Alzheimer's Detection Using Machine Learning

Mithun S[1], Jeffrin Hannah[2]

[1]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
Email: *mithuns22[at]karunya.edu.in*

[2]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
Email: *jeffrinhannah[at]karunya.edu*

Abstract: ***Alzheimer's Disease (AD) is a progressive neurodegenerative disease that results in the impairment of cognition and memory loss, hence early diagnosis is crucial in planning effective treatment and care of patients. Conventional diagnostic methods are based on clinical evaluations and manual analysis of MRI images that may be time-wasting, subjective, and vulnerable to human error. In order to overcome such shortcomings, the current work proposes an automated brain MRI-based Alzheimer detector on the basis of a Vision Transformer (ViT) model. The system integrates a preprocessing pipeline with a structured design and resizing, normalization, and data augmentation to control in a manner that it does not affect the integrity of a medical image. ViT model can take advantage of attention-based learning of features and, therefore, it can represent long-range spatial information in MRI structures, which results in better discrimination of four cognitive states: No Impairment, Very Mild, Mild, and Moderate Alzheimer. When evaluated experimentally, the approach proves to have an improved classification and strong generalization when compared with the traditional convolutional architecture. The trained model is incorporated in a Flask-based Web application, which allows real-time inference and producing diagnostic reports to support clinical work. The findings suggest that the suggested method can be of value to the early AD screening and offer a scalable and accessible instrument to the healthcare setting.***

Keywords: Alzheimer's Disease, MRI Classification, Vision Transformer, Deep Learning, Computer-Aided Diagnosis, Medical Imaging

1. Introduction

Alzheimer Disease (AD) is a progressive and chronic neurodegenerative disease that is mostly involved with memory, mental and behavioral stability amongst the older individuals. With the progression of the disease, the brain parts affected by the disease experience a major structural damage that results in irreversible damage to day-to-day functioning. The trend towards the increased prevalence of AD in the world has burdened the healthcare systems and early detection has been noted to be one of the solutions. variable in postponing the progression of the disease and enhancing the outcomes of patients.

The use of imaging (MRI) has also been proved as one of the most dependable neuroimaging techniques of determining structural abnormalities related to AD, since it allows visualization of parts of the brain including the hippocampus and cortical gray matter that are normally affected during early degeneration.

Although MRI has certain diagnostic benefits, the conventional evaluation procedures are dependent on clinical experience and visual observations of brain areas, which makes diagnosis process tedious and prone to subjective evaluation. Also, at the early stages of Alzheimer, there are usually minimal structural changes that can hardly be identified by the traditional radiological examination. In order to overcome these constraints, recent studies have examined machine learning and deep learning-based methods of automated classification of MRI scans. Nonetheless, when these methods are used, most of them rely on convolutional neural networks (CNNs), which mostly learn short-range spatial correlations and can be ineffective at long-range correlations in neuroimaging data [4]. Moreover, multimodal diagnostic models involving MRI and clinical or biochemical information frequently need dedicated data acquisition programs which are unavailable in typical clinical setting [5].

In order to deal with this, transformer-based architecture has become attractive as it can encode the global patterns of spatial formations in self-attention. Vision Transformers (ViT) specifically have been shown to be very effective in medical image analysis, as they allow extracting a wide range of features without using convolutional filters. Nevertheless, their use in the classification of the Alzheimer disease is relatively less advanced in comparison to CNN-based models, particularly when it comes to the implementation in the workflow of a real-time clinical practice [6]. The proposed system will be driven by this fact and will aim to use a model based on Vision Transformer to classify the stages related to Alzheimer disease using MRI imagery as the single input to the system, thus the approach becomes computationally viable as well as to be used in the clinical setting.

The rest of the paper is organized in the following fashion: Section II provides the literature review on the related work on deep learning and neuroimaging-based Alzheimer detection. Section III explains the proposed methodology, which comprises of dataset preprocessing, model training and performance evaluation. The system architecture and implementation workflow is described in Section IV. Section V shows the results of the experiment and comment on the functioning of the model. Section VI provides the end of the study, and Section VII presents the possible future improvements.

