Explainable Alzheimer Detection Using Transfer Learning and Deep Learning

Boddu Saivardhan[1], Erothu Santhoshi[2], Chintada Sireesha[3], Bammidi Danthiswara Rao[4], A Srikar[5]
[1,2,3,4]*Department of Computer Science and Engineering Avanthi Institute of Engineering and Tech and Technology Cherukupally Village, Vizianagaram – 53116*
[5]*Assistant Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology Cherukupally Village, Vizianagaram – 531162*

***Abstract*—Alzheimer's disease is a progressive neurodegenerative disorder that causes memory loss, cognitive decline, and behavioural impairment. Early detection of Alzheimer's disease is essential for improving treatment outcomes and patient care. Traditional diagnostic approaches rely on manual analysis of brain Magnetic Resonance Imaging (MRI) scans by neurologists, which is time-consuming and prone to human interpretation errors. To address this issue, this paper proposes an automated Alzheimer's disease detection system using transfer learning and Explainable Artificial Intelligence (XAI). The proposed system utilizes the EfficientNetB3 deep learning architecture to classify MRI brain images into four stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. Transfer learning enables efficient feature extraction while improving classification accuracy even with limited medical datasets. To enhance model transparency, Grad-CAM and LIME techniques are integrated to visualize the brain regions influencing the model's prediction. The system is implemented as a web-based application using Stream lit and deployed on a cloud platform, allowing users to upload MRI images and obtain real-time predictions. Experimental results show that the proposed model achieves an accuracy of approximately 91.2%, demonstrating its effectiveness as a decision-support tool for assisting medical professionals in the early detection of Alzheimer's disease.**

***Index Terms*—Alzheimer's Disease, Transfer Learning, EfficientNetB3, Explainable Artificial Intelligence, Grad-CAM, LIME, MRI Image Classification, Deep Learning**

I. INTRODUCTION

Alzheimer's disease is one of the most common neurodegenerative disorders affecting elderly individuals worldwide. The disease gradually damages brain cells, leading to memory loss, cognitive decline, and behavioural changes. According to global health reports, millions of people suffer from Alzheimer's disease, making early diagnosis and treatment extremely important.

Traditional diagnosis of Alzheimer's disease involves clinical evaluation and manual analysis of brain MRI scans by neurologists and radiologists. Although these methods provide valuable medical insights, they are time-consuming and depend heavily on expert interpretation. Detecting Alzheimer's disease in its early stages is particularly challenging because the structural changes in the brain are subtle and difficult to observe.

Recent advancements in artificial intelligence and deep learning have enabled automated analysis of medical images. Convolutional Neural Networks (CNNs) have shown significant success in detecting patterns in complex medical imaging data. However, many deep learning models behave like black boxes, meaning that their predictions cannot easily be interpreted by humans.

To address this issue, this research proposes an automated Alzheimer detection system using transfer learning combined with Explainable Artificial Intelligence techniques. The system uses the EfficientNetB3 deep learning model to classify MRI images into four stages of Alzheimer's disease while providing visual explanations through Grad-CAM and LIME.

II. LITERATURE SURVEY

Several researchers have explored machine learning and deep learning techniques for detecting Alzheimer's disease using brain MRI images.

Traditional machine learning methods rely on manual feature extraction and classifiers such as Support Vector Machines (SVM). Although these approaches provide interpretable results, they require extensive domain knowledge and often achieve lower accuracy.

Recent studies have applied deep learning architectures such as VGG16, ResNet, and Inception networks for Alzheimer classification. These models automatically extract features from MRI images and achieve higher accuracy than traditional methods.

However, many deep learning models lack interpretability, making it difficult for medical professionals to trust their predictions. To overcome this limitation, researchers have introduced Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM and LIME to visualize the regions of the brain responsible for predictions.

This research builds upon these advancements by integrating transfer learning with EfficientNetB3 and incorporating XAI techniques to create an accurate and interpretable Alzheimer detection system.

III. PROPOSED SYSTEM

A. System Overview

The proposed system is an AI-based medical image analysis platform designed to detect Alzheimer's disease from MRI brain scans. The system follows a cloud-based architecture and provides an easy-to-use web interface for uploading MRI images and obtaining predictions.

B. Methodology

1. The system integrates the following components:
2. MRI Image Preprocessing
3. Transfer Learning using EfficientNetB3
4. Alzheimer Stage Classification
5. Explainable AI using Grad-CAM
6. Explainable AI using LIME
7. MRI images are first resized and normalized before being passed to the trained deep learning model. The model then predicts the Alzheimer stage and generates visual explanations.

C. System Workflow

1. User uploads MRI brain image
2. Image preprocessing is performed
3. EfficientNetB3 model extracts features
4. Model predicts Alzheimer stage
5. Grad-CAM heatmap generated
6. LIME explanation generated
7. Prediction result displayed to the user.

