http://doi.org/10.35784/iapgos.6937

received: 12.12.2024 | revised: 21.08.2025 | accepted: 07.09.2025 | available online: 30.09.2025

ALZHEIMER'S DISEASE CLASSIFICATION FROM MRI USING VISION TRANSFORMER

Mohith Reddy Kandi, Sree Vijaya Lakshmi Kothapalli, Sivamsh Pavan Rajanala, Suvarna Vani Koneru, Vishnu Pramukh Vattikunta

Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India

Abstract. Alzheimer's disease (AD) is a progressive neurodegenerative disorder that presents significant challenges for early diagnosis and intervention. Traditional approaches for diagnosing AD using MRI images are labor-intensive and often subjective, resulting in the need for automated, accurate solutions to support clinicians in early-stage detection. This study investigates the use of vision transformer (ViT) for the classification of Alzheimer's disease stages using MRI images. By treating MRI images as sequences of tokens, ViT models capture both global and local spatial dependencies, which enhances their ability to recognize structural brain changes characteristic of AD. The model was trained on a diverse dataset containing four AD categories – Moderate Demented, Mild Demented, Very Mild Demented, and Non-Demented – achieving an overall classification accuracy of 98.9%. This result highlights the efficacy of transformer-based models in distinguishing between subtle structural brain alterations. Future directions for this study include fine-tuning the model on larger datasets and exploring the integration of multi-modal data to further support AD diagnosis and treatment strategies. The findings indicate that vision transformer have the potential to transform diagnostic imaging for neurodegenerative disorders by providing a robust, scalable, and precise tool for early AD detection.

Keywords: Alzheimer's disease, classification, dementia stages, magnetic resonance imaging, neurodegenerative disorders, vision transformer

KLASYFIKACJA CHOROBY ALZHEIMERA NA PODSTAWIE MRI PRZY UŻYCIU TRANSFORMERA WIZYJNEGO

Streszczenie. Choroba Alzheimera (AD) jest postępującą chorobą neurodegeneracyjną, która stanowi poważne wyzwanie dla wczesnej diagnostyki i interwencji. Tradycyjne metody diagnozowania AD przy użyciu obrazów MRI są pracochłonne i często subiektywne, co powoduje potrzebę stosowania zautomatyzowanych, dokładnych rozwiązań wspierających lekarzy w wykrywaniu choroby we wczesnym stadium. Niniejsze badanie dotyczy wykorzystania transformera wizyjnego (ViT) do klasyfikacji stadiów choroby Alzheimera na podstawie obrazów MRI. Traktując obrazy MRI jako sekwencje tokenów, modele ViT wychwytują zarówno globalne, jak i lokalne zależności przestrzenne, co zwiększa ich zdolność do rozpoznawania zmian strukturalnych w mózgu charakterystycznych dla AD. Model został wytrenowany na zróżnicowanym zbiorze danych zawierającym cztery kategorie AD – umiarkowaną demencję, łagodną demencję, bardzo łagodną demencję i brak demencji – osiągając ogólną dokładność klasyfikacji na poziomie 98,9%. Wynik ten podkreśla skuteczność modeli opartych na transformerach w rozróżnianiu subtelnych zmian strukturalnych w mózgu. Przyszłe kierunki badań obejmują dostosowanie modelu do większych zbiorów danych oraz zbadanie możliwości integracji danych multimodalnych w celu dalszego wsparcia diagnostyki i strategii leczenia choroby Alzheimera. Wyniki wskazują, że transformer wizyjny ma potencjał, aby zrewolucjonizować diagnostykę obrazową zaburzeń neurodegeneracyjnych, zapewniając solidne, skalowalne i precyzyjne narzędzie do wczesnego wykrywania choroby Alzheimera.

Słowa kluczowe: choroba Alzheimera, klasyfikacja, stadia demencji, obrazowanie metodą rezonansu magnetycznego, zaburzenia neurodegeneracyjne, transformer wizyjny

Introduction

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders globally, characterized by progressive cognitive decline, memory impairment, and structural brain changes that significantly impact daily life and autonomy in affected individuals. The World Health Organization (WHO) reports that AD and other forms of dementia affect over 55 million people worldwide, with an estimated increase to 139 million cases by 2050 due to aging populations [19]. Early diagnosis is crucial as it can facilitate timely intervention, management, and potential slowing of disease progression, thereby improving patient outcomes and quality of life [20].

One of the primary methods for assessing brain structural changes in AD is Magnetic Resonance Imaging (MRI), a non-invasive imaging technique that provides high-resolution anatomical details and tissue contrast, making it a valuable tool for diagnosing and monitoring AD. MRI allows for the visualization of specific structural changes associated with AD, including hippocampal atrophy and cortical thinning in regions such as the temporal, parietal, and frontal lobes [4]. These structural markers are correlated with cognitive decline and serve as essential indicators of disease progression, making MRI a key component in the clinical evaluation of AD [6]. Traditional methods for interpreting MRI data rely on manual analysis by radiologists and clinicians, which can be time-consuming, subjective, and susceptible to inter-rater variability. Machine learning techniques, particularly deep learning models, have shown promise in automating the classification of MRI images for AD diagnosis, offering objective and reproducible results. Convolutional Neural Networks (CNNs) have been the dominant architecture in computer vision tasks, and studies have successfully applied CNNs to AD diagnosis, achieving notable classification accuracy [11, 13]. However, CNNs have limitations in capturing long-range dependencies and global context due to their reliance on local receptive fields [6].

