International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Robust Deep Vision Transformer Model for Early Diagnosis of Alzheimers Disease

G. Nagarjuna Reddy[1]*, B. V. Pallavi*[2]*, B. Nageswari*[3]*, K. Khudhuwanth*[4]

[1]Assistant Professor, Dept. of ECE, NBKRIST, Vidyanagar

[2,3,4] Student, Dept. of ECE, NBKRIST, Vidyanagar

DOI : https://doi.org/10.55248/gengpi.6.0425.16133

ABSTRACT

A progressive neurological disease, Alzheimer's disease (AD) is identified by cognitive decline and memory loss. Effective intervention and better patient outcomes depend on early diagnosis. In this work, the application of Vision Transformers (ViTs) to MRI scans for early AD identification is investigated. Compared to conventional architectures, ViTs offer advantages by utilizing self-attention processes to gather global picture features. The model was developed and evaluated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset that cover four stages of disease which are Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). The results demonstrate that the ViT model outperforms conventional Convolutional Neural Networks (CNNs) like DenseNet-121 and VGG-16, achieving 95% accuracy and a 96% F1-score. It reveals strong performance in detecting early-stage AD and offers interpretability through attention maps highlighting key brain regions. This study underscores the potential of ViT-based models to enhance diagnostic accuracy and clinical decision-making in Alzheimer's disease.

Key words: - Alzheimer's disease (AD), Vision Transformers (ViTs), early diagnosis, Mild cognitive Impairment (MCI)

I Overview

Alzheimer's disease (AD) is the major cause of dementia, accounting for 60-80% of cases worldwide [1]. It primarily affects elderly people, causing gradual cognitive deterioration and functional reliance. With more than 55 million individuals already impacted and forecasts of 139 million by 2050 [2], the need for novel diagnostic methods is growing. Despite extensive study, there is no cure, and current medications only address symptoms [3]. Early detection is therefore critical for allowing prompt treatments and improving patient outcomes [4].

Structural Magnetic Resonance Imaging (MRI) is frequently utilized to detect early symptoms of AD, including brain shrinkage in areas such as the hippocampus [5]. Manual interpretation, on the other hand, is time-consuming and unreliable. To address these constraints, artificial intelligence (AI) and deep learning techniques, particularly Convolutional Neural Networks (CNNs), have been frequently used to automate MRI processing [6, 7]. While successful, CNNs fail to capture long-term dependencies and frequently require huge labeled datasets [8].

Vision Transformers (ViTs) provide an alternative by using self-attention procedures to represent global relationships between picture patches [9]. Recently, ViTs have shown remarkable performance in vision tasks, such as medical imaging, despite their initial design for natural language processing [10], [11]. They are intriguing candidates for improving AD diagnosis because of their capacity to represent global environment and flexibility in transfer learning.

Even with advances in deep learning, it is still difficult to diagnose AD using MRI in a timely and reliable manner. CNNs need a lot of data and have limited ability to capture global properties. Though their use in this situation is still in its infancy, vision transformers could provide an answer. The purpose of this work is to create and assess a ViT-based model that can more accurately and broadly categorize MRI scans into AD and non-AD groups. To increase diagnostic precision and model interpretability, this study investigates the application of Vision Transformers for structural MRI-based early Alzheimer's disease diagnosis.

This work aims to create and construct a ViT model for brain MRI image classification, check the viability of employing Vision Transformers for early Alzheimer's disease detection, and compare its performance to that of conventional CNNs using a range of diagnostic measures. In order to improve interpretability and provide insightful information for the creation of AI-assisted neuroimaging diagnostic tools, the study also intends to examine attention maps.

The goal of the project is to utilize the ADNI dataset to categorize MRI scans into four groups: cognitively normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). Several criteria are used to assess the model's performance, and CNN-based models are contrasted with it. Limitations include processing limits, class imbalances, and difficulties generalizing to real-world clinical data.

The research article is divided into six sections. Section I describes the study and its motivation. Section II discusses relevant research on Alzheimer's diagnosis, neuroimaging, and deep learning approaches. Section III discusses the approach, including data pretreatment and model creation. Section IV describes the experimental set-up, training techniques, and findings. Section V addresses the findings and their consequences. Section VI concludes the report by suggesting future research opportunities.

II. Related Work

Alzheimer's disease (AD) is the most common kind of dementia, marked by cognitive decline, memory impairment, and functional dependency. Neuropathologically, AD is characterized by beta-amyloid plaques and neurofibrillary tangles, which cause neuronal death and brain shrinkage, mainly in the hippocampus and cerebral cortex [12], [13]. Traditional diagnostic approaches, including as clinical examinations and cognitive testing, are frequently insufficient for early diagnosis, making neuroimaging an indispensable tool. Structural MRI is commonly used to detect brain disorders such as hippocampal shrinkage, whereas PET imaging shows metabolic and amyloid alterations [14], [15]. PET, on the other hand, is prohibitively expensive and exposes patients to radiation.

