Journal of Information Systems Engineering and Management
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Analysis of Skull-Stripped Versus Non-Stripped MRI Scans for CNN-Based Alzheimer's Disease Classification

Atul Mathur[1]*, Rakesh Kumar Dwivedi[2], Rajul Rastogi[3]

*[1]*College Of Computing Sciences & IT, Teerthanker Mahaveer University, Moradabad, India. atulm1979@gmail.com*

[2]College of Computing Sciences & IT, Teerthanker Mahaveer University, Moradabad, India. principal.computers@tmu.ac.in

[3]Teerthanker Mahaveer Medical College & Research Centre, Teerthanker Mahaveer University, Moradabad, India. rajulrst@yahoo.co.in

ARTICLE INFO

Received: 20 Dec 2024

Revised: 11 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Alzheimer's disease (AD) is a debilitating neurodegenerative condition characterized by progressive cognitive decline. Early detection enables experts to initiate preventive treatment at the initial possible stage. The primary focus of the work is to develop an automated detection system using a Convolutional Neural Network (CNN), capable of analyzing the brain magnetic resonance imaging (MRI) scans acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Moreover, skull removal was accomplished on the brain MR scans to create a new dataset. A comparative analysis was conducted on both datasets. Results showed impressive classification accuracies: 98.27% for skull-included scans and slightly higher for skull-stripped scans. The findings of the study demonstrate the successful application of deep learning (DL) techniques in neurological disease detection. The slight improvement in accuracy with the stripped version enables future preprocessing considerations in these areas.

Keywords: Alzheimer's disease, ADNI, CNN, MRI, Deep learning

1 Introduction

AD is a predominant form of dementia that leads to a gradual decline in memory, cognitive abilities, and everyday functioning [1][15]. It stands at the forefront of global health challenges, affecting millions globally, particularly the elderly. The projection indicates a dramatic rise to 82 million dementia cases by 2030. The disorder presents an escalating health crisis that needs immediate attention, creating challenges for the healthcare system and societies at large [18]. Early intervention plays a significant role in slowing down the progression and impacting patient outcomes to preserve life satisfaction for individuals [10][23]. Machine learning has emerged as a revolutionary technique in medical diagnostics. This technological innovation offers promising approaches in the identification of AD and treatment planning [21]. Early diagnosis is crucial as it presents ways to uphold quality standards of living, not only for patients but also for caregivers. Mild cognitive impairment (MCI) occupies a middle ground between cognitively normal (CN) and AD. MCI acts as a warning signal. MCI patients have shown different patterns of disruption in neural networks.

The remarkable capabilities of DL are evident in medical sciences, especially in diagnostics [3][19]. The medical field has already witnessed DL's powerful contributions [14]. The sophisticated applications automatically detect subtle patterns and latent attributes from complex datasets [4][9]. The models have shown high performances in terms of accuracy in feature identification for numerous tasks, such as image classification. CNN is powerful analyzing tool in medical imaging. It has the ability to process MR scans with exceptional results. With classification accuracy surpassing 90%, CNNs are sequentially trained networks consistently achieving outstanding results [5].

Recent advancements in MRI have revolutionized the understanding of AD. Researchers are now able to map the progressive breakdown of neural networks at an early stage, allowing caregivers to monitor treatment efficacy with precision [11][12]. Sample MR scans of CN, MCI, and AD are shown in Figure 1.

CNNs are specialized category of DL architecture and stand for a breakthrough in the domain of image analysis and pattern detection [6][7]. CNNs are instrumental in the interpretation of medical imaging modalities including MRIs. CNNs independently organize features from raw imaging data in a hierarchical fashion [20]. The automated identification of relevant patterns in medical applications provides new possibilities for efficient diagnostic tools [8].

