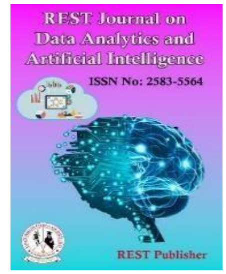
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/16



Enhanced Multi-Class Classification of Alzheimer's Disease Using Ensemble Method on MRI Images

***Hasritha Reddy Jangaon, Preethika Gollapally, Sridhar Reddy Gogu**
School of Engineering Anurag University, Hyderabad, India.
**Corresponding Author Email: reddyhasritha1704@gmail.com*

Abstract. *Alzheimer's disease is a brain disorder that slowly makes it hard for people to remember things, think clearly, and do daily activities. It gets worse over time and can make life very difficult. To help doctors find the disease early and give the right treatment, we need a good way to check brain scans (MRI images) and tell what stage the disease is in. But many computer models used for this have problems. Sometimes, they do not have enough data to learn properly, or the data is not balanced, meaning some stages of the disease have more images than others. Some models are also too complicated and do not work well for all people. This project builds a strong system to classify Alzheimer's disease into four stages: non-demented, very mild demented, mild demented, and moderate demented. The system uses MRI images from the ADNI dataset and improves accuracy by fixing the data imbalance problem using data augmentation. Instead of using only one model, we use an ensemble method, which means we combine multiple pre-trained models to get better results. We use three well-known deep learning models—EfficientNetB3, ResNet50, and DenseNet121—and combine their results using a method called majority voting. This way, if one model makes a mistake, the others can correct it. Our system achieves 99% accuracy, which is better than using any of these models alone. This means doctors can trust the system more to help diagnose Alzheimer's early and correctly. By solving problems with older models, this approach makes Alzheimer's detection better, helping patients get the right treatment sooner and making research on the disease easier.*

Keywords: *Alzheimer's, MRI Images, Data Augmentation, Ensemble Method, Pre-Trained, Majority Voting.*

1. INTRODUCTION

Many people around the world suffer from brain and nerve problems. These problems, called neurological disorders, affect the brain, spinal cord, and nervous system. They can cause serious issues with memory, thinking, emotions, and movement. People who have these problems often find it hard to do daily tasks, and their families also struggle to take care of them. These illnesses also put a big burden on hospitals and doctors because they need a lot of treatment and care [1]. One of the most serious brain diseases is Alzheimer's disease. It is a problem that slowly damages a person's memory and thinking skills [2]. Over time, people with Alzheimer's find it hard to remember things, make decisions, and even recognize their family members. This disease does not have a cure, and once it starts, it keeps getting worse [3]. Alzheimer's is most common in older people, but in some cases, it can also happen to younger adults. Since this disease affects millions of people, scientists and doctors are working hard to find better ways to detect it early and treat it properly [4-5]. Alzheimer's happens because of harmful changes in the brain. Two main problems cause this disease: amyloid plaques and neurofibrillary tangles. These are abnormal structures that build up in the brain and cause brain cells to die. When this happens, the brain starts shrinking, and the person begins to forget things and behave differently [6]. Doctors classify Alzheimer's into different stages based on how much damage has happened in the brain [7]. The four main stages are non-demented, very mild demented, mild demented, and moderate demented. In the non-demented stage, the person does not show clear signs of memory loss yet. In the very mild demented stage, small memory problems start, but they are not very noticeable [8]. In the mild demented stage, the person begins to struggle with daily activities, like managing money or remembering important events. In the moderate demented stage, memory loss and confusion become severe, and the person may need help from family members for simple tasks. Doctors use different tests to find out if someone has Alzheimer's. The most common tests include brain scans like MRI, PET, and CT scans [9-10]. These scans help doctors see changes in the brain. Another method is neuropsychological tests, where the patient answers questions to check their memory, attention, and problem-solving skills. While these methods are useful, they have some problems. Many times, brain scans do not catch the disease in its early stages. By the time it is detected, the brain is already badly damaged. This makes early treatment difficult. Also, these tests can be expensive and are not available in all hospitals. To overcome these problems, scientists are now using Artificial Intelligence (AI) to help doctors find Alzheimer's disease earlier and more accurately [11]. AI allows computers to analyze brain scans and detect tiny changes that humans might miss. One type of AI called Machine Learning (ML) is used for this [12]. ML teaches computers to recognize patterns in large amounts of data, like MRI images of different Alzheimer's stages. Deep Learning (DL), a more advanced type of ML, has been very successful in detecting Alzheimer's. One of the best deep learning methods is Convolutional Neural Networks (CNNs). CNNs are computer models that can analyze pictures and recognize important details. In Alzheimer's research, CNNs are trained to study MRI scans and identify whether a person has Alzheimer's and what stage it is. CNNs have been shown to give better results than human experts in some cases. However, there are still many challenges with using deep learning models for Alzheimer's detection [13]. Some of these challenges include not having enough data, unbalanced data, complex models, and high computational costs. AI models need a lot of images to learn properly. Many hospitals do not have enough MRI scans to train these models well. Some stages of Alzheimer's have more MRI images than others, making it hard for the model to learn all stages equally. Some deep learning models are too difficult to understand and use. Also, training these models requires powerful computers, which not all hospitals can afford. To solve these problems, this study proposes a new method to improve Alzheimer's classification using deep learning. Instead of using just one model, we combine multiple models together. This method is called ensemble learning [14]. By using EfficientNetB3, ResNet50, and DenseNet121, we can take advantage of the strengths of each model. These models are well-known for analyzing images and are already trained on large datasets. We combine their predictions using a technique called majority voting, where the final decision is based on what most models predict. This method improves accuracy and makes the system more reliable. Another way we improve the model is by using data augmentation. This technique helps solve the problem of unbalanced data. It works by creating more training images by rotating, flipping, or slightly changing existing MRI images. This way, the model learns better and gives better predictions for all Alzheimer's stages. The main goal of this study is to create a better and more efficient system for detecting Alzheimer's disease from MRI images. By combining deep learning models and improving data balance, we aim to develop a system that can be used in hospitals to help doctors make quick and accurate diagnoses [15]. This report is divided into different sections. The first section reviews past research on Alzheimer's classification and explains the problems faced by previous methods. The next section describes how we built our system, including the data augmentation methods, deep learning models, and ensemble learning technique. The following section explains how we trained and tested our model. Then, we present the results and show how well our model performed. Finally, we summarize the findings of the study and discuss what can be improved in the future to make Alzheimer's classification even better.

