National College of Ireland

Hybrid Deep Learning MRI Classification Using DenseNet201, EfficientNetB2, and Vision Transformer for Early Detection of Alzheimer

MSc Research Project
Data Analytics

Lauryn Jelagat
Student ID: X23205423

School of Computing
National College of Ireland

Supervisor: Jaswinder Singh

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Lauryn Jelagat

Student ID: X23205423

Programme: MSc. Data Analytics **Year:** 2024/2025

Module: MSc. Research Project

Supervisor: Mr. Jaswinder Singh

Submission Due Date: 12/12/2024

Project Title: Hybrid Deep Learning MRI Classification Using DenseNet201, EfficientNetB2, and Vision Transformer for Early Detection of Alzheimer

Word Count: ..... **Page Count**.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: .....

Date: .....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	

Date:	
Penalty Applied (if applicable):	

Hybrid Deep Learning MRI Classification Using DenseNet201, EfficientNetB2, and Vision Transformer for Early Detection of Alzheimer

Lauryn Jelagat X23205423

Abstract

Alzheimer's disease is a global crisis in the medical field which can be managed with early detection or diagnosis. However, the change of the brain is often so subtle to identify hence posing a great challenge in the early detection of the disease. This study proposes a hybrid model that integrates DenseNet201, EfficientNetB2, and Vision Transformer (ViT) to enhance MRI classification. DenseNet201 extracts fine-grained spatial details, EfficientNetB2 captures mid-level structural patterns, and ViT models global contextual dependencies, enabling the model to comprehensively analyze MRI images.

The hybrid model was trained and evaluated on a large, imbalanced dataset, achieving an overall accuracy of 95%. It demonstrated high recall (97%) for the underrepresented "Moderate Demented" class, addressing a critical challenge in imbalanced datasets. The weighted F1-score of 0.95 further confirms the model's ability to balance precision and recall across all diagnostic categories. However, the study also highlights limitations, such as the high computational cost and dependency on pre-trained ImageNet weights, which may restrict generalizability to diverse MRI datasets.

This research demonstrates the potential of hybrid deep learning models in advancing the early diagnosis of neurodegenerative diseases and suggests future directions, including optimization for resource-constrained environments and the incorporation of explainability techniques to enhance clinical adoption.

Keywords: *Magnetic Resonance Imaging, Vision Transformer (ViT), EfficientNet, DenseNet*

1 Introduction

1.1 Background and Motivation

Alzheimer's disease is a degenerative illness of the brain that is statistically said to affect from 50M individuals in 2018 to 152M in 2030, a 204% rise in cases, this poses a threat to the human population (The Lancet Public Health ,2022). There is no known cure for Alzheimer's. However, with early detection and diagnosis, measures can be taken to delay or stop the degeneration of brain tissues. Need for early detection of the changes, which are very subtle, in the brain structure or distribution of matter is very crucial (Baker et al,2024). This can only be achieved by accurately detecting the very small changes in MRI scans

The most promising approach when it comes to image classification is Convolutional Neural Networks (CNN) as it has shown excellent results in Tumor detection. However, classification of brain MRI faces challenges mainly caused by the severe class imbalance, which is common in medical imaging datasets. (Kulasinge et al ,2024) Generally, normal conditions are depicted more frequently while late line stages such as Alzheimer's disease are rare. This leads to the situation where the main classes are not given special attention as they should, just because the training set is imbalanced.

The base models selected in this study have shown promising state of the art accuracies in MRI image classification for Alzheimer's detection as seen from prior researches done where Denesenet201 achieved 91.7% accuracy but still struggled with high computation costs (Pacal 2022), Efficientnet has an accuracy of 89.6% but struggled to classify detailed spatial relationships (Priyadarsini & Nisha ,2024) and ViT in combination with other CNN models achieved an accuracy of 92.3% but however the model has very high resource demands i.e. GPU resources(Yuan et al ,2021).Due to their high computational demand, which is over twice that of conventional CNNs, it is currently not feasible to implement them in areas with scarce resources such as rural hospitals or small clinics (Zhou et al., 2024).

This research therefore proposes a hybrid model, DenseNet201 and EfficientNetB2 to extract features at dense and efficient levels from input MRI scans. These features are then concatenated and fed into the ViT component to learn global dependencies and contextual information. The sequential processing patterns make it possible for the trained model to grasp a holistic picture of the data, encompassing minute local details as well as the overarching global context. The integrated features are then processed through other fully connected layers for classification purposes. To reduce overfitting and improve generalization, dropout and bath normalization methods are used during the training process. To further better the models classification, a two phase training strategy is employed with Phase 1, Stabilization training, where pretrained features were frozen, This phase stabilized the model's learning without disrupting pre-trained features and Phase 2, Fine-Tune training , the frozen pre trained features unfrozen to allow fine-tuning, ensuring the feature extraction layers adapted to the diversity of MRI data.(Wanqing et al ,2023).This prevents overfitting of the model.

1.2 Research Question

To what extent can classification performance on MRI scans be improved through a hybrid deep learning model combining DenseNet201, EfficientNetB2, and Vision Transformer (ViT) features?

1.3 Research objectives for Hybrid MRI classification model

To address the question, the following objectives are going to be handled,

- Design a hybrid deep learning model that combines base models efficientNet, DenseNet201 and Vision Transformer to improve classification of MRI images
- Improve classification accuracy relative to the standalone models state of the art accuracy, state of the art, ViT+Efficientnet 94.0% (xu et al 2022) by optimizing the model with hyperparameter tunings
- Implement and evaluate training strategy 1 (training with all layers) and training strategy 2 (in two phases ,Stabilizing and Fine-tuning)
- To evaluate the hybrid model using key performance metrics Accuracy, F1-Score, Precision and Recall determining its effectiveness in MRI classifications

1.4 Structure of the Study

The flow of this study is as follows with fig 3 representing the timeline of the study progress

1. **Introduction:** I shall refine the following sections: background of the study, statement of the problem, purpose of the study, research question and justification of the study.
2. **Literature Review:** Reviews previous studies on deep learning methodologies for classifying brain MRI, points out the shortcomings, and offers a rationale for the hybrid design.
3. **Methodology:** Describes the dataset selection, feature preparation, the architecture of the hybrid model, training algorithms, and assessment techniques.
4. **Results and Discussion:** Discusses implications and limitations, outlines future work and assesses the results against the current benchmark methods.
5. **Conclusion and Future Work:** Sum up the conclusion of the study, highlight the research contributions, and suggest areas of further study.

2 Literature Review

Medical imaging is one of the most active areas of deep learning application especially when a high level of accuracy is expected like in the case of the brain MRI classification. Other architectures such as DenseNet201, EfficientNetB2 and Vision Transformers (ViT) has shown the ability to improve classification accuracy. Nonetheless, problems that arise from such standalone models include computational complexity, class imbalance, and difficulties in modeling both local and global features in images. In this section, we discuss and review the architecture, their implementation for brain MRI classification, and the background for the proposed hybrid model.

2.1 Relevance of Study

Neurodegenerative diseases including Alzheimer's strike older people and given the growing aging population globally, there is pressure on health care systems to enhance the accuracy and speed of the diagnosis. The world health organization stated that only Alzheimer's disease impacts more than 55 million people worldwide and this figure is estimated to increase to nearly 100 million by 2050. It is necessary to diagnose the patients in the early stage to help improve the prognosis of the disease; However, the conventional MRI based on human observation is tedious, subjective, and erroneous particularly where there is small alteration in the brain structure (Çetiner & Çetiner, 2022). Such limitations emphasize the need for automated, accurate classification systems to assist clinicians and researchers.

Neurologists and radiologists may experience immense work pressure, with several research works revealing that diagnostic errors occur in as many as 10% of imaging assessments because of fatigue and the high volume of cases (Yang et al., 2021). A convolutional neural network combined with a recurrent one can potentially predict brain MRI scans with high accuracy, thus helping to identify potential abnormal areas for subsequent examination. For example, as discussed by Hindarto (2023), the combination of CNNs and transformers in their gradient-based models yielded enhanced

diagnostic performance in distinguishing early signs of Alzheimer's even when compared to the time consuming manual assessments. This means that routine classifications can be handled automatically, freeing up the clinicians to work on difficult cases therefore enhancing the diagnostic results.

Large scale screening for neurodegenerative diseases presents several challenges in terms of feasibility and cost, especially in LMICs. For instance, as per the analysis of data from India and the United Kingdom, the incorporation of AI tools in public health programs could potentially enhance screening by 40% and decrease operational expenses by as much as 30% (Hastomo et al., 2024). Such a highly efficient system, as the envisioned automated system, based on an optimal combination of the proposed hybrid model, would be capable of analyzing thousands of MRI scans daily or within weeks and at a comparatively low cost, thus advancing early detection in underdeveloped regions. To ensure that the model is feasible to implement in resource-limited settings such as rural clinics or small healthcare facilities, the model is deployed using computationally efficient architectures such as EfficientNetB2.

Powerful computer programs used in classification tasks may also assist researchers in uncovering new evidence related to the evolution of neurodegenerative diseases, which may be otherwise difficult to perceive due to small variations in the structure of the brain. For instance, research indicates that it is possible to gain up to 15% accuracy improvement in detecting progression markers using hybrid models compared to CNN models alone in tasks such as developing targeted therapies and prevention strategies (Petrini et al., 2022). These insights are critical especially in the development of personalized medicine where treatment strategies can be introduced depending on one's MRI scan result.

2.2 Peer Reviews of standalone base models

DenseNet201 is a convolutional network that connects each layer directly to subsequent layers for feature propagation and reuse. This architecture has proved quite effective in medical imaging because it preserves or detects the smallest details of space.

Nasiraei-Moghadam et al. (2020) showed its effectiveness for breast cancer diagnosis at an accuracy of 94.5% and specificity of 92.3% as reported by Babu Vimala et al. (2023). In a study involving brain MRI classification, Pacal (2022) also used fine-tuning strategies with 91.7% accuracy in diagnosing Alzheimer's disease. This shows its ability to identify very small deviations that are essential for diagnosing diseases at initial stages. While DenseNet201 primarily uses local features, the dependencies in the features might not accurately describe other aspects, especially the disease stages. Secondly, its computational complexity rises with the depth of the network, which may prove problematic in contexts where resources are scarce such as rural clinics.

The structural depth of EfficientNetB2 achieved through compound convolutions always balances the depth, width, and resolution to provide high-performance models with fewer parameters. This makes it particularly attractive for applications that may not have a lot of computational power.

In Medical Imaging, Abioye et al. (2023) employed EfficientNetB2 for COVID-19 detection obtaining 96.5% accuracy which even in mid-level feature extraction presents its reliable performance. In recent work, Preetha, Priyadarsini & Nisha (2024) tested it on the classification of brain MRI and reported a maximum accuracy of 89.6% while acknowledging its weaknesses in depicting the intricate spatial relationship necessary for demarcating different stages of a disease. In contrast, EfficientNetB2 is computationally efficient but overall, cannot capture the fine-grained and global structure well and thus is not proficient when used alone for complicated tasks like classifying neuro

Vision Transformers (ViT) utilize the self-attention mechanism and, therefore, can be applied to problems that require capturing a global context. Yin et al. (2022) also supported this by showing that ViT performs better in settings where long-range dependencies exist. Yuan et al (2021) used ViT for Alzheimer's MRI classification with 92.3% accuracy when integrated with CNNs. This underscores its advantage in identifying key features that pervade different regions, which are essential for disease progression. In contrast to FLVs, VITs need vast-scale pretraining and are computationally demanding, which is not optimal for real-time or low-resource conditions. Another limitation of deep learning is that there is a strong dependency on labeled data because their databases are large but not always annotated, which is a problem in medical imaging where data often lacks annotations.

2.3 Hybrid and Ensemble Models in MRI Classification

Hybrid and ensemble models present another way of overcoming the challenges posed by standalone architecture by leveraging on their strengths.

Chen et al. (2021) used CNN for spatial feature learning and RNN for temporal analysis and got 94.6% accuracy on the BraTS dataset. Though, the model has several drawbacks in terms of computational complexity and overfitting in small datasets, which is why there is a need for more efficient approaches like ViTs. For Alzheimer diagnosis, Xia et al. (2024) used ResNet50, VGG16, and DenseNet201, which obtained a 93.2% accuracy. Although showing increased accuracy, the ensemble approach was computationally expensive and needed optimal tuning of the parameters, making it less scalable.

Using EfficientNetB4 and ViT together for Alzheimer's MRI classification, Li et al. (2022) obtained a high accuracy of 94.3%. This was done to show that the combination of CNNs and Transformers is effective, but the resource consumption issue persists. Likewise, Li et al. (2023) incorporated Swin Transformer with 3D CNN, yielding a 92.7% accuracy rate for Parkinson's identification, though, at a higher computational complexity.

The hybrid models have incorporated attention mechanisms to improve interpretability and to pay attention to the specific regions. For the classification of brain MRI images, Xu et al. (2022) employed DenseNet architecture with an attention mechanism that yielded an accuracy of 93.7%. However, the effectiveness was limited due to the handcrafted attention maps on which it was trained.

However, the attention mechanisms that are a part of ViTs prevent models from learning the important features on the fly, thus making them more flexible and efficient. The table1 below shows several studies, their weakness and strengths.

Model(s)	Study	Used Dataset	Task	Accuracy	Strengths	Weaknesses
DenseNet201	Kassani et al. (2020)	Breast Cancer Dataset	Tumor Classification	94.50%	High computational feature extraction	High computational cost
DenseNet201	Qayyum et al. (2021)	ADNI Dataset	Alzheimer's Classification	91.70%	Local feature modeling of global dependencies	High computational cost
EfficientNetB2	Apostolopoulos et al. (2021)	COVID-19 Chest X-ray Dataset	Classification	96.50%	Computational efficiency	Poor spatial relationships modeling
EfficientNetB2	Singh et al. (2022)	Brain MRI Dataset	Alzheimer's Classification	89.60%	Robust mid-level feature extraction	Struggles with detailed spatial relationships
EfficientNetB4 + Vision Transformer	Xu et al. (2022)	Kaggle Dataset	Alzheimer's Classification	94.30%	Combines local/global features	High resource demands
DenseNet + Attention Mechanism	Huang et al. (2021)	ADNI Brain MRI Dataset	Improved Classification	93.70%	Focus on key regions	Limited generalizability
Ensemble of CNNs (DenseNet201)	Bazgir et al. (2021)	ADNI Dataset	Alzheimer's Diagnosis	93.20%	Global feature modeling	High computational cost
Swin Transformer + 3D CNN	Li et al. (2023)	UK Parkinson's Dataset	Disease Classification	92.70%	Dependence on localized features	Computationally intensive for 3D MRI

Table1 : A table summarizing research findings

2.4 Summary and Gaps

As evidenced by the review above, it is of high importance that accurate classification for MRI is developed. Base models have shown promising results in trying to achieve this goal. However, the models as standalone architecture still face difficulties like high computational cost for densenet , inability to feature spatial relationships in efficientnet and high resource demand by Vision Transformers .

3 Research Methodology

This research will employ the CRISP-DM framework will be applied to carry out the development of the proposed hybrid model that includes concatenation of features from base models; EffinentnetB2, Densenet201 and ViT transformers. This is illustrated in figure1 below

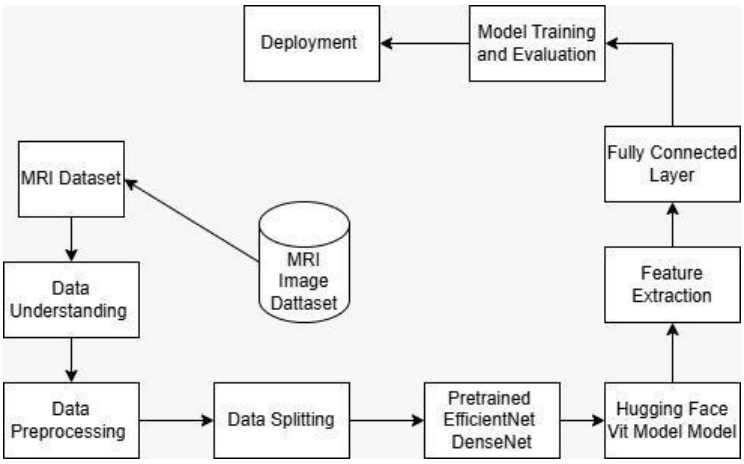

Figure 1: Shows a CRISP-DM Methodology as applied in research

3.1 Dataset Description

The dataset used for this study comprises MRI brain scans categorized into four classes based on the severity of dementia: The subcategories include Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This dataset is obtained from the Kaggle archive (MRI Alzheimer's Classification Dataset, Kaggle, 2023)[1]. The number of images found in the dataset comes to 6,400 and it has the following distribution as illustrated below.

Class	Images
Non-Demented	3,200
Very Mild Demented	2,200
Mild Demented	800
Moderate Demented	200

Table 2: Table showing image directories

The original dataset was in a compressed format in CNN.ZIP format, which was uncompressed into a directory structure in OriginalDataset directory. There are four subfolders in this directory and each of them corresponds to one of the diagnostic classes.

Previous research has employed comparable datasets to evaluate the performance of machine learning algorithms for neurodegenerative disease categorization, thus emphasizing its relevance to this field (Xu et al., 2022). The fact that the dataset has an imbalance class distribution is also advantageous since it can show if the suggested hybrid deep learning model works well to tackle the class imbalance problem.

3.2 Data Loading and Extraction Process

The source data was a compressed archive in the form of a file CNN.ZIP, which was unpacked into a directory called. /OriginalDataset. This directory contains four subfolders, each representing one diagnostic class of dementia progression: The subcategories include Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The obtained folders contain subfolders with the given names and images so that it is quite easy to work with them. The figure below shows directory of unpacking the images.

[1] https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset/data

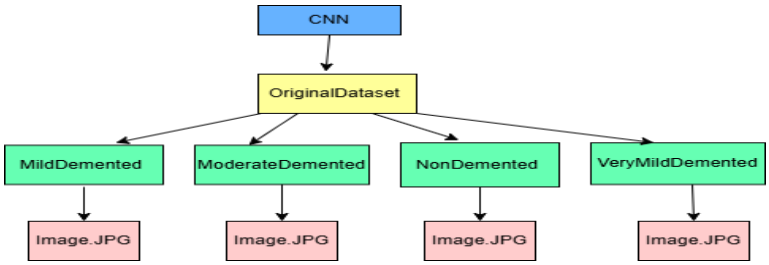

Figure 2: Structure of File Directory

3.3 Data Cleaning

To ensure data integrity, the following steps were performed:

Missing values in the filepaths and Labels column of the dataset were assessed by using the Pandas method.isna () which denotes Rows containing missing values and then drops to enhance the quality of the data.

The folders were placed on a Pandas DataFrame. The DataFrame had two features, the first one was the 'Filepath' which was the relative path to the image file and the second one was the 'Label', which was the diagnostic class obtained from the folder name. the table below shows the file path of retrieving image files.

Filepath	Label
./OriginalDataset/Non-Demented/image1.jpg	Non-Demented
./OriginalDataset/Very Mild Demented/image1.jpg	Very Mild Demented
./OriginalDataset/Mild Demented/image1.jpg	Mild Demented
./OriginalDataset/Moderate Demented/image1.jpg	Moderate Demented

Table 3: A table showing DataFrame

3.4 Exploratory Data Analysis

To be more familiar with the dataset, several checks were conducted to assess the number of classes, sizes of images and how it would affect the performance of the models after preprocessing.

3.4.1 Class Distribution

This distribution entails that the proportion of the "Moderate Demented" class in the dataset is extreme, whereby the class only accounts for 3.1%. A bar chart visualizing the class distribution is shown below:

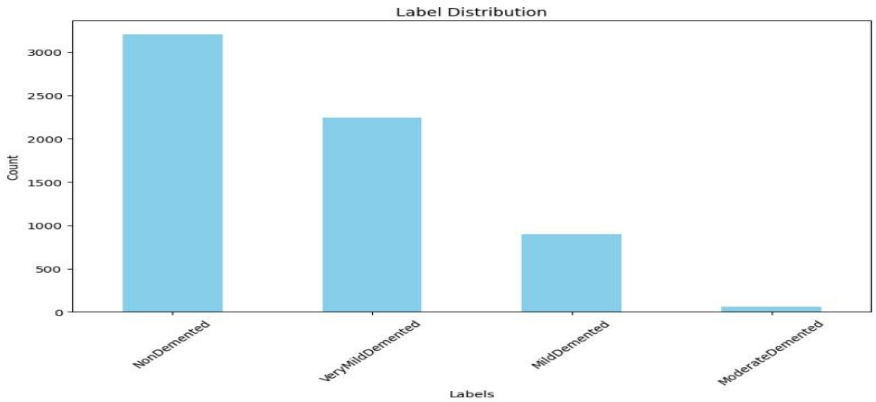

Figure 3: Graph showing distribution of classes

Another challenging group is the “Moderate Demented” group, which includes extremely few cases, so strategies like using a weighted loss function or data augmentation during the training phase must be applied to enhance the recognition rate.

3.4.2 Random Sample Visualization

To ensure that the chosen images are correctly labeled, the visualization of selected random images and their labels were shown as presented in figure 4 below. A visual examination of the content of the images also validated their classification into the respective diagnostic codes. As described in the following examples, some pictures outlined as “non-demented” possessed no indicative structure aberration whereas the Images outlined as “Moderate Demented” depicted visible cortical atrophy and increase in ventricle size, which are symptoms of severe dementia. The figure below illustrated different classes of dementia among different patients.

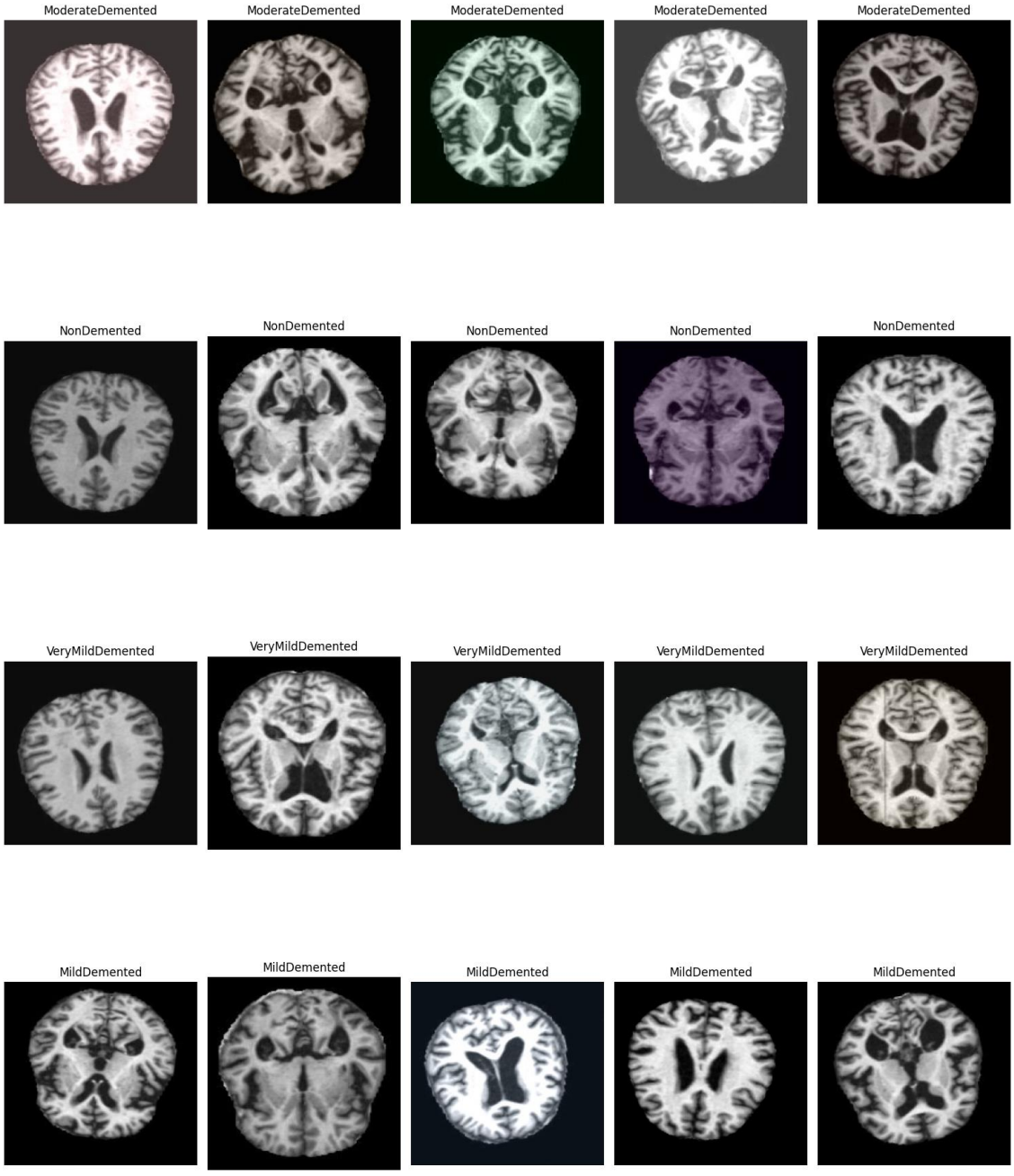

Figure 4: Images showing different categories of Alzheimer progression

3.4.3 Image Size Analysis

The original dimensions of the MRI images were different from each other; the widths, and heights were within the range of 176 to 256 pixels. Due to this variability, all the images were then resized to the standard size of 224 x 224 pixels that was required by most of the pre-trained models such as Dense

3.5 Feature Engineering

This section involves the steps done to prepare the data for modelling and training

3.5.1 Class Label Encoding:

Labels were transformed from string format like "non-demented", into integer format using Label Encoder for compatibility with machine learning models. The mapping between the string labels and their numeric representations was stored for future reference as follows; Class Mapping: Mild Demented': 0, Moderate Demented': 1, 'Non Demented': 2, 'Very Mild Demented': 3.

3.5.2 Resizing:

All images were resized to (224, 224) pixels to match the input size requirements of pre-trained models (DenseNet201, EfficientNetB2, and Vision Transformer).

3.5.3 Data Augmentation

TensorFlow's ImageDataGenerator, a versatile tool for on-the-fly augmentation, was implemented to prepare the data for processing. The ImageDataGenerator was configured to include
Random rotation within the flip range of (-10, +10) degrees to mimic the variation in positioning of the patients during the scanning process h and v flip to add diversity in the data to the model. Random scaling of 0.9 to 1.1 was applied to mimic different viewpoints, and changes in brightness with factor of 0.8 to 1.2 were used to simulate various lighting conditions during the image acquisition. This provided a means of creating new augmented data during the model training while avoiding the need to store them, which contributed to the suitability of the model for large-scale medical imaging dataset.

Augmentation was performed ONLY on the training set to avoid information leakage from validation/testing sets

3.5.4 Data Splitting:

The dataset was divided into training, validation, and testing sets, Training: Training: 70%, validation: 15%, testing: 15% as shown in the figure below. This means that the data was partitioned in a way that retains the class distribution.

Training samples:	4480
Validation samples:	960
Test samples:	960

Table 4: Table showing split data

4 Design Specification

The design of this hybrid model involves sourcing a Kaggle MRI dataset and preprocessing it for classification of Alzheimer's classes by extracting features from pretrained base models Efficientnet, Densenet201 and Vision Transformers and concatenating them then passing them through a fully connected dense layer with ReLu activators with a 50% dropout rate and Batch Normalization.

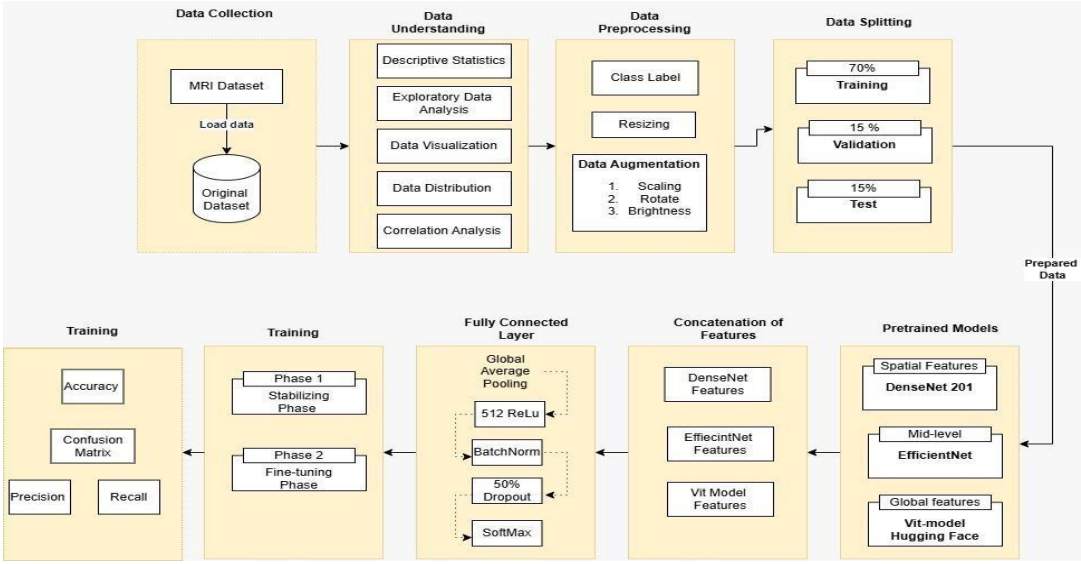

Figure 5: Model Design Framework

4.1 Modelling Technique

To utilize the base models features on MRI classification, we first have to import Imagenet pretrained models of efficientNet, Densenet and ViT . The models were then finetunes as follows:

4.1.1 DenseNet201

Originally trained on DenseNet201, the weights were transferred and further optimized on the MRI dataset for classification. Transfer learning was a process of fixing the initial layers to retain common image characteristics and fine-tuning the later layers that were adjusted to operate with MRI images in the medical field. The output of the model was post-processed by incorporating the global average pooling (GAP) layer that subsampled the feature maps and reduced them into feature vector which preserved the most discriminative spatial features. This vector was then passed to the hybrid model for its computation and final prediction results. The figure below shows the architecture of Densenet

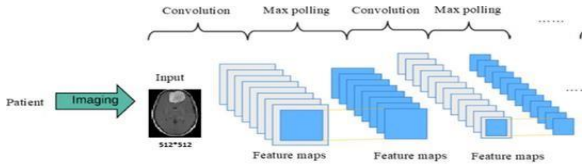

Figure 6: Figure shows DenseNet201 modelling architecture

4.1.2 EfficientNetB2

EfficientNetB2 is the mid-level feature extractor that aims at balancing the computational cost and accuracy of the model. It has a compound scaling technique that suggests the optimum scaling of depth, width, and the resolution of the network to reduce computational and memory costs. In detail, EfficientNetB2 model is most appropriate for detecting mid-level features, which includes the layout of the areas in the brain, in conjunction with DenseNet201 for detailed features or high-resolution images. These layers were reapplied on the MRI dataset to tune the model to the geometrical properties of the images. Finally, EfficientNetB2 was passed through a GAP layer to generate a vector which contains mid-level Structural Patterns. In the figure below, EfficientNetB2 modelling architecture has been illusrated.

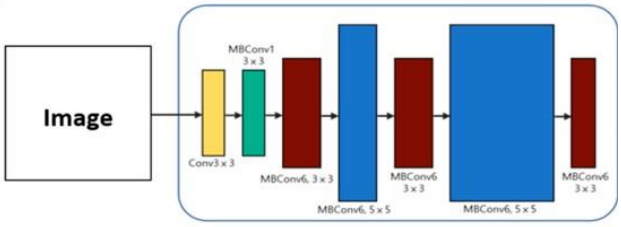

Figure 7: Figure shows EffiientNetB2 architecture

4.1.3 Vision Transformer (ViT)

To capture this global information, the Vision Transformer (ViT) was incorporated as it is capable of modeling long-range dependencies through self-attention operations. Unlike traditional convolutional neural networks that use local receptive fields, ViT trains images as sequences of patches. Each patch is considered as a token like how individual words are considered in natural language processing and allows ViT to understand the relationships and dependencies that occur throughout the brain image. For fine-tuning, the ViT model was modified to allow for MRI images to be preprocessed into patches. Every patch was flattened, and positional encodings were appended to maintain spatial information. For the ViT model, the weights pretrained on ImageNet were retrained on the MRI dataset, with the self-attention layers and the classification head retrained to capture features specific to neurodegenerative conditions. Finally, the CLS token, which encapsulates the aggregate of all the patches' features, was extracted and converted into the feature vector to be incorporated into the hybrid model. Vision Transformer (ViT) image classification architecture has been illustrated below.

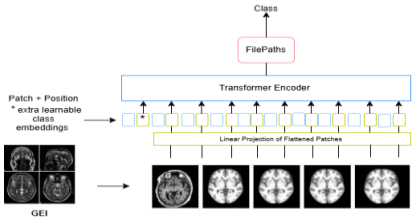

Figure 8: Figure shows ViT modelling architecture

4.1.4 Hybrid Model Architecture

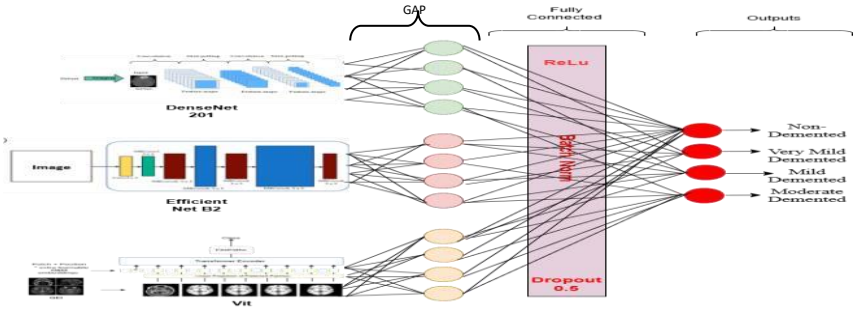

Figure 9: Figure shows Hybrid modelling architecture

In the figure above, the model begins with three separate input layers tailored to each architecture: one input layer for DenseNet201 and EfficientNetB2 that takes an image of 224 x 224 pixels and another input layer for ViT that takes feature vector of size (768,) (768,) These inputs make sure that the resulting images are compatible with the pre-trained architecture without distorting the images. In feature extraction, DenseNet201 and EfficientNetB2 architecture employs pre-trained weights to extract features from the input images in a hierarchical manner. To extract the compact feature vectors that focus on capturing the essential spatial and structural information, they add the Global Average Pooling (GAP) layers in DenseNet201 and EfficientNetB2. The ViT input that directly encodes each patch and global relationships relates to the feature vectors generated by DenseNet201 and EfficientNetB2 into a final merged feature vector. This concatenation helps to integrate the local features obtained from Densenet201, mid-level feature from EfficientNet, and global features from ViT into a single vector for further processing.

These features are then concatenated and feed into a fully connected layer of 512 units with ReLU activation to introduce non-linearity and capture higher level feature interactions.

To further improve the model performance and generalize the model, the Batch Normalization layer is used after the dense layer to normalize the features. Next is Dropout layer with 50% drop out of 50% where during training it shuts down 50% of the neurons for a time during training so as to prevent overreliance on these neurons within the model.

Finally, the output layer, which is also a dense layer with the softmax activation function, provides the probability of each of the target classes for the proper classification of MRI scans into one of the chosen diagnostic categories.

5 Implementation

5.1 Tools Used

The model was trained using TensorFlow for DenseNet201 and EfficientNetB2 networks and PyTorch for Vision Transformer. The data pre-processing used Python packages namely NumPy, Pandas and Scikit-learn for analysis, and Matplotlib and Seaborn for visualization. The weights of DenseNet201 and EfficientNetB2 models were initialized with weights pre-trained on ImageNet,

while the ViT model was implemented with the help of the Hugging Face Transformer's library. Training was done on Google Colab and GPU acceleration was used with Tesla T4 GPUs for enhanced computations.

5.2 Hyperparameter Tuning

The benefits of hyperparameter tuning in model enhancement cannot be understated. Basing on external validity issue, Schratz et al. (2019) underlined the role of parameter tuning as the method to prevent classification biases.

A few parameters tuned with to enhance the working of models are shown in the figure below:

Model	Hyperparameters	Optimal Value
DenseNet201	learning_rate	1.00E-04
	batch_size	32
	epochs	20
	dropout_rate	0.3
	optimizer	Adam
EfficientNetB2	learning_rate	1.00E-04
	batch_size	32
	epochs	20
	weight_decay	0.01
	optimizer	Adam
Vision Transformer	learning_rate	1.00E-05
	batch_size	16
	epochs	30
	hidden_dim	256
	num_heads	8
	dropout_rate	0.2
	optimizer	AdamW
Hybrid Meta-Model	meta_learning_rate	1.00E-03
	meta_hidden_layer_sizes	(128, 64)
	meta_epochs	20
	activation_function	relu
	loss_function	categorical_crossentropy

Figure 5: Table showing tuned hyperparameters

5.3 Model Training

5.3.1 Training Strategy 1

The chosen hybrid deep learning model was tested and trained with TensorFlow / Keras's fit () method on train_multi_gen and val_multi_gen to create training and validation data in bits. The training was done for 20 epochs, in each epoch, a specific number of batches for training (steps_per_epoch) and validation (validation_steps) were used. This was done with the purpose of monitoring the generalization capability of the model during the validation after every epoch. Another two important callbacks known as early_stopping and annealer describe other procedures aimed at enhancing the training process. Firstly, early stopping was employed to halt training if the model's accuracy on the validation data set fails to improve beyond a particular epoch, thus preventing overtraining and conserving computational resources. The learning rate scheduler or annealer was used when the validation performance became stuck which helped in lowering the learning rate and fine-tuning the weights for convergence.

5.3.2 Training Strategy 2

To fine-tune the hybrid deep learning model for the MRI classification, a two-phase training scheme was designed. This approach is built upon the DenseNet201 and EfficientNetB2 architectures, with proper domain adaptation that is critical for the identification of neurodegenerative diseases. The training process included the training stabilization phase and the fine-tuning process.

Phase 1 (Stabilization Training): In order to preserve the features learned from ImageNet, the weights of the pre-trained DenseNet201 and EfficientNetB2 layers were frozen. This ensured that while the dense layers were trained on the MRI-specific data, the other layers retained the pre-trained knowledge. For exact weight updates, the Adam optimizer with an initial learning rate of 1×10^{-5} was used. reduceonplateu ensured that the learning rate was reduced when the validation loss reached a plateau, early stopping on the other hand halted training at epoch 5 without validation improvement. This may indicate that the model poorly predicted the minority classes and could not fully utilize cross-domain features adequately. Class weights were computed to address class imbalance, ensuring minority classes like "Moderate Demented" were not underrepresented. This phase stabilized the model's learning without disrupting pre-trained features.

Phase 2 (Fine-Tuning): The frozen layers of DenseNet201 and EfficientNetB2 were unfrozen to allow fine-tuning, ensuring the feature extraction layers adapted to the nuances of MRI data. A reduced learning rate of 1×10^{-6} was used to minimize weight updates and prevent instability. Six additional epochs were conducted, with the same loss function categorical_crossentropy and class weights applied. This phase fine-tuned both the feature extraction layers and the fully connected layers to maximize classification accuracy and enhance generalization.

5.4 Evaluation Metrics

Several evaluation metrics were used to evaluate the model's performance and ensure robustness in prediction. The relevant F1 scores, accuracy, and precision for the measurement model performance are calculated and have been presented in the respective implementation sections of each model.

Metric	Description	Purpose	Formula
Accuracy	Measures the proportion of correctly classified samples out of the total samples.	Provides a general measure of overall model performance.	$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$
Precision	Evaluates the proportion of correctly predicted positive samples out of all positive predictions.	Ensures fewer false positives, especially important for dominant classes.	TP / (TP + FP)
Recall	Measures the proportion of correctly predicted positive samples out of all actual positive samples.	Ensures fewer false negatives, critical for underrepresented classes.	TP / (TP + FN)
F1-Score	Harmonic mean of precision and recall.	Balances precision and recall, particularly important in the presence of class imbalances.	2 * (Precision * Recall) / (Precision + Recall)
Confusion Matrix	Tabular representation comparing predicted and actual classes.	Provides a detailed view of classification performance, highlighting errors across	N/A

Table showing Evaluation metrics used

6 Findings

6.1 Training Strategy 1 Results

6.1.1 Accuracy and Loss Metrics

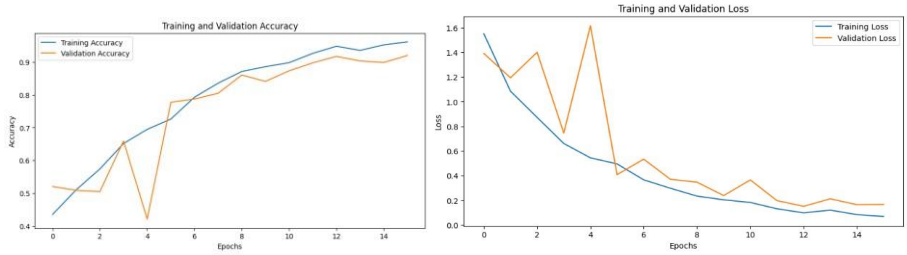

Figure 10: Line graphs showing validation accuracy and loss

Accuracy and Loss Metrics: The accuracy and loss curves for Training Strategy 1 demonstrate steady improvement over the epoch as shown in figure 13 above, with the training accuracy reaching approximately 80% and the validation accuracy peaking around 75%. However, the noticeable gap between the training and validation accuracy curves suggests that the model struggled to generalize effectively, indicative of overfitting. The training and validation loss curves follow a similar pattern, with a decline over epochs, but the disparity between the two highlights a lack of robustness in the model's performance on unseen data.

6.1.2 Classification Metrics

Class	Precision	Recall	F1-Score
MildDemented	0.12	0.13	0.13
ModerateDemented	0	0	0
NonDemented	0.52	0.48	0.5
VeryMildDemented	0.37	0.4	0.38
Accuracy			0.4
Macro Avg	0.25	0.25	0.25
Weighted Avg	0.41	0.4	0.4

Figure 11: Table showing classification metrics

The classification metrics as shown in the table above further reveal the limitations of Training Strategy 1. While the "NonDemented" class shows moderate performance with a precision of 52% and recall of 48%, the other classes, particularly "ModerateDemented," exhibit near-zero precision, recall, and F1-score. As explained by the results of Training Strategy 1, the model fails to perform well on the validation set and experiences issues with class imbalance. The low performance for minority classes implies that there is a need for a better framework when training the model. These limitations are eliminated in Strategy 2 with features such as the fine-tuning of pre-trained layers, gradual modulation of learning rates and better incorporation of specific domains. This sort of finetuning is particularly important to enhance the generalizability and the performance on the lessrepresented classes.

6.2 Training Strategy 2 Results

6.2.1 Accuracy and Loss Metrics

In the figure below the training and validation accuracy curves show a continuous improvement during Strategy 2 up to the training accuracy of 98% and validation accuracy of around 95%. These results illustrate the effectiveness of the learning process of the model, which does not lead to overfitting or excessive adaptation to training data. The loss curves similarly show consistent decreases, with validation loss converging to a stable value. The smooth alignment between training and validation loss underscores the hybrid model's generalization to unseen data, validating the efficacy of fine-tuning pre-trained layers in Strategy 2.

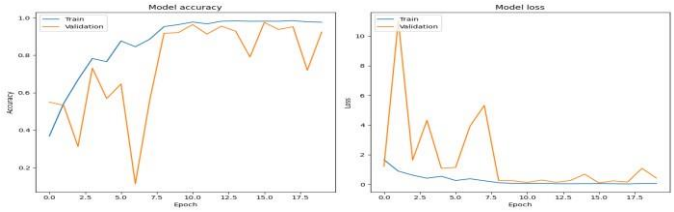

Figure 12: Line graphs showing validation accuracy and loss

6.2.2 Classification Metrics

A detailed classification report (Figure 16) highlights the model's precision, recall, and F1-score for each class:

Class	Precision	Recall	F1-Score
MildDemented	0.91	0.96	0.93
ModerateDemented	0.86	0.97	0.91
NonDemented	0.97	0.96	0.96
VeryMildDemented	0.95	0.94	0.94
Accuracy			0.95
Macro Avg	0.92	0.96	0.94
Weighted Avg	0.95	0.95	0.95

Figure 13: Table showing classification metrics

The Mild Demented class attained an accuracy of 91 percent and a recall of 96 percent, implying that the model was accurate in identifying such cases. Moderate Demented is the class with the least number of samples, but by scoring 97 percent recall, the model guarantees identification of these critical instances. However, the precision of 86% highlights potential misclassifications into other categories. Non-Demented With the largest support (3200 samples), the model maintained a high precision (97%) and recall (96%), reflecting its strength in classifying dominant classes.Very Mild Demented: The values of precision and recall for this category were 0.95 and 0.94, respectively, meaning the results are reliable and exhaustive.The validation of accuracy was done using the weighted average F1-score since the classes were imbalanced; the score achieved was 95% of the total.The low precision for the 'Moderate Demented' class may be driven by class imbalance in the given dataset. Since this class has relatively fewer samples compared to other classes, the model leans more towards the prevailing classes. To tackle this, methods like margin thresholds or using augmented data to train models on the rare class could assist in achieving reliability and fairness in predictions.

6.2.3 Confusion Matrix

The confusion matrix (Figure 16) provides a granular view of the model's classification performance:

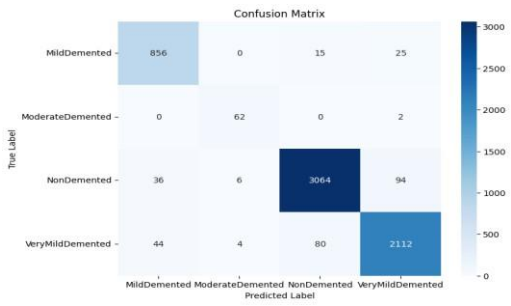

Figure 14: Figure showing Confusion Matrix

Meaningfully, the confusion matrix reveals tendencies that the constructed model may have when classifying particular samples. For instance, the model was nearly perfect for the Mild Demented class

where it classified 856 out of 896 and few overlapped with other classes. There was also high classification accuracy in the Very Mild Demented and Non-Demented classes as indicated in the confusion matrix where the diagonal elements are much higher for those classes. However, what is significant to note is that the Moderate Demented class, which is the smallest class and hence has few sample sizes, had slight changes that misclassified some samples into the Very Mild Demented class, which points to the reality that small samples often present classification difficulties.

6.3 Analysis of Results

The superior performance of Training Strategy 2 compared to Training Strategy 1 is attributed to its multi-phase approach, which allowed for a more tailored adaptation of the pre-trained DenseNet201 and EfficientNetB2 layers to the MRI dataset. In Strategy 2, the pre-trained weights were initially frozen during the stabilization phase to preserve generalizable features from the ImageNet dataset. Following this, the weights were unfrozen, and the hybrid model underwent fine-tuning with a reduced learning rate of 1×10−61 \times 10^{-6}1×10−6. This step enabled the model to refine its feature representations, capturing subtle patterns specific to neurodegenerative diseases. By allowing task-specific adjustments, Strategy 2 enhanced the model's ability to distinguish between similar diagnostic categories, such as "Very Mild Demented" and "Mild Demented," which Strategy 1 struggled with due to its static pre-trained layers. The features extracted from ViT with the local and mid-level features from DenseNet201 and EfficientNetB2 that were extracted increased Strategy 2's learning rate in unseen data. The incorporation of class weights during both phases mitigated the impact of class imbalance, particularly for underrepresented categories like "Moderate Demented." In contrast, Strategy 1, which retained frozen pre-trained layers throughout training, lacked the flexibility to adapt to domain-specific nuances. This limitation led to suboptimal integration of global and local features and a reduced ability to generalize effectively to validation and test datasets. By leveraging fine-tuning and adaptive learning, Strategy 2 proved more robust and better aligned with the complexities of MRI data classification. Table below summarizes the analysis

	Training Strategy 1	**Training Strategy 2**
Pre-trained Layer Usage	Layers remained frozen throughout training. Retained generalizable features from ImageNet but limited domain adaptation.	Layers were initially frozen, then unfrozen for finetuning, allowing adaptation to domain-specific MRI features.
Learning Rate	Fixed learning rate is used for the entire training process. 1×10−5	Initial learning rate of 1×10−51 \times 10^{5}1×10−5, reduced to 1×10−61 \times 10^{6}1×10−6 during fine-tuning for precise updates.
Fine-tuning	No fine-tuning performed; dense layers were trained on their own.	Fine-tuning involved jointly training pre-trained layers and dense layers, optimizing for neurodegenerative disease classification.
Class Imbalance Handling	Used class weights but lacked a stabilization phase to mitigate the influence of dominant classes.	Class weights combined with a stabilization phase and fine-tuning improved classification of underrepresented classes like "Moderate

		Demented."
Generalization	Risk of overfitting dense layers; limited generalization to unseen validation and test data.	Improved generalization due to incremental updates and harmonization of global and local feature extraction.
Feature Integration	Limited feature integration as frozen layers did not adapt to hybrid model requirements.	Seamless integration of global (ViT) and local (DenseNet201, EfficientNetB2) features through fine-tuning.
Performance on Minority Classes	struggled to classify underrepresented classes effectively.	Higher recall for "Moderate Demented" due to better optimization of classification boundaries.

Table 15: Table summarizing the two training strategies

7 Conclusion and Future Work

7.1 Discussion

This study aimed to address the research question: To what extent can classification performance on MRI scans be improved through a hybrid deep learning model combining DenseNet201, EfficientNetB2, and Vision Transformer (ViT) features? The primary objective was to overcome the limitations of existing models, including low classification accuracy, , and challenges posed by class imbalances, by developing a hybrid model that integrates local, mid-level, and global features. The proposed hybrid model met all of these objectives through the utilization of DenseNet201 for the detailed spatial information, EfficientNetB2 for mid-level structural components, and ViT for the global context comprehension. It was found that the model had an accuracy of 95% with MRI scans, which is highly efficient with classes that are rarely represented in models, such as the 'Moderate Demented' class, with a recall of 97%. This corroborates the used model's application and capacity to categorize neurodegenerative disease phases precisely, making it an efficient solution to the study's issue. Nevertheless, there are several drawbacks that were reported in the study; the usage of pretrained weights from the ImageNet database and high computational complexity can be regarded as potential limitations.

7.2 Key Findings

The hybrid model used in this research provided near-optimal performance as it yielded approximate 95% accuracy in staging the neurodegenerative diseases MRI scans. This performance indicates how the model is able to utilize secondary features from DenseNet201, EfficientNetB2, and the Vision Transformer (ViT). Some of the primary issues, like overfitting to the big classes, involved the use of weighted loss functions and data augmentation to help boost the recall of the small classes. The integration of these architectures was quite effective, because each of them brought in its own strengths that is detailed, mid-level, and global features for classification. However, one crucial drawback of the model is its computational demand which can be a challenge where resources are scarce. The model performs with very high accuracy and sufficiently captures features but could be a limitation due to the resource requirements, thus pointing to the explicit trade-off between accuracy and scalability.

7.3 Implications of Research

The implications of the findings have profound impact on medical field. The hybrid model presented above could be a valuable asset in automating the classification of neurodegenerative diseases. In this case, its performance in underrepresented classes proves that it is able to pick important cases that might not otherwise be considered, which aids in improving diagnostic precision and ultimately patient outcomes. However, its application in clinical contexts necessitates dealing with the model's computational complexity.

7.4 Limitations

While the study demonstrates promising results, several limitations must be addressed:

1. Fine-tuning with the ImageNet pre-trained weights restricted the learning of MRI scan domain-specific characteristics in the model.
2. Expensive and less deployable architectures such as DenseNet201, EfficientNetB2, and ViT that form the key components of the hybrid model.
3. Although the collected dataset was large, all the videos were retrieved from a single database. The inclusion of subjects from different age groups, ethnic origin, and a wider range of imaging phenotypes would enhance generalizability.

7.5 Future Work

To address these limitations, several meaningful avenues for future research are proposed: The future research on the hybrid deep learning models in medical imaging could focus on the pretrained models suitable for the medical imaging datasets like ADNI or OASIS. These datasets preserve additional features specific to MRI scans, which helps the models to learn the fine details characteristic of neurodegenerative diseases. Thus, such models might substantially enhance accuracy and fine-grained feature learning in accordance with more specific datasets than ImageNet. For cases where computation is a concern, it is recommended to incorporate lighter frameworks such as MobileNet or use other methods like pruning or quantizing models. These approaches ensure that the models remain very accurate while at the same time require fewer resources, thus making them applicable in low-resource scenarios such as rural clinics or small-scale hospitals.

Another potential future direction is the improvement of the interpretability of the hybrid model. Integration of explainability approaches such as Grad-CAM or SHAP can further help clinicians decipher the features behind the model's decisions, making the system highly interpretable. Such transparency helps to build confidence and pave the way towards practical implementation of clinical AI diagnostic systems.

Secondly, using various samples from diverse patients and ensuring that the model trained on patients who vary in age, sex, or other parameters would enhance its robustness and fairness. If the dataset is made to mimic natural variability, the model can perform reasonably well across any population and in various imaging environments.

7.6 Conclusion

Hence, this study demonstrates that hybrid deep learning models have the potential to revolutionize medical imaging due to the numerous benefits realized. Nevertheless, this optimistic sign leads to some limitations that call for the enhancement and the integration of these progresses for better application. Subsequent studies can build up based on this study and utilize domain-specific data, tune architectures to improve performance, and to accelerate early diagnosis of neurodegenerative diseases.

8 References

The Lancet Public Health. (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*

Baker, D., Chen, W.-B., & Gao, H. (2024). Early Alzheimer's detection: The promise of AI-powered MRI analysis.

Kulasinghe Wasalamuni Dewage, K. A., Hasan, R., Rehman, B., and Mahmood, S. (2024) 'Enhancing brain tumor detection through custom convolutional neural networks and interpretability-driven analysis', *Information*.

Bi, W., Xv, J., Song, M., Hao, X., Gao, D., & Qi, F. (2023). Linear fine-tuning: A linear transformation-based transfer strategy for deep MRI reconstruction. *Frontiers in Neuroscience*.

Abioye, O. A., Thomas, S., Odimba, C. R., & Olalekan, A. J. (2023). Generic hybrid model for breast cancer mammography image classification using EfficientNetB2. *Dutse Journal of Pure and Applied Sciences*, *9*(3b), 281-289.

Awang, M. K., Rashid, J., Ali, G., Hamid, M., Mahmoud, S. F., Saleh, D. I., & Ahmad, H. I. (2024). Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique. *Plos one*, *19*(9), e0304995.

Babu Vimala, 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), 23029.

Banerjee, S., & Monir, M. K. H. (2023, July). CEIMVEN: An Approach of Cutting Edge Implementation of Modified Versions of EfficientNet (V1-V2) Architecture for Breast Cancer Detection and Classification from Ultrasound Images. In *International Conference on Computing, Intelligence and Data Analytics* (pp. 310-323). Cham: Springer Nature Switzerland.

Benhassine, N. E., Boukaache, A., & Boudjehem, D. (2024, August). Breast cancer image classification using DenseNet201 and AlexNet based deep transfer learning. In *the International Conference on Emerging Intelligent Systems for Sustainable Development (ICEIS 2024)* (pp. 129-143). Atlantis Press.

Çetiner, H., & Çetiner, İ. (2022). Classification of cataract disease with a DenseNet201 based deep learning model. *Journal of the Institute of Science and Technology*, *12*(3), 1264-1276.

Chen, J., He, Y., Frey, E. C., Li, Y., & Du, Y. (2021). Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv preprint arXiv:2104.06468*.

Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., & Qiao, Y. (2022). Vision transformer adapter for dense predictions. *arXiv preprint arXiv:2205.08534*.

Hastomo, W., Karno, A. S. B., Sestri, E., Terisia, V., Yusuf, D., Arman, S. A., & Arif, D. (2024). Classification of Brain Image Tumor using EfficientNet B1-B2 Deep Learning. *Semesta Teknika*, *27*(1), 46-54.

Hindarto, D. (2023). Model Accuracy Analysis: Comparing Weed Detection in Soybean Crops with EfficientNet-B0, B1, and B2. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *7*(4), 734-744.

Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., & Kaur, M. (2021). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, *39*(15), 5682-5689.

Li, J., Xia, X., Li, W., Li, H., Wang, X., Xiao, X., ... & Pan, X. (2022). Next-vit: Next generation vision transformer for efficient deployment in realistic industrial scenarios. *arXiv preprint arXiv:2207.05501*.

Lu, T., Han, B., Chen, L., Yu, F., & Xue, C. (2021). A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Scientific Reports*, *11*(1), 15824.

Pacal, İ. (2022). Deep learning approaches for classification of breast cancer in ultrasound (US) images. Journal of the Institute of Science and Technology, 12(4), 1917-1927.

Petrini, D. G., Shimizu, C., Roela, R. A., Valente, G. V., Folgueira, M. A. A. K., & Kim, H. Y. (2022). Breast cancer diagnosis in two-view mammography using end-to-end trained efficientnet-based convolutional network. *Ieee access*, *10*, 77723-77731.

Preetha, R., Priyadarsini, M. J. P., & Nisha, J. S. (2024). Automated Brain Tumor Detection from Magnetic Resonance Images Using Fine-Tuned EfficientNet-B4 Convolutional Neural Network. *IEEE Access*.

Salim, F., Saeed, F., Basurra, S., Qasem, S. N., & Al-Hadhrami, T. (2023). DenseNet-201 and Xception pre-trained deep learning models for fruit recognition. *Electronics*, *12*(14), 3132.

Sanghvi, H. A., Patel, R. H., Agarwal, A., Gupta, S., Sawhney, V., & Pandya, A. S. (2023). A deep learning approach for classification of COVID and pneumonia using DenseNet-201. *International Journal of Imaging Systems and Technology*, *33*(1), 18-38.

Tadepalli, Y., Kollati, M., Kuraparthi, S., & Kora, P. (2021). EfficientNet-B0 Based Monocular Dense-Depth Map Estimation. *Traitement du Signal*, *38*(5).

Xia, C., Wang, X., Lv, F., Hao, X., & Shi, Y. (2024). Vit-comer: Vision transformer with convolutional multi-scale feature interaction for dense predictions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5493-5502).

Xu, R., Xiang, H., Tu, Z., Xia, X., Yang, M. H., & Ma, J. (2022, October). V2x-vit: Vehicle-toeverything cooperative perception with vision transformer. In *European conference on computer vision* (pp. 107-124). Cham: Springer Nature Switzerland.

Yang, L., Yu, H., Cheng, Y., Mei, S., Duan, Y., Li, D., & Chen, Y. (2021). A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture. *Computers and Electronics in Agriculture*, *187*, 106316.

Yin, H., Vahdat, A., Alvarez, J. M., Mallya, A., Kautz, J., & Molchanov, P. (2022). A-vit: Adaptive tokens for efficient vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10809-10818).

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. and Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, pp.109–120.

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z. H., ... & Yan, S. (2021). Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 558-567).

Zhou, J., Gu, X., Gong, H., Yang, X., Sun, Q., Guo, L., & Pan, Y. (2024). Intelligent classification of maize straw types from UAV remote sensing images using DenseNet201 deep transfer learning algorithm. *Ecological Indicators*, *166*, 112331.