JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Kurniasari, D., Dwi Pratama, M., Junaidi, A. & Faisol, A. (2025). Transforming Alzheimer's disease diagnosis: Implementing Vision Transformer for MRI images classification. *Journal of Information and Communication Technology, 24*(1), 130-152. https://doi.org/10.32890/jict 2025.24.1.6

Transforming Alzheimer's Disease Diagnosis: Implementing Vision Transformer for MRI Images Classification

*[1]Dian Kurniasari, [2]Muhammad Dwi Pratama, [3]Akmal Junaidi, & [4]Ahmad Faisol
[1,2&4]Department of Mathematics, Universitas Lampung, Indonesia
[3]Department of Computer Science, Universitas Lampung, Indonesia

*[1]dian.kurniasari@fmipa.unila.ac.id
[2]mdwipratama0@gmail.com
[3]akmal.junaidi@fmipa.unila.ac.id
[4]ahmad.faisol@fmipa.unila.ac.id
*Corresponding author

Received: 25/11/2024 Revised: 22/1/2025 Accepted: 26/1/2025 Published: 31/1/2025

ABSTRACT

Alzheimer's disease (AD) is a progressive neurological disorder and the leading cause of dementia, accounting for 60-80% of cases. Early detection of AD is crucial for timely intervention, as the disease significantly impacts cognitive functions and daily activities. Diagnosing AD in clinical settings remains challenging due to subtle early symptoms, diverse presentations, lengthy diagnostic processes, and inconsistent criteria that heavily rely on medical expertise. Accurate and timely diagnosis during the early stages is essential for effective treatment and intervention. Modern imaging techniques, such as Magnetic Resonance Imaging (MRI), have become essential in diagnosing Alzheimer's by providing detailed insights into structural brain changes. This study explores the application of the Vision Transformer (ViT) model for classifying MRI images of Alzheimer's patients, focusing on enhancing accuracy and reliability through data augmentation during pre-processing. A dataset of 8,000 MRI images, categorised into four groups—non-demented, very mild demented, mild demented, and moderate demented—was used to evaluate the ViT model. The experiment achieved promising results, with an accuracy of 98.19%, sensitivity of 96.34%, specificity of 98.80%, and an F1-score of 96.37%. These findings underscore the model's effectiveness in distinguishing between affected and unaffected individuals, minimising misdiagnosis and enabling timely clinical interventions. However, some challenges remain, particularly in the classification between "Non-Demented" and "Very Mild Demented" cases. Future research should focus on enhancing data augmentation techniques and

increasing data diversity in these categories to improve the model's performance further. The ViT model holds great potential for advancing Alzheimer's diagnosis, offering a valuable tool for early detection and intervention in clinical settings.

Keywords: Alzheimer's, deep learning, image classification, MRI, vision transformer.

INTRODUCTION

Alzheimer's disease is a degenerative neurological condition responsible for 60-80% of dementia cases, according to the Alzheimer's Association. The symptoms of this disease vary in severity and progressively hinder an individual's ability to perform daily tasks, manifesting as apathy, depression, communication difficulties, disorientation, impaired judgment, and challenges with swallowing and walking, alongside notable behavioural changes (Abubakar et al., 2022; Alzheimer's Association, 2021). Initially, these symptoms are mild, but they worsen over time, eventually leading to a need for full-time care.

While Alzheimer's disease predominantly impacts individuals aged 65 and older, it is not an inevitable consequence of ageing and can also manifest in younger people. The progression of the disease is typically categorised into four stages. The initial stage, known as Non-Dementia, is characterised by the absence of symptoms or cognitive impairment. This is followed by Very Mild Dementia, marked by early signs such as challenges in recalling the location of everyday objects. The third stage, Mild Dementia, involves noticeable difficulties in retaining information and changes in behavior. The fourth stage, Moderate Dementia, is distinguished by substantial cognitive decline, impairments in speech, personal care activities like bathing and dressing, and is often accompanied by emotional instability and disturbances in sleep patterns (Alzheimer's Association, 2023; Suganthe et al., 2021). It predominantly affects older individuals, with both incidence and prevalence rising as age increases (Li et al., 2022). While this condition is more prevalent in low- and middle-income countries and regions (Gao & Liu, 2021), Nichols et al. (2019) report highlights that Alzheimer's has escalated into a critical global health concern.

According to Nichols et al. (2022), the number of Alzheimer's patients surged to 43.8 million by 2016, representing a 117% increase from 20.3 million in 1990. Data from the World Health Organization (2017) further highlights the global escalation of Alzheimer's cases, with an estimated 46.8 to 50 million individuals diagnosed and 10 million new cases emerging each year. As Alzheimer's disease is strongly associated with ageing, this trend is expected to intensify as populations grow older and life expectancy rises. By 2050, it is projected that 152 million people will be affected by Alzheimer's disease and related dementias (World Alzheimer Report, 2019).

A significant challenge confronting Alzheimer's researcher is the absence of effective treatments for the disease to date (Irankhah, 2020; Pulido et al., 2020). Existing treatments for Alzheimer's disease, including combination therapies, N-methyl-D-aspartate (NMDA) receptor antagonists, and cholinesterase inhibitors, are thought to mitigate or postpone symptom progression (Dafre & Wasnik, 2023). Nevertheless, the clinical diagnosis of Alzheimer's remains challenging due to the subtlety of early symptoms, the heterogeneity in symptom presentation, the protracted detection process, and the variability in diagnostic criteria, which frequently depend on the clinician's expertise. Therefore, prompt and precise diagnosis during this early phase is crucial to enable swift intervention and treatment (Al Rahbani et al., 2024; Helaly et al., 2022).

Recently, there has been an increasing focus on the application of advanced imaging techniques and computational methods to enhance the precision of Alzheimer's disease diagnoses. Modalities such as Magnetic Resonance Imaging (MRI) are pivotal in examining the structural and functional alterations in the brain associated with Alzheimer's (Wong, Xu, Ayoub, et al., 2023; Wong, Xu, Chen, et al., 2023). MRI serves as a prominent neuroimaging tool in the medical field, extensively employed to capture grayscale brain images with varying contrasts. Specifically, MRI is instrumental for diagnostic purposes, image segmentation, and additional applications (Vimala et al., 2023; Zhou et al., 2020).

This imaging technique offers crucial insights into the pathological alterations, such as atrophy and infarction, that manifest in the brain as a result of Alzheimer's disease. Furthermore, deep learning methodologies, particularly Convolutional Neural Networks (CNNs), have demonstrated considerable promise in the analysis of medical images, including the diagnosis of Alzheimer's disease (Chen et al., 2024). CNNs have played a pivotal role in medical image analysis research for several years. Convolutional filters are employed to analyse and extract critical features from medical images. Various studies have implemented CNNs across a range of applications, including tumour detection and classification (Arevalo et al., 2016), skin lesion identification (Azad et al., 2019; Karimijafarbigloo et al., 2023), and brain tumour segmentation (Azad et al., 2022). Furthermore, CNNs have made substantial contributions to the evaluation of various imaging modalities within clinical medicine, encompassing X-ray radiography, Computed Tomography (CT), MRI, ultrasound, and digital pathology.

However, despite their exceptional performance, CNNs exhibit inherent conceptual limitations; they are unable to model explicit long-range dependencies owing to the constraints imposed by the receptive field of the convolutional kernel. As noted by Kim et al. (2020), CNNs also face challenges in accurately detecting specific patterns in designated locations; to address this issue, patching techniques can be integrated into the CNN architecture. This approach enables CNNs to better capture image details by focusing on more minor features such as edges, textures, or patterns within specific regions. Furthermore, this technique must be implemented with precision to align with the architectural requirements of CNNs.

In response to the challenges outlined earlier, extensive research has concentrated on Transformers and their attention mechanisms. These can be understood as a dynamic process of weight adjustment that depends on input features within CNN-based architectures (Azad et al., 2024; Bello et al., 2019; Karimijafarbigloo et al., 2023; Ramachandran et al., 2019; Vaswani et al., 2021). Transformers have demonstrated superiority across a wide range of Natural Language Processing (NLP) applications, including machine translation, text classification, and question-answering (Vaswani et al., 2017). Their remarkable success in NLP has led to the widespread adoption of Transformer architecture within contemporary Computer Vision (CV) models. Since the introduction of Vision Transformers (ViT) (Dosovitskiy et al., 2021), Transformers have emerged as a superior alternative to Convolutional Neural Networks (CNNs) for various tasks, including image recognition, object detection (Zhu et al., 2021), image segmentation, video comprehension, and image super-resolution (Arnab et al., 2021; Azad et al., 2024). A pivotal aspect of the Transformer architecture is its self-attention mechanism, which adeptly models the relationships among sequence elements, thereby facilitating the exploration of long-range dependencies (Azad et al., 2024). This self-attention mechanism allows for the assignment of varying weights to different segments of the input, which are subsequently integrated with patches and positional embeddings (Dosovitskiy et al., 2021).

