Improving Alzheimer's Disease Diagnosis on Brain MRI Scans with an Ensemble of Deep Learning Models

Md. Sazid Reza
Department of Computer Science and Engineering
Rajshahi University of Engineering and Technology
Rajshahi, Bangladesh
dihansazid@gmail.com

Mir Md. Jahangir Kabir
Data Science and Innovation
TD School
University of Technology Sydney
Sydney, Australia
mmjahangir.kabir@gmail.com

Md. Abdur Rakib Mollah
Department of Computer Science and Engineering
Rajshahi University of Engineering and Technology
Rajshahi, Bangladesh
rakib1703115@gmail.com

***Abstract*—Alzheimer's disease (AD) is a widespread neurological condition affecting millions globally. It gradually advances, leading to memory loss, cognitive deterioration, and a substantial decline in overall quality of life for those affected. AD patients experience memory decline, eroding cherished memories and straining relationships, while daily tasks become challenging. Numerous investigations have been conducted in this field, as the timely identification of Alzheimer's disease at its initial stage is of the utmost importance. A major limitation in this field is the predominant emphasis on using single fine-tuned CNN architecture or comparing pre-trained and custom CNN models for Alzheimer's detection, often on small datasets, which neglects a more comprehensive approach. Using smaller datasets can negatively impact deep learning modeling accuracy due to overfitting, limited representation, and poor generalization. This study addresses the current research problems and proposes an ensemble approach that combines predictions from various pre-trained models, including DenseNet-121, EfficientNet B7, ResNet-50, VGG-19, and also from a Custom CNN. The model averaging ensemble method was applied, a subset of the Stacking Ensemble, to two ADNI datasets, with Dataset-I being the larger. The goal was to assess the efficacy of this ensemble approach for accurate multiclass classification on ADNI datasets, where it successfully identified all classes despite differing sample volumes. A vast experiment was conducted on two distinct and widely recognized real-world datasets, resulting in accuracies of 99.96% and 98.90% respectively. Finally, the outcome of the research compared with recent research findings demonstrates the potential of our approach in advancing Alzheimer's disease detection by outperforming other benchmark approaches by a significant margin.**

***Index Terms*—ADNI, Model Averaging, Predictions, Ensemble, DenseNet-121, EfficientNet B7, ResNet-50, VGG-19.**

I. INTRODUCTION

The complex workings of the human brain have intrigued researchers and scientists for a significant duration, providing a glimpse into the intricacies of cognition, memory, and personality. Alzheimer's disease is only one of the numerous difficulties that the brain may encounter, serving as a sobering reminder of its frailty. Detecting Alzheimer's at an early stage can be challenging due to the subtle fear many patients have about the disease, often attributing memory loss to normal aging, fearing societal judgment and discrimination, and valuing their independence, which can lead to delayed medical treatment. From 1990 to 2019, there was a noticeable increase in both the incidence and the occurrence rates of Alzheimer's disease and other forms of dementia, with a respective rise of 147.95% and 160.84%.[1]. Based on current estimations, the global prevalence of dementia is approximately 50 million individuals, with the United States accounting for approximately 6.5 million of these cases. AD patients display decreased memory network activity and reduced occipital alpha activity, indicating impairment in memory processes and visual processing regions [2]. Neuroimaging has transformed Alzheimer's disease understanding by detecting early brain changes before symptoms. The majority of studies on the detection of Alzheimer's disease use a binary classification system, with the categories AD (Alzheimer's Disease) and CN (Cognitively Normal). The task of multiclass classification involves the classification of several categories, including MD (Mild Demented), ND (Non-Demented), MOD (Moderate Demented), and VMD (Very Mild Demented). These four classes are the extensions of two foundation classes AD and CN. The three classes used by the other fields of research are AD, CN, and MCI, where MCI stands for Mild Cognitive Impairment and can be further subdivided into EMCI, LMCI, and MCI.