2. LITERATURE SURVEY

S. No.	Paper	Data	Classes	Technique	Results	Limitations
1.	G. Dematti et al. (2023) "Alzheimer's Disease Detection Using Brain MRI Images." [16]	OASIS ADNI Type: MRI	Mild Moderate Very Mild Non	- CNN with an 8-layer network structure	Accuracy: 90%	- Data Insufficiency
2.	N. Nasir et al. (2024) "Alzheimer's Magnetic Resonance Imaging Classification Using Deep and Meta-Learning Models." [17]	ADNI Type: MRI	NC MCI AD	- ResNet18, SqueezeNetV1, Vgg11(BN), InceptionV3, MobileNetV2 - Ensemble Stacking, Majority Voting	Accuracy: 90% Precision: 0.90 Recall: 0.89	- High Variance in deeper models - Overfitting in Ensemble method - Limited dataset size
3.	B. S. Rao et al. (2023) "Multi-class Classification of Alzheimer's Disease using Deep Learning and Transfer Learning on 3D MRI Images." [18]	ADNI Type: MRI	NC MCI AD	- 3D CNNs - Transfer Learning with pre-trained models such as ResNet50 and InceptionResNetV2	Accuracy: 91.25%	- Computational Efficiency - Model Interpretability
4.	J. Liu et al. (2021) "Alzheimer's disease detection using depth wise separable convolutional neural networks." [19]	OASIS Type: MRI	HC MCI AD	- Convolutional Neural Networks (CNN) - Depth wise Separable Convolution (DSC) - Transfer Learning with Alex Net and GoogLe Net.	- DSC Model Accuracy: 78.02% - TL with GoogLe Net Accuracy: 93.02%	- Data Insufficiency - High computational costs - Model Complexity
5.	P. Gayathri et al. (2024) "Deep Learning Augmented with SMOTE for Timely Alzheimer's Disease Detection in MRI images." [20]	OASIS Type: MRI	AD Non-AD	- CNNs for feature extraction - Synthetic Minority Over-Sampling Technique (SMOTE) for class imbalance - Spider Monkey Optimization (SMO) for classification	Accuracy: 91% Improved Sensitivity and specificity	- Model Interpretability - Need for broader clinical validation

.

G. Dematti et al. (2023) "Alzheimer's Disease Detection Using Brain MRI Images." This study looks at how to find Alzheimer's disease using brain scans (MRI images). The researchers used data from the OASIS and ADNI datasets and built a deep learning model with eight layers. The model was trained to classify images into four groups: Mild, Moderate, Very Mild, and Non-demented. The model worked well and got an accuracy of 90%. However, the study had a problem because there was not enough data. When there are fewer images to train the model, it might not work well for all patients. N. Nasir et al. (2024) "Alzheimer's Magnetic Resonance Imaging Classification Using Deep and Meta-Learning Models." This research tested different deep learning models to classify Alzheimer's disease using brain MRI images from the ADNI dataset. The models included ResNet18, SqueezeNetV1, Vgg11(BN), InceptionV3, and MobileNetV2. To improve accuracy, the researchers combined multiple models using two methods: Ensemble Stacking and Majority Voting. Their best model got an accuracy of 90%, precision of 0.90, and recall of 0.89. But there were some issues. Some models were too deep, causing them to change too much with new data (high variance). The ensemble method also had overfitting problems. Another issue was that the dataset was small, which made training harder. B. S. Rao et al. (2023) "Multi-class Classification of Alzheimer's Disease using Deep Learning and Transfer Learning on 3D MRI Images." This study focused on using 3D images instead of regular 2D MRI images for better Alzheimer's disease detection. The researchers used the ADNI dataset and trained models like 3D CNNs and pre-trained ResNet50 and InceptionResNetV2. Their best model reached an accuracy of 91.25%. While this approach gave good results, it had some challenges. The models required a lot of computer power to process 3D images, making them slower. Also, it was hard to understand how the model's made decisions, which could make doctors less likely to trust the system. J. Liu et al. (2021) "Alzheimer's Disease Detection Using Depth Wise Separable Convolutional Neural Networks." This research tested different CNN models to find Alzheimer's disease in MRI images from the OASIS dataset. It compared normal CNNs with Depth Wise Separable Convolution (DSC), which makes models faster and smaller. The study also tested Transfer Learning with pre-trained Alex Net and GoogLe Net models. The DSC model got 78.02% accuracy, while Transfer Learning with GoogLe Net did much better with 93.02% accuracy. But the study had some issues. There were not enough images, which made training harder. Also, the best models were expensive to run. P. Gayathri et al. (2024) "Deep Learning Augmented with SMOTE for Timely Alzheimer's Disease Detection in MRI Images." This study used deep learning to find Alzheimer's disease and tried to fix the problem of unbalanced data. The dataset (OASIS) had more images of one type than others, making it hard for the model to learn correctly. To fix this, the researchers used SMOTE, a method that creates more fake images of the underrepresented class so that all groups have similar amounts of data. They used CNNs to find features in the images and Spider Monkey Optimization (SMO) to improve classification. Their model reached 91% accuracy and improved sensitivity and specificity. However, there were some problems. It was hard to understand how the model made its decisions, and it needed more testing with real patients to prove it works well in hospitals

