 3 dense layers, and 1 output layer. The research aims to achieve precise classification of MRI images corresponding to distinct Alzheimer's Disease stages, a pivotal step for early detection and intervention. To achieve this, this research adjusted the structure of the original model and changed some of the parameters and compared their results. Based on the results obtained, the author drew several conclusions. The author found that setting a larger value for the "patience" parameter in the Early Stopping callback helps avoid stopping the training process too early. Additionally, the author observed that the original model with only 2 fully connected layers was not stable enough in terms of gradient descent, indicating the need for additional fully connected layers. In conclusion, this research paper addressed the problem of classifying MRI images of different stages of Alzheimer's Disease using machine learning techniques. The comparative study of different models provided insights into the performance and potential improvements of the classification taskt.

Keywords: CNN, Deep Learning, Alzheimer's Disease.

1. Introduction

Alzheimer's Disease (AD) represents a neurodegenerative pathology characterized by progressive memory deterioration and cognitive dysfunction, predominantly manifesting in the elderly demographic. The irreversible nature of cognitive decline imposes substantial burdens on affected individuals and society as a whole. Despite the absence of a definitive cure for AD, endeavors have been undertaken to mitigate disease progression, particularly emphasizing interventions during the early stages [1]. The annual escalation in the prevalence of Alzheimer's Disease (AD) and the concomitant rise in mortality due to this condition, as opposed to a diminishing trend in fatalities resulting from other diseases, underscore the growing significance of AD within public health. Projections for the United States indicate an anticipated elevation in the afflicted population to 13.8 million by 2050, compared to 5.4 million in 2016. The diagnostic rate is poised to more than double from 66 cases per 1,000 individuals in 2016 to 33 cases per 1,000 individuals by 2050. Notably, AD accounted for 84,767 fatalities in 2013, experiencing a notable surge of 71% in mortality between 2000 and 2013, in stark contrast to a decline observed in mortality attributed to chronic diseases such as stroke and heart disease during the same period [2].

Nowadays, Artificial Intelligence (AI) technology has been widely used in numerous fields [3-7]. However, compared with other aspects, there are fewer scenarios that apply AI to Medicine. The main reason of this situation is that the technology is not mature enough and few patients are willing to trust AI doctors completely. Although AI cannot replace doctors, it could make a great contribution to operations and diagnosis. AI has the potential to serve as a valuable assistant, aiding doctors in the identification of various diseases and demonstrating enhanced accuracy in medical procedures. According to the fact, AI is mainly used in drug research & development, operation robots and intelligent diagnosis.

When doctors diagnose diseases, they mainly rely on the results of medical inspection. For these results, a large number of them are medical images such as X-ray, Ultrasonic imaging, Computed

Tomography (CT) and Magnetic Resonance Imaging (MRI) etc. Facing the same images, doctors may draw different conclusions due to their abilities and experience and for some unexperienced doctors, this may lead to misdiagnosis. Therefore, it is possible to analyze these results by using computer vision and machine learning. That means, if the results from AI and doctor are different from each other, the doctor would reconsider it. The possibility that both AI and doctor are wrong is lower than the possibility one of them make mistakes. While AI may not always be infallible, it serves as a valuable reference for doctors, alleviating some of their workload. This research aims to optimize the CNNs algorithm employed in the present study to make it perform better in classification of AD stages.

2. Method

2.1. Dataset Preparation

The dataset used in this research is Augmented Alzheimer MRI Dataset V2 [8] from Kaggle. The images in the dataset are patience's grayscale MRI images. The input size of single image is 200×190 pixels. The dataset has four classes of images: Mild Demented, Moderate Demented, Non-Demented and Very Mild Demented. Fig. 1 demonstrates examples of the four categories.