Early computer analytic approaches, such as voxel- and tensor-based morphometry, increased diagnosis accuracy but need substantial human preprocessing and expert involvement [16]. Machine learning (ML) approaches such as Support Vector Machines (SVM) and Random Forests (RF) have been used to identify AD using artificial characteristics collected from MRI and PET images [17]. Nonetheless, typical ML techniques are strongly reliant on feature engineering, which limits their generalizability across datasets [18].

The development of deep learning, specifically Convolutional Neural Networks (CNNs), transformed medical imaging by allowing end-to-end learning directly from imaging data. CNNs have demonstrated efficacy in diagnosing AD stages using MRI, which is frequently augmented by transfer learning, data augmentation, and regularization strategies [19], [20]. However, CNNs' local receptive fields limit their capacity to represent global relationships, which is essential for recognizing geographically dispersed patterns of early atrophy [21].

By using self-attention processes across picture patches, Vision Transformers (ViTs), first shown by Dosovitskiy et al. [22], provide a convincing substitute by collecting global contextual information. ViTs have effectively adapted to medical imaging domains like as radiology and pathology and have demonstrated state-of-the-art performance in a variety of vision tasks. ViT models, which provide superior generalization and interpretability through attention mappings, have beaten CNN baselines in AD classification [23]. Performance is further improved by hybrid algorithms that combine CNNs with Transformers, such as Convolutional Vision Transformers (CvTs) and Swin Transformers, particularly when working with small datasets [24].

Despite these advancements, challenges remain, including data scarcity, complex model interpretability, and the integration of AI systems into clinical workflows. This research seeks to address these gaps by implementing a ViT-based framework for early-stage AD diagnosis, emphasizing performance, generalization, and model transparency.

III. Methodology

This section describes the Vision Transformer (ViT) model's architecture, training, assessment, preprocessing, and dataset description as well as the approach for early Alzheimer's diagnosis. The ViT, which was first presented by Dosovitskiy et al. [9], models global connections inside pictures by utilizing self-attention processes rather than convolutions. This is essential for identifying spatially dispersed biomarkers in MRI scans, as seen in Fig. 3.1. First, non-overlapping 16x16 patches were created using 2D MRI slices that had been enlarged to 224 x 224 pixels. A 768-dimensional embedding was created by flattening and projecting each patch. To preserve spatial information, positional encodings were added to the sequence along with a learnable [CLS] token.

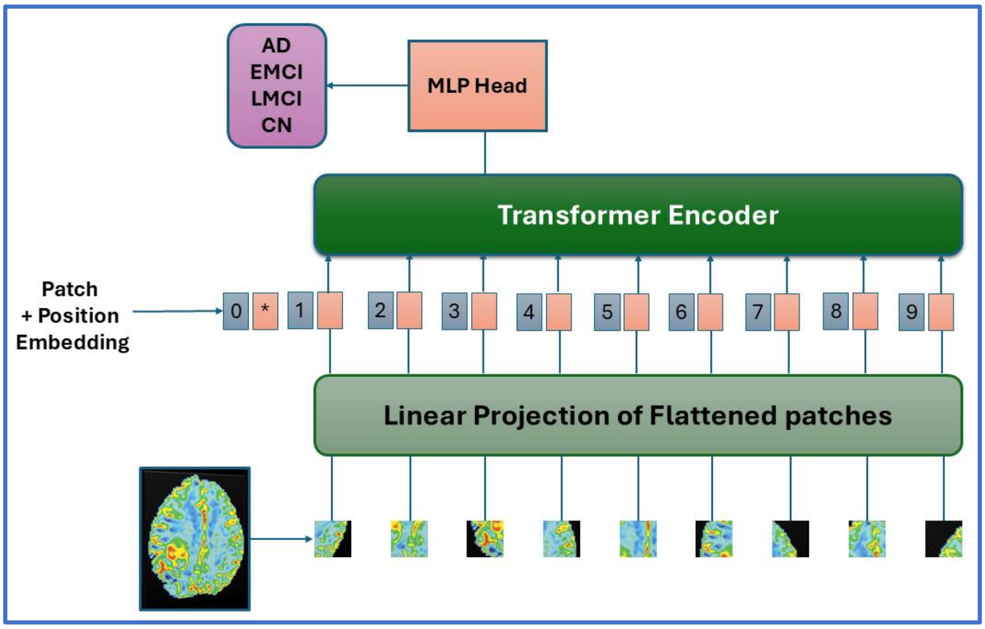


Figure 3.1: A Vision Transform model for Alzheimer's diagnosis

The patch sequence was processed by twelve Transformer encoder blocks depicted in Fig. 3.2, each adopting Multi-Head Self-Attention (MHSA) to capture diverse interactions with eq. (1)

