iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.69045

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Classification of Alzheimer's disease Stages using Vision Tranformers

Bakshish Singh[1], Dhanraj Singh[2], Aditya Vikram Singh[3], Tushar Singh[4]
Faculty of Information Technology, Department of Information Technology, JSSATEN, AKTU, University, Noida

Abstract: ***Alzheimer's disease progresses as a neurodegenerative disorder that creates substantial obstacles during early diagnosis and therapeutic intervention. Brain Magnetic Resonance Imaging (MRI) offers promising diagnostic help but requires advanced analytical methods to detect the disease's subtle structural changes. New deep learning models which include Vision Transformers (ViTs) now receive significant attention due to their superior performance in medical imaging applications. Vision Transformers apply self-attention mechanisms to capture long-range data relationships which enhances their performance in brain MRI scan analysis. Researchers investigated how the ViT base architecture can differentiate between Alzheimer's disease stages.***

Keywords: Vision Transformers, Endmembers, Spectrum, Abundance Estimation, Spectral Analysis

I. INTRODUCTION

The World Health Organization (WHO) reports that dementia currently affects 55 million individuals globally and projections indicate this figure will increase to 78 million by the year 2030. The World Health Organization (WHO 2022) reports that Alzheimer's disease accounts for 60–70% of dementia cases. Alzheimer's disease represents a degenerative brain disorder that affects people older than 65 by causing memory loss and cognitive deterioration. Scientists cannot determine AD's exact cause but agree that genetic factors along with environmental elements play a part [4]. While scientists have yet to discover a cure for the disease treatments exist which can minimize symptoms and improve patients' quality of life. Patients typically experience memory lapses as well as difficulty with tasks alongside language problems disorientation poor judgment abstract thinking issues item misplacement mood swings and decreased motivation [6]. Before any visible changes occur in the brain or blood system AD starts in its preclinical phase. The disease develops for at least 20 years before symptoms become apparent. Individuals progress from Mild Cognitive Impairment (MCI) to dementia which causes substantial interference with their memory abilities and reasoning capacity alongside daily activities. Early diagnosis enables prompt treatment and superior management which may decelerate the progression of the disease [6].

Scientists use two different methods to detect Alzheimer's disease which are invasive and non-invasive procedures. Issue-based diagnostics involve accessing biological markers through medical procedures which include lumbar puncture combined with blood tests to detect disease biomarkers. The procedures for Alzheimer's detection frequently lead to discomfort and carry some level of health dangers. Non- invasive methods offer patients better safety combined with friendliness since they carry minimal risk to patients according to research [9]. Medical imaging technology incorporates X-rays as well as CT scans and MRI tests as its examples. Brain MRI demonstrates superior capabilities than CT imaging by performing better through dense bone structure avoidance and obtaining oblique and transverse slices that help view soft tissues best while precisely separating white and gray matter in the medial temporal lobe [10]. Neurologists can detect Alzheimer's disease more effectively through the use of recent CAD tools. CAD systems embrace traditional and modern deep learning approaches. The deep learning technology differs from standard multi- step pipelines because it needs minimal preprocessing while learning data features automatically without manual selection processes [11]. The medical imaging field mostly utilizes Convolutional Neural Networks (CNNs) Recurrent Neural Networks (RNNs) and Transformers as architectural frameworks to examine 2D and 3D MRI and ultrasound images. Full 3D-CNN models should not be used for 3D brain M I analysis because these images are constructed through stacking multiple 2D planes [12][13][14]. Weighted attention features of Vision Transformers (ViTs) have gained significant achievements in computer vision applications while emerging innovations apply them to medical imaging situations. A distinctive attribute of ViTs separates them from CNNs because their self-attention system effectively tracks extensive spatial relationships across data inputs [2].

Research evidence shows that Vision Transformers yield successful outcomes yet their major practical applications remain restricted to natural image evaluation datasets. The foundation architecture released initially for ImageNet based classification functions [5]. The study of Alzheimer's disease and other neurological disorders primarily uses MRI as their main imaging tool. T1-weighted and T2-weighted types of structural MRI scans serve as the widespread tools in brain research for examining tissues at both gray and white matter levels. The same requirement for deep learning models applies here since ViTs need large datasets to work optimally. Pre-trained and optimized models serve as solutions to lower the dependence on large datasets.

The primary objective of this work is:

- To classify Alzheimer's disease using brain MRI scans across three stages of progression.

