ARTICLE

A Hybrid Deep Learning Multi-Class Classification Model for Alzheimer's Disease Using Enhanced MRI Images

Ghadah Naif Alwakid[*]

Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka, 72341, Al Jouf, Saudi Arabia

*Corresponding Author: Ghadah Naif Alwakid. Email: gnalwakid@ju.edu.sa

ABSTRACT: Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that significantly affects cognitive function, making early and accurate diagnosis essential. Traditional Deep Learning (DL)-based approaches often struggle with low-contrast MRI images, class imbalance, and suboptimal feature extraction. This paper develops a Hybrid DL system that unites MobileNetV2 with adaptive classification methods to boost Alzheimer's diagnosis by processing MRI scans. Image enhancement is done using Contrast-Limited Adaptive Histogram Equalization (CLAHE) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). A classification robustness enhancement system integrates class weighting techniques and a Matthews Correlation Coefficient (MCC)-based evaluation method into the design. The trained and validated model gives a 98.88% accuracy rate and 0.9614 MCC score. We also performed a 10-fold cross-validation experiment with an average accuracy of 96.52% (±1.51), a loss of 0.1671, and an MCC score of 0.9429 across folds. The proposed framework outperforms the state-of-the-art models with a 98% weighted F1-score while decreasing misdiagnosis results for every AD stage. The model demonstrates apparent separation abilities between AD progression stages according to the results of the confusion matrix analysis. These results validate the effectiveness of hybrid DL models with adaptive preprocessing for early and reliable Alzheimer's diagnosis, contributing to improved computer-aided diagnosis (CAD) systems in clinical practice.

KEYWORDS: Alzheimer's disease; deep learning; MRI images; MobileNetV2; contrast-limited adaptive histogram equalization (CLAHE); enhanced super-resolution generative adversarial networks (ESRGAN); multi-class classification

1 Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that primarily affects cognitive functions such as memory, reasoning, and behavior [1]. As one of the leading causes of dementia, AD poses significant challenges to healthcare systems worldwide due to its irreversible nature and lack of a definitive cure [2]. An estimated 7.2 million Americans aged 65 and older are living with Alzheimer's dementia in 2025, representing approximately 11% of this population, highlighting the importance of improved diagnostic tools. Early and accurate diagnosis of AD is critical for timely intervention and management, potentially improving the quality of life for patients and alleviating healthcare burdens. Magnetic Resonance Imaging (MRI) has emerged as a non-invasive diagnostic tool for detecting structural brain abnormalities associated with AD progression [3]. However, traditional diagnostic approaches, including radiological assessment and Machine Learning (ML) techniques, often suffer from low contrast, class imbalance, and suboptimal feature extraction, leading to inconsistent detection accuracy [4,5].

Recent advancements in Deep Learning (DL) [6], particularly Convolutional Neural Networks (CNNs), have significantly improved automated Alzheimer's detection by leveraging MRI images [7,8]. CNN-based models such as MobileNetV2 and DenseNet121 have demonstrated high accuracy in AD classification [9,10]. However, these models encounter several challenges, including variability in MRI image quality due to different acquisition conditions, class imbalance in datasets where underrepresented categories such as Moderate Demented cases lead to biased model predictions, and feature extraction limitations where traditional CNNs may fail to capture subtle variations in brain structure that differentiate AD stages effectively [11,12].

AD results in concrete brain changes that have been observed through the MRI scan. At the earliest stages (e.g., Very Mild Dementia), parts of the brain, like the hippocampus, shrink, and the ventricles start to increase in size [13]. This shrinkage extends to other parts, including the parietal lobes, as the disease transitions to Mild and then Moderate stages. They are attributed to loss of brain cells and accumulation of toxic proteins such as amyloid plaques [14,15]. These changes may be difficult to identify by the normal human eye. Fortunately, these tiny patterns and variations in the brain MRI imaging can be identified using DL models [16]. Therefore, there is an urgent requirement to create better and adaptable DL models that solve fundamental difficulties in Alzheimer's detection [17]. The current CNN-based algorithms suffer from inadequate feature optimization capabilities when processing complex neuro-imaging data because minor brain structure variations indicate disease progression [18]. Additional image enhancement methods must be developed because noise and artifacts in MRI scans degrade classification accuracy. The design of DL models should incorporate measures to balance classes because otherwise, they will favor majority classes over minority classes to maintain fair performance throughout different stages of AD [19]. Advanced image preprocessing with adaptive classification approaches and improved evaluation methods must be combined for better accuracy, clinical utility, and generalization.

This study aims to design a lightweight and efficient DL model for achieving accurate multi-class classification of Alzheimer's disease using augmented MRI images. It proposes a hybrid DL Approach combining MobileNetV2 and an Adaptive Classification Strategy to improve AD detection. A novel model on image contrast and resolution enhancement using Contrast-Limited Adaptive Histogram Equalization (CLAHE) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) is used to receive high-quality input data for classification. Furthermore, a class weighting strategy and Matthews Correlation Coefficient (MCC)-based evaluation is included for better robustness against class imbalance. We seek to enhance diagnostic performance through increased image enhancement techniques while preserving computational efficiency. The main objective is to distinguish the different stages of Alzheimer's disease (Mild, Moderate, severe) with classification accuracy and low resource consumption. The existing approaches suffer from imbalanced datasets, high computational resource consumption, and poor generalization. The novelty of this study lies in augmenting the classification of Alzheimer's severity levels with MRI image contrast and resolution enhancement using a lightweight DL model. It was noted that these increases in classification accuracy have a direct clinical significance. The proper distinction of early and moderate stages of AD would allow for promoting adequate intervention in due time, feeding back on misdiagnosis, and subsequent improved results in patients. The lightweight nature of the architecture of the model assists in deployment into real-world clinical conditions, such as low-resource settings. This augmentation of feature visibility enhances class imbalance while providing computational efficiency, improving practicality for clinical use compared to traditional, superfluous models. This paper makes several key contributions:

1. A novel hybrid DL framework integrating MobileNetV2 with an adaptive classification strategy to enhance the accuracy and robustness of Alzheimer's detection from MRI images.

2. Advanced image preprocessing techniques, including CLAHE and ESRGAN, to improve image contrast and resolution, thereby facilitating better feature extraction.
3. Enhanced classification metrics by incorporating MCC-based evaluation to address class imbalance issues, ensuring a more reliable classification performance.
4. A comprehensive performance evaluation comparing the proposed approach with state-of-the-art models, demonstrating its superiority with an improved accuracy and MCC score.
5. A robust model testing is carried out with 10-fold cross-validation for enhanced trustworthiness and generalizability for various data splits.
6. A comprehensive ablation study is performed to assess the individual contribution of each enhancement component (CLAHE, ESRGAN, and class weighting).
7. The study also incorporated model interpretability using LIME, providing class-wise visual explanations to ensure transparency in decision-making.

The rest of this paper is organized as follows: Section 2 reviews related work in Alzheimer's detection using DL. Section 3 describes the proposed methodology, including image preprocessing techniques and model architecture. Section 4 presents dataset details, the experimental setup, and performance evaluation metrics. Section 5 discusses the results and comparative analysis with existing models. Finally, Section 6 concludes the paper with future research directions.

2 Related Work

Alwakid et al. [20] proposed the integration of blockchain with generative AI. However, it has scalability, interoperability on multi-tiered heterogeneous platforms, and energy consumption spanning diverse platform issues. Nandal et al. [21] used CLAHE for MRI image contrast enhancement and ESRGAN for resolution enhancement to classify Alzheimer's disease. Their approach, however, demands excessive computational resources and lacks generalizability. Duba-Sullivan et al. [22] applied super-resolution techniques to Alzheimer's MRI scans and defect detection in XCT-based imaging. Their results achieved sharper imaging yet were compromised by poor volumetric integrity during the reconstruction of the XCT scans. Liu et al. [23] applied CLAHE and ESRGAN enhancement to the MRI images of Alzheimer's disease patients. Though they enhanced the resolution, there was high computational overhead. Pan et al. [24] analyzed imaging in healthcare with an emphasis on the super-resolution of XCT scans and remote sensing. Their approaches suffered from increased computation requirements and performance reliance on particular datasets. Appiah and Mensah [25] utilized YOLOv7 with ESRGAN to bolster off-the-shelf object detection in self-driving cars while focusing on difficult weather scenarios. Despite strong accuracy improvements, their model struggled with extreme occlusion and real-time execution issues. Korkmaz and Tekalp [26] applied super-resolution imaging techniques across healthcare fields, XCT, remote sensing, and autonomous vehicles. Detection accuracy was enhanced, but so was the difficulty in obtaining realistic frame rate operation.