Alzheimer's disease (AD) develops in stages, starting with Cognitively Normal (CN), or no cognitive impairment. Certain individuals may experience a transition from CN to Early Mild Cognitive Impairment (EMCI). This stage signifies a gradual deterioration in cognitive functioning. As the disease progresses, Late Mild Cognitive Impairment (LMCI) emerges, when deficiencies are increasingly severe as cognitive changes become apparent. Mild Cognitive Impairment (MCI) eventually develops, causing serious cognitive deterioration. In the final stage, people with Alzheimer's (AD) develop moderate to severe dementia, severe communication difficulties, weight loss and need intensive care.

This paper is structured into the following sections: Section

II provides a comprehensive overview of the existing literature on Alzheimer's disease. Section III presents a concise description of the methods and datasets related to our proposed approach. Section IV elaborates on the experimental design and provides a detailed comparative analysis of our proposed ensemble approach. Finally, some conclusions were stated based on the performance of our method in Section V.

II. Related Works

In 2023, Shamrat et al. [3] mri employed a comparative study of several pretrained models on an ADNI dataset of six classes (AD-EMCI-MCI-LMCI-SMC-CN). InceptionV3 was proven to be optimal and later finetuned with RMSprop optimizer resulting in an accuracy of 98.67%. Data augmentation was used to transform the dataset to a balanced dataset of 60000 images of 6 classes.

Garg et al. [4] performed classification on a 4-D dataset transformed into a 2-D dataset through grayscale image conversation. The acquired accuracy was around 97.52%. Rathore [5] performed a 4-way classification with DenseNet-121, DenseNet-169, and DenseNet-201. The obtained accuracy was 91.83%, 93.045%, and 94.079%.

Janghel [6] performed a 3-D to 2-D conversion on the ADNI dataset of two classes and used the VGG-16 model. The study demonstrates a classification accuracy of 99.95% on average for the fMRI dataset and achieves a mean accuracy of 73.46% for the PET dataset.

In 2023, Mujahid et al. [7] employed several ensemble approaches on a four-class dataset. They used augmentation to mitigate the class imbalance issue. On the augmented balanced dataset they acquired accuracies of 97.35% with learning rate of .0001 with the VGG-16+EfficientNet-B2 ensemble. VGG-16-DenseNet-121 ensemble model achieved a 95.56 accuracy in their case.

Tanveer et al. [8] used a 'Deep Transfer Ensemble' technique for Alzheimer's disease detection in 2022. An accuracy of 99.09% on CN-AD AND 98.71% on MCI-AD binary classification was received. A dataset comprising 813 3D-MRI scans was utilized here. The foundational architecture for this transfer-learned approach was built upon the pre-trained model VGG-16.

Jathinsai et al. [9] developed a CNN technique for multiclass classification on an ADNI dataset of 6400 MRI images that is labeled as **Dataset-II** in our study. Data augmentation was avoided here and they achieved an accuracy of 94.92%.

In 2023, Nagaranjan et al. [10] employed an ensemble strategy on the Dataset-II and obtained an accuracy of 98.40%. Image Augmentation was used to augment the number of images from 6400 to 12800. The prediction of the weighted ensemble technique was performed using Inception V3, DenseNet-121, and AlexNet. The model accuracies ranged from 88% to 92% before the ensemble.

The main contributions of this study are as follows: a novel highly efficient approach called model averaging ensemble is proposed to classify the five and four classes of Alzheimer's disease in the ADNI Datasets. Our strategy combines pre-trained models with a Custom Convolutional Neural Network (CNN) model. In this context, we apply transfer learning to pre-trained models such as DenseNet-121, ResNet-50, EfficientNet B7, and VGG-19. These models undergo fine-tuning as feature extractors to adapt their knowledge from pre-training to specific image-related tasks, thereby enhancing their performances. This approach leverages the models' expertise learned from large datasets like ImageNet for improved performance on our target classification task. The majority of research undertaken on this particular topic is primarily centered around a singular model-based accuracy, with a notable absence of the **Ensemble** approach. The absence of the employment of the Custom CNN model alongside pre-trained models for classification tasks was also identified as a gap in the existing literature.