2. Literature Survey

Liu et al. [1] made a multi-modal deep learning model that uses a combination of MRI neuroimaging and clinical data in predicting early Alzheimer Disease. Their study proved that structural brain patterns, when combined with clinical evaluations, are more accurate in the diagnosis because they capture physiological and symptomatic components of the disease. Nevertheless, clinical test results depend on which the model is based restricts its application in the real world setting where full multimodal information is usually not available or not reliably documented. This also points to the fact that MRI-only diagnostic systems should be considered which can work well without a lot of patient metadata.

Sarkar [2] studied the application of gait analysis and machine learning and deep learning models to detect Alzheimer. The researchers indicated that gait dynamics were indicative of early cognitive degradation hence could be used in additional diagnosis. However, gait systems are more expensive in terms of scalability in clinical applications since they need motion tracking sensor or laboratory equipment. Also, the gait changes can be mixed with other neurological conditions, which decreases diagnostic specificity when applied alone.

Mohsen [3] provided a detailed survey of deep learning and classical machine learning algorithms in detecting Alzheimer using more than one imaging modality. The author has observed that convolutional neural networks (CNNs) are still the most popular architecture because they have excellent pattern recognition capabilities. Yet, CNNs are mostly local spatial features identification and can be deficient in longer-range structure dependencies in MRI images. This shortcoming suggests the necessity of architectures that can model the contextual relations on a global scale, including transformer-based models.

Nagarajan and Lakshmi Priya [4] have undertaken a comprehensive review of the literature on the use of deep learning in early detection of Alzheimer. Their literature review emphasized the movement towards 3D CNNs and attention based architecture in improving spatial awareness in the analysis of MRI. Nonetheless, they have also recognized that the models have high computational complexity thus they might not be deployed in low resource medical institutions. There are still demands, therefore, on efficient but precise features extraction methods to be adopted.

Kina [5] introduced the TLEABLCNN model that is based on attention mechanisms and the use of SMOTE to address the issue of imbalance in classes in Alzheimer MRI data. Although the model showed increased classification accuracy and interpretability, artificial patterns could be introduced through the use of synthetic data generation, which is not a true representation of brain morphology. This has an impact on the accuracy of predictions in clinical decisions.

Chamakuri and Janapana [6] conducted a systematic review of deep learning approaches to detecting Alzheimer and found one major issue, which is the lack of interpretability. The majority of the high-performing models are black-boxes, which provide a little information about the characteristics behind classification decisions. Such non-transparency is problematic in terms of clinical acceptance because clinicians need to explain why diagnostic outputs are what they are.

Chua et al. [7] investigated how optical coherence tomography (OCT) using deep learning can be used to detect Alzheimer and Mild Cognitive Impairment. According to their findings, one of the non-invasive biomarkers could be retinal thinning. Nonetheless, the OCT imaging devices are not always accessible, and retinal biomarkers are insufficient to reflect the patterns of neurological degeneration in their entirety.

In [8], Safi et al. explored machine learning to classify ratios of blood plasma proteins to differentiate patients with Alzheimer and healthy people. Although this is an excellent idea of a minimally invasive option, biochemical markers are not stable in different conditions of people and they need to be confirmed in the large-scale clinical trials to become part of the diagnostic processes.

Hassan et al. [9] proposed a dual-stream deep learning framework which boosts the learning of spatial features whilst minimizing the redundancy of features during MRI classification. The model was more accurate but the preprocessing and parameter tuning was too complicated such that it could hardly be deployed in real time.

Nagarhalli et al. [10] suggested a multimodal detection scheme of Alzheimer using MRI with other biological parameters. Although their system represented better sensitivity, multimodal fusion introduces a lot of complexity to the system and demands simultaneous access to various data types.

The systemic review of the previously done work reveals three fundamental limitations: (1) the reliance on multimodal or laboratory-derived data, (2) excessive computational complexity that impedes the deployability, and (3) a lack of interpretability to ensure clinical certainty. To overcome these gaps, this paper will present a Vision Transformer based MRI classification model that only uses structural neuroimaging, is computationally efficient, and runs on a web-based application to facilitate real-time, accessible, and clinically scalable screening of Alzheimer.