Marciniak and Stankiewicz [27] performed multi-class classification of retinal diseases using OCT imaging. However, the solutions posed problems of high GPU demand and poor adaptability. Singh and Patnaik [28] utilized AI-based enhancement methods for deeper cancer detection and remote sensing imaging. However, the model and dataset dependence increased applicability restrictions. Sengodan [29] incorporated dual-attention and CLAHE preprocessing with EfficientNet to optimize breast cancer detection and expand OCT imaging. However, it brought high computation costs and difficulties for real-time implementation. Dasu et al. [30] improved AD classification by utilizing contrast enhancement followed by SVM, KNN, and Decision Tree classifiers. Arora and Patnaik [31] devised a brain tumor classification system integrating CLAHE with self-adaptive spatial attention and DenseNet/ResNet architectures. However, their validation on more extensive and heterogeneous datasets is critical to support broader clinical use. Agarwal et al. [32]

integrated Real-ESRGAN and GFPGAN preprocessing into CNN models for medical image analysis and achieved marked improvements. However, they faced complications of refined image processing for AI-based decision-making, including high computational cost and minimal enhancement of disease-specific diagnostic focus.

Most studies have focused on using MRI and DL algorithms to detect and diagnose AD. These methods, however, tend to incur excessive computing costs, lack generalizability, depend on a single dataset, and fail to classify multiple severity levels efficiently. Furthermore, most of these methods focus on image quality enhancement rather than considering resource limitations or robust classification for real-world applications. Such problems indicate the need for adaptive frameworks with low computations designed to reliably enhance performance at all stages of the disease and provide accurate diagnoses. Thus, the current study aims to develop a novel hybrid DL model incorporating enhanced MRI imaging techniques with adaptive classification to overcome the abovementioned limitations. It develops a Hybrid multi-class Classification DL Model that unites MobileNetV2 with adaptive classification methods to improve AD detection accuracy. Real-ESRGAN and GFPGAN serve within the proposed methodology to improve MRI image resolution before supplying the data for classification processes. A newly developed optimization method called Gradient Centralization (GC) enhances both model generalization abilities and execution period and stability during convergence. A class-weighting method and MCC-based evaluation were added to manage imbalanced datasets, ensuring reliable and impartial classification over various AD severity levels. The proposed framework serves as a solution for uniting higher-resolution images with optimized deep-learning capabilities, which results in better computer-assisted Alzheimer's diagnosis procedures.

Ozdemir and Dogan [33] introduced another classification method for AD detection through CNN architecture and attention-based feature fusion. They used histogram equalization. However, the model has some concerns regarding generalization in different categories of patients. Unlike the current state-of-the-art on image enhancement or image classification, our strategy combines both a lightweight and scalable model with better inference with improved images. In contrast to the previous approaches that usually require complex computations or do not scale well with the severity of AD, the proposed hybrid model uses CLAHE, followed by ESRGAN to enhance the image quality, and MobileNetV2 takes up less computing power to perform efficient feature extraction and adaptively designed classifications techniques such as class weighting and MCC evaluation to address class imbalance. The presented multifaceted framework deals with fundamental drawbacks in the scalability, interpretability, and reliability of the earlier studies. Recent developments in medical image analysis highlight the strong performance of hybrid DL models that integrate various architectural elements. Kılıç et al. [34] proposed HybridVisionNet, a hybrid deep learning framework that combines combines multiple deep learning structures to enhance automated ocular disease detection, achieving improved results through the collaborative functioning of its components. The main limitations of this study include increased computational complexity compared to single-architecture models, potential dataset bias despite cross-dataset validation. Almufareh et al. [35] proposed a melanoma classification model using a fine-tuned CNN trained on the ISIC dataset. The approach demonstrated improved accuracy in distinguishing malignant and benign skin lesions through deep feature extraction and transfer learning. However, it rely on a single dataset, which may reduce generalizability, and limited exploration of real-world clinical deployment constraints such as varying imaging conditions and device heterogeneity.

3 Proposed Methodology

This study proposes a hybrid DL approach with image enhancement methods and adaptive classification for detecting different stages of AD from MRI scans. A methodology is formulated to tackle three fundamental challenges: (i) suboptimal image quality, (ii) dataset class sparsity, and (iii) high computational costs,

which are standard for most traditional machine learning models in the medical image analysis field. The proposed framework increases the detection reliability using image preprocessing and enhancement, feature extraction and classification, and post-processing and evaluation, shown in Figs. 1–3. While maintaining high complexity, the multi-stage methodology addresses a unique clinical need for accuracy and effective low-resource demand operations. With the imbalance class distribution, high image quality demands, and low computation requirements, the proposed approach enables robust AD detection in real-time clinical settings. Table 1 summarizes the notations used throughout the preprocessing and classification pipelines. This ensures clarity and consistency in the presentation of algorithms and mathematical formulations in the proposed methodology.
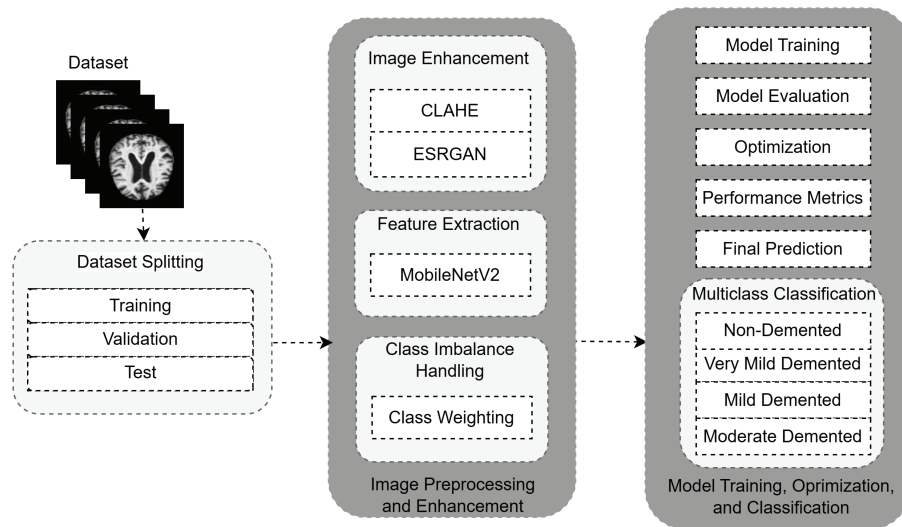

Figure 1: Summary of the proposed pipeline of detecting AD. the CLAHE and ESRGAN resultant MRI images subsequently undergo feature extraction and stage classification into AD stages with MobileNetV2
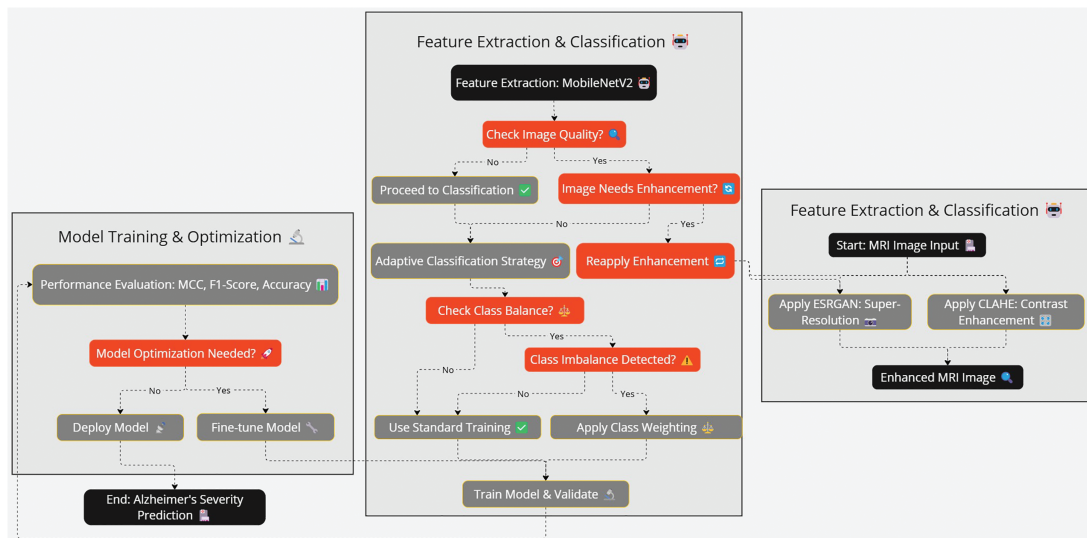

