A Project Report

On

Utilizing machine learning for early diagnosis and severity assessment of Neurodegenerative diseases

---

BY

B V Soma Adithya – SE21UARI192

Jasmitha Bolla – SE21UARI028

Koushik Dupaguntla – SE21UARI162

Under the supervision of

Dr. Satish Chandra

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

DEGREE OF BACHELOR OF TECHNOLOGY

PR 4101: PROJECT TYPE COURSE

ECOLE CENTRALE SCHOOL OF ENGINEERING

MAHINDRA UNIVERSITY

HYDERABAD

(December 2024)

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to all those who contributed to the completion of this project. First and foremost, We are profoundly grateful to Dr. Satish Chandra for his invaluable guidance, insightful feedback, and unwavering support throughout the course of this project.

We extend our heartfelt thanks to our colleagues and peers who offered constructive criticism, shared resources, and provided a motivating and collaborative environment.

Thank you, Professor Satish Chandra, for your patience, wisdom, and for being a beacon of knowledge. We deeply appreciate the opportunity to learn from you.

Ecole Centrale School of Engineering

Mahindra University

Hyderabad

Certificate

This is to certify that the project report entitled "**Utilizing machine learning for early diagnosis and severity assessment of Neurodegenerative diseases**" submitted by Soma Adithya (SE21UARI192), Jasmitha (SE21UARI028), Koushik (SE21UARI162) in partial fulfillment of the requirements of the course PR 4101, Project Course, embodies the work done by him/her under my supervision and guidance.

(Dr. Satish Chandra & Signature)

Ecole Centrale School of Engineering, Hyderabad.

Date: 29th December 2024

ABSTRACT

In our project, we explored different deep learning models for the early detection of neurodegenerative diseases. We used a dataset of three classes: Alzheimer's, Parkinson's, and normal images. We utilized the following models: Vision Transformers, ResNet-101, CNN, and CBAM attention. We evaluated these models for the classification of neurodegenerative diseases. While VIT showed reasonable accuracy, it struggled with detecting subtle disease features. RESNET, with its residual connections, outperformed the others, offering superior accuracy, feature extraction, and generalization. Although CBAM enhanced feature focus, it did not surpass RESNET. In our project, we analyzed the results of these models and explained the reasons behind the accuracies in this report. Overall, RESNET demonstrated the best performance, highlighting its potential for supporting early NDD diagnosis.

CONTENTS

Figures Used in this Report :

1. Introduction

Neurodegenerative diseases (NDDs) represent a class of disorders characterized by the progressive degeneration of neurons in the human brain and nervous system. Unlike other cells in the body, neurons typically do not regenerate or replace themselves, making the damage caused by these diseases largely irreversible. Conditions such as Alzheimer's disease and Parkinson's disease are among the most common neurodegenerative diseases, each presenting unique challenges to patients, caregivers, and healthcare systems worldwide.

The prevalence of neurodegenerative diseases is on the rise, primarily due to the aging global population. According to the World Health Organization (WHO), by 2050, nearly 2 billion people will be aged 60 and above, with a significant portion at risk of developing an NDD. Alzheimer's disease alone affects approximately 55 million people globally, a number projected to triple by mid-century. Beyond the human toll, these diseases pose a massive economic burden. Despite their impact, effective treatments for most neurodegenerative diseases remain elusive. Current therapies primarily focus on symptom management rather than addressing the root causes or halting disease progression. This unmet medical need underscores the urgency of innovative approaches to better understand, diagnose, and treat NDDs.

This project aims to tackle critical challenges in the realm of neurodegenerative diseases, with a focus on improving diagnostic accuracy. Early diagnosis is crucial for managing NDDs effectively, yet many patients experience delays in receiving accurate diagnoses. Leveraging advancements in artificial intelligence, biomarker research, and imaging technologies, this project seeks to bridge gaps in early detection and personalized treatment.In addition to diagnostics, therapeutic innovation is a key area of focus. Neurodegenerative diseases are notoriously complex, often involving a mix of genetic, environmental, and lifestyle factors.

Beyond the immediate benefits to affected individuals, the societal impact of this work cannot be overstated. Families of patients with NDDs often face emotional, physical, and financial stress due to the long-term caregiving demands. Innovations that alleviate disease progression or enhance care can significantly reduce this burden, fostering healthier communities.

The fight against neurodegenerative diseases is not merely a medical endeavor but a moral imperative. This project represents an opportunity to combine cutting-edge science with compassionate care, ultimately reshaping how we understand and treat these devastating conditions. By focusing on translational research and real-world applications, the project aspires to create solutions that are not just theoretically effective but practically impactful.It is by developing diagnostic tools that provide clarity to families early on, the potential impact of this project is vast. It seeks to contribute to a future where neurodegenerative diseases are no longer an inevitable consequence of aging but manageable conditions with promising outcomes.

2. Background and Related Work

In the paper “Parkinson’s Disease Detection from Resting-State EEG Signals Using Common Spatial Pattern, Entropy, and Machine Learning Techniques**”,** they proposed novel common spatial pattern-based approaches for detecting Parkinson's disease (PD) using EEG signals in both off-medication and on-medication cases. The EEG signals are preprocessed to remove artifacts before applying spatial filtering using the common spatial pattern method. Various features are extracted from the spatially filtered signals. After this they have applied machine learning technques for classification. They used Two EEG datasets, the SanDiego (31 participants, 93 min) and UNM (54 participants, 54 min) datasets for testing. The results demonstrate that the proposed methods, particularly using common spatial patterns and log energy entropy, achieve competitive performance compared to other methods in the literature. In off-medication PD detection, classification accuracy reaches around 99%, while for on-medication PD, the accuracy ranges from 95% to 98%. The highest classification accuracy is achieved with features from the alpha and beta bands.

In the paper “Accurate Detection of Alzheimer’s Disease Using Lightweight Deep Learning Model on MRI Data**”** authors have presented an improved lightweight deep learning model for detecting Alzheimer's disease (AD) from magnetic resonance imaging (MRI) scans. The proposed model is designed to accurately detect AD without relying on deep networks or traditional methods like feature extraction and classification, consolidating them into a single stage. With only seven layers, the model is less complex and more time-efficient compared to previous deep learning models. It was tested on a publicly available Kaggle dataset, achieving 99.22% accuracy in binary classification and 95.93% accuracy in multi-classification tasks.

The paper “Deep Learning-Based Classification of Neurodegenerative Diseases Using Gait Dataset: A Comparative Study**”** examines the use of seven deep learning architectures for classifying neuro-degenerative diseases (NDDs) using a gait dynamics dataset. The models analyzed include LSTM, GRU, InceptionTime, ResNet, FCN, TST, and PatchTST. The study evaluates these models across various classification tasks. The findings show that deep learning methods have strong potential for diagnosing and classifying NDDs. ResNet performs best in distinguishing between healthy controls (HC) and individuals with NDDs, as well as between HC and Parkinson's disease (PD). Meanwhile, TST excels in differentiating between Amyotrophic Lateral Sclerosis (ALS) or Huntington's Disease (HD) and healthy controls.

Considering the background work, we have used the ResNet, CNN and ViT models. The accuracy must be very high considering the the problem is about patients health. So we wanted to achieve much higher accuracy which is less comparatively for NDD diseases. We overcame this and achieved an 99.37 percent accuracy for multi class classification of as Alzheimer’s disease and Parkinson’s disease along with normal images.

3. Dataset and Preprocessing

The Neurodegenerative Disease Detection Dataset is designed to facilitate the development and testing of machine learning models for the early detection and classification of neurodegenerative diseases. It consists of MRI images categorized into three distinct groups: Alzheimer's disease, Parkinson's disease, and normal (healthy) controls.

Alzheimer's Disease Dataset:

- Contains MRI scans of patients diagnosed with Alzheimer's disease.Aims to help in identifying patterns and features characteristic of Alzheimer's disease.

Normal Dataset:

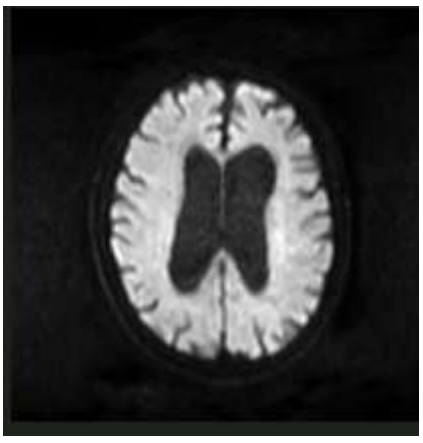
- Contains MRI scans of healthy individuals with no neurodegenerative disease. Serves as a control group to distinguish normal brain patterns from diseased states.

Parkinson's Disease Dataset:

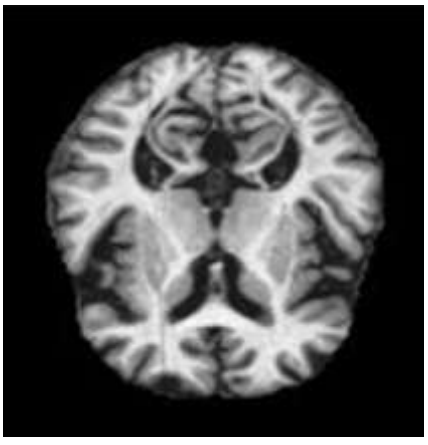
- Contains MRI scans of patients diagnosed with Parkinson's disease. Focuses on identifying early biomarkers and patterns associated with Parkinson's disease.

This augumentation is already done on this dataset and it is balanced. Normal images are 2699, Parkinson images are 2391 and Alzemeirs are 2500 images making a total of 7590 images.

The images in this dataset have been aggregated from various reputable sources to ensure diversity and comprehensiveness. Some of the primary sources include: National Institutes of Health (NIH), Alzheimer's Disease Neuroimaging Initiative (ADNI), Parkinson's Progression Markers Initiative (PPMI), Kaggle Parkinson's Brain MRI Dataset, Kaggle Alzheimer's MRI Dataset.



Parkinson Disease Smaple



Alzheimers Disease Sample

We have splitted the data into train, val and test in the ratio of 80:10:10.

4. Deep Learning Methods

In this section, we will briefly explain the deep learning models used in this project.

4.1 Vision Transformers

The Vision Transformer (ViT) is an innovative deep learning architecture that applies Transformer models, originally designed for natural language processing (NLP), to image recognition tasks. It diverges from traditional convolutional neural networks (CNNs) by treating images as sequences of patches.

