© 2026 IJRTI | Volume 11, Issue 4 April 2026 | ISSN: 2456-3315

A Lightweight Advanced Hybrid Transformer Model for Early Detection of Alzheimer's Disease Using Neurological Medical Images

[1]Saikat Mondal, [2]Venkatesh G
[1]M. Tech Student, [2]Assistant Professor
[1]Department of Artificial Intelligence and Data Science, [2]Department of Artificial Intelligence and Data Science
[1]MVJ College of Engineering, Bangalore, India, [2]MVJ College of Engineering, Bangalore, India
[1]saikat030703@gmail.com, [2]venkateshg.559@gmail.com

***Abstract*—Alzheimer's disease is a progressive neurodegenerative disorder that leads to cognitive decline and memory impairment. Early detection of Alzheimer's disease is crucial for enabling timely clinical intervention and improving patient outcomes. Conventional diagnostic methods rely heavily on manual analysis of brain magnetic resonance imaging scans, which can be time-consuming and prone to subjective interpretation errors. This study presents a deep learning-based framework for automated classification of Alzheimer's disease stages using brain magnetic resonance imaging images. The proposed approach utilizes an efficient MobileViT-based architecture that combines convolutional operations with lightweight transformer modules to capture both local spatial features and global contextual information. Convolutional layers extract fine-grained visual patterns, while transformer components model long-range dependencies across the image. Data augmentation techniques such as random rotation and horizontal flipping are applied to enhance model generalization. The model is trained using transfer learning with selective layer unfreezing and progressive fine-tuning, and optimized using the AdamW optimizer with a cosine annealing learning rate scheduler. Experimental results demonstrate that the proposed model effectively learns discriminative representations from brain magnetic resonance imaging scans and achieves strong classification performance across Alzheimer's disease categories. The framework highlights the potential of MobileViT-based architectures for medical image analysis and supports early diagnosis of neurodegenerative diseases.**

***Index Terms*— Alzheimer disease, deep learning, convolutional neural networks, transformers, medical image classification, brain MRI, MobileViT**

I. Introduction

Alzheimer's disease is one of the most common neurodegenerative disorders affecting millions of individuals worldwide. The condition progressively damages brain cells, leading to memory loss, cognitive decline, and behavioral changes. According to recent studies, early detection of Alzheimer's disease is crucial for slowing disease progression and improving the quality of life for affected individuals. Magnetic resonance imaging has emerged as a widely used imaging modality for analyzing structural brain changes associated with Alzheimer's disease. However, manual interpretation of magnetic resonance imaging scans by medical professionals is a complex and time-consuming process. With the rapid growth of medical imaging data, the need for automated diagnostic systems has become increasingly important. In recent years, deep learning approaches have demonstrated significant success in medical image analysis. Convolutional neural networks have been extensively used for image classification tasks due to their ability to learn hierarchical spatial features. Despite their effectiveness, traditional convolutional models mainly focus on local feature extraction and may not fully capture long-range dependencies present in complex medical images. To overcome these limitations, lightweight vision transformer-based architectures have been introduced, combining convolutional operations with transformer mechanisms. MobileViT is one such architecture that integrates convolutional layers with efficient transformer blocks to capture both local spatial patterns and global contextual relationships while maintaining computational efficiency. This study proposes a MobileViT-based framework for automated classification of Alzheimer's disease using brain magnetic resonance imaging images. The objective of this research is to develop an efficient deep learning model capable of accurately identifying different stages of Alzheimer's disease while ensuring strong generalization across diverse imaging data.

II. Literature Review

Wherever Several research efforts have focused on the application of deep learning techniques for the automated detection of Alzheimer's disease using medical imaging data. These studies demonstrate that machine learning models can significantly assist clinicians by providing objective and fast diagnostic support.

Jabason et al. [1] proposed a semi supervised classification framework based on stacked sparse autoencoders for Alzheimer's disease detection. Their model was designed to address the problem of incomplete clinical and imaging data by estimating missing information during training. This approach improves model robustness and allows the system to function effectively in real clinical scenarios where incomplete data is common.