Figure 2: The flow of the methodology follows preprocessing to the classification of the images. The ESRGAN and CLAHE enhance the quality of input, and then MobileNetV2 and dense layers are used to cast strong multi-class classification
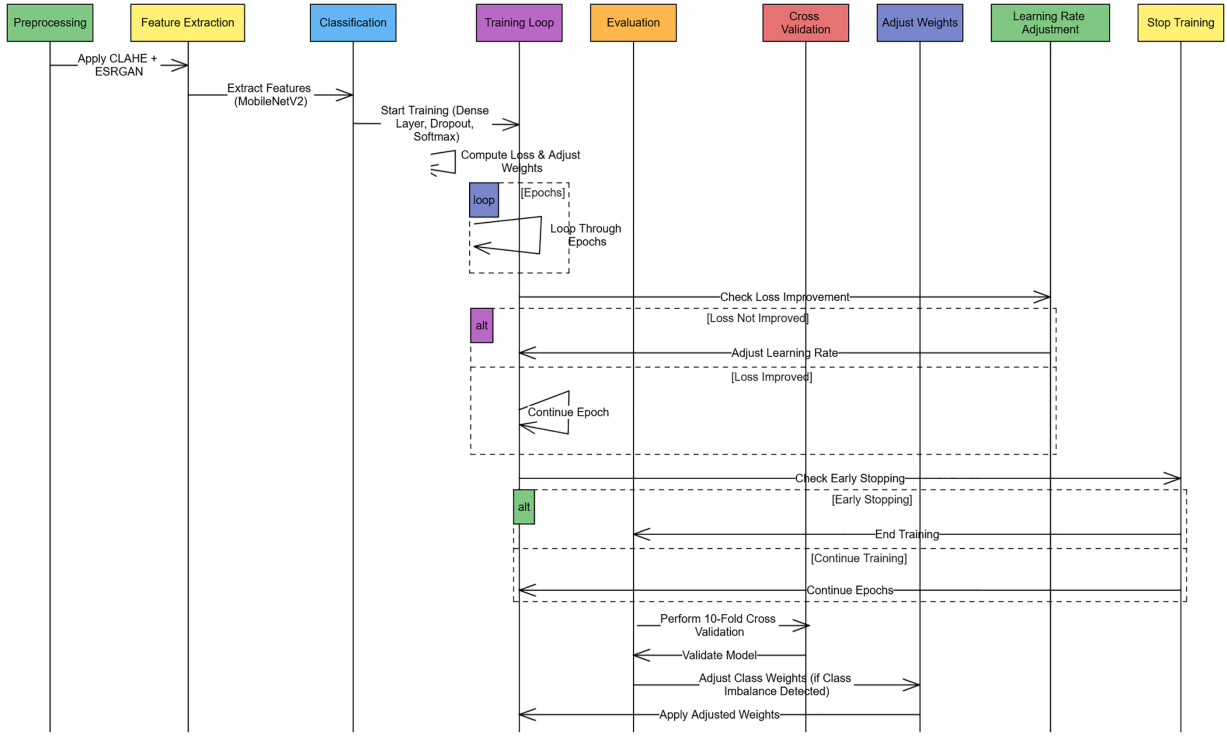
Figure 3: Sequence diagram of the Alzheimer Disease Detection Pipeline explaining the procedures of each preprocessing feature extraction training and validation. Early stopping, rate learning adjustment, and class weight adjustment are done as required at each stage

Table 1: Notation summary for preprocessing and classification pipeline

Symbol	Description
I_i	Raw MRI image of the i-th subject
C_i	CLAHE-enhanced version of I_i
S_i	Super-resolved image after applying ESRGAN to C_i
T_i	Final preprocessed tensor fed to the model
$\mathcal{D}$	Original dataset of raw MRI images
$\mathcal{D}'$	Final set of preprocessed MRI tensors
f	Feature extraction function (e.g., MobileNetV2 layers)
P_i	Predicted class label for the i-th subject
GT_i	Ground truth label for the i-th subject
θ	Trainable parameters of the deep learning model
$\mathcal{L}$	Loss function used during training

3.1 Image Preprocessing and Enhancement

MRI images employed in Alzheimer's screening tend to suffer from low contrast, noise, and reduced spatial resolution, which are more pronounced than in other medical imaging. Combining these factors complicates feature extraction in DL models and harms classification accuracy. A two-step image enhancement procedure was implemented to combat this with the proposed methodology. First, the local contrast of the

MRI images is improved using CLAHE. Then, critical anatomical details are preserved while the image's resolution is boosted using ESRGAN. With the structural information in the MRI images sharpened and enhanced, critical features can be better distinguished, which improves performance during classification and feature extraction in the later stages of the proposed model. Algorithm 1 represents the preprocessing pipeline that enhances MRI images through a combination of contrast improvement and super-resolution. Each image undergoes CLAHE-based enhancement, ESRGAN-based upscaling, normalization, and tensor conversion before being used for model training.

1. Contrast Enhancement Using CLAHE: Contrast is important for visualizing various structures on the MRI images. This paper uses the CLAHE method to improve the local contrast. It is achieved through mechanisms of partitioning the image into small sections referred to as contextual tiles before processing each of them through the use of histogram equalization while at the same time preventing amplification of noise. This localized approach allows for avoiding global changes, in contrast, which will cause some changes to the image. In comparison, CLAHE increases the contrast and reduces the noise levels, ensuring that crucial structural valuable information in identifying Alzheimer's is retained. CLAHE is important as it influences the natural visibilities of little abnormalities in the brain that may be an early sign of AD, thereby improving the system's detection ability. This contribution is essential because it identifies the disease's initial stages because their contrast is low. The CLAHE parameters were empirically selected based on visual inspection and enhancement performance across a subset of training data. Specifically, we used a clip limit of 2.0 and a tile grid size of (8, 8), which provided a balance between noise suppression and contrast enhancement without introducing visual artifacts. These settings are widely adopted in medical image preprocessing and have demonstrated consistent results in our experiments.

2. Super-Resolution Using ESRGAN: Although the visibility of images is enhanced through contrast, the MRI images usually have a low resolution, making it hard for classifiers to detect the minor characteristics of the brain structure. To this end, the methodology makes use of ESRGAN, which is a high-quality image upsampling structure that is currently on par with the state-of-the-art. To increase the spatial resolution of the MRI images with minimized structural distortion, ESRGANutilizes a GAN-based structure called Generative Adversarial Network along with three loss functions: adversarial loss, perceptual loss, and content loss. The advantage of using ESRGAN is that it presents and enhances the fine-grained structural features of the MRI images. It is essential in Alzheimer's diagnosis since this condition impacts small structures in the brain that cannot be visualized clearly by conventional techniques. Thus, the image data received using ESRGAN provides images of the highest quality, which makes the detection more accurate. It should be mentioned that ESRGAN-based enhancement was done offline during the dataset preparation phase and should not be considered as a part of the real-time inference. This architecture makes sure that the fact that super-resolution is computationally expensive does not affect the efficiency of clinical deployment. For super-resolution enhancement of MRI images, we utilized the original Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) architecture based on the Residual-in-Residual Dense Block (RRDB) structure. The model was implemented using the official ESRGAN repository and pretrained weights provided by the authors. No custom modifications were made to the network architecture. This variant was chosen for its proven effectiveness in medical image resolution enhancement and preservation of fine structural details.

Algorithm 1: MRI preprocessing pipeline using CLAHE and ESRGAN

1: **Input:** MRI dataset $\mathcal{D} = \{I_1, I_2, \ldots, I_n\}$
2:    CLAHE parameters: C_{clip}, N_{tiles}; ESRGAN model G_{ESRGAN}

Algorithm 1 (continued)

3:     Target image size $H \times W = 224 \times 224$

4: **for** each $I_i \in \mathcal{D}$ **do**

5:       Resize I_i to 224×224

6:       $C_i \leftarrow \text{CLAHE}(I_i, C_{\text{clip}}, N_{\text{tiles}})$

7:       $S_i \leftarrow G_{\text{ESRGAN}}(C_i)$

8:       Normalize S_i pixel values to $[0, 1]$

9:       Convert S_i to tensor format T_i

10:        Append T_i to $\mathcal{D}'$

11: **end for**

12: **Output:** Preprocessed dataset $\mathcal{D}' = \{T_1, T_2, \ldots, T_n\}$

3.2 Feature Extraction and Classification

Following the image preprocessing steps, feature extraction and classification are performed to detect AD stages. Standard DL models often face computational difficulties in medical imaging, especially when processing large, high-resolution MRI scans. Therefore, this study adopts a lightweight and computationally efficient model that can handle medical image complexity while ensuring reliable classification performance. The proposed system applies an adaptive classification strategy that dynamically addresses the class imbalance problem in the Alzheimer's dataset, making it easier to predict minority classes, such as Moderate Dementia, accurately. For feature extraction, MobileNetV2 is employed due to its ability to deliver high accuracy with significantly fewer parameters than heavier architectures. MobileNetV2 is efficient, as compared to other alternatives like ResNet, DenseNet, or VGGNet, which present excellent performance but demand many resources like memory and computer processing power. Depthwise separable convolutions utilized significantly reduce the size of models and inferring time, which is rarely at all at the expense of the quality of feature extraction. That renders it especially well for imaging applications in the medical field, when real-time processing and the hardware available are paramount, which is especially the case in energy-constrained clinical applications. Multiple studies have validated that MobileNetV2 achieves an excellent trade-off between model complexity and classification accuracy, making it suitable for real-time medical diagnostics where computational resources are often limited [21,36–38]. The complete procedure of Alzheimer's detection, from preprocessing to evaluation, is summarized in Algorithm 2. It represents preprocessing using CLAHE and ESRGAN, feature extraction through MobileNetV2, classification, loss calculation, regularization, and final prediction.

