

Original Research Paper

Ensemble Stacking Method of Classifying the Stages of Alzheimer's Disease by using MRI Dataset

Nagarathna¹, Kusuma², Harsha Huliappa³

¹ Department of Artificial Intelligence and Machine Learning, BNM Institute of Technology, Bangalore, India.

² Department of Information Science and Engineering, Dayananda Sagar Academy of Technology & Management, Bangalore, India.

³ Department of Neurosurgeon, Narayana Multispeciality Hospital, Mysuru, India.

Article History

Received:

17.04.2024

Revised:

27.06.2024

Accepted:

04.07.2024

*Corresponding Author:

Nagarathna

Email

nagarathna.binu@gmail.com

This is an open access article,
licensed under: [CC-BY-SA](#)



Abstract: Alzheimer's Disease (AD) is a progressive neurological disorder that gradually impairs an individual's memory, reasoning, and ability to perform daily tasks. Early and accurate diagnosis of AD is essential for effective intervention, yet remains challenging due to the complexity of its progression. This study explores the use of an ensemble stacking approach to evaluate the effectiveness of transfer learning techniques in classifying various stages of Alzheimer's disease. Unlike traditional methods that directly analyze raw brain images, this research implements a preprocessing technique using the Markov Random Field method to extract the brain tissues specifically affected by AD. These segmented brain tissues are then utilized to train base models, consisting of three convolutional neural networks (CNNs) with varying configurations. The predictions of these base models are ensembled and further refined through a second-level meta-model to enhance classification accuracy. The proposed ensemble stacking framework was evaluated using an MRI dataset obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which contains images categorized into Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), and Healthy Control (HC) groups. The meta-model demonstrated superior performance, achieving an average accuracy of 97%, along with high precision, recall, and F1 scores. This study highlights the potential of ensemble learning and transfer learning in advancing AD diagnosis, offering a robust and efficient approach for categorizing its various stages based on medical imaging data.

Keywords: Alzheimer's Disease, Brain Tissue Segmentation, Classification Accuracy, Ensemble Stacking, Transfer Learning.



1. Introduction

Memory loss is the main feature of Mild Cognitive Impairment (MCI) and it is the transitional state between brain aging and AD [1] [2]. An individual with amnesic MCI has more memory problems than a normal individual of their age, but their symptoms are not as severe as those of an individual with AD. AD is a common progressive brain disease affecting elderly people over the age of 65. In its early-stage loss of memory is minor but with later stages, individuals are unable to convey the information and fail to respond to their surroundings. Despite its significance currently, there is no cure & the current treatments cannot stop the progression of AD. But early detection of this disease is more essential for progressive treatment. The sensitivity of biomarkers and the accuracy of the detection techniques play an important part in the accurate diagnosis of AD. So, to reduce the progression of AD it is more significant to diagnose it at its earliest stage. A combination of various modalities like EEG, MRI, PET, and other modalities can assist in improving the accuracy rate of the analysis [3] - [7]. To find multi-modular bio-markers an exploration is required on the grouping of modalities. There is a need to develop better diagnostic tools, which is what this thesis addresses. Such tools can be efficiently provided by many new emerging machine learning techniques which aim to higher the accuracy of the prediction of AD such that the appropriate treatments can be effectively provided to the patient. In this phase of work proposed ensemble stacking model for automatic detection of AD by using MRI dataset of Kaggle. The proposed model consists three base models of three different convolutional neural networks. Each model is trained with different features of MRI images affect AD. The main features of human brain affected by AD are gray matters, white matters, Cerebrospinal Fluid (CSF) [8]. In one study, people with dementia and participants with AD disease had their grey matter volumes examined. They discovered that while there was a reduction in grey matter in dementia compared to cognitively normal individuals, there was a noticeably smaller reduction compared to those with AD disease. In order to distinguish AD from a variety of other neurodegenerative dementias, we sought to ascertain the diagnostic value of an expanded panel of CSF biomarkers. In this phase of work instead of feed entire MRI dataset into a model extract these three main features affect the disease are extracted by applying Gaussian distribution and a Markov Random Field (MRF) [9]. The segmented brain image is used to train three different convolutional networks.

