I.J. Intelligent Systems and Applications, **2026, 2, 44-63**
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2026.02.04

Transformer-Based vs. CNN-Based Deep Learning for Alzheimer's Disease Classification: Performance and Deployment

Nhat-Kha Nguyen
School of Biomedical Engineering, International University, Vietnam National University HCM City, HCM City, 70000, Vietnam
E-mail: bebeiu19071@student.hcmiu.edu.vn
ORCID iD: https://orcid.org/0009-0004-9724-0613

Thi-Thu-Hien Pham
School of Biomedical Engineering, International University, Vietnam National University HCM City, HCM City, 70000, Vietnam
E-mail: ptthien@hcmiu.edu.vn
ORCID iD: https://orcid.org/0000-0001-5808-3214

Nhat-Minh Nguyen
School of Biomedical Engineering, International University, Vietnam National University HCM City, HCM City, 70000, Vietnam
E-mail: SBEIU19015@mp.hcmiu.edu.vn
ORCID iD: https://orcid.org/0009-0004-6120-1225

Tan-Nhu Nguyen
School of Biomedical Engineering, International University, Vietnam National University HCM City, HCM City, 70000, Vietnam
E-mail: ntnhu@hcmiu.edu.vn
ORCID iD: https://orcid.org/0000-0003-3343-0886

Ngoc-Bich Le*
School of Biomedical Engineering, International University, Vietnam National University HCM City, HCM City, 70000, Vietnam
E-mail: lnbich@hcmiu.edu.vn
ORCID iD: https://orcid.org/0000-0001-7431-0157

Received: 10 November 2025; Revised: 04 January 2026; Accepted: 06 February 2026; Published: 08 April 2026

Abstract: It is well known that diagnosing Alzheimer's disease (AD) accurately and early is a major clinical challenge, especially when using brain MRI data to differentiate between subtle stages of cognitive decline. This study investigated the efficacy of two deep learning models for the classification of AD stages: Vision Transformer (ViT), a transformer-based architecture, and EfficientNetB7, a convolutional neural network. To enhance classification performance and address class imbalance, extensive data preprocessing and augmentation techniques were employed on the publicly accessible 'Alzheimer's Dataset (4 class of Images)' from Kaggle. This dataset comprises 6,400 brain MRI images categorized into four AD stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. Techniques applied included cropping, horizontal and vertical flipping, 20-degree rotations, histogram equalization, Gaussian noise addition, Gaussian blurring, and thresholding, aimed at improving the representation of underrepresented classes. Hyperparameter optimization was executed via a two-phase methodology: an initial grid search to determine parameter ranges, followed by Bayesian optimization employing an upper confidence bound acquisition function to refine learning rates, batch sizes, momentum, and weight decay values. Experimental results indicated that EfficientNetB7 attained a classification accuracy of 93.5% with F1-scores surpassing 92% for early-stage classes, whereas Vision Transformer (ViT) recorded a lower accuracy of 88.7% and exhibited diminished sensitivity to early-stage instances. The performance disparity is due to ViT's dependence on extensive training datasets, which may restrict its generalization when utilized on comparatively smaller medical imaging datasets. The results indicate that, in dataset-constrained

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

scenarios, CNN-based architectures such as EfficientNetB7 may provide more consistent and effective performance. With distinct training, validation, and test datasets, the model's generalization, training stability, and computational efficiency were assessed. With an intuitive user interface, the top-performing model, EfficientNetB7, was implemented as a web-based application to facilitate real-time supportive predictions for research demonstration. This comparative analysis demonstrated that the CNN-based EfficientNetB7 exhibited more robustness with constrained medical imaging data and was computationally economical, but the transformer-based ViT displayed increased sensitivity to dataset size and necessitated extended training to attain similar convergence. The development of a validated and deployable AI-based Alzheimer's disease diagnostic solution showed great promise for clinical use.

Index Terms: Convolutional Neural Networks, Alzheimer's Disease, Deep Learning, EfficientNet, Vision Transformer, Web Application, Early Diagnosis.

1. Introduction

The increasing incidence of Alzheimer's disease (AD) worldwide, along with the significant demand for early and precise diagnosis, has generated an urgent requirement for more dependable diagnostic instruments. Contemporary diagnostic techniques, including manual Magnetic Resonance Imaging (MRI) analysis, are both resource-demanding and imprecise in detecting early-stage AD. Consequently, healthcare professionals are adopting AI-driven technologies, especially Convolutional Neural Networks (CNNs), to enhance the accuracy of AD diagnosis and expedite decision-making processes. These technologies can aid in monitoring disease progression and distinguishing AD from other types of dementia, offering a crucial answer for enhancing patient care.

Machine learning, especially deep learning techniques [1], has demonstrated significant potential in medical analysis [2–8], providing solutions that assist healthcare professionals in achieving more precise and timely diagnoses. Deep learning models, particularly CNNs [9], have been widely employed in image recognition tasks [11–13], instance segmentation of neuronal cells [13], brain tumor segmentation [14,15], dementia detection [16], breast cancer classification [17], AD diagnosis [18,19], diabetic detection [20], stroke detection [21], blood pressure estimation [22], retinal disease detection [23], gender classification [25,26], white blood cell identification [26], plant disease detection [27–29], and rapid COVID-19 detection [30] owing to their capacity to autonomously extract complex patterns and features from visual data.

In recent years, deep learning has experienced substantial progress in CNN designs, resulting in enhanced performance across multiple domains, including medical imaging. This paper examines the utilization of sophisticated deep learning models for Alzheimer's disease detection, specifically EfficientNet [31], a CNN, and Vision Transformer (ViT) [32], a Transformer-based model employing self-attention mechanisms in lieu of convolutional layers.

Recent breakthroughs in artificial intelligence (AI), especially in the utilization of deep learning models such as CNNs, have demonstrated considerable potential in identifying AD using medical imagery, including MRI scans. These models, along with data augmentation and preprocessing methods, have enhanced the precision and dependability of AD diagnosis, facilitating early detection and improved patient outcomes.

Cui et al. [33] presented an improved inception network for predicting the early stages of AD utilizing brain MRI scans. Through the implementation of preprocessing approaches, including histogram equalization and multi-threshold segmentation, the model effectively concentrated on brain regions significantly impacted by AD. This approach attained an overall accuracy of 85.7%, highlighting the efficacy of CNNs in detecting early indicators of AD. Amir Ebrahimi [34] utilized ResNet-18, pre-trained on the ImageNet dataset, for AD identification. The investigation of deep sequence-based models revealed that the temporal convolutional network (TCN) demonstrated superior performance, achieving an accuracy of 91.78%, sensitivity of 91.56%, and specificity of 92%, thereby reinforcing the robustness of CNNs in medical imaging applications.

Waleed Al Shehri et al. [35] conducted a comparison of two CNN architectures—DenseNet169 and ResNet-50—in the diagnosis and classification of AD into distinct stages. DenseNet169 surpassed ResNet-50, attaining a training accuracy of 97.7% and a testing accuracy of 83.82%, whereas ResNet-50 recorded training and testing accuracies of 88.7% and 81.92%, respectively. This underscores the increasing trend of employing various CNN architectures for improved AD detection.

Yildirim et al. [36] conducted an additional study that presented a hybrid model integrating ResNet50, DenseNet201, VGG16, and AlexNet architectures. Their research indicated that the hybrid model attained the best accuracy rate of 90%, surpassing individual CNN models like ResNet50 and DenseNet201, which recorded accuracies of 87% and 78%, respectively. Mohammed BA and colleagues [37] investigated hybrid models that integrate AlexNet with support vector machines (SVM) to mitigate class imbalance challenges in Alzheimer's disease detection. The AlexNet+SVM hybrid model attained an accuracy of 94.8% utilizing MRI data from the OASIS and Kaggle databases, highlighting the significance of integrating several AI methodologies for enhanced Alzheimer's disease identification.

Gharaibeh et al. [38] conducted a study utilizing pre-trained InceptionV3 and DenseNet201 models, implementing data augmentation and feature selection by principal component analysis (PCA). Their work attained an exceptional accuracy of 99%, illustrating the capability of deep learning models in accurate AD diagnosis. H. Ji et al. [39] concentrated

on the early diagnosis of AD utilizing CNN models applied to MRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, attaining a classification accuracy of 97.65% for distinguishing AD from Mild Cognitive Impairment (MCI) and 88.37% for differentiating MCI from Cognitively Normal (CN) subjects, thereby reinforcing the efficacy of CNNs in identifying subtle cerebral alterations.

Asim et al. [40] introduced a unique multi-atlas method that integrates features from many atlas templates and uses PCA for feature selection. Their methodology attained a classification accuracy of 94% for AD against Cognitively Normal and 76.5% for Cognitively Normal versus Mild Cognitive Impairment. The multi-atlas strategy surpassed the single-atlas approach, underscoring the need to employ many data sources and approaches in AD diagnosis. Ultimately, Sisodia et al. [41] examined deep transfer learning models for AD diagnosis, evaluating ResNet50, VGG19, Xception, DenseNet201, and EfficientNetB7. DenseNet201 attained the highest validation accuracy of 96.59%, surpassing other models, hence underscoring the significance of transfer learning and CNN architectures in enhancing AD diagnostic precision.

These results underscore the increasing efficacy of AI in diagnosing AD, especially through the application of CNNs and hybrid models. Through the utilization of data augmentation, transfer learning, and deep learning architectures, these models advance the frontiers of medical diagnostics, providing improved accuracy and reliability in the early identification of AD.

Despite the excellent accuracy of existing CNN-based systems, they continue to encounter technological hurdles. Models like ResNet-18, DenseNet169, and hybrid CNN-SVM approaches exhibit potential; nonetheless, they frequently encounter overfitting, particularly in the context of imbalanced datasets. Data shortage, imbalance, and dependence on outdated models can diminish the efficacy of these strategies. Furthermore, numerous current models have not adequately tackled the necessity for real-world implementation on user-friendly, scalable systems, hence constraining their practical utility in clinical environments. In contrast to conventional CNNs, EfficientNet utilizes a compound scaling approach that effectively equilibrates depth, width, and resolution. This adaptive scaling enables EfficientNet to more efficiently capture complex structural patterns in MRI scans compared to models such as ResNet and DenseNet, which possess less flexible architectures for managing high-resolution images. EfficientNet attained a superior classification rate of 93.5%, surpassing the conventional 85-90% recorded by DenseNet and ResNet in analogous Alzheimer's disease investigations, thus illustrating its robustness and appropriateness for early AD detection. The ViT, while not frequently utilized in medical imaging, provides an alternate method by collecting global dependencies via its self-attention mechanism. Although ViT necessitates greater computer resources, its capacity to analyze long-range spatial patterns may improve the identification of widespread neuropathological alterations linked to AD progression. EfficientNet retains its superiority in scalability owing to its computational efficiency, which is essential for use in practical clinical settings. ViT may be deemed suitable for resource-intensive environments where the capture of intricate relationships is essential.

This study advocates for the implementation of two state-of-the-art models: EfficientNet and ViT, both of which provide enhanced functionalities for the early diagnosis of Alzheimer's disease. EfficientNet, demonstrating superior accuracy with ImageNet weights, will be compared against ViT, an innovative model that, although not being a CNN, has displayed robust performance in image processing. The paper employs data augmentation approaches and meticulous dataset partitioning to enhance model robustness and reduce overfitting, thereby addressing technical deficiencies.

The research will evaluate the diagnostic efficacy of different models and implement the most effective one in a web-based application. This web application, created with Python, HTML, CSS, and JavaScript, will offer healthcare professionals a simple interface for Alzheimer's disease diagnosis. It will also incorporate a MySQL database for safe data management. This holistic strategy seeks to connect machine learning advancements with realistic clinical implementations, providing a scalable and accessible solution for early Alzheimer's disease diagnosis in real-world healthcare settings.

This study has three principal objectives: (1) to evaluate the diagnostic efficacy of a transformer-based model (ViT) versus a CNN-based model (EfficientNetB7) utilizing brain MRI images for Alzheimer's disease classification, (2) to examine the influence of sophisticated preprocessing, augmentation, and hyperparameter optimization on classification performance and generalization, and (3) to illustrate a prototype web-based implementation of the highest-performing model. This study distinguishes itself from previous research by providing a comparative analysis and highlighting the shift from model creation to practical implementation, rather than concentrating on a singular CNN architecture or neglecting end-to-end deployment.

The proposed technology, although demonstrating promising results, is constrained by various restrictions. This encompasses the utilization of 2D axial MRI slices instead of 3D volumetric data, the omission of biomarker integration (such as PET imaging or CSF analysis), and the lack of external clinical validation. These considerations may restrict the generalizability and therapeutic applicability of the suggested method. Nevertheless, the work establishes a robust basis for future investigations into multimodal and clinically validated AI instruments for the early identification of Alzheimer's disease.

This paper's primary contributions are as follows: (1) A comparative assessment of a CNN-based model (EfficientNetB7) and a transformer-based model (Vision Transformer - ViT) for multi-class classification of Alzheimer's disease utilizing T1-weighted 2D brain MRI slices; (2) An elaborate methodology featuring advanced data preprocessing, class-specific augmentation, and a dual-phase hyperparameter optimization framework (grid search and Bayesian optimization), enhancing model robustness and mitigating dataset imbalance; (3) Illustration of practical deployment viability by incorporating the optimal model into a web-based application for real-time inference, complete with

confidence scores and an intuitive interface; (4) A critical examination of the limits inherent in 2D image-based diagnosis and the constraints of transformer models, accompanied by suggestions for future research focused on 3D medical imaging and the integration of multimodal data within existing clinical frameworks.

The subsequent sections of this work are organized as follows: Section 2, Materials and Methodology, delineates the dataset employed in this study, elucidates the data preparation and augmentation approaches, and outlines the architectures and training protocols for both the EfficientNetB7 and Vision Transformer models. Section 3, Results and Discussion, presents the experimental findings, encompassing comparative performance measures and an analysis of each model's behavior. Section 4, Conclusions, encapsulates the principal findings, delineates the study's limitations, and proposes avenues for future research and clinical use.

2. Materials and Methodology

Fig. 1 presents a sequential layout of the experimental methods to offer a clear perspective of the workflow. The pipeline has six fundamental stages, each contingent upon the preceding one to facilitate precise machine learning analysis.

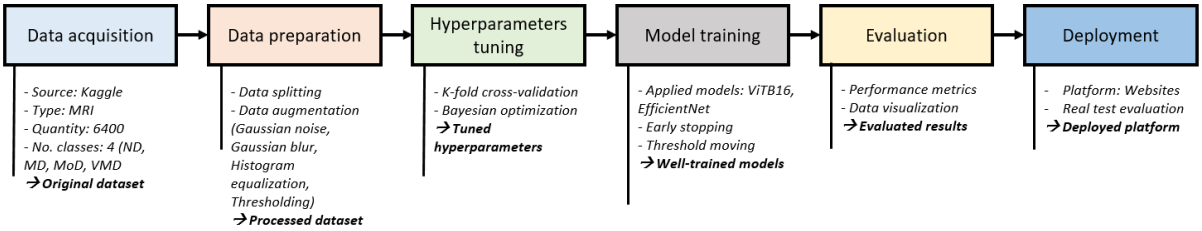


Fig.1. Experimental pipeline

The first phase, Data Selection, outlines the procurement of a high-quality dataset. This dataset comprises 6,400 MRI images in JPEG format, each with a size of 176x208 pixels, classified into four phases of dementia: Non-Demented (ND), Mild Demented (MD), Very Mild Demented (VMD), and Moderate Demented (M). These categories establish the foundation for training a model proficient in differentiating stages of advancing Alzheimer's disease (AD), essential for early diagnosis.

During the Data Preparation phase, the dataset is methodically divided to guarantee equitable class distribution. Data augmentation methods were employed to enhance class equilibrium and alleviate overfitting. Hyperparameter tuning involves optimizing variables including learning rate, model depth, batch size, and epoch count. This iterative procedure, grounded in previous studies and experimentation, guarantees optimal model performance.

Upon fine-tuning hyperparameters, model evaluation is performed utilizing accuracy and loss measures derived from training and testing datasets. Visual instruments such as confusion matrices offer supplementary insights on classification accuracy and model efficacy. The optimized model is implemented as a web application for clinical use, with intentions for ongoing data updates to ensure stability and adaptability.

The methodology advances from data collection—a laborious phase owing to the limited availability of medical datasets—through methodical data reprocessing and augmentation to produce a comprehensive, refined, and stratified dataset. Hyperparameter optimization of two chosen deep learning models guarantees excellent training results, facilitated by early stopping techniques to improve generalization. The objective is to optimize the Area Under the Curve (AUC) for efficient anomaly identification. The trained models are then validated and stored for online deployment, facilitating a smooth transition from development to clinical application.

This comprehensive methodology demonstrates a systematic and methodical strategy for implementing machine learning in AD diagnosis, with the objective of enhancing patient care via improved diagnostic accuracy.

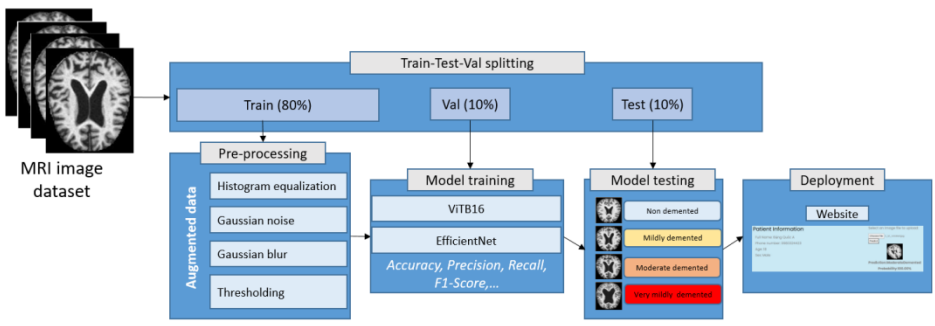


Fig.2. Implementation process

Fig. 2 provides a detailed depiction of the experimental procedure, outlining each successive stage in the methodology. The process commences by partitioning the MRI dataset into three subsets: training (80%), validation (10%), and testing (10%). This partitioning guarantees a uniform distribution of data, enhancing its suitability for training, validation, and testing.

The training set subsequently undergoes a sequence of preprocessing procedures aimed at improving image quality and relevance for model training. Techniques such as histogram equalization, Gaussian noise addition, Gaussian blur, and thresholding were solely implemented on the training set, hence reducing the likelihood of information leakage into the validation and test sets. This regulated method maintains the integrity of the assessment process. Post-training, the models undergo evaluation utilizing typical performance indicators such as accuracy, precision, recall, and F1-score. These metrics offer a comprehensive evaluation of the models' proficiency in categorizing images into four dementia classifications: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented, in accordance with the study's diagnostic aims. In the concluding phase, the trained models are implemented on a web platform to enable their utilization in practical situations. In this deployment phase, healthcare professionals and end-users access the AD diagnosis models, hence enhancing healthcare outcomes.

This study assesses the efficacy of two separate models, Vision Transformer B16 (ViT-B16) and EfficientNet, both trained on a newly preprocessed dataset. Diverse preprocessing techniques and thorough model evaluation were employed to improve diagnostic accuracy in AD detection, hence advancing medical image analysis and clinical decision-making.

EfficientNet and ViT were chosen for this work because their distinctive architectures are particularly adept at addressing the issues associated with diagnosing AD from brain MRI scans. EfficientNet's compound scaling method effectively equilibrates network depth, width, and resolution, allowing it to capture intricate structural features and broader spatial patterns in MRI images. This is especially crucial for identifying nuanced alterations in brain regions, such as the hippocampus, that are impacted in the initial stages of AD. Conversely, ViT employs a self-attention mechanism that enables the recording of long-range dependencies within images, a beneficial attribute for identifying distributed patterns in MRI scans linked to neurodegenerative development. In contrast to conventional CNN models, which often exhibit limited adaptability and efficiency for high-resolution MRI data, EfficientNet and ViT offer a compelling blend of accuracy and computational efficiency for clinical applications, rendering them appropriate for web-based implementation in practical diagnostic environments.

2.1. Data Preparation

A. Dataset

The dataset utilized in this research is the "Alzheimer's Dataset (4 classes of images)" [42], which is freely accessible on Kaggle and has garnered over 15,000 downloads. This dataset, assembled from several sources and manually annotated by a radiologist [43], comprises 6,400 T1-weighted axial MRI images in JPEG format (176×208 pixels). These images were labeled into four diagnostic categories: (1) Non-Demented (ND): 3,200 images; (2) Very Mild Demented (VMD): 2,240 images; (3) Mild Demented (MD): 896 images; (4) Moderate Demented (M): 64 images (see sample images in Fig. 3). It should be noted that the class labels (e.g., 'Very Mild Demented') are based on the original dataset annotations and do not correspond directly to the NIA-AA or ATN diagnostic frameworks. While these labels do not officially correspond to clinical categories like Cognitive Normal (CN) or Mild Cognitive Impairment (MCI), they are somewhat analogous within the framework of MRI-based categorization and are employed in this study in accordance with the dataset's labeling guidelines. The dataset is partitioned into two primary files for training and testing, with the training set consisting of 4,098 images, of which 1,023 are allocated for validation, and the test set including 1,279 images. The dataset is divided into training, validation, and test sets in an 80–10–10% ratio, ensuring a balanced distribution for model training and assessment. To avoid data leakage, a patient-level split was enforced during dataset partitioning. All images from a single subject were restricted to a single subset (training, validation, or test), ensuring that the model did not encounter images from the same individual across different phases of learning and evaluation.

Observable visual changes in brightness and contrast exist among the dementia phases in MRI scans, underscoring abnormalities in brain anatomy linked to each condition. The dataset lacks demographic metadata (e.g., age, gender); however, its extensive utilization and expert annotation establish a solid foundation for model training. To avert data leakage, images from individual patients were confined to a single subset (training, validation, or test), guaranteeing that no patient's images are present in multiple sets. This method reduces the likelihood of the model acquiring patient-specific traits, therefore enhancing its capacity to generalize to unfamiliar data.

Data augmentation methods, such as histogram equalization, Gaussian noise, Gaussian blur, and thresholding, were implemented just on the training set, maintaining the integrity of the validation and test sets. This targeted enhancement bolsters the dependability and generalizability of the model assessment, guaranteeing that performance measures faithfully represent the model's capabilities in practical applications.

The dataset is systematically structured into directories for training, validation, and testing, featuring subdirectories for each classification: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This organized framework optimizes data administration and model training. Additionally, stringent quality control was enforced for all images, retaining only those of significant clinical value and clarity. Distinct patient identifiers (IDs) were allocated to each MRI to systematically organize data and avert cross-contamination among subgroups. For instance, all images from a patient with numerous MRI scans were confined to a single dataset, either for training or validation/testing, to prevent

inflated performance metrics resulting from shared patient data.

The implementation of quality control techniques and systematic dataset curation guarantees the utilization of only clinically pertinent images, hence augmenting the reliability and validity of the AI models developed from this data. This rigorous methodology enhances the model's efficacy and reliability in diagnosing AD.

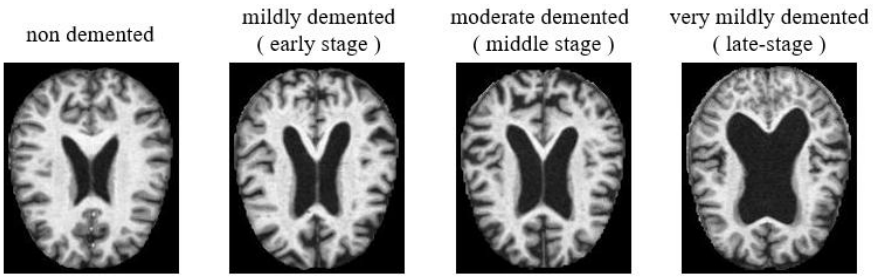


Fig.3. Samples of each class in the "Alzheimer's Dataset (4 classes of images)"

B. Data Augmentation

Initially, all images were cropped to eliminate superfluous features, concentrating on areas essential for analysis. To rectify the significant class imbalance in the initial dataset, comprehensive data augmentation was implemented for the underrepresented classes. Methods like horizontal and vertical flipping, 20-degree rotations, Gaussian noise injection, Gaussian blurring, histogram equalization, and thresholding were strategically employed to augment the datasets for the Mild and Moderate Demented categories.

To further reduce overfitting, histogram equalization and thresholding approaches were utilized, hence augmenting the dataset size and improving image quality. Both approaches are beneficial for enhancing the model's classification efficacy. Gaussian noise was incorporated as a regularization method, bringing stochastic variability that diminishes dependence on particular patterns, hence fostering generalization and improving model accuracy.

Gaussian blur was utilized to mitigate overfitting by softening sharp edges and intricate details, hence reducing high-frequency noise and streamlining the input data. This procedure diminishes model complexity and mitigates undue focus on extraneous features, leading to enhanced generalization and increased accuracy.

Histogram equalization modifies visual contrast and dynamic range by redistributing pixel values to emphasize significant features and patterns. The heightened contrast furnishes the model with clearer information, resulting in increased classification accuracy and less likelihood of overfitting.

Thresholding simplifies images by transforming them into binary representations, categorizing pixels as foreground or background according to intensity. This procedure diminishes data complexity, enabling the model to concentrate on essential regions or objects pertinent to categorization, hence improving accuracy.

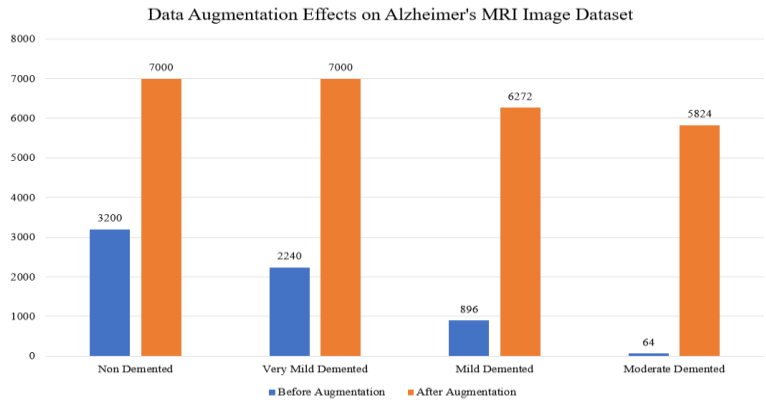


Fig.4. Changes in the dataset after augmentation

Fig. 4 depicts the impact of data augmentation on the distribution of images among the four dementia phases in the dataset: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The dataset initially exhibited a notable imbalance, characterized by a scarcity of images in the Mild and Moderate Demented categories. After augmentation, the dataset was rebalanced to include the following number of images per class: (1) Non-Demented: 7,000 images; (2) Very Mild Demented: 7,000 images; (3) Mild Demented: 6,272 images; (4) Moderate Demented: 5,824 images. The bar chart shown in Fig. 4 illustrates that various augmentation procedures effectively enhanced the sample size in the minority classes, resulting in a more balanced dataset. This enhancement optimizes model training by mitigating the possibility of bias towards more dominant classes and facilitating improved generalization across all stages of dementia.

Subsequent to preprocessing, the dataset was partitioned into training, validation, and testing subsets in an 8:1:1 ratio. This stratification guarantees a balanced data distribution for model training, while the validation set is utilized to track progress and prevent premature cessation. The test set is thereafter allocated for the final evaluation of the model. In the absence of this segmentation, the model may over-fit the training data, hence diminishing its capacity to generalize to novel data. In the stratified splitting phase, the training data was divided into four equal parts to enhance learning efficiency.

The augmentation procedures significantly enhanced the dataset's volume and diversity, facilitating the development of a more balanced dataset. This heterogeneous data distribution is expected to improve the model's generalizability by diminishing dependence on restricted patterns and promoting extensive learning from diverse samples.

C. Handling Class Imbalance

The dataset utilized in this investigation, while extensive, exhibited an imbalance among AD stages, notably with a scarcity of samples in the Very Mild and Moderate Demented categories. To tackle this issue, many solutions were implemented in addition to conventional data augmentation. Stratified sampling was employed in dataset partitioning to provide proportional class representation among training, validation, and test sets. Class weighting was integrated into the loss function, imposing greater penalties on misclassified under-represented samples. Oversampling of under-represented classes in each training batch occurred, enhancing model generalization and mitigating bias towards more prevalent classes such as Non-Demented.

2.2. Model Selection

The meticulous choice of appropriate models is essential for attaining optimal performance and maximizing computing efficiency. This study evaluated many criteria, such as computing efficiency, model complexity, and assessment scores, to create a balanced methodology. Due to the significant computational demands of high-performance models, priority was given to selecting models that provided competitive evaluation scores in relation to their size and complexity. This technique seeks to optimize computing resources while preserving enough diagnostic efficacy in the identification of AD.

Two models were chosen for their efficacy in image classification tasks: (1) EfficientNetB7, a convolutional neural network recognized for its efficient scaling and precision, and (2) ViT, a transformer-based model that employs self-attention to examine global relationships in images. Each model possesses distinct advantages, with EfficientNetB7 emphasizing resource efficiency and ViT presenting an innovative Transformer-based methodology for image processing. This study seeks to evaluate the comparative efficacy of various models based on essential characteristics, such as accuracy, computing efficiency, and stability. The choice of these models demonstrates a calculated equilibrium between performance and resource efficiency, guaranteeing that the study optimizes the capacity for accurate and scalable Alzheimer's disease diagnosis by machine learning methodologies. The objective is to select the most appropriate model(s) for clinical adoption through thorough study and comparison, hence enhancing healthcare outcomes.

2.3. Hyperparameters

The meticulous selection and adjustment of hyperparameters are crucial for enhancing the performance of machine learning models. This study meticulously adjusted essential hyperparameters—namely learning rate, batch size, optimizer, number of epochs, and dropout rate—for the EfficientNetB7 and ViT models (refer to Table 1). The objective was to attain optimal training and assessment results.

Table 1. Initial hyperparameters

	Initial learning rate	**Batch size**	**Optimizer**	**Epochs**	**Dropout rate**
EfficientNetB7	0.01	32	RMSprop	60	0.3
Vision Transformer (ViT)	0.003	64	AdamW	80	0.1

The selection of optimizers, learning rate schedules, and regularization methods was informed by empirical results and recognized best practices. The modifications sought to improve accuracy, stability, and generalization in the model's efficacy for Alzheimer's disease detection from MRI scans. In this setting, stability comprises two essential elements: training consistency and resilience to unfamiliar data. The consistency of training was assessed by analyzing loss and accuracy curves to detect variations or abnormal patterns that suggest instability or overfitting. Robustness was assessed by evaluating the model's performance on test data, with a reliable model anticipated to exhibit consistent performance across training, validation, and test sets, showing low variation in accuracy, precision, recall, and F1-score. This consistency guarantees that the model generalizes effectively to novel data, preventing overfitting to the training dataset.

Stability was quantified by monitoring learning curves and comparing important parameters (accuracy and F1-score) across all three datasets. Significant differences or irregular learning curves were identified as indicators of instability, necessitating additional regularization and hyperparameter optimization to enhance stability.

Early stopping was employed as a precautionary measure against overfitting, enabling the cessation of training when the model exhibited signs of overfitting, thereby conserving time and computational resources. Early stopping is essential for averting overfitting, a phenomenon where the model retains the training data instead of acquiring generalizable

patterns. The training procedure was halted upon evaluating model performance on the validation set, utilizing established criteria including strip length, cool-down duration, patience, and latency. The stability of the learning curves served as a major criterion for hyperparameter tuning and training, functioning as a qualitative parameter similar to a hyperparameter.

Various performance criteria were utilized for model evaluation in multi-class classification. Although accuracy is a fundamental statistic, many advanced metrics offer enhanced insights into model performance, especially in medical diagnosis. This analysis utilized key metrics such as the Confusion Matrix, Precision, Recall, F1-score, and Stability, with particular focus on Learning Curve Stability for its efficacy in managing imbalanced datasets and resistance to threshold fluctuations.

The Confusion Matrix, presented as a tabular summary, provides insights into True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN), juxtaposing model predictions with actual labels in the sample. A high false positive rate indicates a Type I error, whereas a high false negative rate denotes a Type II error. In medical diagnosis, false negatives provide a significant danger by potentially allowing a problem to remain untreated, which may result in inappropriate or insufficient therapy. In such instances, modifying the threshold nearer to zero can facilitate the prioritization of False Positives (FP) over False Negatives (FN), thus achieving an appropriate equilibrium according to diagnostic necessities.

Via the Confusion Matrix, five metrics can be calculated:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{3}$$

$$F1\text{-}score = \frac{2TP}{2TP+FP+FN} \tag{4}$$

2.4. Hyperparameter Tuning

Hyperparameter tuning is crucial for enhancing model performance by determining the optimal configuration of training parameters. This study conceptualized hyperparameter tuning as a sequential process that includes model evaluation via k-fold cross-validation and global optimization through Bayesian approaches (Fig. 5). The loop continued until the desired performance score was attained, at which point the optimal set of hyperparameters was chosen.

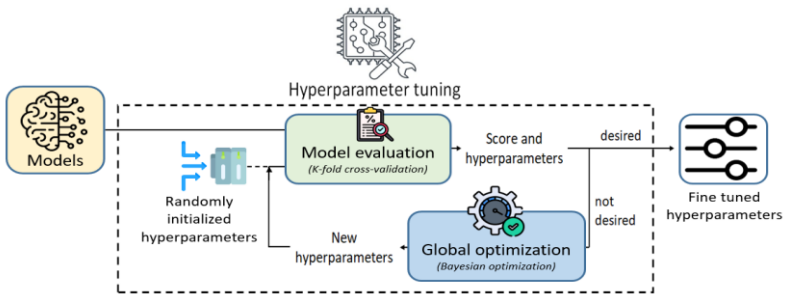


Fig.5. Schematic diagram of hyperparameter tuning

This study primarily modified the hyperparameters, including the loss function, learning rate, batch size, optimizer, epochs, and dropout rate. The parameters established prior to the commencement of training substantially affect the model's learning efficacy and operating efficiency. Choosing an appropriate loss function is essential for classification tasks, as it affects the model's capacity to distinguish between classes. In this context, Categorical Cross-Entropy was employed for multi-class classification, defined as:

$$H(y, \hat{y}) = -\sum_i y_i \log(\hat{y}_i) \tag{5}$$

where y represents the true distribution and $\hat{y}$ the predicted probability distribution.

The Adam optimizer was utilized with an initial learning rate of $1\times10-5$, supplemented by a ReduceLROnPlateau schedule to decrease it to a minimum of $1\times10-6$ as necessary. This method utilizes Adam's capacity to manage sparse gradients and dynamically modify learning rates, facilitating gradual and accurate weight adjustments and ongoing fine-tuning, even as performance enhancements become marginal. This meticulous adjustment enhances the model's learning, facilitating effective generalization to novel data.

Batch size and epochs were optimized by iterative trials to achieve a compromise between efficient learning and the prevention of overfitting. Early stopping was utilized to terminate training when validation performance ceased to

improve, hence preserving resources. Furthermore, weight decay (regularization) was employed to mitigate overfitting by imposing a penalty on the weight magnitudes, hence promoting simpler, more generalizable patterns.

Manual tuning was informed by established best practices and empirical knowledge from prior projects. This iterative method facilitated accurate hyperparameter calibration, yielding a balanced model proficient in learning from training data and generalizing to novel data.

Hyperparameter optimization was conducted in two separate phases to guarantee robust model performance. A preliminary grid search was performed to investigate broad combinations of essential parameters, encompassing batch sizes (16, 32, 64), optimizer types (Adam, SGD, RMSprop), and learning rates ranging from 1e-5 to 1e-2 on a logarithmic scale. Accuracy and F1-score of the validation set were employed as performance indicators to inform selection. This initial phase facilitated the establishment of robust training setups for each architecture: Adam was chosen for ViT due to its efficacy with sparse gradients, whilst SGD demonstrated more stability in convergence for EfficientNetB7 during prolonged epochs. To mitigate overfitting, early stopping was implemented, ceasing training when the validation loss did not enhance after a specified number of epochs.

Subsequent to grid search, Bayesian optimization was utilized for refinement. A Gaussian Process prior utilizing an Upper Confidence Bound (UCB) acquisition function was employed to direct the search, with parameters kappa = 10 and $x_i = 0.1$, prioritizing exploration. The search space encompassed learning rate, dropout rate, weight decay, and momentum, with the optimization target focused on maximizing the F1-score on the validation set. Following 50 iterations, the optimal configurations were evaluated (refer to Table 2), leading to the selection of the most effective option for EfficientNetB7, which comprised a learning rate of 1e-5, a batch size of 32, a dropout rate of 0.4, and a weight decay of 1e-4.

This meticulous method of hyperparameter tuning demonstrates a methodical, evidence-driven strategy designed to achieve optimal model performance while guaranteeing generalizability and dependability.

Table 2. List of tunable hyperparameters and their values

Tunable hyperparameters	Lower bound	Upper bound	Scale	Chosen value
batch size	2^3	2^6	Logarithmic	32 for EfficientNetB7, 64 for ViT
max LR	10^{-5}	10^{-2}	Logarithmic	10^{-5}
base LR/max LR	10^{-2}	10^{0}	Logarithmic	10^{-1}
max momentum	0.85	0.99	Linear	0.9 for EfficientNetB7, 0.85 for ViT
base momentum	0.85 for EfficientNetB7, 0.8 for ViT	0.9 for EfficientNetB7, 0.85 for ViT	Linear	0.8
weight decay	10^{-5} for EfficientNetB7, 10^{-4} for ViT	10^{-4} for EfficientNetB7, 10^{-2} for ViT	Logarithmic	10^{-4} for EfficientNetB7, 10^{-3} for ViT

2.5. Model Training

Upon finalizing hyperparameter tuning, a designated set of training parameters was employed to establish the training loop for each model, encompassing the choice of an optimizer and a cyclic learning rate scheduler, as depicted in Fig. 6. The training set was utilized to improve the model's learning, while the validation set was employed to detect overfitting and initiate early stopping as required. The validation set facilitated the identification of the most appropriate model among two prospective candidates.

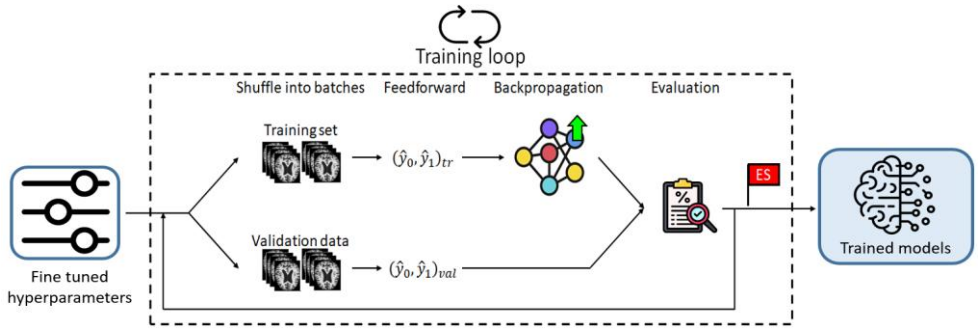


Fig.6. Schematic diagram of training loop

To avert overfitting and optimize computing resources, early stopping was employed, terminating the training process when performance no longer enhanced. After identifying the optimal model, additional adjustments were implemented by modifying the decision threshold. The F1-score is notably resilient among threshold-dependent measures, successfully mitigating challenges associated with imbalanced datasets. The F1-score, which is based on Precision and Recall, can be effectively represented on the same graph as the Precision-Recall curve.

Moreover, modifying the threshold does not affect the model's predictive ability, as the foundational learning is not contingent upon any specific threshold value. Threshold-dependent measures fluctuate solely based on the selected threshold, enabling precise adjustments to the model's performance as required.

2.6. Model Evaluation for Web Platform Deployment

The choice of the optimal model for clinical implementation entailed a thorough assessment of EfficientNetB7 and ViT concerning diagnostic accuracy, computational efficiency, generalization ability, and practical applicability for diagnosing AD using brain MRI images.

The model selection procedure commenced with an extensive evaluation of each model's performance based on critical classification metrics: accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics are essential in medical diagnostics, since they evaluate the precision in distinguishing disease phases and the equilibrium between sensitivity and specificity.

A generalization analysis was performed on an independent test set, which was not utilized during training, to assess robustness. Uniformity in performance among training, validation, and test sets was a key measure of each model's generalization capacity; substantial performance declines between these sets would imply restricted applicability in practical scenarios.

In clinical applications, where swift and accurate diagnosis is crucial, computing efficiency and inference duration are vital. Both models were evaluated for inference speed and resource demands to ascertain their appropriateness for real-time implementation.

The models were also assessed for their capacity to generate confidence scores in conjunction with predictions. These scores improve interpretability, enabling healthcare practitioners to assess the dependability of forecasts and make educated diagnostic judgements.

3. Results and Discussion

3.1. Model Training Results

The training procedure for both EfficientNetB7 and ViT utilized the preprocessed dataset comprising MRI images categorized into four classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This section provides comprehensive findings from the training process, emphasizing critical performance indicators, convergence behaviors, and the impact of hyperparameter adjustment on model efficacy.

Both models were assessed on training and validation datasets, utilizing early stopping techniques to avert overfitting and guarantee optimum generalization.

A. Training Accuracy and Loss

Both models underwent training for specified epochs (EfficientNetB7: 50, ViT: 80) with early termination contingent upon validation loss. The learning curves for accuracy and loss (Fig. 7 and Fig. 8) demonstrated divergent patterns. EfficientNetB7 exhibited a gradual rise in training accuracy, culminating at 95.2%, alongside consistent enhancement in validation accuracy (93.5%) and a concomitant loss of 0.23. This suggests successful learning, probably facilitated by data augmentation, regularization, and early stopping, which mitigated overfitting. Conversely, ViT exhibited more erratic learning curves, with a peak training accuracy of 92.1% and a validation accuracy of 87.4%. The observed oscillations indicate potential overfitting and difficulties with generalization, likely attributable to ViT's sensitivity to hyperparameters and its patch-based design, which may inadequately capture fine-grained characteristics in MRI scans compared to EfficientNetB7.

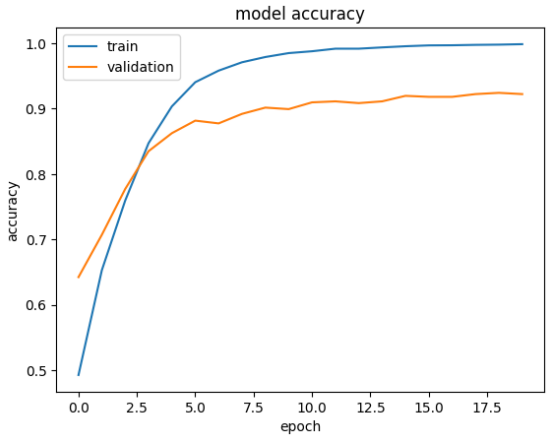


Fig.7. Model training performance of EfficientNetB7

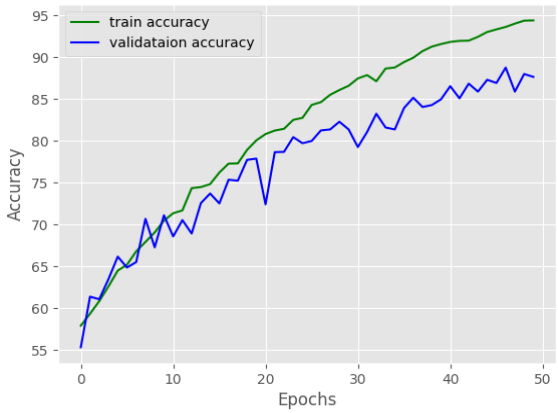


Fig.8. Model training performance of ViT

B. Impact of Hyperparameter Tuning

Hyperparameter adjustment was essential for enhancing the performance of both models. A grid search determined the ideal combination of learning rate, batch size, and number of epochs. EfficientNetB7 demonstrated robustness to hyperparameter variations, sustaining consistent performance with a learning rate of 1×10−5 and a batch size of 32. The stability was further augmented by the dynamic learning rate adjustment approach, ReduceLROnPlateau, which strengthened generalization by decreasing the learning rate according to validation results. Conversely, ViT demonstrated greater sensitivity to hyperparameters, requiring an optimal configuration of a larger batch size (e.g., 64) and a learning rate of 3×10−5. Notwithstanding these modifications, ViT's learning curves exhibited greater variability, suggesting that future enhancements, such as modifying patch sizes or integrating supplementary data augmentation, may be necessary to elevate its efficacy on intricate medical images, such as brain MRIs.

C. Generalization and Overfitting

Both models utilized early stopping contingent on validation loss to avert overfitting, and their generalization abilities were evaluated by contrasting training and validation metrics. EfficientNetB7 exhibited robust generalization with negligible overfitting, evidenced by the proximity of training (95.2%) and validation accuracy (93.5%) and a minimal disparity in loss. The performance was ascribed to the balanced dataset, efficient data augmentation, and the model's scalable architecture. The ViT exhibited a significant disparity between training accuracy (92.1%) and validation accuracy (87.4%), suggesting a likelihood of overfitting. The oscillations in validation loss indicate that ViT's patch-based methodology restricts its capacity to acquire intricate information in brain MRI images, underscoring the necessity for enhancements in its generalization capabilities.

D. Discussion on Model Training Dynamics

The training outcomes underscore the advantages and drawbacks of each model's architecture. EfficientNetB7 exhibited stability and consistency, with elevated accuracy and little loss in both training and validation datasets. Its capacity for effective scaling and the ability to capture complex elements in medical images rendered it particularly suitable for this application. Moreover, EfficientNetB7's robustness to hyperparameter fluctuations, along with proficient data augmentation and regularization, highlights its practical use in medical diagnostics, where scalability and dependability are essential.

Conversely, the ViT, notwithstanding its novel self-attention mechanism, demonstrated less consistent learning dynamics. The transformer-based technique shows potential for tasks requiring the capture of global dependencies; however, additional refining may be necessary for MRI-based AD diagnosis, which necessitates more intricate feature extraction. The inconsistency in training and validation performance indicates that ViT may not be fully optimized for this particular activity. Future research may investigate modifications in patch size or hybrid models that combine CNNs and transformers to capitalize on the advantages of both architectures.

In conclusion, both models showed robust skills in Alzheimer's disease diagnosis; however, EfficientNetB7 consistently surpassed ViT in training stability, generalization, and overall performance. The findings indicate that EfficientNetB7 is more appropriate for clinical application, where accuracy, robustness, and training simplicity are critical.

3.2. Model Evaluation Results

Following the training procedure, both models - EfficientNetB7 and ViT - were assessed on the reserved test set to determine their diagnostic efficacy. A thorough assessment was performed utilizing various metrics, such as accuracy, precision, recall, F1-score, and confusion matrices, to deliver an in-depth comprehension of each model's advantages and drawbacks in detecting AD from brain MRI data.

A. Accuracy

Accuracy, defined as the proportion of correctly categorized cases to the total number of instances, served as a major parameter for assessing overall model performance. EfficientNet and ViT exhibited significant accuracy on the test set, with EfficientNetB7 attaining 93.5% and ViT earning 87.4%. The enhanced precision of EfficientNetB7 is due to its capacity to optimize both the depth and width of the network, facilitating a more effective representation of the intricate features in brain MRI scans. Conversely, ViT, despite its novel transformer-based design, failed to achieve the performance of EfficientNet, possibly because of its dependence on patch-based attention processes, which may not have adequately captured the intricate features crucial for identifying early stages of AD.

B. Precision, Recall, and F1-Score

To evaluate the diagnostic efficacy of the models, precision, recall, and F1-score were computed for each AD classification: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. These measurements are essential in medical diagnostics, as false positives and false negatives can result in significant clinical ramifications. EfficientNetB7 attained a precision of 94% in detecting Very Mild Demented instances, demonstrating a robust capacity to reduce false positives, whereas ViT exhibited a lesser precision of 89%, suggesting possible misclassification. EfficientNetB7 achieved a recall of 92.8% for Mild Demented, effectively identifying the majority of true positives, whereas ViT recorded a recall of 85.6%. The F1-score for EfficientNetB7 was 93.2%, surpassing ViT's 87.1%, hence emphasizing EfficientNetB7's superior equilibrium of precision and recall. These findings underscore the exceptional reliability of EfficientNetB7 for clinical use, especially in the early identification of Alzheimer's disease, where precision and the reduction of diagnostic errors are critical.

C. Confusion Matrix

The confusion matrix for each model offered further insights on classification efficacy across the four dementia stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The confusion matrix for EfficientNetB7 (Fig. 9) demonstrated a significant accuracy in differentiating between the early stages of dementia (Very Mild Demented and Mild Demented) and the Non-Demented class. Early diagnosis of Alzheimer's disease is crucial in clinical applications, since it enables prompt therapies that may decelerate disease progression. The confusion matrix for ViT (Fig. 10) indicated a greater frequency of misclassifications between the Very Mild Demented and Non-Demented classes, perhaps restricting its therapeutic applicability in identifying the initial indications of Alzheimer's Disease. The confusion matrix indicated that ViT misclassified 12% of Very Mild Demented instances as Non-Demented, whereas EfficientNetB7 misclassified 5%. This mismatch suggests that ViT may have difficulty detecting subtle, early-stage symptoms that EfficientNetB7 can more effectively identify.

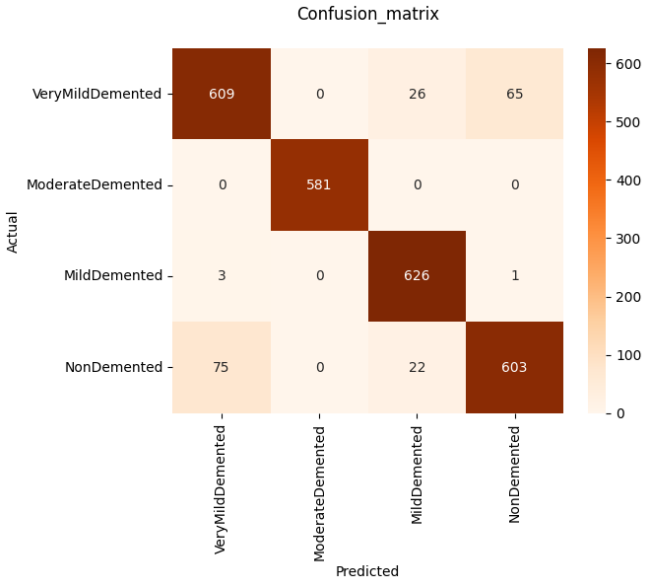


Fig.9. The confusion matrix of EfficientNetB7

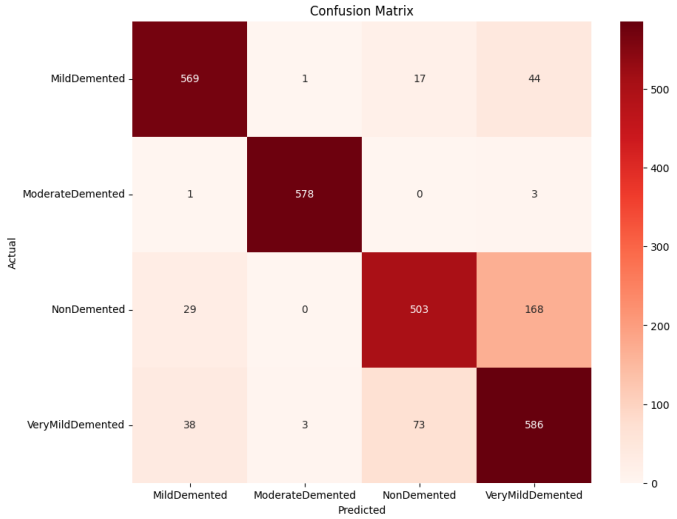


Fig.10. The confusion matrix of ViT

D. Comparative Discussion

EfficientNetB7 surpassed ViT in all parameters, including accuracy, precision, recall, and F1-score, indicating its superiority in diagnosing Alzheimer's Disease. Its optimized architecture effectively scales depth, width, and resolution, extracting significant characteristics from MRI images with low processing costs, making it suitable for real-time clinical applications. Conversely, ViT's patch-based methodology had challenges in the early diagnosis of Alzheimer's disease, underscoring the necessity for enhancements. EfficientNetB7's performance and scalability provide it with a dependable and effective solution for clinical implementation, especially in settings necessitating precise and prompt diagnosis.

A summary of the comparative performance and architectural aspects of the two models was offered in Table 3 for clarity. This table emphasizes critical factors including accuracy, training stability, sensitivity to dataset size, computing efficiency, and deployment appropriateness. The comparison aims to synthesize empirical findings and architectural insights into a useful guide for choosing between CNN-based and transformer-based methods in medical image classification, especially in data-limited situations.

Table 3. Comparative analysis of EfficientNetB7 (CNN-based) and Vision Transformer (ViT)

Aspect	**EfficientNetB7 (CNN)**	**Vision Transformer (ViT)**
Architecture Type	Convolutional Neural Network	Transformer-Based Vision Model
Accuracy	93.5%	88.7%
F1-score (avg)	>92% across classes	Lower F1-score in early-stage classes
Training Stability	High (smooth convergence)	Less stable; slower convergence
Sensitivity to Dataset Size	Low (performs well on small datasets)	High (needs large datasets for optimal performance)
Training Time (epochs)	<20 epochs to convergence	>30 epochs to convergence
Computational Load	Lower (faster and less GPU-intensive)	Higher (longer training and more GPU usage)
Interpretability	Compatible with feature maps and CAMs	Requires attention map visualization
Deployment Suitability	Strong candidate for real-time, clinical applications	Requires optimization for practical deployment

E. Limitations and Future Considerations

A shortcoming of this evaluation is the exclusive dependence on MRI scans for Alzheimer's disease diagnosis. Future research may integrate multimodal data, including PET scans and clinical histories, to enhance diagnostic precision. Another drawback of the dataset is that it does not have demographic information like age, gender, or ethnicity, which makes it hard to see how well the model works for different groups of people. Future research should utilize datasets with more extensive metadata to enhance clinical validation. Moreover, although EfficientNetB7 exhibited robust performance in this study, subsequent research could investigate the incorporation of hybrid models that amalgamate the advantages of both CNNs and transformers, potentially enhancing both fine-grained and global feature extraction. In conclusion, EfficientNetB7 surpassed ViT across all assessed criteria, rendering it a more suitable choice for clinical implementation in the early identification of Alzheimer's Disease. This evaluation's results establish a robust basis for the continued development and enhancement of AI-driven diagnostic tools for neurodegenerative illnesses.

Notwithstanding the application of class weighting and oversampling, the dataset's class imbalance presented difficulties for model generalizability. Although methods such as data augmentation, stratified sampling, and customized loss adjustments enhanced performance, under-represented classes, especially Moderate Demented, exhibited somewhat

worse accuracy owing to insufficient training samples. This disparity may impede the model's dependability in wider clinical contexts where balanced datasets are infrequent. Subsequent research should tackle this issue by broadening datasets throughout all phases or using synthetic data generation methods, such as GANs, to produce authentic samples for under-represented classes. While existing solutions somewhat alleviate imbalance effects, careful interpretation is recommended for less common stages, and additional validation on bigger, balanced datasets is essential for reliable, generalized performance.

The application of MRI-based categorical labeling devoid of biomarker data, including Aβ42, tau, or FDG-PET, constrains the congruence of our study with contemporary Alzheimer's disease diagnostic standards. Future research should integrate biomarker-driven multi-modal datasets and conform to established frameworks like the NIA-AA and ATN system.

This investigation was constrained to two-dimensional axial slices owing to the nature of the publicly accessible dataset. As a result, architectures intended for 3D volumetric data (e.g., 3D CNNs) were not investigated. Subsequent research should utilize volumetric MRI datasets to evaluate the efficacy of these structures. The limited dataset size may have impeded the Vision Transformer's learning ability, as ViT models generally necessitate substantially more data for effective generalization.

Moreover, the effective clinical implementation of the suggested AI system must tackle two principal operational problems. Continuous monitoring of model drift is essential to identify alterations in data distribution that could impact model reliability. This entails establishing procedures for retraining or recalibrating the model with fresh clinical data and employing statistical tests to identify distributional alterations in input features. A strong framework for clinician-AI collaboration is crucial. The AI tool must operate as a decision support system, delivering clear forecasts with accompanying confidence scores and visual elucidations (e.g., saliency maps) to facilitate interpretation. The ultimate diagnostic duty must reside with healthcare professionals to guarantee safety and confidence in clinical decision-making.

F. Model Deployment Results

After the model evaluation phase, the top-performing model, EfficientNetB7, was chosen for implementation in a web-based application to enable its practical application in clinical environments. The deployment method sought to establish an intuitive interface for healthcare practitioners to upload MRI images and obtain real-time diagnostic data. The implementation of the ViT was examined; however, its difficulties in web-based integration prompted the choice to prioritize EfficientNetB7 as the principal model. This section examines the deployment outcomes, technical factors, and practical ramifications of incorporating AI into healthcare workflows.

The web-based application created in this study is exclusively designed for research demonstration and instructional purposes. It has not undergone clinical validation or received approval from regulatory agencies such as the FDA or equivalent entities. Therefore, it is not intended for application in actual clinical environments without additional validation. All diagnostic recommendations produced by the system should be approached with caution and understood under the supervision of certified healthcare practitioners. Future efforts will concentrate on securing ethical permission, executing clinical validation studies, and verifying adherence to medical device laws before any clinical integration.

G. Web Application

EfficientNetB7 was implemented as a web-based diagnostic instrument utilizing Flask, offering a user-friendly and secure interface for healthcare practitioners (Fig. 11). The frontend, constructed using HTML, CSS, and JavaScript, enabled users to upload MRI scans, presenting real-time classifications of Alzheimer's Disease alongside confidence scores. The backend, utilizing Flask, processed images through TensorFlow-integrated EfficientNetB7, while a MySQL database securely held patient information, ensuring confidentiality with user management functionalities.

The implementation of EfficientNetB7 in the web application exhibited robust performance, producing predictions within 1-2 seconds per image on a normal server setup. The swift response time guarantees the system's smooth integration into clinical workflows, preventing delays in the diagnostic process. The rapid inference performance is due to EfficientNetB7's computational efficiency, which balances model complexity with prediction latency. Moreover, the implementation of batch normalization and model optimization strategies during deployment guaranteed that EfficientNetB7 preserved its performance from the training environment upon transitioning to the online platform.

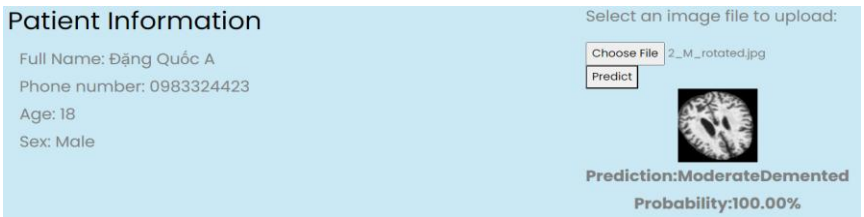


Fig.11. Demonstration of website interface and prediction result

The model preserved its accuracy from the evaluation phase, providing consistent diagnostic outcomes across the test cases submitted during deployment testing. The confidence scores associated with each prediction offered further value, enabling healthcare practitioners to evaluate the trustworthiness of the model's predictions, even in ambiguous circumstances like differentiating between Very Mild Demented and Non-Demented individuals. This further layer of interpretability improves the application's practicality in actual clinical decision-making.

The deployment of the ViT encountered considerable difficulties, in contrast to EfficientNetB7, which integrated effortlessly and operated well inside the Flask context. First, ViT's dependence on patch-based segmentation necessitated particular preprocessing setups that were challenging to reproduce reliably within the web application, leading to diminished inference performance. Secondly, the transformer architecture of ViT resulted in increased processing demands, causing prolonged inference durations (4-5 seconds per image) and necessitating considerable server resources. This elicited scalability apprehensions, especially in resource-constrained environments such as rural hospitals. These constraints hindered the introduction of ViT as the principal model. Nonetheless, with preprocessing optimization and diminished model complexity, ViT continues to be a viable contender for prospective applications.

Although the ViT demonstrated commendable accuracy in categorizing phases of Alzheimer's Disease, its implementation encountered obstacles owing to substantial computational requirements and certain preprocessing necessities for its patch-based input segmentation. Enhancing ViT for practical clinical applications necessitates the resolution of these limitations. One method involves parameter optimization, including the modification of patch size, learning rate, or the quantity of transformer layers. Reduced patch sizes may capture localized features with less processing cost, hence improving sensitivity to tiny MRI alterations. Model simplification methods, such as pruning or knowledge distillation, provide an alternative approach. Pruning eliminates superfluous layers or neurons, hence decreasing computing demands without significant performance degradation. Knowledge distillation instructs a smaller "student" model to emulate the bigger ViT, rendering it appropriate for resource-limited settings. Alternatively, transformer architectures such as Swin Transformers, which employ hierarchical structures and window-based attention, may more effectively capture multi-scale characteristics while necessitating fewer resources. These optimizations may harmonize efficiency and accuracy, facilitating broader clinical implementation of transformer-based AI systems.

The implementation of EfficientNetB7 as a web-based diagnostic instrument presents considerable benefits for healthcare environments, delivering a dependable and instantaneous solution that necessitates minimal AI training. The easy user interface integrates effortlessly into current procedures, enabling healthcare practitioners to upload MRI scans and obtain diagnostic insights without manual analysis. Furthermore, the model's swift inference time and exceptional accuracy render it suitable for high-demand settings such as outpatient clinics and emergency departments. The integrated MySQL database enhances functionality by providing secure storage for diagnostic histories, so improving progress tracking and collaboration with professionals. Moreover, confidence scores augment trust and transparency, enabling doctors to assess diagnostic reliability and make educated decisions, such as ordering additional tests when confidence is diminished.

The implementation of EfficientNetB7 was effective; nonetheless, various restrictions require consideration. The application presently processes individual MRI images and lacks batch processing features, potentially impeding scalability in high-volume environments. Enhancing batch processing optimization and minimizing inference time could increase its efficacy. The algorithm depends exclusively on MRI images for AD diagnosis; including multimodal data such as PET scans or clinical records could enhance accuracy and broaden its clinical relevance. Moreover, although the online deployment guarantees accessibility, empirical testing under multiple situations, encompassing different MRI scanners and workflows, is crucial to confirm robustness and usability.

The use of EfficientNetB7 as a web-based diagnostic instrument for Alzheimer's Disease has demonstrated feasibility and efficacy. The program delivers swift and dependable diagnostic insights, allowing healthcare practitioners to make informed decisions instantaneously. Despite the constraints associated with the deployment of the ViT model, the performance and scalability of EfficientNetB7 render it a compelling candidate for integration into clinical practice. Future endeavors will concentrate on rectifying the observed constraints, augmenting the application's scalability, and integrating other data modalities to further enhance diagnostic precision. This deployment signifies a substantial advancement in the actual application of AI in medical diagnostics, potentially enhancing patient outcomes in the identification of AD.

H. Prediction Results

To assess the empirical accuracy of the models, ten images per category were employed for prediction evaluation within the application. Table 4 summarizes the results of each test, demonstrating the application's comparative diagnostic effectiveness.

The results indicate that both models exhibit very competitive accuracies, with EfficientNetB7 showing exceptional performance, hence designating it as the favored diagnostic instrument. Moreover, real-world diagnostic evaluations, as outlined in Table 4, highlighted the models' proficiency in successfully diagnosing the majority of classes, even with unseen data. Nonetheless, despite ViT's praiseworthy training efficacy, it faced challenges upon online deployment, hindering its diagnostic capabilities. As a result, the model's predictive skills remained untestable, despite its recorded high accuracy.

Table 4. Realistic testing results to preliminarily evaluate the effectiveness of the model

Class/Times	1	2	3	4	5	6	7	8	9	10
Non Demented	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
Very Mild Demented	✓	✓	✓	X	✓	✓	✓	✓	✓	X
Mild Demented	✓	✓	✓	✓	✓	✓	X	✓	✓	✓
Moderate Demented	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

3.3. Discussion

This study highlights how AI can help with early and precise AD diagnosis using brain MRI. The study illustrates the advantages and disadvantages of two sophisticated deep learning models in the context of medical image analysis by contrasting them: ViT and EfficientNetB7. The basis for incorporating deep learning systems into clinical diagnostic workflows is provided by this comparative analysis. The results have wide-ranging effects on algorithm development, practical implementation, and upcoming developments in AI-assisted medical devices.

Across a number of evaluation metrics, such as accuracy, precision, recall, and F1-score, EfficientNetB7 consistently outperformed ViT. Its compound scaling method, which strikes a balance between network width, depth, and resolution, worked especially well for identifying early-stage Alzheimer's symptoms, like those in the Very Mild and Mild Demented classes. This feature emphasizes how useful convolutional neural networks (CNNs) are for tasks that call for the extraction of fine-grained spatial features. A crucial component of clinical reliability, the model also demonstrated strong generalization with little performance variance across training, validation, and testing datasets.

On the other hand, the Vision Transformer encountered difficulties when used to diagnose AD using MRI. ViT's patch-based input representation may mask subtle anatomical cues that are essential for diagnosing Alzheimer's disease, even though its self-attention mechanism is excellent at identifying long-range dependencies in natural images. Additionally, ViT showed training instability and needed more thorough hyperparameter tuning. Furthermore, real-time deployment was hampered by its computational demands, which made it less practical in clinical settings lacking significant hardware resources. Given these drawbacks, hybrid models that combine transformers for global pattern recognition and CNNs for localized feature extraction may provide a more well-rounded approach.

From the standpoint of deployment, EfficientNetB7 was effectively incorporated into a web-based diagnostic tool that provides real-time inference via an intuitive user interface. This demonstrates its suitability for clinical use, particularly in settings with limited resources where speed and ease of use are crucial. On the other hand, ViT's scalability is called into question due to its longer inference time and higher hardware requirements.

However, there were technical issues with the study. Other modalities such as PET scans and clinical data that could improve diagnostic accuracy were not included in the dataset because it was limited to MRI images. Furthermore, performance differed among less-represented classes even though data augmentation assisted in addressing class imbalance. To overcome existing obstacles and improve the clinical viability of AI-driven AD diagnosis tools, future research should concentrate on multimodal data integration, enhanced model interpretability, and the creation of hybrid or lightweight architectures.

Notwithstanding promising outcomes, some restrictions must be recognized. The dataset employed is comparatively small for deep learning, increasing the likelihood of overfitting and constraining the model's capacity to generalize to novel data. Despite the implementation of regularization techniques, data augmentation, and early stopping, overfitting cannot be entirely excluded without validation against an external dataset. The MRI scans were acquired from diverse and undefined locations, potentially introducing scanner-related biases and impacting the consistency of model predictions. Third, the interpretability of models continues to be a significant issue for clinical acceptability. Although convolutional neural networks facilitate visualization methods like class activation maps (CAMs), more efforts are required to incorporate explainability tools (e.g., Grad-CAM, attention maps) into the deployment interface to guarantee that doctors comprehend and trust the model's recommendations. Confronting these issues in forthcoming endeavors will be crucial for progressing towards clinically dependable and ethically robust AI diagnostic instruments.

4. Conclusions

This research examined the utilization of EfficientNetB7 and ViT for the early identification of AD through brain MRI images. EfficientNetB7 consistently surpassed ViT in all essential parameters, including accuracy, precision, recall, and F1-score, showcasing formidable diagnostic powers and excellent generalization, especially in identifying early stages of AD crucial for prompt intervention. Its computational efficiency and smooth integration into a web-based diagnostic tool highlight its appropriateness for practical clinical application. Conversely, ViT, despite its novel self-attention architecture, encountered difficulties in capturing the intricate features crucial for medical imaging, leading to diminished accuracy and prolonged inference durations.

The effective implementation of EfficientNetB7 in an accessible web application underscores the practicality of incorporating AI-driven tools into healthcare procedures. Attributes such as confidence scores improve interpretability, facilitating informed healthcare decision-making. Nonetheless, constraints persist, notably the absence of multimodal data integration and the necessity for empirical validation across varied clinical environments. Future research must

tackle these problems by integrating supplementary data sources, enhancing scalability, and refining transformer-based models for neuroimaging applications. EfficientNetB7 presents a promising approach for the early identification of AD, with the capacity to augment diagnostic precision, diminish variability, and boost patient outcomes via prompt and dependable AI-driven detection. Future initiatives will concentrate on incorporating model drift detection and developing a clinician-centric AI interface that facilitates interpretability, validation, and collaborative decision-making. The proposed AI-based Alzheimer's disease classification system exhibited promising performance in a controlled experimental environment; however, additional measures—such as validation on external datasets, regulatory assessment, and integration into clinical workflows—are required prior to its consideration for real-world clinical application.

Author Contributions Statement

Nhat-Kha Nguyen: Writing (Original Draft): Drafted the initial manuscript, conducted the literature review, and documented the technical background of the study.
Thi-Thu-Hien Pham: Data Curation and Software Development: Performed data acquisition, dataset preprocessing, and implemented the proposed research models.
Nhat-Minh Nguyen: Model Training, Validation, and Performance Evaluation: Led the model training process, validated the results using standard evaluation metrics, and benchmarked performance against existing methods.
Tan-Nhu Nguyen: Writing (Review and Editing) and Project Management: Reviewed and edited the manuscript for clarity and coherence, and coordinated project milestones and timelines.
Ngoc-Bich Le: Conceptualization, Methodology, Supervision, and Formal Analysis: Conceived the research idea, designed the overall methodology and framework, supervised project execution, performed in-depth result analysis, prepared visualizations, and ensured the statistical robustness of the evaluation.

All authors have read and agreed to the published version of the manuscript.

Conflicting Interests Statement

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

Data underlying the results presented in this paper are publicly available on Kaggle [https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images] from [42].

Ethical Declarations

Based on the Alzheimer's dataset (4 class of Images) 2019 dataset published online for the academic community, there is no cause for concern regarding ethical considerations in this study.

Declaration of Generative AI in Scholarly Writing

Generative AI tools were used solely for language refinement and clarity enhancement. The authors remain fully responsible for the scientific content and conclusions of this manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

AI - Artificial Intelligence
AD - Alzheimer's Disease
AUC - Area Under the Curve
CAM - Class Activation Map
CNN - Convolutional Neural Network
CN - Cognitively Normal
CSF - Cerebrospinal Fluid
DL - Deep Learning
FN - False Negative
FP - False Positive
F1-score - Harmonic Mean of Precision and Recall
GPU - Graphics Processing Unit
LR - Learning Rate
MD - Mild Demented
MCI - Mild Cognitive Impairment

MRI - Magnetic Resonance Imaging
ND - Non-Demented
PCA - Principal Component Analysis
PET - Positron Emission Tomography
ReLU - Rectified Linear Unit
SVM - Support Vector Machine
TCN - Temporal Convolutional Network
TN - True Negative
TP - True Positive
UCB - Upper Confidence Bound
ViT - Vision Transformer
VMD - Very Mild Demented

Reference

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[2] A. Feroui, M. Saim, M. E. A. Lazouni, S. A. Lazzouni, Z. A. Elaouaber, and M. Messadi, "Improved diagnosis of lung cancer classification based on deep learning method," *Int. J. Biomed. Eng. Technol.*, vol. 46, no. 2, pp. 138–159, 2024, doi: 10.1504/IJBET.2024.141571.

[3] C. Hu, "An investigation into deep learning for the analysis of medical images," *Applied and Computational Engineering*, vol. 54, no. 1, pp. 7–13, Mar. 2024, doi: 10.54254/2755-2721/54/20241121.

[4] F. R. da Silva, R. Camacho, and J. M. R. S. Tavares, "Federated Learning in Medical Image Analysis: A Systematic Survey," *Electronics (Basel).*, vol. 13, no. 1, p. 47, Dec. 2023, doi: 10.3390/electronics13010047.

[5] P. R. Kharote, M. S. Sankhe, and D. Patkar, "Automated detection and grading of prostate cancer in multiparametric MRI," *Int. J. Biomed. Eng. Technol.*, vol. 40, no. 4, p. 372, 2022, doi: 10.1504/IJBET.2022.128088.

[6] G. K. Mourya, S. Paul, A. Handique, U. Baid, P. V. Dutande, and S. N. Talbar, "Modified U-Net for fully automatic liver segmentation from abdominal CT-image," *Int. J. Biomed. Eng. Technol.*, vol. 40, no. 1, p. 1, 2022, doi: 10.1504/IJBET.2022.125099.

[7] T.-N. Vo, N.-B. Le, Q.-H. Phan, T.-H. Le, and T.-T.-H. Pham, "Detection of COVID-19 from chest X-rays using deep transfer learning," *Health Informatics J.*, vol. 30, no. 4, Oct. 2024, doi: 10.1177/14604582241290724.

[8] B. H. T. Nguyen, B. N. Le, N. H. Trinh, H. T. Le, and T. H. T. Pham, "Breast cancer diagnosis based on detecting lymph node metastases using deep learning," *Science and Technology Development Journal*, 2022, doi: 10.32508/stdj.v25i2.3894.

[9] Z. Lu, G. Qiao, and S. Zhang, "Research Advanced in Technology and Applications of Image Recognition Based on Deep Learning," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 418–425, Mar. 2024, doi: 10.54097/5zyqe776.

[10] A. Gupta, D. Gupta, M. Pathak, and S. K. Wagh, "The advancement of ensemble deep learning architecture for the detection and classification of brain tumours with MRI images," *Int. J. Biomed. Eng. Technol.*, vol. 45, no. 1, pp. 27–44, 2024, doi: 10.1504/IJBET.2024.138619.

[11] H. Shi, N. Zhang, X. Wu, and Y. Zhang, "Multimodal lung tumor image recognition algorithm based on integrated convolutional neural network," *Concurr. Comput.*, vol. 32, no. 21, Nov. 2020, doi: 10.1002/cpe.4965.

[12] J. Huang, "COVID-19 CT Image Detection Based on Convolutional Neural Network," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, IEEE, Dec. 2023, pp. 532–541. doi: 10.1109/ITAIC58329.2023.10409115.

[13] R. Murugesan, G. A. T., V. A., B. Azhaganathan, and P. D. D. Domnic, "Instance segmentation of neuronal cells using U-Net, mask R-CNN, and Detectron2," *Int. J. Biomed. Eng. Technol.*, vol. 45, no. 2, pp. 129–149, 2024, doi: 10.1504/IJBET.2024.138735.

[14] A. Hechri, A. Hamed, and A. Boudaka, "An efficient brain tumour segmentation approach using cascade convolutional neural networks," *Int. J. Biomed. Eng. Technol.*, vol. 44, no. 3, pp. 226–241, 2024, doi: 10.1504/IJBET.2024.137343.

[15] N. Ambily and K. Suresh, "MRI segmentation using deep learning network for brain tumour detection," *Int. J. Biomed. Eng. Technol.*, vol. 43, no. 4, pp. 378–389, 2023, doi: 10.1504/IJBET.2023.135398.

[16] Basavaraj Mali Patil, Megha Rani Raigonda, Sudhir Anakal, Ambresh Bhadrashetty, "Early Detection of Dementia using Deep Learning and Image Processing", International Journal of Engineering and Manufacturing, Vol.13, No.1, pp. 14-22, 2023.

[17] Markos Wondim Walle, Kula Kakeba Tune, Natnael Tilahun Sinshaw, Sudhir Kumar Mohapatra, "Transfer Learning based Breast Cancer Classification via Deep Convolutional Neural Network", International Journal of Engineering and Manufacturing, Vol.13, No.4, pp. 34-43, 2023.

[18] Prateek Keserwani, V. S. Chandrasekhar Pammi, Om Prakash, Ashish Khare, Moongu Jeon,"Classification of Alzheimer Disease using Gabor Texture Feature of Hippocampus Region", International Journal of Image, Graphics and Signal Processing, Vol.8, No.6, pp.13-20, 2016.

[19] Anil Kumar Pallikonda, P Suresh Varma, B. Vivekanandam, "Mask Region-based Convolution Neural Network (Mask R-CNN) Classification of Alzheimer's Disease Based on Magnetic Resonance Imaging (MRI)", International Journal of Image, Graphics and Signal Processing, Vol.15, No.6, pp. 54-66, 2023.

[20] Michael Chi Seng Tang, Huong Yong Ting, Abdulwahab Funsho Atanda, Kee Chuong Ting, "Transfer Learning with EfficientNetV2 for Diabetic Retinopathy Detection", International Journal of Engineering and Manufacturing, Vol.14, No.6, pp. 54-60, 2024.

[21] Rathin Halder, Nusrat Sharmin, "Brain Ischemic Stroke Detection through Deep Learning: A Systematic Review on CT vs MRI vs CTA Images", International Journal of Education and Management Engineering, Vol.14, No.5, pp. 23-34, 2024.

[22] Saloni, R. K. Sharma, Anil K. Gupta,"Classification of High Blood Pressure Persons Vs Normal Blood Pressure Persons Using Voice Analysis", International Journal of Image, Graphics and Signal Processing, vol.6, no.1, pp.47-52, 2014.
[23] Hossein. Abbasi, Ahmed. Alshaeeb, Yasin. Orouskhani, Behrouz. Rahimi, Mostafa. Shomalzadeh, "Advanced Deep Learning Models for Accurate Retinal Disease State Detection", International Journal of Information Technology and Computer Science, Vol.16, No.3, pp.61-71, 2024.
[24] B. Vijayakumari, S. Vidhya, and J. Saranya, "Deep learning-based gender classification with dental X-ray images," *Int. J. Biomed. Eng. Technol.*, vol. 42, no. 1, pp. 109–121, 2023, doi: 10.1504/IJBET.2023.131694.
[25] Ahmad Ilham Gustisyaf, Ardiles Sinaga, " Implementation of Convolutional Neural Network to Classification Gender based on Fingerprint ", International Journal of Modern Education and Computer Science Vol.13, No.4, pp. 55-67, 2021.
[26] Ishwari Singh Rajput, Sonam Tyagi, Aditya Gupta, Vibha Jain, "Advances in Medical Imaging: Using Convolutional Neural Networks for White Blood Cell Identification", International Journal of Image, Graphics and Signal Processing, Vol.16, No.1, pp. 108-125, 2024.
[27] Abdullah Walid, Md. Mehedi Hasan, Tonmoy Roy, Md. Selim Hossain, Nasrin Sultana, "Deep Learning-Based Potato Leaf Disease Detection Using CNN in the Agricultural System", International Journal of Engineering and Manufacturing, Vol.13, No.6, pp. 9-22, 2023.
[28] Q.-H. Phan, V.-T. Nguyen, C.-H. Lien, T.-P. Duong, M. T.-K. Hou, and N.-B. Le, "Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models," *Plants*, vol. 12, no. 4, p. 790, Feb. 2023, doi: 10.3390/plants12040790.
[29] Banothu Balaji, T. Satyanarayana Murthy, Ramu Kuchipudi, "A Comparative Study on Plant Disease Detection and Classification Using Deep Learning Approaches", International Journal of Image, Graphics and Signal Processing, Vol.15, No.3, pp. 48-59, 2023.
[30] S. Panigrahi, U. S. N. Raju, D. Pathak, K. V. Kadambari, and H. Ala, "Rapid detection of COVID-19 from chest X-ray images using deep convolutional neural networks," *Int. J. Biomed. Eng. Technol.*, vol. 41, no. 1, p. 1, 2023, doi: 10.1504/IJBET.2023.128510.
[31] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019.
[32] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020.
[33] Z. Cui, Z. Gao, J. Leng, T. Zhang, P. Quan, and W. Zhao, "Alzheimer's Disease Diagnosis Using Enhanced Inception Network Based on Brain Magnetic Resonance Image," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Nov. 2019, pp. 2324–2330. doi: 10.1109/BIBM47256.2019.8983046.
[34] A. Ebrahimi, S. Luo, and R. Chiong, "Deep sequence modelling for Alzheimer's disease detection using MRI," *Comput. Biol. Med.*, vol. 134, p. 104537, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104537.
[35] W. Al Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Comput. Sci.*, vol. 8, p. e1177, Dec. 2022, doi: 10.7717/peerj-cs.1177.
[36] M. Yildirim and A. Cinar, "Classification of Alzheimer's Disease MRI Images with CNN Based Hybrid Method," *Ingénierie des systèmes d information*, vol. 25, no. 4, pp. 413–418, Sep. 2020, doi: 10.18280/isi.250402.
[37] B. A. Mohammed *et al.*, "Multi-Method Analysis of Medical Records and MRI Images for Early Diagnosis of Dementia and Alzheimer's Disease Based on Deep Learning and Hybrid Methods," *Electronics (Basel).*, vol. 10, no. 22, p. 2860, Nov. 2021, doi: 10.3390/electronics10222860.
[38] M. Gharaibeh *et al.*, "Early Diagnosis of Alzheimer's Disease Using Cerebral Catheter Angiogram Neuroimaging: A Novel Model Based on Deep Learning Approaches," *Big Data and Cognitive Computing*, vol. 6, no. 1, p. 2, Dec. 2021, doi: 10.3390/bdcc6010002.
[39] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, IEEE, Apr. 2014, pp. 1015–1018. doi: 10.1109/ISBI.2014.6868045.
[40] Y. Asim, B. Raza, A. K. Malik, S. Rathore, L. Hussain, and M. A. Iftikhar, "A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning," *Int. J. Imaging Syst. Technol.*, vol. 28, no. 2, pp. 113–123, Jun. 2018, doi: 10.1002/ima.22263.
[41] P. S. Sisodia, G. K. Ameta, Y. Kumar, and N. Chaplot, "A Review of Deep Transfer Learning Approaches for Class-Wise Prediction of Alzheimer's Disease Using MRI Images," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2409–2429, May 2023, doi: 10.1007/s11831-022-09870-0.
[42] "Alzheimer's dataset (4 class of Images)." Accessed: Jan. 20, 2022. [Online]. Available: https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images
[43] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

Authors' Profiles



Nhat-Kha Nguyen received his B.S. degree in Biomedical Engineering and Biomechanics at International University, Vietnam National University Ho Chi Minh City, Vietnam, in 2023. His current research interest is wearable devices, applied Artificial Intelligence (AI) models for diseases early detection.



Pham Thi Thu Hien received the B.S. degree in mechatronics from the Ho Chi Minh City University of Technology-Vietnam National University, Ho Chi Minh City, Vietnam, in 2003 and the M.S. and Ph.D. degrees in mechanical engineering from Southern Taiwan University of Technology and National Cheng Kung University, Tainan, Taiwan, in 2007 and 2012, respectively. She is currently a Head of Biomedical Photonics Lab in School of Biomedical Engineering, International University-Vietnam National University HCMC, Ho Chi Minh City, Vietnam. Her research interests are in the areas of polarized light-tissue studies, polarimetry, optical techniques in precision measurement to determine the optical properties of bio-samples (glucose, collagen, and tumor) or applied Artificial Intelligence (AI) models for cancer detection (skin, liver, blood, and breast), noninvasive glucose measurement, cell/tissue characterization, and laser/LED applications in treatment.



Nhat-Minh Nguyen received his B.S. degree in Biomedical Engineering and Biomechanics at International University, Vietnam National University Ho Chi Minh City, Vietnam, in 2023. He is currently full-time master student in the School of Biomedical Engineering, International University, Vietnam National University Ho Chi Minh City, Vietnam. His current research interest is bio 3D printing technology, wearable devices, applied Artificial Intelligence (AI) models for diseases early detection.



Tan-Nhu NGUYEN received a Ph.D. in Biomedical Engineering and Biomechanics at Université de Technologie de Compiègne, France, in 2020. His current research interest is muscle modeling coupled with a serious game for facial rehabilitation. He is currently a full-time lecturer in the School of Biomedical Engineering, International University, Vietnam National University Ho Chi Minh City, Vietnam.



Ngoc-Bich Le received his B.S. degrees at Bach Khoa University, Vietnam, and his Master's and Ph.D. in Mechatronics Science from Southern Taiwan University of Science and Technology – Taiwan in 2004, 2007, and 2010, respectively. He is currently a full-time lecturer in the School of Biomedical Engineering, International University, Vietnam National University Ho Chi Minh City, Vietnam. He published several papers in preferred Journals such as J. Biomedical Microdevices, J. Microfluidics and Nanofluidics, J. Sensors and Actuators, and many Vietnamese Engineering Books in automation, CAD, and mold design. He also presented various academic as well as research-based papers at several national and international conferences. His articles focus on Medical devices, MEMs, Microfluidics, Robotics, and AI.

How to cite this paper: Nhat-Kha Nguyen, Thi-Thu-Hien Pham, Nhat-Minh Nguyen, Tan-Nhu Nguyen, Ngoc-Bich Le, "Transformer-Based vs. CNN-Based Deep Learning for Alzheimer's Disease Classification: Performance and Deployment", International Journal of Intelligent Systems and Applications(IJISA), Vol.18, No.2, pp.44-63, 2026. DOI:10.5815/ijisa.2026.02.04