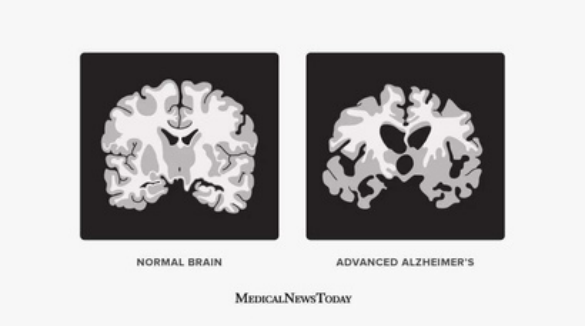
Deep Learning-Based Approach for Alzheimer's Disease Detection and Explainability

Sai Saranya Mulukutla

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that significantly impacts cognitive functions, necessitating early detection for timely intervention. However, the intricate structural changes in the brain make MRI-based diagnosis challenging. This study proposes a deep learning-based approach incorporating a DenseNet121 model enhanced with a Convolutional Block Attention Module (CBAM) to refine feature extraction. Additionally, Local Interpretable Model-Agnostic Explanations (LIME) are utilized to improve the interpretability of model predictions, enabling clinicians to understand the basis of classification decisions. The proposed model is evaluated on the OASIS dataset, demonstrating high classification accuracy while effectively handling class imbalance and overfitting challenges. The findings indicate that integrating attention mechanisms and explainable AI significantly enhances the model's reliability, making it a promising tool for computer-aided Alzheimer's diagnosis from MRI scans.



Introduction

Alzheimer's disease (AD) is the leading cause of dementia, affecting over 55 million people globally—a number expected to reach 139 million by 2050. It severely impacts memory and daily functioning, placing a growing economic burden on healthcare systems. Early detection is crucial for effective management. While MRI is useful for identifying brain abnormalities linked to AD, manual analysis is slow and subjective. This study proposes a deep learning-based solution using transfer learning, attention mechanisms, and Explainable AI (LIME) to improve the accuracy, interpretability, and automation of AD diagnosis from MRI scans.

Problem Statement

Early and accurate detection of Alzheimer's disease (AD) remains a critical yet challenging task due to the subtle and complex anatomical changes in brain MRI scans. Traditional diagnostic methods are time-consuming, subjective, and often lack consistency, while existing AI-based solutions face limitations such as high computational cost, overfitting, lack of interpretability, and poor generalization across datasets. There is a pressing need for a robust, lightweight, and explainable deep learning framework that can effectively identify early signs of AD from MRI data, while ensuring transparency and trustworthiness to support clinical decision-making.

Methodology

This study adopts a comprehensive deep learning-based pipeline for the early detection of Alzheimer's Disease (AD) using MRI scans. The methodology consists of five main stages: dataset preparation, model architecture design, training setup, evaluation, and interpretability.