In this research paper, we evaluated the promising aspects of CNN-based approaches for detecting AD with brain images. We developed the model based on CNN employed on the ADNI dataset of brain MR scans. Initially, we applied the model on ADNI dataset with-skull brain MR scans and observed the performance measures, then applied the skull removal methods and obtained a skull-stripped dataset of brain MR scans, employed the model for classification, followed by a comparative performance analysis of both approaches [8].

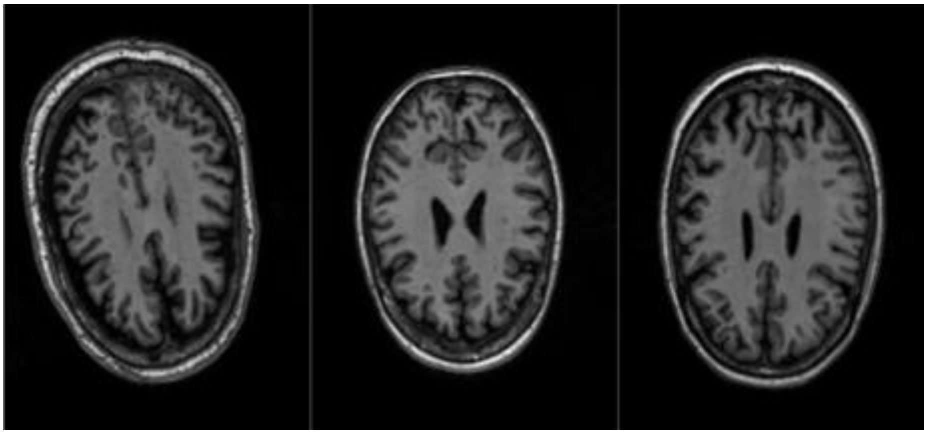

Figure 1. Sample MRI brain scans of CN, MCI, and AD

2 Materials and Methods

2.1 Workflow

Figure 2 illustrates the workflow of the research. Patients' 3-D MR scans were integrated into the framework. These scans were then processed to extract 2-D slices. In preprocessing, data augmentation procedures were deployed to enrich the dataset for the accuracy of the results. For skull removal, the dataset was duplicated to create two versions. On one version, a series of image processing techniques were performed on raw brain scans including: Convert BGR to RGB, Thresholding, Gaussian Blur, Subtract Processed Image from Original, Convert Processed Image to Grayscale, Adaptive Thresholding, Convert Grayscale to BGR, and Inverting Image. The output of each step is shown in Figure 3. After skull removal, the datasets were then divided into training (80%) and testing (20%) subsets. A CNN was then employed for feature extraction and to classify the scans into three classes: CN, MCI, and AD.

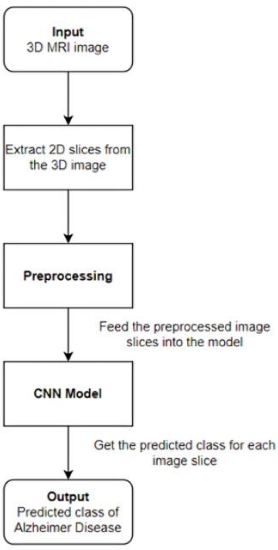

Figure 2. Workflow diagram of proposed model

2.2 Data Sets

The ADNI data set encompasses MR brain scans, containing 5164 3-D MRI images. From these, 2-D slices were extracted. The scans underwent data augmentation using various techniques: 10-degree rotations, flips along both horizontal and vertical axes, and brightness adjustments ranging from 0.3 to 1.7. After augmentation, the total images reached 19292 across three classes CN, MCI, and AD.

2.3 Proposed Method

The structured design of the suggested framework is depicted in Figure 4. This 2-D CNN model employs the 27 layers, trained with both with-skull and skull-stripped datasets separately to extract features. The model has five layers of convolutional, activation function, batch normalization, max pooling, one flatten layer, three dense layers, two dropout layers, and one output layer. The description of the layers is given below.

