

Early Detection of Alzheimer's Disease Using Deep Learning

Dr.R. Ahila¹, Ms. Dhanyashri K², Ms. Saru Varsini V M³, Ms. Sivaranjani M⁴, Ms. Tharsini K⁵

¹Associate Professor,

^{2, 3, 4, 5}Student, Department of Computer Science & Engineering,

Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu

Abstract

Alzheimer's disease is one of the most common neurodegenerative illnesses, affecting the patient's memory, cognitive, and decision-making abilities. It is essential to diagnose the condition at an early stage for effective intervention to prevent further deterioration. Most traditional methods of diagnosing the disease depend on analyzing pictures obtained with an MRI scanner. This is a technique that needs much energy and, at the same time, cannot be used to detect the small variations taking place during the various stages of AD. In the proposed model, an attempt was made to design a detection algorithm that will help in detecting the early stage of Alzheimer's disease through MRI scans. ViT architecture is used in the proposed model, which facilitates the extraction of local as well as global features in MRI images through the self-attention process. Various image processing techniques, such as resizing, normalization, and augmentation of images, have been applied to the model. The proposed algorithm can classify images according to the AD stage. A web application has been designed to help users upload the MRI images to get predictions through an easy-to-use interface. With the suggested approach, not only has the manual work been minimized, but also the time needed to diagnose is reduced, which helps doctors make decisions. The results show that the phases of AD in MRI scans may be categorized using the ViT model.

Key Words : Alzheimer's Disease, Vision Transformer (ViT), Magnetic Resonance Imaging (MRI), Deep Learning, Early Disease Detection, Medical Image Classification, Flask, Artificial Intelligence.

1. Introduction

Alzheimer's disease AD refers to a neurological disorder affecting countless people around the world. Due to impaired mental functions, the illness subtly impairs a person's capacity to recall experiences and engage in cognitive activities. As the global population grows older, the prevalence rate of Alzheimer's disease keeps rising, thereby posing serious challenges to the medical sector.

Early diagnosis is crucial because it enables the provision of prompt medical attention to the patients. MRI image analysis by healthcare providers is one of the conventional techniques used when diagnosing AD. Though efficient, this technique may take much time, not forgetting that the smallest alterations in the brain structure might go unnoticed when the disease is in its early stages. Automated methods for identifying neurological conditions have been made possible by deep learning.

The Vision Transformer model, designed specifically for image analysis, was utilized to develop an intelligent system for automatically diagnosing symptoms of AD. The foundation for developing such a system is training a deep neural network using MRI pictures of people who have received an Alzheimer's diagnosis. The system can thus

recognize characteristics linked to different phases of the illness.

It is aimed at improving the diagnosis process as well as minimizing the analysis period. Moreover, the suggested solution may serve as an additional instrument that will assist doctors in making decisions. The effectiveness of the algorithm was evaluated according to its performance characteristics.

2. Literature Survey

Azhar Hatem Jebur Albaidhani et al. proposed an early Alzheimer's disease detection framework using Vision Transformers on Continuous Wavelet Transform (CWT)-transformed EEG signals. The model employed self-attention mechanisms to learn complex neurological patterns and achieved promising classification performance. However, the study utilized a limited number of EEG channels and a single dataset, which may affect its generalization capability across different clinical environments [1].

Mihan Muhammad Sadiq Fareed et al. developed ADD-Net, a deep learning model for Alzheimer's disease detection using MRI brain images. The framework automatically extracted disease-related features and achieved high classification accuracy. However, class imbalance in the

dataset could reduce prediction reliability and model robustness when applied to unseen data [2].

Ghadah Naif Alwakid et al. introduced an Alzheimer's disease detection system using MobileNetV2 and DenseNet121 combined with image enhancement techniques such as CLAHE and ESRGAN. The enhanced MRI images improved diagnostic accuracy significantly. However, MobileNetV2 required careful parameter tuning, while DenseNet121 was computationally expensive and susceptible to feature loss due to its bottleneck architecture [3].