Key Steps in ViT:

1. Patch Embedding:
    - The input image is divided into fixed-size patches (e.g., 16x16 pixels).
    - Each patch is flattened into a 1D vector and linearly projected into a higher-dimensional embedding space.
2. Positional Encoding:
    - Since Transformers are permutation-invariant, positional encodings are added to the patch embeddings to retain spatial information.
    - These encodings can be fixed (predefined sinusoidal patterns) or learnable parameters.
3. Transformer Encoder:
    - The encoder consists of multiple layers of **multi-head self-attention** and **feed-forward neural networks**. Each layer processes the sequence of patch embeddings:
        - **Multi-Head Self-Attention (MHSA)**: Captures relationships between all patches, enabling global feature learning.
        - **Feed-Forward Neural Network (FFN)**: Applies non-linear transformations to enhance feature representation.
    - Layer normalization and residual connections are used to stabilize training and improve model performance.
4. Classification Head:
    - A learnable class token is appended to the patch sequence before it is passed to the Transformer encoder.

After processing, the class token contains a summarized representation of the entire image, which is passed to a classification head (usually a dense layer) for the final prediction.

4.2 ResNet – 101

ResNet 101 (Residual Network with 101 layers) is a deep neural network designed to solve the vanishing gradient problem in very deep networks. It uses **residual blocks**, where the input is added back to the output via a skip connection, allowing the network to learn identity mappings. This architecture facilitates training deeper networks efficiently by reusing learned features. ResNet 101 consists of convolutional layers, batch normalization, ReLU activation functions, pooling layers, and skip connections.

4.3 Convolutional Neural Networks (CNN)

CNNs are neural networks primarily used for image and spatial data processing. They work by using convolutional layers to extract features from input data (e.g., edges, shapes, patterns), followed by pooling layers to reduce dimensionality while retaining important information. Key components include convolutional layers, pooling (e.g., max pooling), fully connected layers, and activation functions like ReLU. CNNs excel at hierarchical feature learning, starting with simple patterns and progressing to complex structures.

4.4 Convolutional Block Attention Module (CBAM)

CBAM (Convolutional Block Attention Module) is an attention mechanism that enhances convolutional neural networks (CNNs) by focusing on important features. It consists of two components: the **Channel Attention Module (CAM)** and the **Spatial Attention Module (SAM)**. The CAM emphasizes useful channels by applying global average and max pooling to create a channel attention map, which is then multiplied with the feature map. The SAM, on the other hand, focuses on important spatial locations by pooling across spatial dimensions to generate a spatial attention map. Both attention maps refine the feature map by highlighting relevant information in both channels and spatial areas.

5. Implementation

In this section we will explain our code implementations for each model.

We have utilized PyTorch framework throughout the project. To prevent overfitting and ensure efficient use of computational resources, early stopping is employed in all the following models. . This mechanism halts training if the validation loss does not improve for a specified number of epochs, indicating that the model has likely reached its optimal performance. The model state corresponding to the best validation loss is saved for later evaluation in .pth file. DGX A100 Super computer is utilized for this project.

5.1 Vision Transformers (ViT)

We loaded the dataset and convert each image into 224x224 size images suitable for ViT model. Then we normalize the dataset based on the mean and standard deviation values of the whole dataset. Following this we splitetd the dataset into 80:20 into train and test sets. The test set is further divided into 50:50 as validation and test sets.

We loaded the train, validation and test sets into respective loaders. To use computational resources effectively, we used a batch size of 16x16 and 4 tasks per node in the Super Computer. We the verified the size of each loader. After ViT config is modified accordingly i.e., patch size and number of classes. We ran the model on DGX A100 Super computer.

The training process involves optimizing the ViT model on the training dataset while monitoring its performance on the validation set. Cross-entropy loss is employed as the objective function to measure the difference between predicted and true labels. The Adam optimizer is used for weight updates, offering efficient convergence with minimal hyperparameter tuning. As explained earlier, early stopping criteria is employed.

After training, the model is tested on the test set to evaluate its final performance. Accuracy is calculated as the overall percentage of correct predictions. Additionally, a detailed classification report is generated, providing precision, recall, and F1-score for each class. These metrics offer a comprehensive understanding of the model's strengths and weaknesses, particularly in handling different categories. The confusion matrix is also analyzed to visualize the model's performance in distinguishing between classes.

We experimented the model with 3 different varients.

1. 16x16 Patch Size
2. 8x8 Patch Size
3. 4x4 Patch Size

Foe each of the model, below are the loss evolution graphs.

5.1.1 ViT 16x16 Patch Size

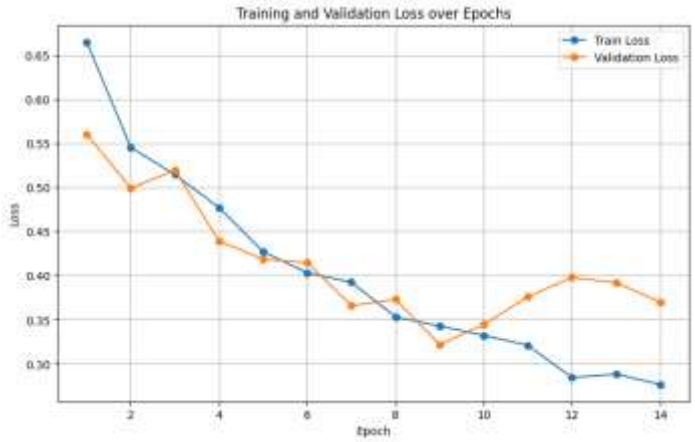


Figure: 1 Loss Curve of ViT 16x16 Patch size

5.1.2 ViT 8x8 Patch Size

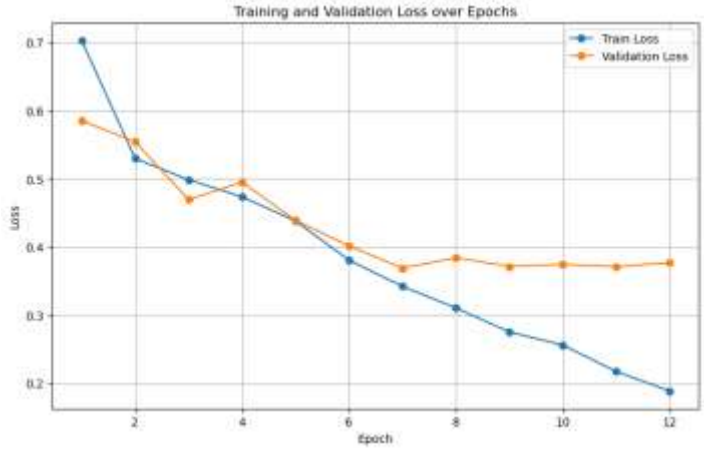


Figure: 2 Loss Curve of ViT 8x8 Patch size

5.1.3 ViT 4x4 Patch Size

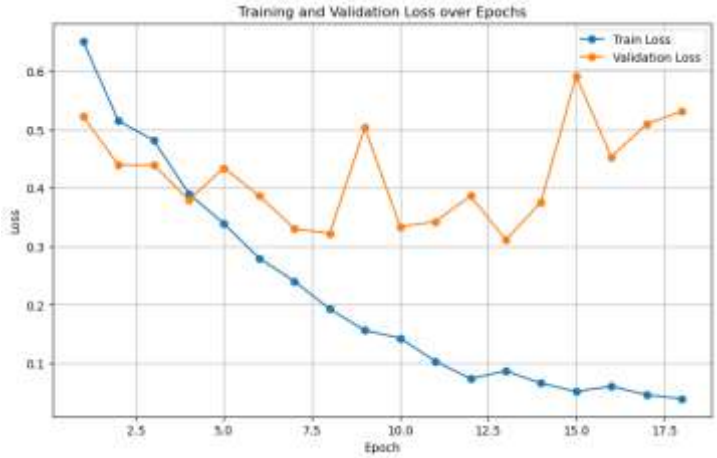


Figure: 3 Loss Curve of ViT 4x4 Patch size

5.2 ResNet-101

We loaded the dataset and convert each image into 224x224 size images for ResNet model. Then we normalize the dataset based on the mean and standard deviation values of the ImageNet dataset as the ResNet model is pretrained on the Imagenet dataset. Following this we splitetd the dataset into 80:20 into train and test sets. The test set is further divided into 50:50 as validation and test sets.

Then we change the last layer of the ResNet-101 model according to our no of classes. The training process involves optimizing the ResNet model on the training dataset while monitoring its performance on the validation set. Cross-entropy loss is employed as the objective function to measure the difference between predicted and true labels. The Adam optimizer is used for weight updates, offering efficient convergence with minimal hyperparameter tuning. As explained earlier, early stopping criteria is employed.

After training, the model is tested on the test set. Classification metrices are calculated and the conduison matrix is visuvalized.

5.2.1 ResNet-101 (Batch = 16 & Learning Rate = 0.000001)

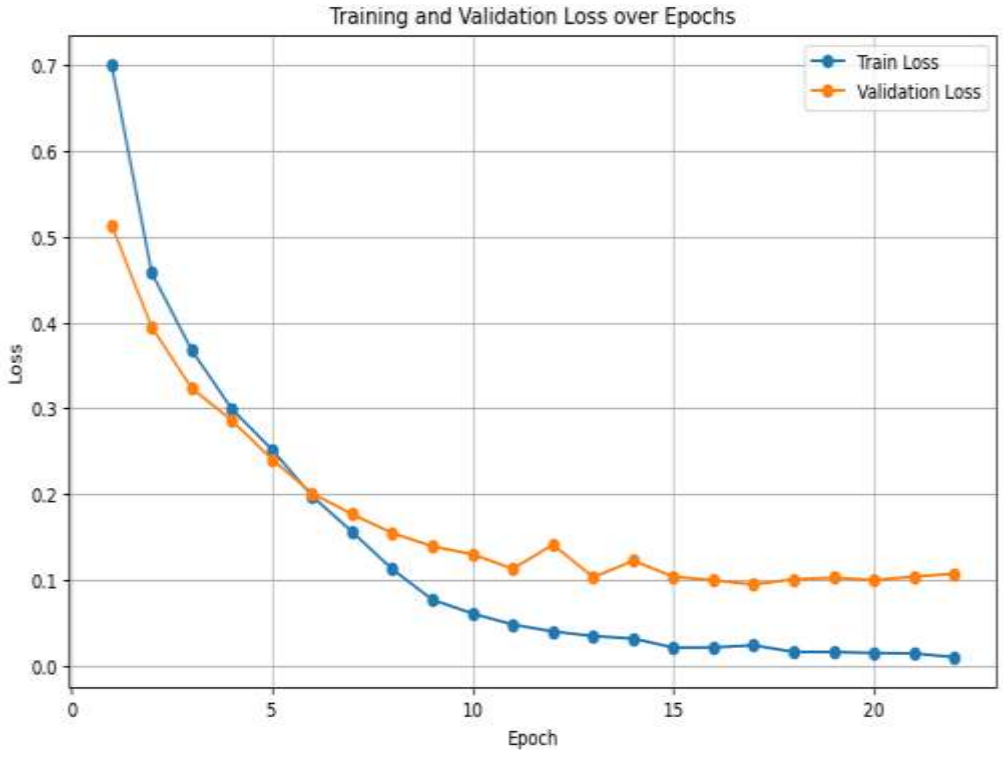


Figure: 4 Loss Curves of ResNet-101 (16 batch size)