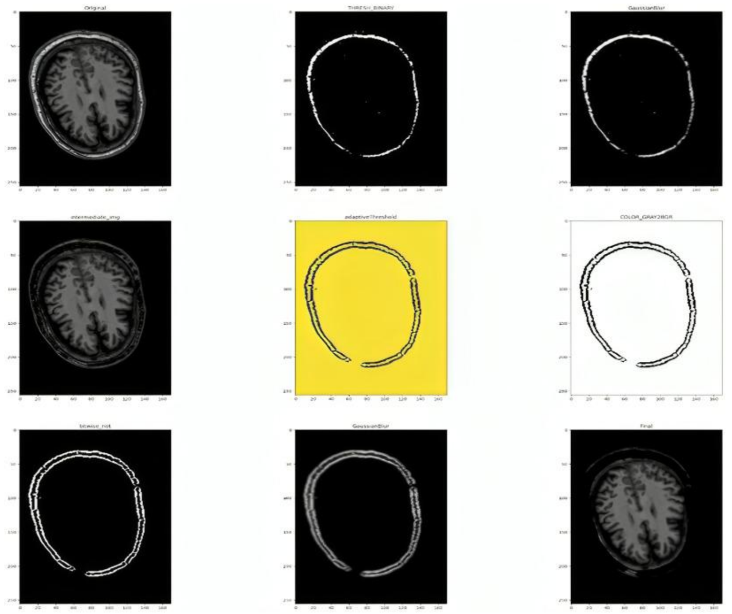

Figure 3. Different stages of skull removal from a raw MR brain scan

2.3.1 Convolutional Neural Network

CNNs automatically learn meaningful patterns from visual data such as photos and videos [17][22]. This architecture is inspired by cortical neurobiology and comprises convolutional layers. CNNs are particularly effective on datasets with numerous nodes and parameters.

2.3.2 Activation Layer

An activation function introduces non-linear properties to the network, which is essential for learning complex patterns through backpropagation. This layer takes an input signal and applies a non-linear transformation, generating output that serves as input for the next layer. Common non-linear activation functions include Sigmoid, Tanh, ReLU, and ELU. The model used ReLU as its activation function.

2.3.3 Batch Normalization

This technique standardizes data by normalizing the outputs of the previous layer across each batch, enhancing the network's training efficiency and stability. In this model's implementation, batches of 16 images were processed together. This standardization process accelerates learning and creates a more robust neural network [2].

2.3.4 Max Pooling Layer

Max pooling serves to reduce the spatial dimensions of feature maps through downsampling [2]. This layer compresses the convolutional block by selecting the most prominent features, thereby decreasing the model's computational load and number of parameters. Max pooling was implemented after batch normalization in each convolution block, which helps prevent overfitting while preserving essential information.

2.3.5 Flatten Layer

This layer transforms three-dimensional array inputs into one-dimensional vectors. By restructuring the data in this way, the flatten layer prepares the output for processing by subsequent fully connected layers, serving as a bridge between convolutional and dense layers.

2.3.6 Dense Layer

In the model architecture, the dense layer connects every neuron to all neurons in the next layer, creating a web of connections. These layers can capture complex non-linear relationships in the data. The dense layer's comprehensive connectivity makes it powerful for sorting data into different classes. This type of deep neural network architecture is particularly suited for classification tasks [2].

2.3.7 Output Layer

The network culminates in an output layer containing three neurons, each corresponding to a different classification category. This layer implements the SoftMax activation function, which calculates probability distributions across all classes, with the highest probability indicating the predicted class. This configuration is specifically designed for multi-class classification problems.

For model optimization, Adam (Adaptive Moment Estimation) optimizer is employed to handle backpropagation when losses occur. The learning rate parameter is configured at 10^{-4}.