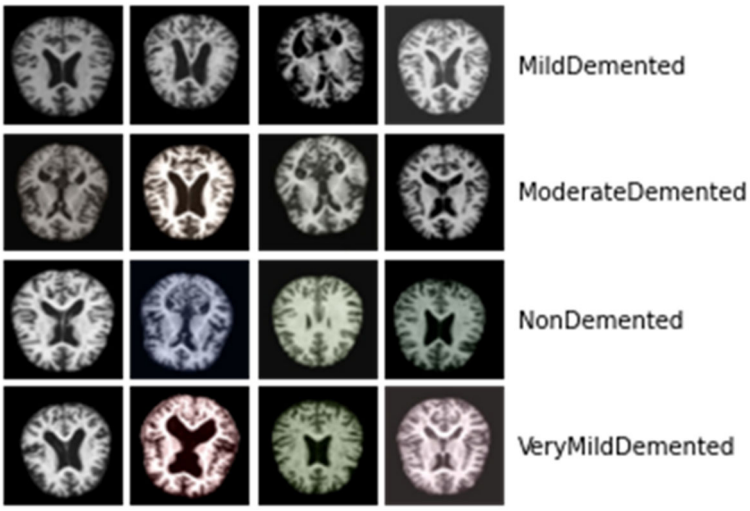


Figure 1. Example images of the collected dataset [8].

In the training dataset, the distribution of images across the four categories is as follows: MildDemented with 8960 images, ModerateDemented with 6464 images, NonDemented with 9600 images, and VeryMildDemented with 8960 images. In addition, the input size is 224x224x3, so an image needs to be reformed to the target size 224x224. Then it will be transformed as an array and be divided by 255 (normalization). After that use the function "expand_dims" to augment the array, let parameter "axis" = 0.

2.2. Optimized CNN

CNNs usually consist of three layers: convolutional layer, max pooling layer and fully connected layer. Convolutional layer is used to extract feature(s) of images; max pooling layer is used for dimension reduction; it reduces the amount of calculation. The two layers' role is data preprocessing, which is essential to classification. Fully connected layer is used for classification.

The purpose of the research is to detect "Mild Demented", "Moderate Demented", "Non-Demented" and "Very Mild Demented" in patient's MRI, to train the CNN model by using the images in train data and classify as many images in test data correctly as possible. In this research, some of the parameters were adjusted to make the model perform better. The CNN based Alzheimer MRI images classification algorithm is developed using ResNet152V2 architecture.

ResNet152 is a residual learning framework that has 152 layers but have lower complexity than VGG nets [9].

He et al. introduced the ResNet architecture as a solution to address the challenge of vanishing gradients encountered in the training of exceptionally deep convolutional networks. ResNets have demonstrated superior performance compared to earlier models across diverse tasks, including object detection and semantic image segmentation. This advancement has led to the gradual displacement of VGGNets within the domain of computer vision [10].

Initially, calling tf. keras. applications. resnet_v2. ResNet152V2() function to create a ResNet152V2 model. The model uses weights that are pre-trained on ImageNet and does not include a fully connected layer on top. The input shape is (224, 224, 3), which means that the size of the input image is 224x224 pixels and the number of channels is 3. Subsequently, by traversing the first 100 layers of the ResNet152V2 model, these layers are set as untrainable. This is because the first few layers of the pre-trained ResNet152V2 model are convolutional layers used to extract low-level features, and they usually do not need to be retrained. Next, set the rest of the model's layers as trainable (starting at layer 101). The output of the ResNet152V2 model was further taken as input, and the feature map was converted into a vector by the global average pooling layer tf. keras. layers. GlobalAveragePooling2D (). In the next step, adding a dropout layer to prevent overfitting. The dropout layer reduces the relevance of the input by randomly setting the input element to 0, the parameter is 0.2. By adding a fully connected layer tf. keras. layers. Dense(1024, activation='relu'), the eigenvectors are mapped to the 1024-dimensional hidden layers, and the ReLU activation function is used for nonlinear transformation. Add another dropout layer with a parameter of 0.2. A fully connected layer tf. keras. layers. Dense (256, activation=’relu’) was added again to map the hidden layer to 256 dimensions and use the ReLU activation function to perform a nonlinear transformation. Add another dropout layer with a parameter of 0.2. Finally, an output layer tf.keras. layers. Dense(4, activation='softmax') is added to map the output of the hidden layer to the probability distribution of the 4 categories, and the softmax activation function is used for multi-classification.