3. DATASET

The data acquisition process began by downloading the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset from Kaggle, a widely recognized source for medical imaging data. The dataset comprised over 6,400 MRI images, which were categorized into four distinct stages of Alzheimer's disease: non-demented, very mild demented, mild demented, and moderate demented. This classification was based on the severity of cognitive impairment, with each category representing a different level of disease progression. The dataset's division into these stages allowed for a multi-class classification approach, making it suitable for training machine learning models to detect and differentiate between various levels of cognitive decline. Additionally, the diversity within the dataset posed a challenge due to class imbalance, as some stages had fewer samples compared to others, necessitating data augmentation techniques to ensure balanced model training.

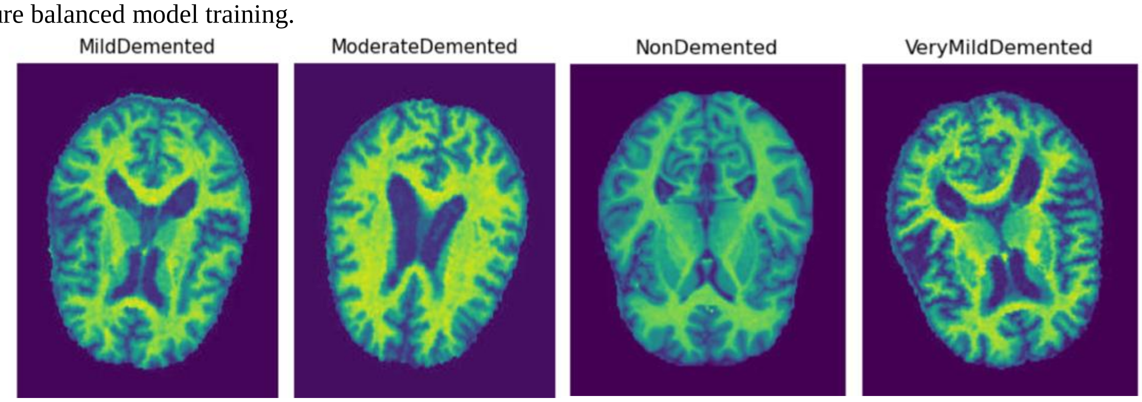


FIGURE 1. Sample Images

TABLE 2. Dataset Distributions

Data before Augmentation	Train	Test
Mild Demented	717	179
Moderate Demented	52	12
Very Mild Demented	1792	448
Non Demented	2560	640

Data after Augmentation	Train	Test
Mild Demented	2560	179
Moderate Demented	2560	12
Very Mild Demented	2560	448
Non Demented	2560	640

Data after splitting	Train	Validation	Test
Mild Demented	2076	484	179
Moderate Demented	2025	535	12
Very Mild Demented	2002	558	448
Non Demented	2089	471	640

The data preprocessing process began with image normalization to standardize the pixel values across all MRI images. Following normalization, various data augmentation techniques were applied to expand the dataset and mitigate the effects of class imbalance. The augmentation methods included rotation to different angles, horizontal flipping, and adjustments to image properties such as brightness, contrast, saturation, and hue. These transformations simulated different imaging scenarios and improved the model's ability to generalize across diverse real-world conditions. Following data augmentation, the dataset was divided into training and validation sets so that the model could learn from the training set and be assessed on the validation set, which acted as a stand-in for data that was not visible. This process helped in fine-tuning the model's hyperparameters and assessing its generalization capability before testing it on the final test set