III. Methodologies

A. Dataset Description

The ADNI database was the most prevalent during our analysis. Two ADNI datasets were used in our approach due to their larger size to maximize accuracy. ADNI data are typically accessible via the LONI Image and Data Archive (IDA), a secure repository for research data. Furthermore, due to the nature of our task being a classification challenge rather than a segmentation problem, we were compelled to utilize preprocessed magnetic resonance imaging (MRI) images. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multisite longitudinal observational research of Alzheimer's disease, mild cognitive impairment, and healthy elderly people [11]. Axial plane MRI scans were utilized from ADNI as they enable the identification of lesions, ventricular enlargement, and global brain changes. The MRI scans were T1-weighted, with enhanced fatty tissue signals while suppressing water signals, resulting in black fluid shading.

1) Data acquisition: A five-class classification system consisting of the following classes: Alzheimer's Disease (AD), Cognitively Normal (CN), Mild MCI (MCI), Late MCI (LMCI), and Early MCI (EMCI) was employed. About 18775 Brain MRI images were obtained from **Dataset-I** from the ADNI database [11]. There were 8346 AD cases, 8650 CN cases, 1155 MCI cases, 480 EMCI cases, and 144 LMCI cases. 6400 MRI Scans were acquired from **Dataset-II** for four-class classifications, which was a public dataset collected from ADNI by various contributors [12]. About 896 were for cases of mild demented (MD), 64 were for cases of moderate demented (MOD), 2240 were for cases of very mild demented (VMD), and 3200 were for cases of non-demented (ND). Data augmentation can distort brain MRI images, potentially compromising the accuracy of disease detection and progression monitoring. Our study avoided augmentation because we wanted to construct an ensemble on imbalanced datasets, which are more prevalent in real-world practical data.

2) Data Preprocessing: Preprocessed T1-weighted MRI images were selected from ADNI as they are essential for identifying gray and white matter in brain scans, detecting abnormalities, and estimating brain volume. Initially, the images were of 224*224 size, but they were resized to 150*150 as resizing MRI images to 150x150 in Alzheimer's detection studies offers computational efficiency, reduces noise, and strikes a balance between accuracy and resource efficiency.

B. Convolutional Neural Network

Deep learning models known as Convolutional Neural Networks (CNNs) are specifically crafted for the processing of images and videos They comprise distinct layers: convolution, pooling, dense (fully connected), dropout, and fully connected layers. Convolution layers execute convolutions, a specific linear operation for feature extraction. This involves using a small numerical array called a kernel on an input represented as a tensor [13]. After convolution, an activation function introduces non-linear transformations for capturing intricate data patterns and relationships. [14]. The pooling layer gradually reduces input image spatial size to simplify network computations. Pooling reduces space and provides only important data to CNN layers [15]. A dense layer, known as a fully connected layer is used in the latter stages of a neural network to change the dimensionality of the output from the previous layer. Dropout is ignoring neurons during random neuron training. Ignoring means not passing these units during forward or backward pass. Our study used five CNN-based models: four pre-trained on ImageNet (DenseNet-121, EfficientNet B7, VGG-19, and ResNet-50) and one Custom CNN model.

C. DenseNet-121

DenseNet, a deep learning architecture developed by Huang et al. [16], has greatly influenced computer vision research. DenseNet addresses the challenge of vanishing gradients during deep network training by incorporating dense connections. It is a 121-layer deep architecture with around 7 million parameters which signifies the total number of learnable weights and biases within the network. In our research, DenseNet-121 was fine-tuned by rendering its layers non-trainable, followed by the addition of a dropout layer for feature extraction and overfitting prevention. Afterward, feature maps were compressed into singular values for each channel through a global average pooling layer, and classification was executed by incorporating a dense layer with softmax activation. The model underwent transfer learning to adapt its classification capabilities to datasets containing five and four classes.