The primary aim of this study is to assess the efficacy of the ViT architecture in the multiclass classification of MRI images for Alzheimer's disease, specifically distinguishing between Non-Dementia, Moderate Dementia, Mild Dementia, and Very Mild Dementia. The study emphasises enhancing accuracy and reliability through data augmentation during the pre-processing phase. Furthermore, this research seeks to:

- Enhance the dependability of early Alzheimer's disease detection utilising the ViT model.
- Minimise diagnostic errors by improving the precision of MRI image analysis.
- Support clinical practice by delivering effective models for early diagnosis and timely intervention.

RELATED WORK

A patient suspected of having Alzheimer's disease should undergo a comprehensive clinical assessment and detailed medical history review. During this evaluation, healthcare professionals assess the clinical symptoms, conduct interviews with patients and caregivers to gather information about their concerns and analyse alterations in memory, language, and other cognitive abilities. Common cognitive assessments utilised in this context include the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA). Furthermore, the patient's medical history, medication usage, and familial health background are examined to validate the Alzheimer's diagnosis, as numerous other medical conditions may present similar symptoms. Additionally, brain imaging modalities, such as MRI, play a crucial role in diagnosing Alzheimer's disease. However, these imaging techniques can be costly and necessitate specialised equipment and trained professionals. The high expenses are further exacerbated by the need for neuropsychologists or other specialists who are tasked with administering and interpreting the results of these assessments (Breijyeh & Karaman, 2020; Juganavar et al., 2023).

Numerous studies have undertaken efforts over the past few decades to mitigate the substantial costs associated with diagnostic procedures. Recently, researchers have begun using transformer models in this domain to assist in the diagnostic process and facilitate cost reductions. A comprehensive overview of prior research pertaining to Alzheimer's is presented in Table 1.

Table 1

A Comprehensive Overview of Prior Research

Researcher	Data	Research Objective
Lyu et al. (2022)	ImageNet-21K consists of MRI images	Improving brain imaging classification with transfer learning techniques dengan model ViT.
Almurafeh et al. (2023)	MRI images sourced from Kaggle	Classifying MRI images for Alzheimer's disease detection using transformers (ViT).
Shin et al. (2023)	PET 18F-Florbetaben	Evaluating the Effectiveness of ViT and CNN-VGG19 in Classifying Alzheimer's Images (Binary & Ternary).

Researcher	Data	Research Objective
Shaffi et al. (2024)	OASIS & ADNI (MRI)	Enhancing Alzheimer's disease classification using a ViT ensemble.
Tang et al. (2024)	ADNI (PET & MRI)	Advancing Alzheimer's disease diagnosis via multi-modal data integration (PET & MRI).
The proposed studied	MRI images sourced from Kaggle	Precisely identifying and categorising MRI images of Alzheimer's disease in multiclass classification.

Lyu et al. (2022) addressed the challenge of data limitations in brain imaging by implementing a cross-domain transfer learning approach. Their study employed the ViT, initially trained on the ImageNet-21K dataset, before transferring it to a brain imaging dataset. This approach incorporated a slice-wise convolutional embedding technique aimed at enhancing standard patch operations. The findings indicated that this methodology effectively transferred knowledge from the natural imaging domain to brain imaging, achieving classification performance that is on par with recent research efforts. Almufareh et al. (2023) devised machine learning methodologies for the detection of Alzheimer's disease through the analysis of neuroimaging data, particularly utilising MRI scans. Their research employed attention-based mechanisms alongside a ViT approach, commencing with the pre-processing of MRI images prior to classification by the network. The model was trained on publicly available datasets from Kaggle, achieving remarkable results with an accuracy of 99.06%, precision of 99.06%, recall of 99.14%, and an F1-score of 99.1%. Furthermore, the study included comparative analyses demonstrating that this approach outperformed alternative techniques, establishing it as an effective model for the expedited and precise diagnosis of Alzheimer's disease, thereby enhancing the quality of life for patients.

Shin et al. (2023) used a technique for detecting dementia-related images using the ViT on PET scans with 18F-Florbetaben, comparing its performance to that of the CNN VGG19 model. The ViT was chosen due to its ability to establish direct relationships among images, which is particularly beneficial for analysing the complexities of the brain. The study evaluated both binary classifications (normal versus abnormal) and ternary classifications (healthy controls, moderate cognitive impairment, and Alzheimer's disease). The results revealed that ViT outperformed VGG19 in binary classification tasks but did not exhibit the same advantage in ternary classification. Thus, it can be concluded that ViT has not consistently shown superiority over CNN in classifying Alzheimer's disease when utilising PET imaging.

Research by Shaffi et al. (2024) proposed the implementation of ViT models to enhance the classification efficiency of Alzheimer's disease using MRI images. They constructed an ensemble framework comprising four fundamental ViT models, incorporating both hard-voting and soft-voting methodologies. The evaluation was carried out using the OASIS and ADNI datasets, specifically addressing challenges associated with imbalanced data. The ViT ensemble demonstrated a 2% increase in accuracy relative to individual models, and when compared to CNNs and traditional machine learning models, the ViT models exhibited improvements in accuracy of 4.14% and 4.72%, respectively.

Tang et al. (2024) integrated multi-modal data to enhance the diagnostic process for Alzheimer's disease. By employing MRI and PET imaging, they utilised a 3D convolutional neural network (3D CNN) to extract significant features, subsequently scaling the Transformer model to investigate the

global correlations among these features. This combined multi-modal information was visualised to pinpoint brain regions associated with Alzheimer's disease. Consequently, the model achieved an accuracy of 98.1% on the ADNI dataset, identifying the left parahippocampal region as a consistently significant area pertinent to the diagnosis of Alzheimer's disease.

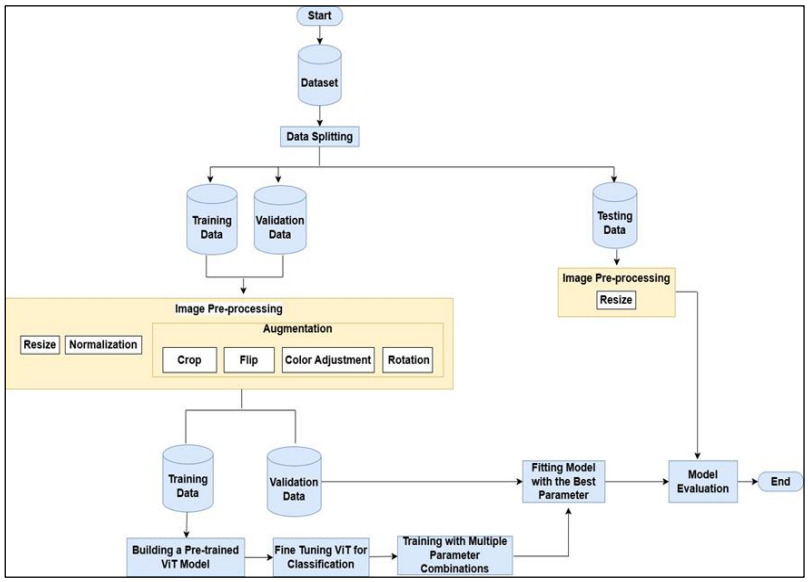
The research findings outlined above generally demonstrate promising outcomes in classifying MRI and PET images for Alzheimer's diagnosis. The ViT was selected for its capability to capture intricate relationships in medical imaging data. For instance, Lyu et al. (2022) indicated that transfer learning from natural to brain imaging domains enhances classification performance. Almufareh et al. (2023) underscored the efficacy of attention-based mechanisms in ViT for Alzheimer's detection, while Shaffi et al. (2024) reported improved accuracy using a ViT ensemble framework. Conversely, Shin et al. (2023) identified limitations in ViT's performance in ternary classification compared to CNNs, suggesting the need for further refinement in multi-class categorisation strategies. This study preferred ViT over CNNs due to its ability to capture long-range pixel relationships, advantageous for analysing complex brain structures, and the effectiveness of its pre-trained models on large datasets like ImageNet for transfer learning in medical imaging.

THE PROPOSED METHOD

The proposed method employs a deep learning approach utilising the ViT for computer vision tasks. The process begins with the collection and pre-processing of an MRI Alzheimer's image dataset. Data are split into 90% for model development (further divided into training and validation sets) and 10% for testing. Pre-processing includes normalisation, resizing to 224×224, and augmentation techniques like Resize, Cropping, Flipping, Rotation, and Color Adjustment. A pre-trained ViT model is fine-tuned, optimised with various learning rates and batch sizes, and evaluated using metrics such as accuracy, specificity, sensitivity, and F1-score. The research methodology proposed in this study is shown in Figure 1.