Recent advances in deep learning have introduced transformer-based models, particularly vision transformer, which address some limitations of CNNs by treating images as sequences of patches and leveraging self-attention mechanisms to capture spatial dependencies across the entire image [8]. The vision transformer model, initially proposed by Dosovitskiy et al. [3], has demonstrated impressive performance in various vision tasks, such as image classification, object detection, and segmentation. Unlike traditional CNNs, ViT models can process image data holistically, capturing both global and local context, which is particularly beneficial in medical imaging where subtle spatial patterns are often indicative of disease [2]. This study explores the application of vision transformer for the classification of AD stages using MRI images. By leveraging the unique capabilities of ViT, our research aims to improve classification accuracy and provide a scalable, automated solution for AD diagnosis. We trained and evaluated the ViT model on a dataset of MRI images from individuals at various AD stages, including moderate demented, mild demented, very mild demented, and non-demented. The primary objectives of this study include assessing the ViT model's performance in accurately classifying these stages and comparing its efficacy against traditional CNN-based models. The findings of this study underscore the potential of transformer-based architectures as effective tools in the early diagnosis and management of Alzheimer's disease.

1. Literature review

Almadhoun and Abu-Naser [1] utilized pre-trained Convolutional Neural Networks (CNNs), specifically the Xception model, to classify AD stages. Their approach achieved a notable classification accuracy of 96.86% across four stages of AD using MRI data sourced from ADNI and OASIS datasets. This study highlighted the utility of CNN-based models in reducing training time by leveraging pre-trained weights. However, the reliance on pre-trained models posed limitations in adaptability, particularly for datasets with unique characteristics specific to AD imaging.

Pradhan et al. [10] investigated deep learning architectures VGG19 and DenseNet169 to classify AD stages in MRI images, achieving classification accuracies of 86% and 88%, respectively. Their study demonstrated that established CNN architectures could yield robust results in medical imaging. Nevertheless, the study was limited by a lack of interpretability, with minimal insights into the decision-making process of the deep learning models. Additionally, their reliance on a relatively small dataset (approximately 6,000 images) may restrict generalizability to larger populations.

Samhan et al. [12] introduced a CNN-based classification model utilizing the VGG16 architecture. This model achieved a high accuracy of 97% on a dataset of 10,432 MRI images, classified into four AD stages. The study underscored the effectiveness of CNNs in capturing spatial features relevant to AD diagnosis, providing a fast and reliable approach for clinicians. However, the model's dependency on large datasets and high computational resources for training presents challenges in scalability, particularly in resource-constrained environments.

Tanveer et al. [17] proposed a Deep Transfer Ensemble (DTE) method that employed transfer learning, random hyperparameter search, and snapshot ensembles to classify AD stages. The DTE model demonstrated high classification accuracy, reaching 99.05% and 85.27% on ADNI dataset splits for Normal Controls (NC) vs. AD and Mild Cognitive Impairment (MCI) vs. AD, respectively. Although the ensemble approach achieved superior results, it required significant computational resources, which could limit its application in real-time clinical settings.

AlzheimerNet, proposed by Shamrat et al. [15], achieved an accuracy of 98.67% using a modified InceptionV3 model. AlzheimerNet demonstrated advanced classification capabilities by employing state-of-the-art techniques like CLAHE enhancement and data augmentation. This study represents a significant advancement in AD classification; however, it provided limited details on the specific hyperparameters used, which could restrict reproducibility.

Sorour et al. [16] introduced a hybrid deep neural network model for AD classification, combining elements from LeNet and AlexNet. This model achieved an accuracy of 93.58%, showcasing the potential of hybrid architectures to enhance performance while maintaining computational efficiency. The study highlighted the benefits of merging traditional CNN architectures to capture complex features in MRI data, though it lacked comprehensive comparisons with more complex architectures.

Hazarika et al. [5] provided a comprehensive study on different classification techniques for Alzheimer's Disease (AD) using brain images, with a particular focus on deep learning methodologies such as Artificial Neural Networks (ANNs). Their review highlighted that ANN-based approaches achieved an average accuracy of 93% across various AD classification tasks. This study underscored the effectiveness of neural networks in capturing complex patterns within MRI images, suggesting that ANNs hold promise for accurate AD diagnosis. However, the study noted limitations in the scalability of ANN models, particularly in terms of computational resource requirements, and offered limited guidance on the specific neural network architectures that may be most effective for AD classification. This analysis serves as a valuable resource for understanding the comparative strengths of different deep learning approaches in medical imaging.