Harsh Vardhan Bansal et al. proposed a multimodal framework that integrated ResNet-101, Firefly-based feature selection, and Support Vector Machine classification for Alzheimer's disease diagnosis. The approach improved feature selection and classification performance. However, the feature extraction process could lead to the loss of important information, affecting overall prediction accuracy [4].

Yashar Sarbaz et al. developed a deep learning framework for predicting the progression of Mild Cognitive Impairment using MRI scans. The model analyzed structural brain changes and demonstrated the potential of deep learning for early disease prediction. Nevertheless, it required significant computational resources and faced challenges in handling complex multi-class classification tasks [5].

Dosovitskiy et al. introduced the Vision Transformer (ViT), which processes images as sequences of patches using self-attention mechanisms. The model effectively captured global contextual information and achieved state-of-the-art image classification performance. However, Vision Transformers require large datasets and high computational resources for effective training [6].

Ayoub Benali Amjoud et al. presented a review of Deep Learning, CNN, and Vision Transformer-based approaches for image analysis. The study highlighted the ability of Vision Transformers to capture global contextual relationships through self-attention mechanisms. However, their implementation often requires large-scale datasets and substantial computational power [7].

Jose Mauricio et al. conducted a comparative study of Vision Transformers and CNNs for image classification. The results showed that Vision Transformers often outperform CNNs by learning long-range dependencies and global image features more effectively. However, they require greater computational resources and larger training datasets than traditional CNN models [8].

Aakash Varma et al. reviewed recent advancements in Vision Transformer architectures for image classification. The study emphasized the superior feature representation and scalability of transformer-

based models. Nevertheless, increased computational complexity and longer training times remain significant challenges for practical deployment [9].

From the reviewed literature, it is evident that transformer-based architectures provide better global feature extraction and classification performance than conventional CNN models. Their ability to capture long-range dependencies through self-attention mechanisms makes them highly suitable for medical image analysis. Therefore, the proposed work utilizes a Vision Transformer-based framework for the early detection of Alzheimer's disease.

Table 1. Comparison of Existing and Proposed System

Features	Existing System	Proposed System
Disability Coverage	Moderate	Fine-Grained
Features Extraction	Local	Global
Web Platform	No	Yes
Confidence Score	Rarely Provided	Provided with Every Prediction
Report Generation	Manual	Automated
Accuracy	affected by Class Imbalance	98% across all 4 classes
Class Imbalance	Limited	Handled
Training Speed	Slower	Faster

3. Proposed Methodology

The proposed system is designed to analyse Alzheimer's disease at an early stage through analysis of the brain MRI images. It uses a Vision Transformer (ViT) model to evaluate MRI scans and categorize them into various stage of Alzheimer's : Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented.

The MRI images employed are taken from the Kaggle Alzheimer's MRI dataset. First these images are pre-processed by data augmentation like rotation and flipping and normalization is applied and resized before training. The preprocessing procedures are required for the improvement of the image quality to train the model.

We extract salient features of MRI scans using a pre-trained vision transformer model. The program searches several brain regions for patterns that might indicate Alzheimer's.

The system predicts the stage of AD based on the identified patterns. The model has been integrated into a web application created using Flask. Users can upload MRI brain scans and get predictions on the stage of AD using this app.

The suggested method provides an automated, dependable, and user-friendly

method to identification of disease. It helps doctors in early detection, the system can promote timely treatment and improve patient care.

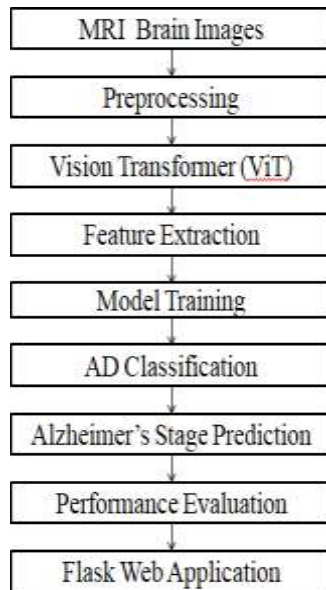


Fig. 1. Workflow Diagram

4. System Implementation

The proposed system consists of five main modules Data Collection and Preprocessing, Model Training, Classification, Evaluation Metrics, and Web Application. The modules are designed to collaborate to identify and classify various phases of Alzheimer's disease in MRI brain scans.