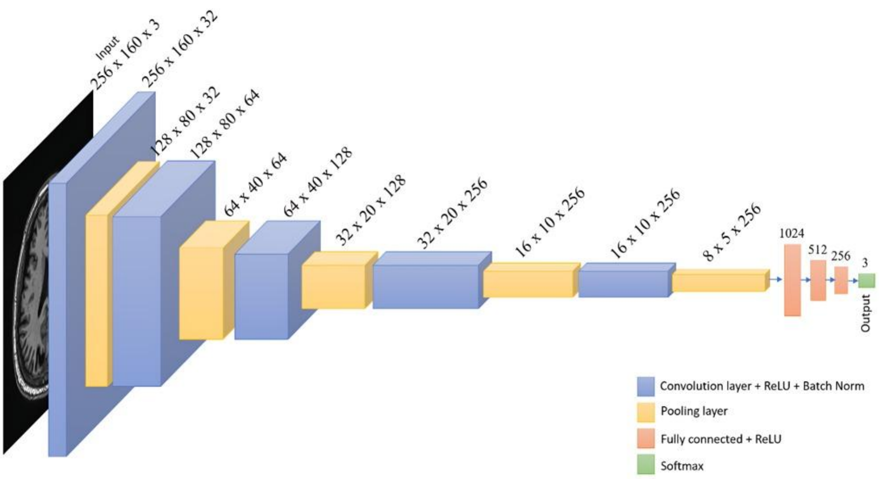

Figure 4. Model architecture diagram for the proposed system

In the CNN architecture, filter sizes are (3,3) and (2,2), varying through the network for the convolutional and max-pooling layers respectively. The total parameters in the model are 12,288,963 including trainable and non-trainable parameters. Details of layers, output shape, and parameters are shown in Table 1.

Table 1. Model Layers with output shape and parameters.

Layer	Output Shape	Param #
Conv1 (Conv2D)	(None, 256, 160, 32)	896
Relu1 (ReLU)	(None, 256, 160, 32)	0
BatchNorm1 (Batch Normalization)	(None, 256, 160, 32)	128
MaxPool1 (MaxPooling2D)	(None, 128, 80, 32)	0

Conv2	(None, 128, 80, 64)	51264
Relu2	(None, 128, 80, 64)	0
BatchNorm2	(None, 128, 80, 64)	256
MaxPool2	(None, 64, 40, 64)	0
Conv3	(None, 64, 40, 128)	204928
Relu3	(None, 64, 40, 128)	0
BatchNorm3	(None, 64, 40, 128)	512
MaxPool3	(None, 32, 20, 128)	0
Conv4 (Conv2D)	(None, 32, 20, 256)	295168
Relu4 (ReLU)	(None, 32, 20, 256)	0
BatchNorm4	(None, 32, 20, 256)	1024
MaxPool4	(None, 16, 10, 256)	0
Conv5	(None, 16, 10, 256)	590080
Relu5	(None, 16, 10, 256)	0
BatchNorm5	(None, 16, 10, 256)	1024
MaxPool5	(None, 8, 5, 256)	0
Flatten (Flatten)	(None, 10240)	0
Dense1 (Dense)	(None, 1024)	10486784
Dropout1	(None, 1024)	0
Dense	(None, 512)	524800
Dropout2	(None, 512)	0
Dense3 (Dense)	(None, 256)	131328
Output (Dense)	(None, 3)	771

2.4 Performance Measures

The classification report offers a detailed assessment of a model's performance, especially in the case of multi-class problems [16]. It comprises key metrics like precision, recall, and F1-score, which help in evaluating how well the algorithm performs for each class [13].

2.4.1 Precision

Precision measures the accuracy of the model's positive predictions for an individual class. It is the ratio of the number of true positives and the sum of true positives and false positives.

Journal of Information Systems Engineering and Management
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2.4.2 Recall

To measure the ability of the model to correctly identify all instances of an individual class, recall is used. Recall, also referred to as sensitivity or true positive rate, is the ratio of true positives to the sum of true positives and false negatives.

2.4.3 F1-Score

To get the balance between precision and recall, the F1-score is used. It is the harmonic mean of precision and recall, which is particularly useful when dealing with imbalanced datasets. It is calculated as in (1)

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (1)$$

2.4.4 Accuracy

Accuracy can be calculated as the ratio of correctly predicted instances to the total number of instances.