Figure 1

The Research Method for the Study



Diagnostic Modalities for Alzheimer's Disease

Medical professionals employ two primary methods to differentiate between samples with and without dementia: neuroimaging and non-neuroimaging techniques. Neuroimaging modalities, including CT, MRI, and Positron Emission Tomography (PET), are integral to the diagnosis of Alzheimer's disease. CT scans are used to assess brain abnormalities such as size variations, injury, or the presence of tumors. MRI offers high-resolution images of brain structures, aiding in the detection of atrophy in specific regions. PET, on the other hand, utilises radioactive tracers to evaluate the brain's metabolic activity, providing a comprehensive assessment of its condition. Among these techniques, MRI is widely regarded as the most effective for identifying early physical changes associated with Alzheimer's disease. Furthermore, non-neuroimaging methods such as blood tests, genetic profiling, and neuropsychological assessments contribute to disease diagnosis, though neuroimaging remains the foremost diagnostic tool (Malik & Singh, 2024; Mi et al., 2024).

Dataset

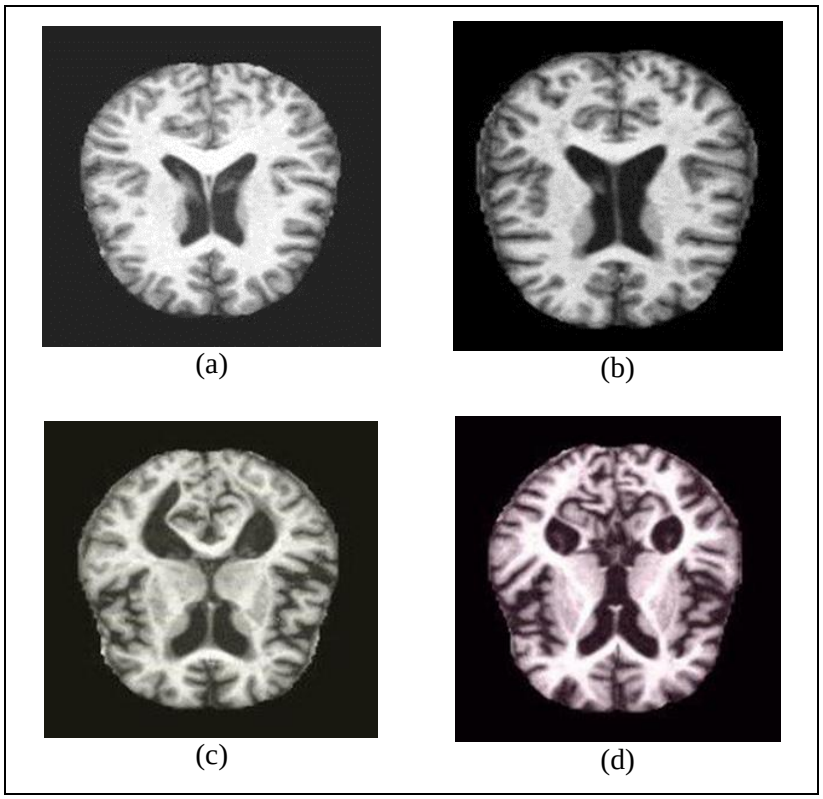
This study employs image data sourced online from the Alzheimer's Disease Neuroimaging Initiative (ADNI), specifically focusing on MRI scans related to Alzheimer's disease. The dataset, accessible via the Kaggle platform Azheimer's MRI, comprises a total of 8,000 MRI images of the human brain, categorised into four classifications: non-demented, mildly demented, very mildly demented, and moderately demented. Each class contains 2,000 images, ensuring a balanced distribution across all categories. The distinctions between the various types of dementia are illustrated in Figure 2, which displays the MRI results.

Data Splitting

Data splitting refers to the process of partitioning a dataset into distinct subsets for the purposes of training and model evaluation. Although this might appear straightforward, dataset partitioning demands a precise methodology, as both the dataset's size and the partition ratio can significantly influence the model's overall outcomes and performance. Consequently, partitioning the data into training, testing, or validation sets is a widely adopted strategy for determining model hyperparameters and assessing generalisation performance (Muraina, 2022).

Figure 2

MRI Images Illustrating (a) Non-Demented, (b) Very Mild Demented, (c) Mild Demented, and (d) Moderate Demented Classifications



The division of data into separate subsets is a fundamental strategy to prevent overfitting. Muraina (2022) highlights that partitioning the dataset into three distinct segments is vital for minimising both overfitting and bias during model selection. He advocates for the largest portion of the data to be used for training, while the validation (or development) set, and the test set should be of equal size. In a similar vein, Joseph and Vakayil (2021) suggest that the training set itself can be further subdivided, incorporating a validation subset. Conversely, some researchers choose to forgo a separate test set, opting instead to split the data into just two parts: the development and the test set. In this framework, the model is trained on the training set, hyperparameters are optimised using the development set, and final performance assessments are made based on the outcomes from this process.

Data Augmentation

Data augmentation encompasses a variety of techniques applied at the data level rather than modifying the model architecture. These methods enhance deep learning model performance by generating diverse and balanced samples for the training dataset. Optimal model accuracy depends on the dataset's quality and quantity, requiring sufficient diversity and size, both of which data augmentation can provide (Alomar et al., 2023). Augmentation techniques can be categorised by their purpose—either increasing dataset size or diversity—or by the specific problem they address. For instance, random erasing tackles occlusion (Zhong et al., 2020), rotation and flipping mitigate viewpoint variations (Divon & Tal, 2018), brightness adjustments handle lighting changes (Liu et al., 2021) and cropping and zooming address scaling and background issues.

Vision Transformer Frameworks

Deep learning is a subset of Artificial Neural Networks (ANNs) characterised by its use of multiple layers to process intricate data and execute machine learning tasks typically performed by humans to obtain specific knowledge. Fundamental deep learning models, such as the Multi-Layer Perceptron (MLP), operate by mapping inputs to outputs through mathematical functions (Bengio et al., 2021). In contrast to traditional algorithms, such as those in machine learning that necessitate manual feature engineering, deep learning demonstrates superiority by automatically analysing complex data. This capability renders it exceptionally valuable across various domains, including healthcare, finance, and social sciences. Notably, deep learning has been applied in areas such as object detection, stock price forecasting, and disease classification.

One of the significant advancements in deep learning is the ViT, an architecture specifically designed for image processing. ViT is based on the transformer model—an attention-driven mechanism initially developed for natural language processing tasks, which has since been adapted for various computer vision applications, including image classification, segmentation, and object detection (Dosovitskiy et al., 2021; Gheflati & Rivaz, 2022). By leveraging the attention mechanism, ViT efficiently processes image data compared to earlier models such as CNNs. In numerous tasks, ViT has demonstrated performance that is either comparable to or exceeds that of CNNs, which have been traditionally employed in image processing (Khan et al., 2022).

The ViT architecture comprises three key components: the segmentation of the image into smaller patches (patch embeddings), the incorporation of positional embeddings, and the subsequent processing through transformer encoders, as illustrated in Figure 3. In the initial step, the image is segmented into smaller patches, which are then represented as vectors. Each patch typically measures 16×16 pixels and is subsequently flattened linearly according to Equation 1 (Dosovitskiy et al., 2021).

$$n = \frac{hw}{h_p w_p} \tag{1}$$

In this equation, (*h, w*) denotes the resolution of the input image, where *h* represents the height and w the width. Conversely, (hp, wp) specifies the resolution of each square patch, which is measured at $m \times m$, and n denotes the total number of patches generated. Subsequently, positional embeddings are applied to convey information regarding the relative positioning of each patch (Vaswani et al., 2017). This step is crucial, as the ViT lacks an inherent mechanism to identify natural sequences within images, unlike CNNs. The arrangement of the patch embeddings is determined using Equation 2:

$$z_0 = [V_{class}; X_1E; X_2E; \ldots; X_nE] + E_{pos}, \quad E \in \mathbb{R}^{(p^2c)\times D} E_{pos} \in \mathbb{R}^{(n+1)\times D} \tag{2}$$

In this context, $Vclass$ refers to the one-hot encoding of a matrix token generated by the computer, while Xn denotes the nth patch presented in matrix form. Additionally, E represents the embedding matrix corresponding to the patch, and $Epos$ signifies the position encoding, also formatted as a matrix. After the image has been processed into patches and positional embeddings, the outcomes are compiled into a $z0$ vector, which is subsequently inputted into the transformer encoder block (Dosovitskiy et al., 2021). The transformer encoders are fundamental components of the ViT architecture, with an illustration of the process within the transformer encoder depicted in Figure 4.