After determining the structure of the model, creating a new model res_model with the input of the ResNet152V2 model and the output as the output of the layer added above. This builds a complete network model. Subsequently, the model is compiled, using categorical_crossentropy as the loss function, the ADAM optimizer, and the accuracy as the evaluation metric.

3. Results and Discussion

Table 1 is the compared result of different models. The training and validation performance of different models were shown in Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7. Model 2 and 4 add a fully connected layer tf. keras. layers. Dense(512, activation='relu') to map the eigenvectors to the 512-dimensional hidden layer between the two fully connected layers of the original model, and use the ReLU activation function to perform a nonlinear transformation. Add another dropout layer with a parameter of 0.2. Also, the learning rate of “adam” optimizer is 0.01.

Table 1. The results of different models

Models/ parameters	original	1	2	3	4	5
Batch size	32	32	64	64	64	64
patience	5	10	10	10	10	10
epochs	100	500	20	200	200	200
Step per epoch	default	default	425	100	425	100
Validation steps	default	default	107	default	107	default
Train loss	0.27	0.32	0.19	0.36	0.17	0.22
Train accuracy	92.31 %	90.75 %	94.31 %	92.97 %	95.50 %	94.63 %
Test loss	0.18	0.16	0.03	0.16	0.04	0.03
Test accuracy	94.69 %	95.78 %	98.95 %	96.00%	98.69 %	99.27 %

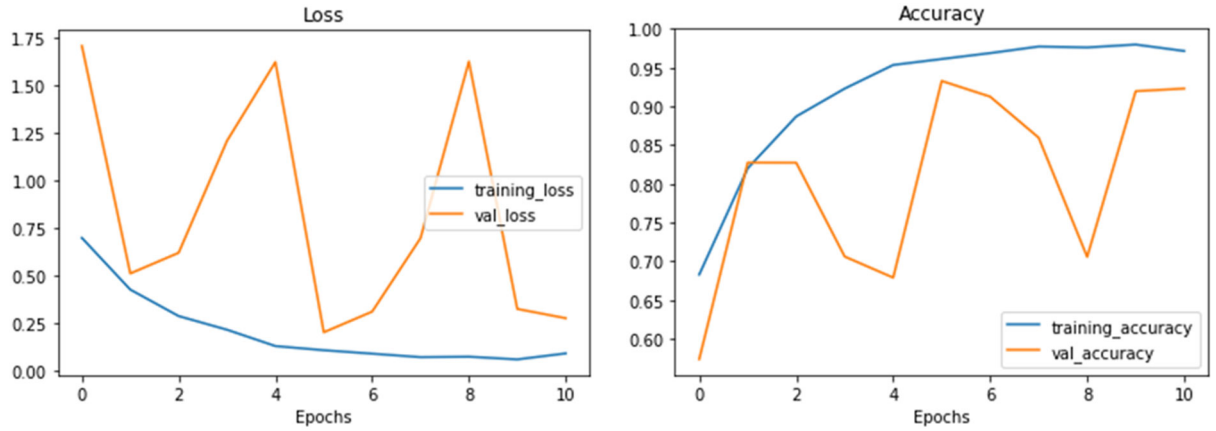


Figure 2. The training and validation performance of the original model (Photo credit: Original).

