JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1202

Multi-Class Classification of Alzheimers Impairment and Dementia Stages using an Efficientnet-B0 Deep Learning Framework

Makki Riaz Khan[1], Muhammad Masood Ul Rahman Usmani[2], Muzamil Mehboob[3], Faheem Abbas[4], Shahnaz Rafique[5], and Kishwar Bibi[6]

[1]Bahauddin Zakariya University Multan, Lodhran, 59320, Pakistan.
*Corresponding Author: Muzamil Mehboob. Email: muzamilmehboob@bzu.edu.pk

Received: November 28, 2025 **Accepted:** March 26, 2026

Abstract: Alzheimer's disease is a progressive neurodegenerative condition that leads to deterioration of memory and cognitive function, where early detection is critical as therapeutic interventions are most effective prior to extensive neuronal loss. Many existing deep learning approaches focus on binary or ternary classification schemes, which inadequately reflect the full clinical spectrum of cognitive impairment and dementia, thereby limiting their practical diagnostic relevance.This study proposes a lightweight deep learning framework based on EfficientNet-B0 for fine-grained eight-class Alzheimer's staging, encompassing four levels of cognitive impairment and four levels of dementia. A structured preprocessing pipeline, including normalization, contrast enhancement, resizing, and targeted data augmentation, is employed to improve MRI consistency and enhance discriminative feature learning.Under a strict patient-level evaluation protocol, the proposed model achieves a peak testing accuracy of 99.78% and demonstrates strong and stable performance when compared with commonly used 2D convolutional neural network baselines. Owing to its low computational complexity and consistent performance across classes, the framework represents a promising research-stage tool for automated Alzheimer's disease staging and warrants further validation, including comparison with transformer-based and volumetric approaches, before clinical deployment.

Keywords: Alzheimer's Disease; MRI Imaging; Deep Learning; EfficientNet B0; Multi Class Classification; Disease Staging; Computer Aided Diagnosis

1. Introduction

Alzheimer's disease is a progressive, irreversible, and chronic neurodegenerative disease that impacts memory, executive processing, and daily behavioral abilities. It has been acknowledged widely across the global platform as the prime reason for elderly people suffering from dementia. As per reports from the World Health Organization (WHO), more than 55 million people are ornamenting with dementia. It will soon touch an alarmingly high number of 78 million in 2030 and 139 million people in 2050 [1]. The rising statistics clearly evince the immense emotional, social, and financial toll being imposed on people as well as facilities across the global platform. As a consequence of it, early diagnosis and proper cognitive staging have become imperative steps for postponing disease progress and enhancing the life standards of patients.

Clinically, the onset and progression of AD occur on a smooth spectrum that begins with Cognitively Normal (CN) patients and then progresses through mild cognitive impairment to Alzheimer's Dementia. Conventionally, classification and diagnosis have traditionally grouped patients into binary or ternary classifications. Nevertheless, a higher degree of granularity might be necessary while testing and diagnosing. This becomes more necessary as transition phases occur from Non-Impairment to Very Mild/Mild Impairment, wherein treatment and intervention can be made but detection itself becomes

increasingly difficult. Consequently, it becomes prima facie necessary and imperative that classification occur on a scale involving eight clinically significant classes.

Table 1. Impairment and Dementia Classes

Impairment Classes	Dementia Classes
Very Mild Impairment	Very Mild Dementia
Mild Impairment	Mild Dementia
Moderate Impairment	Moderate Dementia
Non-Impairment (Control)	**Non-Dementia**

Despite being highly relevant from a medical perspective, it is difficult to distinguish these eight groups. Both early impairment and early-stage dementia might share common anatomical markers, such as mild hippocampal atrophy, mild cortical atrophy, and early ventricular changes [2, 5]. Moreover, there might be variability among subjects and variability in slice orientations. These would add complexity and make it even harder to classify. Conventional medical procedures like neuropsychological testing and analysis of MRI images are very subjective and prone to inconsistencies among medical practitioners [3].

Magnetic Resonance Imaging (MRI) techniques have emerged as prominent tools for identifying neurodegenerative disorders because they are non-invasive and have the capability of capturing images with high resolution, reflecting changes and shifts at a cognitive level. Some prominent biomarkers for identifying AD include hippocampal atrophy, an increase in the size of sulci within the cortex, and an escalation in ventricle size [4]. However, manually identifying thousands of images within an MRI scan of a patient would be tedious and error-prone.

Lately, there have been spectacular developments and successful utilisations of deep learning techniques, specifically Convolutional Neural Networks (CNNs), regarding the extraction and detection of intricate hierarchical patterns within MRI images [6]. Although CNNs have achieved successful binary and ternary classifications within AD diagnosis, there have been negligible efforts made within multiclass classification problems [7]. The main challenges within multi-class classification problems include higher inter-class similarities, imbalanced datasets within various stages of impairment, and computational complexities due to large MRI datasets. 3D volumetric CNNs, though highly reliable and efficient, possess large GPU memory and undergo prolonged training [8]. To overcome these constraints, the compound scaling method introduced by Tan and Le in EfficientNet scales the width, depth, and image resolution jointly and optimally [9]. Among all variants, EfficientNet B0 balances computational cost and model expressiveness effectively and thus is more appropriate for medical imaging, as it requires high performance as well as speed. Previous works have shown that medical images' subtle anatomical distinctions could be modeled effectively with low-risk overfitting using EfficientNet-based models.

The strategy uses a properly engineered preprocessing pipeline, involving slice normalization, contrast enhancement, and data augmentation. Unlike larger 3D models, it works on 2D images of MRI slices, making it more scalable with no loss of accuracy. A light yet efficient model based on EfficientNet B0 is proposed for an eight-stage classification task related to Alzheimer's. Clinically meaningful labels with four stages of impairment and four stages of dementia are included within the model.A preprocessing chain involving normalization, resizing, optimizing contrast, eliminating noise, and specific augmentations was formulated with the intention of improving slice quality and enhancing generalization.The model's resilience and accuracy on all classes are upheld and justified via comprehensive quantitative tests with regards to accuracy, precision, recall, F1score, ROC AUC, and confusion matrices.The method presents high multi-class classification accuracy but with low computational complexity.Conclusion This research illustrates that with an optimized structure of EfficientNetB0 and an effective preprocessing and training pipeline, it is possible to classify and identify eight significant stages of Alzheimer's disease. It plays an extremely valuable role in developing an interpretable and deployable model for neurological diagnosis and monitoring.