Liu et al. [2] introduced a deep learning-based system for early detection of Alzheimer's disease using structural magnetic resonance imaging scans. Their model demonstrated strong generalization ability across different datasets, which is an important requirement for practical medical applications. The study highlighted the importance of building models that can adapt to variations in imaging protocols and patient populations.

Khvostikov et al. [3] developed a three dimensional convolutional neural network architecture based on the Inception model to analyze structural MRI and diffusion tensor imaging data simultaneously. By combining multiple imaging modalities, the model was able to capture more detailed structural changes associated with Alzheimer's disease progression.

Song et al. [8] further explored multimodal data fusion by integrating magnetic resonance imaging and positron emission tomography scans. Their approach utilized a feature fusion strategy to combine complementary information from both imaging modalities, which resulted in improved diagnostic performance compared to single modality approaches.

Several studies have also explored the use of convolutional neural network architectures for Alzheimer's disease classification. Zaabi et al. [4] implemented transfer learning based convolutional neural networks for brain MRI classification. Their findings showed that pretrained deep learning models can significantly improve classification accuracy when applied to medical imaging datasets.

Janghel and Rathore [5] proposed an early detection system based on deep convolutional neural networks that learns hierarchical feature representations from MRI scans. Their work emphasized that deep architectures can effectively capture subtle anatomical patterns associated with neurodegenerative diseases.

Gunawardena et al. [10] applied convolutional neural networks to detect early signs of Alzheimer's disease from structural MRI images. Their research demonstrated the capability of deep learning models to identify complex brain abnormalities that may not be easily visible through manual analysis.

Liu et al. [6] introduced a hybrid deep learning framework combining convolutional neural networks with recurrent neural networks to analyze fluorodeoxyglucose positron emission tomography images. This architecture was able to capture both spatial and sequential information within the imaging data, resulting in improved classification accuracy.

Prakash et al. [9] conducted a comparative study of various transfer learning models for Alzheimer's disease classification. Their results indicated that fine tuned deep neural networks outperform traditional machine learning algorithms in terms of accuracy and robustness.

Nawaz et al. [7] designed a deep convolutional neural network model specifically for MRI based Alzheimer's disease classification. Their work showed that deep learning based approaches provide superior performance compared to conventional image processing methods.

Although convolutional neural networks have achieved promising results, they mainly focus on local spatial features. Recent research has therefore explored hybrid architectures that combine convolution operations with attention mechanisms to capture both local and global image patterns. This study builds upon these advancements by implementing a hybrid convolution transformer architecture based on MobileViT for improved Alzheimer's disease classification.

III. METHODOLOGY

The dataset used for training and evaluation of the proposed deep learning model consists of brain magnetic resonance imaging images representing different stages of Alzheimer's disease. The dataset includes MRI scans corresponding to various cognitive conditions such as mild cognitive impairment, Alzheimer's disease, and healthy controls. Prior to training, several preprocessing operations are applied to standardize the input data and improve model performance. All MRI images are resized to a resolution of 224 × 224 pixels to match the input requirements of the deep neural network architecture. Normalization is performed using predefined mean and standard deviation values to stabilize the training process and ensure consistent pixel intensity distributions. In addition, data augmentation techniques such as horizontal flipping and small angle rotation are applied to increase the diversity of training samples and improve the generalization capability of the model.

An efficient deep learning architecture is employed in this study (Figure. 1), which is based on the MobileViT framework. MobileViT integrates convolutional neural network operations with lightweight transformer modules to effectively capture both local spatial features and global contextual relationships within MRI images. In the initial stages of the network, convolutional layers are used to extract low-level and mid-level features such as edges, textures, and structural patterns from the brain images. These layers efficiently capture local spatial information and anatomical structures associated with Alzheimer's-related brain changes.

The later stages of the architecture incorporate transformer-based blocks that enable the model to capture long-range dependencies between different regions of the MRI images. The attention mechanism allows the network to model relationships among spatial features and focus on critical brain regions that may indicate neurodegenerative changes. By combining convolutional operations with efficient transformer modules, the MobileViT architecture achieves a balance between computational efficiency and strong representation learning capability.