The stacking method of ensemble machine learning is one [10]. It comprises combining the predictions from several machine learning models on the same dataset using strategies like bagging and boosting. The answer to this challenge will be another machine learning model that learns when to utilise or trust each model in the ensemble. Instead of using samples from the training dataset, stacking frequently employs other models that are appropriate for the same dataset (e.g., not all decision trees). In contrast to boosting, stacking uses a single model to determine the best way to combine the predictions from the contributing models (instead of a series of models that adjust the predictions of earlier models, for example). Two or more base models, sometimes referred to as level-0 models, make up the architecture of a stacking model, together with a meta-model, also known as a level-1 model, which incorporates the predictions of the base models.

As opposed to Level-0 Models (Base-Models), these models are fitted to the training data and have their predictions created. Model at Level 1 (Meta-Model): A model that determines the best approach to combine the predictions of the basic models. The meta-model is trained using extrapolations of the underlying models using non-sample data. In other words, the base models receive data that wasn't used to train them, which they then use to make predictions and generate the expected outcomes, which are then utilised as input and output pairs in the training dataset that the meta-model is fitted to. The outputs from the underlying models were fed into the meta-model. As a result, the meta-learner model will yield more accurate results than the sub model. Albright [11] and the author, used five CNN architectures to create ensemble models as the foundation model. The accuracy of any model that was taken into consideration by CNN was 93.18%.

2. Literature Review

2.1. Image Preprocessing

In this projected work considered the MRI images collected from ADNI. Before feed images into proposed ensemble model pre-process the image to improve the process performance [12] [13]. The details about dataset preprocessing methods like bias correction, normalization and skull stripping are

discussed are presented in one of our papers sections [14]. In this section focus on segmenting processed brain images into WM, GM and CSF.

The Gaussian distribution and MRF were employed in the proposed work to extract WM, GM, and CSF. The genuine image is reconstructed using the available data and the MRF, a stochastic process that identifies the local properties of an image. A significant technique for modelling spatial continuity and other properties is the MRF of prior contextual data, and even straightforward modelling of this kind can yield insightful results for the segmentation procedure [15]. The MRF is a conditional probability model in which a voxel's likelihood depends on its surroundings. It is equivalent to an energy-function-based Gibbs joint probability distribution [9]. In order to reduce misclassification errors caused by picture noise, MRF models have been effectively included into a variety of brain MRI segmentation techniques. One node in the lattice P can represent each pixel (or voxel) in a picture. Let m be the total number of image elements (for both a 2D and a 3D image), and let x_i represent the intensity value of a single pixel (or voxel) with a position i in an image constructed over a finite lattice P . Let N_i for all $i \in P$ symbolise a neighbouring system for a lattice P , with N_i standing for a restricted area surrounding i that excludes x_i . The group of nodes that are situated around node i and are separated from its centre by a distance equal to square root of x are known as its neighbours. $N_i = \{i' \in P \mid \lfloor \text{dist}(\lfloor \text{pixel} \rfloor_i, \lfloor \text{pixel} \rfloor_{i'}) \rfloor^2 \leq r \wedge i' \neq i\}$ Where $r: r > 0$ is an integer number and $\text{dist}(a, b)$ is the Euclidean distance between adjacent pixels a and b .

Where $\text{dist}(a, b)$ is the Euclidean distance between neighboring pixels a and b and $r \in \mathbb{Z}: r > 0$ is an integer number. The first and the second order neighborhoods are the most commonly used neighborhoods in image segmentation. The first order neighborhood consists of 4 nearest nodes in a 2D image and 6 nearest nodes in a 3D image, while the second order neighborhood consists of 8 nearest nodes in a 2D image and 18 nearest nodes in a 3D image Markov random field model can be represented with a graph $G=(P, N)$, where P represents the nodes and N determines the links (also called edges) that connect the nodes according to the neighborhood relationship.

2.2. Constructing the Training Model

Three distinct CNN architectures for the level 0 base model were taken into consideration in this experiment. Each model is designated as EM1, EM2, or EM3. EM1, EM2, and EM3 are the top, middle, and bottom networks, respectively. The deep learning network is taken into account for the meta model. These architectures have a good classification performance because they have already been trained on ImageNet. Figure depicts the ensemble stacking model that has been proposed. The suggested approach took the preprocessed photos into account. With a ratio of 0.2, the preprocessed images are split into a train dataset and a test dataset. The validation and train subsets of the train dataset are created. For the training dataset, the three base models are each trained independently.