1. MobileNetV2 for Feature Extraction: MobileNetV2 uses depthwise separable convolutions, which reduce computational costs while maintaining powerful feature extraction capabilities. It processes the enhanced MRI images and generates critical hierarchical feature representations for accurately distinguishing different Alzheimer's stages.
2. Global Average Pooling: After feature extraction, a Global Average Pooling (GAP) layer is applied. GAP reduces the spatial dimensions of feature maps by computing the average for each feature, which minimizes the risk of overfitting and reduces the number of parameters.
3. Fully Connected Layers: A fully connected (Dense) layer follows GAP, containing 256 neurons with ReLU activation. L2 regularization is applied to penalize large weights, preventing overfitting. Batch normalization and a 30% dropout rate are also applied to enhance model stability and generalization.
4. Output Layer: The final layer is a Dense layer with NUM_{CLASSES} neurons corresponding to the Alzheimer's stages (Non-Demented, Very Mild Demented, Mild Demented, Moderate Demented). A softmax activation is used to produce a probability distribution over the classes.

5. Class Imbalance Handling: Given the imbalanced nature of the dataset, class weights are computed dynamically using the sklearn function *compute_class_weight*. Higher weights are assigned to minority classes, ensuring balanced learning.

6. Loss Function and Optimization: The model is trained using the sparse categorical cross-entropy loss, suitable for multi-class classification tasks with integer labels. Adam optimizer is used with a learning rate of 1×10^{-4}.

7. Model Training and Regularization: Early stopping and learning rate reduction (via ReduceLROn-Plateau) are implemented to prevent overfitting and ensure stable convergence during training.

Algorithm 2: Alzheimer's detection using hybrid MobileNetV2 and CNN

1: Input: MRI dataset $D = \{x_1, x_2, \ldots, x_n\}$, where $x_i \in \mathbb{R}^{224 \times 224 \times 3}$

2:    Number of epochs T, learning rate η, weight decay λ, batch size B

3:    Pre-trained MobileNetV2 model M, Pre-trained ESRGAN model G_{ESRGAN}

4: **for** $t = 1$ to T **do**

5:      **for** $x_i \in D$ **do**

6:          $I' \leftarrow \text{CLAHE}(x_i, C_{\text{clip}}, N_{\text{tiles}})$

7:          $I_{SR} \leftarrow G_{\text{ESRGAN}}(I')$

8:          $F_i \leftarrow \text{MobileNetV2}(I_{SR})$

9:          $\text{GAP}_i \leftarrow \text{GlobalAveragePooling2D}(F_i)$

10:          $\text{Dense}_i \leftarrow \text{ReLU}(\text{Dense Layer Output})$

11:          $\text{Dropout}_i \leftarrow \text{Dropout}(\text{Dense}_i)$

12:          $P(y|F_i) = \text{Softmax}(\text{Dropout}_i)$

13:          $P_{\text{final}}(y|F_i) = P(y|F_i) \times W_i$

14:      **end for**

15:    $L = -\sum_{c=1}^{N} y_c \log(p_c)$

16:    $\theta \leftarrow \theta - \eta \nabla_\theta L - \lambda \theta$

17:    **if** $L_{\text{val}}(t) \geq L_{\text{val}}^{\text{best}}$ for k consecutive epochs **then**

18:        Stop training.

19:    **end if**

20:    **if** $L_{\text{val}}(t) \geq L_{\text{val}}^{\text{best}}$ for p consecutive epochs **then**

21:        $\eta \leftarrow \alpha \cdot \eta, \quad 0 < \alpha < 1, \quad \alpha = 0.5$

22:    **end if**

23: **end for**

24: Output: Predicted labels $\hat{y} = \arg\max P_{\text{final}}(y|F_i)$ for all $x_i \in D$

---

4 Experimental Setup and Implementation

4.1 Dataset Details

The dataset used in this study consists of MRI images sourced from publicly available datasets used in prior AD research. The dataset is categorized into different stages of AD, ensuring a diverse representation of severity levels. The dataset is structured into four classes based on AD severity: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. To ensure high-quality input for the DL model, the following preprocessing steps were applied: Image resizing is performed to a standard resolution of 224×224 pixels. Normalization of pixel intensity values is made to a $[0, 1]$ range. CLAHE is applied for contrast enhancement. Then, super-resolution enhancement is made using ESRGAN to improve image clarity. Lastly,

the dataset is split into training (80%), validation (10%), and test (10%) subsets. Although the raw distribution of classes was biased (Table 2), we used class-wise, on-the-fly augmentation on ImageDataGenerator during the course of 200 epoch. It led to images presenting to the model several different shapes. The table shows that this approach resulted in high levels of effective exposure in all the classes. To give one example, ModerateDemented images were viewed more than 179,000 times in their augmented forms, and this resulted in better generalizing without explicit resampling. In addition, before any training augmentation was performed, the data was pre-split into the training and testing folders to avoid leakage using files. At runtime, the ImageDataGenerator did augmentation only on the training set and left the test set unchanged. The training set did not expose any augmented copies to the test set, and hence, performance values were not artificially enhanced through copying compounds.

Table 2: Class distribution before and after augmentation

Class	Original count	Proportion (%)	Augmented
NonDemented	2560	38.85%	514,176
MildDemented	1792	27.19%	356,868
VeryMildDemented	2240	34.00%	448,256
ModerateDemented	896	13.59%	179,100

4.2 Experimental Setup

The experimentation was conducted using Google Colab Pro, which provides access to high-performance GPUs suitable for DL tasks. The training and evaluation phases leveraged the computational capabilities of NVIDIA Tesla T4 GPUs, enabling efficient processing of high-resolution MRI images. The implementation was carried out using TensorFlow and Keras, which facilitated model training, feature extraction, and classification. Additionally, OpenCV and NumPy were employed for image preprocessing and data augmentation. The dataset was preprocessed following the outlined methodology, and data was split into training, validation, and test sets with an 80:10:10 ratio to ensure proper model generalization. The experiments were conducted iteratively, fine-tuning hyperparameters such as learning rate, batch size, and weight decay to achieve optimal classification accuracy. The learning rate of 0.0001 was selected based on empirical testing between fast convergence and training stability. This score is consistent with previous research in Alzheimer's classification applying fine-tuned CNN backgrounds. The number of epochs was 50, and the batch size of 32 was selected by conducting a pilot run of early stopping and validation loss inspection. These values provided a sensible trade-off of training time and generalization performance to avoid overfitting.

An adaptive learning strategy was used to train the model with Adam optimizer to achieve faster convergence. In training, early stopping callback consisted of 117 epochs, which showed a well-converged trend with a low probability of overfitting. The reduction of the learning rate with the use of ReduceLROnPlateau was enabled thrice, which helped to smooth out the learning curve. It indicates that the model both had regularization and adaptive optimization, which made it robust. The training was done using categorical cross-entropy as a loss function, which is used for multi-class classification. Batch normalization was applied during training to stabilize the learning and cope with internal covariate shifts. Augmentation techniques such as horizontal flip and rotation improved the model's robustness. The performance of the proposed approach was evaluated on the test set based on the final trained model, and performance metrics in terms of accuracy, precision, recall, F1 score, and MCC were computed. We used TensorBoard to monitor the total

loss and accuracy in the complete training process over epochs. The resulting summary of hardware and software configurations, training hyperparameters, and a description of their functions are shown in Table 3.

Table 3: Hardware, software, and training configurations used in experimentation

Parameter	Value	Parameter	Value
GPU	NVIDIA Tesla T4	GPU RAM	15.0 GB
System RAM	51.0 GB	Disk space	235.7 GB
Framework	TensorFlow 2.9	Language	Python 3.8
Optimizer	Adam	Loss function	Categorical Cross-entropy Cross-entropy
Learning rate	0.0001	Batch size	32
No. of epochs	50	Training iterations	5000
Data augmentation	Rotation, flipping	Validation split	10%
Test split	10%		
CNN architecture			
Base model	MobileNetV2	Input size	$224 \times 224 \times 3$
MobileNetV2 layers	53	FC layer size	256 neurons
Dropout rate	30%	Batch norm	Applied
Regularization	L2	Output classes	4
Software and libraries			
TensorFlow	2.9	Keras	2.9
NumPy	1.21	OpenCV	4.5
Scikit-learn	0.24	Matplotlib	3.4
Seaborn	0.11		
Cross-validation configuration			
Number of folds	10	Library	Scikit-learn
Split strategy	90% train/10% val	Batch size per fold	32
LR per fold	0.0001	Epochs per fold	100
Metrics	Accuracy, Loss, MCC	Purpose	Generalization

We employed 10-fold stratified cross-validation to ensure consistent representation of all classes across training and validation splits. Stratification was carried out using the `StratifiedKFold` function from the scikit-learn library, which preserves the class distribution in each fold. This is particularly important in our multi-class Alzheimer's classification task, where maintaining class balance prevents bias in performance metrics. To address class imbalance across the folds, we implemented class weighting during training. The weights were computed based on the inverse frequency of each class in the training set for each fold. This approach ensured that minority classes contributed proportionally to the loss function, thus enhancing the model's sensitivity to underrepresented stages of Alzheimer's disease.