4.1 Data collection and preprocessing

The data in this research was from the Alzheimer's MRI Dataset available on Kaggle. The overall number of images in the dataset are 44,000 images, divided into four classes: Non-Demented (12,000 images), Very Mild Demented (11,200 images), Mild Demented (10,000 images), and Moderate Demented (10,000 images). These images are labeled with a class label, representing the stage of the Alzheimer's disease (AD) spectrum. The dataset was split into training, validation, and testing sets with a ratio of 70%, 15%, and 15%, respectively, to make sure that the datasets were used appropriately for model training and model evaluation. This division enables the model to learn on a part of the data, validate and test on unseen images. All the MRI images were preprocessed before training. The images were resized to 224×224 pixels to conform to the input size of the Vision Transformer model. To ensure consistency and stability in training, pixel values were normalized. To enhance the dataset diversity and the model's generalization, data augmentation techniques were also performed. These methods were horizontal flipping, random rotation, random zooming and contrast adjustment. The original MRI images were then preprocessed into a standard format with consistent size, normalized pixel values, and multiple variations, to aid the model in learning more effectively.

4.2 Model training

The ViT model was trained using MRI images that had been preprocessed. Unlike CNNs, ViT breaks down an image into segments and employs self attention to process them. This method helps the model to grasp not only local image details but also Global Image Features. The training utilized the pre-trained Google ViT-Base-Patch16-224 model. Before processing each MRI image with transformer encoder, patches measuring 16 by 16 pixels were generated. By focusing on self-attention, the model effectively identified key structural abnormalities and patterns in the brain linked to Alzheimer's. The training was conducted with a batch size of 16 over five epochs. The Adam optimizer was employed to efficiently update model parameters, and using the Cross-Entropy Loss function prediction errors were calculated during training. Early stopping was implemented to eliminate overfitting and enhance generalization. Furthermore, whenever the validation loss decreased, the model with the high performance was automatically saved.

4.3 Classification

The training method entailed the model learning key features from MRI pictures, which increased its capacity to categorize the four phases of Alzheimer's disease. This trained model was utilized for categorization purposes. During this phase,

MRI brain images from the test dataset were input into the trained model. Consistency in pre-processing was ensured by applying the same procedure. Vision Transformer was employed to pinpoint important features in the brain's MRI and to produce probability scores for each disease category. From the highest-probability class, the system has been able to classify the MRI image into one of four classes: Non-Demented, Very Mild Demented, Mild Demented, or Moderate Demented. The trained model successfully classified all the stages with an overall accuracy of about 98%. The high accuracy suggest that the model captured crucial patterns in MRI brain images and can make accurate predictions.

4.4 Web application

As the system to be accessible and user friendly, the trained Vision Transformer model was embedded in a Flask based Web App. The application has a user-friendly interface that allows users to upload MRI brain images for analysis. After uploading an image, the system automatically preprocesses the image and feeds it into the trained model for prediction. Then the app shows the stage of Alzheimer's disease and a confidence rating about the accuracy of the prediction. Besides disease prediction, the app offers graphical output like graphs and AI-generated report summarizing the output. These features assist users in understanding the prediction better and also aid medical professionals make appropriate decisions. The web application allows to avoid complex

technical knowledge, which makes the system easier to use in real life scenarios. The application offers fast and automatic analysis of MRI scans which helps to detect and diagnose Alzheimer's disease early.