D. EfficientNet B7

EfficientNet B7 is an 813-layer architecture with around 64 million parameters. The compound scaling approach of EfficientNet B7 involves simultaneous adjustments of the model's depth, width, and resolution. The inverted bottleneck structure optimizes computing efficiency and representation power for feature extraction. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), EfficientNet B7 excelled with an accuracy of 85.8% for the first category and 98.8% for the fifth category [17]. The base EfficientNet B7 model was fine-tuned with these custom layers, employing the Adam optimizer and categorical cross-entropy loss, and performance was assessed using specified metrics.

E. Custom CNN

The model began with a Convolutional Layer using 64 filters to extract initial features like edges, corners, and simple textures. A MaxPooling Layer followed downsampling data and maintaining essential traits while controlling complexity. A second Convolutional Layer with 32 filters enhanced the model's ability to recognize complex patterns indicative of early signs of Alzheimer's disease, such as subtle changes in the shape or size of specific brain structures or the presence of abnormal structures like amyloid plaques. Another MaxPooling Layer further sharpened the focus on important features. The third Convolutional Layer, with 32 filters, was

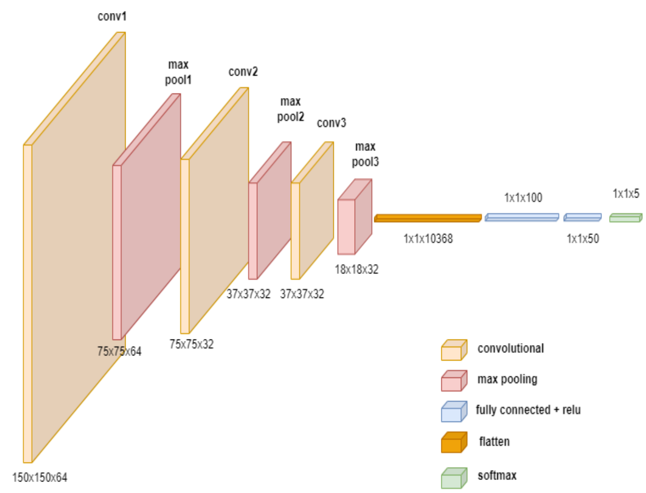

Fig. 1: Custom CNN Architecture

used to catch more advanced disease patterns like biomarkers or unique spatial correlations among diseased brain regions. After flattening the data, two dense layers with 100 neurons, 50 neurons, and ReLU activation facilitated complex feature combinations and pattern recognition. The last dense layer incorporated the softmax activation function for our classification task. Fig.1 illustrates the model architecture along with its filter size and layer dimensions.

F. ResNet-50

ResNet-50 is a 50-layer-deep network trained on 23 million parameters. It was pre-trained on the ImageNet Dataset. Unique features include a residual block that allows information to avoid utilizing skip connections or shortcuts. ResNet-50 steps in to solve the vanishing gradient issue since VGG-19 is unable to. Also, it possesses a less complex architecture than VGG-19. ResNet-50 was also finetuned like the previous models for our specific classification tasks.

G. VGG-19

VGG-19 consists of 16 convolutional layers and 3 fully connected layers. The layers were made non-trainable for feature extraction, integrated into a sequential model with flattening, softmax dense layer, and optimized using Adam with categorical cross-entropy loss. The problem with VGG-19 was the vanishing gradient problem (when gradients become too small during deep neural network training, hindering effective weight updates in early layers).