4. METHODOLOGY

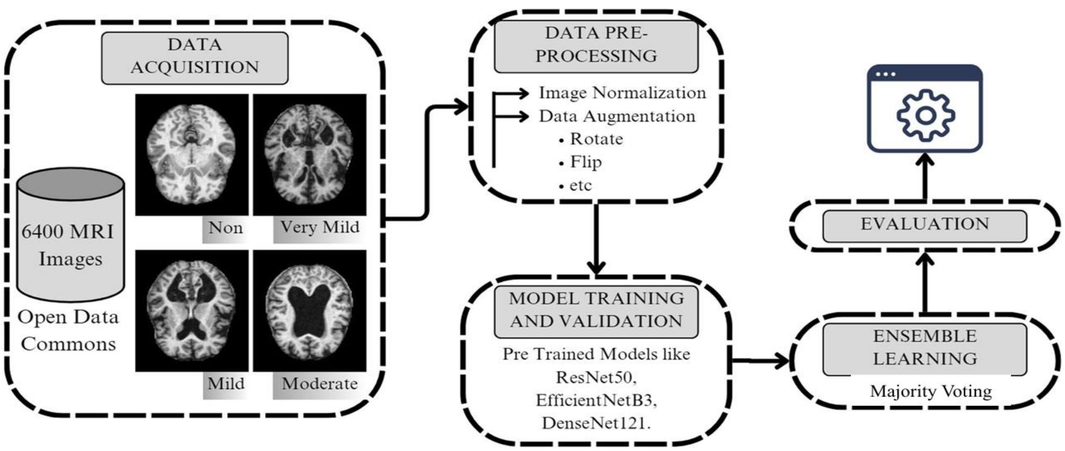


FIGURE 2. Methodology

Model Training and Validation: For the classification of Alzheimer's disease stages using MRI images, three advanced deep learning models—EfficientNetB3, ResNet50, and DenseNet121—were employed. These models, known for their powerful feature extraction capabilities and pre-trained on large-scale datasets, were fine-tuned to suit the specific requirements of multi-class classification in this project. Each model's architecture was modified to output predictions for the 4 stages of Alzheimer's disease, allowing for effective training and evaluation on the given dataset. EfficientNetB3 is a deep learning model known for its efficient scaling capabilities, which balance model depth, width, and resolution to achieve high performance with fewer parameters. Pre-trained on ImageNet, it has a robust architecture suited for image classification tasks while being computationally efficient. In this project, the EfficientNetB3 model was fine-tuned by replacing the final fully connected layer with a linear layer containing four output neurons, corresponding to the four Alzheimer's disease stages. This modification allows the model to perform multi-class classification on the MRI dataset. ResNet50, a 50-layer deep residual network, is one of the most commonly used architectures in image classification. It introduces skip connections, enabling the model to handle deeper layers without the vanishing gradient problem. For this task, ResNet50 was employed with pre-trained weights, and the final fully connected layer was replaced with a four-neuron output layer to accommodate the Alzheimer’s disease classification task. ResNet50’s depth and residual connections make it particularly adept at learning complex patterns in medical images like MRI scans. DenseNet121 is another powerful deep learning model known for its densely connected layers, where each layer receives the outputs of all preceding layers, enhancing feature reuse and reducing the number of parameters. It can handle complex medical imaging data because of its dense connection, which enhances gradient flow and reduces the vanishing gradient problem. DenseNet121 was optimized for this purpose by changing its last fully connected layer to produce four classes, enabling the multi-stage categorization of Alzheimer's illness. The training and validation process for each model followed a structured workflow aimed at optimizing the classification of Alzheimer's disease stages. Initially, the pre-trained models (EfficientNetB3, ResNet50, and DenseNet121) were fine-tuned by replacing their final fully connected layers with a four-class output layer to match the task at hand. The Adam optimizer was chosen for optimizing the model parameters with a learning rate of 0.001, while the Cross Entropy Loss function was used to calculate the classification loss for each prediction. For each model, training was performed over 10 epochs. The training process aimed to minimize the loss and maximize the classification accuracy. For evaluation, after each epoch, the model was switched to evaluation mode to process the validation data. The validation process calculated metrics like validation loss, accuracy, precision, recall, and F1-score, along with generating a confusion matrix and a detailed classification report for further performance analysis. Finally, the training and validation loss/accuracy trends were plotted to visualize the model's learning progress over time. Each trained model was saved as a checkpoint using the torch. Save () function, allowing for future reuse and deployment.

Ensemble Learning: By merging predictions from several models, ensemble learning is a machine learning technique that improves model performance. This approach leverages the diversity in model architectures and learning strategies to reduce errors and improve generalization. Common techniques include bagging, which involves training models on different data subsets; boosting, where models are sequentially trained to correct errors from previous models; stacking, which uses predictions from various models as inputs for a higher-level model; and majority voting, where the final output is determined by the majority vote of the models. In this project, ensemble learning was implemented using a majority voting strategy by combining three pre-trained deep learning models: EfficientNetB3, ResNet50, and DenseNet121. We used metrics like validation loss, accuracy, precision, recall, and F1-score to assess the ensemble model on the validation set. Furthermore, a classification report and a confusion matrix were produced to offer comprehensive insights into the model's performance during the four phases of Alzheimer's disease. The ensemble approach outperformed individual models by aggregating their strengths, enhancing the accuracy and reliability of predictions. Finally, the trained ensemble model was saved for future use in the diagnostic system