The ensemble predictions are then made using the stacking classifier ensemble approach with meta-learner [12] – [14]:

1) EM1

Three 2D convolutional layers with 32, 64, and 128 filters of size 2×2 make up the EM1 algorithm. Maxpooling, dropout, and batch normalisation layers came after each convolutional layer. To prevent an overfitting of a model, dropout the 0.2 percent of neurons here. For four various classifications of AD disease, including non-demented, very mildly demented, mildly demented, and moderately demented, the classification portion of the system is made up of flat, dense layers with four neurons. Use ReLU and sigmoid activation in convolutional and dense layers, respectively.

2) EM2

The EM2 has three 2D convolutional layers, each with 64, 128, and 256 3×3 -sized filters. Maxpooling layers came after each convolutional layer. At the end of the feature extraction layers, the dropout layer is introduced. The system's classification component is divided into two flat, dense layers, each with 500 and 4 neurons. Convolutional and dense layer employ Relu and Softmax.

3) EM3

Four convolutional blocks with 32, 64, 128 and 256 filters are placed after the EM3's base layer's two convolutional layers with 16 filters. Four convolutional layers with a ReLU activation function make up the convolutional blocks. Four dense layers with 512, 256, 128 and

four neurons with sigmoid activation function make up the classification component of the model.

The layer description of base zero models and meta learners are given in Tables 2,3 and 4 respectively.

Table 2. EM1 Layer Description

Layers	Filter, Stride and Padding	Input Image Size	Output Image Size
Image Input		206x176x3	-
Convolution2D	32 filters with stride 1 and padding 1	206x176x3	206x174x32
Max Pooling1	2x2 max pooling with stride 2 and padding 0	206x174x32	103x87x32
Dropout	0.2 dropout	103x87x32	103x87x32
Batch Normalization	Batch Normalization	103x87x32	103x87x32
Convolution2D	64 filters with stride 1 and padding 1	103x87x32	101x85x64
Max Pooling2	2x2 max pooling with stride 2 and padding 0	101x85x64	50x42x64
Dropout	0.2 dropout	50x42x64	50x42x64
Batch normalization	Batch Normalization	50x42x64	50x42x64
Convolution3D	128 filters with stride 1 and padding 1	50x42x64	48x40x128
Max Pooling3	2x2 max pooling with stride 2 and padding 0	48x40x128	24x20x128
Dropout	0.2 dropout	24x20x128	24x20x128
Batch Normalization	Batch Normalization	24x20x128	24x20x128

Table 3. EM2 Layer Description

Layers	Filter, Stride and Padding	Input Image Size	Output Image Size
Image Input	-	208x176x3	
Convolution2D	64 filters with stride 1 and padding 1	208x176x3	206x174x32
Max Pooling1	2x2 max pooling with stride 2 and padding 0	206x174x32	104x88x64
Convolution2D	128 filters with stride 1 and padding 1	104x88x64	104x88x128
Max Pooling2	2x2 max pooling with stride 2 and padding 0	104x88x128	52x44x128
Convolution3D	256 filters with stride 1 and padding 1	52x44x128	52x44x256
Max Pooling3	2x2 max pooling with stride 2 and padding 0	52x44x256	26x22x256
Dropout	0.2 Dropout	26x22x256	26x22x256

Table 4. EM 3 Layer Description

Layers	Filter, Stride and Padding	Input Image Size	Output Image Size
Image Input	-	208x176x3	
Convolution2D1	32 filters with stride 1 and padding 1	206x174x3	206x174x32
Max Pooling1	2x2 max pooling with stride 2 and padding 0	206x174x32	103x87x64
Convolution2D2	64 filters with stride 1 and padding 1	103x87x64	101x85x64
Max Pooling2	2x2 max pooling with stride 2 and padding 0	101x85x64	50x42x64
Convolution3D	128 filters with stride 1 and padding 1	50x42x64	48x40x128
Max Pooling3	2x2 max pooling with stride 2 and padding 0	48x40x128	24x20x128
Convolution3D	256 filters with stride 1 and padding 1	24x20x128	22x18x256
Max Pooling3	2x2 max pooling with stride 1 and padding 0	22x18x256	11x9x256
Dropout	0.2 Dropout	11x9x256	11x9x256
Batch Normalization	-	11x9x256	11x9x256

Meta model the suggested system's meta model is constructed of four convolutional blocks, each of which has three convolution layers and 32, 64, 128 or 256 filters. Following each convolutional block were the maxpooling layers, then the convolutional layers. Make the model run for different optimizers, such as Adam, Adagrad and Rmsprop, to reduce the system loss. The Adam displays strong performance by providing a 97% accuracy rate.