To further improve training efficiency, transfer learning is applied by initializing the model with pretrained weights obtained from large-scale image datasets. During training, most of the early layers of the network are frozen to retain previously learned visual representations, while the final stages and classification head are fine-tuned using the Alzheimer's MRI dataset. Additionally, a progressive unfreezing strategy is adopted, where deeper layers are gradually unfrozen during later training stages to enhance feature adaptation. This approach reduces training time while enabling the model to better capture domain-specific patterns in medical imaging data.

The model is trained using supervised learning with cross-entropy loss to measure classification errors between predicted labels and ground truth categories. Label smoothing is incorporated to improve model generalization and reduce prediction overconfidence. The AdamW optimizer is used to update network parameters due to its effectiveness in handling weight decay and stabilizing deep network training. In addition, a cosine annealing learning rate scheduler dynamically adjusts the learning rate during training to ensure smooth convergence and improved optimization performance. Regularization techniques such as dropout, gradient clipping, and weight decay are also applied to reduce overfitting.

The proposed system ultimately produces probability scores corresponding to different Alzheimer's disease categories, enabling automated classification of MRI scans. By integrating convolution-based feature extraction with lightweight transformer-based global context modeling, the proposed model demonstrates strong potential for assisting clinicians in the early detection and diagnosis of Alzheimer's disease.