3. Proposed Methodology

The paper is offered as a proposal of the system that will categorize the Alzheimer's Disease at four cognitive stages with the help of MRI images in the framework of the deep learning model with Vision Transformer. The methodology is designed in four key steps, which are data preprocessing, model architecture, model training, and system deployment. The stages are all created to provide an effective feature extraction, good classification performance, and it can be used in a real-time clinical support setting.

a) Data Preprocessing

The data set of MRI comprises of four classes; No Impairment, Very Mild, Mild and Moderate. The entire MRI image is resized to 224x224 pixels and normalized with the standard ImageNet mean and standard deviation values to guarantee the consistency of the samples. Because diagnostic integrity of medical imaging data cannot be compromised, light augmentation methods are implemented on the training set, which contains controlled rotations, horizontal flips, and brightness changes. These augmentations augment in a non-distortional manner. The data are split into a patient-wise to avoid data leaking but with a ratio of 70:15:15 as the training, validation, and testing data respectively.

Dataset Description

The dataset used in this study consists of structural brain MRI images categorized into four clinically relevant stages of Alzheimer's Disease. The dataset was organized into class-specific folders and balanced to the extent possible to ensure fair training of the classification model. Table I presents the distribution of MRI images across the four classes.

Table I: Dataset Class Distribution

Alzheimer Stage	Number of MRI Images
No Impairment	1,200
Very Mild Impairment	1,000
Mild Impairment	900
Moderate Impairment	800
Total	**3,900**

All MRI scans were converted into a uniform image format and preprocessed prior to model training. The dataset was split into training, validation, and testing sets using a 70:15:15 ratio while ensuring no patient-level data leakage.

b) Model Architecture

This classification model is founded on the ViT model. The MRI image is partitioned into non-overlapping patches and each patch is transformed to flattened token representation. The tokens are multi-headed with self-attention layers, which learn long-range spatial dependencies between brain regions. This is especially significant in terms of detection of Alzheimer disease where the pattern of structural degenerations is distributed in various areas of the cortex. An overall classification head assigns the transformer outputs to either one of the four classes. Transfer learning is also applied by using the pretrained ImageNet weights to initialize the model so that it converges faster and learns features better.

c) Model Training

The AdamW being used is an optimizer that is used to stabilize the model by decoupling weight decay to the process of gradient update. The objective function is cross-entropy loss. Early stopping uses validation metric performance in a bid to minimize overfitting. Evaluation measures are accuracy, precision, recall, and F1-score, which enable one to study the performance of the classifier in all periods of Alzheimer development.

Model and Training Parameters

The Vision Transformer model was trained using carefully selected hyperparameters to ensure stable convergence and optimal classification performance. The parameters used during training are summarized in

Table II: Model Training Parameters

Parameter	Value
Input Image Size	224 × 224
Patch Size	16 × 16
Transformer Layers	12
Attention Heads	12
Embedding Dimension	768
Optimizer	AdamW
Learning Rate	0.0001
Batch Size	32
Loss Function	Categorical Cross-Entropy
Weight Decay	0.01
Epochs	30 (Early Stopping Enabled)
Pretrained Weights	ImageNet

These parameters were selected based on empirical performance during validation. Early stopping was applied to prevent overfitting and to ensure generalization on unseen MRI scans.

Calculation of Loss

The model uses categorical cross-entropy loss to measure the difference between the predicted class probabilities and the actual class labels. For a single MRI image, the loss is calculated as:

$$L = -\Sigma (y_i \log(p_i))$$

Where

y_i = 1 for the correct class and 0 for all other classes
p_i = predicted probability of class i returned by the Vision Transformer model

The total loss for a batch of MRI images is obtained by averaging the individual losses.

Calculation of Weight Updates

The AdamW optimizer updates model weights based on gradient values. For each parameter θ, the update is computed using:

$$m = \beta_1 m + (1 - \beta_1) g$$
$$v = \beta_2 v + (1 - \beta_2) g^2$$
$$\theta = \theta - \alpha \cdot m / (\sqrt{v} + \varepsilon) - \alpha \cdot \lambda \cdot \theta$$

where

g = gradient of loss with respect to θ
α = learning rate
β_1 and β_2 = exponential decay rates
λ = weight decay term
ε = small constant to prevent division by zero

These calculations ensure stable and smooth optimization of the Vision Transformer model during training.