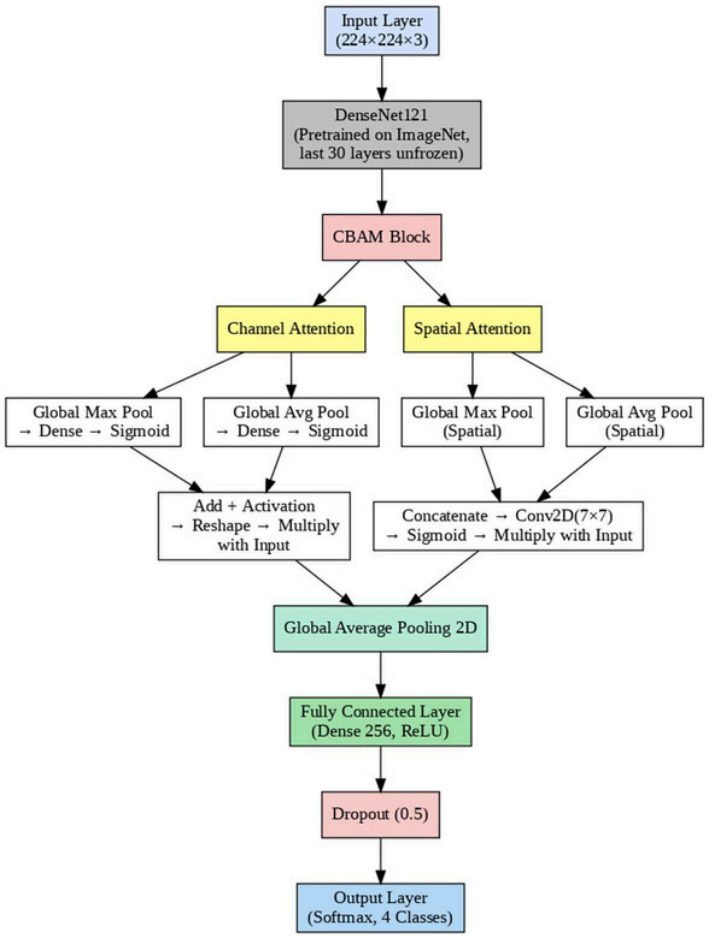
A. Dataset Preparation

The OASIS MRI dataset was used for classification, ensuring class balance to prevent model bias. The dataset has 4 classes representing four stages of one of the symptoms of Alzheimer's disease, dementia. The classes are Non-demented, Very Mild Demented, Mild, and Moderate demented. Each class was capped at 4000 samples, except "Moderate Dementia," which had 488 samples. An 80/20 train-validation split was used with stratification to preserve class distribution. Images were resized to 224×224 pixels and normalized. Data augmentation techniques such as horizontal flipping and contrast adjustment were applied to increase variability and improve generalization. The TensorFlow tf.data API was used to construct an optimized data pipeline that included shuffling, batching, and prefetching.

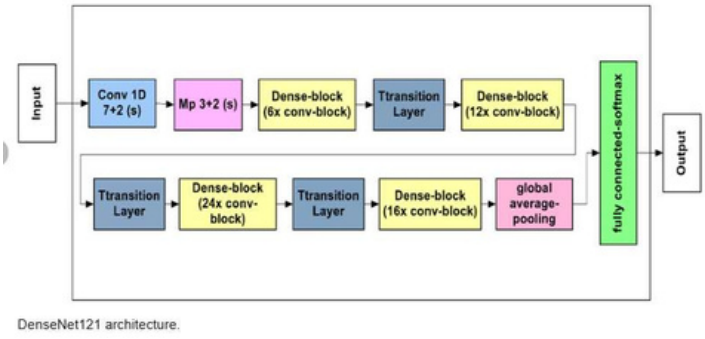
B. Model Architecture

The core model is based on a pre-trained DenseNet121 backbone, with the top layers excluded and the last 30 layers unfrozen for fine-tuning. The Convolutional Block Attention Module (CBAM) was integrated to improve feature extraction. CBAM applies both channel and spatial attention mechanisms to focus on critical regions in the input MRI images. The final architecture includes global average pooling, several dense layers, dropout regularization, and a softmax output layer for multi-class classification.

- DenseNet121 Backbone: Extracts features from MRI scans.
- CBAM Layers: multiply_1, concatenate for attention mechanisms.
- Global Pooling Layers: Includes Global Average and Max Pooling.
- Dense Layers: Dense_4 to Dense_9 for classification.
- Dropout Layer: Prevents overfitting.
- Lambda Layers: Custom operations (for attention mechanisms)

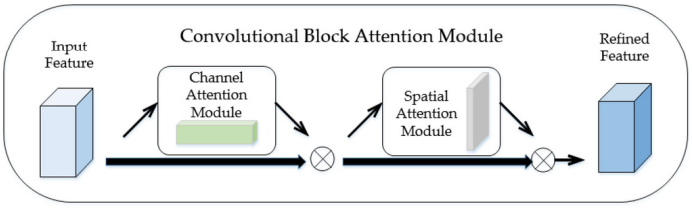


Model Architecture

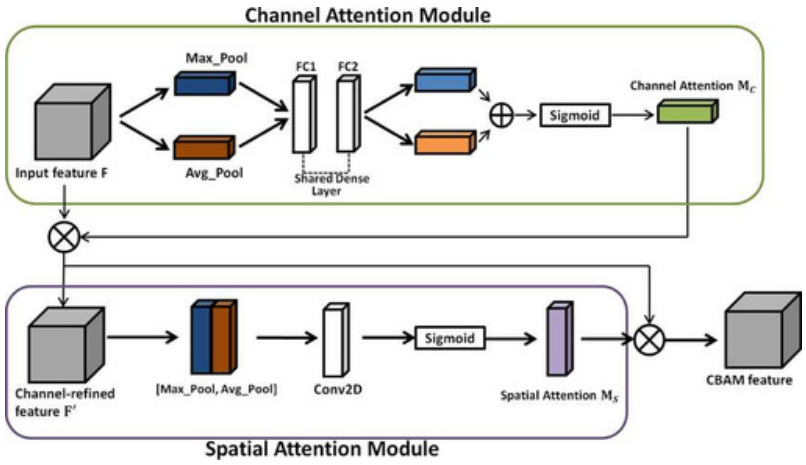


DenseNet121 architecture.