Mondal et al. [9] introduced xViTCOS, an explainable vision transformer-based model for COVID-19 screening, which they applied to radiography images (X-ray and CT scans). The model demonstrated high accuracy in detecting COVID-19 cases while also providing interpretable visualizations that highlighted regions of interest, aiding radiologists in diagnosis. While focused on COVID-19, this study showcased the potential of vision transformers in medical image analysis, emphasizing their ability to process spatial information holistically and provide explainable outputs. The xViTCOS model, however, encountered limitations due to data scarcity in COVID-19 radiography and significant computational requirements associated with transformer architectures. Despite these challenges, the success of xViTCOS indicates that ViTs could be adapted to other medical imaging tasks, such as AD classification, where explainability and spatial relationship recognition are equally critical.

Wen, Junhao, et al. [18] conducted an extensive review and reproducible evaluation of convolutional neural networks (CNNs) for Alzheimer's disease classification using MRI data. They categorized approaches into four types: 2D slice-level, 3D patch-level, region-of-interest (ROI) based, and 3D subject-level CNNs. Their findings highlighted that while 3D approaches outperformed 2D slice-based models, issues like data leakage and inconsistencies in validation practices affected performance accuracy. This study emphasizes the need for standardized frameworks to enhance reproducibility and comparability across AD studies.

Sethi, Monika, et al. [14] investigated CNN applications for AD classification, applying 2D CNNs to structural MRI data. Their work focused on understanding how CNNs perform across different AD stages and MRI features, achieving high accuracy in binary classifications (e.g., AD vs. healthy controls). While effective, this approach faced challenges due to class imbalances and limited interpretability, particularly regarding the specific regions of the brain associated with AD markers. Their research highlights the promise of CNNs, despite some limitations in explainability.

2. Proposed methodology

The primary objective of the proposed system architecture is to establish a robust and efficient framework that ensures the project's success. This architecture emphasizes critical aspects such as scalability, adaptability, and security by leveraging modern technologies and adhering to industry best practices to meet the project's objectives effectively. The process flow diagram, illustrated in Fig. 1, provides a visual representation of the sequential steps involved in the system's operation, promoting a clear and comprehensive understanding of the overall process.

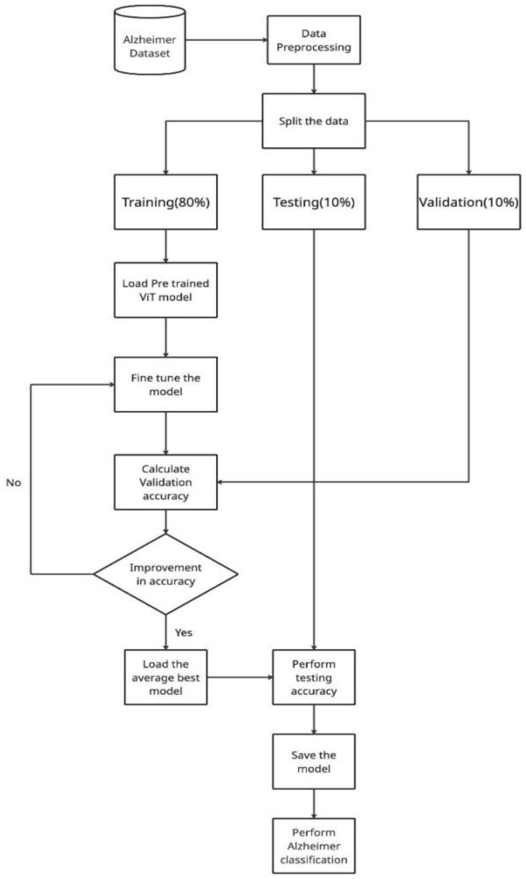


Fig. 1. Process flow model

The proposed model consists of several essential modules. Firstly, the Alzheimer Disease Dataset Collection module focuses on gathering relevant MRI image data. Secondly, the Pre-processing Dataset module handles important tasks, including resizing, normalization, and converting grayscale images to RGB. Following this, the dataset is divided into training, testing, and validation sets in an 8:1:1 ratio. The next step involves loading the ViT model and fine-tuning it using the training data. If the validation accuracy no longer improves, the fine-tuning process is halted, and the best average model is loaded. Subsequently, the Testing Accuracy is performed, and the model is saved. At this stage, the model is fully prepared to accurately classify Alzheimer's images.

2.1. Data collection

The Alzheimer MRI Dataset is a targeted collection of Magnetic Resonance Imaging (MRI) data, specifically curated to aid in Alzheimer's disease research. This dataset is composed of images from diverse sources, including the Open Access Series of Imaging Studies (OASIS), the Alzheimer's Disease Neuroimaging Initiative (ADNI), and other public databases. Such a variety of sources ensures a comprehensive representation of Alzheimer's pathology, making the dataset particularly suited for machine learning and deep learning applications aimed at classifying stages of dementia.

The dataset consists of 6,400 images, categorized into four distinct classes, as shown in Fig. 2, to reflect different stages of Alzheimer's disease progression. The four classes are as follows: "Mild Demented," comprising 896 images; "Moderate Demented," with 64 images; "Non-Demented," the largest category containing 3,200 images; and "Very Mild Demented," with 2,240 images. This distribution provides a balanced yet varied sample that enhances the model's ability to distinguish among the multiple stages of the disease. The images are standardized to a resolution of 128×128 pixels, ensuring uniformity in input size and compatibility with a range of model architectures, while remaining computationally efficient.