3 RESULTS & DISCUSSION

The brain images with and without a skull were analyzed for AD on a 2-D CNN model. The experimental findings demonstrate that our proposed CNN model delivers state-of-the-art performance in detecting AD on both datasets. The model achieved excellent accuracy, underscoring its potential for clinical AD diagnosis applications.

The CNN classifier achieved slightly higher accuracy on skull-stripped brain scans in the testing dataset. The achieved accuracy for the with-skull dataset was 98.78%, while for the skull-stripped brain MR scans dataset, it was 98.97%. Table 2 shows the comparison of classification reports for with-skull and skull-stripped brain scan datasets.

Table 2. Comparison of prediction results (a) with-skull and (b) skull-stripped brain MR scans

(a)

Class	Precision	Recall	F-1 score
CN	0.98	0.99	0.99
MCI	0.98	0.99	0.99
AD	1.00	0.98	0.99

(b)

Class	Precision	Recall	F-1 score
CN	0.99	0.98	0.99
MCI	0.99	0.99	0.99
AD	0.99	0.99	0.99

We also compared both of the approaches in terms of loss and accuracy. The loss and accuracy graphs for with-skull dataset are shown in Figure 5 and Figure 6, while for skull-stripped dataset are presented in Figure 7 and Figure 8 respectively.