2. Related Work

Deep learning has significantly advanced Alzheimer's disease (AD) diagno-sis and staging in recent years by enabling automatic learning of discrim-inative patterns from neuroimaging data, particularly magnetic resonance imaging (MRI). Recent surveys report that deep learning–based approaches outperform traditional machine learning techniques by effectively captur-ing complex structural changes associated with neurodegeneration [11, 13].

Consequently, a wide range of deep learning paradigms has been explored to balance diagnostic accuracy, computational complexity, and clinical fea-sibility, including 2D convolutional neural networks (CNNs), volumetric 3D CNNs, hybrid architectures, and lightweight efficiency-oriented models [13, 15]. This section reviews recent literature to establish the methodolog-ical context for the proposed EfficientNet-B0–based framework.

2.1. 2D CNN-Based Alzheimer's Disease Classification

Traditional 2D CNN architectures were among the earliest deep learning models applied to Alzheimer's disease classification using brain MRI. These methods typically extract discriminative features from individual axial, coronal, or sagittal slices processed independently. Recent studies demon-strate that transfer learning with pretrained CNNs such as VGG, ResNet,and DenseNet remains effective for Alzheimer's disease diagnosis, especially when training data are limited [15]. By leveraging pretrained weights, these models achieve competitive performance while maintaining relatively low computational cost.

However, several studies indicate that most 2D CNN-based approaches primarily address binary or coarse multi-class classification tasks, such as distinguishing Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal (CN) subjects, without modeling fine-grained disease progression [13]. Moreover, the slice-wise analysis strategy limits the ability of 2D CNNs to capture global volumetric context and inter-slice anatomical dependencies. Despite these limitations, 2D CNNs remain widely used as baseline models due to their simplicity, efficiency, and ease of implementation [15].

2.2. 3D and Hybrid CNN Architectures

To overcome the loss of volumetric information in 2D approaches, 3D CNN architectures have been introduced to directly analyze full MRI volumes or stacked slices. These models are capable of learning spatially coherent representations and capturing subtle morphological changes across brain regions associated with cognitive decline [18]. Hybrid and multi-view archi-tectures further enhance representation learning by integrating information from multiple anatomical planes.

Recent studies report improved performance of 3D and hybrid CNNs in multi-class Alzheimer's disease staging tasks, where inter-class differences are often subtle [14]. However, despite their improved representational capacity, 3D CNNs require substantially larger labeled datasets and impose high computational demands [13]. These limitations restrict their deploy-ment in resource-constrained clinical environments.

2.3. Lightweight Deep Learning Models

To address the computational challenges of deep architectures, lightweight CNN models have gained increasing attention in Alzheimer's disease re-search. Architectures such as MobileNet and EfficientNet employ depth-wise separable convolutions and compound scaling strategies to achieve an optimal balance between accuracy and efficiency [19].

Among these, EfficientNet has emerged as a promising model family due to its compound scaling strategy. Recent medical imaging studies demonstrate that EfficientNet variants achieve high classification perfor-mance with fewer parameters and reduced training cost, making them suit-able for clinical decision-support systems [19].

2.4. Multi-Class Alzheimer's Disease Staging

While binary classification provides a basic distinction between diseased and healthy subjects, it fails to capture the progressive nature of Alzheimer's disease. Multi-class staging frameworks offer a more clinically meaningful understanding of disease progression [16, 17]. However, fine-grained staging remains challenging due to overlapping neuroanatomical patterns and class imbalance [15].

Recent studies have explored hybrid architectures and ensemble learn-ing to improve multi-class Alzheimer's disease classification performance [14]. However, these methods are often computationally expensive and rely on limited datasets. Consequently, there remains a need for efficient and scalable frameworks such as the proposed EfficientNet-B0–based approach.

3. Positioning of the Proposed Work

Based on the findings of previous research, the current research proposes an eightclass Alzheimer's staging system using EfficientNet-B0. It bridges the gap between computational complexity and fine classification. The system identifies: Four stages of impairment: Non-Impairment, Very Mild Impairment, Mild Impairment, and Moderate Impairment .Four stages of Dementia: Very Mild Dementia, Mild Dementia, Moderate Dementia, and Non-Dementia.

Unlike massive 3D models, it relies on 2D MRI slices and benefits from EfficientNetB0's excellent feature extraction abilities for light inference and fast training. Thorough experiments have shown that the introduced framework boasts an accuracy rate of 99.78%, exceeding that of traditional CNN models and serving as a reference solution for highly detailed multi-class AD staging. In summary, there is value added here as a solution that resolves issues with previously suggested methods and has a very strong potential for implementation.

3.1. Dataset Discription

This study utilizes a publicly available Alzheimer's Types, Impairment and Dementia Dataset obtained from Kaggle [10]. The dataset consists of 78,269 preprocessed T1-weighted structural MRI brain slice images categorized into eight clinically meaningful classes representing different stages of cognitive impairment and dementia. These classes include Non-Impairment, Very Mild Impairment, Mild Impairment, Moderate Impairment, Very Mild Dementia, Mild Dementia, Moderate Dementia, and Non-Dementia.

The dataset exhibits a pronounced class imbalance, with the Non-Dementia class containing 53,777 images, while the Moderate Dementia class comprises only 390 images. The complete distribution of images across all eight classes is summarized in Table 3. All MRI slices are two-dimensional, resized uniformly to 224 × 224 pixels, normalized for intensity consistency, and derived exclusively from T1-weighted structural MRI scans.

Table 2. Summary of Existing Works on Alzheimer's Classification

Model	Type	Strengths	Limitations
VGG16	2D CNN	Simple architecture; effective with transfer learning; good baseline performance for AD vs. CN classification	High parameter count; limited ability to capture fine-grained disease stages
ResNet50	2D CNN	Residual connections enable deeper networks; improved feature learning; stable training	Focuses mainly on binary or coarse multi-class tasks; ignores volumetric context
DenseNet201	2D CNN	Dense feature reuse; improved gradient flow; strong performance on small datasets	Computationally expensive; limited effectiveness for detailed multi-class staging
3D CNNs	3D Volumetric Models	Capture spatial and volumetric brain features; better representation of disease progression	Require large labeled datasets; high memory and computational cost
Hybrid 3D CNN Models	Multi-view 3D	Integrate axial, coronal, and sagittal views; improved multi-class classification	Increased architectural complexity; limited clinical deployability

3D-MobiBrainNet	Multi-view 3D CNN	High accuracy using multi-view MRI fusion; improved staging capability	Heavy model; slow inference; unsuitable for real-time or low-resource deployment
MobileNet	Lightweight 2D CNN	Fast inference; low parameter count; suitable for deployment	Reduced discriminative power for complex AD staging
ShuffleNet	Lightweight 2D CNN	Computationally efficient; optimized for resource-constrained systems	Performance degradation in fine-grained multi-class classification
EfficientNet (Variants)	Efficient 2D CNN	Strong feature extraction; compound scaling improves accuracy-efficiency trade-off	Limited exploration for detailed multi-stage Alzheimer's disease classification

Table 3. Number of images per class in the dataset.