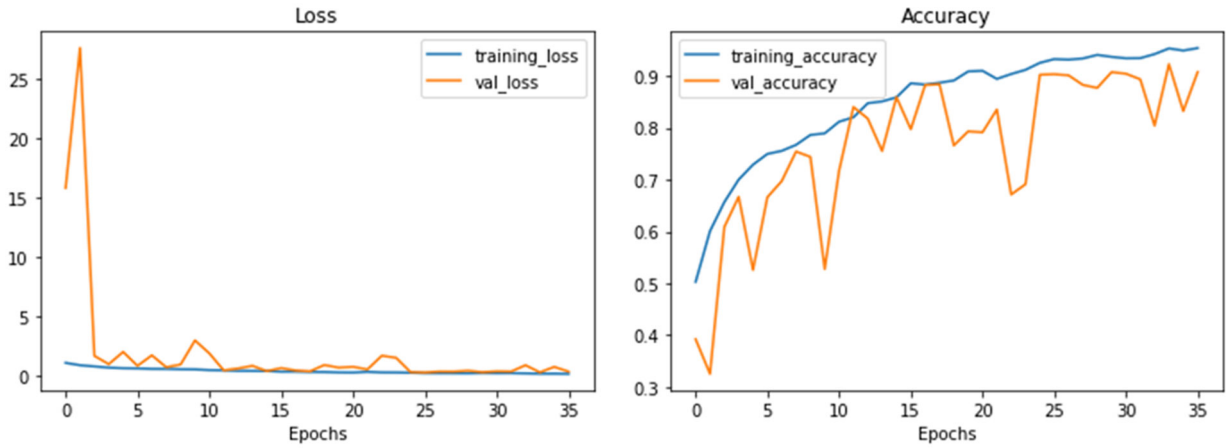


Figure 3. The training and validation performance of model 1 (Photo credit: Original).

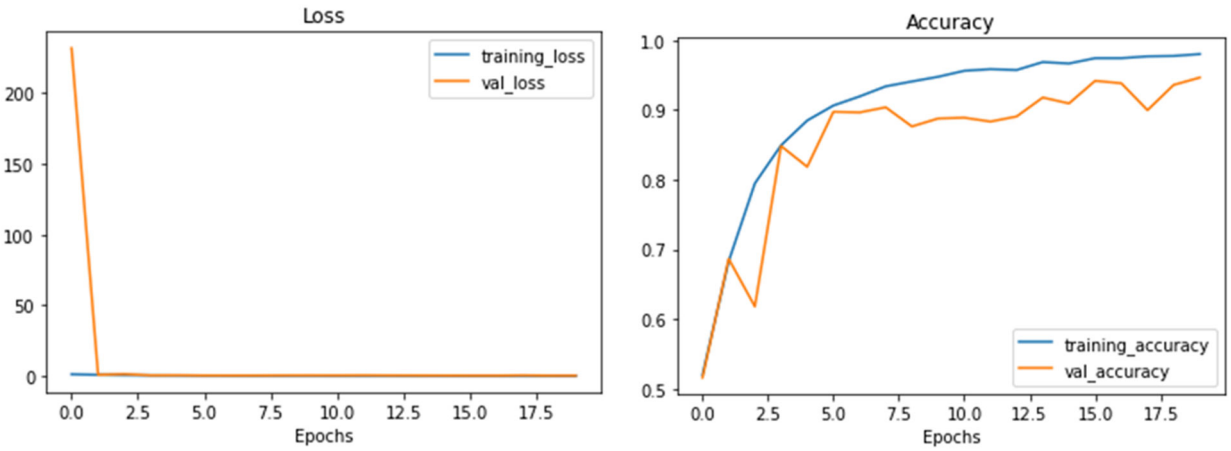


Figure 4. The training and validation performance of model 2 (Photo credit: Original).

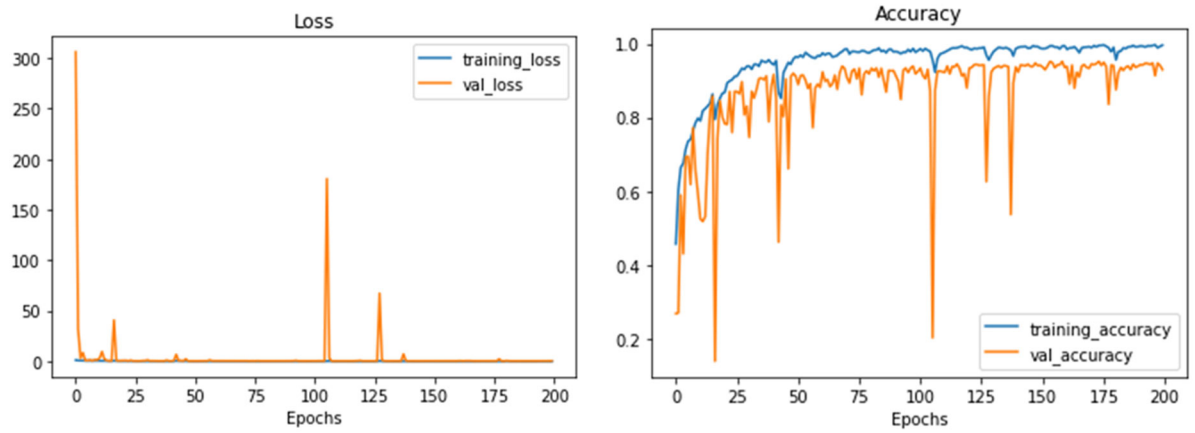


Figure 5. The training and validation performance of model 3 (Photo credit: Original).