5.2.2 ResNet-101 (Batch = 32 & Learning Rate = 0.0001)

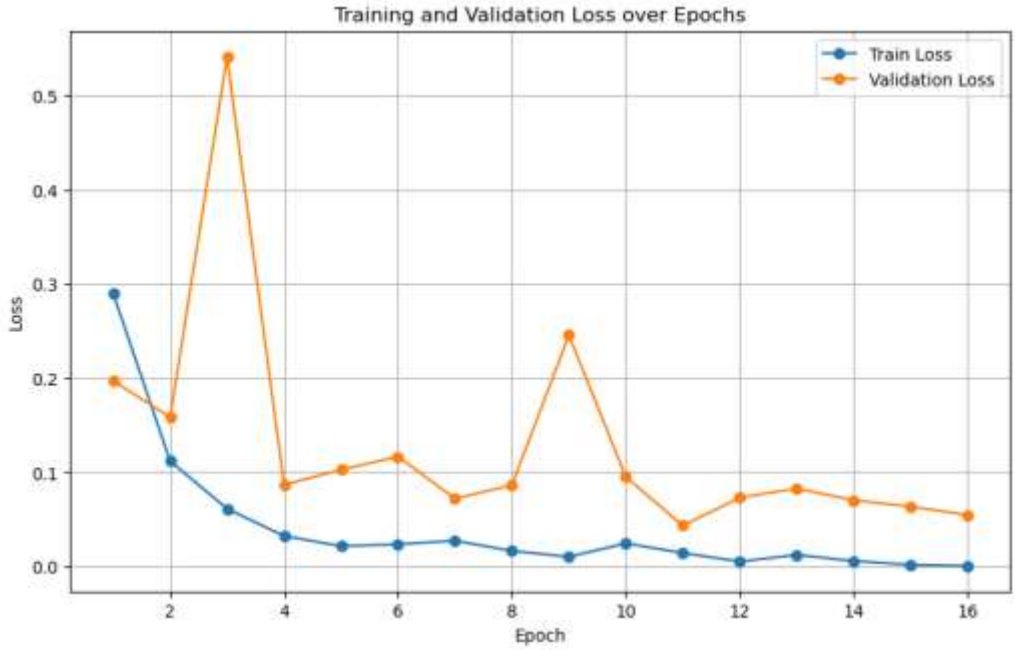


Figure: 5 Loss Curves of ResNet-101 (32 batch size)

5.3 Convolutional Neural Network (CNN)

We loaded the dataset and convert each image into 224x224 size images for CNN model as each image should be of same shape. Then we normalize the dataset based on the mean and standard deviation values of the data. Following this we splitetd the dataset into 80:20 into train and test sets. The test set is further divided into 50:50 as validation and test sets.

We loaded the train, validation and test sets into respective loaders. To use computational resources effectively, we used a batch size of 32x32 and 4 tasks per node in the Super Computer. We then verified the size of each loader. We ran the model on DGX A100 Super computer.

We designed a CNN consisting of 6 convolutional layers each followed by ReLU activation layer and one Maxpool layer for every two convolutional layers. Atlast we have an total of 512 feature maps. Then these are flattend and passed through 3 linear layers and two ReLu and Dropout layers respectively. We have used an learning rate of 0.0001 for this model.

The training process involves optimizing the CNN model on the training dataset while monitoring its performance on the validation set. Cross-entropy loss is employed as the objective function to measure the difference between predicted and true labels. The Adam optimizer is used for weight updates, offering efficient convergence with minimal hyperparameter tuning. As explained earlier, early stopping criteria is employed.

After training, the model is tested on the test set. Classification metrices are calculated and the conduison matrix is visuvalized.

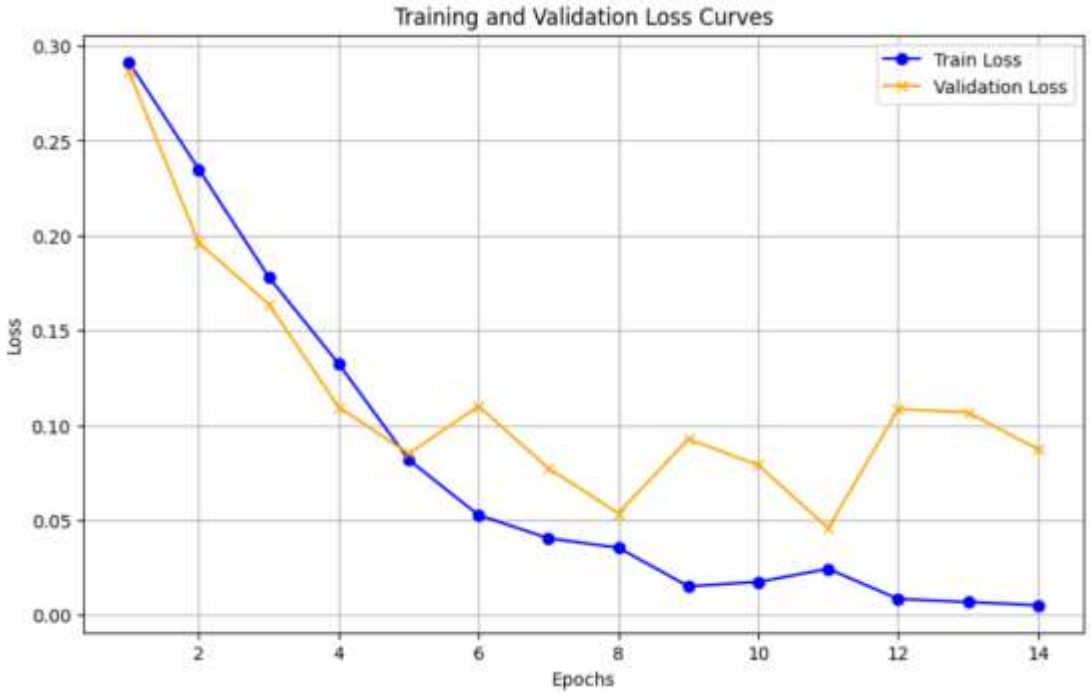


Figure: 6 Loss Curves of CNN

5.4 CNN+CBAM

We loaded the dataset and converted each image into 224x224 size images for this model as each image should be of same shape. Then we normalize the dataset based on the mean and standard deviation values of the data. Following this we splitetd the dataset into 80:20 into train and test sets. The test set is further divided into 50:50 as validation and test sets.

We loaded the train, validation and test sets into respective loaders. To use computational resources effectively, we used a batch size of 8x8 and 4 tasks per node in the Super Computer. We then verified the size of each loader. We ran the model on DGX A100 Super computer.

We used the same CNN model defined above and added an CBAM module after the last Maxpool Layer. The training process involves optimizing the CNN+CBAM model on the training dataset while monitoring its performance on the validation set. Cross-entropy loss is employed as the objective function to measure the difference between predicted and true labels. The Adam optimizer is used for weight updates, offering efficient convergence with minimal hyperparameter tuning. As explained earlier, early stopping criteria is employed.

After training, the model is tested on the test set. Classification metrices are calculated and the conduison matrix is visuvalized.

We have implemented CNN+CBAM in 3 variants.

1. Standard CBAM
2. CBAM (1 input channel)
3. CBAM (3 input channel)

5.4.1 Standard CBAM

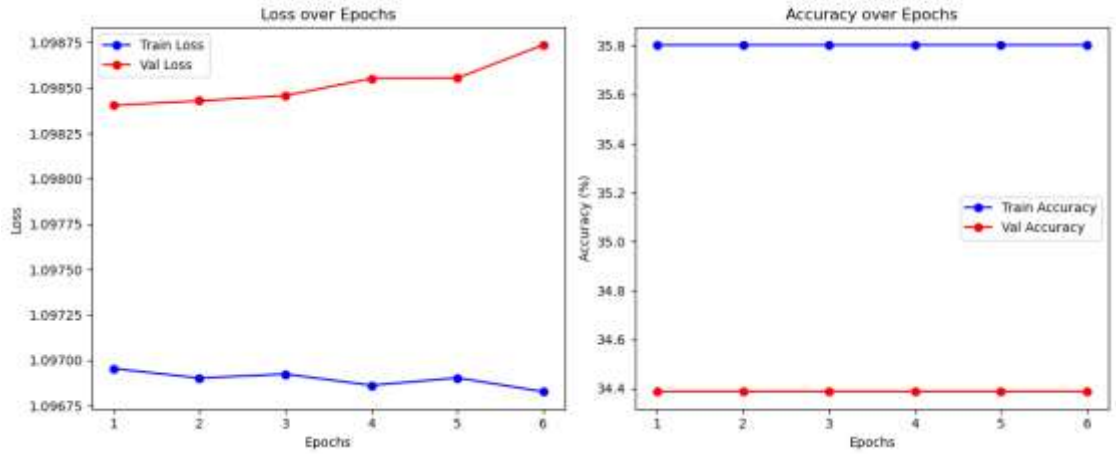


Figure: 7 Loss and accuracy Curves of Standard CBAM

5.4.2 CBAM (1 input channel)

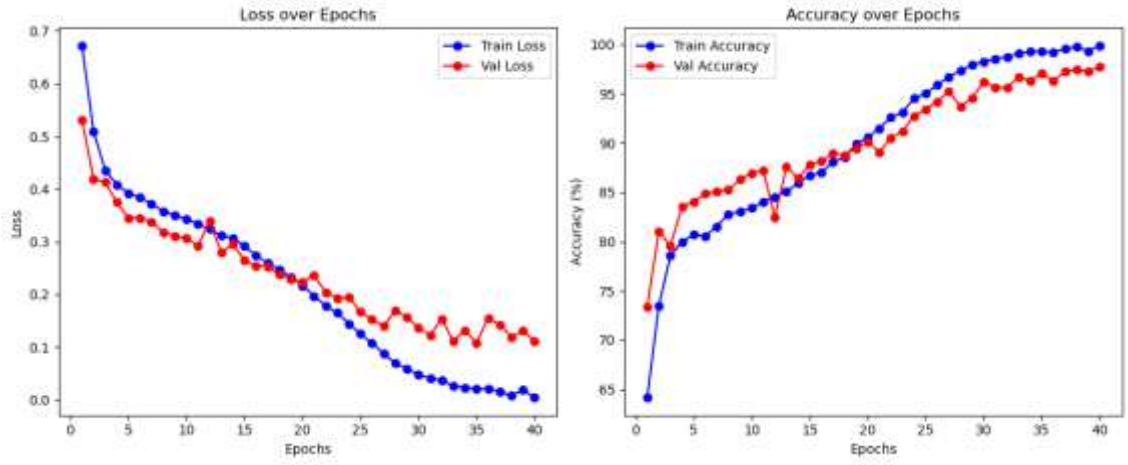


Figure: 8 Loss and accuracy Curves of CBAM (1 channel)