- After setting up the models, fitting each of the base models to train dataset. The sample snippet for model fitting is given below:
`EM1.fit(x_traindataset,y_traindataset)`
`EM2.fit(x_traindataset,y_traindataset)`
`EM3.fit(x_traindataset,y_traindataset)`
- After training the base models evaluate their accuracy by using the following code sample. Call prediction of EM1, EM2 and EM3 as PrEM1,PrEM2 and PrEM3 respectively.
`PrEM1=EM1.predict_accuracy(x_test)`
`PrEM2=EM2.predict_accuracy(x_test)`
`PrEM3=EM3.predict_accuracy(x_test)`
- Finally compute the average prediction by taking the average of PrEM1,PrEM2 and PrEM3:
`Avgpred=(PrEM1+PrEM2+PrEM3)/3`
- Further train the meta model by AvgPred and Test original dataset and it is given in below:
`Meta.fit(Avgpred,x_traindataset)`
- At last evaluate the meta model by using test dataset:
`pred3=EM3.predict_accuracy(Avgpred ,x_test)`

3. Methodology

The proposed ensemble stacking model is shown in figure 1. The NIFTY MRI brain images collected from Alzheimer's Disease Neuroimaging Initiative (ADNI) [12] are preprocessed by following the process bias correction, Normalization, Skull stripping and segmentation and are discussed in paper [13]. The dataset consists three classes of images AD, MCI and NC. The total number of images belonging to each class are given in Table 1.

Table 1. ADNI MRI Dataset

Classes	ADNI Dataset Count	Male	Female	Female < 70 YO	Male < 70 YO
AD	472	353	119	35	84
MCI	703	369	334	98	236
HC	1108	713	395	3	392

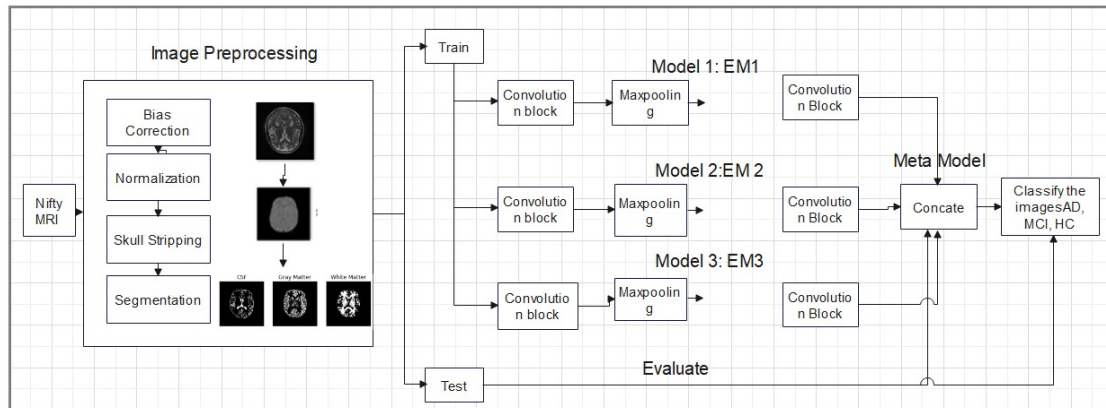


Figure 1. Proposed Model

Further preprocessed segmented images are divided into train and test dataset. The train dataset are used to train three base models of convolutional neural networks differ by configurations. Finally results of three models are fusion by meta model, it also learning from test dataset and classify the input images into AD, MCI, HC. Finally evaluate the system performance by finding statistical parameters.

4. Finding and Discussion

This roposed work was carried out using Google Colab. A kaggle MRI dataset with four distinct classes—non-demented, mildly demented, very mildly demented, and moderately demented—was employed in this experiment. The dataset is split into the train dataset and the test dataset. The base models undergo 10 iterations with a 0.0001 learning rate using the train dataset to train them. The categorical cross entropy function is used to change the network weight in the level 0 models after they have been trained using images of AD disease using the Adam optimizer. The parameters considered in this experiment are given Table 5.