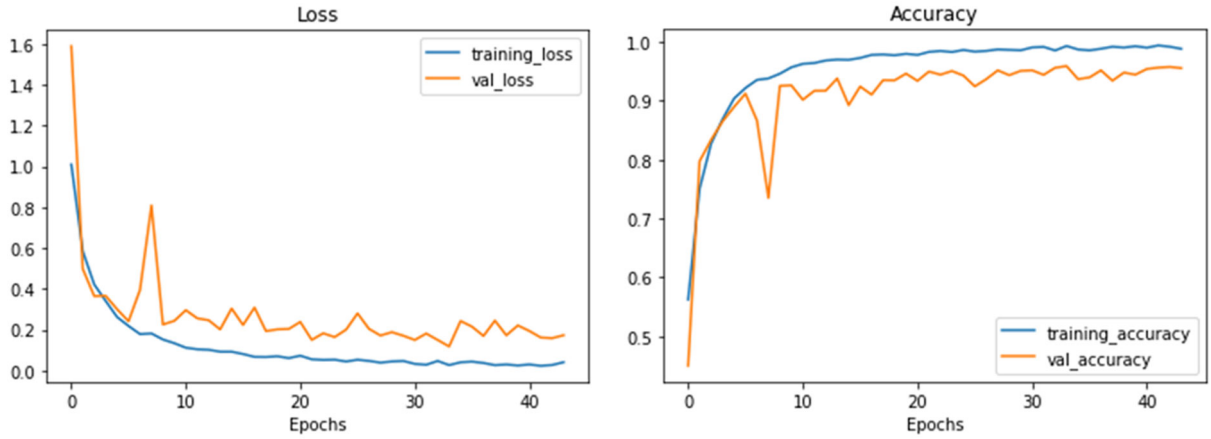


Figure 6. The training and validation performance of model 4 (Photo credit: Original).

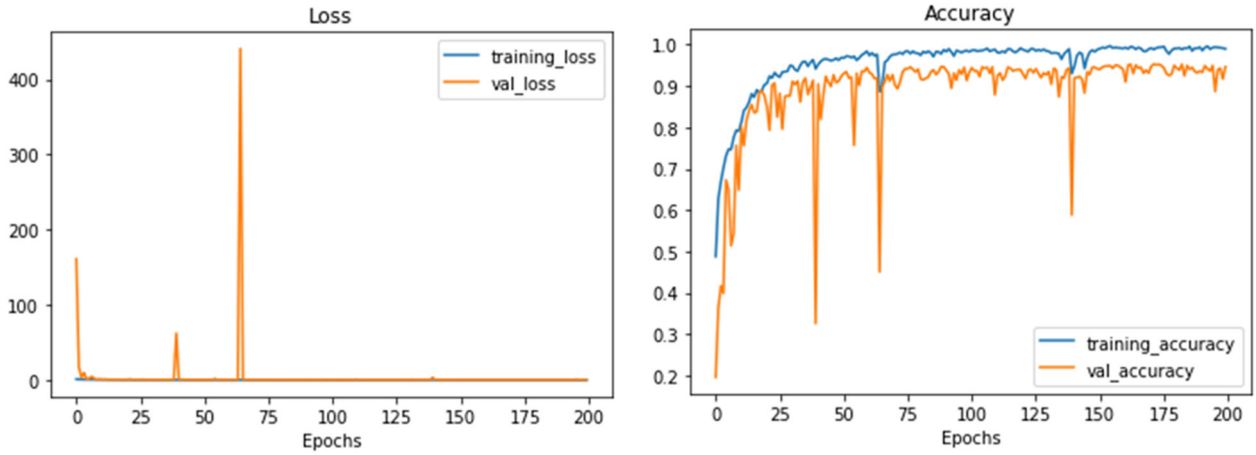


Figure 7. The training and validation performance of model 5 (Photo credit: Original).