The transformer encoder is composed of several critical components, one of which is the Multihead Attention mechanism. This mechanism enables the model to concentrate on various regions of the image concurrently. Through Multihead Attention, the model can identify multiple significant areas within the image without requiring sequential processing. This capability is a primary advantage of the ViT compared to CNNs, which tend to depend more on the spatial arrangement of images. The structure of Multihead Attention is illustrated in Figure 5.

Figure 3

ViT Architecture

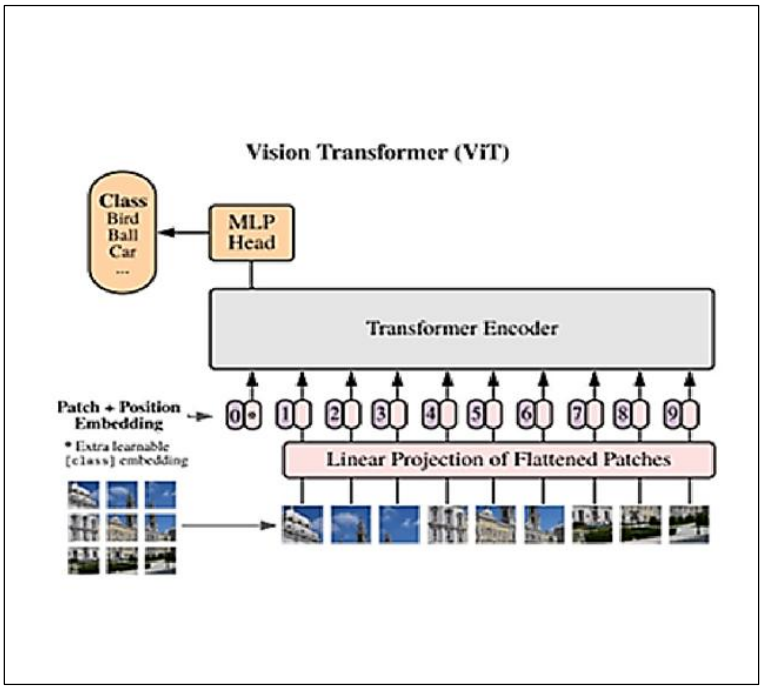


Figure 4

Transformer Encoder

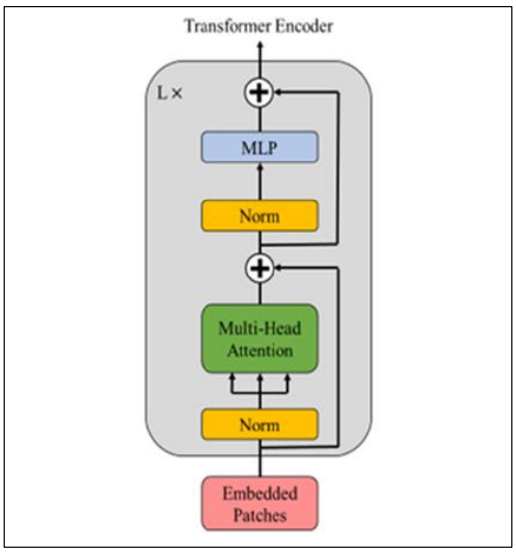
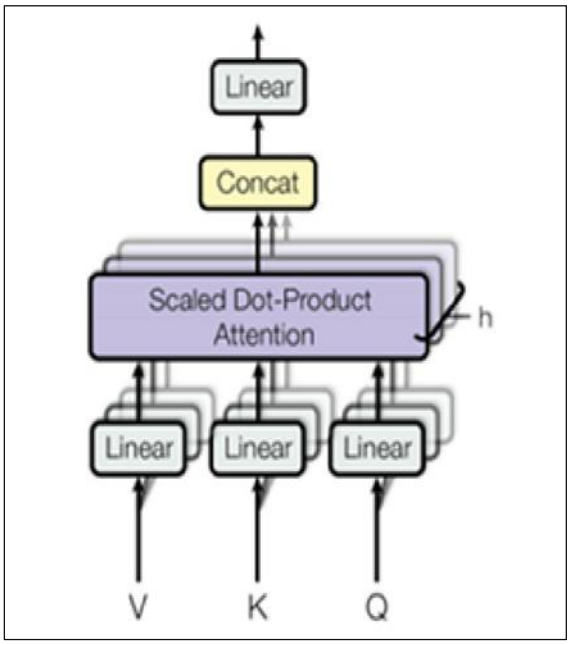


Figure 5

Multihead Attention Structure



According to the depiction of the Multi-Head Attention architecture, the query (Q), key (K), and value (V) are linearly projected multiple times, with distinct projections corresponding to each dimension of dq, dk, and dv. The calculations for the matrices Q, K, and V are detailed in Equations 3, 4, and 5 (Fan et al., 2021):

$$Q = W_q Z \tag{3}$$

$$K = W_k Z \tag{4}$$

$$V = W_v Z \tag{5}$$

Where Wq, Wk, and Wv represent the weights associated with linear transformations, typically characterised by small magnitudes. These weights are initialised randomly from suitable distributions, such as Gaussian, Xavier, or Kaiming distributions, and this initialisation occurs only once prior to the commencement of the training process.

Each projection is subsequently processed concurrently via the attention mechanism, yielding a dv-dimensional output. The calculation of the output matrix within the attention model is defined by Equation 6 (Han et al., 2023):

$$Attention\ (\mathrm{Q}, \mathrm{K}, \mathrm{V}) = softmax\left(\frac{\mathrm{Q} \cdot \mathrm{K}^{\mathrm{T}}}{\sqrt{d_k}}\right) \cdot \mathrm{V} \tag{6}$$

The outcomes of all these attention mechanisms are integrated and reprojected to generate the final output. Multi-head attention enables the model to simultaneously capture information from various representations across different subspaces and positions, as detailed in Equation 7, utilising the W parameter matrix for the queries (Q), keys (K), values (V), and the final output (Vaswani et al., 2017).

$$\text{MultiHead} = \text{Concat}(\text{head}_1, .. , \text{head}_h) \cdot W^o \quad (7)$$
$$\text{with} \quad \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

An additional element that enhances the transformer encoder is the MLP, which operates through a dense layer utilising GeLU activation. Researchers employ residual connections and layer normalisation to maintain and stabilise information from the preceding layer before it proceeds to the next layer (Bazi et al., 2021). This methodology enables the ViT to process images with greater efficacy and efficiency.

Performance Evaluation

A widely employed metric for assessing multi-class or single-label classification models—where each data instance can be linked to only one class at a time—is the confusion matrix (Krstinic et al., 2023). The confusion matrix offers a detailed overview of the classifier's performance. Typically, it encompasses four key metrics: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These metrics are instrumental in determining whether a patient is correctly diagnosed with a particular disease (Hasnain et al., 2020).

The TP value indicates a scenario in which the prediction of the disease aligns with the actual condition, confirming that an individual is indeed affected by the disease. Conversely, FP refers to a situation where an individual is forecasted to have the disease despite being in a state of good health. In the context of predictive analysis, the TN value indicates that the model predicts an individual as healthy, and this prediction is accurate. On the other hand, FN occurs when the model predicts an individual as healthy, but the individual is, in fact, suffering from a disease (Juneja et al., 2020).

The specified values of TP, TN, FP, and FN can be utilised to calculate various evaluation metrics for assessing the performance of the implemented methods. The metrics consist of accuracy, sensitivity, specificity, and F1-score. Accuracy measures the degree to which the predicted outcomes align with the actual values, reflecting the proportion of samples that are correctly and incorrectly classified (Xhumari & Haloci, 2023). Sensitivity evaluates the model's capacity to correctly identify positive instances, a critical factor in disease prediction (Juneja et al., 2020). Specificity assesses the model's proficiency in correctly recognising negative classes, often used in medical contexts for disease diagnosis (Juneja et al., 2020). The F1 score represents the harmonic mean of precision and recall (Xhumari & Haloci, 2023). A high F1 score indicates the model's strong performance in label prediction. A classification model is deemed to exhibit excellent performance if its evaluation metrics exceed 90%. Performance is classified as good for scores ranging from 81-90%, moderate for 71-80%, poor for 61-70%, and failed if the score falls below 60%. The calculation of these evaluation metrics is formally defined by Equations 8, 9, 10, and 11:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \; x \; 100\% \quad (8)$$

$$Sensitivity = \frac{TP}{(TP + FN)} x \; 100\% \quad (9)$$

$$Specifity = \frac{TN}{(TN + FP)} x \; 100\% \quad (10)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} x\ 100\% \quad (11)$$

