IST 691: Deep Learning in Practice
Final Project Report

DEEP LEARNING- BASED DETECTION AND SEVERITY PREDICTION OF DEMENTIA USING MRI SCANS WITH REGION-SPECIFIC ANALYSIS

Group 4
Resham Bahira
Kulveen Kaur
Nithin Kumar H G
Sudhanshu Pawar

Project Overview and Goals

Millions of individuals worldwide suffer from dementia, including Alzheimer's, and the number of cases is predicted to rise. Effective intervention requires an accurate and timely diagnosis. However, conventional approaches frequently use simple deep learning or machine learning models to categorize or forecast dementia, which are unable to pinpoint the precise brain regions impacted or provide an explanation for their conclusions. This study overcomes these constraints by employing Grad-CAM for visual explanations and CNNs for classification, which makes it particularly helpful for medical imaging.

This project focuses on using Deep Learning concepts like convolutional neural networks(CNN) and Grad-CAM(Gradient weighted Class activation Mapping) to detect and predict severity of dementia using MRI scans. The CNN model classifies the class of dementia(class0: No dementia, class1: very mild demented, class2:mild demented, class3: moderately demneted) and followed by this Grad-CAM offers region-specific analysis by point out the most important brain regions affecting the models' predictions. This combination increases confidence in the model's decision-making process and improves the interpretability of the outcomes.

This capability aligns directly with our obejctive to identify and understand specific the specific brain regions impacted by the disease, aiding medical professionals in gaining deepr insights into dementia's progression.

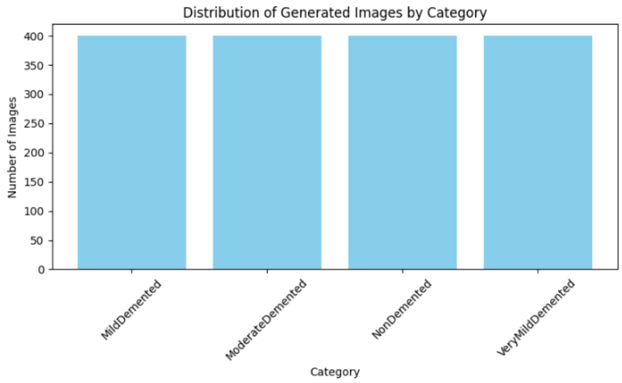
Importance of Grad Cam in Medical diagnosis:

In modern medicine, medical imaging technologies such as computed tomography (CT), X-ray, ultrasound, magnetic resonance imaging (MRI), nuclear medicine, etc., have been proven to provide useful diagnostic information by displaying areas of a lesion or tumor not visible to the human eye, and may also help provide additional recessive information by using modern data analysis methods. These methods, including Artificial Intelligence (AI) technologies, are based on deep learning architectures, and have shown remarkable results in recent studies. However, the lack of explanatory ability of connection-based, instead of algorithm-based, deep learning technologies is one of the main reasons for the delay in the acceptance of these technologies in the mainstream medical field. One of the recent methods that may offer the explanatory ability for the CNN classes of deep learning neural networks is the gradient-weighted class activation mapping (Grad-CAM) method, which produces heat-maps that may offer explanations of the classification results.