According to Table 1, model 2, 4, 5 performed well in this classification problem. Model 1 has higher test accuracy but lower train accuracy compared with the original model. Model 3 is slightly better than the original model.

Model 1, 2, 4 have relatively smooth curves of the validation loss and accuracy while model 3 and 5's are choppy.

1) The original model sets an early stopping, whose parameter "patience" is 5. It means if the validation loss does not decrease in 5 epochs, it would stop training the model. And the plot figures show that it stopped after 10 epochs, which means it stopped early. In the study, the parameter

"patience" was modified to a value of 10, resulting in enhanced test accuracy across all adjusted models compared to the original model. Notably, for Models 2, 4, and 5, with set epochs of 20, 200, and 200, respectively, only Model 4 experienced early termination. This suggests the possibility that the original model might not have completed its training regimen when the process was halted.

Top of Form

2) Model 1, 3, 5 has the same structure of the original model. Model 1 only adjusted patience and epochs but it did not train over 100 epochs and didn't perform better significantly than the original model so some other parameters need to be changed. Model 3 and 5 set step per epoch as 100 and performed better.

3) The only difference between model 2 and 4 is epochs. Model 2 does not stop early while model 4 stopped after training approximately 45 epochs. Hence the ideal number of epochs of the models that have an extra fully connected layer is more than 45. Also, the extra layer makes the curves of loss and accuracy smoother.

4) The parameters of models 3 and 5 are the same. However, while the loss and accuracy curves exhibit similarities, they are not entirely identical. Strikingly, Model 5 demonstrates superior performance, showcasing higher test accuracy than both Model 2 and Model 4. While this outcome may be coincidental, it underscores the inherent instability in the results of such models.

4. Conclusion

This research tackles the intelligent diagnosis challenge in Alzheimer's Disease by utilizing machine learning to categorize MRI images representing different stages of the disease. In light of the presented results, several noteworthy conclusions can be derived. To prevent premature termination of the training process and ensure thorough model training, it is advisable to set a larger value for the "patience" parameter. The original model, featuring only two fully connected layers, exhibits instability in gradient descent. Addressing this issue involves augmenting the model with additional fully connected layers. While this modification can enhance performance, especially compared to models with more layers, the associated risks in a medical context are considerable, as even a minor error could adversely impact patient outcomes and prove challenging to rectify. Despite achieving the highest accuracy rates, Models 4 and 5 leave some unresolved issues due to time constraints. The optimal parameters for Model 4 remain uncertain, and there is a potential for further improvements by exploring alterations or optimizations to the model's structure. These outstanding challenges underscore the complexity of the problem and emphasize the need for continued research and refinement in the pursuit of accurate and reliable classification of MRI images in the context of medical diagnoses.

References

[1] Zhao Z, et al. Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: A review. Frontiers in Computational Neuroscience 17, 2023, 1038636.

[2] Salehi W, et al. An approach to binary classification of alzheimer's disease using LSTM. Bioengineering 10.8, 2023, 950.

[3] Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. arXiv preprint arXiv: 1701.03056, 2017.

[4] Qiu Y, Wang J, Jin Z, et al. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 2022, 72: 103323.

[5] Chen H, et al. Kernel-mask knowledge distillation for efficient and accurate arbitrary-shaped text detection. Complex & Intelligent Systems, 2023: 1 - 12.

[6] Zhou X, Li Y, Liang W. CNN-RNN based intelligent recommendation for online medical pre-diagnosis support. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2020, 18 (3): 912 - 921.

[7] Cao L. Ai in finance: challenges, techniques, and opportunities. ACM Computing Surveys (CSUR), 2022, 55 (3): 1 - 38.

[8] Kaggle Augmented Alzheimer mri dataset v2 https: //www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset-v2, 2023.

[9] He K, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[10] Wu Z, et al. Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition 90, 2019, 119 - 133.