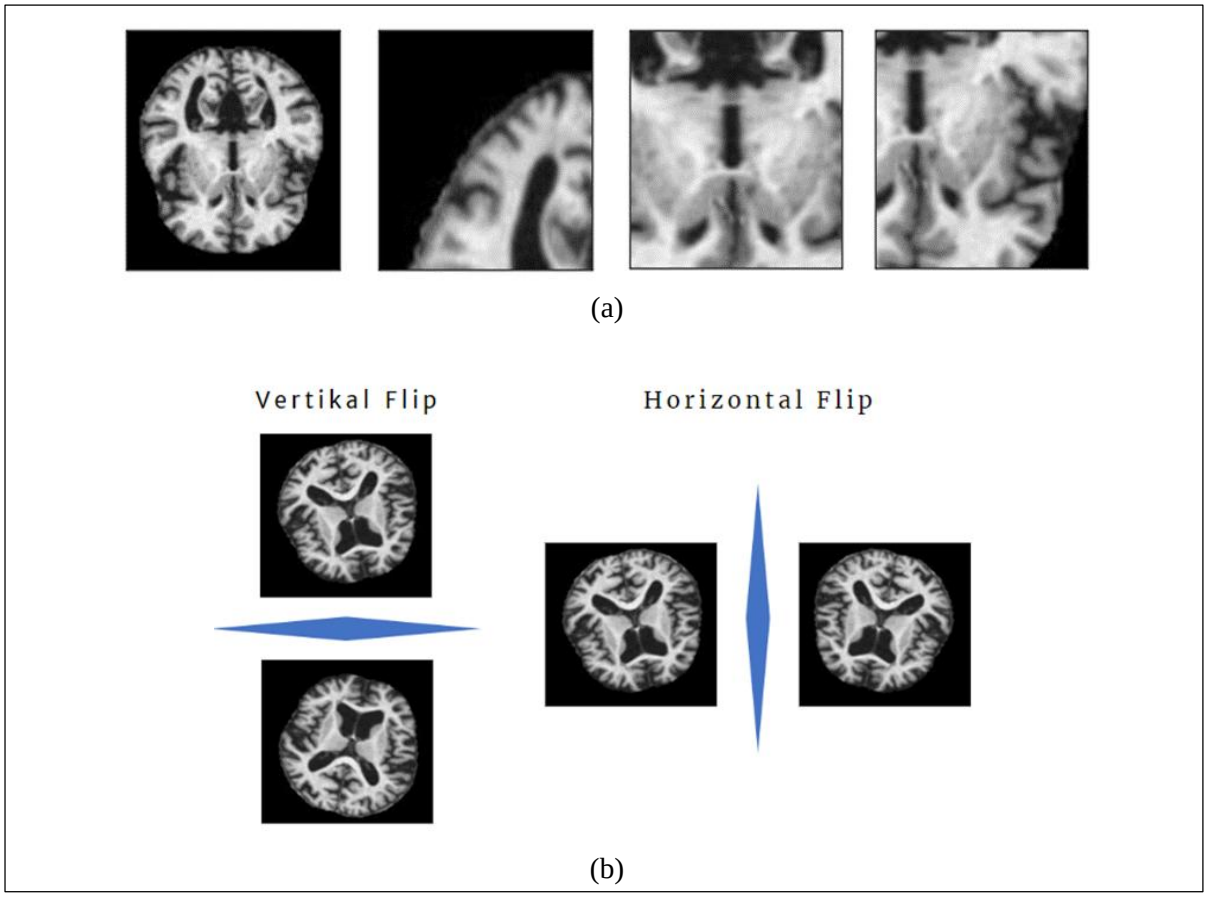
RESULTS AND DISCUSSION

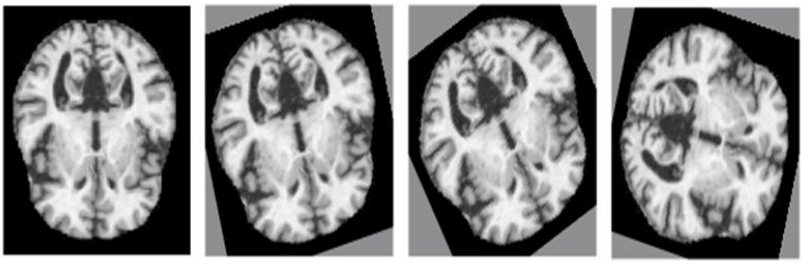
The initial phase of this study requires the input of data into the Python directory to enable further processing. The dataset consisted of 8000 MRI images associated with Alzheimer's disease, organised into four specific categories: non-demented, mildly demented, very mildly demented, and moderately demented. The pre-processing techniques employed encompass data augmentation, including modifications through methods such as resizing, flipping, color adjustments, and the transformation of images to grayscale. Figure 6 provides a detailed explanation of the augmentation process.

Initially, all images with a resolution of 200 x 190 pixels were resized to 224 x 224 pixels. This adjustment was made to align the size of Alzheimer's images with the input dimensions of a pre-trained model, "google/vit-base-patch16-224". Furthermore, the normalisation procedure was applied to the images, altering their pixel values to ensure a mean of zero and a variance of one, as derived from the pre-trained model. The normalising step is crucial forenabling faster and more uniform convergence of the model during training.

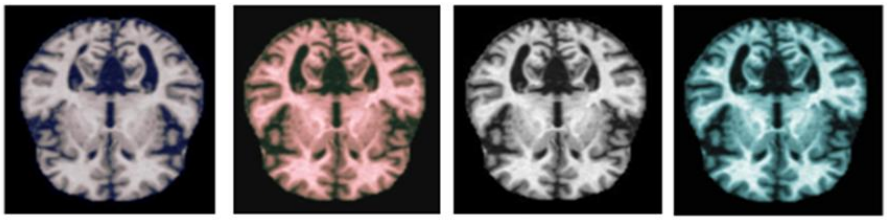
Figure 6

Illustration of the Augmentation Process(a) Resize, (b) Flip, (c) Color Adjustment, and (d) Grayscale





(c)



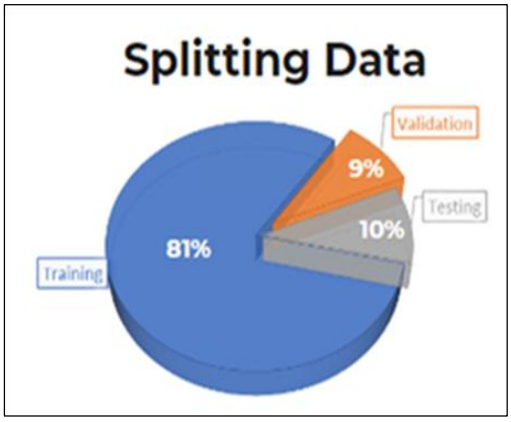
(d)

Following the augmentation and normalisation procedures, the image is prepared for subsequent analysis by converting it into a suitable format for processing by the model. Additionally, this research involves fine-tuning the "google/vit-base-patch16-224" model by retraining it. Originally, the model was trained on a diverse dataset comprising both general and medical images, which enabled it to adapt to a wide range of image types. This model is specifically optimised for the classification of Alzheimer's disease MRI images into four distinct categories. Moreover, fine-tuning is conducted with a strategy to mitigate overfitting, which includes setting the dropout rate to 0.2.

The training process utilised a dataset comprising 8,000 samples, which were partitioned into three distinct subsets: training, validation, and testing. A total of 90% of the dataset, equating to 7,200 samples, was designated for model training, while the remaining 10%, or 800 samples, was reserved for evaluation purposes. Within the 7,200 training samples, 90% (6,480 samples) were allocated for model training, and 10% (720 samples) for validation. Thus, the data was segmented into 6,480 samples for primary training, 720 samples for validation, and 800 samples for final testing. The distribution of these datasets is depicted in Figure 7.

Figure 7

Distribution of MRI Image Data for Alzheimer's Disease



Throughout the training phase, various parameter combinations were employed to achieve optimal performance, specifically by testing different pairings of learning rates and batch sizes: 0.01 with a batch size of 32, 0.01 with a batch size of 64, 0.001 with a batch size of 32, 0.001 with a batch size of 64, 0.0001 with a batch size of 32 and 0.0001 with a batch size of 64, each trained over 50 epochs. The results of training the model with these various parameters are presented in Table 2.

Table 2

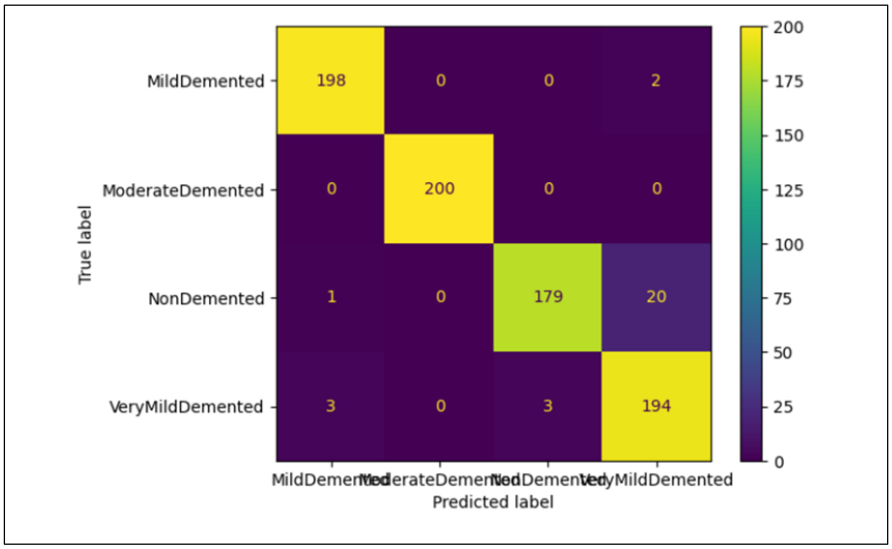
Comparison of Training Outcomes Across Various Parameter Combinations

Learning Rate/Batch Size	Accuracy	Sensitivity	Specificity	F1-Score
0.01/32	62.50%	25.00%	75.00%	10.00%
0.01/64	62.50%	25.00%	75.00%	10.00%
0.001/32	75.12%	50.25%	83.41%	48.40%
0.001/64	75.18%	50.37%	83.45%	50.57%
0.0001/32	97.37%	94.75%	98.25%	94.78%
0.0001/64	**98.19%**	**96.34%**	**98.80%**	**96.37%**

The analysis results presented in Table 2 indicate that the most effective combination of parameters for the classification model is a learning rate of 0.0001 and a batch size of 64. This parameter configuration yields superior performance across all evaluation metrics, achieving an accuracy of 98.19%, sensitivity of 96.34%, specificity of 98.80%, and an F1 score of 96.37%. These outcomes highlight the model's proficiency in accurately classifying images, with minimal error. Following the identification of the optimal parameters, the model was evaluated using a test set comprising 10% of the total dataset—800 images distributed across four categories, with 200 images per category. The classification results are summarised in the confusion matrix shown in Figure 8, which provides a breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class.

Figure 8

Confusion Matrix for Predicting Alzheimer's Disease



Based on the confusion matrix depicted in Figure 8, the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values for each class are determined. For the "Mild Demented" class, the model identifies 192 cases as TP and 595 cases as TN, with 7 cases classified as FP and 8 as FN. In the "Moderately Demented" class, the model accurately classifies all cases, yielding TP and TN values of 200 and 600, respectively, with no FP or FN. In the "Non-Demented" class, the model reports 178 TP, 595 TN, 5 FP, and 8 FN. Lastly, for the "Very Mild Demented" class, the TP value is 19, TN is 595, FP is 7, and FN is 22. A summary of the evaluation results for each class is presented in Table 3.

Table 3

Evaluation Metrics Results for Each Class

Class	Accuracy	Sensitivity	Specificity	F1-Score
Mild Demented	0,99	0,99	0,99	0,98
Moderate Demented	1,00	1,00	1,00	1,00
Non Demented	0,97	0,89	0,99	0,93
Very Mild Demented	0,96	0,97	0,96	0,93
Average	0,98	0,96	0,98	0,96

This evaluation revealed that the "Mild Demented" category attained an accuracy of 0.99, accompanied by high sensitivity and specificity. The "Moderate Demented" category demonstrated flawless performance, with all metrics scoring 1.00. For the "Non-Demented" category, the model achieved an accuracy of 0.97, a sensitivity of 0.89, and a specificity of 0.99. The "Very Mild Demented" category exhibited an accuracy of 0.96, with a sensitivity of 0.97 and specificity of 0.96, reflecting solid performance despite some misclassification instances. These errors were primarily attributed to the inherent difficulty in distinguishing features across categories, particularly between "Non-Demented" and "Very Mild Demented," which share overlapping characteristics. Furthermore, the incorrect application of augmentation techniques may introduce bias into the model training process. Figures 9 and 10 present examples of both correctly classified and misclassified MRI images.

Figure 9

Correctly Classified Sample Set

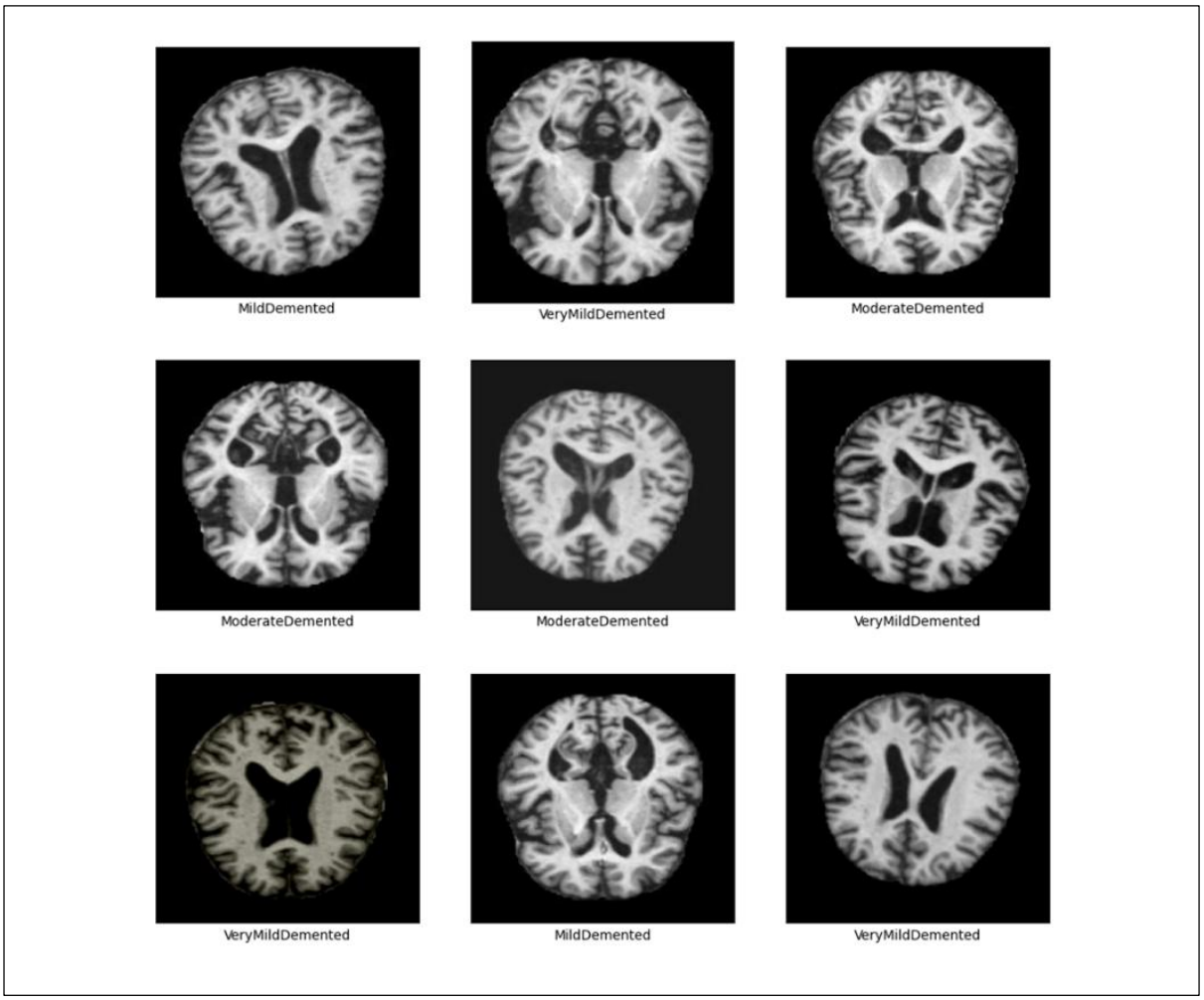
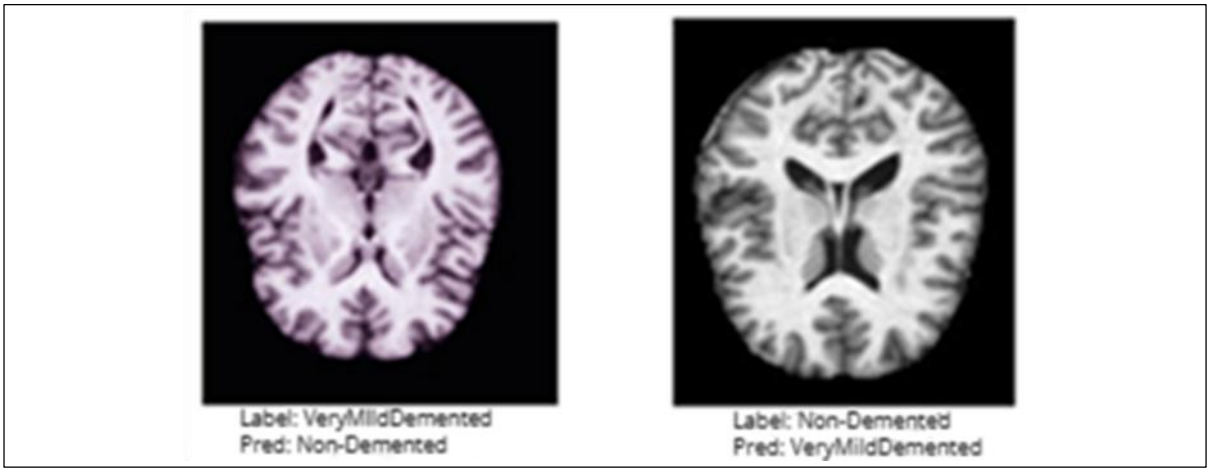