II. LITERATURE REVIEW

The early identification of medical conditions has experienced promising results from deep learning techniques during the recent period. Science tools based on deep learning make important contributions to both early-stage diagnosis techniques and deeper knowledge expansion regarding Alzheimer's disease. The section presents important breakthroughs within this specific field.

[1] introduced a lightweight ViT architecture incorporating Multi-Head Self-Attention (LMHSA), Inverted Residual Units (IRU), and Local Feed-Forward Networks (LFFN) to reduce training costs. The approach was evaluated on the ADNI dataset, specifically tailored for binary classification between AD and CN. Expanding on this, [2] investigated the use of ViTs for Alzheimer's diagnosis using multi-modality PET scans from the ADNI dataset. This study compared CNNs and ViTs by converting 3D PET data into 2D slices to lower computational demands, demonstrating the effectiveness of ViTs in PET- based diagnosis.

The study by [3] proposes the use of Vision Transformers (ViTs) and a compact variant called Neuroimage Transformer (NiT) for detecting Alzheimer's disease. Utilizing the self- attention mechanism of ViTs, the research achieves competitive accuracy in AD diagnosis while reducing training time and resource usage by employing pre-trained models. The NiT model, a scaled-down version of ViT/16, is evaluated on both the ADNI and OASIS datasets for performance comparison across architectures. Taking a step further, [4] presents OViTAD—an optimized Vision Transformer model tailored for Alzheimer's prediction using resting-state fMRI and structural MRI scans. This work uses Mild Cognitive Impairment (MCI) as the basis for classification and performs two tasks: distinguishing MCI from AD + CN (Cognitively Normal) and differentiating AD + MCI from CN.

[5] proposes a novel technique that integrates Vision Transformers (ViT) with Bidirectional Long Short-Term Memory (Bi-LSTM) networks for diagnosing Alzheimer's disease using 3D MRI data. This hybrid model captures both spatial and temporal features from MRI scans, utilizing ViTs for spatial analysis and Bi-LSTM for sequential data classification. The study focuses solely on distinguishing between Cognitively Normal (CN) and Alzheimer's Disease (AD) cases.

[6] presents a convolution-based enhancement of Vision Transformer (ViT) models aimed at predicting the progression from Mild Cognitive Impairment (MCI). By optimizing patch extraction, the proposed ConViT model demonstrates strong diagnostic performance, highlighting the practical applicability of ViTs in Alzheimer's diagnosis. However, the model focuses exclusively on the MCI stage and does not cover other AD stages.

[7] proposes a hybrid model combining AlexNet and transfer learning to enhance both binary and multi-stage classification performance of Alzheimer's disease. The model processes both segmented and unsegmented MRI scans from the OASIS dataset and demonstrates effective results for multi-stage classification. In a separate study,

[8] [15] presents a Fusion Transformer architecture specifically designed for AD detection. By incorporating 2D coronal slices as input, this method improves diagnostic accuracy and robustness across varied patient groups. The approach integrates a Transformer with GFNet, training each model separately and merging their outputs to enhance performance on 16×16 image patches. Although this study focuses solely on 2D coronal slices, future work aims to extend the fusion technique to other imaging modalities.

III. METHODOLOGY

The team consistently followed the original ViT framework since it allowed precise assessment of transformer model advantages and ViT scalability. We selected the vit-base- patch16-224 model which consists of a ViT-B/16 variant with dropout set to 0.1 and patch size of 16, hidden size of 768 alongside 12 encoder layers and 12 self-attention heads. Training commenced with ImageNet-21k which contains 224×224-resolution images and subsequently the model received fine-tuning with ImageNet (ILSVRC2012). These pre-trained models display strong performance after minimal fine-tuning because they were trained on various types of image data. The architecture consists of 12 successive layers where each component advances the input embeddings through a process which discovers hierarchical features to create semantic representations with increasing depth. The combined layers generate MRI scan data which becomes suitable for classification needs.

We applied additional 224×224 resolution resizing to the preprocessed dataset images before continuation of our work. A random split divided the data into training (80%), validation (10%) and testing (10%) parts. The optimization was performed using Adam with a 0.001 learning rate. As part of its training process the ViT model acquires the capability to assign correct labels to images under supervision. The model seeks to optimize its parameters consisting of transformer layers and patch embeddings and classification head through the reduction of the cross- entropy loss function that calculates prediction versus actual label discrepancies. The optimization process implements Adam as its primary gradient-based method throughout training. The model learns to predict correctly by using an adequate amount of pre-labelled training data. Information entered the model through twenty complete passes before the conclusion of training. The studies took place on a MacBook equipped with an Apple M2 processor consisting of an 8-core CPU, 10-core GPU and 16GB unified memory.