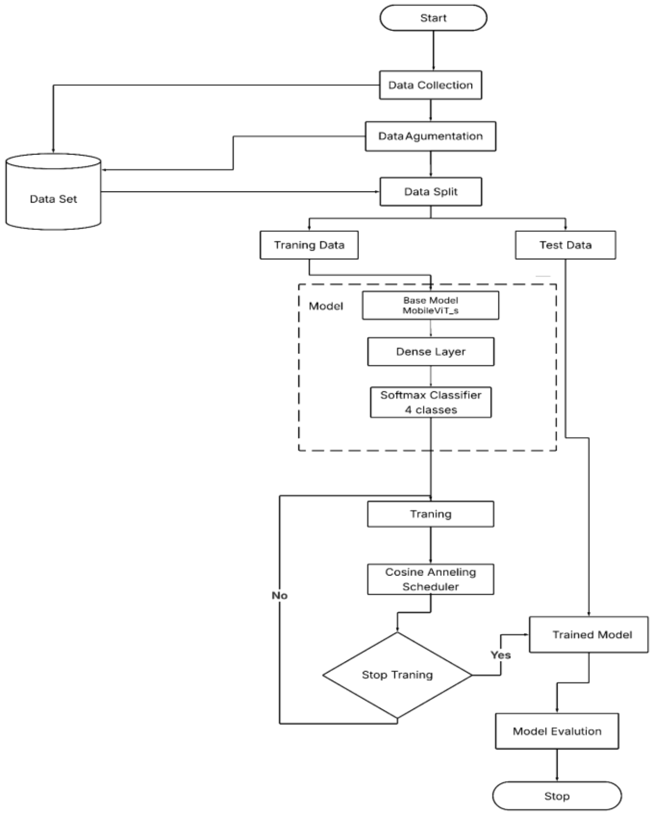

Figure 1: Proposed System

IV. RESULT

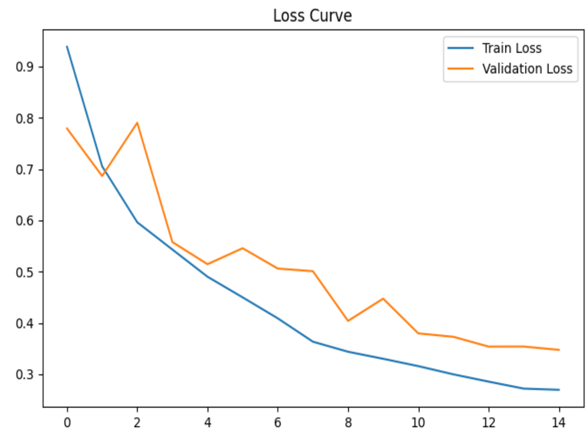

Figure 2: Loss Curve

The loss convergence behavior of the proposed MobileViT model over 15 epochs demonstrates stable and effective training, as illustrated in Figure. 2. The training loss shows a smooth and consistent decline from approximately 0.94 at the initial epoch to around 0.27 at the final epoch, indicating efficient learning of feature representations. Similarly, the validation loss decreases from approximately 0.78 to 0.35, following an overall downward trend. By the end of training, the gap between training and validation loss remains narrow, with a final difference of approximately 0.08, suggesting strong generalization capability and minimal overfitting. Overall, the results confirm stable convergence and reliable performance of the model on unseen data.

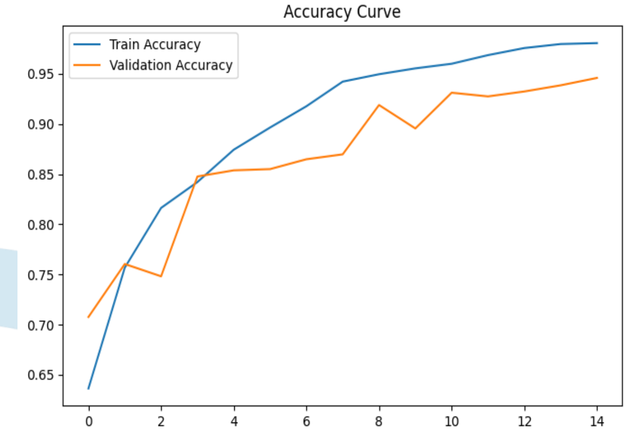

Figure 3: Accuracy Curve

The accuracy performance of the proposed MobileViT model over 15 epochs, as illustrated in Fig. 3, demonstrates strong learning and convergence behavior. The training accuracy increases steadily from approximately 0.64 at the initial epoch to around 0.98 at the final epoch, with rapid improvement observed in the early stages followed by diminishing gains after epoch 10. Similarly, the validation accuracy shows an overall upward trend, improving from approximately 0.71 to 0.945, although minor fluctuations are observed at epochs 2, 9, and 11. The model maintains consistent performance across unseen data. The final generalization gap between training and validation accuracy is very less, indicating good generalization with limited overfitting. Furthermore, the validation accuracy stabilizes during the final epochs, suggesting that the model has reached convergence and achieved a mature performance state.

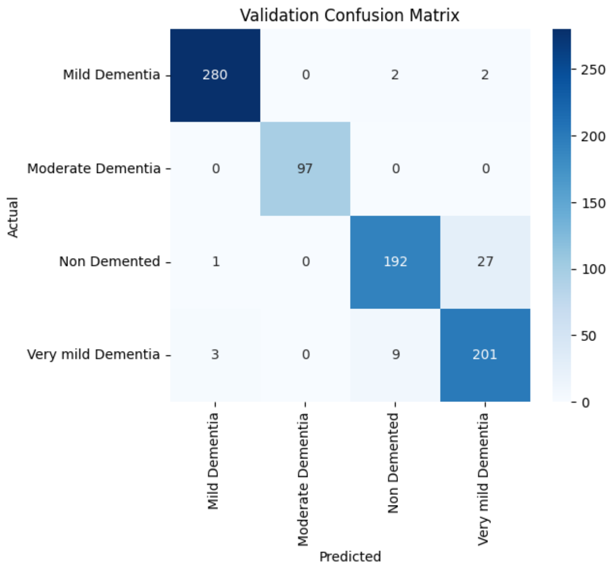

Figure 4: Confusion Matrix

The class-wise performance of the proposed MobileViT model, as shown in Fig. 4, demonstrates strong and consistent classification capability across different stages of Alzheimer's disease. The model achieves high accuracy across all categories, with particularly strong performance in Moderate and Mild Dementia classes, while also maintaining reliable results for Very Mild Dementia and Non Demented cases. Overall, the results indicate that the model effectively captures discriminative features and provides robust classification performance across multiple disease stages.