DenseNet121 Architecture



CBAM



Spatial and Channel attention module of the CBAM

C. Training Setup

The model was trained using the Adam optimizer with an initial learning rate of 1e-3. Due to integer-encoded class labels, sparse categorical cross-entropy was chosen as the loss function. To ensure efficient and robust training, callbacks such as EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint were implemented to monitor validation performance and prevent overfitting.

D. Performance and Evaluation

The model showed rapid accuracy improvement during training, achieving 97% validation accuracy by Epoch 11 and a final validation accuracy of 98.44%. The AUC-ROC score reached 0.9984, indicating excellent classification performance. The confusion matrix revealed precision and recall values above 96% for all classes, including “Mild,” “Moderate,” and “Very Mild” dementia, as well as “Non-Demented” cases.

Results

Epoch results

Epoch 1/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 122s 59ms/step - accuracy: 0.5662 - loss: 0.9767 - val_accuracy: 0.8235 - val_loss: 0.4823 - learning_rate: 0.0010 Epoch 2/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 31s 24ms/step - accuracy: 0.8239 - loss: 0.4582 - val_accuracy: 0.8995 - val_loss: 0.2671 - learning_rate: 0.0010 Epoch 3/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 22ms/step - accuracy: 0.9002 - loss: 0.2697 - val_accuracy: 0.8983 - val_loss: 0.2674 - learning_rate: 0.0010 Epoch 4/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 23ms/step - accuracy: 0.9319 - loss: 0.1930 - val_accuracy: 0.9492 - val_loss: 0.1499 - learning_rate: 0.0010 Epoch 5/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 23ms/step - accuracy: 0.9551 - loss: 0.1390 - val_accuracy: 0.9488 - val_loss: 0.1414 - learning_rate: 0.0010 Epoch 6/30 1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 23ms/step - accuracy: 0.9624 - loss: 0.1107 - val_accuracy: 0.9596 - val_loss: 0.1350 - learning_rate: 0.0010

Epoch 7/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 23ms/step - accuracy: 0.9735 - loss: 0.0799 - val_accuracy: 0.9536 - val_loss: 0.1427 - learning_rate: 0.0010
Epoch 8/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 23ms/step - accuracy: 0.9737 - loss: 0.0739 - val_accuracy: 0.9412 - val_loss: 0.1958 - learning_rate: 0.0010
Epoch 9/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 23ms/step - accuracy: 0.9850 - loss: 0.0430 - val_accuracy: 0.9632 - val_loss: 0.1235 - learning_rate: 5.0000e-04
Epoch 10/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 24ms/step - accuracy: 0.9910 - loss: 0.0272 - val_accuracy: 0.9704 - val_loss: 0.0976 - learning_rate: 5.0000e-04
Epoch 11/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 23ms/step - accuracy: 0.9934 - loss: 0.0205 - val_accuracy: 0.9724 - val_loss: 0.0991 - learning_rate: 5.0000e-04
Epoch 12/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 30s 23ms/step - accuracy: 0.9922 - loss: 0.0215 - val_accuracy: 0.9728 - val_loss: 0.1067 - learning_rate: 5.0000e-04
Epoch 13/30
...
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 23ms/step - accuracy: 0.9995 - loss: 0.0014 - val_accuracy: 0.9828 - val_loss: 0.0878 - learning_rate: 6.2500e-05
Epoch 24/30
1249/1249 ━━━━━━━━━━━━━━━━━━━━━━━━ 29s 22ms/step - accuracy: 0.9995 - loss: 0.0012 - val_accuracy: 0.9844 - val_loss: 0.0817 - learning_rate: 3.1250e-05
313/313 ━━━━━━━━━━━━━━━━━━━━━━━━ 21s 39ms/step

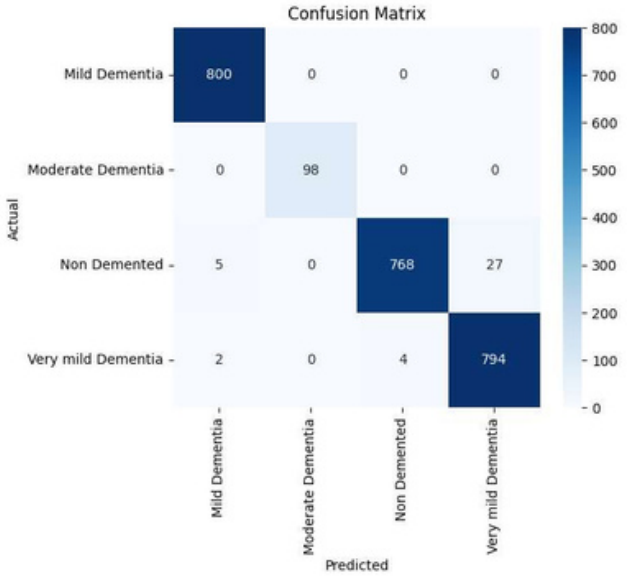
Peak Performance:

- Final Epoch (24):
  - Training Accuracy: 99.95%
  - Validation Accuracy: 98.44%
  - Validation Loss: 0.0817

	precision	recall	f1-score	support
Mild Dementia	0.99	1.00	1.00	800
Moderate Dementia	1.00	1.00	1.00	98
Non Demented	0.99	0.96	0.98	800
Very mild Dementia	0.97	0.99	0.98	800
accuracy			0.98	2498
macro avg	0.99	0.99	0.99	2498
weighted avg	0.99	0.98	0.98	2498

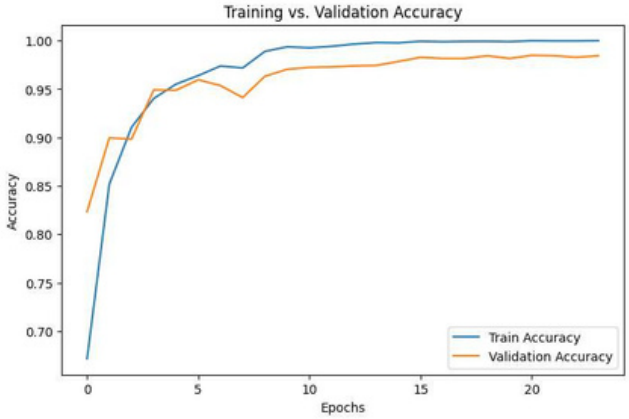
AUC-ROC Score: 0.9984374079799765

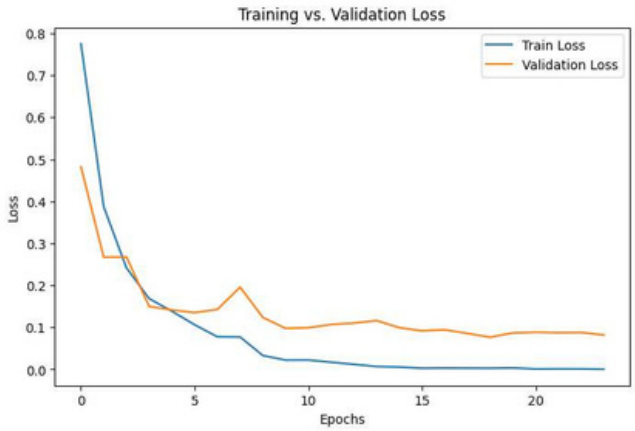
Confusion matrix

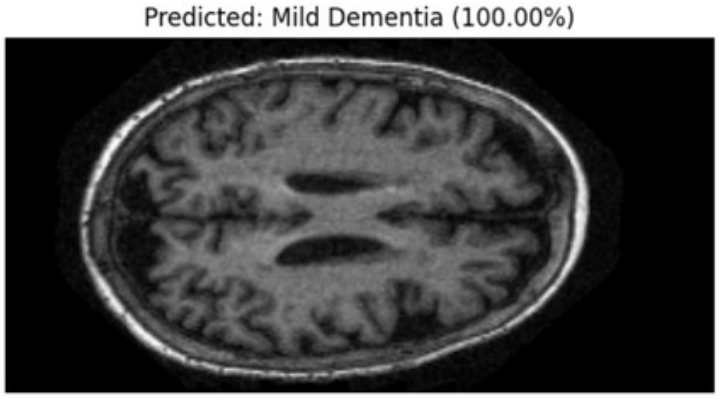


• The model showed rapid improvement in validation accuracy, reaching 97%+ by epoch 10.

• Early stopping was prevented as validation loss remained stable.