Figure 1. Training vs Testing graph

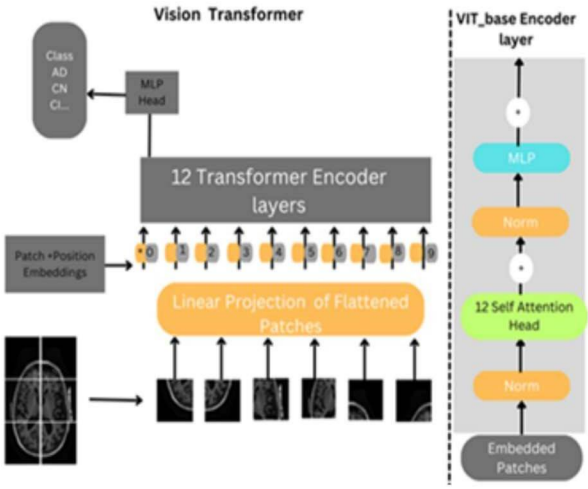


Figure 2. Architecture of Vision Transformer

The standard transformer layer operates with two fundamental components known as Multi-Layer Perceptron (MLP) and Multi-Head Self-Attention (MSA). The MSA mechanism divides its input into smaller units before performing scaled dot-product attention operations simultaneously on each segment. Attention output formation happens through the concatenation of output data from every attention head. The next stage of the processing chain is the MLP block that contains linear layers divided by the Gaussian Error Linear Unit (GeLU) activation function. The MSA and MLP modules rely on layer normalization together with residual connections which function similarly to skip connections to improve training stability and performance

A. *Vision Transformer Architecture*

The main goal of this research project is to use ViT Vision Transformers (ViT) for image classification tasks. ViT architecture builds upon the work of [17] which views images through a patch-based sequence structure to deploy Transformer processing methods and obtain results that sometimes best CNN approaches for big datasets. The ViT image processing method takes a 2D image through an initial flattening step before dividing it into rectangular 2D patches which match the format expected by Transformers that operate on one-dimensional token sequences. The trainable linear projection transforms the patch inputs into D-dimensional vectors. The ESL tool applies Multi-Head Self- Attention and Multi-Layer Perceptrons (MLPs) to process the patch embeddings for learning meaningful representations.

.

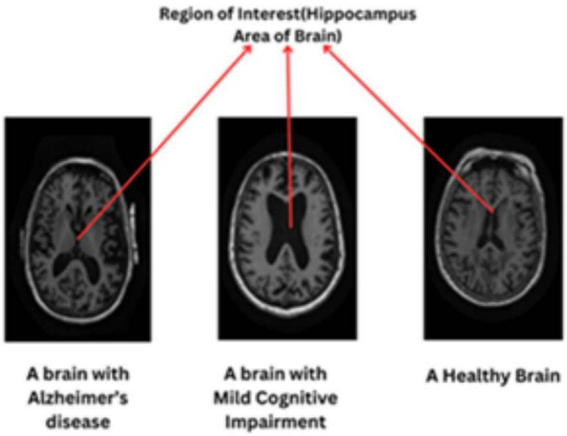


Figure 3. Samples from Dataset

B. *Dataset*

The study retrieved MRI images from ADNI and ADNI- Axial database collections which are components of the Alzheimer's Disease Neuroimaging Initiative (ADNI) project [16]. This research makes use of a collection of diagnostic instruments including positron emission tomography (PET) and clinical assessments together with biological markers and MRI scans to monitor Alzheimer's Disease (AD) and Cognitive Normal (CN) and Mild Cognitive Impairment (MCI). This initial dataset divides participants into five stages which include Cognitive Normal (CN), Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI) as well as Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). The second dataset includes three distinctive categories which are CN and CI and AD.

Images from all classes			
Dataset	Train Images	Test Images	Val Images
ADNI	1036	260	207
Axial	4123	1031	825

Feature extraction for classification purposes required an enhancement of the original MRI data from the ADNI database. The MRI data from the ADNI database received preprocessing procedures. The initial process separates crucial brain regions from other information through removal. extraneous information such as voxel coordinates and skull data. We conducted an extensive review of the dataset to gain better insight into its contents. The research performed an exploratory investigation which concentrated on the hippocampus region. The hippocampus region consists of grey matter structures in the human brain, critically associated with Alzheimer's disease.

