Deep Learning–Based Multi-Class Alzheimer's Disease Classification from Brain MRI

Defining the Problem

Alzheimer's Disease (AD) is a progressive brain disorder characterised by a reduction in cognitive ability (e.g., thinking, reasoning, learning, and remembering) in addition to memory loss as a result of brain degeneration or damage. The detection and diagnosis of Alzheimer's Disease will be more clinically useful if we are able to identify the disease stage and intervene early and clinically.

Structural brain changes due to Alzheimer's Disease can be detected through the use of MRI technology. MRI, however, is currently performed by trained radiologists, who interpret MRI scans manually, which is labour-intensive and prone to inter-observer variability when interpreting scans from multiple radiologists.

This project will primarily focus on the development of a method of classifying Alzheimer's Disease into four clinically significant categories by analyzing brain MRI images using an automated and progressive approach. The proposed categories of MRI scan classification are Non-demented, highly mildly demented (very mild dementia), mildly demented (early stage dementia), and moderately demented (mid-stage dementia). The goal of the project is to provide Clinical Decision Support using Deep Learning techniques to produce a more efficient and predictable method of assisting in clinical decisions.
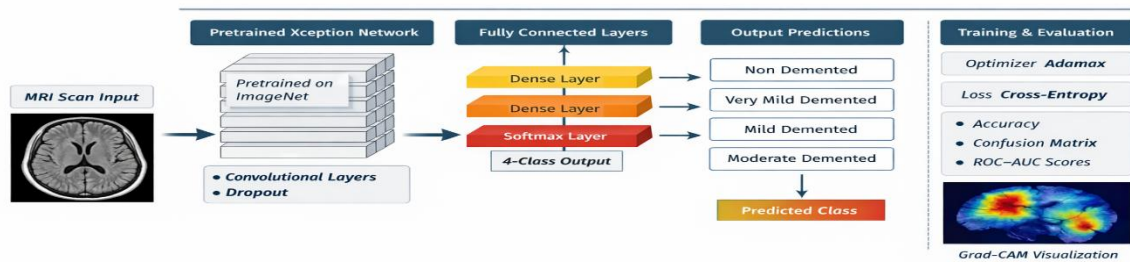
Methods and Evaluation

Dataset and Preprocessing

The research work is based on an enlarged dataset of Alzheimer's MRIs which consists of labeled brain MRI pictures of four different stages of the disease. A structured dataset was made by organizing the image file paths and labels. The dataset was divided into training, validation, and testing sets by stratified sampling to guarantee unbiased competition and to avoid class imbalance bias. All MRI images were beforehand training resized to a constant resolution and normalized.

Model Architecture

The transfer learning strategy was utilized with the Xception convolutional neural network, which had been pretrained on the ImageNet dataset. The pretrained feature extractor was kept unchanged, and the final classification layers were substituted with

fully connected layers specific to the task for the four-class prediction. Overfitting was reduced through the application of dropout regularization.

Fig-1 Model Architecture

Training Strategy

- Optimizer: Adamax with learning rate 32-

- Loss function: Cross-entropy loss

- Training setup: GPU/CPU compatible PyTorch pipeline

- Epoch-wise monitoring of training and validation performance
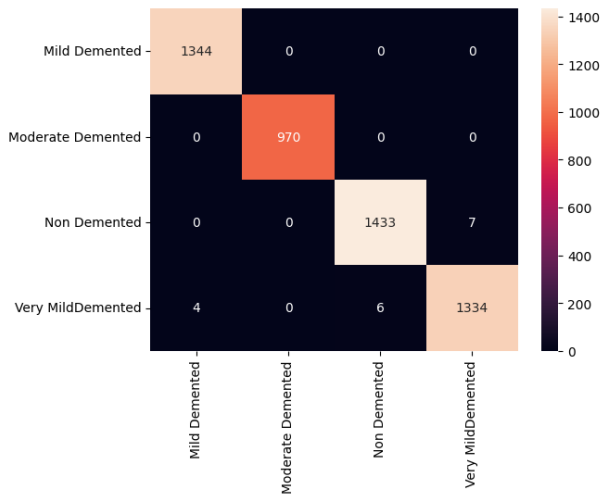
Evaluation Metrics

Model performance was evaluated using:

- Overall classification accuracy

- Confusion matrix for class-wise analysis

- Multi-class **ROC–AUC (one-vs-rest)** to measure discriminative performance
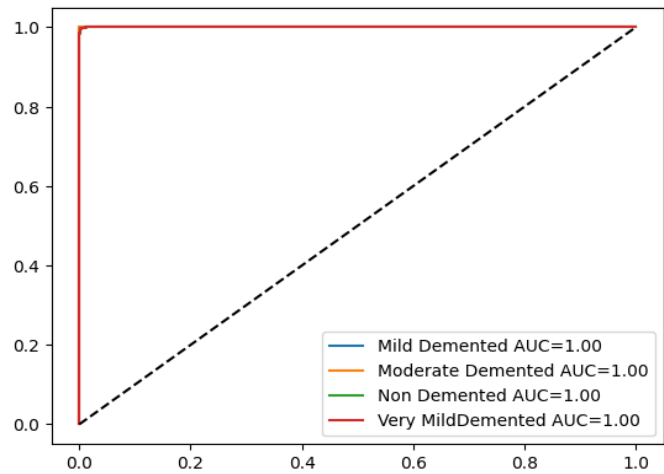
Grad-CAM was used only as a **qualitative, post-hoc analysis tool** to verify model attention and was not included as a core evaluation metric.

Results:

The proposed Xception-based deep learning model achieved excellent performance in multi-class Alzheimer's disease classification from MRI scans. Confusion matrix as in Fig-2 analysis showed near-perfect classification across all four stages, with only minimal confusion between Non-Demented and Very Mild Demented classes. The model attained a ROC–AUC score of 1.00 for each class using a one-vs-rest strategy as clearly inferred in Fig-3, indicating strong discriminative capability. Grad-CAM visualizations in Fig-4 further confirmed that the model attends to meaningful brain regions, supporting the validity of its predictions.
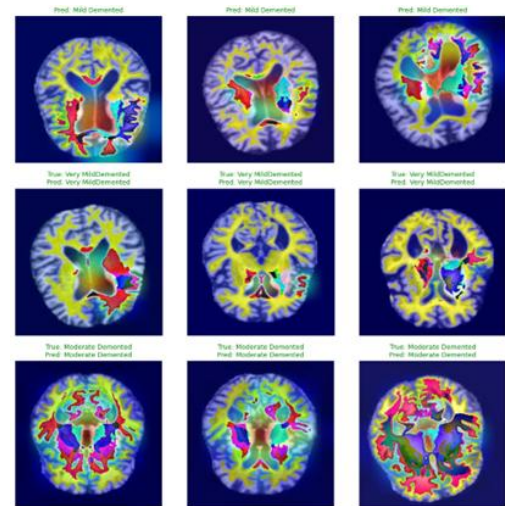
Fig-2 Confusion Matrix   **Fig-3** AUc_ROC curve

Limitations

Despite promising results, several limitations remain:

- The model was trained and evaluated on a single dataset, which may limit generalization to data from other scanners or clinical centers.
- Subtle inter-class differences in early-stage Alzheimer's disease remain challenging even for deep learning models.

Fig-4 Gradcam evaluation

These limitations indicate that further validation and dataset diversity are required before clinical deployment

Next Steps and Societal Impact:

The upcoming research will be aimed at not only but extending the model to 3D MRI volumes and longitudinal scans which will lead to the richer and more complex capturing of spatial and temporal disease patterns. Besides, there will be performance validation across clinical datasets from different centers, thus improving generalization. Moreover, the addition of clinical metadata and the optimization of lightweight architectures will further allow the real-time clinical deployment. From the societal point of view, such automated classification systems for Alzheimer's disease have the ability to support early screening, reduce diagnostic time, and even lighten the workload of the clinicians dealing with the patients in large numbers. If they are used responsibly as decision-support tools, they can not only help to standardize the diagnosis but also improve access to neurological care and thus contribute to more timely and effective patient management.