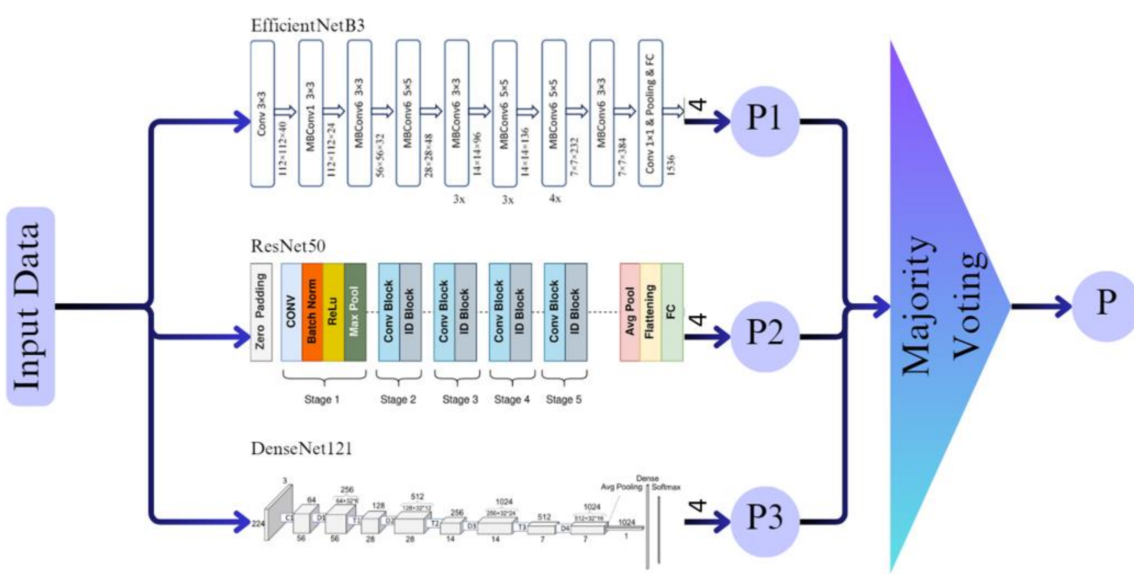


FIGURE 3. Proposed Ensemble Approach

Evaluation: The evaluation process for the Alzheimer's disease classification project involved a different assessment of the model's performance using various metrics and techniques. After training each individual model (EfficientNetB3, ResNet50, and DenseNet121), as well as the ensemble model, the evaluation was conducted on a separate validation set to assess the model's ability to generalize on unseen data. Validation loss and accuracy were the main evaluation metrics, and they were monitored during the training phase to keep an eye on the model's learning behavior. Loss and accuracy charts were made for the training and validation stages for every epoch in order to show the model's learning progress. Metrics including F1-score, accuracy, and recall were also computed for every model. A more thorough grasp of the model's performance across classes was made possible by precision, recall, and F1-score, particularly when working with unbalanced datasets. While recall evaluated the model's sensitivity in identifying all real positive cases, precision rated the model's capacity to accurately identify positive predictions for each class. When it came to class imbalance, the F1-score—which is the harmonic mean of precision and recall—provided a fair assessment metric. In order to pinpoint certain regions where the model performed poorly or effectively, the confusion matrix was also utilized to show the distribution of samples that were correctly and wrongly classified across the four phases of Alzheimer's disease. For the ensemble model evaluation, predictions from the three individual models were majority voted to provide a final output. The ensemble model's performance was measured using the same metrics as the individual models (validation loss, accuracy, precision, recall, F1-score, and confusion matrix), allowing for a direct comparison. This comparative analysis showed how the ensemble approach improved classification accuracy and consistency across the different Alzheimer's stages. Overall, the evaluation process played a crucial role in validating the effectiveness of the proposed ensemble learning approach and in ensuring the model's readiness for real-world clinical applications in Alzheimer's disease detection and classification.

Web Page Development: An open-source Python framework called Stream lit makes it easy for programmers to create dynamic, intuitive web applications for data science and machine learning projects. It is particularly well-suited for deploying machine learning models and visualizing data, as it provides built-in functionalities for displaying various types of content, including charts, images, text, and interactive widgets. In this project, Stream lit is used to develop a user-friendly web interface that allows users to interact with the Alzheimer's disease classification model. The primary goal of the Stream lit app is to provide a platform where users, such as clinicians or researchers, could upload MRI images and receive real-time classification results

5. RESULTS AND DISCUSSION

Performance of EfficientNetB3 Model:

FIGURE 4. Train vs Validation loss curves & Train vs Validation accuracy curves of EfficientNetB3 model

```
Confusion Matrix:
[[486   0   6  37]
 [  0 522   0   0]
 [  0   0 489   3]
 [  2   0   2 501]]
```

FIGURE 5. Confusion matrix for EfficientNetB3 model

Classification Report:

	precision	recall	f1-score	support
MildDemented	1.00	0.92	0.96	529
ModerateDemented	1.00	1.00	1.00	522
NonDemented	0.98	0.99	0.99	492
VeryMildDemented	0.93	0.99	0.96	505
accuracy			0.98	2048
macro avg	0.98	0.98	0.98	2048
weighted avg	0.98	0.98	0.98	2048

FIGURE 6. Classification Report of EfficientNetB3 model

The EfficientNetB3 model displayed strong classification capabilities, achieving an overall accuracy of 98%. It performed exceptionally well for the Moderate Demented class, where it reached perfect precision, recall, and F1-score, correctly classifying all instances. The non-demented class also showed high precision (0.98) and recall (0.99), with minimal misclassifications. However, for the Very Mild Demented class, the precision was slightly lower at 0.93, though it had a high recall of 0.99, indicating that while the model was effective at identifying this stage, it occasionally confused other classes with Very Mild Demented. For the Mild Demented class, the model achieved a precision of 1.00 and a recall of 0.92, suggesting some errors in distinguishing Mild Demented cases. Overall, the EfficientNetB3 model exhibited robust performance across all classes, with minor misclassification issues in closely related stages.

Performance of ResNet50 Model:

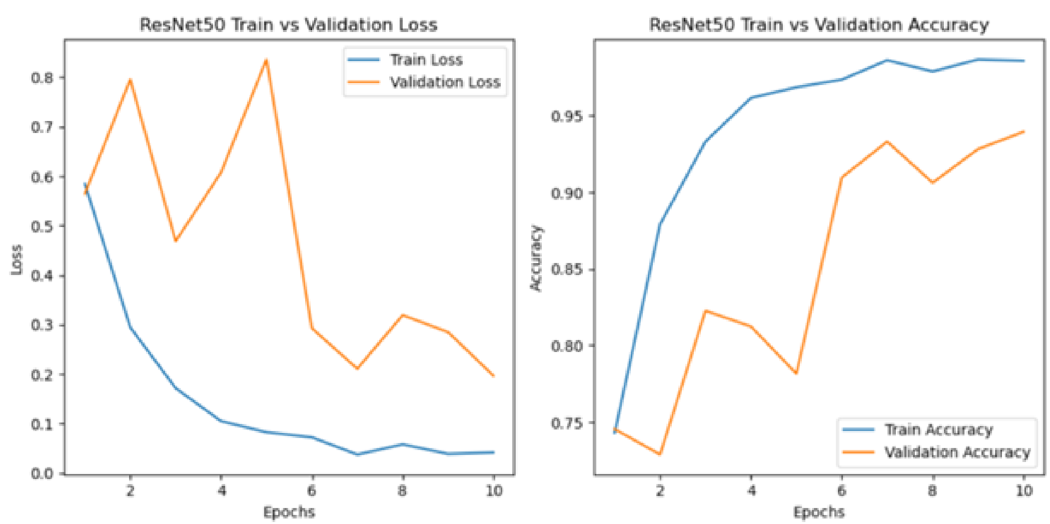


FIGURE 7. Train vs Validation loss curves & Train vs Validation accuracy curves of ResNet50 model

```
Confusion Matrix:
[[506   6   3  14]
 [  0 522   0   0]
 [  3   0 475  14]
 [ 78   0   6 421]]
```

FIGURE 8. Confusion matrix for ResNet50 model

Classification Report:

	precision	recall	f1-score	support
MildDemented	0.86	0.96	0.91	529
ModerateDemented	0.99	1.00	0.99	522
NonDemented	0.98	0.97	0.97	492
VeryMildDemented	0.94	0.83	0.88	505
accuracy			0.94	2048
macro avg	0.94	0.94	0.94	2048
weighted avg	0.94	0.94	0.94	2048

FIGURE 9. Classification Report of ResNet50 model

ResNet50 achieved an accuracy of 94%, indicating reliable performance but with some limitations compared to EfficientNetB3. The model excelled in classifying the Moderate Demented class with precision, recall, and F1-score values close to 1.00, showing consistency in detecting this stage. The non-demented class also performed well, with a precision of 0.98 and recall of 0.97. However, the model faced challenges in the Very Mild Demented class, with a recall of 0.83, indicating difficulty in correctly identifying all cases of this stage, as reflected in some misclassifications with Mild Demented. For the Mild Demented class, the model achieved a precision of 0.86 and a recall of 0.96, showing some tendency to misclassify other stages as Mild Demented. ResNet50 demonstrated a solid performance but showed slightly lower accuracy and greater confusion between stages compared to EfficientNetB3.

Performance of DenseNet121 Model:

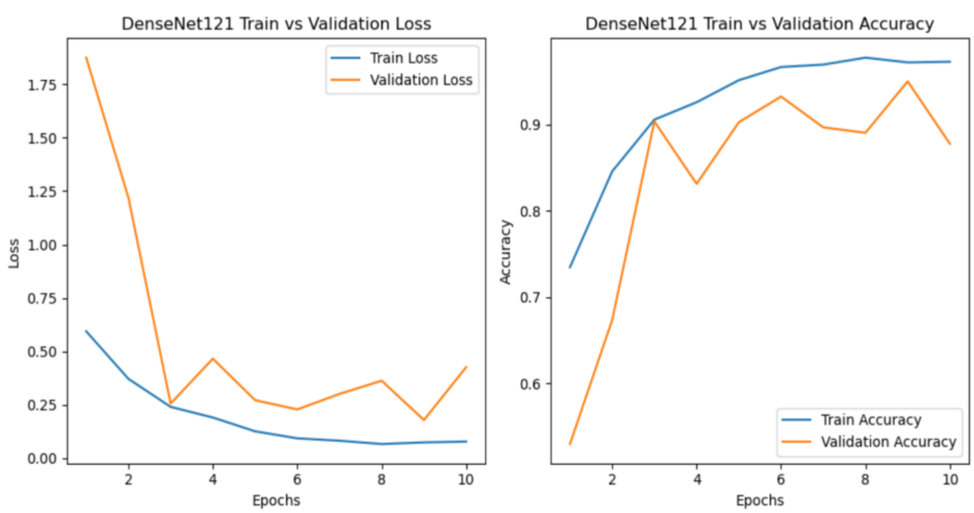


FIGURE 10. Train vs Validation loss curves & Train vs Validation accuracy curves of DenseNet121 model

```
Confusion Matrix:
[[519   0   3   7]
 [112 403   0   7]
 [ 17   0 463  12]
 [ 89   0   4 412]]
```

FIGURE 11. Confusion matrix for DenseNet121 model

```
Classification Report:
                  precision    recall  f1-score   support

    MildDemented       0.70      0.98      0.82       529
ModerateDemented       1.00      0.77      0.87       522
     NonDemented       0.99      0.94      0.96       492
VeryMildDemented       0.94      0.82      0.87       505

        accuracy                           0.88      2048
       macro avg       0.91      0.88      0.88      2048
    weighted avg       0.91      0.88      0.88      2048
```

FIGURE 12. Classification Report of DenseNet121 model

The DenseNet121 model achieved an overall accuracy of 88%, indicating some limitations in classifying the four Alzheimer's stages. It performed exceptionally well for the Mild Demented class, achieving a high recall of 0.98, meaning the model correctly identified almost all Mild Demented cases. However, the precision for this class was only 0.70, indicating a higher rate of false positives. The model struggled with the Moderate Demented class, with a recall of 0.77, meaning that a significant number of Moderate Demented instances were misclassified, primarily as Mild Demented. For the non-demented class, the model showed high precision (0.99) and recall (0.94), reflecting its ability to correctly classify most non-demented cases with few misclassifications. The Very Mild Demented class had a precision of 0.94 and a recall of 0.82, showing some difficulty in distinguishing this stage from the others. Overall, DenseNet121 demonstrated strong performance in some areas but had noticeable weaknesses in handling class confusion, particularly between closely related stages.