Table 5. Parameters

Parameters	Model-1
Optimizer	Adam, Rmsprop, Adagrad,
Activation Function	ReLU, Softmax and Sigmoid
Loss function	Categorical cross entropy
Batch size	15
Dataset	ADNI
Epoch	10
Learning Rate	0.0001
Normalization	Batch Normalization
Pooling	Maxpooling

The output of the stackclassifier, which stacks these three model predictions, is used as an input by the meta learner. Run the meta learner for ten iterations using various optimizers, activation functions, and cross entropy loss functions to make it learn from the base model's prediction output and the test data of the original dataset. According to the performance table, the EM1 model provides 96% accuracy, 97.5% precision, 90.75 percent recall, and 93.7 percent f1 score. The EM2 has an accuracy of 87.5%, a precision of 92.4%, a recall of 83.25 percent, and an F1 score of 85%. However, EM3 displays 95% accuracy, 965 precision, 95% recall, and 96% F1 score. The table below contains the suggested ensemble model's performance measures. The accuracy of the meta model significantly increased as it was trained using ensemble models for prediction and test data, yielding an accuracy of 97%, with 98.25% precision, 96.5% recall, and 97% f1 score. The confusion matrix of ensemble method is given in Table 6.

Table 6. Proposed Model Confusion matrix

HC	MCI	AD	Accuracy	Precision	Recall	F1 Score
506	2	0	97%	95%	100%	97%
8	131	2	97%	99%	92%	95%
17	1	358	97%	99%	94%	96%
Average			97%	98.25%	96.5%	97%

To lower the loss efficiency, the meta model is iterated across 25 epochs using various optimizers, including Adam, Rmsprop, and Adagrad. Adam has superior performance in this regard, and Figure 2 contains the parameter values for accuracy, recall, precision, and f1 score for random epochs with various optimizers

(75, 10)

	Epochs	Optimizer	Activation_Function	Train_Loss	Train_Accuracy	Train_Recall	Val_Loss	Val_Accuracy	Val_Precision	Val_Recall
19	20	Adagrad	relu	0.563182	0.766893	0.947626	1.077080	0.608861	NaN	0.924684
15	16	Rmsprop	relu	0.573549	0.936228	0.972863	1.485669	0.863291	NaN	0.986076
13	14	Rmsprop	relu	0.679027	0.918317	0.973406	1.503256	0.853165	NaN	0.956329
4	5	Adagrad	relu	0.833755	0.662144	0.897965	1.257996	0.577848	NaN	0.887342
22	23	Adam	relu	0.291263	0.949254	0.993216	0.490513	0.936076	NaN	0.985443

Figure 2. Screenshot Generated for Different Optimizers

The screenshots of accuracy and loss history for meta-models are shown in Figure 3 and Figure 4 respectively.

```
[29] history_df = pd.DataFrame(history.history)
      history_df.loc[:, ['loss', 'val_loss']].plot();
```

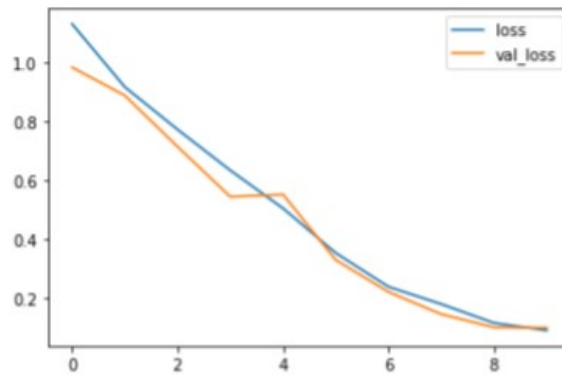


Figure 3. Meta Loss History Graph

```
[30] history_df = pd.DataFrame(history.history)
      history_df.loc[:, ['accuracy', 'val_accuracy']].plot();
```