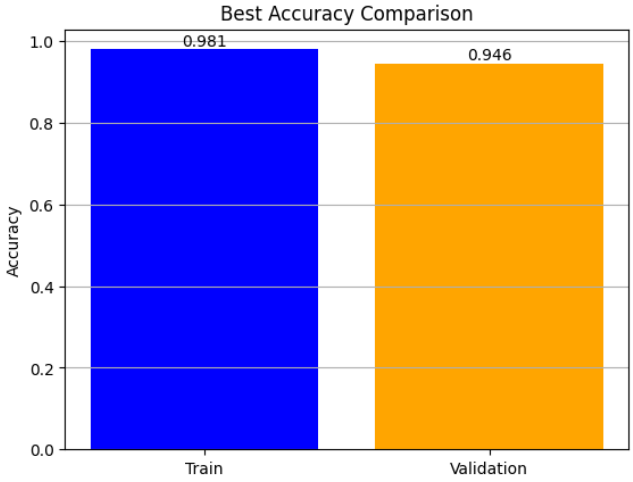

Figure 5: *Validation vs Training Accuracy*

The proposed MobileViT model achieves strong overall performance, with a peak training accuracy of 98.1% and a peak validation accuracy of 94.6%. The high validation accuracy, closely matching the training performance, indicates that the model possesses strong predictive capability and generalizes effectively to unseen data. These results demonstrate the robustness and reliability of the proposed approach for Alzheimer's disease classification.

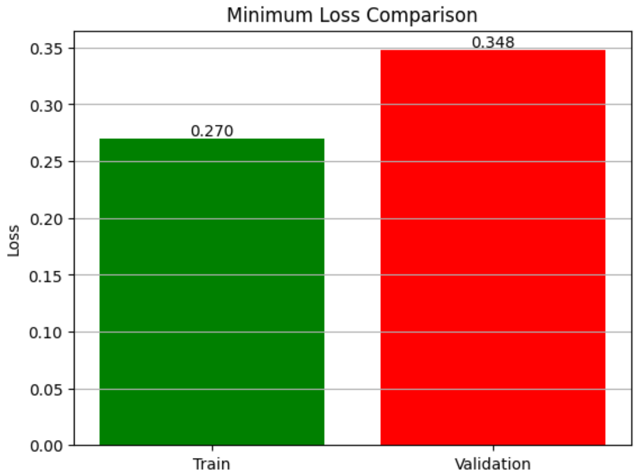

Figure 6: Validation vs Training Loss

The loss analysis of the proposed MobileViT model indicates effective optimization during training, with a minimum training loss of 0.270 and a minimum validation loss of 0.348. The relatively close values between training and validation loss further suggest good generalization performance and stable convergence of the model.

Table 1 Classification Report

Class	*Precision*	*Recall*	*F1 Score*	*Support*
Mild Dementia	0.99	0.99	0.99	284
Moderate Dementia	1.00	1.00	1.00	97
Non Demented	0.95	0.87	0.91	220
Very mild Dementia	0.87	0.94	0.91	213

V. Conclusion

The proposed Alzheimer's disease detection and classification system using a MobileViT-based architecture demonstrates the effectiveness of deep learning in supporting medical professionals with accurate and efficient diagnosis. By utilizing brain magnetic resonance imaging images, the system can effectively classify different stages of Alzheimer's disease and provide confidence scores, facilitating early detection and timely clinical intervention. The use of MobileViT ensures that the model remains computationally efficient while still capturing both local and global feature representations, making it suitable for deployment in both advanced clinical environments and resource-constrained settings. The implementation of preprocessing, data augmentation, and feature extraction techniques ensures that the input data is standardized and enriched for optimal model performance. This contributes to improved classification accuracy while minimizing misclassification across disease stages. The generated classification outputs can serve as a valuable decision-support tool for neurologists and healthcare practitioners, enhancing diagnostic efficiency and reducing manual workload. Although the system achieves strong performance, there is scope for further enhancement, including the incorporation of 3D MRI analysis, advanced attention mechanisms, explainable AI techniques, and multi-modal data integration. These improvements could further increase model interpretability and diagnostic reliability, supporting broader adoption in real-world clinical applications. In conclusion, this work emphasizes the growing impact of artificial intelligence in healthcare and demonstrates how efficient architectures such as MobileViT can be effectively applied to complex medical imaging tasks. With continued advancements and validation, the proposed system has the potential to evolve into a scalable and reliable tool for early detection of Alzheimer's disease and improved patient care outcomes.