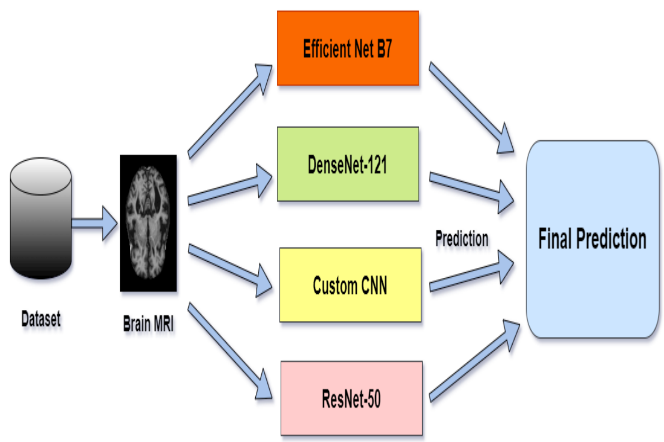

Fig. 2: Proposed Averaging Ensemble Prediction Model

H. Ensemble Learning

Ensemble learning leverages multiple diverse models to generate a more accurate and robust prediction [18]. The final prediction is achieved using predictions from the base models and the corresponding target values. It learns to combine these predictions into a single forecast that surpasses the aggregate. Harnessing the collective strengths of multiple CNN architectures, we employed a model averaging ensemble to finalize predictions for each test set example, assigning equal importance to each prediction from the four best-performing models. This process involved taking the average prediction for each instance in our test dataset for each model. For each example, we chose the class exhibiting the highest average probability, a selection subsequently validated by the predictions made by the models. Each model used in the final ensemble under this ensemble technique holds an equivalent and substantial role. Fig. 2 illustrates the model averaging ensemble technique employed in our study and also specifies the top four predictions used in our case.

IV. Experimental Analysis

A. Experimental Setup

In this study, version 3.7 of the Python programming language was utilized, and the Kaggle Cloud platform was utilized to run the GPU Tesla P-100.

B. Experimental Design

The default batch size, 32 was used in our case. The models were trained for 50 epochs because larger epochs can assist neural networks in catching intricate patterns in difficult datasets, improving accuracy and generalization. A dropout rate of 0.5 was chosen to compel neural networks to learn more robust and broad features by randomly deactivating half of their neurons during training. To prevent overfitting, the categorical cross-entropy was chosen as the loss function. Adam (Adaptive Moment Estimation) optimizer was utilized with a default learning rate of 0.001. This optimization algorithm is employed for training machine learning models, especially neural networks. It adapts the learning rate for each parameter by considering both its prior gradients and squared gradients, leading to quicker convergence and efficient management of various feature scales. For the two datasets, a fully connected layer was added with five and four neurons to all our models to finetune them for five and four-class classifications.

C. Performance metrics

Five performance metrics were employed for a comprehensive performance evaluation. Precision and Recall were used to examine the balance between true positives and false positives/negatives in the evaluation, assessing the compromises between these outcomes. Recall evaluated the model's capacity to identify true positives, whereas Precision evaluated the accuracy of positive predictions. Additionally, the F1 Score provided a fair evaluation of a model's performance because it combines Precision and Recall into a single metric. In addition, the Loss metric, which was mainly used during training, measured prediction errors and directed attempts to reduce them, ultimately improving model accuracy.

D. Result Analysis

A split of 80:10:10 was employed on Dataset-I, with 80% of images going to training, 10% to testing, and the rest to validation. An 80:20 train-test split was employed on Dataset-II and in this case, 10% of the train data went into the validation set. Conversely, the test set has no influence on training and is entirely independent. Shuffling was used to randomly organize training data before feeding it to a model to avoid biases. The final predicted values on **Dataset-I** are depicted below. TABLE I shows that ResNet-50 was the best-performing model in our case. Although ResNet-50 is 50 layers deep whereas EfficientNet B7 is 813 layers deep and DenseNet-121 is 121 layers deep and they both are more deeper and complex architectures. Both DenseNet-121 and EfficientNet couldn't outperform ResNet-50 as they are known for achieving high performance on large and varied datasets. Our Custom CNN model was 9 layers deep and simple in architecture.

TABLE I: Model performance metrics (Dataset-I)

Model	F1 Score	Recall	Precision	Accuracy
DenseNet-121	0.995495	0.989498	0.987203	0.987961
Custom CNN	0.997670	0.994175	0.994218	0.994175
EfficientNet B7	0.997410	0.996276	0.995084	0.995293
ResNet-50	0.998291	0.995728	0.996291	0.995728
VGG-19	0.986175	0.965437	0.966175	0.965437

Compared to pre-trained models, our Custom CNN achieved the third-best accuracy of 99.41% proving custom structures can be advantageous in some situations. Additionally, VGG-19 produced poor performance, therefore it was not included in the final prediction. Fig. 3 and Fig. 4 depict the accuracy curves of Custom CNN and ResNet-50.