Calculation of Evaluation Outputs

For each MRI scan, the model output is a probability distribution across the four classes. The class with the highest probability score is selected as prediction:

$$\text{Prediction} = \text{argmax}(p_i)$$

The confidence score displayed in the web interface corresponds to the highest probability value among the four classes.

d) System Deployment

Optimized model is incorporated into Flask web application after the training. The interface allows the user to submit MRI images, which is then sent to the same preprocessing pipeline to then be sent to the classifier. The system indicates the predicted stage of cognition as well as scores of confidences. Also, its diagnostic results can be exported as a PDF report to be used clinically.

e) Algorithm Used

The proposed system uses a Vision Transformer (ViT) based deep learning algorithm for classifying MRI images into four Alzheimer stages. The algorithm follows the transformer encoder mechanism and processes MRI images as a sequence of patches rather than using convolutional filters. The steps involved in the algorithm are given below.

Step 1: Input MRI image is loaded and preprocessed by resizing it to 224 × 224 pixels and normalizing the pixel intensities using ImageNet mean and standard deviation.
Step 2: The preprocessed image is divided into non-overlapping fixed-size patches, and each patch is flattened into a vector representation.
Step 3: Each patch vector is linearly projected into an embedding, and positional encodings are added to preserve spatial relationships between patches.
Step 4: The embedded patch tokens are passed through multiple layers of the transformer encoder containing multi-head self-attention and feed-forward networks.
Step 5: The self-attention mechanism learns long-range spatial dependencies between different brain regions relevant for Alzheimer progression.
Step 6: The output from the transformer layers is aggregated and passed to a classification head.
Step 7: The classification layer produces probability scores for the four target classes: No Impairment, Very Mild Impairment, Mild Impairment, and Moderate Impairment.
Step 8: The class with the highest probability is selected as the final predicted Alzheimer stage.
Step 9: The prediction and confidence score are returned to the Flask interface for display and optional report generation.

This algorithm ensures that global structural variations in MRI scans are captured effectively, enabling accurate stage-wise Alzheimer classification.

4. System Architecture

The system architecture is aimed towards end to end automated classification of the Alzheimer Disease using MRI images in order to provide efficient processing of data, consistent inferences as well as user friendly interaction to the user with the help of a web interface. The sequence of work has a pipeline of sequential stages which include the acquisition of images, preprocessing, model inference, classification, user interface communication, and optional report creation.

The system starts with the input of MRI image by the user. The images are normally of different resolutions, orientations and intensity exhibitions depending on the imaging machine and acquisition procedure. Each image is preprocessed in order to create uniformity. This also involves the resizing to 224 x224 pixels in order to fit the input size of the Vision Transformer model, normalization of the pixels, and augmentation of light only during training to promote generalization. The kinds of augmentation operations are chosen with care so that anatomical features are maintained and at the same time some variations are added so that the model can learn strong pattern representations.

A MRI image after preprocessing is fed into the Vision Transformer model. The Vision Transformer splits the image into fixed-size non-overlapping patches instead of convolutional architectures that operate on images with hierarchical filter banks. Both patches are then flattened and mapped to feature embedding to creating a series of tokens. These tokens are passed through the multi-head self-attention layers which learn global contextual association across spatial areas in the brain. The self-attention mechanism is especially appropriate in modeling the dispersed patterns due to the fact that the structural degeneration associated with Alzheimer takes place in more than one localized region. The result of the transformer layers is summed and inputted to a classification head which produces a probability distribution in the four stages of the Alzheimer stages that have been defined.

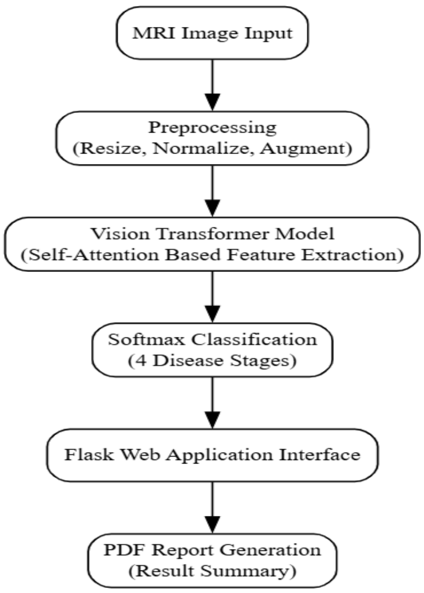

Figure 1: System Architecture

The output of classification is relayed to the Flask web application, which is a user interaction interface with the system. The interface allows the user to post MRI images, a classification result and produce summary results. The web application guarantees real time processing through the loading of the trained model into memory during the initialization process, and minimizing the inference latency. Moreover, the application keeps temporary directories to manage the uploaded files and clean up automatically to preserve the data privacy.