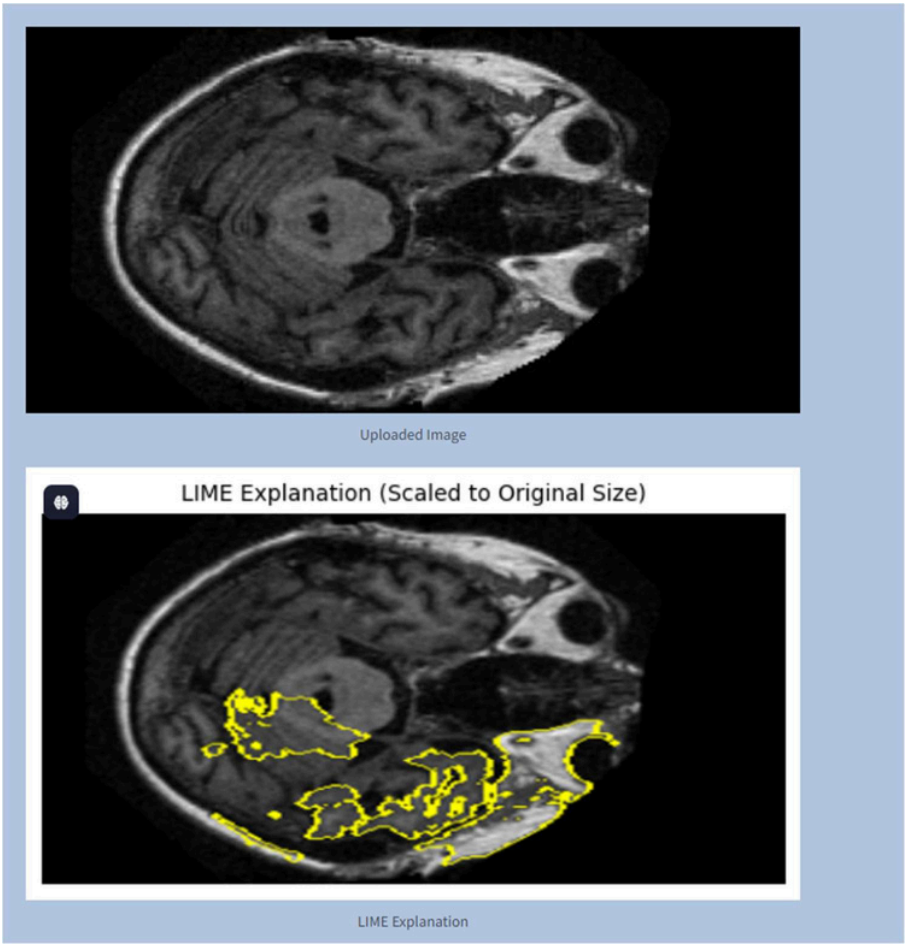
Predicted output for the mild dementia-based test image

Explainability Using LIME

Local Interpretable Model-Agnostic Explanations (LIME) was employed to improve model transparency and clinical trust. LIME helped visualize which regions of MRI scans influenced predictions the most. The top 5 most important segments per prediction were highlighted, aligning with AD's pathological features. This interpretability layer ensures clinicians can better understand and validate the model’s output.

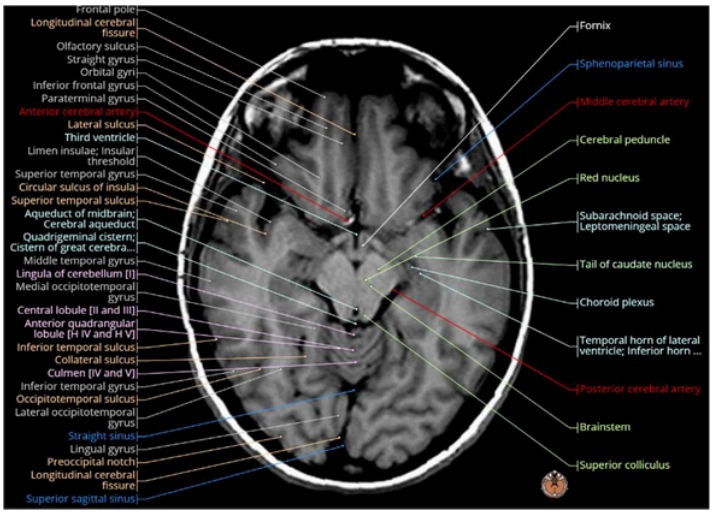
Steps taken to produce the lime-based outcome

 Image Preprocessing: The uploaded MRI image is resized to 224×224 and normalized for model input.
 Model Prediction Wrapper: A custom prediction function is defined for LIME using the trained DenseNet121 + CBAM model.
 LIME Explanation Generation: LIME perturbs the image, evaluates predictions, and identifies the most influential superpixels.
 Brain Region Segmentation: The output image from LIME is converted to grayscale, and Otsu's thresholding is applied to segment brain tissue.
 Mask Refinement: Small artifacts are removed using morphological filtering (remove_small_objects) to retain only relevant brain regions.
 Visualization: The final LIME mask is restricted to brain areas and overlaid on the image using mark_boundaries() for visual explanation



Lime-based areas were predicted by the model for very mild dementia, mainly focusing on the temporal lobe. Corresponding info taken from the info table on the same app page.