IV. EVALUATION

Both datasets contain 1,296 MRI images which are divided into five classes and 5,154 images distributed among three classes. A first dataset holds 1,296 MRI images across five categories, and the second dataset shows 5,154 images distributed across three classes. For performance evaluation, we evaluated performance through accuracy as well as precision and F1 score, and AUC curve, and the confusion matrix.

Initially, the model was trained using 16 items as the batch amount. Further experimentation involved testing batch sizes of 8, 16, 32, and 64. In the case of the first dataset, the model began with an accuracy measurement of 50%. The accuracy began at 50% but reached an optimized level of 68%. The model performed better with batch size 16 while adjusting learning rate parameters. However, the small size of this dataset limited overall performance.

In contrast, the second dataset produced an 88% accuracy level. The ViT base model operated with pre- trained weights at a batch size of 16. Weights from ImageNet-1K demonstrated the best classification performance. Experimental results indicate that this particular batch size proved to be optimal for the model. The MRI image features were best extracted by this model configuration.

The model reached an F1 score of 0.81% together with a precision score of 0.81% in its performance. The model extracts essential brain information from the input data through its operation, targeting regions in accordance with the defined batch size during prediction. The performance of the model deteriorates when batch sizes are elevated.

The model becomes underfitted because of too small batch sizes, which results in both feature loss and model overgeneralization. Two different approaches emerge for selecting batch sizes. The reduction of image size when small leads to semantic details being lost from MRI images.