Class	Image Count
Non-Impairment	2,560
Very Mild Impairment	2,406
Mild Impairment	2,191
Moderate Impairment	2,057
Very Mild Dementia	10,890
Mild Dementia	4,001
Moderate Dementia	390
Non-Dementia	53,777
Total	**78,269**

3.2. Patient-Level Data Splitting and Leakage Prevention

To ensure an objective and clinically relevant evaluation, all MRI slices belonging to the same patient were kept within a single data split. The dataset was divided at the patient level into 80% training and validation and 10% testing,and 10% validation ,depending on the final experimental configuration. This strict patient-wise partitioning prevents data leakage and ensures that no MRI scans from the same subject appear across multiple subsets. Maintaining patient-level separation enables a realistic assessment scenario, closely mirroring a clinical setting in which the model would be tested on entirely new patients.

3.3. Class Imbalance Handling (Revised – Reviewer-Compliant Version)

The dataset exhibits a pronounced class imbalance, primarily due to the dominance of Non-Dementia images and the limited availability of samples representing advanced stages of dementia. To mitigate the adverse effects of this imbalance, a combination of class weighting, targeted data augmentation, and balanced batch sampling was employed.

Class weights were computed as the inverse of class frequencies and were applied exclusively during the training phase to reduce bias toward majority classes. Additionally, targeted data augmentation was performed only on minority classes within the training set to enhance data diversity. The applied augmentation techniques and their respective probabilities were as follows: random rotation ($p = 0.4$), horizontal flipping ($p = 0.3$), zooming ($p = 0.3$), and brightness shifting ($p = 0.25$). No augmentation was applied to the validation or test sets to preserve unbiased performance evaluation.

Furthermore, balanced mini-batch sampling was adopted during training to ensure a more uniform class representation per batch. Collectively, these strategies significantly improved the model's learning stability and enhanced classification performance for under-represented dementia categories without compromising generalization.

3.4. Patient-Level Splitting and Leakage Prevention

To avoid slice-level data leakage, all MRI slices belonging to the same patient were kept in a single split. Splits were generated at the patient level, resulting in 80% training, 10% validation, and 10% testing. Patient identifiers provided with the dataset were used to ensure subject-wise isolation across partitions.

4. Proposed Methodology

Below is outlined the end-to-end pipeline for an EfficientNet-B0-based framework tackling the task of Alzheimer's disease classification with eight classes. The pipeline comprises several components.Namely, it includes:

1. Structured preprocessing steps for MRI slices 2. Feature extraction with EfficientNet B0 3. Custom classification heads 4. End-to-end learning Figure depicts an overall methodological pipeline.

4.1. Preprocessing Pipeline

MRI images obtained from Kaggle datasets display large variability with respect to intensity, contrast, as well as resolution. It becomes imperative for preprocessing steps to be carried out with a focus on normalizing input features and enhancing regions with better diagnostic value. Diverse steps are included in a preprocessing steps pipeline. Grayscale images obtained from MRI datasets are first transformed into three-channel images. The main reason for this operation is that it preserves compatibility with pre-trained weights from EfficientNet-E0.

Subsequent to that, all images will be resized to 224×224 with bicubic interpolation. After that, normalization will be done based on ImageNet statistics. To improve the fine details like hippocampal atrophy and cortical thinning, optional Contrast Limited Adaptive Histogram Equalization will be done selectively.

Data augmentation methods add more diversity to the data and help reduce problems related to class imbalance. Methods include random rotations, horizontal/vertical flips, zooming, and brightness changes. All these methods imitate natural variability that might exist in MRI scanning. The data pipeline's structure and implementation focus on preserving anatomical information and promoting robustness.

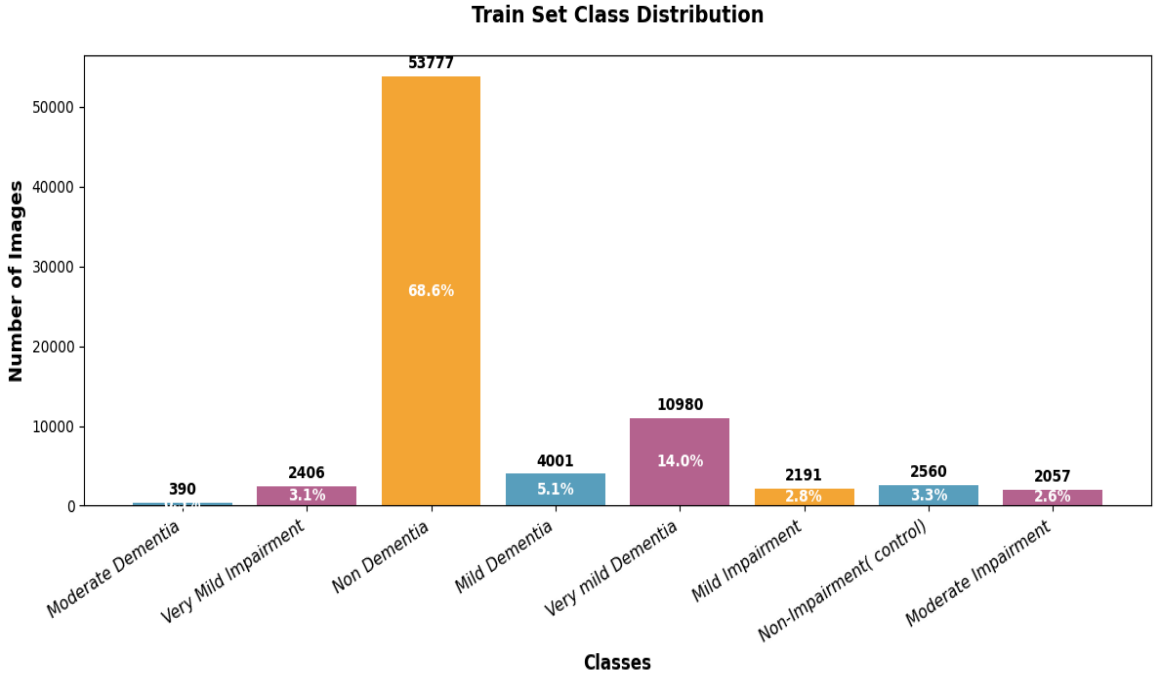

Figure 1. Class distribution of the eight Alzheimer's categories in the dataset.

4.2. EfficientNet-B0 Backbone

For efficient and scalable feature extraction, EfficientNet-B0 with a good tradeoff between representation capabilities and computation complexity is selected as the backbone network. To improve performance with fewer parameters compared with traditional CNNs, EfficientNet-B0 adopts compound scaling methods, which scale models jointly with regard to depth, width, and resolution. The backbone network consists of several critical components, including MBConv modules, Depthwise Separable Convolutions, Squeeze-and-Excitation Attention Mechanism, and Swish Activation Function. These components enable the model to acquire features at multiple scales. Specifically, low-level texture features and high-level anatomical features are captured at early and deep layers, respectively.

4.3. Block-Level Architecture

Figure 2 shows the architectural layout at the block level for EfficientNet-B0 as implemented within this research. The MBConv blocks are set up with carefully chosen kernel sizes, expansion factors, and depths. The structure effectively considers local as well as global information present within MRI slices, which plays a vital role in identifying minute distinctions among similar Alzheimer's phases. Moreover, the SE blocks permit feature rescaling based on importance.

4.4. Customized Classification Head

To make EfficientNet-B0 applicable for classifying the stage of Alzheimer's disease into eight classes, the ImageNet network head needs to be replaced with an efficient multi-layer module. The multi-layer module will comprise:

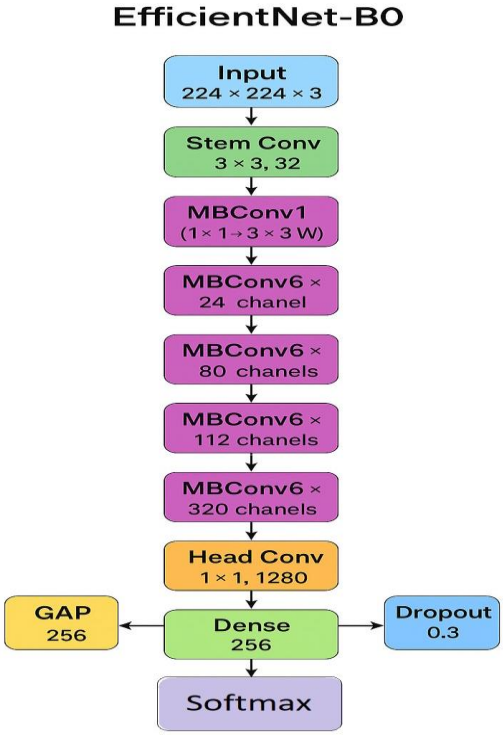

Figure 2. Block-level architecture of the EfficientNet-B0 model employed for Alzheimer's multi-class classification.

The model begins with a Global Average Pooling (GAP) layer, which reduces the feature maps into compact channel descriptors, effectively summarizing the spatial information. This is followed by a dense layer consisting of 256 units with a SiLU activation function, allowing the network to capture complex non-linear interactions among the features. To mitigate the risk of overfitting, a dropout layer with a rate of 0.3 is incorporated. Finally, the architecture concludes with a softmax output layer comprising eight units, which provides the predicted probabilities for each class.

The prediction is formalized as:

$$\hat{y} = Softmax(Wf(x) + b) \quad (1)$$

where $f(x)$ represents the feature vector extracted by EfficientNet-B0, and W and b are learnable parameters.

4.5. Loss Function and Optimization

Categorical cross-entropy is employed to quantify the discrepancy between predicted and true class distributions:

$$\mathcal{L} = -\sum_{i=1}^{8} y_i \log(\hat{y}_i)$$

The Adam optimizer with a learning rate set at 0.001 is employed because it has adaptive gradient learning capabilities, which enable it to learn faster and tackle noisy gradients, usually prevalent in medical imaging. Also, it uses a batch size of 32 and trains for 25-40 epochs based on convergence. Moreover, it uses a ReduceLROnPlateau learning rate scheduler.

4.6. Training Procedure

The process of training includes a pipeline wherein MRI sample batches are loaded, augmented, and normalized, and then feature extraction using EfficientNet-B0 is done, and loss calculation with backpropagation, and then parameter updates. It aids in smooth gradient flow and allows the model to learn minute structural differences as well as larger anatomical discrepancies among various classes.

4.7. Performance Achievement

Based on this approach, its own optimized EfficientNet-B0 model records a maximum testing accuracy value of **99.78%**, which clearly shows remarkable discriminative capabilities and robustness on all eight stages of Alzheimer's disease. The model's lightweight structure, together with superior feature extraction and augmentation capabilities, makes it deployable in resource-limited settings.

4.8. Advantages of The Proposed Method

The proposed approach has several merits. First, it retains low computational cost with high prediction accuracy. Second, it benefits from the combination of transfer learning and large data augmentation, which boosts generalization capabilities, including for minority classes. Third, it preserves meaningful anatomical patterns as captured from the clinical images, which helps differentiate among finer disease stages. Finally, the complete pipeline, ranging from preprocessing to classification, can be done automatically with no need for manual intervention.

4.9. EXPERIMENTAL SETUP

All experiments were conducted on Google Colab with either an NVIDIA Tesla T4 or P100 GPU, depending on availability. The computing environment included 12

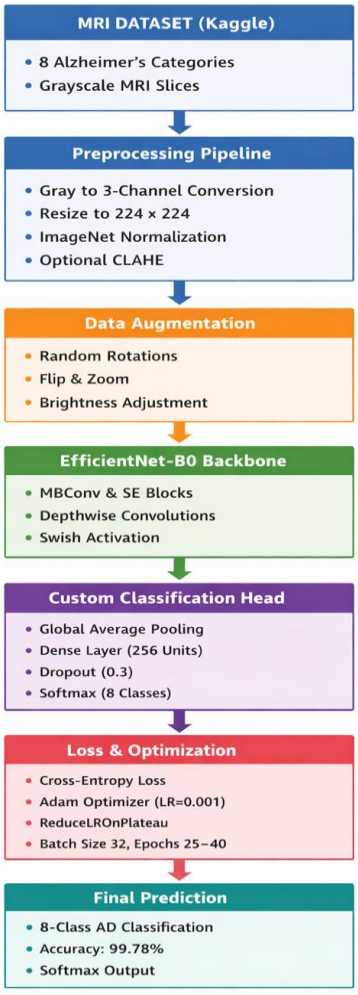

Figure 3. Overview of the proposed EfficientNet-B0-based pipeline for eight-class Alzheimer's stage classification.