Performance of Ensemble Model

```
Confusion Matrix:
[[518   0   3   8]
 [  0 522   0   0]
 [  0   0 490   2]
 [ 16   0   1 488]]
```

FIGURE 13. Confusion matrix for Ensemble model

```
Classification Report:
                  precision    recall  f1-score   support

    MildDemented       0.97      0.98      0.97       529
ModerateDemented       1.00      1.00      1.00       522
     NonDemented       0.99      1.00      0.99       492
VeryMildDemented       0.98      0.97      0.97       505

        accuracy                           0.99      2048
       macro avg       0.99      0.99      0.99      2048
    weighted avg       0.99      0.99      0.99      2048
```

FIGURE 14. Classification report of Ensemble model

The ensemble model outperformed all individual models, achieving a high overall accuracy of 99%. It showed perfect performance for the Moderate Demented class, with precision and recall of 1.00, and near-perfect scores for the non-demented class (0.99 precision and recall). The model also achieved strong results for the Very Mild Demented (precision 0.98, recall 0.97) and Mild Demented classes (precision 0.97, recall 0.98), demonstrating its ability to accurately classify nearly all instances. The ensemble model's ability to combine the strengths of EfficientNetB3, ResNet50, and DenseNet121 contributed to its superior performance across all stages, minimizing misclassifications and achieving the highest precision, recall, and F1-scores overall. The outputs generated by the images of respective classes are given below. An image taken from the Mild Demented folder is correctly classified as Mild Demented only. An image taken from the Very Mild Demented folder is correctly classified as Very Mild Demented only. An image taken from the Moderate Demented folder is correctly classified as Moderate Demented only. An image taken from the non-demented folder is correctly classified as non-demented only. This shows that the model performs very well on the data.

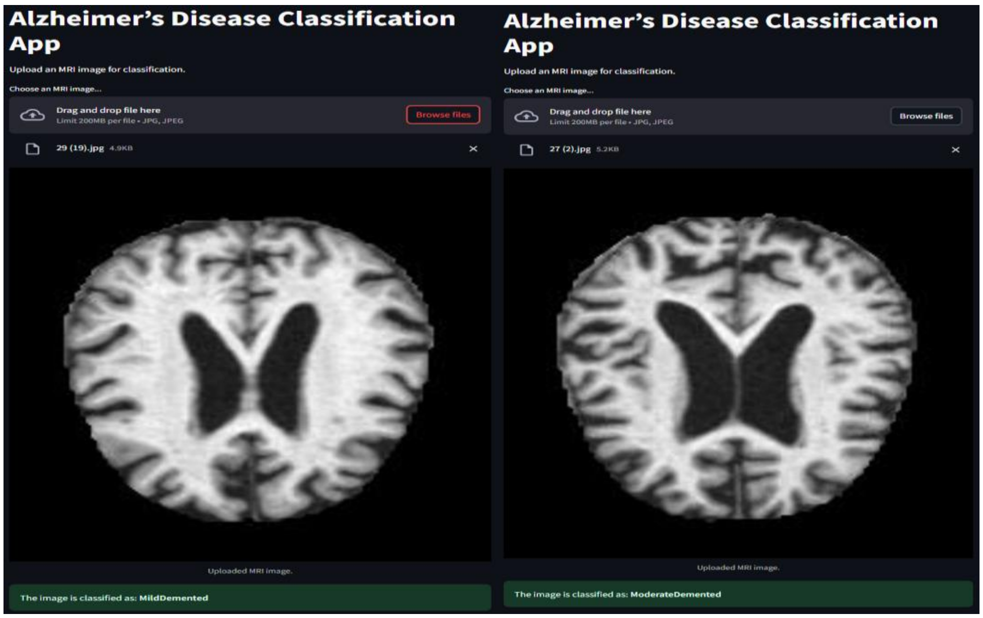


FIGURE 15.

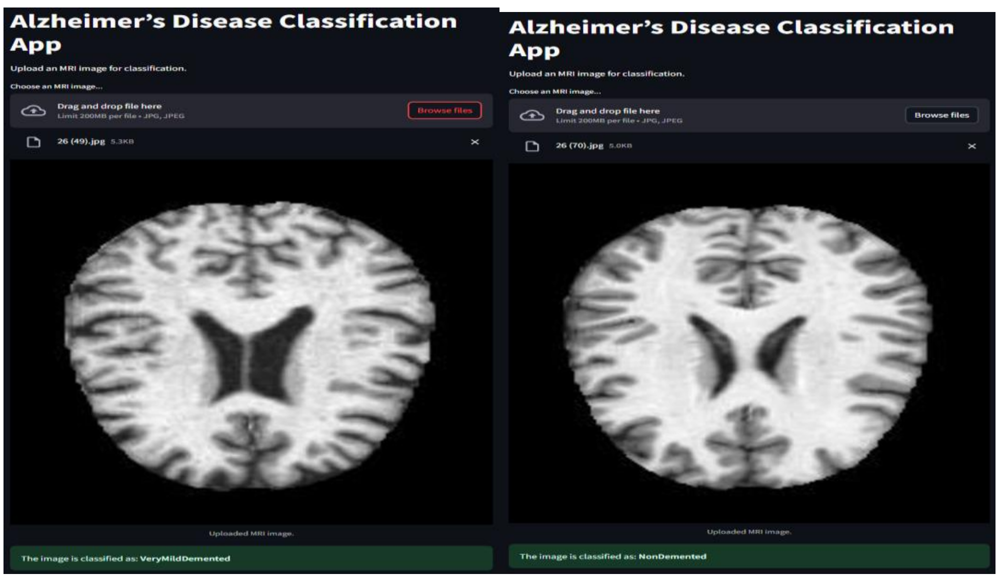


FIGURE 16.