5. Role of Vision Transformer in Project

A vision transformer is highly important in the system aimed at diagnosing early symptoms of Alzheimer's disease. Its main task is to analyze MRI scans of the patient's brain in order to detect the patterns emerging in different phases of the emergence of this disease. Different from other image analysis models, ViT uses an approach known as self-attention that helps it create relations between diverse parts of the brain. The project's Vision Transformer would be able to analyze the brain MRIs and classify patients as Very Mild Demented, Mild, or Non-Demented. The ability to focus on vital parts of the brain image enables it to spot minor changes in it. The capability of ViT to utilize the information gathered from the vast amount of data stored in image databases enables this approach to become more efficient through the transfer learning process. The increased level of efficiency is the result of accuracy and reduced training time.

6. Result and Discussion

6.1 Performance Analysis

The proposed framework's accuracy and practical effectiveness were evaluated based on its ability to classify brain MRI scans into four stages of Alzheimer's disease: Non-Demented, Very Mildly Demented, Mildly Demented, and Moderately Demented. The model utilized the Kaggle Alzheimer's MRI dataset for training and evaluation, with data split into training, validation, and test sets through stratified sampling to maintain class balance across all subsets.

The model underwent training over several epochs, with continuous monitoring of training and validation accuracy. Both accuracy curves exhibited a steady upward trend, and the early stopping mechanism guaranteed that the optimal weights were saved once the validation loss stopped improving. The confusion matrix on the test set confirmed reliable performance across all four categories, with NonDemented and ModerateDemented exhibiting the highest classification accuracy. Some minor misclassifications were noted between the VeryMildDemented and MildDemented stages, which is expected clinically due to the subtle structural similarities between these two intermediate classifications.

One of the main strengths of this work is the use of the Vision Transformer

architecture. Unlike traditional convolutional neural networks, the Vision Transformer can capture both local and global information from MRI images through its self-attention mechanism. This allows the model to identify subtle structural changes in different regions of the brain, which are important for detecting the progression of Alzheimer's disease. The integration of the trained model into a Flask-based web application enhances its practical usefulness. The application allows users to upload MRI brain images and obtain predictions in real time through a simple interface. This makes the system easy to use and provides quick support for disease classification.

The proposed system contributes to the field of medical image analysis by providing an automated and reliable approach for Alzheimer's disease detection. By supporting early diagnosis and reducing manual effort, the system can assist healthcare professionals in making faster and more informed clinical decisions. Overall, the obtained results demonstrate that the proposed approach is an effective tool for the early detection and classification of Alzheimer's disease using MRI brain images.