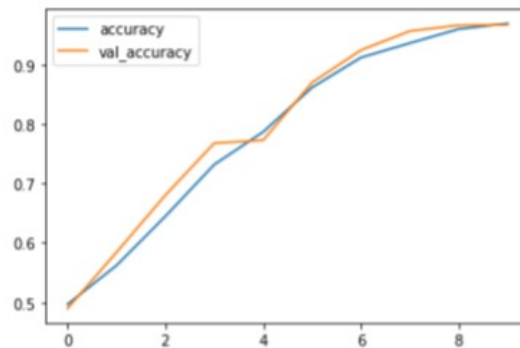


Figure 4. Meta Learner Accuracy History Graph

5. Conclusion

Few samples of the MRI dataset was utilized for testing the projected algorithm. Here implemented ensemble stacking model to improve the performance of the system by extracting the feature of input images at two level of classification. At first level of prediction base model is implemented with three convolutional models and at the final level meta model is implemented for final prediction. The model shows an accuracy of 97% and 98.25% of precession. Projected model is executed utilizing python that was accessible openly. This investigation was led on Intel ® center ™ i7-8750H, 16GB RAM, 64bit Operating framework, NVIDIA GPU.

References

- [1] H. D. Yang, S. B. Lee, D. H. Kim, and L. D. Young, "History of Alzheimer's Disease," *Dementia and Neurocognitive Disorders*, vol. 15, no. 4, pp. 115, 2016.
- [2] Medscape, "What is mild cognitive impairment (MCI) in patients with Alzheimer disease (AD)," [Online]. Available: <https://www.medscape.com/answers/1134817-22560/what-is-mild-cognitive-impairment-mci-in-patients-with-alzheimer-disease-ad>. [Accessed: Jan. 18, 2024].
- [3] J. Lieff, "Neuron Networks in Healthy and Diseased Brains," [Online]. Available: <https://jonlieffmd.com/tag/prominent-brain-networks-are-rich-club>. [Accessed: Jan. 18, 2024].
- [4] D. Vaghari, E. Kabir, and R. N. Henson, "Late combination shows that MEG adds to MRI in classifying MCI versus controls," *Neuroimage*, vol. 252, 2022.
- [5] A. Alberdi, A. Aztiria, and A. Basarab, "On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey," *Artificial Intelligence in Medicine*, vol. 71, pp. 1–29, 2016.
- [6] E. Teng, J. M. Ringman, L. K. Ross, R. A. Mulnard, M. B. Dick, G. Bartzokis, et al., "Diagnosing depression in Alzheimer disease with the national institute of mental health provisional criteria," *The American Journal of Geriatric Psychiatry*, vol. 16, no. 6, pp. 469–477, 2008.
- [7] R. İnce, S. S. Adanır, and F. Sevmez, "The inventor of electroencephalography (EEG): Hans Berger (1873–1941)," *Child's Nervous System*, pp. 1–2, 2020.
- [8] M. C. Norton, C. J. Clark, J. T. Tschanz, P. Hartin, E. B. Fauth, J. A. Gast, et al., "The design and progress of a multidomain lifestyle intervention to improve brain health in middle-aged persons to reduce later Alzheimer's disease risk: The Gray Matters randomized trial," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 1, no. 1, pp. 53–62, 2015.
- [9] P. Zhang, S. Lin, J. Qiao, and Y. Tu, "Diagnosis of Alzheimer's Disease with Ensemble Learning Classifier and 3D Convolutional Neural Network," *Sensors (Basel)*, vol. 21, no. 22, pp. 7634, 2021.
- [10] A. Chandra and S. Roy, "On the detection of Alzheimer's disease using support vector machine based majority voter classifier," *Proceedings of the 2021 Conference*, 2021.
- [11] J. Albright, "Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, pp. 483–491, 2019.
- [12] Alzheimer's Disease Neuroimaging Initiative, [Online]. Available: <https://adni.loni.usc.edu/>. [Accessed: Dec. 11, 2023].
- [13] C. R. Nagarathna and M. Kusuma, "Classifying the stages of Alzheimer's disease by using multi-layer feed forward neural network," *Procedia Computer Science*, vol. 218, pp. 1845–1856, 2023.
- [14] B. Solano-Rojas and R. Villalón-Fonseca, "A low-cost three-dimensional DenseNet neural network for Alzheimer's disease early discovery," *Sensors*, vol. 21, no. 4, 2021.
- [15] S. F. Ahmed, M. S. B. Alam, M. Hassan, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, pp. 13521–13617, 2023.