12	Cerebellum	Balance, coordination	Affected in some dementias (ataxia symptoms)	Balance tests, MRI	Physical therapy
13	Temporal Lobe	Hearing, memory, speech	Highly involved in early Alzheimer's	Cognitive tests, MRI	Memory therapy, speech therapy
14	Lateral Ventricles (in	CSF flow	Enlargement due to cortical atrophy	MRI	Monitor for NPH



App Deployment

Key Features of the Alzheimer's Detection App

1. User-Friendly Interface
   Built using Streamlit, the app offers an intuitive and interactive layout with sidebar navigation for easy access to all features.

2. MRI Scan Upload & Prediction
   Users can upload brain MRI images (JPG/PNG), and the app predicts the Alzheimer's stage (Mild, Moderate, Very Mild, Non-Demented) with confidence scores.

3. CBAM-Enhanced Deep Learning Model
   The backend model integrates DenseNet121 with a CBAM attention module to focus on critical brain regions for improved accuracy.

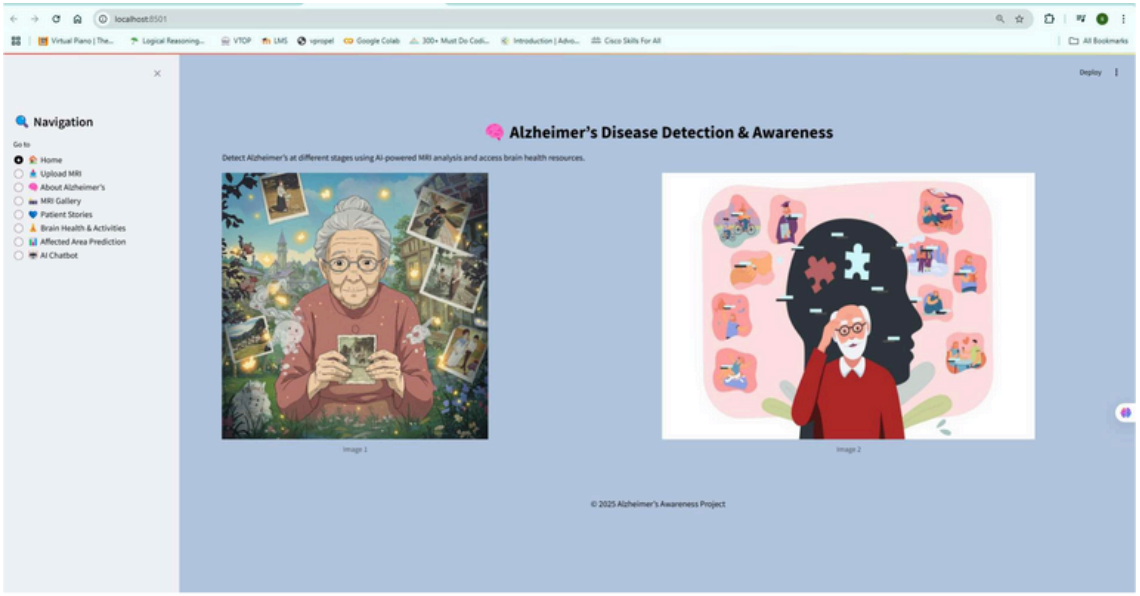
4. LIME-Based Affected Area Visualization
   Uses LIME to highlight and visually explain which brain areas influenced the model's prediction, refined by brain segmentation masks.

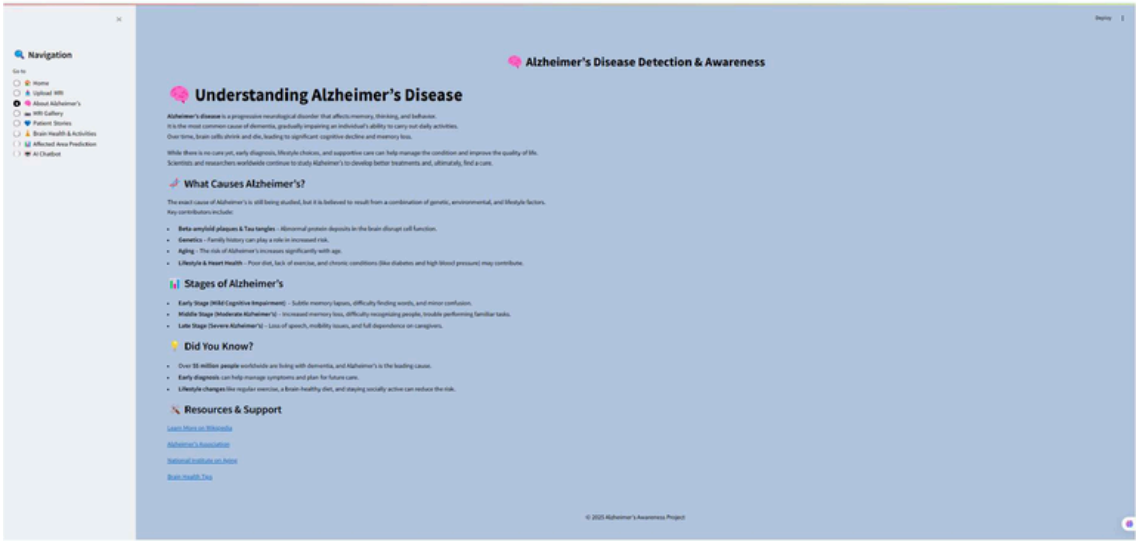
5. Labeled Brain Anatomy Display
   Side-by-side labeled anatomical brain images help users understand affected regions and compare healthy vs diseased areas.