5.4.3 CBAM (3 input channel)

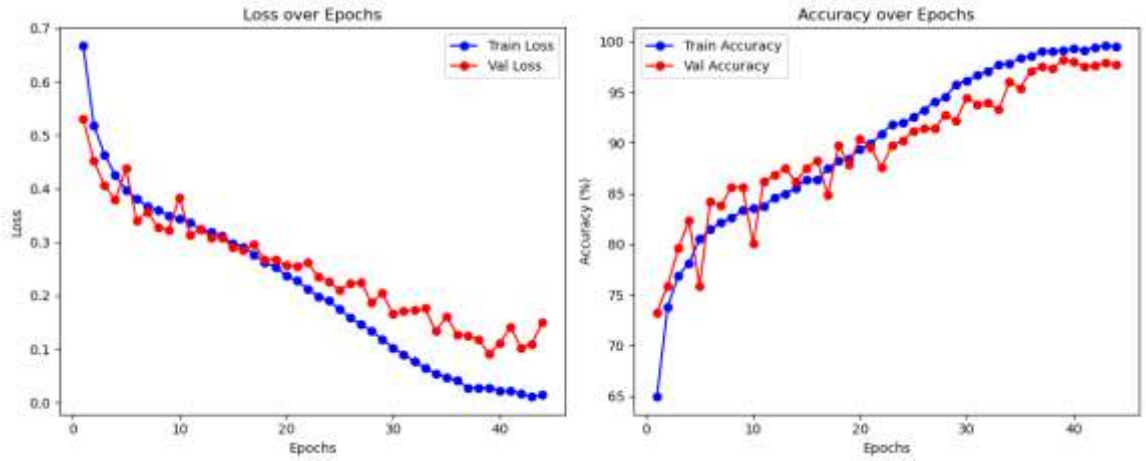


Figure: 9 Loss and accuracy Curves of CBAM (3 channel)

5.5 ViT + CNN (Hybrid)

In this model, we use the same ViT and CNN model defined above but both are used for feature extraction for the input data.

We call the ViT and CNN models consecutively, extracting the local and global features. Now we have the features maps. Both the feature maps are concatenated and flattend. Then they are passed through the 3 linear layers and two ReLU and Dropout layers.

The training and testing is same as the above models.

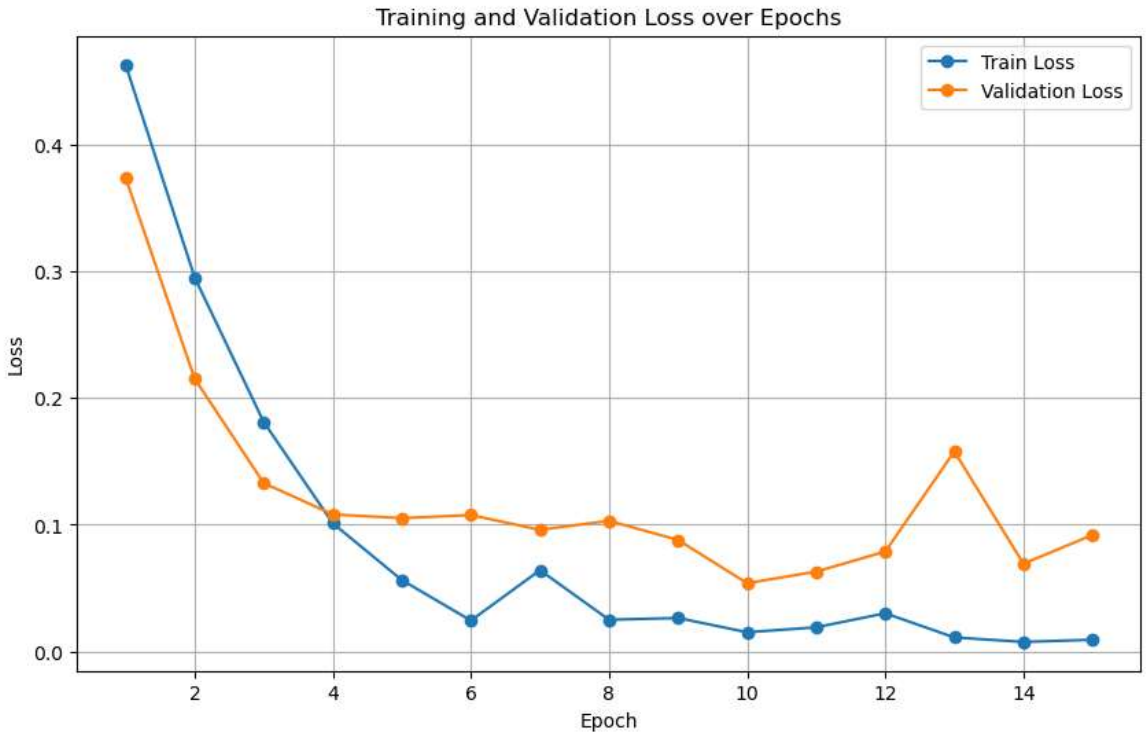


Figure: 10 Loss Curves of Hybrid model

6. Results

Vision Transformers (16 x 16 Patch Size, 7 Epochs)

Accuracy: 0.8248

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.69	0.93	0.79	250
normal	0.90	0.59	0.71	270
parkinsons	0.96	0.98	0.97	239
accuracy			0.82	759
macro avg	0.85	0.83	0.82	759
weighted avg	0.85	0.82	0.82	759

Figure:11 Classification Report of ViT (16x16)

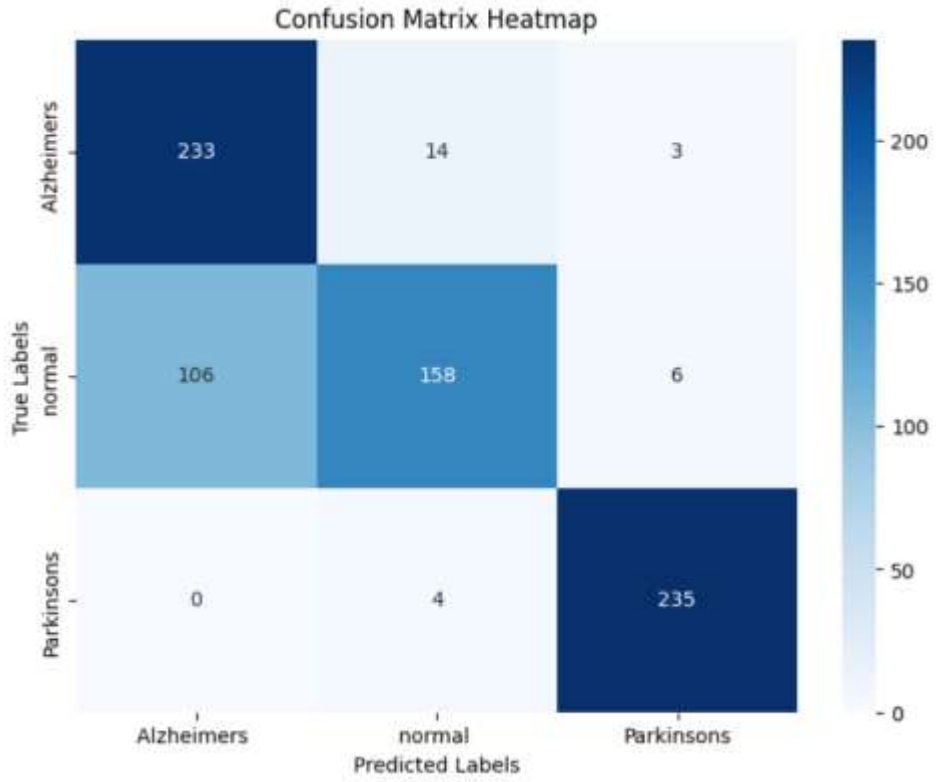


Figure: 12 Confusion matrix of ViT (16x16)

Vision Transformers (8 x 8 Patch Size, 7 Epochs)

Accuracy: 0.8656

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.78	0.88	0.82	250
normal	0.87	0.75	0.81	270
parkinsons	0.97	0.98	0.98	239
accuracy			0.87	759
macro avg	0.87	0.87	0.87	759
weighted avg	0.87	0.87	0.86	759

Figure:13 Classification Report of ViT (8x8)

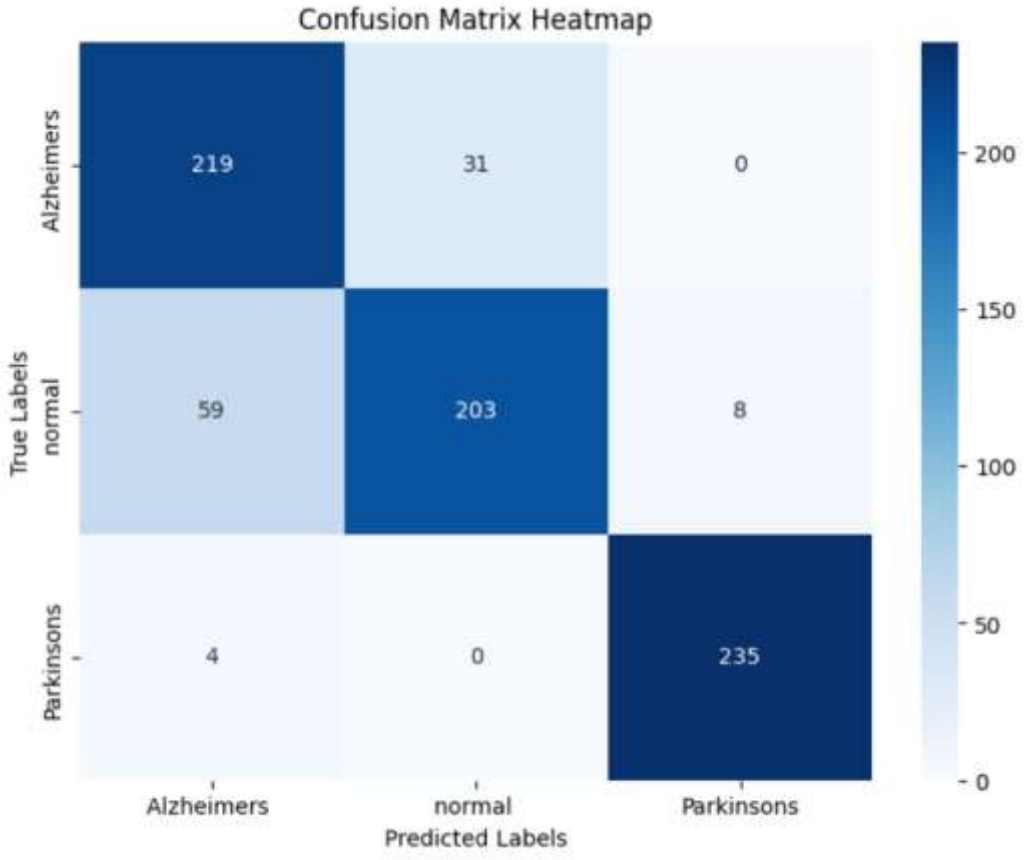


Figure: 14 Confusion matrix of ViT (8x8)