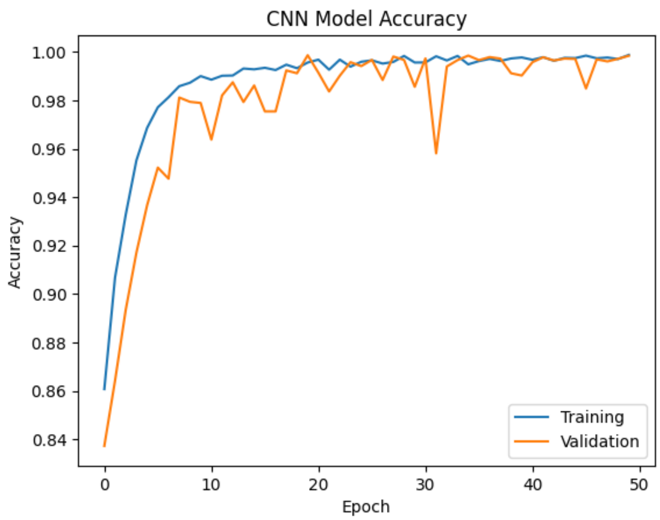

Fig. 3: Training and validation accuracy of Custom CNN model (Dataset-I)

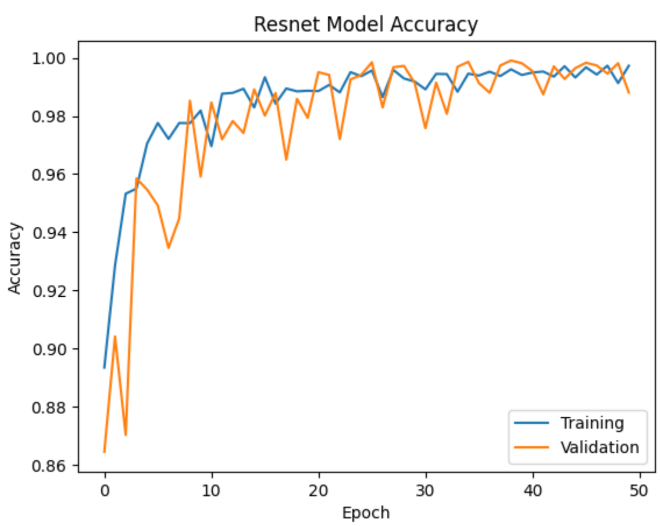

Fig. 4: Training and validation accuracy of ResNet-50 model (Dataset-I)

The validation accuracy exhibits minor oscillations in both models, but these fluctuations diminish over time, indicating a convergence toward the optimal solution after 50 epochs. Because our models responded likewise, we decided to adopt a model averaging ensemble method rather than a weighted averaging ensemble. This process involved averaging the predicted probabilities for each class from the four models. We selected the class with the highest average probability for each example, representing the most likely class as validated by all the models in the ensemble After the merging of predictions of the four models, the final prediction on the test dataset obtained an astonishing **99.96%** accuracy on **Dataset-I** which surpassed all the five initial models and also the approaches from related works. Fig. 5 depicts the confusion matrix of the final ensemble on Dataset-I.

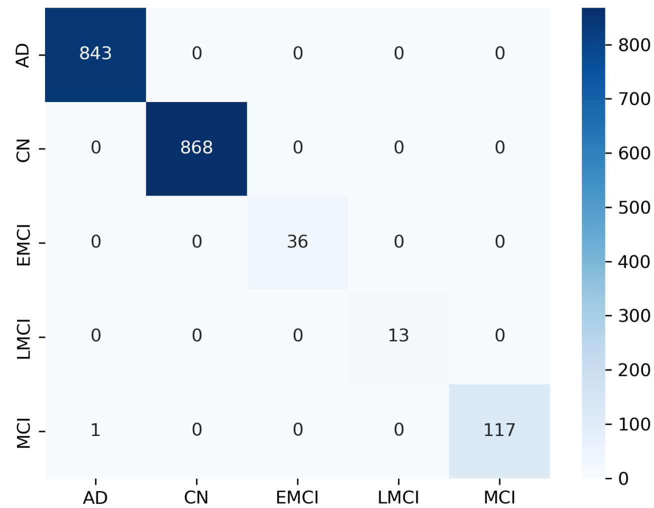

Fig. 5: Ensemble Confusion Matrix(Dataset-I)