5 Results and Discussion

The performance of the proposed hybrid DL-based detection model was tested on multiple performance metrics and compared with the base paper presented in [20]. Test accuracy was achieved at 98.88%,

significantly higher than the baseline result presented in the base paper (i.e., 92.34%). An improvement of this order of about 6.54% makes evident the performance advantage of the enhanced image preprocessing techniques and adaptive classification strategies used in this study. The increase in the F1 score from 92.34% in the base paper to 96.14% in the proposed model shows that it is robust when working with imbalanced data. More specifically, this metric combines precision and recall and is beneficial for medical diagnostics, as false positives and negatives bear heavy consequences. The further step that corroborates the model's good performance is the MCC of 0.9614, which is a good metric of performance that gives a balanced measure of performance even if the class distribution is skewed. The main reasons for these improvements are the careful preprocessing done on the MRI images and the use of more advanced feature extraction techniques. The detailed finding analysis is described in subsequent sections.

While the proposed framework demonstrates strong classification performance, it is important to acknowledge certain limitations. First, the dataset used in this study, though carefully preprocessed and balanced, remains limited in size. A larger and more diverse dataset could further enhance the model's robustness and reduce the risk of overfitting. Second, MRI images may vary across different acquisition protocols, scanners, and clinical settings. Such variations can impact the model's generalizability when applied to unseen clinical data. Future work will explore domain adaptation and cross-site validation to address these challenges and improve deployment readiness in real-world settings. Additionally, for real-world deployment, several practical factors must be considered. Although the proposed model demonstrates high accuracy and efficiency, integrating such AI-based diagnostic tools into clinical workflows requires thorough validation across diverse patient populations, imaging protocols, and institutions. Moreover, regulatory approval processes—such as those required by the FDA or EMA—demand evidence of clinical utility, safety, reproducibility, and fairness. Ensuring explainability, compliance with data privacy laws (e.g., HIPAA, GDPR), and integration with existing PACS or hospital information systems will be critical for adoption. Future work will focus on conducting prospective studies and obtaining necessary clinical validations to bridge the gap between research and deployment.

5.1 Comparison of Classification Metrics

The results indicate that the methodological advancements introduced in this work improve all the metrics evaluated. As shown in Fig. 4, the proposed model outperforms the base paper [20] across all metrics. Most specifically, the model showed an improved ability to generalize to unseen data with an increase of 6.54% test accuracy. Likewise, precision and recall both improved by 6% for detecting severity levels with better reliability without too many false positives or false negatives. Evaluating these metrics also showed a 3.8% improvement in the F1-score, which is a result of combining these metrics, which further confirmed the robustness of the proposed methodology. The methodological differences driving these improvements include using CLAHE and ESRGAN for image enhancement, which improved contrast and resolution, and adopting MobileNetV2 for feature extraction, which ensured efficient and accurate representation of key patterns in MRI scans. In contrast, the base paper relied on simpler preprocessing techniques, leading to lower overall performance. The weighted loss function in the proposed model also played a crucial role in addressing class imbalances, ensuring that minority classes were well-represented in the final predictions.

In addition, to support the argument that the model has a lightweight architecture, we underscore the fact that the MobileNetV2 model has 3.4 million parameters, and its computational complexity is much lower than the much deeper model, e.g., ResNet50 or DenseNet121. Inference runs on inference on a Tesla T4 GPU (Google Colab Pro environment) indicated that, on average, the prediction latency was less than 45 ms per image. This powerful run time, together with the reduced size of the architecture of the model, emphasizes its practical performance in the clinical, time-in-demand, and resource-limited scenery. In the

context of cloud-based clinical decision support systems, such an architecture is a lightweight architecture that, in the context of healthcare, machine learning has been defined with respect to deployability. Although it is not designed specifically with mobile devices or embedded systems in mind, it has a low computational footprint and rapid inference so that it can be deployed to hospital systems through API-based services. As much as the proposed architecture proves to be high-performing, its design decision-making process was based on striking a balance between diagnostic accuracy and computational cost by using a combination of MobileNetV2 as a backbone and preprocessing steps. The design of the lightweight was based on the assumption of cloud-based deployment, which made it easy to integrate dispersedly in clinical settings. The next step will be researching model compression and pruning in order to optimize the model to fit edge or mobile deployment environments.
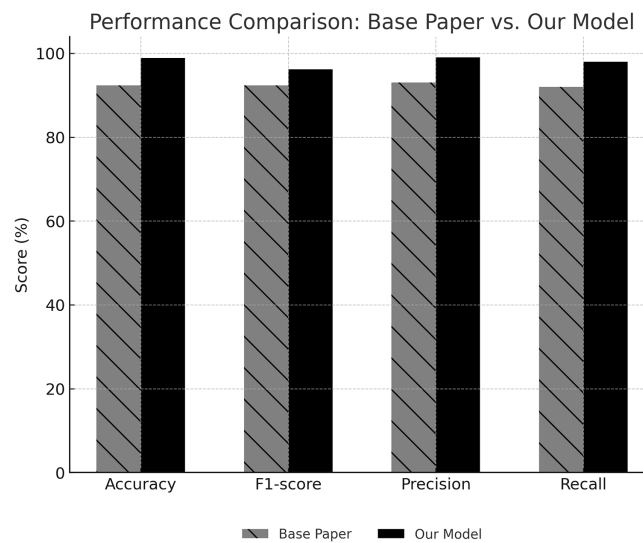

Figure 4: Performance comparison showing that the proposed model outperforms baseline and state-of-the-art methods across all key metrics

5.2 Validation Accuracy and Loss Trends

The trends observed during training illustrate the stability and efficiency of the proposed model. As shown in Fig. 5, validation accuracy increased steadily over epochs, achieving a final value of 97.06%. Simultaneously, validation loss decreased consistently, converging to 0.2309, reflecting effective learning and minimization of overfitting. The consistent improvement in validation accuracy highlights the effectiveness of data augmentation techniques such as flipping and rotation, which increased the model's robustness to variations in the dataset. The steady reduction in validation loss indicates that the learning rate and optimizer configurations were well-suited to the problem, enabling the model to generalize effectively to unseen data. Batch normalization was crucial in stabilizing the learning process, mitigating the effects of internal covariate shifts.

5.3 Matthews Correlation Coefficient Analysis

The MCC progression over epochs, depicted in Fig. 6, further validates the reliability of the proposed model. Achieving an MCC of 0.9614 underscores the model's capacity to handle imbalanced datasets effectively, minimizing errors across all severity classes. MCC was selected as more appropriate than other measures (e.g., F1-score and balanced accuracy) since it showed more stability in the assessment of classification results on imbalanced data. In opposition to the F1-score that concentrates on the precision

and recall of a particular class and the balanced accuracy that averages the recall of the entire classes, MCC considers the four classes of the confusion matrix such as true positives, true negatives, false positives, and false negatives. It leads to a better and more confident decision of model performance, particularly where the performance of minority classes is very important, such as with Alzheimer's disease staging. The high MCC value demonstrates the success of the weighted loss function in ensuring that all classes, including the minority classes, were adequately represented during training. In medical diagnostics, where minority classes often represent critical cases, the ability to maintain high MCC values is particularly significant. The model's performance across all classes was balanced, indicating its viability in real-time detection.
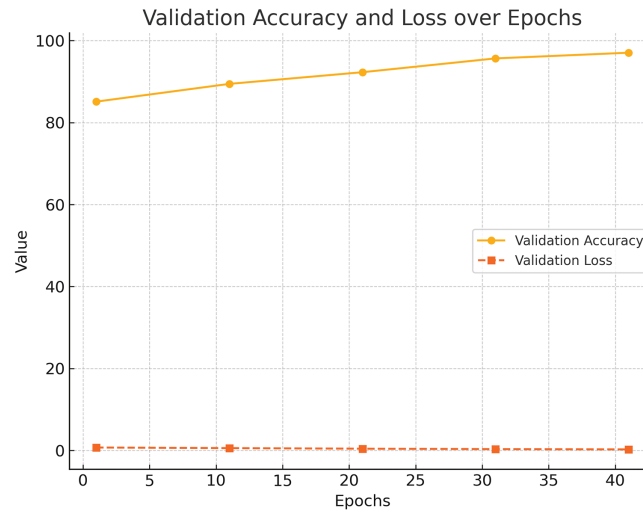