6.2 Output Visualization

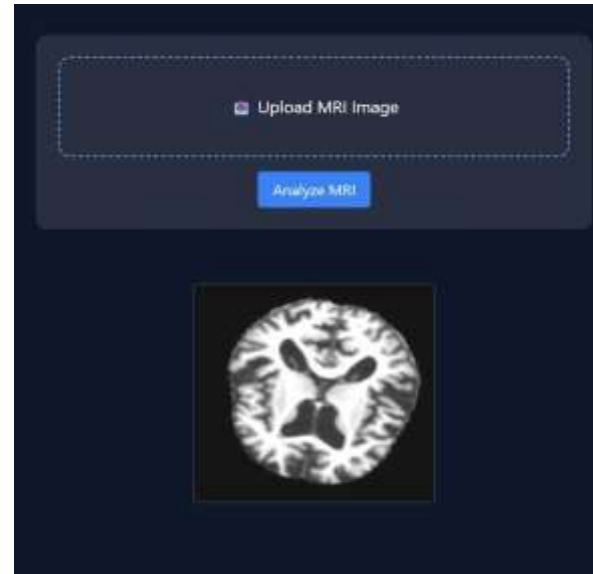


Fig. 2. Uploaded MRI image

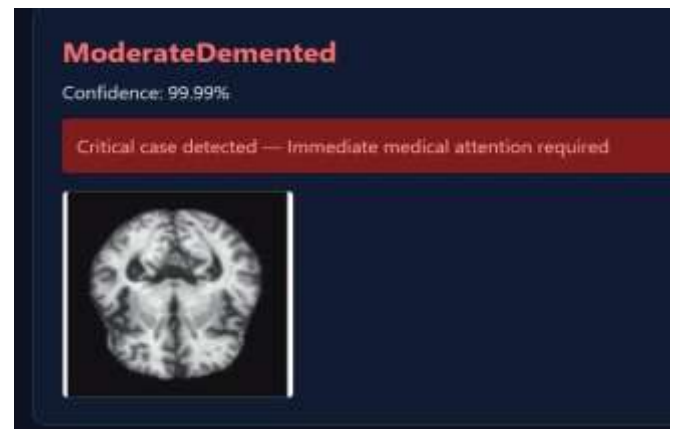


Fig 3. Confidence score

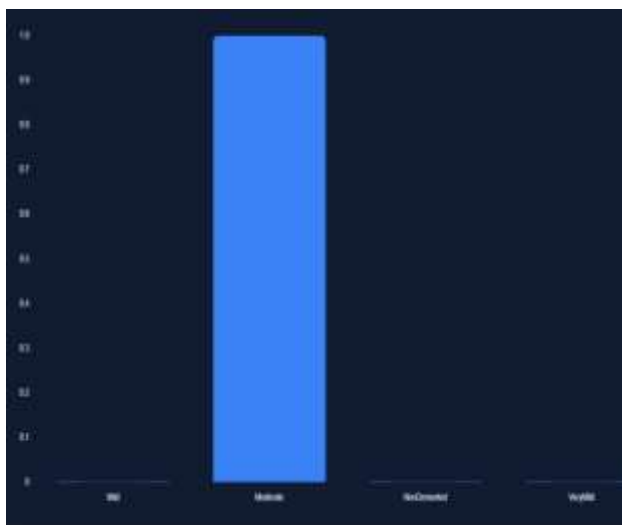


Fig.4. Confidence Score graph



Fig 5. Prediction Report

The Flask web application successfully generated predictions for uploaded MRI images. The system provides the predicted disease stage along with confidence scores, making it easy for users to interpret the results.

7. Conclusion

Given that Alzheimer's is among the most prevalent neurodegenerative illnesses, early diagnosis is crucial for efficient treatment planning and patient care. The goal was to create an automated system that uses deep learning techniques which uses MRI brain images to detect Alzheimer's illness brain scans. The suggested method uses a Vision Transformer-based model for more accurately diagnose various illness stages.

The choice of ViT as the core architecture draws from Dosovitskiy et al. [6], who showed that transformers applied directly to image patches can match or outperform CNNs when pretrained on sufficient data. This is especially valuable in medical imaging, where understanding global brain structure matters more than local texture patterns. As Mauricio et al. [8] further support, ViT's multi-head self-attention mechanism contributes to stronger classification accuracy, making it well-suited for the spatial complexity of MRI-based Alzheimer's staging.

The proposed system incorporates typical preprocessing steps, enhances data through augmentation, and involves fine-tuning of a pretrained ViT model, and an early stopping mechanism to retain the best-performing weights. A Flask-based web application was also built on top of the trained model, allowing MRI images to be

uploaded and classified with confidence scores — making it practically usable in a clinical environment.

Despite promising results, the system has a few notable limitations. First, the model was trained and evaluated on a Kaggle-sourced dataset, which may not fully represent the variability seen in real-world clinical MRI data. Differences in scanner hardware, acquisition protocols, and patient demographics could influence overall generalization effectiveness. Second, the current system processes only 2D MRI slices and does not exploit the volumetric information present in full 3D brain scans, which could carry additional diagnostic value. Third, ViT models generally demand more data to train effectively compared to CNNs, as noted by Dosovitskiy et al. [6], and performance on smaller datasets may be more dependent on the quality of pretraining. Finally, the system currently lacks explainability features — clinicians are not shown which regions of the MRI influenced the prediction, which is an important requirement for real-world medical deployment.