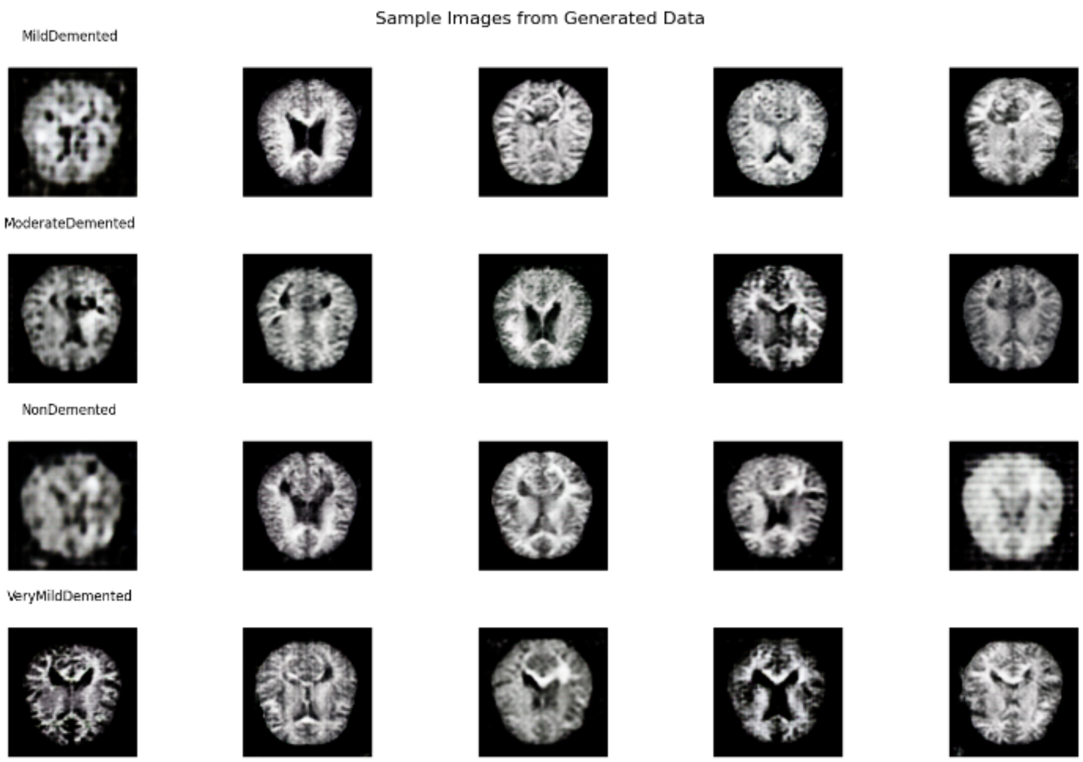
Computer-aided diagnosis technology based on the deep learning greatly increases the efficiency of medical diagnosis tasks. However, the black-box nature of deep neural networks reduces the reliability of auxiliary clinical diagnosis, so it is necessary to explain the medical imaging diagnosis model based on deep learning. The Gradient-weighted Class Activation Mapping (Grad-CAM) can determine the contribution of input features to the result of the classification task and visualize these contributions as heatmaps.

Dataset Summary and Visualizations

- **Source of Data:** The dataset is sourced from Kaggle that includes original MRI images for dementia detection having 4 classes.
- **Original Dataset Size:** Initially, the dataset contains 6,400 images categorized into four dementia severity classes: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented having significant class imbalance.
- **Augmented Dataset Size:** Augmentation via Deep Convolutional GAN (DCGAN) expanded the dataset to 6400 images, with 400 images per class for better balance.



Sample Images from Generated Augmented Data



Results

Performance metrics

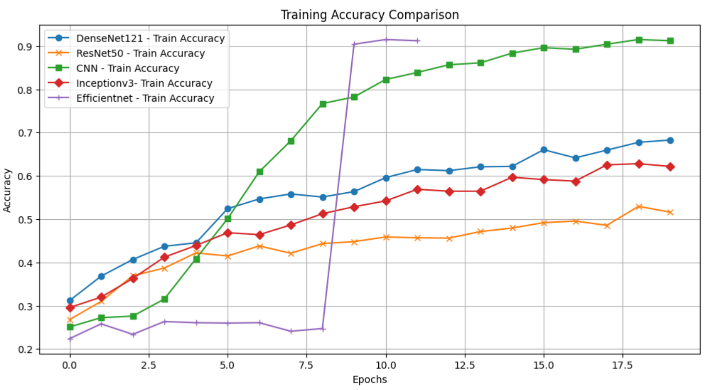
1. **Precision (Macro Average)-** Precision is a metric that measures how often a machine learning model correctly predicts the positive class. Macro average calculates each class's performance metric and then takes the arithmetic mean across all classes.
2. **Recall -** Recall is a metric that measures how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset.
3. **F1-Score** – F1 score computes the average of precision and recall, where the relative contribution of both of these metrics are equal to F1 score
4. **Accuracy -** Accuracy shows how often a classification ML model is correct overall.