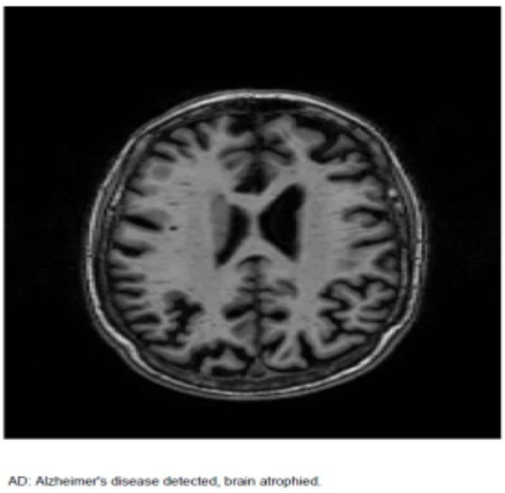


Figure 5. Example of generated PDF

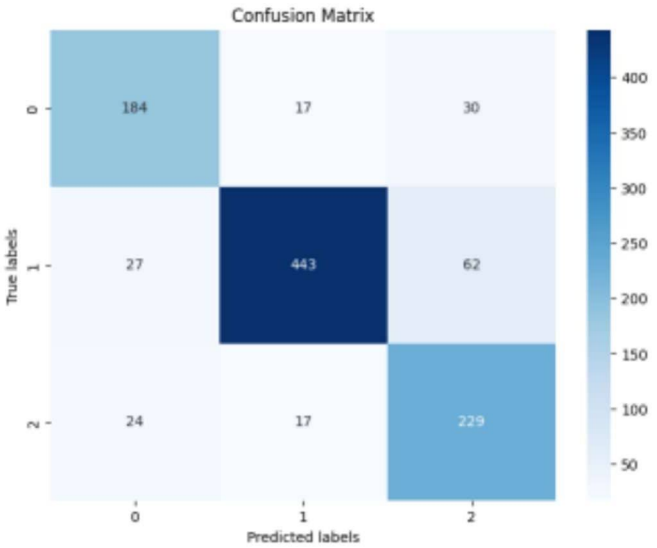


Figure 6. Confusion Matrix

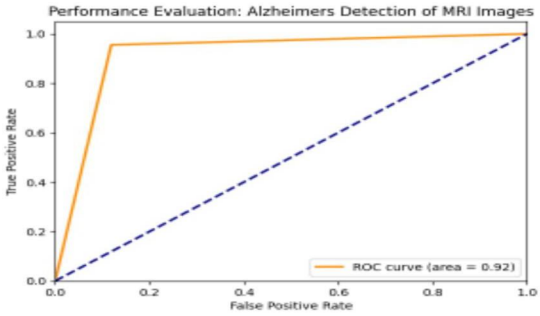


Figure 7. ROC Curve

V. DISCUSSION

Predicting Alzheimer's Disease (AD) with reliable accuracy is important because it helps in starting treatment early. That's why many studies have tried to improve how AD is diagnosed and classified. In this work, we carried out a comparison to see how well Vision Transformers (ViTs) perform. The focus was on how ViTs handle mid-sagittal MRI slices from the ADNI dataset. These MRI images were taken from the ADNI database. Our model reached 88% accuracy, with a Precision Score of 0.82 and an F1 Score of 0.82, which is better than the results from recent MRI- based classification methods. The results show that attention-based models like ViT can perform better than standard CNN models. With a larger set of 5,154 MRI images and more attention heads in the transformer encoder, the model performed best— achieving a ROC-AUC score of 0.92, shown in Figure 6.

Other studies have shown that ViT models rely heavily on tuning hyperparameters. Factors like batch size, learning rate schedules, warmups, and data augmentation played a big role. Comparison results are shown in Figures 7, 8, and 9.

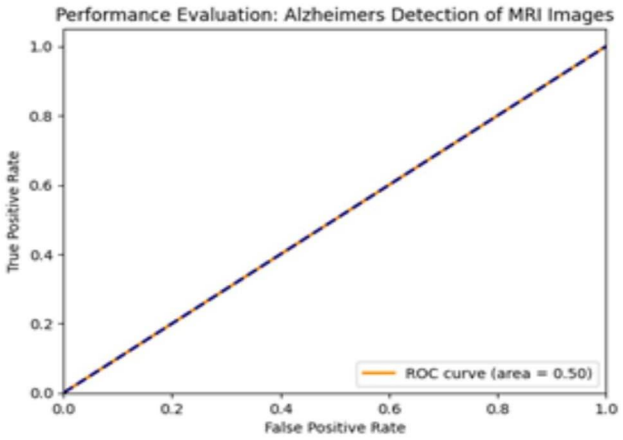


Figure 8. ROC curve with of ADNI dataset

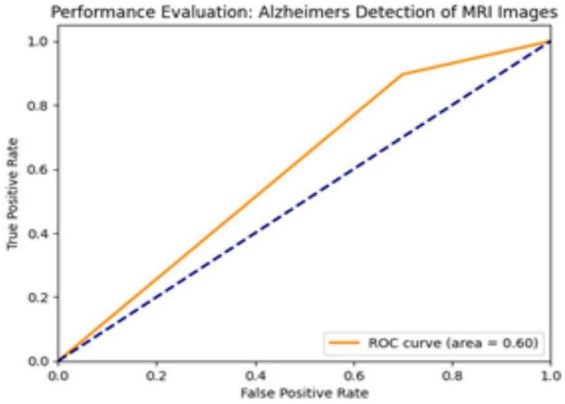


Figure 9. ROC curve of ADNI dataset after changing batch size

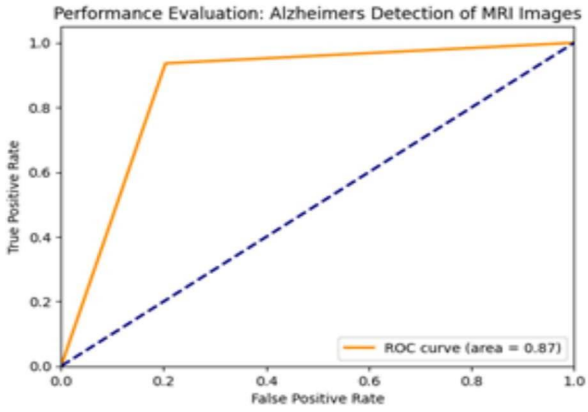


Figure 10. ROC curve of Axial Dataset

Using a patch size of 16 helped the model pick up better features. ViT-B gave the highest accuracy among all the tested versions. But with the smaller dataset, the more complex transformer models didn't perform as well.

The benchmark tests of our proposed methodology are compared with recent studies, as shown in Table 2. The table includes a summary of those previous works. Accuracy values along with Area Under the Curve (AUC) scores are the main points of comparison across these studies.

While earlier works documented their outcomes using AUC values, our method demonstrated strong performance using the ViT-B architecture. Previous studies, such as those by [1] and [3], explored different versions of ViT models and reported promising results. In contrast, our evaluation focused specifically on the ViT-B variant.