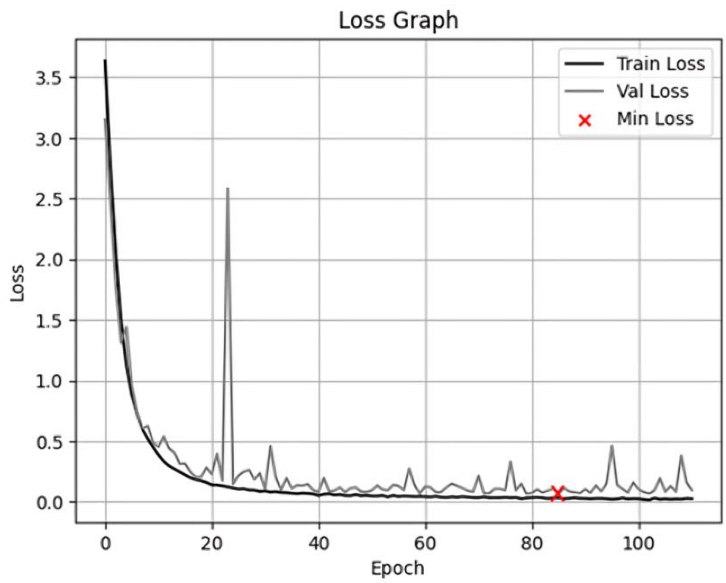

Figure 5. Loss graph for with-skull brain MR scans

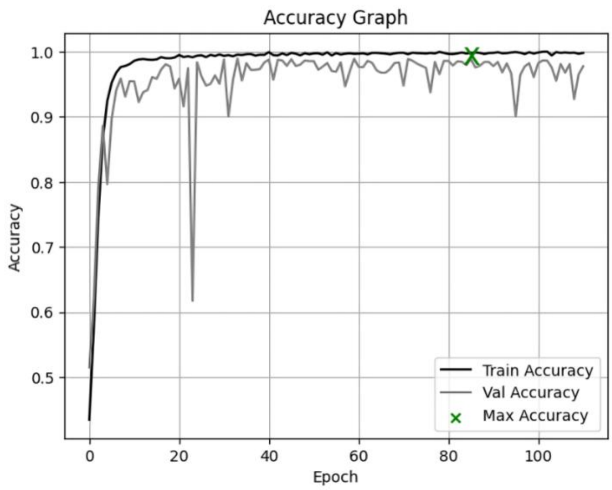

Figure 6. Accuracy graph for with-skull brain MR scans

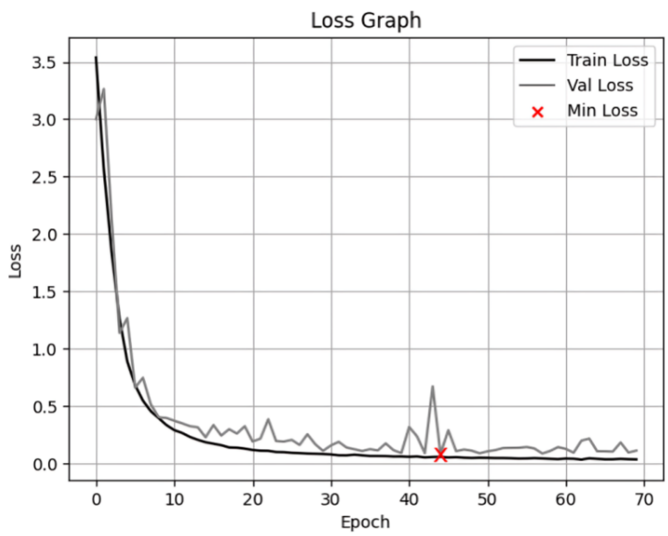

Figure 7. Loss graph for skull-stripped brain MR scans

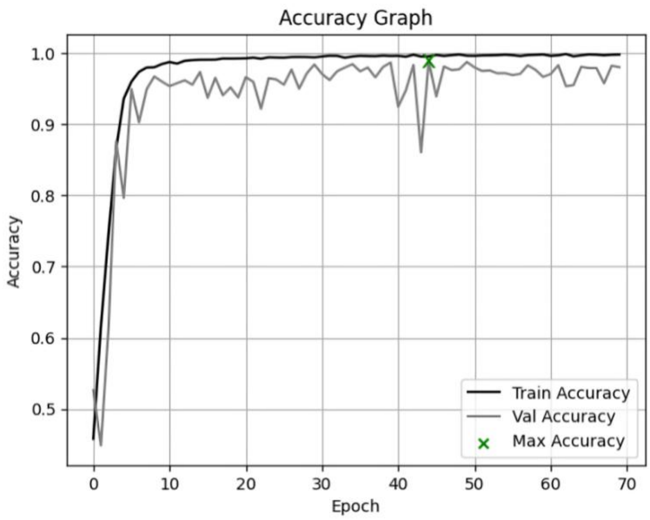

Figure 8. Accuracy graph for skull-stripped brain MR scans

4 CONCLUSION

In this paper, we explored the potential applications of DL for early diagnosis of AD, leveraging CNN to analyze neuroimaging data. The algorithm automatically extracts disease-related biomarkers from brain scans, enabling timely diagnosis and monitoring. Model performance across datasets demonstrates the validity of the approach for predicting AD and MCI from MR brain images. We implemented skull removal techniques on MR images and evaluated the performance of the model on both the original and skull-stripped datasets. A comparative study was performed on the classification reports. The classification metrics indicated varying levels of performance across accuracy, sensitivity, specificity, and F1-score.

REFRENCES

[1] AbdulAzeem, Y., Bahgat, W. M., & Badawy, M. (2021). A CNN based framework for classification of Alzheimer's disease. *Neural Computing and Applications*, *33*(16), 10415-10428.

[2] Amini, M., Pedram, M. M., Moradi, A., & Ouchani, M. (2021). Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN). *Computational and Mathematical Methods in Medicine*, *2021*(1), 5514839.

[3] Coelho, M., Cerny, M., & Tavares, J. M. R. (2025). Deep learning methods to detect Alzheimer's disease from MRI: A systematic review. *Expert Systems*, *42*(1), e13463.

[4] Cui, R., Liu, M., & Alzheimer's Disease Neuroimaging Initiative. (2019). RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, *73*, 1-10.

[5] Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. (2020). Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Computer methods and programs in biomedicine*, *187*, 105242.