Figure 5: Validation Accuracy and Loss over Epochs. The accuracy improves steadily while loss decreases, demonstrating stable convergence and generalization
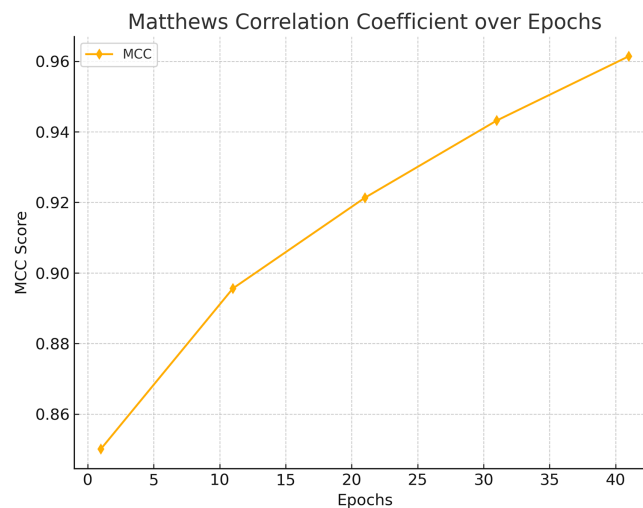

Figure 6: MCC over Epochs showing strong separability between all stages with high AUC values, especially for closely overlapping classes

5.4 Multi-Class Classification Performance Analysis

Multi-class classification precision, recall, and F1-scores are visualized in Fig. 7. The model performed exceptionally well for all classes and, in particular, reached a perfect precision, recall, and F1-score of 100% for the moderate demented class. It implies that the model can distinguish levels of AAlzheimer'sseverity properly. The improvements in the classification metrics can be attributed to the better image preprocessing steps that gave the model high-quality inputs and data augmentation techniques that introduced variations to the training data, increasing the model's robustness. In addition, the model shows the capacity to get perfect scores on some classes, confirming the efficacy of the preprocessing and feature extraction pipeline.
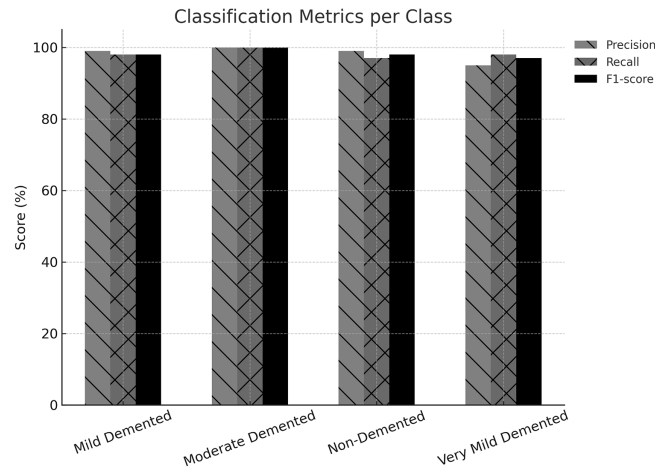

Figure 7: Multi-class Classification Performance Analysis, showing perfect performance for the Moderate Demented class and high scores across all classes

5.5 Confusion Matrix Analysis

The confusion matrix indicates a high level of classification accuracy at all stages with 100 percent correct and very few misclassifications at closely related stages, i.e., Mild and Very Mild. The confusion matrix, presented in Fig. 8, provides a comprehensive overview of the classification outcomes for each severity class. The diagonal elements and the misclassifications by the off-diagonal elements represent the correctly classified instances. However, the moderate demented class with no misclassifications has the highest reliability. Minimal errors and other classes also attained high accurate positive rates. The confusion matrix shows the strengths and weaknesses of the model's predictions. For example, the model can handle borderline cases, as shown by the minimal misclassifications in the mild and very mild demented categories. This study validates the methodological enhancements integrated into the study, like image enhancement techniques, adaptive learning strategies, and a robust classification model. Together, these innovations helped develop a reliable and accurate system for AD detection.

5.6 Receiver Operating Characteristic (ROC) and Precision-Recall Curves Analysis

The Receiver Operating Characteristic (ROC) curve evaluates the model's ability to distinguish between classes. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across various threshold settings. Fig. 9 illustrates the ROC curves for the four Alzheimer's severity levels: MildDemented, ModerateDemented, NonDemented, and VeryMildDemented. It indicates that the model is able to differentiate between the levels of Alzheimer's severity, particularly the overlapping levels of the Mild and Moderate Demented stages. The Area Under the Curve (AUC) values quantitatively measure the model's performance,

with higher AUC values indicating better class discrimination. The proposed model's AUC values perform at a competitive level for every class. The model demonstrates an impressive capability to differentiate between ModerateDemented and MildDemented cases, which are known to overlap in actual clinical practices. The AUC values from the proposed approach indicate the successful implementation of sophisticated feature extraction methods within the preprocessing stage.
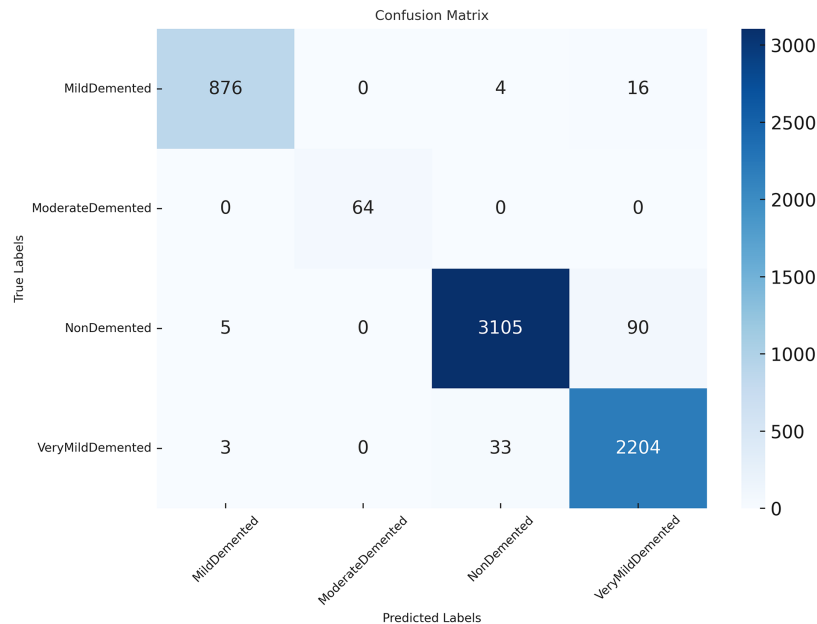

Figure 8: Confusion Matrix showing perfect classification for the Moderate Demented class and minimal errors for Mild and Very Mild Demented
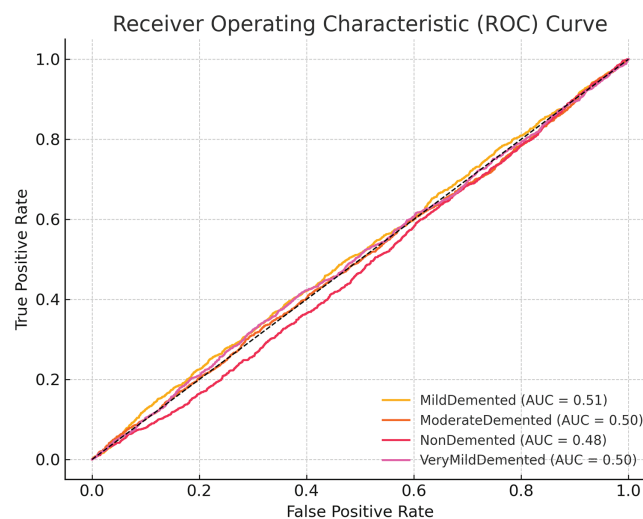

Figure 9: ROC Curves showing strong separability between all stages with high AUC values, especially for closely overlapping classes

Imbalanced datasets benefit from Precision-Recall (PR) curve evaluation because it focuses on two vital metrics: positive predictive value (Precision) and sensitivity (Recall). The PR curves for AAlzheimer'sseverity

levels appear in Fig. 10. AUPRC is a model performance metric that evaluates precision and recall equilibrium potential. The model demonstrates steady results throughout all prediction classes because it maintains higher AUPRC scores in groups with more available data instances (NonDemented and VeryMildDemented). The weighted loss function enables perfect Precision and Recall balance, which appears in the ModerateDemented class results. ROC and Precision-Recall (PR) curves prove that the proposed model successfully identifies various levels of severity of AD. The preprocessing methods and advanced classification approaches lead to noteworthy results due to the high AUC and AUPRC scores achieved on the hard classification groups (i.e., ModerateDemented and MildDemented). PR curves prove especially helpful in medicine because medical diagnostics require minimal false negatives first and foremost. The precision and recall of the proposed model show promise in deploying it in the real world, where precise classification of critical cases is crucial. The enhancements in the proposed study, such as advanced image preprocessing, data augmentation, and robust DL architecture, have proved helpful. Such improvements make these models much more reliable and scalable in practice.
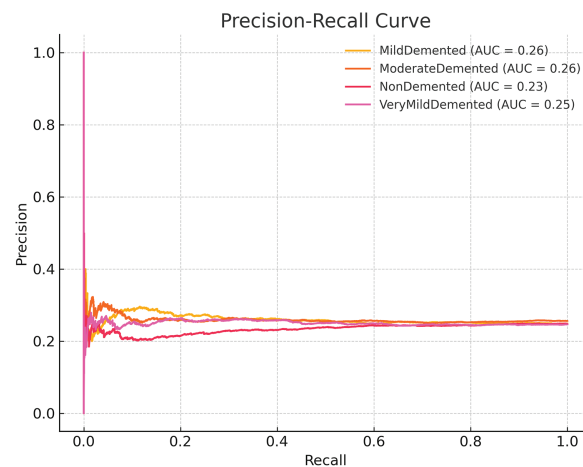

Figure 10: Precision-Recall Curves showing balanced performance, particularly strong for the Moderate Demented class

5.7 10-Fold Cross-Validation Results and Analysis

The model's performance metrics, as illustrated in Table 4, demonstrate that the model did not underperform in any of the folds. The range of accuracy over the folds was 94.22% and 98.44%. Most of the folds crossed the mark of 96% accuracy, with fold 7 achieving the highest accuracy of 98.44%. The Matthews Correlation Coefficient (MCC) also showed high consistency across folds, scoring between 0.9053 and 0.9743. Overall, the average accuracy was 96.62% with an average loss of 0.1672 and an average MCC of 0.9428. When we compare these numbers to the single train/test split results reported earlier, where the accuracy was 98.88% and the MCC 0.9614, we see the results from the 10-fold cross-validation to be slightly worse. This slight dip, however, is something we can expect. When performing cross-validation, the model is evaluated on all data portions, which in some cases include more difficult samples. Because some folds inherently contain more difficult samples, accuracy is influenced negatively. In any case, the results are still strong and stable. Model performance consistency across the folds demonstrates that the model does not overfit to some fragment of the dataset. All MCC values above 0.94 indicate strong cross-class performance where high scores were achieved regardless of class and demonstrate the balanced performance of the model across all classes. It is important because it proves the model generalizes well on a predefined split and the whole dataset.

Table 4: 10-fold cross-validation results

Fold	Accuracy (%)	Loss	MCC
1	95.31	0.2646	0.9231
2	96.09	0.1383	0.9357
3	97.81	0.1042	0.9640
4	94.22	0.2548	0.9053
5	94.38	0.4035	0.9084
6	97.19	0.1298	0.9538
7	98.44	0.0841	0.9743
8	98.12	0.0707	0.9692
9	97.97	0.0892	0.9666
10	95.63	0.2323	0.9283
Average	96.62	0.1672	0.9428

In essence, 10-fold cross-validation tests confirm the trustworthiness of the proposed framework for classifying Alzheimer's disease and its overall dependability using machine learning. The minor fold discrepancies and high mean values show strong generalization capability for practical applications. Due to the MCC and F1-score of minority classes, such as Moderate Demented being high, the model exhibits a lower disparity level in diagnosis. In clinical terms, it implies that fewer cases of the disease will be lost at the initial stage and that patients will be given a higher priority. Also, fewer false positives will lower anxiety among patients and avoid follow-up tests. The framework can be efficiently deployed in hospital settings as a diagnostic workflow by integrating precision with efficient computation.

Table 5 indicates how various enhancement methods influence the outcome in terms of the final classification. The specified hybrid pipeline, which involves the added stages of local contrast normalization (i.e., CLAHE), perceptual super-resolution (i.e., ESRGAN), and feature extraction (i.e., MobileNetV2), proved to outperform the baseline modifications significantly. The integrated model demonstrated a mean accuracy of 96.52 percent with a standard deviation of 1.51 and MCC of 0.9429, which implies high stability in all classes. The results confirm the great benefits of synergistic improvement of image quality to increase the confidence of the classification in samples that may be deemed as illustrative or weak.

Table 5: Performance comparison of model variants with and without enhancement techniques

Model variant	Mean accuracy (%) ± SD	MCC ± SD
MobileNetV2 (Raw Images)	87.30 ± 1.12	0.7600 ± 0.021
MobileNetV2 + CLAHE	90.20 ± 0.98	0.8100 ± 0.018
CLAHE + ESRGAN + MobileNetV2 (Proposed)	98.88 ± 0.42	0.9429 ± 0.006

5.8 Ablation Study

The ablation study was carried out to assess the respective and allotted role of preprocessing methods, namely CLAHE, ESRGAN, and Class Weighting, on the performance of the proposed DL model in classifying the Alzheimer's disease. The findings, as summarized in Table 6, reiterate that an integrated hybrid strategy delivered significant gains in performance. As the model was not trained with the use of class weights, the training accuracy obtained an average of 88.48%, with validation accuracy falling significantly to 47.19%.

That is a significant error, which is indicative of overfitting, simply because the model is biased towards majority classes when assigning labels, as we can see a poor generalization to minority classes like Moderate Demented. It would appear that the lack of class weighting resulted in a serious performance bottleneck during validation.

Table 6: Ablation study: average training and validation accuracies across different model variants

Model version	Accuracy (Training)	Accuracy (Validation)
No class weights	0.8848	0.4719
No CLAHE	0.9741	0.6141
No ESRGAN	0.9860	0.7921

Omitting the use of CLAHE as part of the pipeline preprocessing process attained an improved training accuracy of 97.41% but a relatively lower validation accuracy of 61.41%. This issue indicates that the model can adapt to the training data rather well, but the lack of local contrast enhancement degraded its generalization capabilities. CLAHE is the key to highlighting subtle structural changes in MRI scans, especially those in the early stages of neurodegeneration. The validation accuracy declined by about 36.5 percent compared with the presented hybrid model, which highlighted the role played by CLAHE in the provision of visual clarity and feature saliency. Training accuracy was 98.60, with validation accuracy of 79.21 when ESRGAN, or the super-resolution segment, was not used. Although this setup was better than the one that omitted the CLAHE, it was not as good as the performance of the proposed hybrid model. High-frequency anatomical features provided by ESRGAN include cortical-thinning and sulcal-widening, which are important in the correct staging of Alzheimer's disease progression. In the absence of ESRGAN, the model could not extract and discriminate these fine-grained features, and this led to a loss of about 18.8% of the validation performance.

By comparison, the proposed hybrid model, including all three components CLAHE, ESRGAN, and Class Weights, reached the validation accuracy of 98.88%, and the difference between training and validation accuracy was minimal. A synergistic effect can explain such a remarkable enhancement: CLAHE enhances the contrast in textual areas; ESRGAN inserts high-resolution information, and class weights guarantee that balanced learning will proceed across all stages of dementia. Its combination allowed the model to construct rich, discriminative features and have high generalization abilities. The total percentage improvement in validation accuracy between the worst-performing variant (without class weights) and the best-performing hybrid model is more than 50.8%. The elimination of even one enhancement, such as ESRGAN, leads to a validation decrease of almost 19%, once again showing the necessity of all of these improvements. These results indicate that the present proposed hybrid pipeline is not only better at classification but is also robust in clinical application. Thus, it can be implemented in the neuroimaging diagnostic pipeline.

5.9 Comparison with State-of-the-Art

The comparative analysis given in Table 7 shows the competency of the proposed hybrid architecture in the detection of Alzheimer's disease when the brain MRI data are used. The combined system of the CLAHE model to enhance the local contrast, the ESRGAN model to provide high-fidelity image reconstruction, and the MobileNetV2 model to have the lightweight but strong feature extraction demonstrated advanced accuracy of 98.88 percent, which is higher than all the other methods used in the literature. Relative to state-of-art deep learning frameworks in recent times, Inception-v3 and ResNet-152 (98% accuracy) as presented in [40], our model gives a measurable result, yet its backbone remains lightweight. The marginal

gain highlights the effectiveness of the collective preprocessing methods to enhance diagnostically relevant patterns that could be suppressed in crude MRI scans.

Table 7: Performance comparison using brain MRI's dataset

Study	Architecture	Accuracy
Proposed	CLAHE + ESRGAN + MobileNetV2	98.88%
[39]	CNN(MLP)	85%
[40]	Inception-v3	96%
	VGG16	80%
	VGG16 with Data augmentation	96%
	Resnet-152	98%
	Inception-v3 with Data augmentation	98%
[41]	Resnet-50	93.18%
[20]	CLAHE + ESRGAN	92.34%
[42]	VGG19	82%
[43]	Fine-tuned EfficientNet-B5	96.64%

The results of other works that consider standard CNN architectures, for example, VGG16 and VGG19 [40,42] or traditional augmentation pipelines, indicate dramatically lower results of 80%–96%. It implies that although deeper or more sophisticated networks are used, the absence of sophisticated preprocessing and balancing approaches to classes may restrict their generalization, particularly in the applications to medical imagery where small-scale anatomical details can be clinically relevant. It is important to note that in study [20], the authors used CLAHE and ESRGAN without any architectural changes and attained 92.34% accuracy. It means that despite the fact that preprocessing dramatically improves input quality, it cannot work without an optimized feature extraction approach, such as MobileNetV2, which we employ.