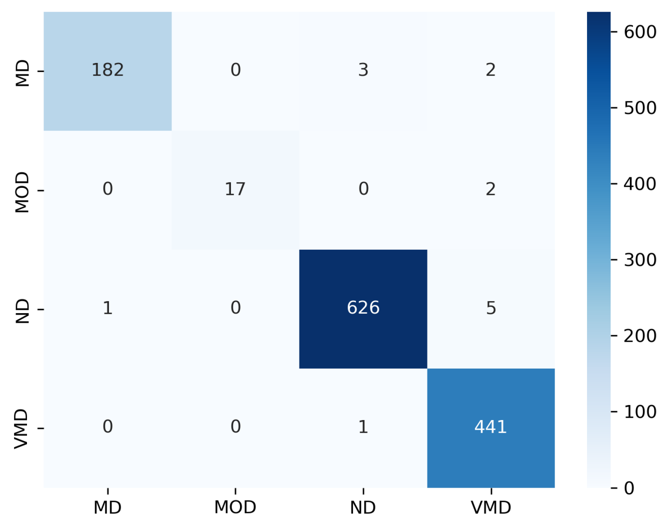

Fig. 6: Ensemble Confusion Matrix(Dataset-II)

In the scenario of **Dataset-II**, it was observed that all the models exhibited comparable behavior, resulting in a final ensemble accuracy of **98.90%** which is higher than the resultant accuracies from the related studies that have been conducted on the same dataset after augmenting the dataset. The performance of our models was superior on Dataset-I due to the availability of a bigger training set, which allowed for better model fitting. Fig. 6 depicts the confusion matrix of the final ensemble on Dataset-II. The confusion matrix validates the accuracy. Table II shows a comparison table with relevant research, and it is crystal clear from this table that our approach has outperformed all approaches in terms of accuracy. The majority of the relevant papers were restricted to classifying only three classes whereas we ran classification tasks on five (AD-LMCI-MCI-EMCI-CN) and four (MD-MOD-ND-VMD) classes.

Method	Classes	Accuracy
Shamrat et al. [3]	(AD-EMCI-MCI-LMCI-SMC-CN)	98.67%
Garg et al. [4]	(AD-CN)	97.52%
Rathore [5]	(AD-LMCI-MCI-NC)	94.08%
Janghel [6]	(AD-CN)	99.95%
Tanveer et al. [8]	(AD-CN)	99.09%
Mujahid et al. [7]	(MD-MOD-ND-VMD)	97.35%
Jathinsai et al. [9]	(MD-MOD-ND-VMD)	94.92%
Nagaranjan et al. [10]	(MD-MOD-ND-VMD)	98.40%
Ensemble Approach (Dataset-II)	**(MD-MOD-ND-VMD)**	**98.90%**
Ensemble Approach (Dataset-I)	**(AD-MCI-LMCI-EMCI-CN)**	**99.96%**

TABLE II: Comparision of performances among our proposed approach and existing approaches