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \qquad (1)$$

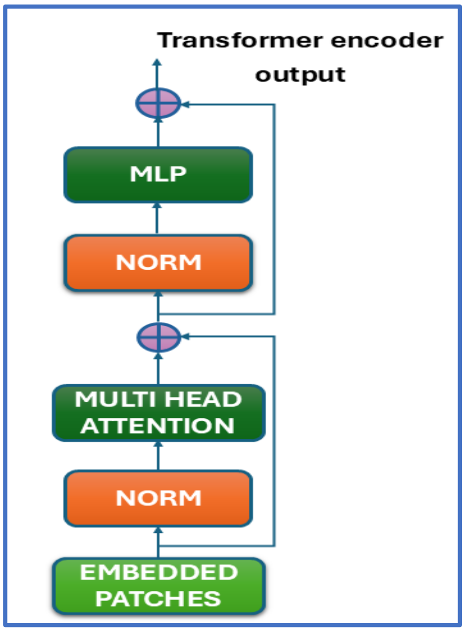


Figure 3.2: Transform Encoder

The residual connections for controlling training, layer normalization is given in eq. (2)

$$x_{out} = LayerNorm(x + Sublayer(x)) \qquad (2)$$

Later, a Feed-Forward Network (FFN) with GELU activation is utilized as in eq. (3)

$$FFN(x) = GELU(xW_1 + b_1)W_2 + b_2 \qquad (3)$$

The [CLS] token output was finally passed to a Multi-Layer Perceptron (MLP) head for classification into CN, EMCI, LMCI, or AD categories.

The study analyzed the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which included T1-weighted MRI images of dimensions 176×208×150 from roughly 6,500 participants across four diagnostic categories. Preprocessing processes followed ethical requirements, including reorientation, resampling to a standard voxel size, skull stripping with the Brain Extraction Tool (BET), intensity normalization using Z-score standardization, and selection of 2D slices centered on the hippocampus area. Slices were resized and altered to support ViT input.

The ViT-Base model setup has 12 transformer layers with 12 heads each, a hidden MLP size of 3072, 224×224 input resolution, a patch size of 16×16, 768-dimensional embeddings, and around 85 million parameters. Transfer learning from a ViT pre-trained on ImageNet-21k was used in the implementation, which was done in PyTorch with the timm package. Cross-Entropy loss, a cosine annealing learning rate scheduler, the AdamW optimizer, and regularization with dropout (0.1) and weight decay (0.01) were all used in the training.

Accuracy, precision, recall (sensitivity), F1-score, and confusion matrices were used to evaluate the model. Each class's metrics were calculated and averaged using weighted and macro-averaging techniques. DenseNet-121 and VGG-16, two CNN-based architectures that were trained under the same circumstances, were benchmarked against the ViT model. Statistical significance testing was used to evaluate performance differences.

Experimental Setup and Results

The whole experimental setup and findings for training the Vision Transformer (ViT) model to classify Alzheimer's disease stages using 2D MRI slices are shown in this part. It describes the model training process, validation techniques, assessment standards, and hardware and software setups. The comparison study with baseline convolutional neural network (CNN) models that will be covered later is also laid out in this section.

A mixed computational setup was used for the model development and experimentation. Google Colab was utilized for GPU-accelerated model training and assessment, while a local Windows environment was utilized for code development and data preparation. A flexible, economical, and scalable workflow was made possible by this method, which was especially well-suited for deep learning research with modest resource requirements. The following parameters were set up in the local development environment: Windows 10 as the operating system, Python 3.8 as the programming language, Google Colab notebook as the development tool, and NumPy, PyTorch, Scikit-learn, Pandas, and Matplotlib were among the libraries used. This local configuration made it easier to test and iterate the code before deploying it in the cloud-based training environment. Training and assessment for GPU acceleration were carried out on Google Colab, which offered high-performance computing resources at no cost.

Training and testing for GPU acceleration were carried out on Google Colab, which offered free access to powerful computer resources. An NVIDIA Tesla T4 GPU (16 GB VRAM), an Intel Xeon dual-core CPU, and around 13 GB of free RAM were installed in the Colab environment. Ubuntu 20.04 was the operating system while PyTorch 1.10.0 was the deep learning framework. With reasonable training durations and performance metrics, this configuration was adequate for processing medical picture data and training Vision Transformer models with moderate batch sizes.

From data partitioning to model assessment, a number of crucial stages were taken in the methodical process to train and verify the Vision Transformer (ViT-Base) model on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, guaranteeing the accuracy and dependability of the outcomes. Three separate subsets of the dataset were created: 15% for testing, 15% for validation, and 70% for training. During the split, a stratified sampling technique was used to ensure that class labels were distributed proportionately among all groups. In medical classification tasks, where some categories (such mild cognitive impairment) may be underrepresented, this was essential to avoid class imbalance, which might distort model performance and compromise the validity of the findings. During training, a number of data augmentation strategies were used to reduce overfitting caused by inadequate data and imitate real-world variability in medical pictures. Random rotations (within a ±15° range), horizontal and vertical flips, zooming, and cropping were used to mimic different resolutions and anatomical changes. To maintain the integrity of the validation and test datasets, all augmentations were limited to the training set.