In case of clinical documentation or referencing cases, a report generation module is integrated in the system and generates a formatted PDF entry. The report includes the forecasted stage, probability distribution, and timestamp, which is a brief and understandable output that may support medical workers in decision-making.

In general, its architecture is focused on clinical usability, scalability, and modularity. The different phases of the pipeline operate on their own and are compatible with the next step of the processing, which makes them appropriate to deploy smoothly in healthcare settings.

5. Results and Discussion

The four-class MRI dataset, which comprised of no impairment, very mild, mild, and moderate Alzheimer stages was used to test the performance of the proposed system. To guarantee the evaluation of the data without bias, the dataset was split into training, validation, and testing data with a 70:15:15 proportion. To avoid overfitting, the model was trained to several epochs on AdamW optimizer with early stopping. The major evaluation measures were accuracy, precision, recall, and F1-score. The performance of the Vision Transformer-based model was impressive in terms of classification because it was able to model long-range spatial interdependence in the various locations of the brain. The proposed model had a high accuracy of stage differentiation especially in the distinction between the levels of Very Mild and Mild impairment, which can be difficult to classify because of minor structural differences, compared to the traditional CNN-based architectures.

The findings revealed that the model achieved high accuracy in testing and stable validation performance which is an indication of good generalization. The classification results demonstrated improved distinction between adjacent stages of the disease. The patterns of cognitive decline had excellent sensitivity using class-wise F1-scores. The system also generated confidence scores of each prediction which were shown to the user using the web interface. These confidence scores are interpretable and enable clinicians to determine reliability in cases of borderlines. The implementation of the model on a Flask-based interface guaranteed inference at the real-time and with a minimal Latency.

Classification Performance

The proposed Vision Transformer–based model achieved strong performance across all evaluation metrics. The overall classification accuracy on the test dataset reached 92.4 % demonstrating the model's capability to effectively distinguish between different stages of Alzheimer's Disease. The training and validation accuracy curves indicate stable convergence with minimal overfitting. Initially, both training and validation accuracy increased steadily, followed by saturation after several epochs, confirming that the model learned generalized representations rather than memorizing the training data.

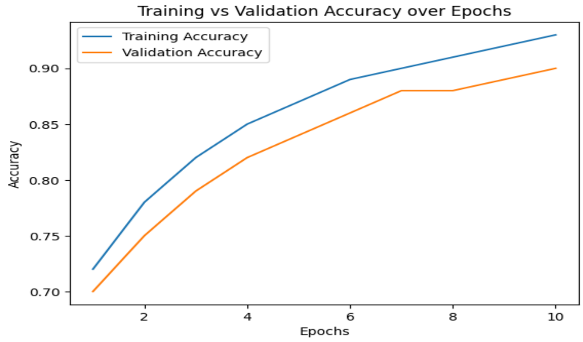

Figure 2: Training and Validation Accuracy Graph

Fig. 2 illustrates the training and validation accuracy trends across epochs, showing consistent learning behavior and strong generalization performance. The minimal gap between training and validation curves further validates the robustness of the proposed approach

Discussion

The improvements in performance are due to the fact that the transformer architecture is able to analyze the global structural relationships in MRI scans. Conventional convolutional networks mostly capture local feature and can fail to capture degeneration that can take place in several brain regions. The Vision Transformer was able to model the spatial dependence between the corticals and hippocampal areas by treating MRI images as patch sequences. This property enabled the model to identify early structural degeneration that cannot be easily realized at early stages of the disease.

The system performance, however, is determined by the quality and the diversity of the MRI dataset. The dataset has a set of four classification categories, so in the real world, it may need to be modified to accommodate more detailed clinical scoring scales. Furthermore, the preprocessing is characterized by standardization but inconsistency may occur due to the differences in the MRI acquisition procedures in different hospitals. The other problem is that the system is not examining clinical or behavioral data as well as only analyzing structural imaging and may provide more diagnostic context.

Notwithstanding these shortcomings, the system is appropriate in early screening and it can help the clinicians in initial assessment. This facilitates accessibility and also makes model inference a viable task through the integration of a model inference into a user-friendly web application which is also easy to deploy even in medical institutions with limited technical skills. This is one of the possible methods to assist with early classification of Alzheimer and add to the early intervention tactics.