Tests confirmed that the ViT-B model can effectively detect Alzheimer's disease. For each sample evaluated, our approach generated classification outputs. A supplementary check on the test cases further validated the reliability of the results.

TABLE 1. COMPARISON WITH OTHER WORKS

Studies	Model	Acc	AUC
Hoang et al[1](2023)	Vit-B,Vit-S,Vit-L	72%	0.87
Maqssod et al(2019)[2]	CNN(ALexnet)	92%	0.95
Dhinagar et al(2023)[3]	Vit-B/Nit	79%	0.82
Kushol et al (2023)[4]	Fusion ViT	88%	0.92
Zing et al(2022)[15]	ADVIT(PET SCAN)	91%	0.95
Saraff et al(2023)[5]	OViTAD	88%	0.92
Akan et al(2024)[6]	ViT-Bi-LSTM	95%	0.97
Khatri et al(2024)[7]	Cvit	92%	0.94
Ours	Vit-B	88%	0.92

VI. CONCLUSION

The authors introduce a classification system designed to identify the stages of Alzheimer's disease by analyzing brain MRI images using Vision Transformers (ViTs). This research applies ViTs to leverage their capabilities for detecting these stages with notable accuracy. The approach demonstrated strong performance, achieving 88% accuracy and an AUC of 0.92 on the Axial dataset. To enhance interpretability, we also generated detailed PDF reports corresponding to each predicted class.

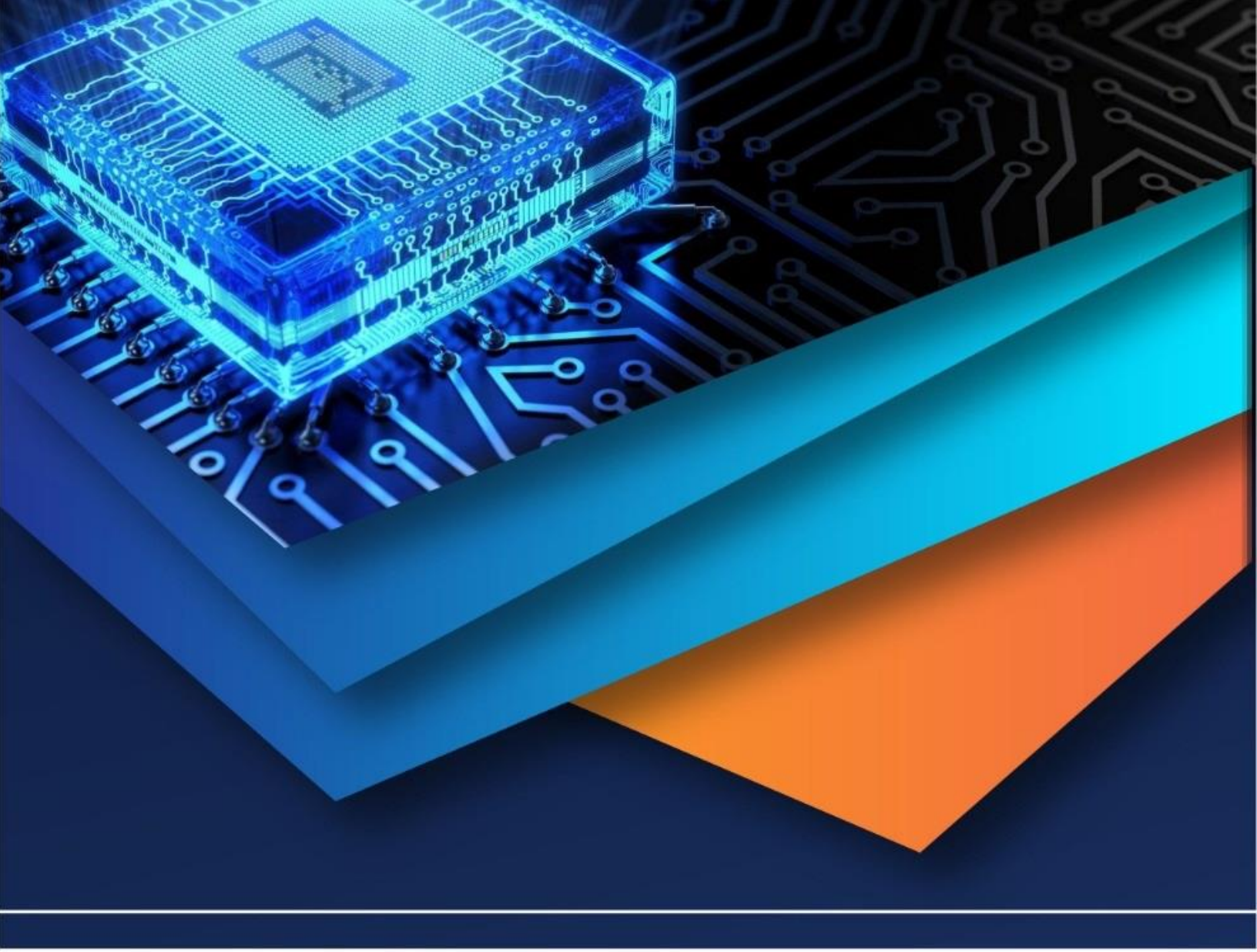
These findings highlight the diagnostic potential of Vision Transformers in identifying Alzheimer's disease. Supported by deep learning optimization techniques and attention maps, our method simplifies the understanding of key brain regions involved in the disease.