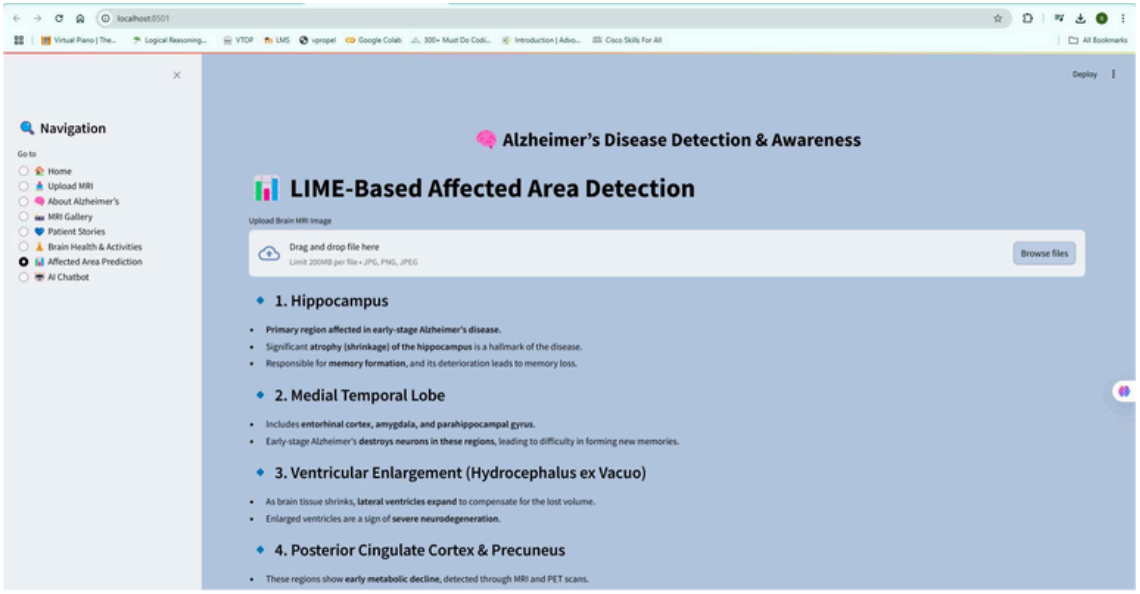
6. AI Chatbot Integration
   A built-in chatbot answers Alzheimer-related queries, with potential for further expansion using LangChain/Groq. It supports three LLMS available for community usage in the GROQ cloud - llama3-8b-8192, llama3-70b-8192, gemma-7b-it

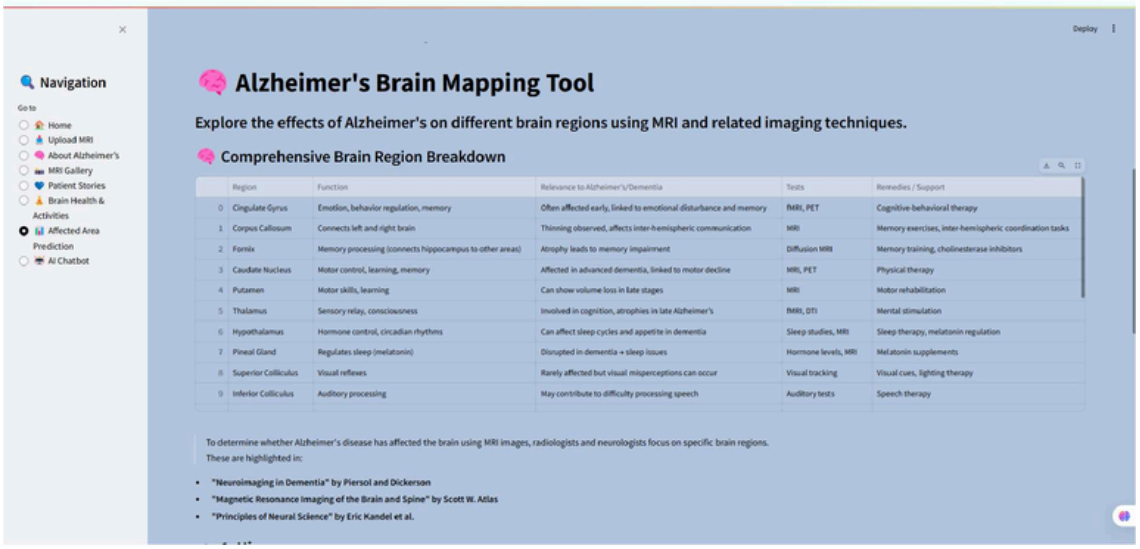
7. Educational Content & Resources
   Includes informative sections about Alzheimer's disease, brain health activities, patient stories, and links to trusted external resources.