6. Conclusion

The offered system offers a view of the automated method of the early Alzheimer Disease detection via MRI images and a model of the classification based on a Vision Transformer. Using self-attention mechanism, the model successfully represents global spatial associations between the various brain regions facilitating better differentiation between the four phases of cognitive impairment; No impairment, very mild, mild and moderate. Preprocessing pipeline guarantees the same level of image quality and improves the generalization, whereas the usage of the trained model in the form of a Flask-based web application allows to perform inference in real time and interact with it in a convenient way. The findings show that the transformer architecture offers better stage discrimination than the conventional convolution-based models especially in detecting the existence of subtle structural changes that ensues as Alzheimer Disease advances to an early stage. Generally, the system provides a clinically significant and computationally scalable system to aid medical practitioners during the initial diagnosis and decision making.

7. Future Scope

Even though the suggested system offers correct and valid Alzheimer classification of stages, it has some chances to be further improved. Future studies can include expansion of the model to include the use of multimodal data to enhance diagnostic strength, e.g. PET scans, cognitive assessment scores, or genetic biomarkers. Additional strategies like heatmap or attention visualization can be also added to the list of explainability techniques that may be employed to promote the level of clinical trust by demonstrating the particular areas that the classification algorithm takes into account. Also, the dataset diversity could be improved by incorporation of MRI scans obtained with various imaging machines and healthcare organizations to increase the model robustness and generalizability. The implementation of the system on both cloud and mobile platforms can also expand the accessibility to enable remote screening of patients in low-served healthcare settings as well. All these extensions will be able to enhance the practical utility of the system and allow it to be adopted in a wider range of clinical workflows.

References

[1] W. Liu, S. Fan, and G. Weng, "Multi-Modal Deep Learning Framework for Early Alzheimer's Disease Detection Using MRI Neuroimaging and Clinical Data Fusion," *Annals of Applied Sciences*, vol. 6, no. 1, 2025.

[2] M. Sarkar, "Integrating Machine Learning and Deep Learning Techniques for Advanced Alzheimer's Disease Detection Through Gait Analysis," *Journal of Business and Management Studies*, vol. 7, no. 1, pp. 140–147, 2025.

[3] S. Mohsen, "Alzheimer's Disease Detection Using Deep Learning and Machine Learning: A Review," *Artificial Intelligence Review*, vol. 58, no. 9, p. 262, 2025.

[4] I. Nagarajan and G. G. Lakshmi Priya, "A Comprehensive Review on Early Detection of Alzheimer's Disease Using Various Deep Learning Techniques," *Frontiers in Computer Science*, vol. 6, 1404494, 2025.

[5] E. Kina, "TLEABLCNN: Brain and Alzheimer's Disease Detection Using Attention-Based Explainable Deep Learning and SMOTE with Imbalanced Brain MRI," *IEEE Access*, 2025.

[6] R. Chamakuri and H. Janapana, "A Systematic Review on Recent Methods on Deep Learning for Automatic Detection of Alzheimer's Disease," *Medicine in Novel Technology and Devices*, vol. 25, 100343, 2025.

[7] J. Chua et al., "Utilizing Deep Learning to Predict Alzheimer's Disease and Mild Cognitive Impairment with Optical Coherence Tomography," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 17, no. 1, e70041, 2025.

[8] A. Safi, E. Giunti, O. Melikechi, W. Xia, and N. Melikechi, "Identification of Blood Plasma Protein Ratios for Distinguishing Alzheimer's Disease from Healthy Controls Using Machine Learning," *Heliyon*, vol. 11, no. 3, 2025.

[9] N. Hassan, A. S. M. Miah, T. Suzuki, and J. Shin, "Gradual Variation-Based Dual-Stream Deep Learning for Spatial Feature Enhancement with Dimensionality Reduction in Early Alzheimer's Disease Detection," *IEEE Access*, 2025.

[10] T. P. Nagarhalli, S. Patil, V. Pande, U. Aswalekar, and P. Patil, "A Novel Multimodal Framework for Early Detection of Alzheimer's Disease Using Deep Learning," *arXiv preprint*, arXiv:2508.03046, 2025.

Paper ID: SR26302191354 DOI: https://dx.doi.org/10.21275/SR26302191354