By integrating interpretive reports with classification outputs, the system offers valuable insights. Overall, this research presents an innovative approach to Alzheimer's diagnosis, showing promise as a computer-aided tool for improved clinical decisions and better patient outcomes.

REFERENCES

[1] Vision transformers for the prediction of mild cognitive impairment to Alzheimer's disease progression using mid-sagittal sMRI,Hoang, Gia Minh and Kim, Ue-Hwan and Kim, Jae Gwan,2023
[2] Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans, Maqsood, Muazzam and Nazir, Faria and Khan, Umair and Aadil, Farhan and Jamal, Habibullah and Mehmood, Irfan and Song, Oh-young, 2019
[3] Efficiently Training Vision Transformers on Structural MRI Scans for Alzheimer's Disease Detection, Dhinagar, Nikhil J and Thomopoulos, Sophia I and Laltoo, Emily and Thompson, Paul M,2023
[4] Addformer: Alzheimer's disease detection from structural mri using fusion transformer ,Kushol, Rafsanjany and Masoumzadeh, Abbas and Huo, Dong and Kalra, Sanjay and Yang, Yee-Hong ,2023
[5] OViTAD: Optimized vision transformer to predict various stages of Alzheimer's disease using resting-state fMRI and structural MRI data, Sarraf, Saman and Sarraf, Arman and DeSouza, Danielle D and Anderson, John AE and Kabia, Milton and Alzheimer's Disease Neuroimaging Initiative, 2023
[6] Vision Transformers and Bi-LSTM for Alzheimer's Disease Diagnosis from 3D MRI, Akan, Taymaz and Alp, Sait and Bhuiyanb, Mohammad AN, 2024
[7] Diagnosis of Alzheimer's disease via optimized lightweight convolution-attention and structural MRI, Khatri, Uttam and Kwon, Goo-Rak, 2024
[8] Deep learning for Alzheimer's disease diagnosis: A survey, 2022
[9] Gradient-based learning applied to document recognition, 1998

[10] Redmon, Joseph and Divvala, Santosh and Girshick, Ross and Farhadi, Ali, You Only Look Once: Unified, Real-Time Object Detection, 2016
[11] Deep learning based pipelines for Alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method, 2022
[12] Silva, Iago R. R. and Silva, Gabriela S. L. and de Souza, Rodrigo G. and dos Santos, Wellington P. and de A. Fagundes, Roberta A, Model Based on Deep Feature Extraction for Diagnosis of Alzheimer's Disease, 2019
[13] Zhao, Bendong and Lu, Huanzhang and Chen, Shangfeng and Liu, Junliang and Wu, Dongya, Convolutional neural networks for time series classification,2017
[14] Vaswani, Ashish and Shazeer, Noam and Parmar, Niki and Uszkoreit, Jakob and Jones, Llion and Gomez, Aidan N and Kaiser, Łukasz and Polosukhin, Illia, Attention is All you Need, 2017
[15] Advit: Vision transformer on multi-modality pet images for alzheimer disease diagnosis, Xing, Xin and Liang, Gongbo and Zhang, Yu and Khanal, Subash and Lin, Ai-Ling and Jacobs, Nathan, 2022
[16] https://ida.loni.usc.edu/login.jsp?project=ADNI
[17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. (2020).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY