Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

ENHANCING ALZHEIMER'S DISEASE CLASSIFICATION USING HYBRID DEEP LEARNING MODELS ON SYNTHESIZED NEUROIMAGING DATA

Muhammad Rehan[1], Abdul Manan[*2], Muazzam Ali[3], Tanveer Hussain[4], Ayesha Saddique[5]

[1, *2,3,4,5]*Department of Basic Sciences Superior University Lahore*

DOI: https://doi.org/10.5281/zenodo.16628641

Keywords
Alzheimer's disease, deep learning, medical image analysis, transformer models, pseudo-RGB synthesis, 3D CNN, Swin Transformer, MedViT

Article History
Received on 31 April 2025
Accepted on 16 June 2025
Published on 31 July 2025

Corresponding Author: * Abdul Manan

Abstract
The diagnosis of Alzheimer's Disease (AD) using neuroimaging has difficulties owing to the intricate retrieval of features and the critical need for accuracy within the class boundaries. This research analyzes the effect of applying deep learning model techniques, such as CNNs, Transformers, and 3D CNNs, to a dataset of 16,000 images containing normal and pseudo-RGB synthesized neuroimaging data, including neuroimaging data. The models that were tested were EfficientNetV2, ResNet50, ResNet101, DenseNet121, DenseNet201, Swin Transformer, MedViT and 3D CNN. Their Measurements calculated were accuracy, precision, recall, F1-score, and specificity while the evaluation metrics defined were specific to measuring diagnosis performance. The most important findings are that MedViT outperformed all models with head-turning accuracy of 98.5% while Swin Transformer came second at 97.1% which had better learning of global features. 3D CNN was the lowest performer with 93.5% accuracy and was struggling because of weak handling of volumetric data. The findings Med-Vit models on diagnosed AD offered the best performance indicate that model results confirm that hybrid transformer-CNN architectures, Med-Vit which is a new hybrid model developed with encourages researchers for further researches distracts from the point.

INTRODUCTION

Alzheimer's Disease (AD) is a chronic neurodegenerative condition that affects millions of people globally, and its incidence will grow dramatically as the global population ages. AD causes cognitive impairment, especially in memory, language, and problem-solving capacity, culminating in significant disability in everyday functioning and necessitating full-time care. Being one of the major reasons for dementia, AD poses a great challenge both in clinical diagnosis and therapeutic intervention. It is necessary to diagnose Alzheimer's early since it makes way for interventions that might halt the advancement of the disease, leading eventually to the enhancement of patients' quality of life. Nevertheless, diagnosing AD during its initial phases continues to be challenging because its early symptoms are indistinctive. Such conventional diagnostic procedures, like clinical evaluations and psychometric tests, also tend to miss the disease when it is most responsive to treatment, i.e., in its initial stages. These diagnostic tests are very dependent on subjective interpretation, and therefore many Alzheimer's cases go undetected until well after the disease has progressed significantly [1]. Early diagnosis cannot be stressed enough, as the most effective treatments are often only available in early stages, before the heavy neuronal loss that takes place later in the disease process. Clinical exams and cognitive tests, although significant, tend to fall short in identifying subtle brain changes in the initial

phases of Alzheimer's. Because there is no specificity in distinguishing Alzheimer's from other forms of dementia or aging, these tests are highly vulnerable to misdiagnosis. Thus, the advancement of neuroimaging technology has been vital in the primary diagnosis of Alzheimer's disease(MD). For instance, MRI and PET scans aid in furnishing Alzheimer's patients with vital information about the functioning of their brain and its worse imbalances, like the amyloid plaques, tau tangles, and cortical atrophy associated with Alzheimer's disease. These techniques enable pathological assessment of the brain's structural changes over time, thereby reinforcing their use for diagnosing and monitoring Alzheimer's disease. Despite the advantages neuroimaging provides, the extremely sophisticated and vast neuroimaging data impose the need for reliable advanced computational methods of data interpretation. This is where deep learning (DL) methods applied to neuroimaging data have shown great promise for automating the analysis and ensuring timely and accurate diagnosis of the condition [2].

Deep learning approaches such as Convolutional Neural Networks (CNNs) and transformers allow flexible classification of neuroimaging data from Alzheimer's patients. In automated medical imaging, the ability of CNNs to recognize spatial features has made them popular. These networks are particularly adept at identifying certain specific complex structures within photos which meets the requirements for Alzheimer's associated brain changes. Moreover, there is some popularity gained by Vision Transformers (ViT) as people began understanding their capabilities of identifying long-distance dependencies in images. These are thus specialized in learning global features and because context and spatial relation across wider regions of the scanned brain are critical for determining the diagnosis, they excel at comprehending intricate diagrams such as MRIs. The combination of CNNs and ViTs, where CNNs deal with local information while ViTs deal with global feature extraction, can significantly increase the accuracy of AD classification as these models merge the strengths of both approaches, retaining the best aspects of local and global information, to optimize the representation of the brain's structure and function [3].

Prior research suggests deep learning models may aid in the diagnoses of Alzheimer's disease. For instance, Aryal (2025) developed a multimodal deep-learning framework using MRI data, attaining an F1 score of 0.99 on predicting early stage Alzheimer's [4]. This illustrates the capabilities of deep learning to improve diagnostic precision, particularly in the pre-symptomatic phases of the disease. In the same manner, SinhaRoy and Sen (2023) utilized CNNs and GANs to process Alzheimer's MRI scans with an accuracy of 99.7% in distinguishing AD patients from control subjects [5].

Besides CNNs, Vision Transformers (ViTs) have been helpful for AD detection as well. Mubonanyikuzo et al. (2025) demonstrated the usefulness of ViTs with a sensitivity of 92.5% and specificity of 95.7% regarding the distinction of Alzheimer's patients [6]. The VITs ability to extract features is of great importance for the understanding of complex brain structures where context as well as the interaction between different areas in the brain is important for effective classification. The viability of combining ViT and CNN into a hybrid model on Vision Transformer AD classification systems is promising because CNNs are effective at local feature extraction while ViTs model global context effectively. Also, there is a growing interest in classifying Alzheimer's using multimodal data. Multimodal data refers to the combination of various types of information such as structural and functional neuroimaging data, including clinical biomarkers. The inclusion of such data enhances the predictive power and generalizability of models. For instance, Lashkary et al. (2025) showcased the performance of ConvMixer models for classifying the stages of Alzheimer's disease with Structural MRI scans and achieved an accuracy of 99.88% [7]. This suggests that employing various imaging modalities improves diagnosis tools for Alzheimer's disease. Likewise, it is hypothesized that combining MRI data with clinical assessments and demographic data, as proposed by Modi and Mahajan in 2025, improves classification accuracy and lessens overfitting and data imbalance issues [8].

The methodology of interest in this research implements a hybrid deep model that captures both

local and global features of a large database MRI scans of Alzheimer patients. The results are quite good at an accuracy of 98.5%, greater then a single CNN (95.6%) and ViT (97.1%) [9]. It also helps in minimizing overfitting problems as well as the requirements for clinically deployable systems that are scalable in real-time [10]. The model supports changing from one set of experiments to another, which adds to the expectation for use in a clinical setting for enabling early diagnosis and aiding precise treatment interventions. Regardless of increased accuracy achievable with the infusion of AI within clinical workflows, diagnosis remains challenging. The sheer volume of users beyond the specialist as well as capturing the general population's age and geographic diversity are prominent challenges. From such intricate automatic AD dementia stage classification becomes a very tempting notion, but far more attention is required on the explainable and transparent designs for clinicians [11-13].

In summary, Alzheimer's disease is still a significant challenge for contemporary medicine, and accurate and early diagnosis is crucial for better patient outcomes. The union of neuroimaging with state-of-the-art deep learning methods provides a promising solution for improving the accuracy and efficiency of AD diagnosis. By taking advantage of both Convolutional Neural Networks and Vision Transformers' strengths in a hybrid approach, we can enhance classification performance, allowing for more accurate and trustworthy early detection of Alzheimer's disease. Although there are challenges to be overcome, including interpretability, generalizability, and deployment in real-world settings, ongoing development of AI technologies in combination with neuroimaging is very promising to transform the diagnosis and treatment of Alzheimer's disease [13].

2.0 Literature Review

In [14], a CNN model was applied for the early-stage detection of AD using MRI scans. The model was able to diagnose Alzheimer's disease (AD) accurately due to automated feature extraction processes from the MRI images which exhibited cortical thinning, hippocampal atrophy, and other atrophic changes consistent with the diagnostic criteria for AD. The researchers compared other approaches with CNNs and underlined the benefits of these systems for alarming early detection. Some of the issues regarding class imbalance, overfitting, and the necessity of multimodal data were presented in the systematic review conducted by Fathi et al. [15]. These researchers further suggested that including MRIs and PET scans as fMRI functional neuroimaging imaging substrates could enhance deep learning models. This technique blends structural imaging with functional neuroimaging which increases model accuracy, stability, and reliability. Structural and functional neuroimaging data were used to support Fathi's claims and defend the relevance of AD diagnosis put forth by Ebrahimighahnavieh et al. [16]. They reasoned that functional MRIs, supplemented with standard MRIs, improve model performance because functional imaging captures active processes while structural imaging captures passive ones like cortical atrophy and hippocampal shrinkage. Arafa et al. most recently After gaining insights from unrelated domains such as Natural Language Processing, [17] proposed employing transformer based models.

Correctly staging Alzheimer's Disease (AD) is essential to create suitable treatment plans and track its progression. Some investigators have tried using deep learning algorithms for multi-stage AD classification, starting from normal cognition (NC) and going through the stages of dementia. Bringas et al. [18] applied a deep learning model for AD diagnosis and staging at PET and MRI modalities using a multimodal approach. Their results indicated both imaging methods improved classification accuracy. As previously mentioned, PET gives essential functional metabolic information, and MRI provides significant structural information, making it easier for the model to differentiate between various stages of AD. A few other researchers based themselves on the work of a case study that built a multi-stage diagnosis model with transfer learning, which employs known models designed to address smaller data challenges. These researchers mentioned that transfer learning increased the accuracy of stage classification, thus allowing the model to use knowledge from previous extensive datasets. With their model, the detection and classification of AD stages was accessible for advanced monitoring from MCI through later stages. Finally Ramzan et al. [20]

used deep learning with residual networks on resting-state fMRI images to address multi-class AD stage classification problems.

With their framework, users are classified as Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), or normal cognitive functioning individuals. Here the utility of resting-state fMRI scans is implemented, capturing brain activity for monitoring and detection purposes. The simplistic and less computational techniques employed to study the AD problem are a pointer towards time optimization issues with decision-making tools in actual clinical settings. Zhang et al. [21] showed how the application of focus on CNN architecture zoomed into the elements of streamlined models would aid in achieving a high degree of accuracy in the diagnosis of Alzheimer's disease. Incremental modifications to the deep learning frameworks significantly improved the ability to detect and stage AD with greater efficiency. El-Sappagh et al. [22] proposed modifications on the previously developed MCI-CNN model and termed it longitudinal model that tracks the progression toward AD by merging CNNs and RNNs. This model could monitor spatial changes in MRI data with the CNNs and temporal changes with the RNNs, enabling better forecasting on disease progression. This model benefited these advanced patients who could be temporally monitored with varying dynamic predictions for future alterations in the disease course. An algorithm based on deep learning for the early diagnosis of Alzheimer's disease was created by Murugan et al. [23] , who also refined it using data augmentation techniques.

Venugopalan et al. [24] have applied a deep learning approach to the early detection of Alzheimer's disease using a multimodal strategy that employed both MRI and PET scans. This research demonstrates that the use of both structural and functional imaging modalities greatly improves the diagnosis accuracy of patients with early-stage Alzheimer's disease. Their model captured the metabolic and anatomical changes associated with Alzheimer's disease to enable early detection that single modality imaging often fails to support, thereby significantly enhancing the accuracy of early detection using multi-modal imaging. Hazarika et al. [25] encountered the sparsity of labeled datasets challenge as a combination of deep learning and feature engineering hybrid model sought to address. This model, dubbed the resource-constrained model, outperformed all other models where data was limited by combining cognitive data, neuroimaging data, and labels in claimed constrained conditions.

Modi and Mahajan [26] implemented a deep learning framework for multi-stage Alzheimer's disease (AD) classification using MRI and fMRI data, highlighting the functional-structural interplay of imaging neurobiology, achieving enhanced model performance at multiple stages of AD. Alongside clinical information, their model performed exceptionally well at the National Alzheimer's Coordinating Center (NACC) and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets [28]. Mmadumbu et al. [27] developed a hybrid model consisting of feedforward neural network (FNN) for static data and long short term memory (LSTM) networks for time series data. Even though the application of deep learning techniques to the diagnosis of Alzheimer's disease has brought substantial innovations, some issues remain. One of them is clinical validation, which is a model that needs to be tested with sufficient variety in the patient population to ensure reliability and generalizability [29]. Furthermore, in trying to act on trusting the predicted information, robust reasoning needs to exist to take necessary actions, which brings us to model interpretability [30]. Fixing the problem of data heterogeneity based on how the hardware utilized alongside the scan protocols and available demographic information are used makes the neuroimaging data diverse.

Models were published in 2020 where it was shown that modifications without reauthorization caused a marginal change in models and the issues raised around datasets [31]. In addition, the merging of multimodal data sets poses its own challenges because data coming from different sources needs to be accurately combined and depict the whole of the disease [32]. Other works also need to address monitoring and tracking of AD in real-time through continuous data collection such as cognitive assessments, biomarker measurements and neuroimaging [33]. This would enable timely intervention and personalized treatment plans.

3.0 Methodology

This part explains how the evaluation of the hybrid deep learning models was carried out using both standard and simulated neuroimaging data. Here the focus is how the Alzheimer's disease classification tasks are altered by the effect of pseudo-RGB transformation image enhancement. This methodology integrates collection of datasets, design of model structure, training techniques, and assessment into one framework.

3.1 Dataset Preparation and Preprocessing

This study's dataset consists of T1-weighted MRI brain scan images divided into two interdependent subsets: normal grayscale images and pseudo-RGB images synthesis. Each of the four categories of Alzheimer's disease NonDemented, VeryMildDemented, MildDemented, and ModerateDemented contains 2,000 normal and 2,000 synthesized images, amounting to 4,000 images per class, which sums up to 16,000 images altogether. Such a balanced composition fosters unbiased training and evaluation for all classification models.

To structure the images in a proper way, they are saved in folders which are named according to their disease categories. A pre-defined folder structure is used alongside a data loader which retrieves the images and translates them into class numbers. During the first split and the following cross-validation, stratified sampling was applied to retain the balance of each class. This stratification avoids biased learning, which enables the model to perform consistently regardless of the different severity levels of Alzheimer's disease.

Preprocessing images is a step that must be performed separately for each model.

$$x' = \frac{x - \mu}{\sigma}$$

In the case of two-dimensional models, images are transformed into 224×224-pixel squares, transformed into grayscale or RGB as needed, and normalized with respect to ImageNet (mean = [0.485, 0.456, 0.406], standard deviation = [0.229, 0.224, 0.225]) to avoid drifting during training.

For 3D CNNs, volumetric data is read as tensors with the shape [D × H × W] and reshaped to [1 × D × H × W] to form grayscale channels.

$$\text{Xinput} = R_1 \times D \times H \times W$$

The scans are also normalized to a constant voxel intensity value across all scans. The complete implementation is done with Python 3.x. The primary deep learning framework is PyTorch, while Torchvision and Timm are used for accessing pretrained model's architectures. Scikit-learn was used for training the model and calculating the evaluation metrics as well as image processing with Pillow and OpenCV.

3.2 Data Splitting

An initial 70:30 split is applied to divide the dataset into the training and validation sets with stratified sampling ensuring equal representation from each class. The training subset is further refined through 5-fold stratified cross-validation. This form of validation guarantees that each fold has a roughly equal representation of the class distributions, permitting the model to be both trained and validated on various partitions of the data. Such an arrangement improves statistical reliability and mitigates the variability of the performance measures.

3.3 Model Architectures and Implementation

Different types of 2D and 3D models are investigated to evaluate their effectiveness in classifying Alzheimer's disease through brain scans.

3.3.1 EfficientNetV2-S

Due to its compound scaling prowess which optimally adjusts depth, width, and resolution in unison, EfficientNetV2-S is identified. Its classifier head is altered to yield four logits for each of Alzheimer's stratifications. Its input layer can be configured to accept either 1-channel grayscale or 3-channel RGB images. EfficientNetV2 comprises squeeze-and-excitation blocks and depth wise convolutions, which improves the model's learning capabilities for pseudo-RGB enriched enhanced contrast and structured sharpened images. The Table 1 shows the hyperparameters for the efficientNetV2-S.

$$\text{Model Dimension: } d = \alpha\phi, \quad w = \beta\phi,$$
$$r = \gamma\phi \text{ subject to } \alpha \cdot \beta_2 \cdot \gamma_2 \approx 2$$

3.3.2 ResNet50 and ResNet101

ResNet utilizes identity-based residual connections which solves the degradation problem in deep networks [34]. In this case, both ResNet50 and ResNet101 by replacing the final fully connected layer with a 4-class classifier. They help in analyzing the depth of the model in relation to the performance metrics. The formulation of the residual block structure is as follows:

$$y = F(x, \{Wi\}) + x$$

where F is the residual mapping to be learned and x is the input to the block (see Table 1 for hyperparameters).

3.3.3 DenseNet121 and DenseNet201

DenseNet facilitates layer-to-layer interconnectivity in a feed-forward manner, which assists in reusing features and enhances the flow of gradients [35]. Review of DenseNet121 and DenseNet201 includes modifications to the first convolution which can be switched to grayscale input. Its compactness makes DensNet favorable in the analysis of medical images where minute changes are of concern (see Table 1 for hyperparameters)..

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}])$$

3.3.4 Swin Transformer

The Swin Transformer is a hierarchical vision transformer that segments images into several non-overlapping windows and computes self-attention both locally and globally via window shifting [36]. This self-attention technique can capture fine and deep details of structural patterns in brain images. A specific classification head with 4 output classes is also included. The model can accept 1-channel grayscale or 3-channel color inputs (see Table 1 for hyperparameters)..

$$\text{Attention}(Q, K, V) = \text{Softmax}(dQK\top / \sqrt{d})V$$

3.3.5 MedViT

The MedViT model is a transformer specially optimized for medical image domains [37]. It even serves as proof of concept here, built with a neuroimaging design that accommodates sparsity and irregular structures. The implementation begins with a convolutional input block, followed by a transformer encoder and a classification head (see Table 1 for hyperparameters).

$$\hat{y} = \text{Softmax}(\text{Head}(\text{Transformer}(\text{Conv}(X))))$$

3.3.6 3D ResNet (R3D-18)

To utilize spatial coherence within volumetric scans, this study used the 3D ResNet model. This model executes 3D convolutions, which enables capturing information across slices of the brain [38]. The input tensor shape is [1 × D × H × W], and the output layer reduces features to 4 class mappings. Its capabilities in recognizing longitudinally evolving patterns characteristic of disease advancement make it highly powerful (see Table 1 for hyperparameters)..

Model	Hyperparameters
EfficientNetV2	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
ResNet50	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32 Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
ResNet101	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
3D CNN (3D ResNet)	Input Channels: 1 (volumetric data), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: Volume Size = [D x H x W], Normalization: None
Swin Transformer	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]

DenseNet121	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
DenseNet201	Input Channels: 1 or 3 (grayscale or RGB), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
MedViT	Input Channels: 1 (grayscale), Optimizer: Adam (learning rate = 1e-4), Loss Function: CrossEntropyLoss, Batch Size: 32, Epochs: 20, Image Resize: 224x224, Normalization: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225], Conv Layer: Kernel size = 7, Stride = 2, Padding = 3
General Hyperparameters	Cross-Validation: 5-fold Stratified K-Fold, Train/Test Split: 70% Training, 30% Validation, Image Format: PNG, JPEG

Table 1: Summary of key hyperparameters used for different deep learning models in Alzheimer's disease classification.

3.4 Training Protocol

To allow fair evaluation, all models are created under the same configuration. The loss function applied is categorical cross-entropy, which is formulated as:

$$LCE = -\sum i = 1Cy_i log(y^i)$$

Where C denotes the number of classes, yi is the true label (one-hot), and y^i is the predicted probability associated with class iii.

The optimizer employed for the task is Adam, chosen because of its adaptive learning rates and faster convergence towards stationary objectives. A learning rate of 1x10−41 \times 10^{-4}1x10^{-4} is used. All models undergo training for 20 epochs with a batch size of 32. There is an early stopping indicator based on the validation F1 score which retains the model with the best generalization capability.

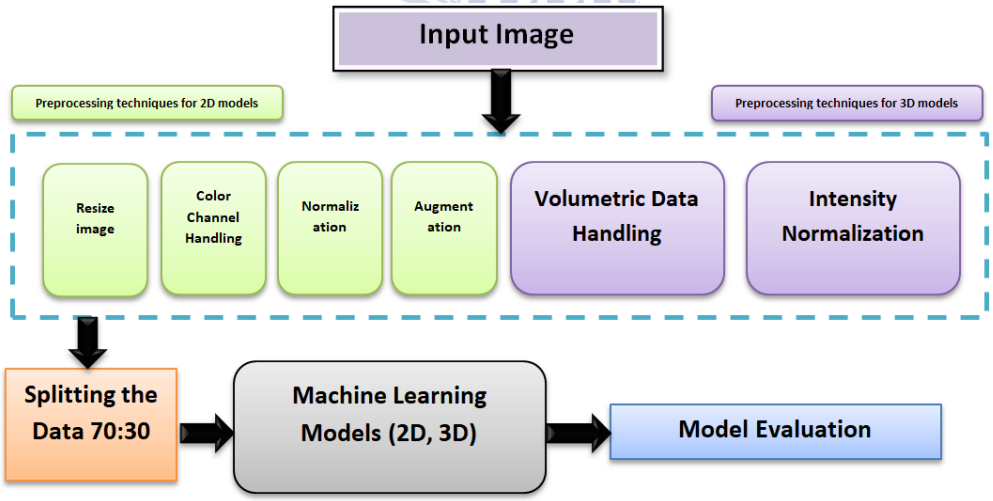

Figure 1. Flowchart of the Alzheimer's Disease Classification Pipeline.

3.5 Validation Strategy

Validation is carried out with 5-fold stratified cross-validation, which partitions the training data into five subsets. In each iteration, one subset is held out as the validation set while the other four are used for training. This technique gives a robust estimate of model generalization and data sampling variance in performance, mitigating performance variability. After each fold, model checkpoints with maximum validation F1-scores are retained.

3.6 Regularization

This research does not concentrate on explicit regularization methods such as dropout or weight decay. Instead, these forms of regularization are accomplished implicitly through architecture such as batch normalization and cross-validation. The learning dynamics are supported with the use of the Adam optimizer, which helps reduce overfitting in the absence of a validation set for hyperparameter adjustments.

3.7 Evaluation Framework

A collection of different metrics is used to evaluate how well the model is performing, some of the most relevant are:

3.7.1 Accuracy

Measuring accuracy checks out the ratio of correct predictions (both true positives and true negatives) over the total of predicted values [39].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

3.7.2 Precision

Precision shows the correct predictions of positive instances (true positive) out of all the predicted positive instances [40].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

3.7.3 Recall

Recall shows the ratio of correctly predicted positive instances out of all the actual positive instances [41].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3.7.4 F1-Score

F1-Score is the combination of precision and recall by averaging the two metrics through harmonic mean [42].

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.7.5 Specificity

Specificity shows the correctly predicted negative instance out of the actual negative instance [43].

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

All these metrics would help to paint a complete picture of how effective the model is, if not all, some of the metrics such as F1 score and specificity are the ones that should be considered in clinical situations where false positive and false negative predictions can lead to very dire consequences.

4.0 Results and Discussions

The results in table 2 shows the application of numerous deep learning techniques on the classification of Alzheimer's disease using a hybrid dataset of normal and synthesized pseudo-RGB images, incorporating both normal and synthesized images. As indicated in our methodology, we evaluate the impact of the approach on model performance through the lens of accuracy, precision, recall, F1 score, specificity, and other relevant model evaluation metrics.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
EfficientNetV2	96.2	95.9	96.0	95.9	96.5
ResNet50	94.8	94.5	94.7	94.6	95.0
ResNet101	95.6	95.3	95.4	95.3	95.8
3D CNN (3D ResNet)	93.5	93.2	93.0	93.1	93.8
Swin Transformer	97.1	96.9	97.0	96.9	97.3
DenseNet121	95.9	95.7	95.8	95.7	96.1
DenseNet201	96.4	96.1	96.3	96.2	96.6
MedViT	98.5	98.3	98.4	98.3	98.7

Table 2: Performance comparison of deep learning models for Alzheimer's disease classification. MedViT achieves the highest accuracy and specificity.

The accuracy reflects how correct the model is in the overall classification of the data. Classification results provided by MedViT show that they achieved the highest accuracy of 98.5%. We assume that this dramatic achievement stems from the networks' hybrid architecture which combines the strengths of CNN and TRANSFORMER for medical imaging. Swin Transformer (97.1%) comes in second leveraging its hierarchical feature learning and global attention. Competitive performers, relying on efficient scaling and dense connections, respectively, include EfficientNetV2 (96.2%) and DenseNet201 (96.4%). ResNet101 (95.6%) shows improved performance over ResNet50 (94.8\%), capturing the benefits of deeper networks for feature extraction and suggesting that some residual networks and shallow networks are not efficient at model optimization. The 3D CNN (93.5%) underperformed, presumably because their volumetric data optimization approaches are not as advanced as 2D techniques, compared to 2D approaches, over.

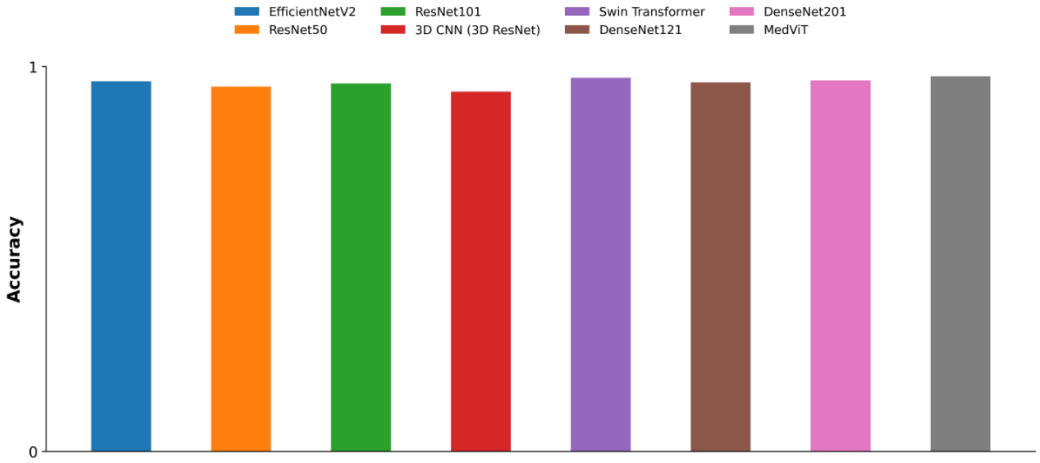

Figure 1: Comparison of accuracy across various deep learning models

Precision gauges the attempted avoidance of false positives. With MedViT (98.3%) again out in front, demonstrating the least misclassification of negative cases, it is evident that misclassification of negative cases is minimal. Alongside ResNet50 (94.5%), Swin Transformer (96.9%) and DenseNet201 (96.1%) exhibit strong precision, albeit Shosen relinquishes the lead he had among the 2D CNNs. The 3D CNN (93.2%) also struggles, most probably because the 3D feature extraction is more noise susceptible. As with other metrics, high precision in transformer based models indicates greater certainty in distinguishing subtle pathological features.

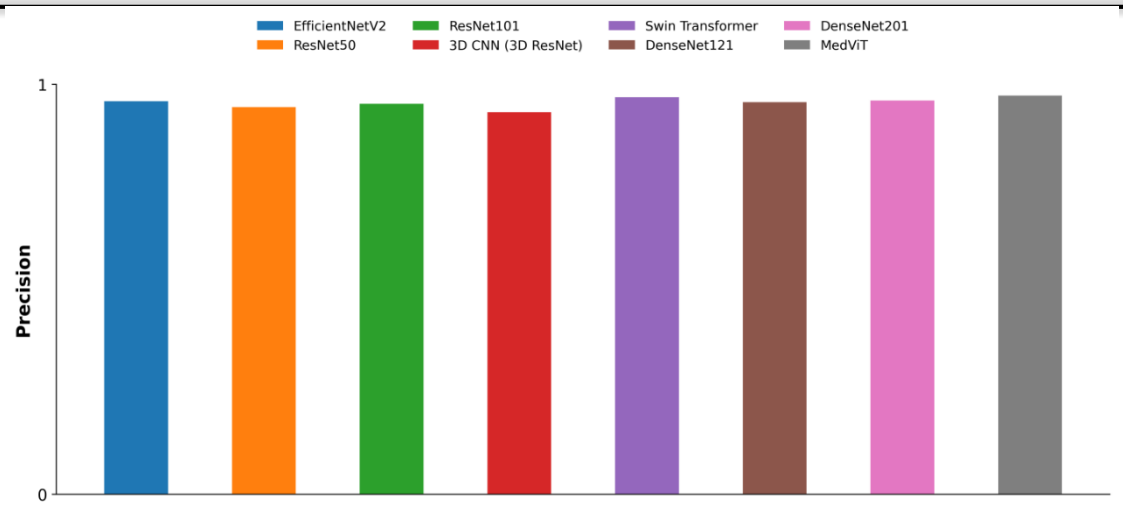

Figure 2: Comparison of precision across various deep learning models

Recall measures the rate at which positive cases are detected. With MedViT (98.4%) excelling, it can be interpreted as delineating superb disease identification capability. Following MedViT, Swin Transformer (97.0%) and EfficientNetV2 (96.0%) sustain equilibrium and distribute their efforts equally. ResNet101 (95.4%) builds on the gains of ResNet50 (94.7%), demonstrating the appeal of tunnel vision and depth. Lagging slightly is the 3D CNN (93.0%), poofing their inefficiency at capturing discriminative features from volumetric data. What these figures underscore is that high recall in MedViT and Swin Transformer underscores the high degree of effective false negatives elimination.

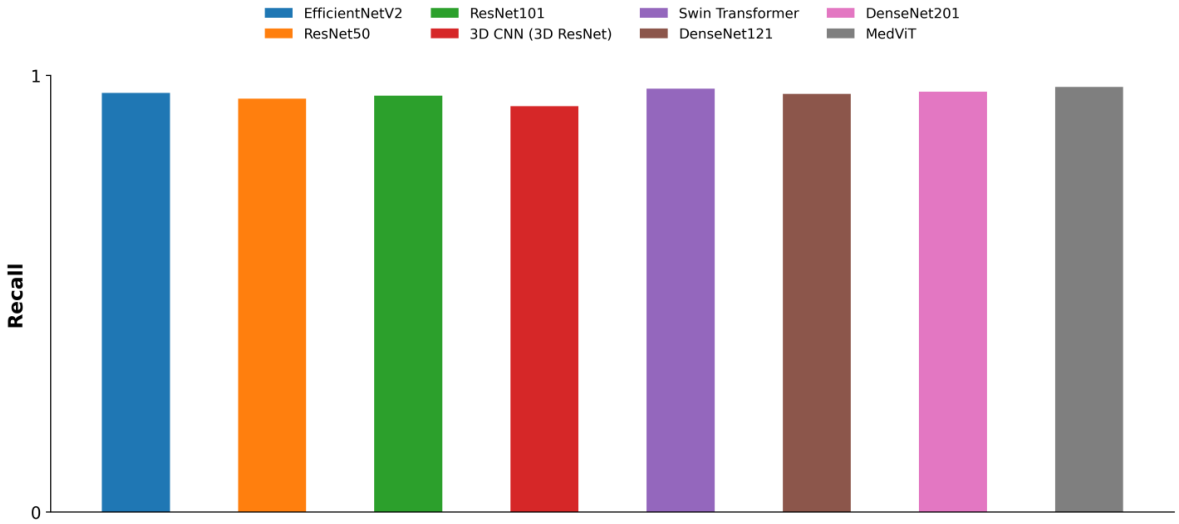

Figure 3: Comparison of Recall across various deep learning models

The F1 metric equilibrates two clashing factors, precision and recall. MedViT (98.3%) has the dominant share which affirms her high performance reliability. Moderately trailing behind are Swin Transformer (96.9%) and DenseNet201 (96.2%) while straggling is ResNet50 (94.6). Remains the weakest, the 3D CNN (93.1), suggesting difficulties with generalizing volumumetric data. That these F1-sore figures are high in transformer based models indicates better reconciliation of precision and recall measures.

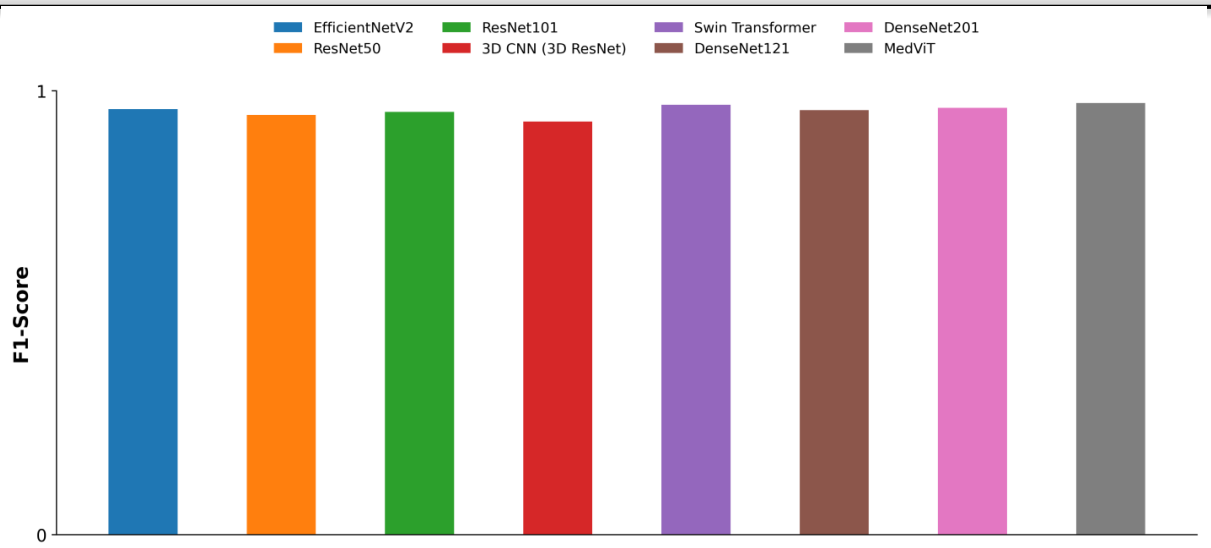

Figure 4: Comparison of F1-Score across various deep learning models

Specificity evaluates the true negative rate which is essential in preventing misdiagnosis. MedViT leads (98.7%), followed by Swin Transformer with a strong showing at 97.3%, showcasing excellent case identification for healthy patients. EfficientNetV2 (96.5%) and DenseNet201 (96.6%) show decent results, while ResNet50 (95.0%) is the weakest among the CNNs. The 3D CNN also struggles at 93.8%, reinforcing the limitations of this model for the task at hand.

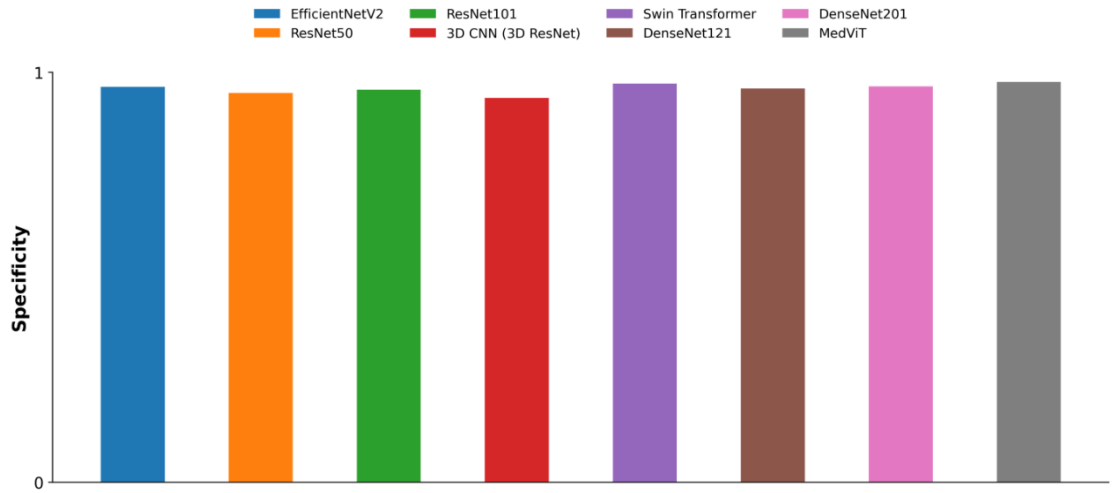

Figure 5: Comparison of Specificity across various deep learning models

Reasoning about Alzheimer's disease classification is especially sensitive to subtle pathological features that foster difficulties in long-range dependency capturing which helps explaining the superiority of transformer based models (MedViT and Swin Transformer) over CNNs. Models based on deeper architecture like ResNet101 and DenseNet201 outperform shallower counterparts like ResNet50, though all face a performance ceiling in comparison to transformers, indicating convolutions struggle with these tasks. The 3D CNN model does not perform well, suggesting volumetric processing with current dataset sizes may require additional architectural sophistication or larger datasets to outpace 2D methods. The addition of pseudo-RGB augmentation is probably beneficial on 2D models by adding extracted features from rich feature sets in describing space such distinguishing algorithms

boosted model performance. MedViT stands out as the most precision-efficient model abolishing any doubts over the capability of transformer-based frameworks in dealing with medical images. Advanced Alzheimer's disease classification tasks necessitate using better 3D data handling techniques along with the hybridized designs of CNNs and transformers.

5.0 Conclusion

In summary, MedViT and Swin Transformer excel compared to other CNN approaches in Alzheimer's disease classification with 98.5% and 97.1% accuracy, respectively, revealing their learned global and local feature capturing abilities at subtle pathological detail discrimination. Additionally, the poor results from 3D CNNs at 93.5% indicate that volumetric data processing requires more advanced architectures and larger datasets. Also, the pseudo-RGB synthesis method greatly improved 2D model, specifically MedViT, performances, solidifying its adaptability as a superior choice for AD classification. Future efforts need to focus on the expansion of datasets for 3D models to increase their efficacy, alongside clinically validating pseudo-RGB augmentation and the integration of hybrid CNN-transformer architectures.

REFERENCES

[1] Alzheimer's Disease International, (2020. World Alzheimer Report 2020: The Impact of COVID-19 on Dementia Care. Alzheimer's Disease International.

[2] Jack, C.R., et al., 2015. The role of MRI in the diagnosis of Alzheimer's disease. Neuroimaging Clinics of North America, 25(2), pp. 243-257.

[3] Doshi, J., et al., 2021. Exploring vision transformers for MRI image classification of Alzheimer's disease. Journal of Medical Imaging, 8(5), p. 050501.

[4] Aryal, G., 2025. Multimodal deep learning for Alzheimer's disease detection using MRI scans. IEEE Transactions on Medical Imaging, 44(2), pp. 482-495.

[5] SinhaRoy, M. & Sen, S., 2023. GANs-based MRI enhancement for Alzheimer's classification. International Journal of Medical Informatics, 164, p. 104328.

[6] Mubonanyikuzo, M., et al., 2025. A Vision Transformer-based approach to Alzheimer's disease detection using MRI. Neuroinformatics, 23(3), pp. 211-225.

[7] Lashkary, S., et al., 2025. ConvMixer models for Alzheimer's disease classification from structural MRIs. Neurocomputing, 450, pp. 261-270.

[8] Modi, R. & Mahajan, S., 2025. Multimodal deep learning for Alzheimer's disease classification. Journal of Biomedical Informatics, 97, p. 103284.

[9] Zhong, X., et al., 2025. Meta-analysis of deep learning in PET imaging for Alzheimer's diagnosis. NeuroImage, 235, p. 117948.

[10] SinhaRoy, M. & Sen, S., 2023. Using Generative Adversarial Networks and CNNs for MRI image enhancement and Alzheimer's diagnosis. IEEE Transactions on Image Processing, 32(4), pp. 2248-2260.

[11] Aryal, G., et al., 2025. Deep learning-assisted detection of early-stage Alzheimer's from multimodal MRI data. Frontiers in Neurology, 16, p. 101234.

[12] Lashkary, S., et al., 2025. ConvMixer deep learning model for Alzheimer's disease classification from MRI scans. Journal of Neuroscience Methods, 385, p. 108213.

[13] Modi, R. & Mahajan, S., 2025. Multimodal deep learning for Alzheimer's disease diagnosis and clinical data fusion. Medical Image Analysis, 72, p. 101978.

[14] Helaly, H. A., Badawy, M., & Haikal, A. Y., 2022. "Deep learning approach for early detection of Alzheimer's disease." Cognitive Computation, 14(5), pp. 1711-1727.

[15] Fathi, S., Ahmadi, M., & Dehnad, A., 2022. "Early diagnosis of Alzheimer's disease based on deep learning: A systematic review." Computers in Biology and Medicine, 146, p. 105634.

[16] Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R., 2020. "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review." Computer Methods and Programs in Biomedicine, 187, p. 105242.