Vision Transformers (4 x 4 Patch Size, 7 Epochs)

Accuracy: 0.8722

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

  alzheimers       0.78      0.92      0.84       250
      normal       0.90      0.73      0.80       270
  parkinsons       0.96      0.99      0.97       239

    accuracy                           0.87       759
   macro avg       0.88      0.88      0.87       759
weighted avg       0.88      0.87      0.87       759
```

Figure:15 Classification Report of ViT (4x4)

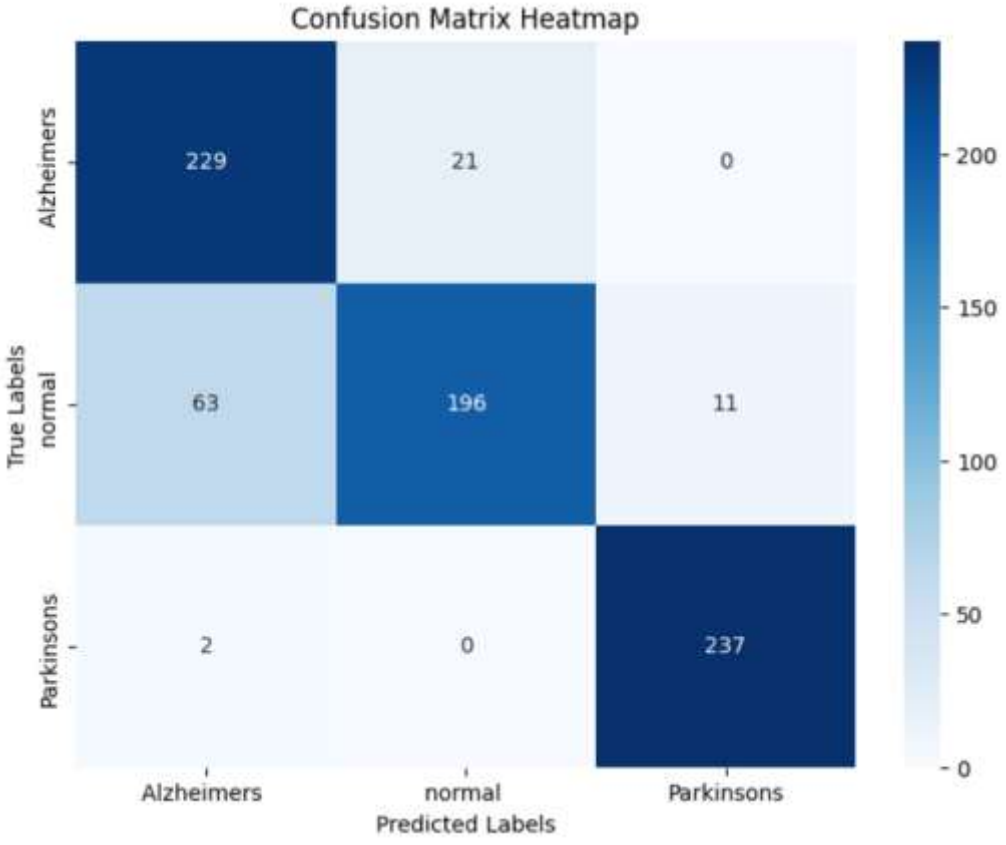


Figure: 16 Confusion matrix of ViT (4x4)

ResNet-101

7 Epochs (Batch = 32 & Learning Rate = 0.0001)

Accuracy: 0.993

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.98	1.00	0.99	250
normal	1.00	0.98	0.99	270
parkinsons	1.00	1.00	1.00	239
accuracy			0.99	759
macro avg	0.99	0.99	0.99	759
weighted avg	0.99	0.99	0.99	759

Figure:17 Classification Report of ResNet-101 (Batch size 32)

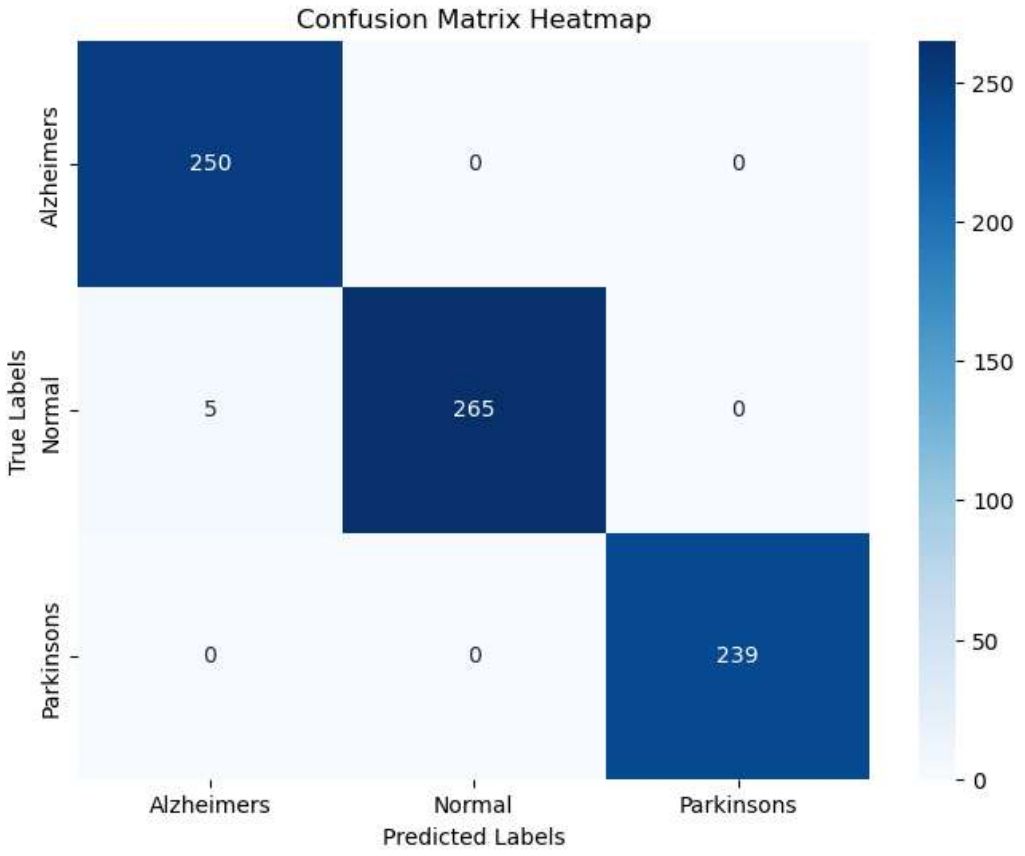


Figure: 18 Confusion matrix of ResNet-101 (Batch size 32)

ResNet-101

7 Epochs (Batch = 16 & Learning Rate = 0.000001)

Accuracy: 0.969

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.93	0.98	0.95	250
normal	0.98	0.94	0.96	270
parkinsons	1.00	1.00	1.00	239
accuracy			0.97	759
macro avg	0.97	0.97	0.97	759
weighted avg	0.97	0.97	0.97	759

Figure: 19 Classification Report of ResNet-101 (Batch size 16)

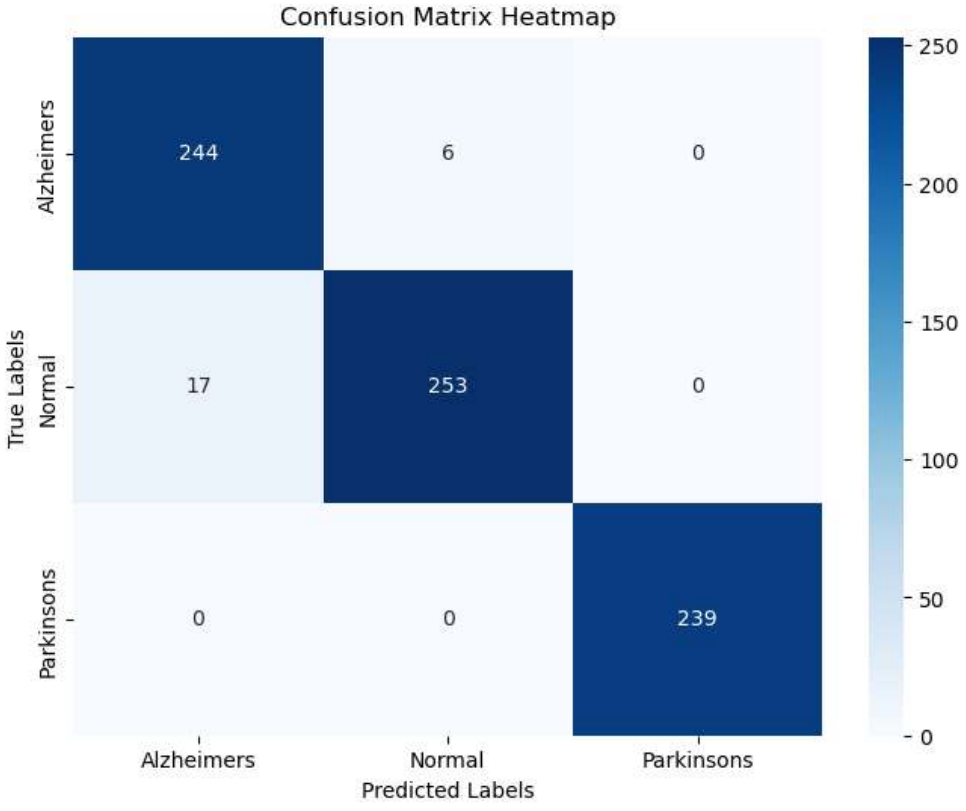


Figure: 20 Confusion Matrx of ResNet-101 (Batch size 16)

CNN

15 Layers 14 Epochs

Accuracy: 0.9829

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.96	0.99	0.97	240
normal	0.99	0.97	0.98	286
parkinsons	1.00	1.00	1.00	234
accuracy			0.98	760
macro avg	0.98	0.98	0.98	760
weighted avg	0.98	0.98	0.98	760

Figure: 21 Classification Report of CNN

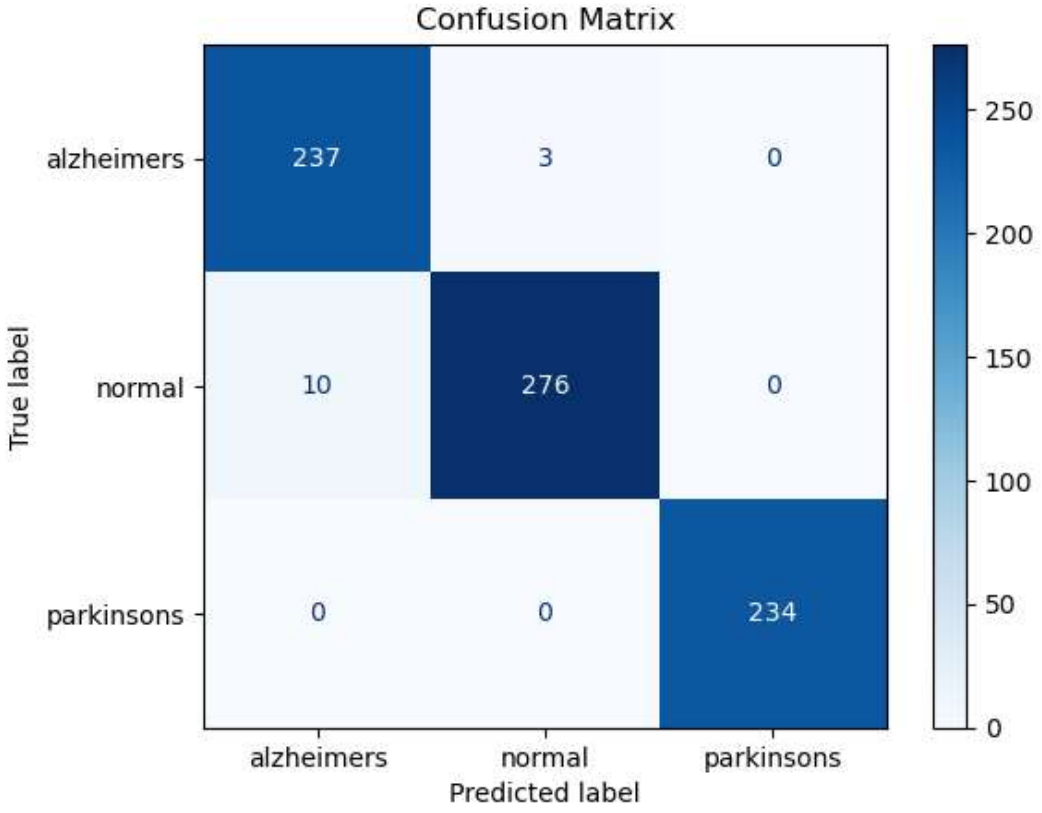


Figure: 22 Confusion Matrix of CNN

CNN+CBAM (Standard)

6 Epochs

Accuracy: 0.3473

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

  alzheimers       0.00      0.00      0.00       231
      normal       0.35      1.00      0.52       264
  parkinsons       0.00      0.00      0.00       265

    accuracy                           0.35       760
   macro avg       0.12      0.33      0.17       760
weighted avg       0.12      0.35      0.18       760
```

Figure: 23 Classification Report of CNN+CBAM (Standard)

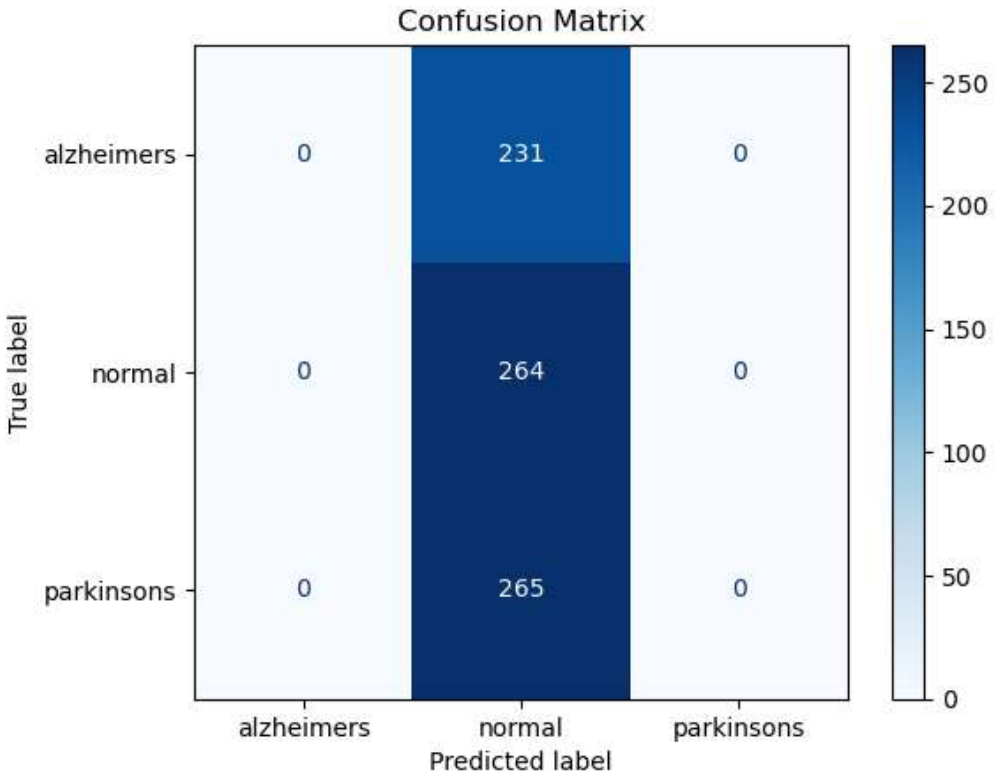


Figure: 24 Confusion Matrix of CNN+CBAM (Standard)

CNN+CBAM (Modified 1 channel)

40 Epochs

Accuracy: 0.971

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

  alzheimers       0.94      0.98      0.96       247
      normal       0.98      0.94      0.96       268
  parkinsons       1.00      0.99      1.00       245