[6] Farooq, A., Anwar, S., Awais, M., & Rehman, S. (2017, October). A deep CNN based multi-class classification of Alzheimer's disease using MRI. In *2017 IEEE International Conference on Imaging systems and techniques (IST)* (pp. 1-6). IEEE.

[7] Folego, G., Weiler, M., Casseb, R. F., Pires, R., & Rocha, A. (2020). Alzheimer's disease detection through whole-brain 3D-CNN MRI. *Frontiers in bioengineering and biotechnology*, *8*, 534592.

[8] Franciotti, R., Nardini, D., Russo, M., Onofrj, M., Sensi, S. L., & Alzheimer's Disease Neuroimaging Initiative. (2023). Comparison of machine learning-based approaches to predict the conversion to Alzheimer's disease from mild cognitive impairment. *Neuroscience*, *514*, 143-152.

[9] Helaly, H. A., Badawy, M., & Haikal, A. Y. (2022). Deep learning approach for early detection of Alzheimer's disease. *Cognitive computation*, *14*(5), 1711-1727.

[10] Hon, M., & Khan, N. M. (2017, November). Towards Alzheimer's disease classification through transfer learning. In *2017 IEEE International conference on bioinformatics and biomedicine (BIBM)* (pp. 1166-1169). IEEE.

[11] Kruthika, K. R., Maheshappa, H. D., & Alzheimer's Disease Neuroimaging Initiative. (2019). CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. *Informatics in Medicine Unlocked*, *14*, 59-68.

[12] Li, H., Habes, M., Wolk, D. A., Fan, Y., & Alzheimer's Disease Neuroimaging Initiative. (2019). A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia*, *15*(8), 1059-1070.

[13] Li, Q., & Yang, M. Q. (2021). Comparison of machine learning approaches for enhancing Alzheimer's disease classification. *PeerJ*, *9*, e10549.

[14] Marwa, E. G., Moustafa, H. E. D., Khalifa, F., Khater, H., & AbdElhalim, E. (2023). An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alexandria Engineering Journal*, *63*, 211-221.

[15] Nanni, L., Interlenghi, M., Brahnam, S., Salvatore, C., Papa, S., Nemni, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2020). 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.

[16] Puente-Castro, A., Fernandez-Blanco, E., Pazos, A., & Munteanu, C. R. (2020). Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. *Computers in biology and medicine*, *120*, 103764.

[17] Roy, S. S., Sikaria, R., & Susan, A. (2019). A deep learning based CNN approach on MRI for Alzheimer's disease detection. *Intelligent Decision Technologies*, *13*(4), 495-505.

[18] Srivastava, A., Dhall, A., Patiyal, S., Arora, A., Jarwal, A., & Raghava, G. P. (2023). Prediction of alzheimer's disease from single cell transcriptomics using deep learning. *bioRxiv*, 2023-07.

[19] Tang, X., & Liu, J. (2021). Comparing different algorithms for the course of Alzheimer's disease using machine learning. *Annals of Palliative Medicine*, *10*(9), 9715724-9719724.

[20] Wu, C., Guo, S., Hong, Y., Xiao, B., Wu, Y., Zhang, Q., & Alzheimer's Disease Neuroimaging Initiative. (2018). Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. *Quantitative imaging in medicine and surgery*, *8*(10), 992.

[21] Yiğit, A., & Işik, Z. (2020). Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease. *Turkish Journal of Electrical Engineering and Computer Sciences*, *28*(1), 196-210.

[22] Yildirim, M., & Cinar, A. (2020). Classification of Alzheimer's Disease MRI Images with CNN Based Hybrid Method. *Ingénierie des Systèmes d Inf.*, *25*(4), 413-418.

[23] Zhou, Q., Wang, J., Yu, X., Wang, S., & Zhang, Y. (2023). A survey of deep learning for Alzheimer's disease. *Machine learning and knowledge extraction*, *5*(2), 611-668.