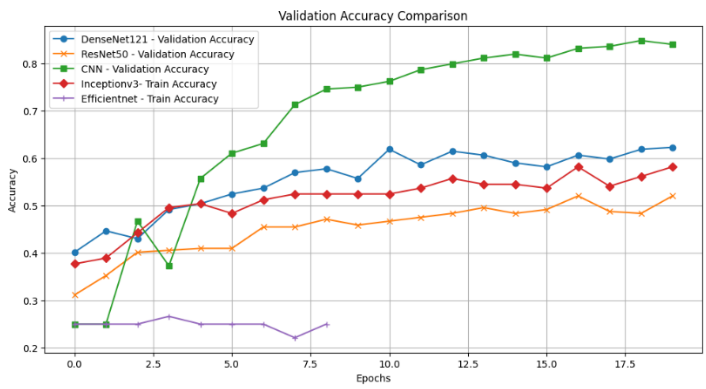
We can observe from the metrics that Cnn is the most effective model for classifying our dementia into 4 classes followed by InceptionV3 and Densenet

Model	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)	Accuracy
CNN	0.87	0.87	0.87	0.87
ResNet	0.49	0.49	0.49	0.49
DenseNet	0.55	0.55	0.55	0.55
InceptionV3	0.57	0.57	0.56	0.57
EfficientNet	0.15	0.28	0.16	0.28

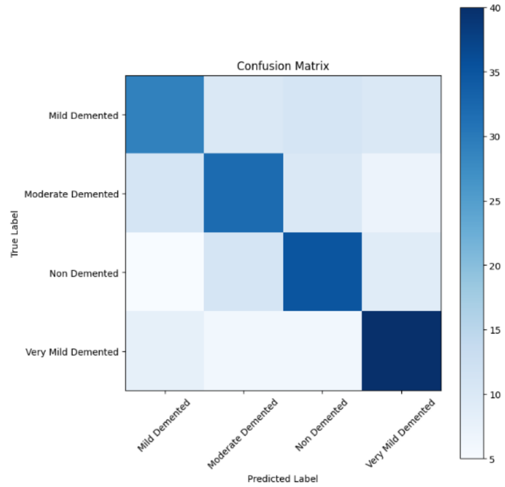
Training And validation phase

1. We can see from the first graph that most of the models are improving over epochs in their training phase apart from EfficientNet
2. Validation Accuracy graph signifies that CNN and DenseNet performs very well on unseen data while efficient and Resnet struggles in generalization.
3. There is a small gap between the training and validation accuracies of Cnn and DenseNet which tells us that they are performing a balanced learning while others undergo overfitting or underfitting,