    accuracy                           0.97       760
   macro avg       0.97      0.97      0.97       760
weighted avg       0.97      0.97      0.97       760
```

Figure: 25 Classification Report of CNN+CBAM (1 channel)

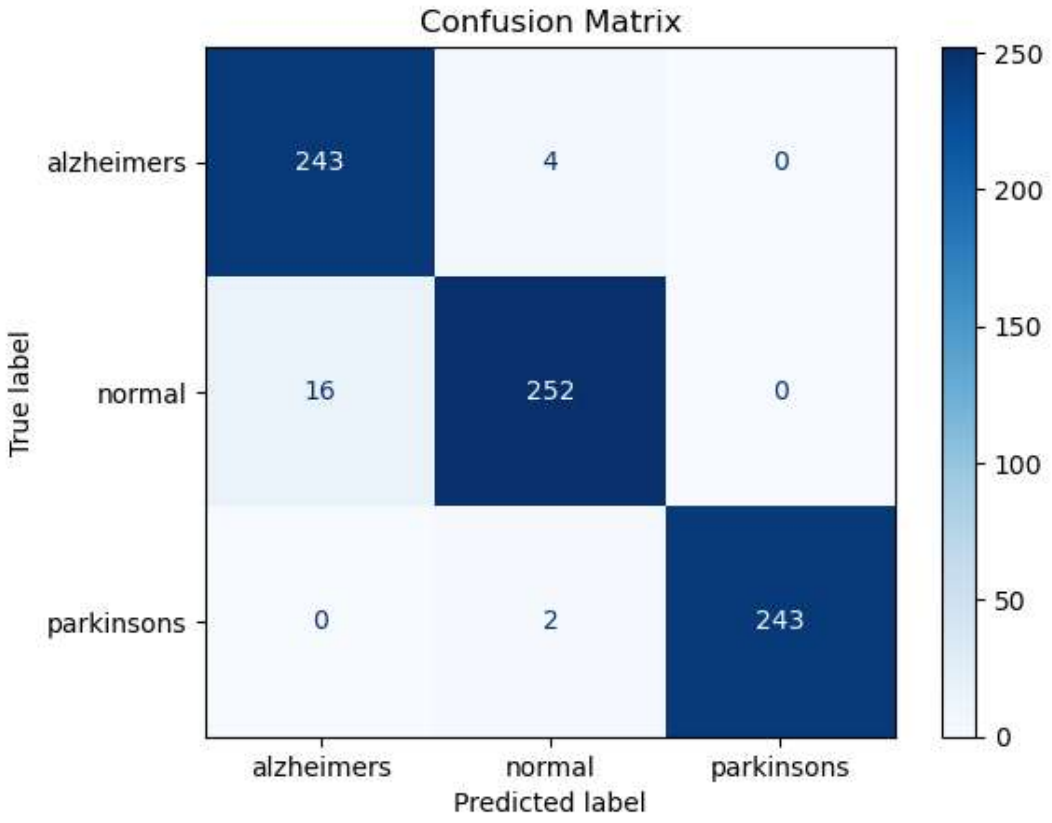


Figure: 26 Confusion Matrix of CNN+CBAM (1 channel)

CNN+CBAM (Modified 3 channel)

25 Epochs

Accuracy: 0.975

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.95	0.98	0.96	243
normal	0.97	0.96	0.97	277
parkinsons	1.00	1.00	1.00	240
accuracy			0.97	760
macro avg	0.98	0.98	0.98	760
weighted avg	0.98	0.97	0.98	760

Figure: 27 Classification Report of CNN+CBAM (3 channel)

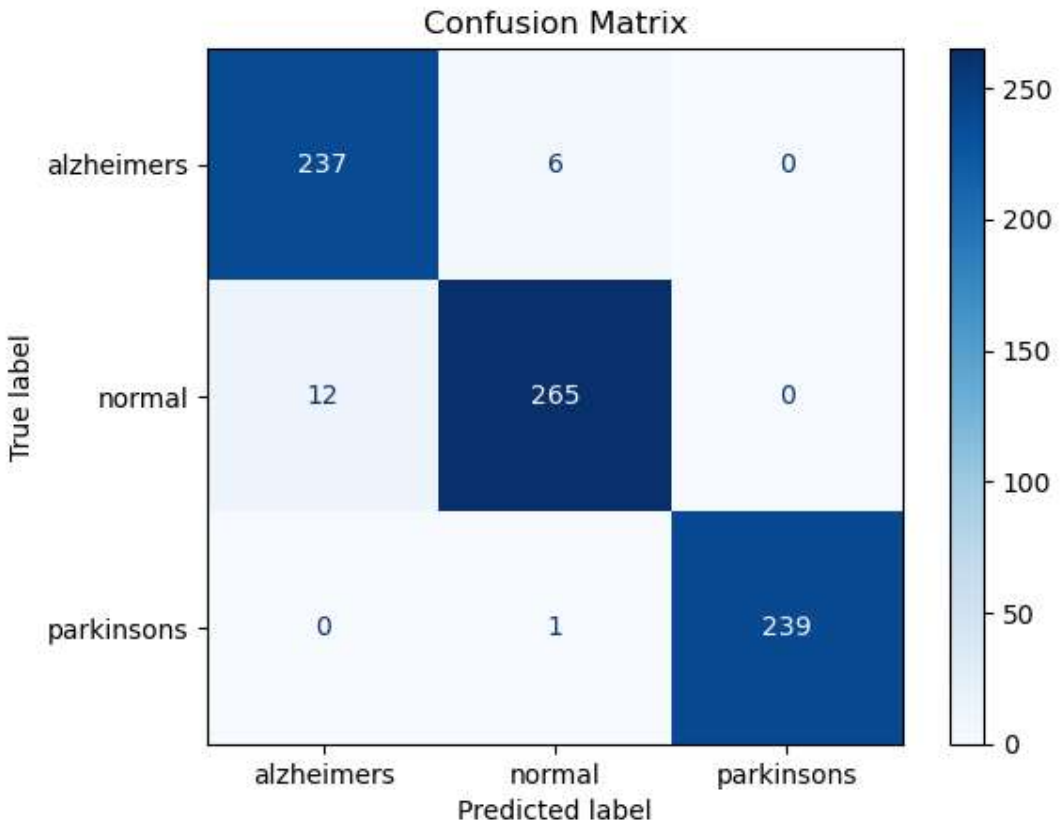


Figure: 28 Confusion Matrix of CNN+CBAM (3 channel)

ViT + CNN (Hybrid)

15 Epochs

Accuracy: 0.9671

Classification Report:

	precision	recall	f1-score	support
alzheimers	0.96	0.94	0.95	250
normal	0.95	0.96	0.95	270
parkinsons	1.00	1.00	1.00	239
accuracy			0.97	759
macro avg	0.97	0.97	0.97	759
weighted avg	0.97	0.97	0.97	759

Figure: 29 Classification Report of Hybrid Model

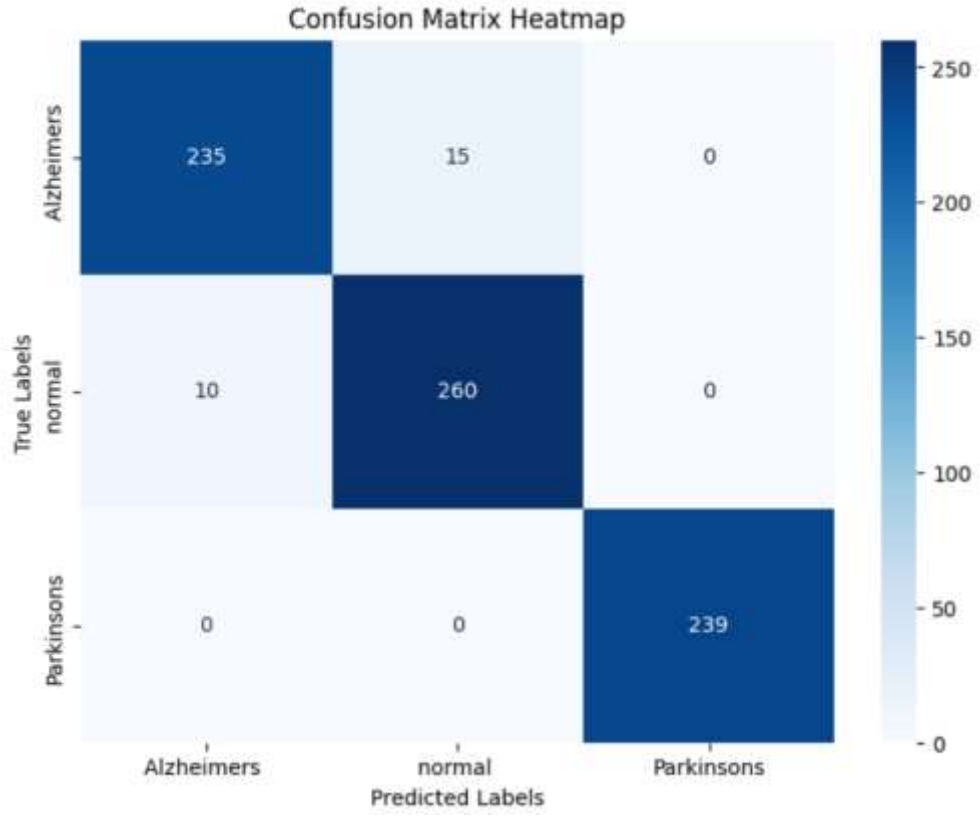


Figure: 30 Confusion Matrix of Hybrid Model

7. Comparision of Results

Here we compared our models results with other models results available online.

Model	Accuracy (%)	Precision (%)	Epochs	Comparison with VGG-16
ResNet 101	99.3	98	7	**+9.3% accuracy**, **+7% precision**, trained in **-29 epochs**. Superior across all metrics compared to VGG-16.
CNN	98.3	98	14	**+8.3% accuracy**, **+7% precision**, trained in **-22 epochs**. Substantial improvement over VGG-16.
CNN+CBAM (Variant 1)	97.1	97	40	**+7.1% accuracy**, **+6% precision**, trained in **+4 epochs**. Less efficient than VGG-16.
CNN+CBAM (Variant 2)	97.25	98	25	**+7.25% accuracy**, **+7% precision**, trained in **-11 epochs**. Balanced improvement over VGG-16.
CNN + ViT (Hybrid)	96.7	95.8	15	**+6.7% accuracy**, **+4.8% precision**, trained in **-21 epochs**. Slight improvement over VGG-16.
ViT 16x16 Patch Size	82.5	81	7	**-7.5% accuracy**, **-10% precision**, trained in **-29 epochs**. Inferior to VGG-16 despite fewer epochs.
ViT 8x8 Patch Size	86.5	87	7	**-3.5% accuracy**, **-4% precision**, trained in **-29 epochs**. Still worse than VGG-16 but closer.
ViT 4x4 Patch Size	87.5	87	7	**-2.5% accuracy**, **-4% precision**, trained in **-29 epochs**. Slightly below VGG-16 but better than other ViT models.
VGG-16	90.0	91	36	**Baseline** for comparison.

Figure: 31 Comparison of Results

8. Analysis of Results

Transformers have revolutionized the NLP field with their self-attention mechanism and outstanding results. This success has extended into the image processing domain through Vision Transformer (ViT) models.

So we initially experimented with ViT on our dataset using different variants, including 16x16, 8x8, and 4x4 patch sizes. We observed a steady increase in accuracy from 82% to 87%. As ViT converts images into patches for processing, reducing the patch size led to better accuracy. This improvement highlighted a limitation of ViT: while it excels at extracting global features, it struggles with local feature extraction. Local features are critical for detecting neurodegenerative diseases such as Parkinson's and Alzheimer's. Consequently, we decided to explore convolutional neural networks (CNNs), which are better suited for extracting local image features.

We began with pre-trained models, leveraging transfer learning to take advantage of their additional knowledge for improved disease detection. We experimented with the ResNet-101 model, opting for 101 layers to enhance feature extraction compared to models with fewer layers. After hyperparameter tuning, we achieved a good accuracy of **99.37%**. ResNet's performance can be attributed to its skip connections, which mitigate overfitting and improve training effectiveness.
Following this, we designed a custom CNN with 15 layers, incorporating dropout layers to prevent overfitting. This model achieved a commendable accuracy of **98.3%**, further confirming that ViT struggled with local feature extraction—a task that both ResNet and our CNN handled better.

Next, we tried of adding attention mechanism and observed that CBAM is computationally efficient than other mechanism like Self attention. we integrated the CBAM (Convolutional Block Attention Module) into the CNN to enhance performance. CBAM applies attention to emphasize important regions where local features are extracted. Initially, we used the standard CBAM code from the 2018 research paper, which yielded a disappointing accuracy of 35%. We then modified CBAM to suit our neurodegenerative dataset.

- For **1 input layer** (standard for grayscale images), we directly fed the data into the model and achieved an accuracy of **97.1%**.
- For **3 input layers**, we duplicated the input layer using a transformation function and fed the data into the model. This approach slightly improved accuracy to **97.5%**.

The addition of CBAM led to a slight decrease in accuracy overall, likely due to the increased model complexity.

After experimenting with CNNs and ViT, we developed a hybrid model combining both approaches. ViT is adept at global feature extraction, while CNN excels in local feature extraction. Combining the two aimed to leverage their strengths for improved performance. We

extracted image features separately using ViT and CNN, concatenated these features, and passed them through multiple fully connected layers. The hybrid model achieved an accuracy of **96.7%**, but this was lower than expected due to feature redundancy, where overlapping features negatively impacted the model's performance.

In conclusion, among all the models we tried, **ResNet-101** achieved the highest accuracy of **99.37%**, demonstrating its superiority in detecting neurodegenerative diseases.

We also observed images from the ResNet-101 model's predictions, including true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Images are attached below.

True Positive Images:

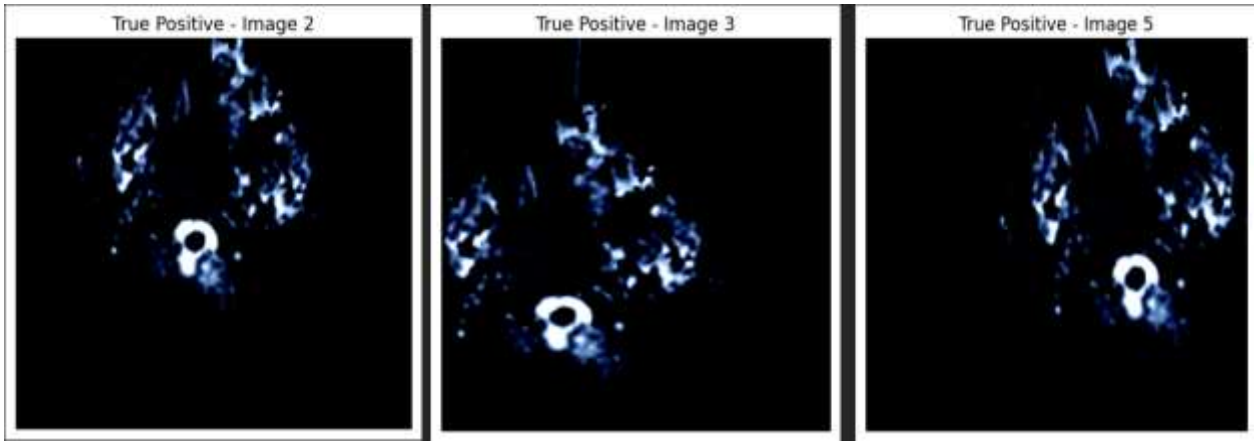


Figure: 32 True Positive Images

These kind of images are of Neurodegenerative diseases. The ResNet model correctly predicts them.

False Positive Images:

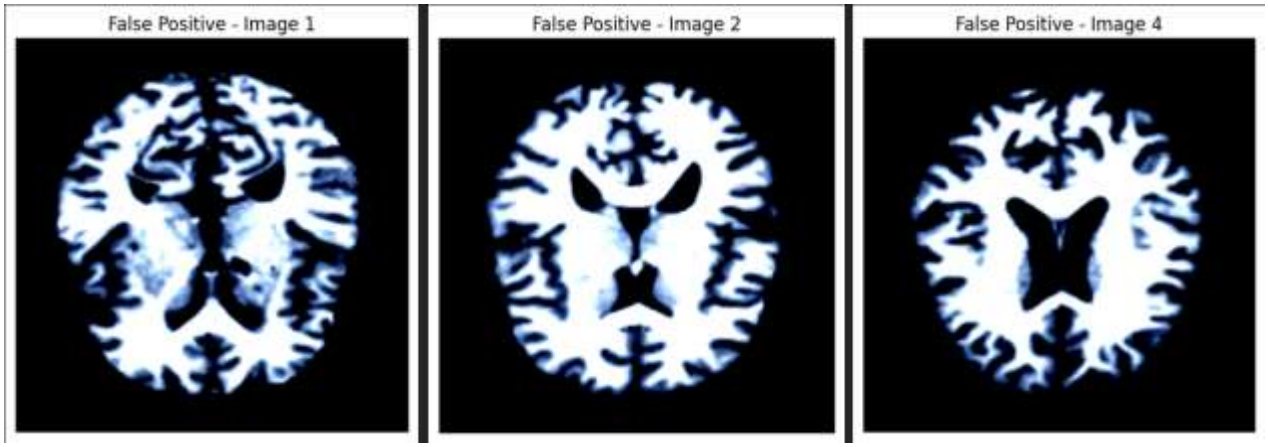


Figure: 33 False Positive Images

It predicts the patient has neurodegenerative disease but actually the patient is healthy.