To facilitate rigorous training and evaluation, the dataset is partitioned into training, validation, and testing subsets in an 80-10-10 ratio. This split is designed to support effective model training, with 80% of the data allocated for model training, 10% for validation to monitor overfitting and fine-tune parameters, and the remaining 10% set aside for testing to assess final model performance. By maintaining this distribution, the dataset not only provides an optimal framework for model development but also allows for consistent evaluation across different dementia classification models.

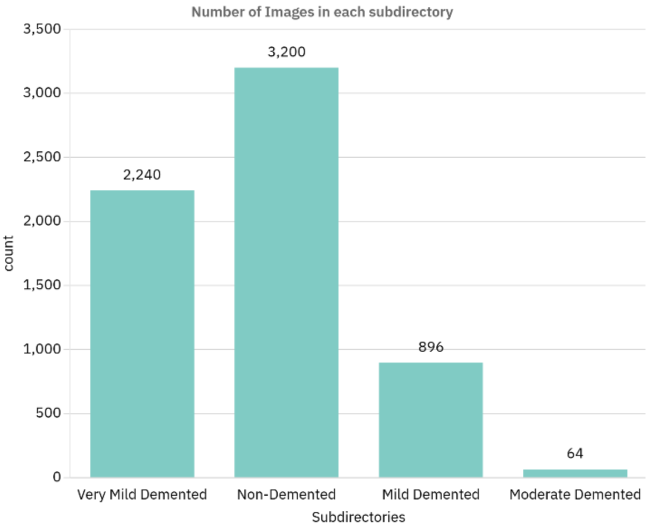


Fig. 2. Data distribution

2.2. Data preprocessing

Data preprocessing is a critical step in preparing the Alzheimer MRI dataset for effective model training and evaluation. Initially, each MRI image is resized to a uniform resolution of 224×224 pixels to ensure consistency across the dataset, facilitating model compatibility and reducing computational requirements. Following resizing, images undergo normalization, scaling pixel values to a standardized range (e.g., 0 to 1), which aids in accelerating model convergence by minimizing variance between pixel intensities. Additionally, since the original MRI images are grayscale, they are converted to RGB format, aligning with the input requirements of the vision transformer model, which is optimized for three-channel images. These preprocessing steps ensure that the dataset is in a consistent and high-quality format, enabling the model to accurately capture structural differences associated with Alzheimer's disease across various stages. Finally, the dataset is divided into training, validation, and test sets with an 8:1:1 split, maintaining a balanced representation of each class to support robust model training, validation, and evaluation.

2.3. Data augmentation

Data augmentation is a crucial technique for enhancing the robustness and generalization of the model, especially when working with limited datasets. In this study, data augmentation was employed to artificially expand the Alzheimer MRI dataset, thereby improving the model's ability to generalize across various transformations. Augmentation techniques such as random rotations, horizontal and vertical flipping, and slight zooming were applied to the images. These transformations simulate different image orientations and variations, which helps the model learn invariant and discriminative features that are essential for accurate classification. By incorporating these augmentation methods, the dataset was diversified, reducing the risk of overfitting and improving the model's performance in identifying and classifying the different stages of Alzheimer's disease under diverse conditions.

2.4. Vision transformer architecture

The vision transformer architecture, introduced in the paper "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale" by Dosovitskiy et al., represents a significant shift in the way image data is processed for computer vision tasks. Unlike conventional convolutional neural networks, which operate directly on pixel grids, ViT treats an image as a sequence of fixed-size patches. These patches, typically of size 16×16 pixels, are linearly embedded into lower-dimensional feature vectors, transforming the 2D spatial information of the image into a sequence of token embeddings. This approach allows the model to leverage the power of transformers, which were initially designed for natural language processing.

To address the lack of spatial information inherent in transformer models, positional encodings are added to the patch embeddings. These encodings provide the model with information about the relative positions of each patch within the image, allowing the ViT to capture spatial relationships between different regions of the image.

The core of the ViT model consists of multiple transformer encoder layers. Each encoder layer includes a multi-head self-attention mechanism, which enables the model to attend to different parts of the image simultaneously. This mechanism helps capture long-range dependencies between the patches, allowing the model to learn complex relationships. After the attention layer, the model applies a feedforward neural network to process the output of the attention mechanism.

This network performs nonlinear transformations to refine the feature representations further. To stabilize training and improve information flow, layer normalization and residual connections are incorporated into the architecture.

The final output of the ViT model is passed through a classification head, which typically includes one or more fully connected layers. A softmax activation function is applied to the output of the last layer to predict the probability distribution across the various classes, thus providing the model's classification results. The vision transformer has shown impressive results across various computer vision tasks, including image classification, object detection, and segmentation, demonstrating its ability to effectively model spatial dependencies and capture complex features in images. The below Fig. 3 shows the ViT architecture.

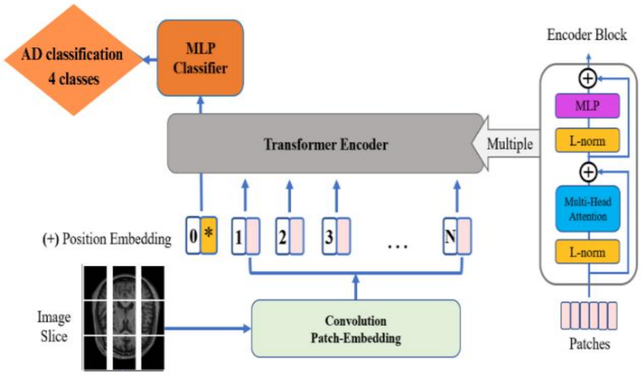


Fig. 3. ViT Architecture

2.5. Model training

The training of the vision transformer model for Alzheimer's disease classification was carried out with a set of carefully defined hyperparameters, optimized for the task at hand. The AdamW optimizer was selected for the model's training, which is a variant of the Adam optimizer that incorporates weight decay, improving generalization and preventing overfitting. The optimizer used the default parameter values provided by PyTorch, which were found to be efficient for training vision transformers.

To stabilize the training process and enhance the model's robustness, exponential moving average (EMA) was employed. EMA averages the model's weights at each step, which helps in maintaining a smooth learning trajectory and reduces the impact of any noise in the optimization process. This technique improves the model's ability to generalize by gradually stabilizing its learned parameters.

In addition to EMA, label smoothing cross-entropy loss was implemented to prevent the model from becoming overly confident in its predictions. By smoothing the target labels, the model is encouraged to make more calibrated predictions, avoiding extreme confidence in incorrect classifications. A smoothing factor of 0.30 was used, ensuring that the model did not become overly biased toward the training examples.

The learning rate was initially set to 0.0001, which was gradually reduced during the training to allow for a more refined learning process. Early stopping was also employed to prevent overfitting by halting training if no improvement in validation accuracy was observed over 7 consecutive epochs, ensuring that the model did not continue training beyond the point of maximum performance.

Hyperparameters were encapsulated in a dictionary, providing a clear configuration for the training procedure. The model was set to train for a maximum of 200 epochs, with accuracy as the primary evaluation metric for both training and validation sets.

For model initialization, pre-trained weights from the ImageNet dataset were used, enabling the vision transformer to benefit from prior knowledge learned on a large, diverse image dataset. This transfer learning approach is crucial in ensuring efficient learning, especially for smaller, specialized datasets like the Alzheimer's MRI dataset.

The training process was managed by SuperGradients' trainer, which facilitated the tracking of metrics, saving model checkpoints, and logging the progress throughout the training. The trainer allowed for easy integration of the various components, including the model, data loaders, and training parameters, ensuring an organized and reproducible training pipeline.

This setup was designed to optimize model performance while preventing overfitting and ensuring stable learning, thus making the training process efficient and effective for Alzheimer's disease classification.

Algorithm 1: Vision Transformer (Vit) Model Training For Alzheimer's Disease Classification

Input: Pre-processed MRI image dataset for Alzheimer's disease classification.

Output: Trained Vision Transformer model for Alzheimer's disease classification.

1. Import necessary libraries: PyTorch, SuperGradients, NumPy, Matplotlib, etc.
2. Load the Vision Transformer model with pre-trained weights (ImageNet).
3. Define training parameters: AdamW optimizer, label smoothing cross-entropy loss with a smoothing factor of 0.30, learning rate 0.0001, and maximum epochs of 200.
4. Implement data preprocessing and augmentation pipelines (resizing, normalization, random rotations, flips, elastic transformations).
5. Split the dataset into training, validation, and test sets (80%, 10%, 10% respectively).
6. Initialize the Trainer object with experiment name and checkpoint directory.
7. Train the model using the training and validation dataloaders, monitoring accuracy and loss at each epoch.
8. Apply early stopping with a patience of 7 epochs to prevent overfitting.
9. Save model checkpoints and maintain an Exponential Moving Average (EMA) of the best-performing model.
10. Evaluate the model on the test set after training and save the final model.

2.6. Model evaluation metrics

In this section, we discuss the evaluation metrics used to assess the performance of the vision transformer model in classifying Alzheimer's disease stages based on MRI images. The following metrics were utilized to gauge the model's classification accuracy and its ability to detect the various stages of dementia: accuracy, confusion matrix, precision, recall, and F1-score.

2.6.1. Accuracy

Accuracy is a standard evaluation metric used to measure the overall performance of a classification model. It is defined as the proportion of correctly classified samples relative to the total number of samples in the dataset. In our study, accuracy provides an indication of how well the ViT model performed in classifying MRI images into the correct dementia stages.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (1)$$

2.6.2. Confusion matrix

The confusion matrix is a valuable tool for understanding the performance of a classification model in greater detail. It shows the relationship between predicted labels and actual labels for each class, providing insight into how well the model distinguishes between each dementia stage. Each cell in the confusion matrix represents true positives, false positives, true negatives, and false negatives, enabling an in-depth evaluation of the model’s ability to classify each dementia stage correctly.

The confusion matrix aids in computing other important performance metrics such as precision, recall, and F1-score, offering a comprehensive understanding of the model's classification capabilities.

2.6.3. Precision

Precision, also known as positive predictive value, measures the proportion of true positive predictions relative to the total number of predicted positive samples. In the context of Alzheimer’s disease classification, precision indicates the model’s ability to avoid false positives. A high precision value suggests that the model is effective at identifying true instances of a given dementia stage without misclassifying non-demented cases.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

2.6.4. Recall

Recall, or sensitivity, measures the proportion of true positive predictions relative to the total number of actual positive samples. Recall assesses how well the model identifies all instances of a particular dementia stage. A high recall value indicates that the model is effective at capturing all true instances of a given class, minimizing the number of false negatives.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3)$$

2.6.5. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model’s performance. It accounts for both false positives and false negatives, making it particularly useful when dealing with imbalanced datasets or when both precision and recall are important. A high F1-score indicates that the model is both precise and capable of identifying most true positives.

$$F1\ score = 2\ \frac{Precision \cdot Recal}{Precision + Recall} \quad (4)$$

3. Results and analysis

3.1. Model results

The ViT model demonstrated strong performance in classifying MRI images into different dementia stages, including Moderate Demented, Mild Demented, Very Mild Demented, and Non-demented. By leveraging its ability to capture spatial dependencies within the images, the ViT model accurately distinguished between various stages of Alzheimer’s disease with high precision. Through extensive training on a diverse dataset comprising MRI images from multiple sources, the ViT model learned to recognize subtle structural brain changes associated with different stages of Alzheimer’s disease. This allowed the model to make precise predictions and effectively categorize MRI scans into the appropriate dementia stage. Overall, the ViT model’s classification results underscore its potential as a valuable tool in automating the diagnosis of Alzheimer’s disease from MRI images. By providing accurate and timely assessments of dementia stages, the ViT model has the potential to aid clinicians in making informed decisions and improving patient outcomes. In this section, we present the results of our study on automating the classification of Alzheimer’s disease (AD) from MRI images using vision transformer models. We discuss the performance of the ViT model, compare it with baseline models, visualize the results, and provide qualitative analysis. The Fig. 4 shows the classification results of random plotting on test data.

3.2. Model performance

The vision transformer model achieved an impressive overall accuracy of 98.9% in classifying MRI images across different dementia stages. This high accuracy indicates that the model correctly predicted the dementia stage for nearly all images, reinforcing its potential as an effective diagnostic tool. In terms of precision, the model scored exceptionally well, achieving an overall precision of 98.6%, which suggests its strong ability to avoid false positives across the classes. The confusion matrix shown in Fig. 5 provides a comprehensive view of the ViT model’s classification performance across four dementia stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented.

Each cell in the matrix displays the number of predictions for each category, with correct classifications appearing along the diagonal. For the Mild Demented category, the model accurately classified 89 out of 91 images, with only two misclassifications. In the Moderate Demented category, all 7 images were perfectly classified, demonstrating flawless performance. The Non-Demented category achieved high accuracy, with 319 out of 320 images correctly classified.

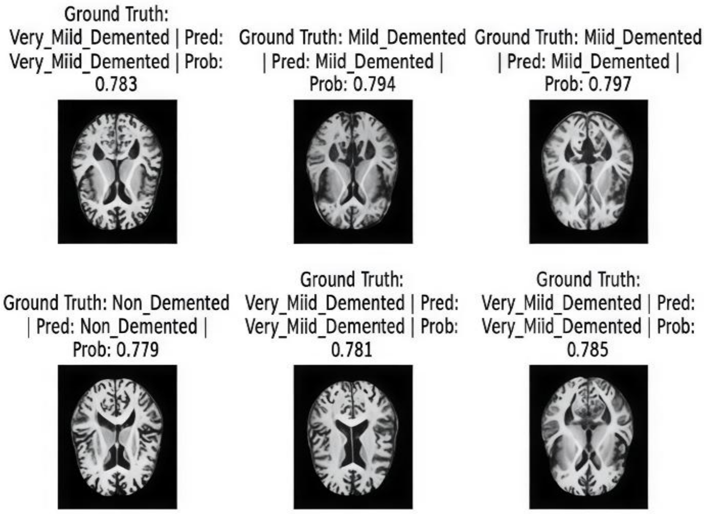

Ground Truth: Very_Mild_Demented | Pred: Very_Miid_Demented | Prob: 0.783
Ground Truth: Mild_Demented | Pred: Miid_Demented | Prob: 0.794
Ground Truth: Miid_Demented | Pred: Miid_Demented | Prob: 0.797
Ground Truth: Non_Demented | Pred: Non_Demented | Prob: 0.779
Ground Truth: Very_Miid_Demented | Pred: Very_Miid_Demented | Prob: 0.781
Ground Truth: Very_Mild_Demented | Pred: Very_Miid_Demented | Prob: 0.785

Fig. 4. The classification results of random plotting on test data

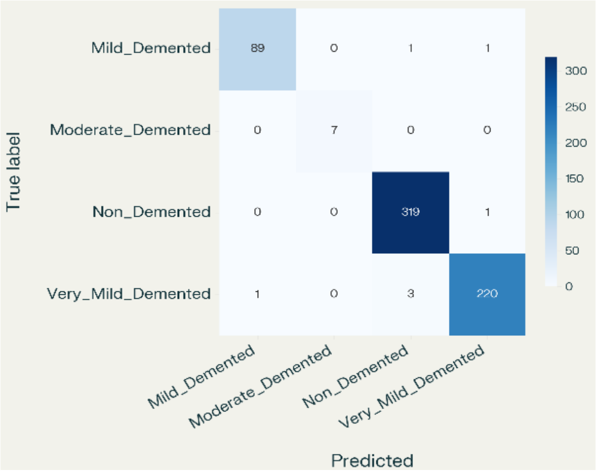


Fig. 5. Confusion matrix

Similarly, in the Very Mild Demented category, the model correctly identified 220 out of 224 images, indicating strong classification capability with only a few misclassifications. This confusion matrix highlights the model's high precision and recall across all stages, reinforcing the robustness and reliability of the ViT model in accurately detecting and classifying various stages of Alzheimer's disease.

The model's overall recall reached 99.2%, signifying its effectiveness in capturing nearly all true cases across dementia stages. This high recall rate means the model successfully identified the vast majority of actual cases, further enhancing its credibility as a reliable classification model. Lastly, the ViT model obtained an overall F1-score of 98.9%, which balances both precision and recall, illustrating its consistent performance across varying levels of dementia.

The table 1 summarizes the individual performance metrics for each dementia class.

Table 1. Evaluation metrics of the model

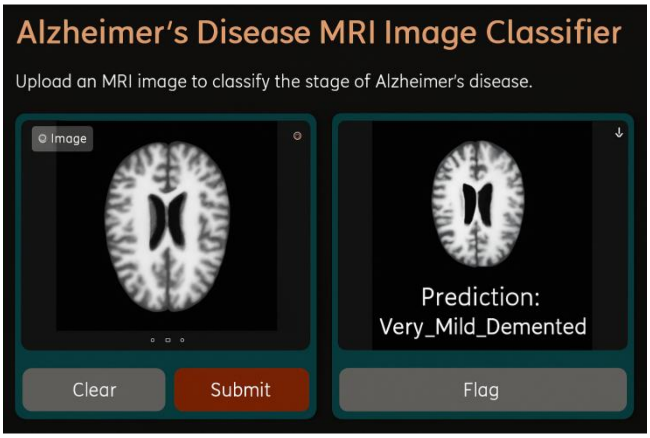
Dementia Stage	Precision	Recall	F1-Score
Mild Demented	97.8%	98.9%	98.3%
Moderate Demented	100%	100%	100%
Non-Demented	100%	100%	100%
Very Mild Demented	98.2%	98.2%	98.8%

The detailed breakdown in the table demonstrates that the ViT model performs consistently across all dementia stages, making it a promising tool for supporting clinical decision-making in Alzheimer's disease diagnosis and monitoring.

3.3. User interface

The user interface for our ViT-based Alzheimer's disease classification model was developed using Gradio, a Python library that enables easy creation of customizable interfaces for machine learning models. Gradio provides an interactive and user-friendly platform, allowing users to upload MRI images and obtain real-time classification results for dementia stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented.

Through the Gradio interface as shown in Fig. 6 users can simply drag and drop an MRI image into the designated upload area. Once the image is uploaded, the model processes it and displays the predicted dementia stage along with confidence scores. This accessible and intuitive interface ensures that healthcare practitioners, researchers, and even non-technical users can interact with the model effortlessly, making it a practical tool for aiding in the early detection and diagnosis of Alzheimer's disease. Additionally, Gradio's flexibility supports easy customization and potential expansion of the interface for future model updates or additional functionalities.


Alzheimer's Disease MRI Image Classifier

Upload an MRI image to classify the stage of Alzheimer's disease.

Prediction: Very_Mild_Demented

Fig. 6. User interface

4. Conclusion

In conclusion, our study demonstrates the effectiveness of vision transformer in accurately classifying Alzheimer's disease based on MRI images. With an overall accuracy of 98.9%, our model showcases strong performance across all classes, including mild, moderate, non-demented, and very mild dementia cases. The high precision and recall values obtained for each class underscore the model's ability to make accurate predictions and capture the majority of actual cases. By leveraging the power of transformer-based architectures, we have opened up new possibilities for non-invasive diagnosis and disease monitoring, offering a promising approach to understanding complex neurological disorders like Alzheimer's disease. Moving forward, there are several avenues for future research and improvement in the realm of vision transformers for medical image analysis. One potential direction is to explore the fine-tuning of pre-trained transformer models on larger and more diverse datasets to further enhance their performance and generalization capabilities. Additionally, investigating novel attention mechanisms and model architectures tailored to medical imaging tasks could lead to further improvements in classification accuracy and interpretability. Furthermore, integrating multi-modal data, such as clinical information and genetic markers, with vision transformers holds promise for developing comprehensive diagnostic tools for neurodegenerative diseases. Continued research in this field has the potential to revolutionize diagnostic imaging, enabling early detection and personalized treatment strategies for neurodegenerative disorders like Alzheimer's disease.

References

[1] Almadhoun H. R., Abu-Naser S. S.: Classification of alzheimer's disease using traditional classifiers with pre-trained CNN. International Journal of Academic Health and Medical Research (IJAHMR) 5(4), 2021, 17–21.
[2] Azad R., et al.: Advances in medical image analysis with vision transformers: a comprehensive review. Medical Image Analysis 91, 2024, 103000.
[3] Dosovitskiy A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
[4] Frisoni G. B., et al.: The clinical use of structural MRI in Alzheimer disease. Nature reviews neurology 6(2), 2010, 67–77.
[5] Hazarika R. A., et al.: An approach for classification of Alzheimer's disease using deep neural network and brain magnetic resonance imaging (MRI). Electronics 12(3), 2023, 676.
[6] He K., et al.: Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition, 2016.
[7] Jack Jr C. R., et al.: NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & dementia 14(4), 2018, 535–562.
[8] Khan S., et al.: Transformers in vision: A survey. ACM computing surveys (CSUR) 54(10s), 2022, 1–41.
[9] Mondal A. K., et al.: xViTCOS: explainable vision transformer based COVID-19 screening using radiography. IEEE Journal of Translational Engineering in Health and Medicine 10, 2021, 1–10.
[10] Pradhan A., Gige J., Eliazer M.: Detection of Alzheimer's disease (AD) in MRI images using deep learning. Int. J. Eng. Res. Technol 10(3), 2021, 580–585.
[11] Saleem T. J., et al.: Deep learning-based diagnosis of Alzheimer's disease. Journal of Personalized Medicine 12(5), 2022, 815.

[12] Samhan L. F., et al.: Classification of Alzheimer's disease using convolutional neural networks. International Journal of Academic Information Systems Research (IJAISR) 6(3), 2022, 18–23.

[13] Sarraf S., et al.: DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. BioRxiv, 2016, 070441 [https://doi.org/10.1101/070441].

[14] Sethi M., et al.: [Retracted] An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network. BioMed Research International 2022(1), 2022, 8739960.

[15] Shamrat F. M. J. M., et al.: AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access 11, 2023, 16376–16395.

[16] Sorour S. E., et al.: Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques. Journal of King Saud University-Computer and Information Sciences 36(2), 2024, 101940.

[17] Tanveer M., et al.: Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. IEEE Journal of Biomedical and Health Informatics 26(4), 2021, 1453–1463.

[18] Wen J., et al.: Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical image analysis 63, 2020, 101694.

[19] World Health Organization: Dementia. 2021 [https://www.who.int/news-room/fact-sheets/detail/dementia].

[20] 2020 Alzheimer's disease facts and figures. Alzheimer's dementia 16, 2020, 391–460 [https://doi.org/10.1002/alz.12068].

Eng. Mohith Reddy Kandi
e-mail: mohith7618@gmail.com

Mohith Reddy Kandi is a fourth-year B. Tech student specializing in Computer Science and Engineering at V.R. Siddhartha Engineering College, Vijayawada, India.
He is passionate about Machine Learning, the Internet of Things, and web programming.

https://orcid.org/0009-0001-2928-4110

Ph.D. Sree Vijaya Lakshmi Kothapalli
e-mail: vijaya@vrsiddhartha.ac.in

Sree Vijaya Lakshmi Kothapalli is an assistant professor in the Department of Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. She is pursuing her Ph.D. in Computer Science and Engineering from Jawaharlal Nehru Technological University, Kakinada, and holds an M.Tech in the same field from the same university, completed in 2009. With over 28 years of experience, her research focuses on Data Mining and Machine Learning. A member of ISTE and the International Association of Engineers, she also holds a Certificate of Merit from the DOEACC Society.

https://orcid.org/0000-0003-1595-3931

Eng. Sivamsh Pavan Rajanala
e-mail: sivamsh.r@gmail.com

Sivamsh Pavan Rajanala is a fourth-year B.Tech. student specializing in Computer Science and Engineering at V.R. Siddhartha Engineering College, Vijayawada, India.
His interests lie in the fields of Internet of Things and Machine Learning.

https://orcid.org/0009-0005-2851-5641

Prof. Suvarna Vani Koneru
e-mail: suvarnavanik@vrsiddhartha.ac.in

Dr. Suvarna Vani Koneru is a professor in the Department of Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. She earned her Ph.D. from the University of Hyderabad in 2014 and has over 24 years of experience in teaching and research.
Her areas of interest include bioinformatics, Machine Learning, and GIS. A recipient of the National Level Summer Research Fellowship (2010) and the Venus International Women Award (2017). Dr. Koneru is a recognized research supervisor and a member of ISTE and the Institution of Engineers (India).

https://orcid.org/0000-0003-4608-6308

Eng. Vishnu Pramukh Vattikunta
e-mail: vishnupramukh0007@gmail.com

Vishnu Pramukh Vattikunta is a fourth-year B. Tech. student specializing in Computer Science and Engineering at V.R. Siddhartha Engineering College, Vijayawada, India.
His interests lie in the fields of Machine Learning and web programming.

https://orcid.org/0009-0009-2902-7999