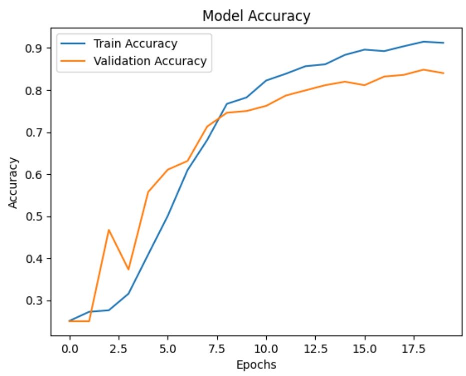


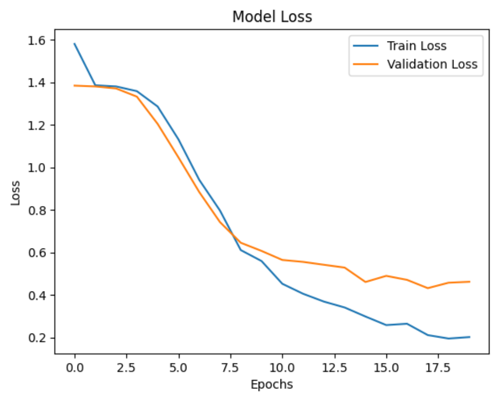


Model Results Summary

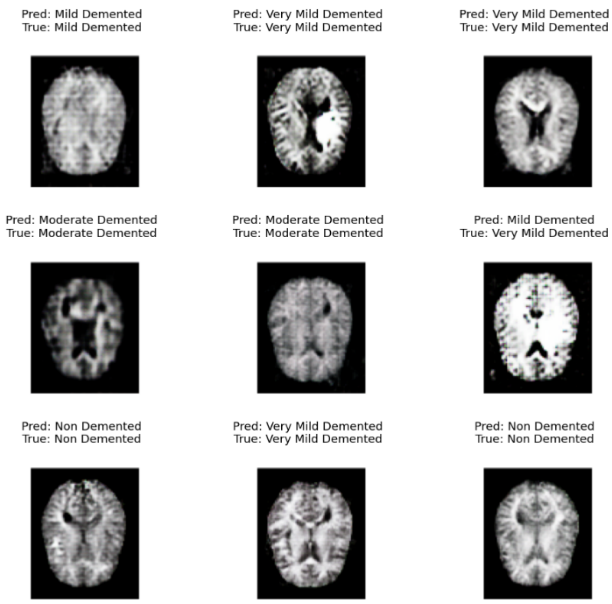


- After reviewing the performance metric, Cnn proves to be the most effective model among all giving us the best performance metric on dementia classification model.
- The confusion matrix of CNN model tells us that among all classes, CNN model is classifying the '**Very Mild Demented** 'class most accurately.
- The upward trend in the model accuracy graph of training reaching 90% and validation 80% signifies that the model exhibits strong performance.
- Also, the validation accuracy is very close to the training accuracy which indicates that the model is performing well on the unseen data making it capable for generalization.
- The overall performance from both the graphs suggests that the model is performing well without going underfitting or overfitting.



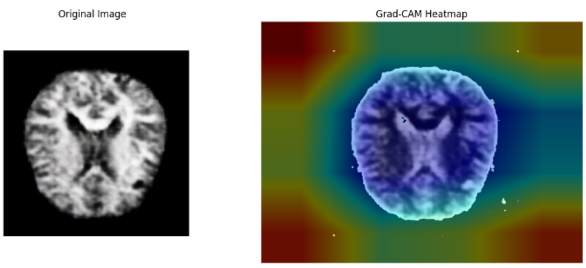


Below is the prediction visualization showing what class the model is predicting and what the true class is. We can see that out of 9 only one prediction is incorrect which implies model is correctly predicting the data**.**

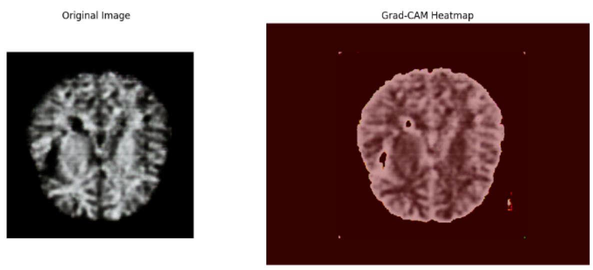


GRADCAM RESULTS

- We achieved the highest accuracy (87%) and precision using CNN model and, making it highly effective for diagnosing and predicting the severity of dementia
- To visualize our model, we generated the Grad-CAM heatmap that uses blue color to indicate areas of high importance, highlighting brain regions that significantly contribute to the model's predictions about dementia severity or progression.
- The figure below shows the image of very mild dementia in which GradCAM is highlining the blue area of brain suggesting that this region is highly influencing the models' predictions.
- The second image is the non demented image which has red heatmap signifying that since is it non demented it is not much influencing the dementia prediction.
- This helps the clinicians and researchers to look at these areas for further analysis investigation and analysis of the disease for better results.
- This also ensures that how well our deep learning model combines with the clinical knowledge.



Very Mild Demented Image



Non-Demented Image

METHODS

Model Architecture:

- Our project used Convolutional Neural Network (CNN) with convolutional layers, each followed by pooling layers, batch normalization, and dropout regularization.
- To maximize performance, we experimented with various CNN architectures, including ResNet, DenseNet , InceptionV3 and EfficientNet. By comparing accuracy and computational efficiency across these architectures, we selected the best model optimal for our multiclass task.

Augmentation and Data Preprocessing:

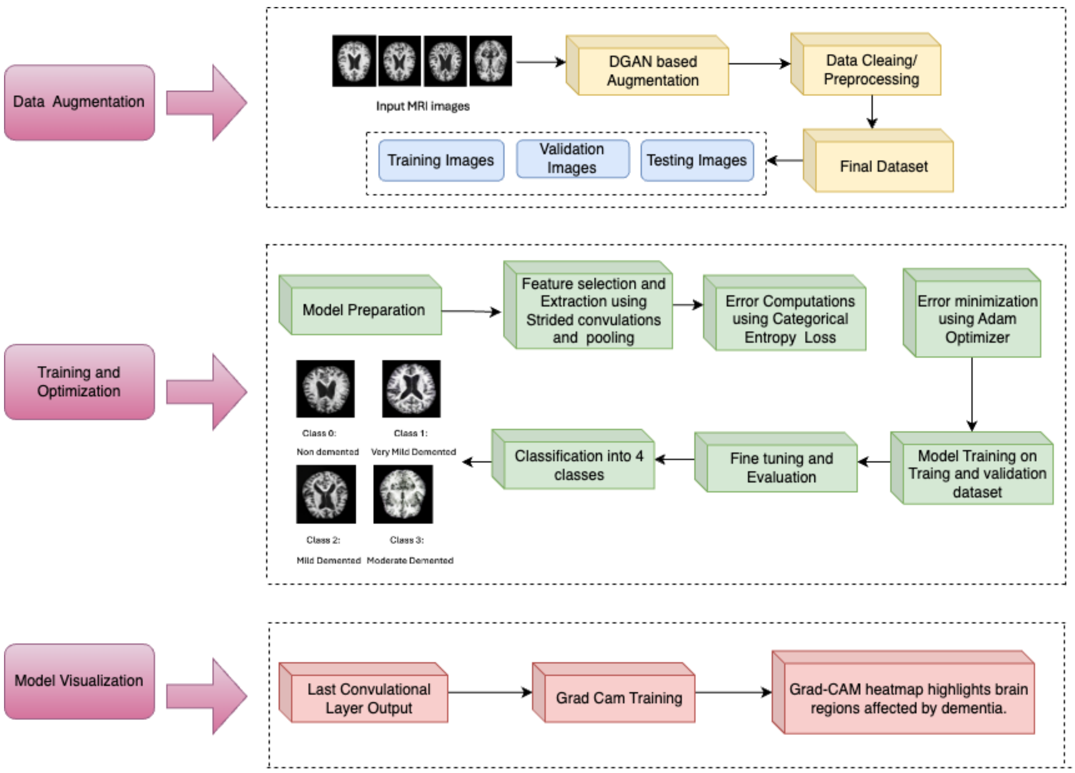
- We used DCGAN (Deep Convolutional GAN) to generate synthetic MRI to address the class imbalance in the original data.
- DCGAN helped in maintaining the spatial features required for effective classification, such as brain regions that may indicate dementia severity.

GradCAM Visualization:

- To provide transparency and interpretability, the model integrated with the GradCAM-based visualization.

- It highlighted the areas in MRI scans that influence model predictions, visually mapping the stages of dementia across heatmaps.
- This capability will help medical professionals understand model decisions, facilitating trust and potential use in clinical settings.

Evaluation Metrics:
Our evaluation metrics includes accuracy, precision, recall, and F1-score giving us the moe=st effective model for our task.
Below is the pipeline used in our project.



CHALLENGES

- The original dataset was extremely unbalanced and contained only a small number of MRI images, which made it difficult to train the model efficiently. To solve this, we used DGAN (Deep Generative Adversarial Networks) for data augmentation; however, this procedure was quite time consuming and needed a large amount of GPU power to produce high-quality augmented images.
- Initially, we struggled with the augmentation process due to a lack of access to powerful GPUs. This limitation significantly slowed down our progress and delayed the expansion of the dataset. Eventuallyy, with proper GPU access provided to us (Azure for ML), we were able to efficiently augment the images, increase the dataset size, and achieve better model accuracy. This resource improvement was critical to the success of the project.

ACHIEVEMENTS

1. We have achieved 87% accuracy on classifying dementia into its 4 classes.
2. We have successfully implemented GradCAM heatmap visualization to show model interpretability and prove clinical relevance shows regions influencing model predictions.
3. We have addressed the issue of class imbalance by using Deep Convolutional GAN (DCGAN) augmentation technique.
4. We have made the model generalized by having less gap between training and validation accuracy.

FUTURE WORK

1. We can use other augmented techniques like StyleGAN to further improve accuracy.
2. We can also use multimodal techniques to combine non-image data (Patient History) and image data (MRI Scans) for better predictions.
3. We can also expand our model and generalize it to predict other neurodegenerative diseases like Parkinson's or Huntington’s.

CITATIONS

- https://www.sciencedirect.com/science/article/pii/S1110016822005191
- https://www.sciencedirect.com/science/article/pii/S2543106424000152
- https://www.nature.com/articles/s41467-022-31037-5
- https://www.nature.com/articles/s41598-022-20674-x?fromPaywallRec=false
- https://www.nature.com/articles/s41598-020-79243-9?fromPaywallRec=false
- https://www.nature.com/articles/s41467-022-31037-5?fromPaywallRec=false
- https://www.mdpi.com/2076-3417/12/15/7748
- https://ieeexplore.ieee.org/document/9587953