Figure 10

Misclassification Sample Set



In summary, the model demonstrated exceptional performance, achieving an average accuracy of 0.98, along with sensitivity, specificity, and an average F1 score of 0.96, 0.98, and 0.96, respectively. These results underscore the model's strong capability in accurately classifying MRI images across various stages of Alzheimer's disease. For future studies, it is advisable to enhance the datasets for the "Non-Demented" and "Very Mildly Demented" categories, thereby enabling the model to better capture the distinguishing features of each class. Furthermore, a comprehensive assessment of the augmentation methods employed, particularly cropping, is recommended to ensure that critical information within the images is preserved and not inadvertently lost. These refinements aim to reduce the likelihood of classification inaccuracies in the process.

Following the training and performance evaluation of the model using the ViT architecture, the subsequent task is to interpret the results and contrast them with findings from prior research. Table 4 provides a comparative analysis of the ViT model's performance against various methods previously employed for classifying Alzheimer's MRI images, as documented in the literature.

Table 4

Comparative Analysis of the Proposed Model with Previous Research

Researcher	Accuracy	Sensitivity	Specificity	F1-score
Lyu et al. (2023)	95.3%	94.4%	-	93.20%
Almufareh et al. (2023)	99.06%	99.06%	99.14%	99.1%
Shin et al. (2023)	80% & 56.67%	60% & 56.67%	-	66.67% & 55.45%
Shaffi et al. (2024)	99.29%	97.54%	99.68%	-
Tang et al. (2024)	98.1%	95.82%	99.09%	97.81%
Proposed Method	**98,19%**	**96,34%**	**98,80%**	**96,37%**

The results indicate a notable enhancement in accuracy and other evaluation metrics, with accuracy reaching 98.19%, sensitivity at 96.34%, specificity at 98.80%, and an F1 score of 96.37%. In contrast, a study by Lyu et al., which utilised data from ImageNet-21K, reported an accuracy of 95.3% and a sensitivity of 94.4%, thereby underscoring the advancements made with the proposed approach. Almufareh et al. reported an accuracy of 99.06% and a specificity of 99.14% in classifying Alzheimer's MRI images. These findings are notably competitive, particularly given that their analysis was conducted on a dataset of approximately 80,000 MRI images—roughly ten times the size of the dataset employed in the present study. In comparison, Shaffi et al. achieved an accuracy of 99.29%, although the absence of reported F1 scores limits a comprehensive evaluation of their performance. Additional research, including studies by Tang et al. and Shin et al., has demonstrated promising results; however, the methodology presented in this study surpasses others in terms of specificity.

The strength of this approach is rooted in the deployment of a meticulously optimised ViT model, which attains an impressive specificity of 98.80%, significantly reducing diagnostic errors in identifying individuals without Alzheimer's disease. Furthermore, the F1 score of 96.37% reflects a robust equilibrium between precision and sensitivity—two critical metrics in medical diagnostics. While several prior studies have yielded remarkable results, the methodology presented herein achieves competitive performance, with accuracy nearing the highest previously reported figures, while also demonstrating substantial effectiveness in classifying Alzheimer's-related MRI images.

CONCLUSION

This study highlights the significance of early detection of Alzheimer's disease through the application of advanced imaging techniques, particularly MRI, in conjunction with ViT models. An analysis of 8,000 MRI images, categorised into four distinct groups, resulted in an accuracy of 98.19%, sensitivity of 96.34%, specificity of 98.80%, and an F1-score of 96.37%. These results indicate that the employed model is not only effective in classifying images but also proficient in identifying individuals who are unaffected by Alzheimer's, which is essential for reducing misdiagnoses and facilitating timely interventions.

While the findings of this study demonstrate competitive performance in comparison to prior research, several challenges remain to be addressed, particularly the misclassification between the 'Non-Demented' and 'Very Mild Demented' categories. It is recommended that future research focuses on enhancing the data volume in these categories and assessing the applied augmentation techniques. These measures aim to improve both the accuracy and reliability of the model, as well as enhance the effectiveness of early detection in clinical practice for patients with Alzheimer's disease.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Abubakar, M. B., Sanusi, K. O., Ugusman, A., Mohamed, W., Kamal, H., Ibrahim, N. H., Khoo, C. S., & Kumar, J. (2022). Alzheimer's disease: An update and insights into pathophysiology. *Frontiers in Aging Neuroscience*, *14*. https://doi.org/10.3389/fnagi.2022.742408

Al Rahbani, R. G., Ioannou, A., & Wang, T. (2024). Alzheimer's disease multi-class detection through deep learning models and post-processing heuristics. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualisation*, *12*(1). https://doi.org/10.1080/21681163.2024.2383219

Almufareh, M. F., Tehsin, S., Humayun, M., & Kausar, S. (2023). Artificial cognition for detection of mental disability: A vision transformer approach for Alzheimer's disease. *Healthcare (Switzerland)*, *11*(20). https://doi.org/10.3390/healthcare11202763

Alomar, K., Aysel, H. I., & Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies. *Journal of Imaging*, *9*(2), 46. https://doi.org/10.3390/jimaging9020046

Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, *17*(3), 327–406. https://doi.org/10.1002/alz.12328

Alzheimer's Association. (2023). 2023 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, *19*(4), 1598–1695. https://doi.org/10.1002/alz.13016

Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Guevara Lopez, M. A. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine*, *127*, 248–257. https://doi.org/10.1016/j.cmpb.2015.12.014

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., & Schmid, C. (2021). ViViT: A Video vision transformer. *Proceedings of the IEEE International Conference on Computer Vision*, 6816–6826. https://doi.org/10.1109/ICCV48922.2021.00676

Azad, R., Asadi-Aghbolaghi, M., Fathy, M., & Escalera, S. (2019). Bi-directional ConvLSTM U-net with densley connected convolutions. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 406–415. https://doi.org/10.1109/ICCVW.2019.00052

Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., Jose, A., Roy, R., & Merhof, D. (2024). Advances in medical image analysis with vision transformers: A comprehensive review. *Medical Image Analysis*, *91*. https://doi.org/10.1016/j.media.2023.103000

Azad, R., Khosravi, N., & Merhof, D. (2022). SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities. *Proceedings of Machine Learning Research*, *172*, 48–62.

Bazi, Y., Bashmal, L., Al Rahhal, M. M., Dayil, R. Al, & Ajlan, N. Al. (2021). Vision transformers for remote sensing image classification. *Remote Sensing*, *13*(3), 1–20. https://doi.org/10.3390/rs13030516

Bello, I., Zoph, B., Le, Q., Vaswani, A., & Shlens, J. (2019). Attention augmented convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 3285–3294. https://doi.org/10.1109/ICCV.2019.00338

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM*, *64*(7), 58–65. https://doi.org/10.1145/3448250

Breijyeh, Z., & Karaman, R. (2020). Comprehensive review on Alzheimer's disease: Causes and treatment. In *Molecules* (Vol. 25, Issue 24). https://doi.org/10.3390/MOLECULES25245789

Chen, Y., Wang, L., Ding, B., Shi, J., Wen, T., Huang, J., & Ye, Y. (2024). Automated Alzheimer's disease classification using deep learning models with Soft-NMS and improved ResNet50 integration. *Journal of Radiation Research and Applied Sciences*, *17*(1), 100782. https://doi.org/10.1016/j.jrras.2023.100782

Dafre, R., & Wasnik, P. (2023). Current diagnostic and treatment methods of Alzheimer's disease: A narrative review. *Cureus*. https://doi.org/10.7759/cureus.45649

Divon, G., & Tal, A. (2018). Viewpoint Estimation---Insights & Model. *European Conference on Computer Vision*, 252–268.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16X16 words: Transformers for image recognition at scale. *ICLR 2021 - 9th International Conference on Learning Representations*.

Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., & Feichtenhofer, C. (2021). Multiscale vision transformers. *Proceedings of the IEEE International Conference on Computer Vision*, 6804–6815. https://doi.org/10.1109/ICCV48922.2021.00675

Gao, Y., & Liu, X. (2021). Secular trends in the incidence of and mortality due to Alzheimer's disease and other forms of Dementia in China from 1990 to 2019: An age-period-cohort study and joinpoint analysis. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.709156

Gheflati, B., & Rivaz, H. (2022). Vision transformers for classification of breast ultrasound images. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2022-July*, 480–483. https://doi.org/10.1109/EMBC48229.2022.9871809

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A Survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, *8*, 90847–90861. https://doi.org/10.1109/ACCESS.2020.2994222

Helaly, H. A., Badawy, M., & Haikal, A. Y. (2022). Deep learning approach for early detection of Alzheimer's disease. *Cognitive Computation*, *14*(5), 1711–1727. https://doi.org/10.1007/s12559- 021-09946-2

Irankhah, E. (2020). Evaluation of early detection methods for Alzheimer's disease. *Bioprocess Engineering*, *4*(1), 17. https://doi.org/10.11648/j.be.20200401.13

Joseph, V. R., & Vakayil, A. (2021). SPLIT: An optimal method for data splitting. *Technometrics*, *64*(2), 166–176. https://doi.org/10.1080/00401706.2021.1921037

Juganavar, A., Joshi, A., & Shegekar, T. (2023). Navigating early Alzheimer's diagnosis: A comprehensive review of diagnostic innovations. *Cureus*. https://doi.org/10.7759/cureus.44937

Juneja, M., Thakur, N., Thakur, S., Uniyal, A., Wani, A., & Jindal, P. (2020). GC-NET for classification of glaucoma in the retinal fundus image. *Machine Vision and Applications*, *31*(5). https://doi.org/10.1007/s00138-020-01091-4

Karimijafarbigloo, S., Azad, R., Kazerouni, A., & Merhof, D. (2023). MS-Former: Multi-scale self-guided transformer for medical image segmentation. *Proceedings of Machine Learning Research*, *227*, 680–694.

Karimijafarbigloo, S., Azad, R., Kazerouni, A., Ebadollahi, S., & Merhof, D. (2023). MMCFormer: Missing modality compensation transformer for brain tumor segmentation. *Proceedings of Machine Learning Research*, *227*, 1144–1162.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys*, *54*(10) 1-41. https://doi.org/10.1145/3505244

Kim, W. S., Lee, D. H., Kim, Y. J., Kim, T., Hwang, R. Y., & Lee, H. J. (2020). Path detection for autonomous traveling in orchards using patch-based CNN. *Computers and Electronics in Agriculture*, *175*. https://doi.org/10.1016/j.compag.2020.105620

Krstinic, D., Seric, L., & Slapnicar, I. (2023). Comments on "MLCM: Multi-label confusion matrix." In *IEEE Access* (Vol. 11, pp. 40692–40697). https://doi.org/10.1109/ACCESS.2023.3267672

Li, X., Feng, X., Sun, X., Hou, N., Han, F., & Liu, Y. (2022). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019. *Frontiers in Aging Neuroscience*, *14*. https://doi.org/10.3389/fnagi.2022.937486

Liu, H., Wang, C., & Peng, Y. (2021). Data augmentation with illumination correction in sematic segmentation. *Journal of Physics Conference Series*, *2025*(1), 012009. https://doi.org/10.1088/1742-6596/2025/1/012009

Lyu, Y., Yu, X., Zhu, D., & Zhang, L. (2022). Classification of Alzheimer's disease via vision transformer. *ACM International Conference Proceeding Series*, 463–468. https://doi.org/10.1145/3529190.3534754

Malik, P., & Singh, S. (2024). Alzheimer's disease classification using neuroimaging modalities and deep learning. *Journal of Harbin Engineering University*, *45*(6), 433-448.

Mi, J., Liu, C., Chen, H., Qian, Y., Zhu, J., Zhang, Y., Liang, Y., Wang, L., & Ta, D. (2024). Light on Alzheimer's disease: From basic insights to preclinical studies. In *Frontiers in Aging Neuroscience* (Vol. 16). Frontiers Media SA. https://doi.org/10.3389/fnagi.2024.1363458

Muraina, I. O. (2022). Ideal dataset splitting ratios in Machine learning algorithms: General concerns for data scientists and data analysts. In *7th International Mardin Artuklu Scientific Researches Conference*.

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Nichols, E., Szoeke, C. E. I., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., Aichour, M. T. E., Akinyemi, R. O., Alahdab, F., Asgedom, S. W., Awasthi, A., Barker-Collo, S. L., Baune, B. T., Béjot, Y., Belachew, A. B., Bennett, D. A., Biadgo, B., Bijani, A., Bin Sayeed, M. S., … Murray, C. J. L. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *18*(1), 88–106. https://doi.org/10.1016/S1474-4422(18)30403-4

Pulido, M. L. B., Hernández, J. B. A., Ballester, M. Á. F., González, C. M. T., Mekyska, J., & Smékal, Z. (2020). Alzheimer's disease and automatic speech analysis: A review. In *Expert Systems with Applications* (Vol. 150). https://doi.org/10.1016/j.eswa.2020.113213

Ramachandran, P., Bello, I., Parmar, N., Levskaya, A., Vaswani, A., & Shlens, J. (2019). Stand-alone self-attention in vision models. *Advances in Neural Information Processing Systems*, *32*.

Shaffi, N., Viswan, V., & Mahmud, M. (2024). Ensemble of vision transformer architectures for efficient Alzheimer's Disease classification. *Brain Informatics*, *11*(1). https://doi.org/10.1186/s40708-024-00238-7

Shin, H., Jeon, S., Seol, Y., Kim, S., & Kang, D. (2023). Vision transformer approach for classification of Alzheimer's disease using 18F-Florbetaben brain images. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13063453

Suganthe, R. C., Geetha, M., Sreekanth, G. R., Gowtham, K., Deepakkumar, S., & Elango, R. (2021). Multi-class Classification of Alzheimer's Disease Using Hybrid Deep Convolutional Neural Network. *Volatiles & Essent. Oils*, *8*(5), 145–153.

Tang, Y., Xiong, X., Tong, G., Yang, Y., & Zhang, H. (2024). Multi-modal diagnosis model of Alzheimer's disease based on improved Transformer. *BioMedical Engineering Online*, *23*(1). https://doi.org/10.1186/s12938-024-01204-4

Vaswani, A., Ramachandran, P., Srinivas, A., Parmar, N., Hechtman, B., & Shlens, J. (2021). Scaling local self-attention for parameter efficient visual backbones. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 12889–12899. https://doi.org/10.1109/CVPR46437.2021.01270

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*, 5999–6009.

Vimala, B. B., Srinivasan, S., Mathivanan, S. K., Mahalakshmi, Jayagopal, P., & Dalu, G. T. (2023). Detection and classification of brain tumor using hybrid deep learning models. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-50505-6

Wong, K. K. L., Xu, J., Chen, C., Ghista, D., & Zhao, H. (2023). Functional magnetic resonance imaging providing the brain effect mechanism of acupuncture and moxibustion treatment for depression. *Frontiers in Neurology*, *14*. https://doi.org/10.3389/fneur.2023.1151421

Wong, K. K. L., Xu, W., Ayoub, M., Fu, Y. L., Xu, H., Shi, R., Zhang, M., Su, F., Huang, Z., & Chen, W. (2023). Brain image segmentation of the corpus callosum by combining Bi-Directional Convolutional LSTM and U-Net using multi-slice CT and MRI. *Computer Methods and Programs in Biomedicine*, *238*. https://doi.org/10.1016/j.cmpb.2023.107602

World Alzheimer Report. (2019). World Alzheimer report 2019, attitudes to dementia. *Alzheimer's Disease International: London*.

Xhumari, E., & Haloci, S. (2023). A comparative study of credit scoring and risk management Techniques in Fintech: Machine Learning vs. Regression Analysis. *CEUR Workshop Proceedings*, *3402*, 13–20.

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(07), 13001–13008. https://doi.org/10.1609/aaai.v34i07.7000

Zhou, T., Fu, H., Chen, G., Shen, J., & Shao, L. (2020). Hi-Net: Hybrid-fusion network for multi-modal MR image synthesis. *IEEE Transactions on Medical Imaging*, *39*(9), 2772–2781. https://doi.org/10.1109/TMI.2020.2975344

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2021). Deformable Detr: Deformable transformers for end-to-end object detection. *ICLR 2021 - 9th International Conference on Learning Representations*.