8. Future Scope

While the proposed system shows promising results, several directions remain open for further improvement. Integrating Explainable AI techniques would generate visual attention maps highlighting the brain

regions influencing predictions, improving clinical transparency and helping physicians understand the model's decisions — a critical requirement for real-world medical deployment. Extending the framework to process full 3D MRI volumes rather than 2D slices could also capture richer volumetric and structural changes associated with disease progression, potentially improving classification accuracy across all four stages.

In addition, the current model was trained on a single publicly available dataset, which may limit its generalization across different clinical environments. Future work could focus on validating the system using MRI data collected from diverse hospitals, scanners, and patient populations to assess real-world robustness. Exploring multimodal inputs such as combining MRI with cognitive assessment scores or genetic markers could further enhance diagnostic accuracy, and deploying the system on cloud-based or mobile platforms could extend its accessibility to remote and resource-limited healthcare settings.

9. Reference

- [1] A. H. J. Albaidhani, G. Serbes and H. O. Ilhan, "Early Detection of Alzheimer's Disease Using Vision Transformers on CWT-Transformed EEG Signals in Response to Olfactory Stimuli," in *IEEE Access*, vol. 14, pp. 6968-6985, 2026, doi: 10.1109/ACCESS.2026.3652249.

- [2] M. M. S. Fareed et al., "ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI Scans," in *IEEE Access*, vol. 10, pp. 96930-96951, 2022, doi: 10.1109/ACCESS.2022.3204395.
- [3] G. N. Alwakid, S. Tahir, M. Humayun and W. Gouda, "Improving Alzheimer's Detection With Deep Learning and Image Processing Techniques," in *IEEE Access*, vol. 12, pp. 153445-153456, 2024, doi: 10.1109/ACCESS.2024.3481238.
- [4] H. V. Bansal, P. Gupta and V. Juneja, "A Multimodal Deep Learning Framework Using ResNet-101 and Firefly-Based Feature Selection for Early Diagnosis of Dementia and Alzheimer's Disease," in *IEEE Access*, vol. 13, pp. 184709-184721, 2025, doi: 10.1109/ACCESS.2025.3621157.
- [5] A. Baktash, Y. Sarbaz, S. Meshgini and R. Afrouzian, "Advancing Early Diagnosis: Predicting Mild Cognitive Impairment Progression in Normal Individuals Using Deep Learning on MRI Features," in *IEEE Access*, vol. 13, pp. 122591-122602, 2025, doi: 10.1109/ACCESS.2025.3588801.
- [6] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations (ICLR)*, 2021, doi: 10.48550/arXiv.2010.11929.
- [7] A. Baktash, Y. Sarbaz, S. Meshgini and R. Afrouzian, "Advancing Early Diagnosis: Predicting Mild Cognitive Impairment Progression in Normal Individuals Using Deep Learning on MRI Features," in *IEEE Access*, vol. 13, pp. 122591-122602, 2025, doi: 10.1109/ACCESS.2025.3588801.
- [8] J. Maurício, I. Domingues, and J. Bernardino, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review," *Applied Sciences*, vol. 13, no. 9, p. 5521, 2023, doi: 10.3390/app13095521.
- [9] A. Varma et al., "Vision Transformers in Image Classification: A Comprehensive Review of Architectures and Applications," *arXiv preprint arXiv:2302.09462*, 2023, doi: 10.48550/arXiv.2302.09462.
- [10] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
- [11] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 574–584, 2022.
- [12] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in Medical Imaging: A Survey," *Medical Image Analysis*, vol. 88, p. 102802, Aug. 2023.
- [13] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proceedings of the IEEE International*

Conference on Computer Vision (ICCV), pp. 618–626, 2017.

[14] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations (ICLR), 2021.

[15] J. Mauricio, I. Domingues, and J. Bernardino, "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review," Applied Sciences, vol. 13, no. 9, p. 5521, 2023.