True Negative Images:

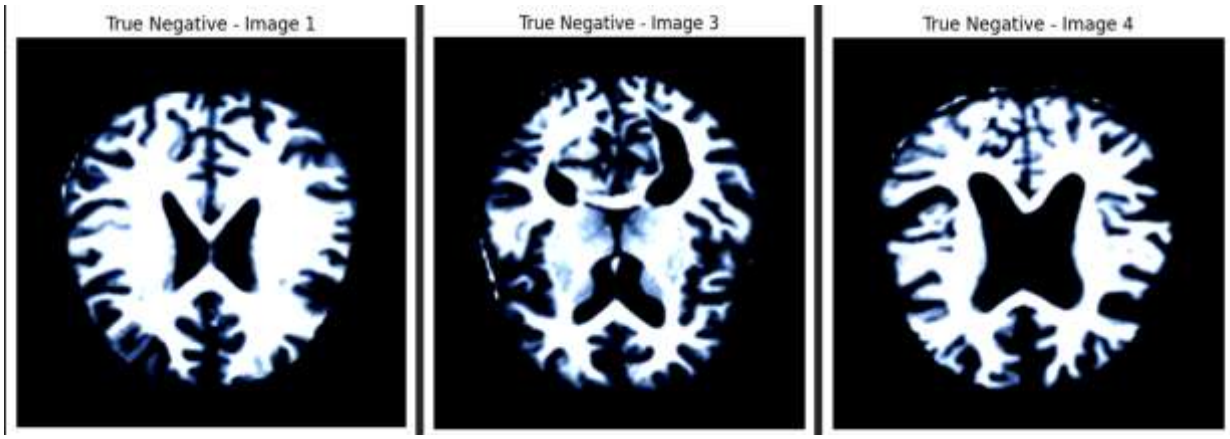


Figure: 34 True Negative Images

Here model correctly predicts an healthy patient as not having the disease.

False Negative Images:

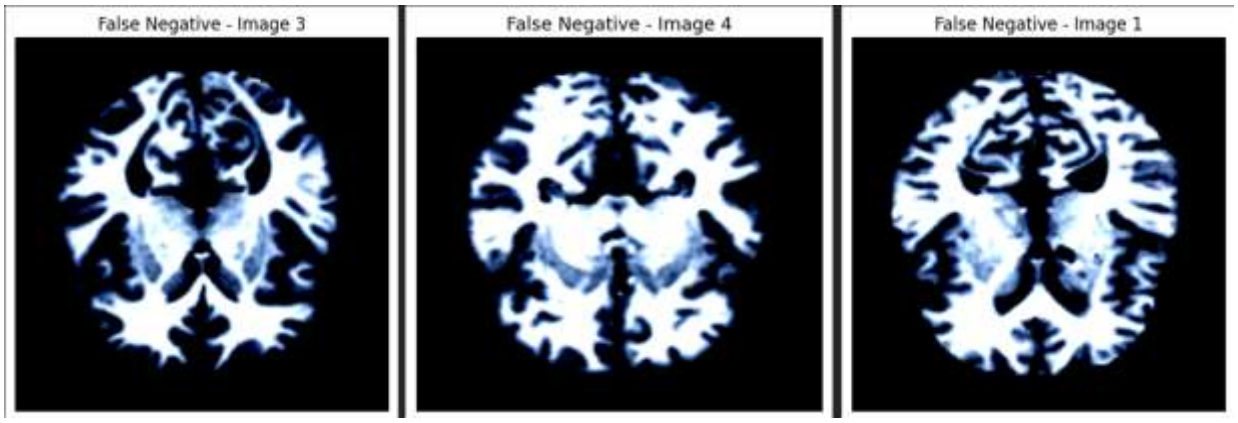


Figure: 35 False Negative Images

This class is very important and it should be prevented. Here model predicts the diseased patient as healthy patient.

9. Conclusion

To conclude, In this study, three deep learning models—Vision Transformer (VIT), RESNET, and CNN+Convolutional Block Attention Module (CBAM)—were implemented and compared in terms of their ability to accurately classify images related to NDDs.

The first model, VIT, leverages self-attention mechanisms to capture long-range dependencies in image data. While the Vision Transformer has shown impressive results in various computer vision tasks, its application to NDD detection posed several challenges. VIT demonstrated a reasonable performance in terms of classification accuracy, especially in scenarios where long-range dependencies in the data were critical. However, when compared to other models, its performance in detecting subtle features related to neurodegenerative diseases was not as robust.

The second model examined was RESNET, a deep convolutional neural network known for its residual connections, which help mitigate the vanishing gradient problem and allow the model to learn deeper representations of data. RESNET was particularly effective in the task of NDD classification due to its ability to capture intricate details in the images, which is crucial when distinguishing between different stages or types of neurodegenerative diseases. RESNET demonstrated superior accuracy and robustness in the classification task compared to VIT and CBAM.

The third model explored was CBAM, a novel attention mechanism that focuses on refining feature maps by adding spatial and channel-wise attention. The goal of CBAM is to improve the representational power of convolutional neural networks by emphasizing the most relevant features in the input data. While CNN+CBAM provided an improvement over traditional convolutional models, its performance did not surpass that of RESNET in the context of NDD classificationAtlast, we developed an hybrid model comprising of CNN and ViT, We call the ViT and CNN models consecutively, extracting the local and global features. And then they are concatened and passed the linear layers. Although the accuracy of hybrid model is good enough, it did not surpass the ResNet-101.

Upon comparing the performance of all models, it was evident that RESNET outperformed both VIT, CNN+CBAM and Hybrid models in terms of accuracy, computational efficiency, and ability to generalize. This success can be attributed to the deep architecture of RESNET, which allows the model to learn rich feature representations from the data. Additionally, the residual connections in RESNET help in retaining information across layers, making it more effective for tasks requiring detailed image analysis, such as detecting minute differences in medical images of the brain or other affected regions. Finally, this study demonstrates that deep learning models, particularly RESNET, have great potential in the automated detection of neurodegenerative diseases.

10. Future Scope

Building on the insights and findings gained through this research, several key directions can be pursued to enhance the depth and applicability of this project. The following future scope outlines potential areas for expansion:

Transition to Real-World Medical Data One of the limitations of this study was the reliance on synthetic or publicly available datasets for training the deep learning models. While these datasets provided valuable insights, real-world medical data from prominent medical hospitals are far more representative of the complexity and diversity found in clinical settings. In the future, we plan to transition to real images and medical data collected from hospitals specializing in neurodegenerative diseases. This would involve collaborating with healthcare institutions to obtain high-quality, annotated datasets containing brain scans, MRI images, and other diagnostic materials. These real-world datasets will enable the models to perform under more challenging, clinically relevant conditions, improving their generalization ability and clinical applicability.

Integration of Explainable AI (XAI) Although deep learning models such as RESNET have shown great promise in the classification of neurodegenerative diseases, one of the significant challenges in deploying these models in medical settings is the lack of interpretability. For deep learning models to be trusted and used effectively by healthcare professionals, it is crucial that their decision-making process is transparent and explainable. To address this, we plan to incorporate Explainable AI (XAI) techniques into this project. These methods will help demystify the model's predictions, enabling healthcare professionals to understand why a particular diagnosis was made.

11. References

1. Aljalal, M., Aldosari, S. A., AlSharabi, K., Abdurraqeeb, A. M., & Alturki, F. A. (2022). Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics*, *12*(5), 1033.
2. Zhou, Z., Kanwal, A., Chaturvedi, K., Raza, R., Prakash, S., Jan, T., & Prasad, M. (2023, July). Deep learning-based classification of neurodegenerative diseases using gait dataset: A comparative study. In *Proceedings of the 2023 International Conference on Robotics, Control and Vision Engineering* (pp. 59-64).
3. El-Latif, A. A. A., Chelloug, S. A., Alabdulhafith, M., & Hammad, M. (2023). Accurate detection of Alzheimer's disease using lightweight deep learning model on MRI data. *Diagnostics*, *13*(7), 1216.
4. Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
5. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).
6. Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.
7. Sheinerman, K. S., & Umansky, S. R. (2013). Early detection of neurodegenerative diseases: circulating brain-enriched microRNA. Cell Cycle, 12(1), 1-2.
8. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.
9. Hansson, O. (2021). Biomarkers for neurodegenerative diseases. *Nature medicine*, *27*(6), 954-963.
10. Hansson, O. (2021). Biomarkers for neurodegenerative diseases. *Nature medicine*, *27*(6), 954-963.