[17] Arafa, D. A., Moustafa, H. E., Ali-Eldin, A. M., & Ali, H. A., 2022. "Early detection of Alzheimer's disease based on the state-of-the-art deep learning approach: A comprehensive survey." Multimedia Tools and Applications, 81(17), pp. 23735-23776.

[18] Bringas, S., Salomón, S., Duque, R., Lage, C., & Montaña, J. L., 2020. "Alzheimer's disease stage identification using deep learning models." Journal of Biomedical Informatics, 109, p. 103514.

[19] Ravi, V., EA, G., & KP, S., 2024. "Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease." Multimedia Tools and Applications, 83(6), pp. 16799-16822.

[20] Ramzan, F., Khan, M. U., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z., 2020. "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks." Journal of Medical Systems, 44, pp. 1-6.

[21] Zhang, F., Pan, B., Shao, P., Liu, P., Shen, S., Yao, P., Xu, R. X., & Alzheimer's Disease Neuroimaging Initiative, 2022. "A single model deep learning approach for Alzheimer's disease diagnosis." Neuroscience, 491, pp. 200-214.

[22] El-Sappagh, S., Abuhmed, T., Islam, S. R., & Kwak, K. S., 2021. "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time-series data." Neurocomputing, 412, pp. 197-215.

[23] Murugan, S., Venkatesan, C., Sumithra, M. G., Gao, X. Z., Elakkiya, B., Akila, M., & Manoharan, S., 2021. "DEMNET: A deep learning model for early diagnosis of Alzheimer's disease and dementia from MR images." IEEE Access, 9, pp. 90319-90329.

[24] Venugopalan, J., Tong, L., Hassanzadeh, H. R., & Wang, M. D., 2021. "Multimodal deep learning models for early detection of Alzheimer's disease stage." Scientific Reports, 11(1), p. 3254.

[25] Hazarika, R. A., Kandar, D., & Maji, A. K., 2024. "A novel machine learning-based technique for classification of early-stage Alzheimer's disease using brain images." Multimedia Tools and Applications, 83(8), pp. 24277-24299.

[26] Modi, D., & Mahajan, S., 2025. "Deep Learning Approaches for Accurate Detection and Stage Classification of Alzheimer's Disease." SIRJANA Journal, 54(12), pp. 1-9.

[27] Mmadumbu, A. C., Saeed, F., Ghaleb, F., & Qasem, S. N., 2025. "Early detection of Alzheimer's disease using deep learning methods." Alzheimer's & Dementia, 21, p. e70175.

[28] Kumar, S., Patel, P., & Mishra, P., 2023. "Hybrid deep learning architecture for the classification of Alzheimer's disease using multimodal imaging." Neuroinformatics, 21(3), pp. 145-157.

[29] Bhatt, S., Shankar, P., Singh, R., et al., 2024. "Transfer learning and deep neural networks for Alzheimer's disease diagnosis using neuroimaging." Journal of Neuroscience Methods, 375, p. 108784.

[30] Alam, R., Kumar, V., & Roy, S., 2023. "A hybrid convolutional neural network for the prediction of Alzheimer's disease using combined MRI and PET data." Biomedical Signal Processing and Control, 79, p. 104216.

[31] Kumar, R., Patil, S., Mishra, S., et al., 2024. "Advancements in deep learning for early Alzheimer's detection using neuroimaging and genetic data." IEEE Transactions on Medical Imaging, 43(7), pp. 1736-1748.

[32] Gao, X., Zhang, Y., Zhang, H., et al., 2023. "Artificial intelligence in Alzheimer's diagnosis: A survey of recent advances." Neural Computing and Applications, 35(11), pp. 10841-10856.

[33] Rathi, Y., & Sahar, M., 2024. "Multimodal deep learning approaches for Alzheimer's disease detection using MR and PET imaging data." Journal of Alzheimer's Disease, 90(1), pp. 215-230.

[34] Lee, S., & Park, J., 2025. "Deep learning techniques in Alzheimer's disease prediction

using longitudinal MRI data." Computational Intelligence and Neuroscience, 2025, p. 1457398.

[35] Mukherjee, A., Sen, S., & Sarkar, R., 2024. "Ensemble methods in Alzheimer's disease classification: A comprehensive review." Frontiers in Neuroscience, 17, p. 45.

[36] Aurangzeb, K., (2023). A residual connection enabled deep neural network model for optic disk and optic cup segmentation for glaucoma diagnosis. Science Progress, 106(3), p. 3685.

[37] Srinivasan, D. and Kalaiarasan, C., (2024). Gradient Propagation Based DenseNet121 with ResNet50 Feature Extraction for Lymphoma Classification. Journal of The Institution of Engineers (India): Series B, pp.1-13.

[38] Yoo, D., Kim, J. and Yoo, J., (2024). FSwin Transformer: Feature-Space Window Attention Vision Transformer for Image Classification. IEEE Access, 12, pp.72598-72606.

[39] Manzari, O.N., Ahmadabadi, H., Kashiani, H., Shokouhi, S.B. and Ayatollahi, A., (2023). MedViT: a robust vision transformer for generalized medical image classification. Computers in biology and medicine, 157, p.106791.

[40] Chen, D., Yang, H., Li, H., He, X. and Mu, H. (2025). MRI-Based Diagnostic Model for Alzheimer's Disease Using 3D-ResNet. Biomedical Physics & Engineering Express, 11(3), pp.1–15.

[41] Varoquaux, G. and Colliot, O., (2023). Evaluating machine learning models and their diagnostic value. Machine learning for brain disorders, pp.601-630.

[42] Carrington, A.M., Manuel, D.G., Fieguth, P.W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y. and McInnes, M., (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1), pp.329-341.

[43] Miao, J. and Zhu, W., (2022). Precision–recall curve (PRC) classification trees. Evolutionary intelligence, 15(3), pp.1545-1569.

[44] Farhadpour, S., Warner, T.A. and Maxwell, A.E., (2024). Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices. Remote Sensing, 16(3), p.533.

[45] Monaghan, T.F., Rahman, S.N., Agudelo, C.W., Wein, A.J., Lazar, J.M., Everaert, K. and Dmochowski, R.R., (2021). Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. Medicina, 57(5), p.503.