8. Gallery and Comparison View
   MRI gallery showcases comparisons between healthy and Alzheimer-affected brains across different stages.


localhost:8501
Navigation
Go to
Home
Upload MRI
About Alzheimer's
MRI Gallery
Patient Stories
Brain Health & Activities
Affected Area Prediction
AI Chatbot
Deploy
Alzheimer's Disease Detection & Awareness
Detect Alzheimer's at different stages using AI-powered MRI analysis and access brain health resources.
Image 1
Image 2
© 2025 Alzheimer's Awareness Project


Navigation
Alzheimer's Disease Detection & Awareness
Understanding Alzheimer's Disease
What Causes Alzheimer's?
Stages of Alzheimer's
Did You Know?
Resources & Support
© 2025 Alzheimer's Awareness Project





Summary

The proposed methodology integrates a well-balanced MRI dataset with an advanced attention-enhanced convolutional neural network (DenseNet121 + CBAM) and robust interpretability using LIME. This AI-powered diagnostic framework accurately classifies brain MRI scans into four stages of Alzheimer's disease: Non-Demented, Very Mild Dementia, Mild Dementia, and Moderate Dementia. To ensure transparency, LIME-based explainability is refined through brain segmentation, highlighting only clinically significant regions and avoiding irrelevant areas. A user-friendly Streamlit web application was developed to facilitate real-time predictions, interpretability visualizations, educational content, and an AI chatbot, making the system accessible to both healthcare professionals and the public.

Future Work

- Dataset Expansion for better generalization across populations.
- Multimodal Integration with PET scans and cognitive tests.
- Clinical Deployment through mobile or real-time hospital systems.
- Model Optimization for faster inference on edge devices.
- Enhanced Explainability using SHAP or Grad-CAM++.
- Prognostic Capabilities to monitor disease progression over time.

References

[1] World Health Organization, “Dementia,” 2023. [Online]. Available: https://www.who.int/news-room/factsheets/detail/dementia.

[2] Alzheimer’s Association, “2023 Alzheimer’s Disease Facts and Figures,” 2023. [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures.

[3] M. Liu, D. Cheng, K. Wang, and Y. Wang, “Multi-modality cascaded convolutional neural networks for Alzheimer’s disease diagnosis,” Neuroinformatics, vol. 16, no. 3-4, pp. 295-308, 2018.

[4] A. Payan and G. Montana, “Predicting Alzheimer’s disease: A neuroimaging study with 3D convolutional neural networks,” arXiv preprint arXiv:1502.02506, 2015.

[5] J. Qiu, Y. Zhang, H. Zhu, and P. Zhang, “A domain adaptation approach for Alzheimer’s disease classification from MRI using deep learning,” IEEE Transactions on Medical Imaging, vol. 41, no. 2, pp. 412–423, 2022.

[6] Y. Li, S. Han, J. Chen, and M. Liu, “Handling imbalanced data in Alzheimer’s disease classification using focal loss and oversampling techniques,” IEEE Access, vol. 9, pp.110453–110465, 2021.

[7] M. Islam, S. A. Nasrin, and H. Lee, “Explainable AI for Alzheimer’s disease detection: A Grad-CAM-based approach,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2310–2319, 2023.

[8] H. Wang, X. Jin, and L. Zhao, “Attention-enhanced convolutional neural networks for Alzheimer’s disease classification,” Computers in Biology and Medicine, vol. 140, p.105098, 2022.

[9] A. Kumar, B. Singh, and R. Gupta, “Lightweight deeplearning models for Alzheimer’s detection: A mobile healthcare perspective,” IEEE Sensors Journal, vol. 22, no. 13, pp.12744–12753, 2022.

[10] R. Jack, C. Petersen, Y. Xu, and D. Knopman, “Prediction of Alzheimer’s disease using structural MRI and machine learning techniques,” IEEE Transactions on Medical Imaging, vol. 40, no. 5, pp. 1152–1164, 2021.