References

[1] Hoang, Vo Duc, Cao The Thang, and Nguyen Nang Hung Van. "Applying the Vision Transformer Algorithm to Diagnose the Stages of Alzheimer's Disease." In *2024 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 287-291. IEEE, 2024.

[2] Kaushik, Pratham, Vinay Kukreja, and Amanveer Singh. "Deep Learning-Based Detection and Severity Classification of Alzheimer's Disease Stages Using a Custom CNN Model." In *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*, pp. 1-5. IEEE, 2024.

[3] Balasubramanian, M., T. Vishnu Sriram, G. Gautam Kumar, J. Jeo Godson Samuel, and M. Kanishka. "A CNN-Based Framework for Alzheimer's Disease Diagnosis and Prognosis from Multimodal Imaging Data." In 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), pp. 1-6. IEEE, 2024.

[4] Nair, Aswathy R., and Monisha Mohan. "A Study on Various Machine Learning Techniques for the Early Detection of Alzheimer Disease Using Neuro Images." In *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, pp. 998-1003. IEEE, 2024.

[5] Bagade, Virendra, and S. P. Godse. "Early detection of Alzheimer's disease based on the state-of-the-art deep learning approach." In *2024 IEEE Pune Section International Conference (PuneCon)*, pp. 1-7. IEEE, 2024.

[6] Salahudin, Amna, Eeman Qadeer, Ayesha Ahmed, and Labiba Gillani Fahad. "Enhanced Classification of Alzheimer's Disease Using Deep Learning and Machine Learning Models with MRI Scans." In *2024 International Conference on Frontiers of Information Technology (FIT)*, pp. 1-6. IEEE, 2024.

[7] Sambana, Bosubabu, Kathrigalla Keerthana, Bala Sandeep, Kothapalle Vishnu Vardhan, Ballari Sai Kiran, and Makkina Surya Prakash. "EfficientNet and Vision Transformers for Alzheimer's Detection and Diagnosis." In *2025 5th International Conference on Expert Clouds and Applications (ICOECA)*, pp. 489-494. IEEE, 2025.

[8] Li, Jiaqiang, Yian Gao, Peng Yang, Zhenghua Guan, Tianfu Wang, Guolin Ma, and Baiying Lei. "Multi-Modal Feature Fusion Using Transformer for Early Alzheimer's Disease Diagnosis." In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1-4. IEEE, 2025.

[9] Yeasir, Md Rafsan, Fatin Nibbrash Nakib, Md Arafat Ibna Mizan, Md Musfiqur Rahman Mridha, Md Shahid Ahammed Shakil, Sakib Imtiaz, and Hafijur Rahman Hemel. "Innovative Lightweight CNN Model for Stage-Wise Alzheimer's Disease Detection." In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, pp. 1-5. IEEE, 2025.

[10] Yadav, Vivek, Sushil Kumar, and Umang Rastogi. "Progressing Alzheimer's Diagnosis with Ensemble CNN Model." In *2025 3rd International Conference on Disruptive Technologies (ICDT)*, pp. 1134-1138. IEEE, 2025.

[11] Ravichandran, Vinaya Varshini, Priyan Malarvizhi Kumar, and Balasubramanian Prabhu Kavin. "Deep Learning-Driven Early Detection of Alzheimer's Disease using CNN." In *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, pp. 69-74. IEEE, 2025.

[12] Li, Hanzhi. "Vision Mamba for Accurate and Efficient Alzheimer's Disease Classification via Brain MRI." In *2025 6th International Conference on Electronic Communication and Artificial Intelligence (ICECAI)*, pp. 409-413. IEEE, 2025.