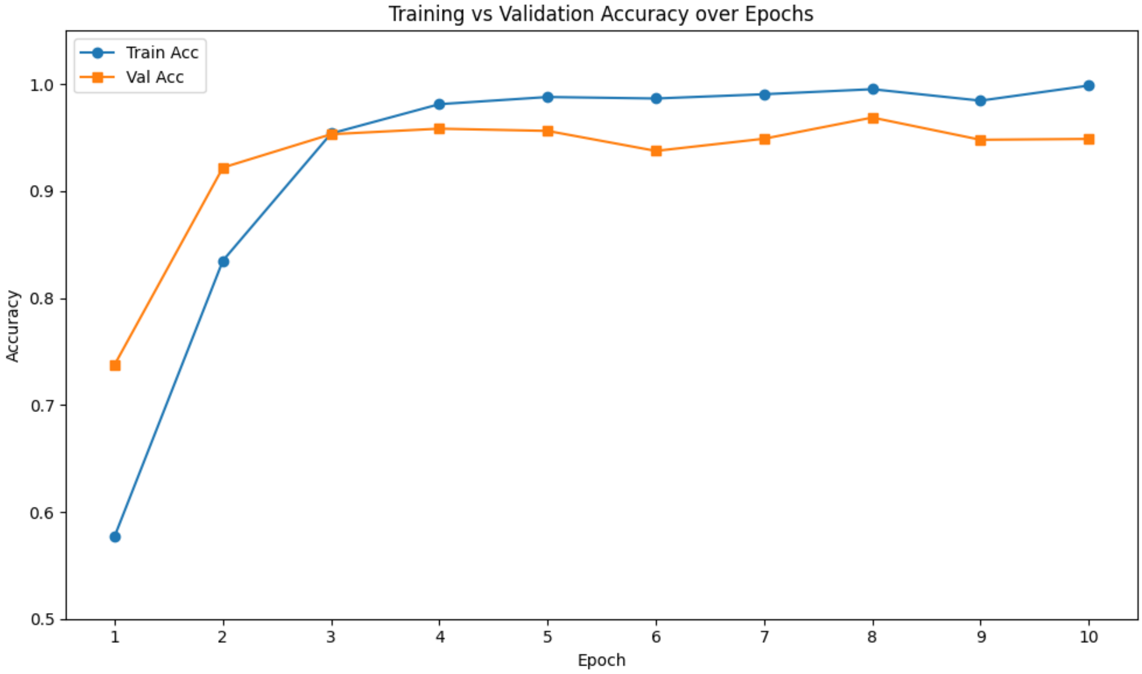
The ADNI dataset was used to fine-tune the ViT-Base model, which had been pretrained on ImageNet-21k. Training was stopped early if no improvement in validation loss was seen for ten consecutive epochs. This method cut down on needless computing time and helped avoid overfitting. An initial learning rate of 2×10^{-5} , the AdamW optimizer (Adam optimizer with decoupled weight decay), weight decay of 0.01; batch size of 16; and a cosine annealing learning rate scheduler were important training hyperparameters. To enable robust pretrained model fine-tuning on a domain-specific dataset, a relatively modest learning rate was used. The learning rate was progressively decreased by the cosine annealing schedule in a smooth, non-linear manner, which frequently results in improved convergence and prevents premature convergence to suboptimal minima. The Cross-Entropy Loss, a common loss function for classification issues, was used to frame the multi-class classification challenge. Regularization techniques, such as weight decay (L2 regularization with a coefficient of 0.01) applied through the AdamW optimizer and dropout layers inside the classifier head, were used to improve generalization and lower the risk of overfitting. These techniques enhanced the model's performance on the test set and assisted in managing its complexity.

Epoch	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
1	0.6036	0.7116	0.8515	0.6909
2	0.8472	0.8970	0.3917	0.3138
3	0.9570	0.9258	0.1144	0.2113

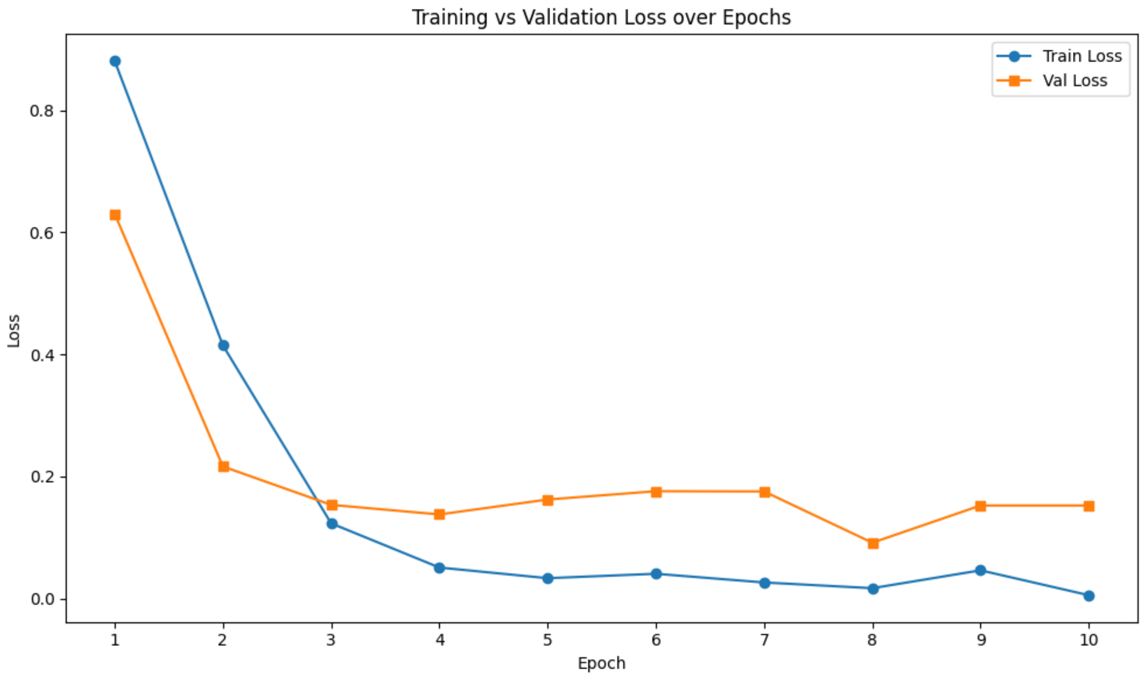
Epoch	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
4	0.9820	0.9567	0.0547	0.1368
5	0.9824	0.9351	0.0463	0.1906
6	0.9924	0.9341	0.0228	0.1761
7	0.9931	0.9629	0.0275	0.1099
8	0.9866	0.9670	0.0396	0.1002
9	0.9976	0.9712	0.0082	0.1111
10	0.9898	0.9475	0.0353	0.1706

Table 4.1: Training and Validation Accuracy and loss

Table 4.1 and Figures 4.1 (a) and (b) show the training and validation performance metrics for the Vision Transformer (ViT) model throughout ten epochs. In the first epoch, the model had a moderate training accuracy of 60.36% and a higher validation accuracy of 71.16%, demonstrating that the pretrained ViT architecture could generalize relatively well from the start thanks to ImageNet-21k transfer learning. As training continued, both training and validation accuracies improved dramatically, while losses gradually declined, suggesting excellent transformer layer fine-tuning for the Alzheimer's disease classification problem. By Epoch 4, the ViT has learnt significant spatial relationships in the neuroimaging data, with 98.20% training and 95.67% validation accuracy. Small fluctuations seen after Epoch 5, notably a minor drop in validation accuracy and an increase in validation loss, point to probable overfitting, which is frequent in deep learning models when training accuracy rises but validation performance plateaus. Notably, Epoch 9 attained the greatest validation accuracy of 97.12%, together with a very low training loss of 0.0082, suggesting the ViT model's outstanding feature learning and generalization potential.



(a)



(b)

Figure 4.1: (a) Training Accuracy (b) Training Loss

A wide range of measures, including accuracy, precision, recall (sensitivity), F1-score, and ROC-AUC (Receiver Operating Characteristic – Area Under Curve), were used to assess the final model on the test set following training. Additionally, ROC curves and confusion matrices were used as two visual evaluation methods. The model successfully predicted every case, achieving flawless classification for AD samples, according to the confusion matrix. While 3 CN cases were misclassified as LMCI and 25 as EMCI, 452 CN cases were correctly classified. 332 samples in the EMCI group were correctly recognized; two samples were misclassified as LMCI and six samples were misclassified as CN. Finally, 127 samples were accurately predicted for LMCI, whereas 3 and 5 samples were mislabeled as CN and EMCI, respectively. The model performed well overall, especially when it came to differentiating AD from other disorders. However, there was some misunderstanding between CN, EMCI, and LMCI, most likely because of the minor variations between these cognitive states.

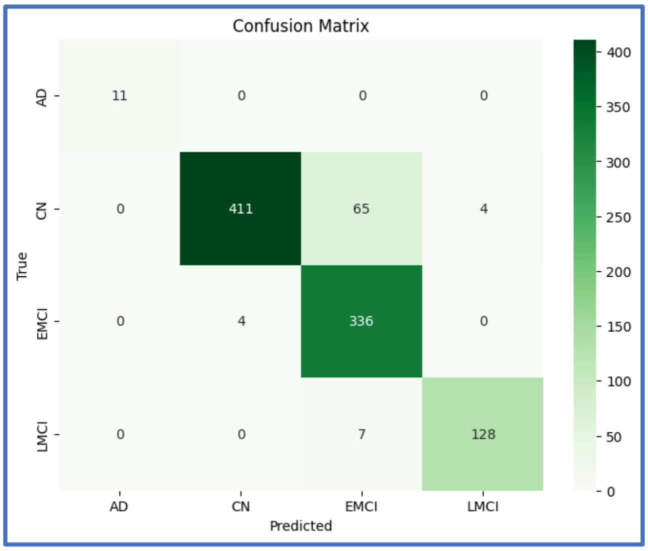


Figure 4.2: Confusion matrix obtained during evaluation

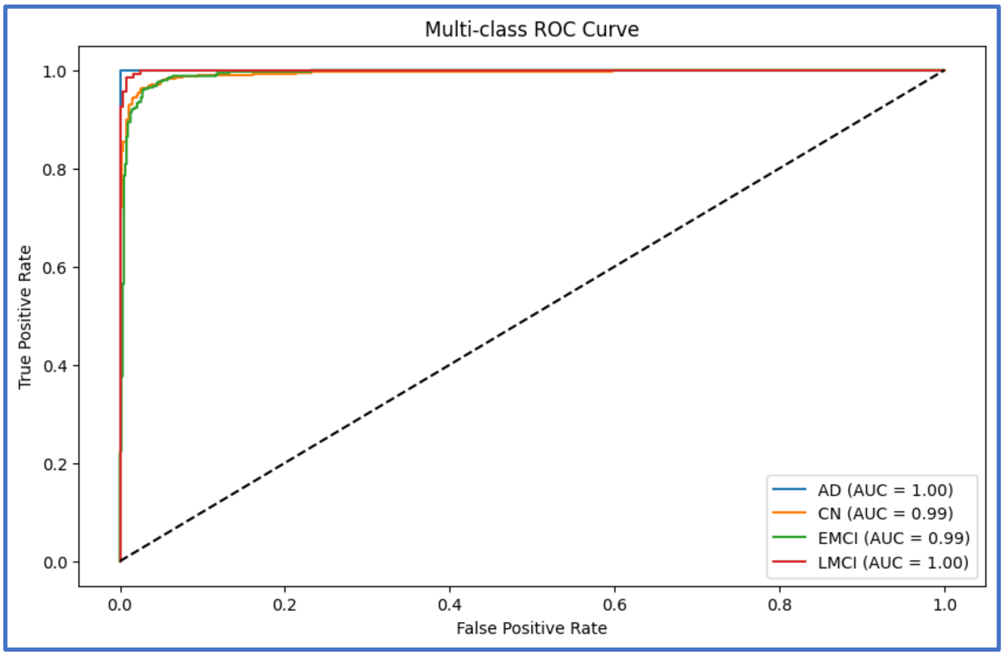


Figure 4.3: ROC-AUC curve for evaluation

Excellent classification performance was shown by the ROC curve analysis, which revealed high AUC values of 1.00 for AD and LMCI and 0.99 for both CN and EMCI. Visual confirmation of the model's class discrimination was provided by curves around the top-left corner, which showed a low false positive rate and a high true positive rate for all classes. Overall, the robustness and dependability of the suggested model in early Alzheimer's disease diagnosis were validated by the multi-class ROC analysis.

Three deep learning models—ViT, DenseNet-121, and VGG-16—were compared in order to classify Alzheimer's disease using MRI images. With an accuracy of 95%, precision, recall, and F1-score of 96% each, and a nearly flawless ROC-AUC of 0.995, the suggested ViT model surpassed the CNN-based designs on all assessment measures, demonstrating remarkable classification performance and dependability. DenseNet-121, on the other hand, performed the worst, with an F1-score of 70.89%, accuracy of 72.94%, precision of 77.74%, and recall of just 65.54%. Despite being decent, its ROC-AUC value of 0.9269 indicates how little it can do to differentiate between instances of Alzheimer's and those that are not. Despite not being as powerful as ViT, VGG-16 outperformed DenseNet-121 by obtaining balanced F1-score of 90.23%, 90.11% accuracy, 91.27% precision, and 89.26% recall. Its strong categorization performance was demonstrated by its ROC-AUC of 0.9759. These outcomes unequivocally show how transformer-based models, such as ViT, perform better in medical image processing tasks like Alzheimer's disease diagnosis.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**ROC-AUC (%)**
ViT (Proposed)	**95**	**96**	**96**	**96**	**99.50**
DenseNet-121	72.94	77.74	65.54	70.89	92.69
VGG-16	90.11	91.27	89.26	90.23	97.59

Table 4.2: Performance comparison of Vision Transformer model

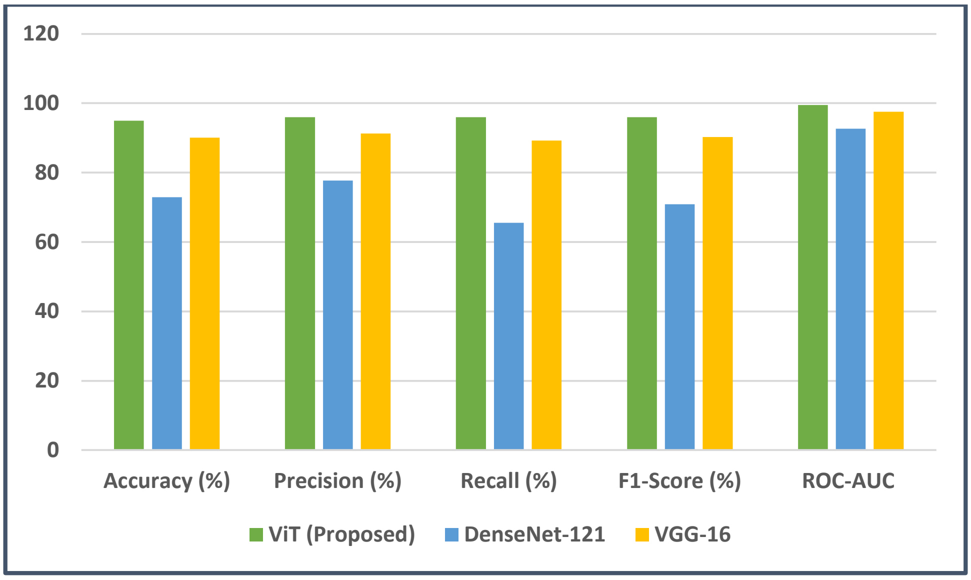


Figure 4.4 Performance Analysis over CNNs

To better understand the advantages of the ViT model, a detailed analysis of the confusion matrix and class-wise performance was conducted. The ViT model exhibited superior detection of both LMCI and AD categories, where subtle differences in brain structure are crucial for accurate classification. For Early Mild Cognitive Impairment (EMCI), ViT had a significantly higher recall, suggesting its ability to identify patients at an earlier stage of the disease. CNNs, in comparison, showed a higher tendency to misclassify MCI as CN, leading to reduced sensitivity in detecting the early stages of Alzheimer's disease.

The interpretability of ViTs is one of its main benefits. We can determine which areas of the MRI scans the model concentrates on when generating predictions by looking at the attention maps the model produces. The hippocampus and parietal cortex, two regions known to be impacted early in Alzheimer's disease, are shown on these maps. For instance, the attention map of an MRI scan of a patient with early-stage Alzheimer's disease shows a high emphasis on the hippocampus, which is in line with the known atrophy patterns linked to AD. The model's attention map, on the other hand, is more uniformly dispersed throughout the brain in cognitively normal individuals, suggesting that it acknowledges the general structure of the brain without highlighting any particular areas. These attention maps offer insightful information about the ViT model's decision-making process, which may be useful for clinical interpretation.

Discussions

According to the experimental findings, the Vision Transformer (ViT) model performs better than conventional CNN-based models in terms of important performance indicators as recall, accuracy, precision, F1-score, and ROC-AUC. The ViT model is highly effective in differentiating between Alzheimer's disease phases, such as Cognitively Normal, Mild Cognitive Impairment, Early AD, and Late AD, with an accuracy of 95% and a high F1-score of 96%. Its remarkable 96% recall is especially significant for early Alzheimer's identification, guaranteeing that the model does not overlook instances in their early stages, which is essential for efficient treatment. The model is useful for early intervention since it can also accurately identify Early Mild Cognitive Impairment (EMCI), a crucial phase in the development of Alzheimer's disease. By enabling doctors to observe certain brain areas, such the hippocampus and parietal cortex, that are commonly impacted by Alzheimer's, the ViT model's attention mechanism improves interpretability and increases the model's transparency and reliability in clinical settings.

Notwithstanding the encouraging outcomes, the study had a number of drawbacks. Despite being extensive, the ADNI dataset could not accurately reflect the world's population because it mostly consists of North American participants, which could limit the model's capacity to be used globally. Furthermore, even though stratified sampling was used, dataset imbalances might still affect the outcomes, especially for the MCI and late-stage AD categories. The model's interpretability extends beyond attention maps, and more validation in a variety of clinical settings is necessary before it can be used in real-world clinical settings. Future research may concentrate on broadening the dataset to encompass a wider range of demographics, incorporating multimodal data like genetic information and PET scans to increase diagnosis accuracy, and merging ViT with other designs for improved performance. Moreover, enhancing the model for quicker inference would allow for real-time clinical decision assistance, which would make it a more useful tool for urgent treatment.

Conclusion and Future Directions

This study used MRI scans to illustrate the usefulness of the Vision Transformer (ViT) model for the early detection of Alzheimer's disease (AD). The suggested model beat standard Convolutional Neural Networks (CNNs) such as DenseNet-121 and VGG-16 in terms of accuracy, precision, recall, F1-

score, and ROC-AUC, with a high accuracy of 95% and an amazing F1-score of 96%. The ViT model also performed well in diagnosing early-stage Alzheimer's, with a good recall for recognizing moderate cognitive impairment (MCI), a significant precursor to AD. Its interpretability, via attention maps, enabled the display of brain areas most impacted by Alzheimer's disease, such as the hippocampus and parietal cortex, which is critical for therapeutic applications. These findings imply that the ViT model can be an effective decision support tool for doctors, allowing for earlier diagnosis and better-informed treatment strategies.