6. CONCLUSION

The project successfully developed a robust system for classifying different stages of Alzheimer's disease using MRI images, leveraging the collaborative strengths of multiple models. By implementing a range of data augmentation techniques—such as rotations, flips, and brightness and contrast adjustments—the team effectively addressed challenges associated with data imbalance, enhancing the diversity of the training dataset. The EfficientNetB3 model demonstrated impressive performance, achieving a remarkable accuracy of 98%. This high accuracy indicated that the model effectively learned from the features of the training data, making it well-suited for tasks like image classification. Similarly, the ResNet50 model showed strong performance with an accuracy of 94%. Its architecture, which included residual connections, allowed it to learn deeper representations without suffering from the vanishing gradient problem, thus contributing to its reliability in classifying complex images. The DenseNet121 model, while slightly lower in performance, still achieved an accuracy of 88%. Its dense connections between layers helped in feature reuse and mitigated the vanishing gradient issue, although it may not have been as effective as the previous models for this particular task. Finally, the ensemble model, which combined the predictions of EfficientNetB3, ResNet50, and DenseNet121 using a majority voting approach, achieved the highest accuracy of 99%. This significant improvement illustrated the benefits of ensemble learning, as it harnessed the strengths of multiple models, leading to better overall performance in accurately classifying images. The ensemble model clearly outperformed the individual pre-trained models, enhancing its ability to perform the classification task effectively. This innovative system not only facilitates the early detection of Alzheimer's disease, which is vital for improving treatment and care, but also represents a meaningful advancement in the integration of AI within healthcare. By enhancing diagnostic capabilities for Alzheimer's and similar diseases, the project contributes to the broader goal of leveraging artificial intelligence to transform medical practice and patient outcomes.

REFERENCES

[1]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[2]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[3]. Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[4]. R. S. Kumar, K. K. Srinivas, A. Peddi and P. A. H. Vardhini, "Artificial Intelligence based Human Attention Detection through Brain Computer Interface for Health Care Monitoring," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 42-45, doi: 10.1109/BECITHCON54710.2021.9893646.

[5]. Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.

[6]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[7]. Manoranjan Dash, "Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment," International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[8]. Manoranjan Dash et al.," Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique", Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[9]. Manoranjan Dash et al.," Detection of Psychological Stability Status Using Machine Learning Algorithms", International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[10]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., “Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection”, Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[11]. Suresh. M, A. M. Reddy, "A Stacking-based Ensemble Framework for Automatic Depression Detection using Audio Signals", International Journal of Advanced Computer Science and Applications (IJACSA), vol. 14, no. 7, pp. 603-612, 2023, doi: 10.14569/IJACSA.2023.0140767.

[12]. Mallikarjuna A. Reddy, Sudheer K. Reddy, Santhosh C.N. Kumar, Srinivasa K. Reddy, “Leveraging bio-maximum inverse rank method for iris and palm recognition”, International Journal of Biometrics, 2022 Vol.14 No.3/4, pp.421 - 438, DOI: 10.1504/IJBM.2022.10048978.

[13]. V. NavyaSree, Y. Surarchitha, A. M. Reddy, B. Devi Sree, A. Anuhya and H. Jabeen, "Predicting the Risk Factor of Kidney Disease using Meta Classifiers," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972392.

[14]. B. H. Rao et al., "MTESSERACT: An Application for Form Recognition in Courier Services," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 848-853, doi: 10.1109/ICOSEC54921.2022.9952031.

[15]. P. S. Silpa et al., "Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 759-767, doi: 10.1109/ICOSEC54921.2022.9951883.

[16]. G. Dematti, S. P., S. Jinni, S. G. Borkar, and S. P. Naik, "Alzheimer’s Disease Detection Using Brain MRI Images," Int. J. Res. Pub. Rev., vol. 4, no. 3, pp. 2409-2416, Mar. 2023.

[17]. N. Nasir, M. Ahmed, N. Afreen, & M. Sameer, "Alzheimer’s Magnetic Resonance Imaging Classification Using Deep and Meta Learning Models," arXiv preprint, arXiv: 2405.12126, 2024.

[18]. B. S. Rao, M. Aparna, S. S. Kolisetty, H. Janapana, and Y. V. Koteswararao, "Multi-class Classification of Alzheimer’s Disease Using Deep Learning and Transfer Learning on 3D MRI Images," Trans. Sci. Technol., vol. 41, pp. 328-335, 2023.

[19]. J. Liu, M. Li, Y. Luo, S. Yang, W. Li, & Y. Bi, "Alzheimer’s disease detection using depthwise separable convolutional neural networks," Comput. Methods Prog. Biomed. 106032, 2021.

[20]. P. Gayathri, N. Geetha, M. Sridhar, R. Kuchipudi, K. S. Babu, L. P. Maguluri, & B. K. Bala, "Deep Learning Augmented with SMOTE for Timely Alzheimer’s Disease Detection in MRI Images," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 2, pp. 253-260, 2024.