V. Conclusion

In this study, ADNI datasets were utilized. The obtained accuracy scores for the models DenseNet-121, Custom CNN, EfficientNet B7, ResNet-50, and VGG-19 were 98.79%, 99.41%, 99.52%, 99.57%, and 96.54%, correspondingly. The ResNet 50 model exhibited superior performance, while the VGG-19 model revealed comparatively inferior performance on the datasets. In this study, a model-averaging ensemble methodology was employed by incorporating the top four models. The resultant accuracy achieved was an astounding **99.96%** on **Dataset-I** with five classes and **98.90%** on **Dataset-II** with four classes. After comparing with the related works it was clear that our final ensemble approach had outperformed all the benchmark approaches. Also, our research proposes a unique approach, as pre-trained CNN models were evaluated with a Custom CNN model on two datasets, then combining the best predictions created the optimal ensemble accuracy that outperformed the studies conducted on ADNI. For future work, there is a new window of research opportunity if we conduct our research with T2-weighted images, as ours was conducted on T1-weighted images only. Also, To create a more reliable system, clinical data from patients can be added in addition to imaging data and multimodal data. Addressing the overfitting issue is also a part of our future research plan.

References

[1] X. Li, X. Feng, X. Sun, N. Hou, F. Han, and Y. Liu, "Global, regional, and national burden of alzheimer's disease and other dementias, 1990–2019," *Frontiers in Aging Neuroscience*, vol. 14, p. 937486, 2022.

[2] Y. Aoki, R. Takahashi, Y. Suzuki, R. D. Pascual-Marqui, Y. Kito, S. Hikida, K. Maruyama, M. Hata, R. Ishii, M. Iwase *et al.*, "Eeg resting-state networks in alzheimer's disease associated with clinical symptoms," *Scientific Reports*, vol. 13, no. 1, p. 3964, 2023.

[3] F. J. M. Shamrat, S. Akter, S. Azam, A. Karim, P. Ghosh, Z. Tasnim, K. M. Hasib, F. De Boer, and K. Ahmed, "Alzheimernet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images," *IEEE Access*, vol. 11, pp. 16 376–16 395, 2023.

[4] R. Garg, R. R. Janghel, and Y. Rathore, "Enhancing learnability of classification algorithms using simple data preprocessing in fmri scans of alzheimer's disease," in *Advances in Automation, Signal Processing, Instrumentation, and Control: Select Proceedings of i-CASIC 2020*. Springer, 2021, pp. 1055–1063.

[5] Y. K. Rathore and R. R. Janghel, "8 prediction of stage of alzheimer's," *Next Generation Healthcare Systems Using Soft Computing Techniques*, vol. 8, p. 4, 2022.

[6] R. Janghel and Y. Rathore, "Deep convolution neural network based system for early diagnosis of alzheimer's disease," *Irbm*, vol. 42, no. 4, pp. 258–267, 2021.

[7] M. Mujahid, A. Rehman, T. Alam, F. S. Alamri, S. M. Fati, and T. Saba, "An efficient ensemble approach for alzheimer's disease detection using an adaptive synthetic technique and deep learning," *Diagnostics*, vol. 13, no. 15, p. 2489, 2023.

[8] M. Tanveer, A. H. Rashid, M. A. Ganaie, M. Reza, I. Razzak, and K.-L. Hua, "Classification of alzheimer's disease using ensemble of deep neural networks trained through transfer learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1453–1463, 2022.

[9] D. Jathinsai, S. Narisetti, A. V. Pilla *et al.*, "Cnn technique for detecting alzheimer's disease by using mri images," 2022.

[10] V. Nagarajan, L. Venkataramana, V. Prasad, V. Narasimhan, and S. Santhanakrishnan, "Deployable and weighted ensemble-based deep learning model for alzheimer's disease detection," 2023.

[11] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[12] Alzheimer's Dataset—Kaggle Available online: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images [Accessed: july 6,2023].

[13] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, pp. 611–629, 2018.

[14] R. Tharsanee, R. Soundariya, A. S. Kumar, M. Karthiga, and S. Sountharrajan, "Deep convolutional neural network–based image classification for covid-19 diagnosis," in *Data Science for COVID-19*. Elsevier, 2021, pp. 117–145.

[15] V. Lakkavaram, L. Raghuveer, C. Satish Kumar, G. Sai Sri, and S. Habeeb, "A review on practical diagnostic of tomato plant diseases," *Suraj Punj J Multidiscip Res*, vol. 9, pp. 432–435, 2019.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[18] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.