D. Advantages

- Automated detection of Alzheimer's disease
- Faster diagnosis compared to manual analysis
- Improved accuracy using transfer learning
- Explainable AI for medical interpretation
- Cloud-based real-time access

IV. IMPLEMENTATION

A. Backend

• Python
• TensorFlow / Keras
• OpenCV and NumPy

B. Frontend

• Streamlit Web Framework

C. Development Tools

• Visual Studio Code
• Google Colab

D. Modules

1. MRI Image Upload Module
2. Image Preprocessing Module
3. Deep Learning Prediction Module
4. Grad-CAM Visualization Module
5. LIME Explanation Module

V. RESULTS AND DISCUSSION

- The proposed Alzheimer detection system was tested using MRI brain images from a publicly available dataset. The EfficientNetB3 transfer learning model achieved an overall classification accuracy of approximately 91.2%.
- A. Performance Analysis
- Key observations include:
- High classification accuracy across four Alzheimer stages
- Faster prediction time using transfer learning

- Reduced computational cost compared to training models from scratch

- B. Explainability Analysis
- Grad-CAM heatmaps successfully highlighted the brain regions responsible for the model's prediction. LIME explanations further helped identify the important image segments influencing the classification result.

- C. Comparison with Traditional Models

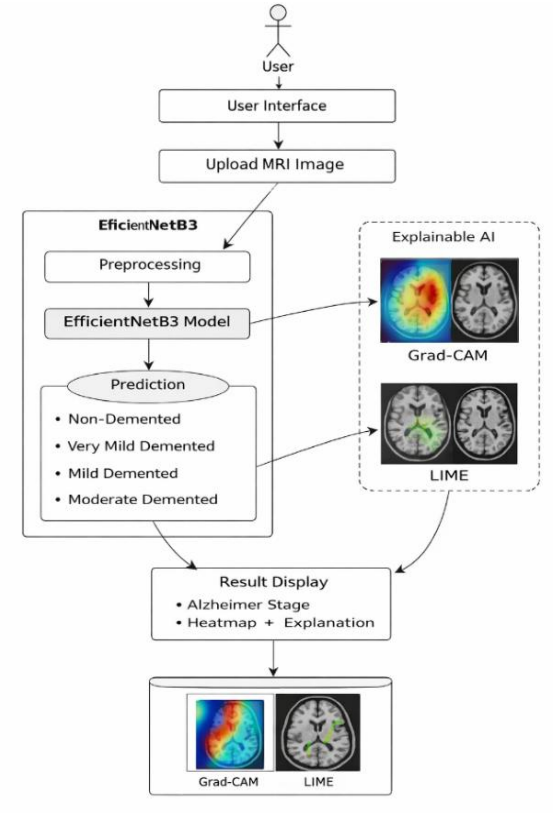
Model	Accuracy
SVM	82.5%
VGG16	88.7%
ResNet50	90.1%
EfficientNetB3 (Proposed)	91.2%

- These results demonstrate that the proposed model achieves improved performance compared to traditional machine learning and earlier deep learning approaches.

VI. SYSTEM ARCHITECTURE DIAGRAM

Alzheimer detection system flowchart

Alzheimer detection system flowchart



- Diagram Explanation
- User uploads MRI image through web interface
- Backend processes image using EfficientNetB3 model
- Grad-CAM and LIME generate visual explanations
- Prediction result displayed to user

VII. CONCLUSION

The proposed system demonstrates the effectiveness of transfer learning and explainable artificial intelligence in detecting Alzheimer's disease from MRI brain images. By using the EfficientNetB3 deep learning architecture, the system achieves high classification accuracy while maintaining computational efficiency.

The integration of Grad-CAM and LIME provides transparency and interpretability, enabling medical professionals to understand the reasoning behind model predictions. The web-based implementation allows easy access and real-time diagnosis support.
Overall, the proposed system provides an accurate, interpretable, and scalable solution for assisting doctors in the early detection of Alzheimer's disease.

VIII. FUTURE SCOPE

The system can be further improved by
• Training the model using larger MRI datasets
• Integrating Vision Transformer models
• Supporting additional medical imaging techniques such as PET scans
• Developing mobile applications for remote diagnosis
• Integrating with hospital information systems

REFERENCES

[1] . I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[2] J. Islam and Y. Zhang, "Brain MRI Analysis for Alzheimer's Disease Diagnosis Using Deep Convolutional Neural Networks," Brain Informatics, 2018.
[3] R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks," IEEE ICCV, 2017.
[4] M. Tan and Q. Le, "Efficient Net: Rethinking Model Scaling for CNNs," ICML, 2019.