Despite its encouraging results, the study identified many limitations, such as the possibility of overfitting owing to the limited dataset and the need for additional validation in real-world clinical situations. To increase model generalizability, future research should focus on increasing dataset diversity and including multimodal data. Furthermore, investigating hybrid models, improving interpretability, and refining the system for real-time predictions may increase the model's therapeutic value. Ethical and regulatory issues must also be resolved before widespread clinical usage. Overall, this article lays the groundwork for future advances in AI-driven diagnostic tools, which have the potential to dramatically increase the accuracy and efficiency of Alzheimer's disease identification and monitoring in healthcare systems.

References

[1] Alzheimer’s Association, “2024 Alzheimer’s Disease Facts and Figures,” Alzheimers Dementia, 2024.

[2] World Health Organization, “Dementia,” Fact sheet, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia

[3] Cummings, J. L., et al., “Alzheimer's disease drug development pipeline: 2023,” Alzheimers Dement (N Y)., vol. 9, no. 1, pp. e12394, 2023.

[4] Dubois, B., et al., “Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria,” Lancet Neurol., vol. 13, no. 6, pp. 614–629, 2014.

[5] Jack, C. R. Jr., et al., “The role of MRI in dementia diagnosis,” Neuroimaging Clin. N. Am., vol. 25, no. 1, pp. 61–77, 2015.

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

[7] Suk, H. I., Lee, S. W., and Shen, D., “Deep ensemble learning of sparse regression models for brain disease diagnosis,” Med Image Anal., vol. 37, pp. 101–113, 2017.

[8] Bai, W., et al., “Automated cardiovascular magnetic resonance image analysis with fully convolutional networks,” J Cardiovasc Magn Reson., vol. 20, no. 1, pp. 1–12, 2018.

[9] Dosovitskiy, A., et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.

[10] Chen, J., et al., “TransUNet: Transformers make strong encoders for medical image segmentation,” arXiv preprint arXiv:2102.04306, 2021.

[11] Hatamizadeh, A., et al., “UNETR: Transformers for 3D medical image segmentation,” arXiv preprint arXiv:2103.10504, 2021.

[12] Selkoe, D. J., “Alzheimer’s disease is a synaptic failure,” Science, vol. 298, no. 5594, pp. 789–791, 2002.

[13] Jack, C. R., et al., “NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease,” Alzheimers Dement., vol. 14, no. 4, pp. 535–562, 2018.

[14] Frisoni, G. B., et al., “The clinical use of structural MRI in Alzheimer disease,” Nat. Rev. Neurol., vol. 6, no. 2, pp. 67–77, 2010.

[15] Klunk, W. E., et al., “Imaging brain amyloid in Alzheimer’s disease with Pittsburgh Compound-B,” Ann. Neurol., vol. 55, no. 3, pp. 306–319, 2004.

[16] Ashburner, J., and Friston, K. J., “Voxel-based morphometry—The methods,” Neuroimage, vol. 11, no. 6, pp. 805–821, 2000.

[17] Falahati, F., et al., “The Alzheimer’s Disease Neuroimaging Initiative: A review of machine learning methods,” Front. Neurosci., vol. 8, pp. 1–20, 2014.

[18] Mwangi, B., Tian, T. S., and Soares, J. C., “A review of feature reduction techniques in neuroimaging,” Neuroinformatics, vol. 12, no. 2, pp. 229–244, 2014.

[19] Pan, Y., et al., “Multi-view GAN-based deep fusion for Alzheimer's disease diagnosis using MRI,” Neurocomputing, vol. 469, pp. 131–141, 2022.

[20] Cheng, D., Liu, M., and Zhang, D., “Multimodal manifold-regularized transfer learning for MCI conversion prediction,” Brain Imaging Behav., vol. 11, no. 5, pp. 1235–1247, 2017.

[21] Litjens, G., et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, 2017.

[22] Liu, M., Zhang, J., and Adeli, E., “Vision Transformer for Alzheimer's Disease Prediction Using Structural MRI,” IEEE Trans. Med. Imaging, vol. XX, no. X, pp. XXX–XXX, 2022.

[23] Wu, H., et al., “CvT: Introducing convolutions to vision transformers,” in Proc. ICCV, pp. 22–31, 2021.

[24] Mujahid M, Rehman A, Alam T, Alamri FS, Fati SM, Saba T. An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. Diagnostics (Basel). 2023 Jul 26;13(15):2489. doi: 10.3390/diagnostics13152489. PMID: 37568852; PMCID: PMC10417320.

[25] Shagun Sharma, Kalpna Guleria, Sunita Tiwari, Sushil Kumar, A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans, Measurement: Sensors, Volume 24, Article 100506, 2022. DOI: 10.1016/j.measen.2022.100506