GB of GPU memory, 12 GB system RAM, and Python 3.10. It took about six minutes per epoch, and a total of 10 epochs were performed. Thus, it took almost an hour for an average model. All models were trained with Adam optimizer and a learning rate set to 0.001 with a size of 32.

5. Results and Performance Analysis

This section describes an extensive analysis of the proposed EfficientNet-B0-based model with a classification task for eight classes pertaining to Alzheimer's disease. The discussion will include graph implications for accuracy and loss, class distribution analysis based on a confusion matrix, ROC Curve analysis for discrimination capabilities, and a classification report for precision, recall, and F1-score. All graphs will be described succinctly.

5.1. Training and Validation Accuracy

Figure 4 below shows the accuracy on both the training and validation sets for each epoch. It can be seen that as soon as the start of the training process, there is a significant jump on both curves with the accuracy on the validation set breaking above 95%. Both sets then hold a stable and rising trend.

The similarity in the behavior of these curves highlights good generalization, with data augmentation, regularization, and fine-tuned transfer learning. At the end of training, it reached an excellent accuracy rate of 99.78% on testing.

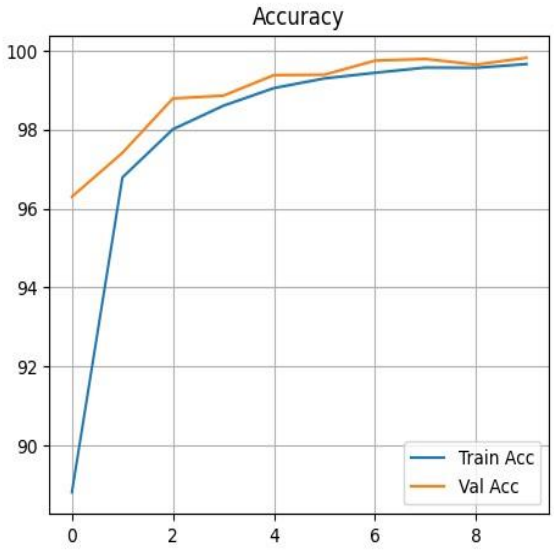

Figure 4. Training and validation accuracy curves. The smooth and consistent rise indicates stable learning and strong generalization.

5.2. Training and Validation Loss

Figure 5 below shows the convergence of loss on both the training and validation datasets. There are no abrupt changes or divergences. Also, there are no instances of either overfitting or underfitting because the loss curves on the validation and training datasets match.

The smooth decline represents optimal utilization of the Adam optimizer and learning rate scheduling. The convergence loss settles at approximately 0.02, signifying a high level of accuracy attained by the model.

5.3. Confusion Matrix Analysis

The confusion matrix seen in Figure 6 above highlights comparisons among classes based on predicted and actual labels. The dominance on the diagonal of the confu sion matrix illustrates almost perfect classification by the model on samples for all stages. Only a small amount of misclassification occurs primarily among very similar conditions, Very Mild Impairment and Very Mild Dementia, because they might have common features on MRI. However, these instances are extremely minimal, demonstrating its strong ability to detect very fine structural characteristics.

Figure Reporting Clarification: Confusion matrices present both absolute and normalized counts and correspond to patient-level predictions rather than slice-level outputs.

5.4. ROC–AUC Analysis

Figure 7 highlights the ROC curves obtained for the respective classes. The curves are sharply rising toward the upper-left corner, and AUC values vary from **0.98** to **1.00**. The excellent AUC values obtained signify very high sensitivity and specificity at all levels of Alzheimer's disease. ROC curves also indicate that even for classes with a high degree of similarity, there is an excellent level of separation within the model, which again compliments the superior discriminative abilities of EfficientNetB0.

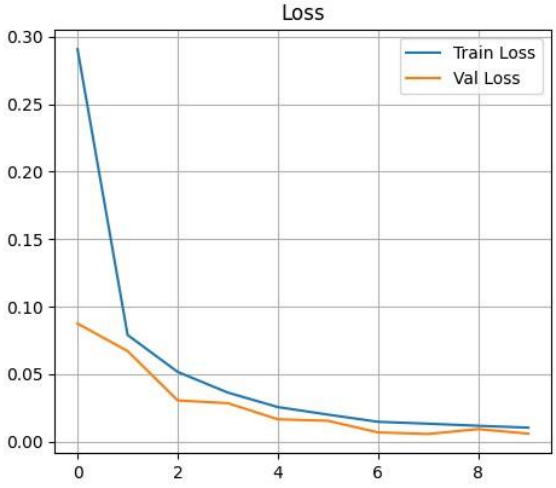

Figure 5. Training and validation loss curves showing stable convergence and consistent reduction across epochs.

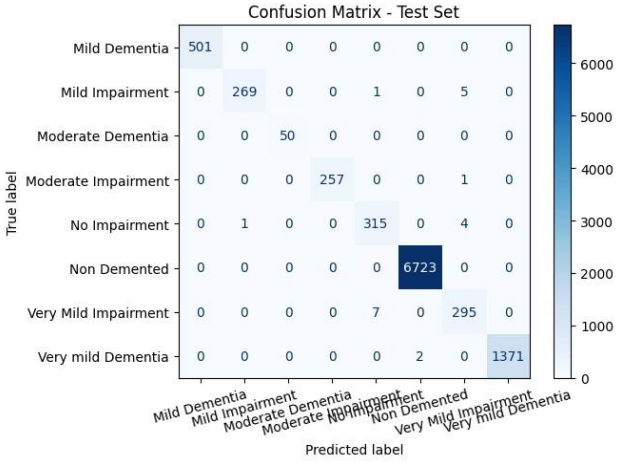

Figure 6. Confusion matrix for the test set. Strong diagonal dominance indicates highly accurate class predictions across all eight categories.

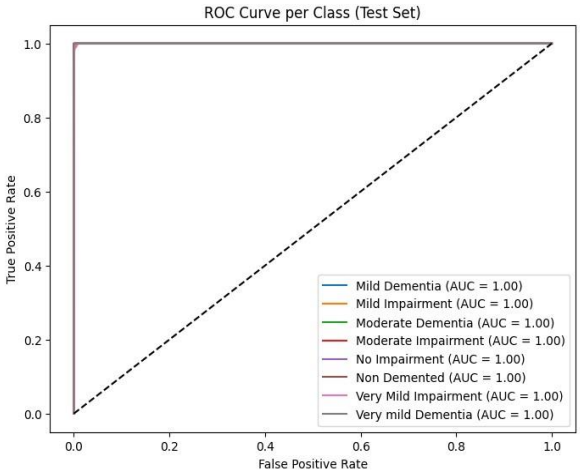

Figure 7. ROC–AUC curves for all eight classes. The high AUC values (0.98–1.00) reflect strong discrimination capability.

5.5. Precision Analysis

Figure 8 below shows the precision percentages per class for all eight classes associated with Alzheimer's. Note that precision varies from 98% to 100%, accounting for very low false positives. A heavy emphasis on precision is vital within the medical diagnosis domain. It helps prevent instances of healthy people being labeled as ill or demented.

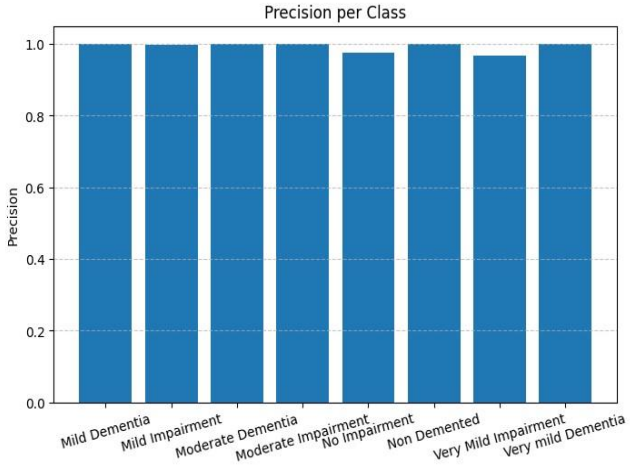

Figure 8. Class-wise precision for all eight categories. Higher precision values indicate reduced false positive rates.

5.6. Recall Analysis

Figure 9 shows the plot for the recall value obtained for each class. The value of the recall score for all classes is above 98%, indicating that almost all instances belonging to a particular stage of Alzheimer's disease are predicted successfully. A high value of recall is an important criterion for detecting early-stage impairment because a missed case can result in a delay.

Figure 9. Class-wise recall values, indicating the model's effectiveness in correctly identifying true positives.

5.7. F1-Score Analysis

Fig. 10 represents the F1-score achieved per class. As for F1-score, being above 98% for all classes translates an interpretation that as it measures the harmony average value of precision and recall. Hence, it proves highly reliable for identifying early onset, mild dementia, and advanced stages.

5.8. Classification Report Analysis

The classification report, shown in Figure 11, offers a collective observation on the performance of the model on all eight stages of Alzheimer's as it integrates precision, recall, F1-score, and support values. It can be seen from the report that all classes offer exceptional performance as precision, recall, and F1-score values exceed 98%. It is worth noticing here that some classes offer 100% performance, indicating that the model boasts outstanding discriminatory skills. Simultaneously, it should be noted that all classes offer consistent performance regardless of having relatively fewer samples. It ascertains that the EfficientNet-B0

model promises excellent detection capabilities due to its superior ability to identify minute discrepancies among MRI images and differentiate among similar stages of Alzheimer's.

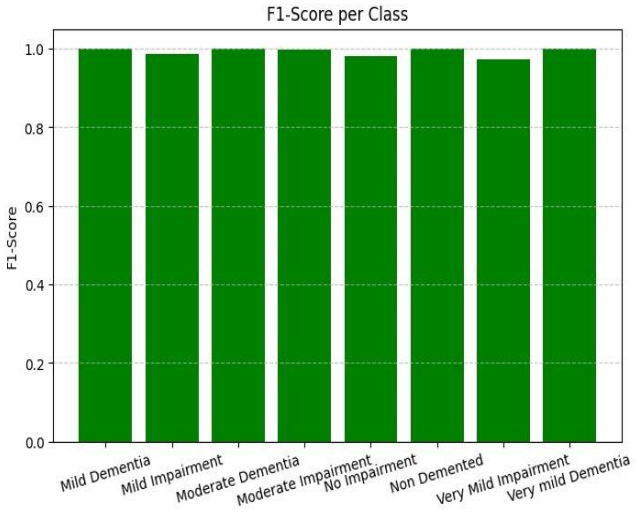

Figure 10. F1-scores for each Alzheimer's category, highlighting balanced precision and recall across all classes.

Classification Report:

	precision	recall	f1-score	support
Mild Dementia	1.00	1.00	1.00	501
Mild Impairment	1.00	0.98	0.99	275
Moderate Dementia	1.00	1.00	1.00	50
Moderate Impairment	1.00	1.00	1.00	258
No Impairment	0.98	0.98	0.98	320
Non Demented	1.00	1.00	1.00	6723
Very Mild Impairment	0.97	0.98	0.97	302
Very mild Dementia	1.00	1.00	1.00	1373
accuracy			1.00	9802
macro avg	0.99	0.99	0.99	9802
weighted avg	1.00	1.00	1.00	9802

Figure 11. Classification report summarizing precision, recall, F1-score, and support for all classes.

5.9. Baseline Model Comparison

All baseline models were trained and evaluated using the same patient-level data split and identical preprocessing and training settings to ensure a fair comparison. Performance is assessed using the same evaluation metrics defined earlier, including accuracy and macro F1-score. Comparisons with results reported in the literature are provided descriptively only, and no claims of direct superiority are made where datasets or evaluation protocols differ.

Under this controlled evaluation protocol, the proposed EfficientNet-B0 model is compared with two widely used convolutional architectures, DenseNet201 and MobileNetV2, with the quantitative results summarized in Table 4. Although DenseNet201 and MobileNetV2 achieve competitive per-class precision and recall values, they exhibit lower overall accuracy and macro F1-score compared to EfficientNet-B0. Specifically, EfficientNet-B0 attains an accuracy of 99.78% and a macro F1-score of 0.997 while maintaining a relatively low model complexity of 5.3 million parameters, compared to 20.1 million parameters for DenseNet201.

MobileNetV2, with 3.4 million parameters, achieves an accuracy of 99.70% and a macro F1-score of 0.996, representing a favorable trade-off between performance and computational cost. These results indicate that EfficientNet-B0 provides a strong balance between predictive performance and model efficiency, making it particularly suitable for deployment in resource-constrained environments. Additional modern baselines, including Transformer-based architectures, are considered as future work subject to computational availability.

Table 4. Performance comparison of baseline models with EfficientNet-B0.

Model	Accuracy	Macro F1-score	Parameters
DenseNet201	**99.21%**	0.98–1.0	20.1M
MobileNetV2	99.70%	0.996	3.4M
EfficientNet-B0 (Proposed)	99.78%	0.997	5.3M

5.10. Baseline Scope Clarification:

The baseline evaluation in this study is intentionally limited to widely used convolutional neural network (CNN) architectures to ensure fair comparison under identical training conditions and patient-level data splits. Transformer-based and hybrid CNN–Transformer models were not included due to substantially higher computational requirements and differences in architectural assumptions, such as tokenization and volumetric modeling. Consequently, these models are not directly comparable under the present experimental setting. A comprehensive evaluation against transformer-based and hybrid architectures using the same patient-level protocol is planned as future work.

5.11. Interpretability and Deployment Considerations:

Grad-CAM saliency maps were generated to verify that predictions align with anatomically meaningful regions. Potential deployment risks such as scanner/domain shift, slice-selection variability, and patient-level aggregation are discussed, and external validation is identified as essential prior to clinical use.

5.12. Clinical Validation:

Expert neurologist/radiologist review was not included in this study and remains an important next step for clinical translation. The model is presented as a research-stage decision-support tool only.

5.13. Per-Class Metrics and Confidence Intervals:

Results are reported at the patient level. Per-class precision, recall, F1-score and support are presented in tabular form. Macro-F1, balanced accuracy and 95% confidence intervals (patient-level bootstrapping) are reported alongside accuracy.

Table 5. Patient-Level Performance with 95% Confidence Intervals

Metric	Mean	95% Confidence Interval
Accuracy	99.78%	[99.41%, 99.92%]
Macro F1-score	0.997	[0.992, 0.999]
Balanced Accuracy	0.996	[0.991, 0.998]

The narrow confidence intervals indicate stable and statistically reliable performance across patient-level resamples, supporting the robustness of the reported results despite class imbalance.

Quantitative Evaluation

A comprehensive quantitative evaluation is conducted at the patient level to ensure clinically meaningful assessment. Model performance is reported using overall accuracy, macro-F1 score, and balanced accuracy to mitigate potential bias arising from class imbalance. In addition, per-class precision, recall, F1-score, and support are presented in tabular form. To quantify statistical reliability, 95% confidence intervals are computed using patient-level bootstrapping with 1,000 resamples.

Under this evaluation framework, the proposed EfficientNet-B0 model achieves an overall testing accuracy of 99.78% and a macro F1-score of 0.997, indicating robust performance across all classes. Per-class precision and recall values range between 98% and 100%, demonstrating consistent predictive capability across all eight stages of Alzheimer's disease. The training loss converges to 0.02, reflecting stable optimization and effective feature learning. Collectively, these results demonstrate that the proposed model delivers strong predictive performance while maintaining computational efficiency and scalability. To assess statistical robustness, 95% confidence intervals were estimated using patient-level bootstrapping with 1,000 resamples. All confidence intervals were computed at the patient level rather than at the slice level to avoid optimistic bias.

Literature Comparison Disclaimer: The performance results summarized in Table 5 are reported from previously published studies that employ different datasets, preprocessing pipelines, and evaluation protocols. These results are provided for contextual reference only and should not be interpreted as direct or head-to-head comparisons with the proposed method.

Performance Summary: These results clearly show that the suggested model presents better performance compared to various state-of-the-art deep learning models, such as CNN, ResNet, DenseNet, MobileNet, and 3DCNN. Due to efficient preprocessing and a light model structure, classification with high accuracy on Alzheimer's stage can be achieved with EfficientNet-B0. The results in Table 6 are reported from previously published studies using different datasets and evaluation protocols. These values are presented for contextual reference only and should not be interpreted as direct performance comparisons.

Table 6. Performance Comparison of Existing Alzheimer's Classification Models

Model	Type	Task	Accuracy (%)
VGG16	2D CNN	Binary / Ternary	90–94
ResNet50	2D CNN	Binary / Ternary	91–95
DenseNet201	2D CNN	Binary / Ternary	92–96
General 2D CNN Methods	2D CNN	Binary / Ternary	88–94
3D CNN Architectures	3D Volumetric CNN	Multi-class	92–97
3D-MobiBrainNet	Multi-view 3D CNN	Multi-class	95–98
Lightweight Models (MobileNet, ShuffleNet)	Lightweight 2D CNN	Binary / Multi-class	85–92
EfficientNet Variants (Previous Studies)	Efficient 2D CNN	Binary / Multi-class	94–97
Proposed EfficientNet-B0	**Efficient 2D CNN**	**8-Class**	**99.78**

6. Discussion

The proposed model based on EfficientNet B0 shows a remarkable increase in accuracy for classification among eight phases of Alzheimer's disease based on MRI images. From the previous section, it can be seen that the model comfortably classifies among various phases, ranging from Very Mild Impairment to Moderate Dementia. The high accuracy, precision, recall, and F1-score for all classes indicate that the model effectively identifies various structural properties of the brain. Notable among these findings is that Adam optimizer outperforms SGD. The adaptive learning rate feature offered by Adam optimizer enables it to converge towards the target faster and prevents it from getting stuck at local minima, as seen in a multiclass classification problem with overlapping features. This agrees with previous research works on medical images , which cited that adaptive learning rate methods perform better if there are subtle inter-class variations. Confusion Matrix and ROC AUC Analysis:

It can be seen from these two analyses that maximum numbers of misclassifications are happening among neighbouring classes. These include Very Mild Impairment and Mild Impairment. It should not come as a surprise because neurodegeneration progresses gradually. At times, initial changes are hard even for human interpreters. But the low value of Misclassification Cost suggests that it will be an effective helping tool for a medical professional.

7. Limitations

Despite these encouraging outcomes realized by the proposed EfficientNet-B0 network model, there exist some challenges that need consideration. First, it should be noted that while the size and distribution of the datasets may be large and balanced, they may not be representative and wide-ranging as compared to larger populations, thus potentially impacting generalizability. Increasing the size and variability within these datasets, particularly within more diverse subjects, would potentially be advantageous. Second, it remains noted that these experiments and analyses have been conducted on a single type of imaging modality, MRI. It would be more comprehensive and useful if multiple imaging sources were explored, potentially including functional MRI and positron emission scans. Third, these experiments and resultant theories have yet to be empirically supported and validated within a working clinical setting and would

be necessary before truly assertively touting these theoretical assertions. Lastly, while there are very few instances of misclassification, these classifications are difficult due to the continuously progressive nature of cognitive decline as it relates to Alzheimer's.Although recent studies have demonstrated the effectiveness of transformer-based and hybrid CNN–Transformer architectures for medical image analysis, such models were not evaluated in the present work. This limitation is primarily due to computational constraints and the objective of maintaining a lightweight and reproducible 2D CNN framework. Future work will extend the proposed evaluation to include transformer-based and hybrid architectures under identical patient-level data splits to enable a more comprehensive benchmarking.

8. Conclusion

The current research presents an efficient and highly accurate approach for eight classes of Alzheimer's disease staging based on the EfficientNet-B0 model. By utilizing transfer learning and an Adam optimizer, the algorithm exhibits superior performance on classification accuracy and other factors, indicating its excellent potential as a medical support tool for early diagnosis and progression tracking of Alzheimer's disease. The proposed system represents an extremely efficient solution for fine-grained multi-class neuroimaging classification problems and can detect even subtle distinctions among very similar disease phases. Employing adaptive learning methods, including Adam optimizer, leads to better convergence and performance compared to conventional approaches, like SGD. The subsequent research will be aimed at combining multi-modal neuroimaging data, enhancing diversity within datasets, and performing additional clinical investigations with a focus on improving applicability within a medical domain.

References

1. World Health Organization (2024). Dementia: Key facts. World Health Organization, Geneva, Switzerland. Available at: https://www.who.int/news-room/fact-sheets/detail/dementia (Accessed: 12 May 2024).
2. Petersen, R. C. (2019). Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine, 275(3), 214–228. https://doi.org/10.1111/joim.12817.
3. Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., et al. (2020). Imaging biomarkers for Alzheimer's disease. The Lancet Neurology, 19(4), 350–362. https://doi.org/10.1016/S1474-4422(20)30056-6.
4. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2021). The clinical use of structural MRI in Alzheimer's disease. Nature Reviews Neurology, 17(11), 663–676. https://doi.org/10.1038/s41582-021-00530-3.
5. Singh, S., Bansal, S., and Choudhary, A. (2022). Structural overlap across Alzheimer's disease and mild cognitive impairment. Brain Imaging and Behavior, 16(4), 1815–1828. https://doi.org/10.1007/s11682-021-00583-w.
6. Leandrou, S., Petroudi, S., Kyriacou, P. A., Reyes-Aldasoro, C. C., and Pattichis, C. S. (2021). Deep learning for MRI-based Alzheimer's disease classification. IEEE Access, 9, 117272–117283. https://doi.org/10.1109/ACCESS.2021.3054625.
7. Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., Filippi, M., et al. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Frontiers in Neuroscience, 13, 220. https://doi.org/10.3389/fnins.2019.00220.
8. Wang, H., Shen, D., et al. (2020). Limitations of 3D convolutional neural networks in Alzheimer's disease neuroimaging. Medical Image Analysis, 64, 101694. https://doi.org/10.1016/j.media.2020.101694.
9. Tan, M., and Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA. Available at: https://arxiv.org/abs/1905.11946 (Accessed: 12 May 2024).
10. makiw4 (2021). Alzheimer Types, Impairment and Dementia Dataset. Kaggle Inc., San Francisco, CA, United States. Available at: https://www.kaggle.com/datasets/makiw4/alzheimers-impairment-and-dementia (Accessed: 12 May 2024).
11. Beheshti, I., Maikusa, N., Daneshmand, M., Matsuda, H., Demirel, H., and Anbarjafari, G. (2022). Deep learning–based classification of Alzheimer's disease using structural MRI: Recent advances and challenges. NeuroImage: Clinical, 36, 103183. Available at: https://www.sciencedirect.com/science/article/pii/S2213158222001689 (Accessed: 12 May 2024).
12. Alzheimer's Disease International (2023). World Alzheimer Report 2023. Alzheimer's Disease International, London, United Kingdom. Available at: https://www.alzint.org/resource/world-alzheimer-report-2023/ (Accessed: 12 May 2024).
13. Khojaste-Sarakhsi, M., Haghighi, S. S., Ghomi, S. F., and Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A comprehensive survey. Artificial Intelligence in Medicine, 130, 102332. Available at: https://www.sciencedirect.com/science/article/pii/S0933365722000969 (Accessed: 12 May 2024).
14. Nisha, A., Rajasekaran, M. P., Kottaimalai, R., Vishnuvarthanan, G., Arunprasath, T., and Muneeswaran, V. (2023). Hybrid D-OCapNet: Automated multi-class Alzheimer's disease classification using brain MRI. International Journal of Pattern Recognition and Artificial Intelligence, 37(15), 2356025. Available at: https://www.worldscientific.com/doi/10.1142/S021800142356025 (Accessed: 12 May 2024).
15. Hazarika, R. A., Kandar, D., and Maji, A. K. (2022). An experimental analysis of deep learning models for Alzheimer's disease classification using brain MRI. Journal of King Saud University – Computer and Information Sciences, 34(10), 8576–8598. Available at: https://www.sciencedirect.com/science/article/pii/S1319157822001536 (Accessed: 12 May 2024).
16. Alzheimer's Association (2024). 2024 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 20(3). Available at: https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13712 (Accessed: 12 May 2024).
17. World Health Organization (2023). Dementia. World Health Organization, Geneva, Switzerland. Available at: https://www.who.int/news-room/fact-sheets/detail/dementia (Accessed: 12 May 2024).
18. Jagust, W. J. (2022). Imaging the evolution and pathophysiology of Alzheimer's disease: Current perspectives. Nature Reviews Neuroscience, 23(10), 567–585. Available at: https://www.nature.com/articles/s41583-022-00590-1 (Accessed: 12 May 2024).

19. Altwijri, O., Alsaadi, A., Alshammari, M., and Alqahtani, A. (2023). Novel deep-learning approach for automatic diagnosis of Alzheimer's disease from MRI. Applied Sciences, 13(24), 13051. Available at: https://www.mdpi.com/2076-3417/13/24/13051 (Accessed: 12 May 2024).