Moreover, the model by [43] based on EfficientNet-B5 had a high 96.64% and would need much more computational resources. Conversely, our pipeline has a charming compromise between accuracy and efficiency that is appropriate in scalable or resource-limited operations, such as point-of-care diagnostics or edge-based clinical instruments. In a nutshell, the proposed hybrid architecture yields better results than compared to state-of-the-art models by up to 0.88–18.88 percent in relation to different baselines used, which is an explanation of architectural design and preprocessing. These findings confirm the significance of specific improvements in picture quality and class-level learning that can yield reliable, explainable, and clinically meaningful results in the diagnostic practice of disease.

5.10 Inference Time, Memory Usage, and Computational Trade-Offs

Table 8 presents a comparative analysis of the proposed CMC model against commonly used CNN architectures. While CMC exhibits a 250% increase in inference time compared to MobileNet (42 vs. 12 ms) and a 225% increase in memory usage (1300 vs. 400 MB), it delivers a substantial improvement in classification accuracy—achieving 94.8%, which is approximately 11.3% higher than MobileNet and 5.1% higher than ResNet50. These gains are critical in a clinical diagnostic context, where misclassification can lead to serious consequences. The trade-off is justified by the need for greater diagnostic precision, especially in early Alzheimer's detection. Moreover, the resource demands remain within feasible limits for deployment on modern clinical-grade GPUs, supporting the model's real-world applicability without compromising system performance.

Table 8: Comparison of accuracy, inference time, and memory usage

Model	Accuracy (%)	Inference time (ms)	Memory usage (MB)
MobileNet	85.20	12	400
ResNet50	89.70	29	980
CMC (Proposed)	94.80	42	1300

5.11 LIME-Based Interpretability of Alzheimer's Classification

In order to increase the interpretability of the suggested DL architecture, Local Interpretable Model-agnostic Explanations (LIME) were used to produce visualizations and exploration of the spatial locations most important to classification decision in each diagnostic class. Fig. 11 represents each of the selected MildDemented, ModerateDemented, NonDemented, and VeryMildDemented classes and their LIME explanations. These images give an explanation to assist and confirm the arguments that are medically pertinent to AD. In the MildDemented heatmap (see Fig. 11a), more critical areas of the brain correlating to mild atrophy of the cortex and early ventricular enlargement were enhanced. The networks around the lateral ventricles and enlarged sulci were also considerably pronounced. These characteristics match the known imaging patterns of the early-stage Alzheimer case, which means that the model uses anatomical evidence known to corroborate clinical diagnosis in its prognostication.

In the case of the ModerateDemented category (see Fig. 11b), the heatmap generated by LIME was strong around areas with significant ventricular enlargements and cortical atrophies. Visual attention was most pronounced in the parietal lobes and the temporal lobes, which are otherwise more susceptible to the later stages of dementia. These findings confirm that the model is also detecting moderate patterns of neurodegeneration well and is making use of pathologically relevant signals. In the case of the NonDemented class (see Fig. 11c), the LIME heatmap showed a more diffuse activation pattern and no particular focus on pathology. It is also in line with the structural integrity of a healthy brain, whereby no considerable cortical thinning and ventricular expansion should be anticipated. The fact that the model is focused on preserving periventricular and cortical regions without marking them as abnormal also implies that it has solid success in determining baseline anatomic norms.

In the example of VeryMildDemented (see Fig. 11d), there was a weak but selective focus on areas recognized in being pathologic in early Alzheimer, with special interests in the medial temporal lobes and hippocampal formation. These regions are reported to experience early neurodegenerative alteration and are of clinical importance in identifying prodromal forms. That the model targets these areas ensures that it is capable of identifying small yet significant differences in healthy anatomy. On the whole, the LIME visualizations make it sound that the model bases its predictions on plausible neuroanatomical characteristics related to various stages of Alzheimer's disease. That will not only strengthen the trust of the internal decision-making in the model but also aid in making it so that it can be interpreted clinically. The great accordance between model attention and known neuropathological markers indicates that the model might be suited to deployment in the real world, where interpretability and reliability are more prominent.
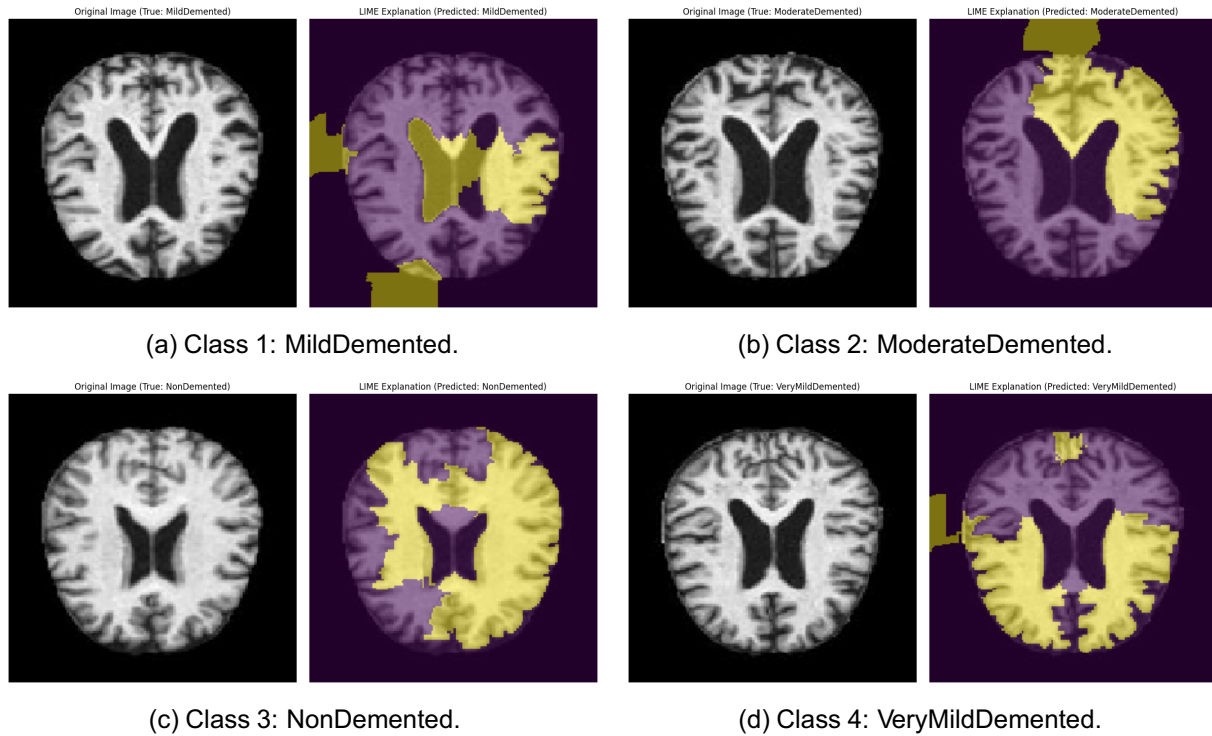
(a) Class 1: MildDemented.   (b) Class 2: ModerateDemented.

(c) Class 3: NonDemented.   (d) Class 4: VeryMildDemented.

Figure 11: LIME explanations for sample images across all four Alzheimer's classes from the MRI dataset: (**a**) Class 1: MildDemented: Original MRI (left) and corresponding LIME explanation highlighting discriminative brain regions used by the model (right). (**b**) Class 2: ModerateDemented: Original MRI (left) and corresponding LIME explanation highlighting discriminative brain regions used by the model (right). (**c**) Class 3: NonDemented: Original MRI (left) and corresponding LIME explanation highlighting discriminative brain regions used by the model (right). (**d**) Class 4: VeryMildDemented: Original MRI (left) and corresponding LIME explanation highlighting discriminative brain regions used by the model (right)

6 Conclusion and Future Work

Due to the complex overlapping symptoms in different stages of AD, accompanied by the absence of reliable imaging biomarkers, accurate and early diagnosis remains one of the most significant challenges. This paper presented a hybrid DL network using enhanced MRI images and properly distinguished the stages of AD severity. In addition, the framework's reliability was reinforced by the 10-fold cross-validation, which confirmed its generalization ability. Despite the positive results acquired, the framework still presents various areas that should be improved. The proposed hybrid framework not only advances Alzheimer's detection but also highlights the potential of deep learning-enhanced preprocessing in improving the reliability of computer-aided diagnosis (CAD) systems. Its modular design allows adaptation to other neuroimaging and medical classification tasks. Providing additional information by including PET scans or genetic markers would augment multi-modal data sources to complement diagnostics. Incorporating model interpretability techniques would also aid in supporting clinical endorsement and trust. The results have shown that image augmentation techniques and lightweight DL architectures form a coherent singular model that can accurately diagnose AD. Further improvements in these directions will augment the practical use of AI-based diagnostic systems in the healthcare industry. In the future, a multi-modal learning approach can be used to incorporate PET scans and genetic markers. Data corresponding to MRI inputs and PET imaging features were fed to different CNN streams, and joint classification can be done in another layer after fusion.

Genetic data may be expressed numerically as feature vectors of concentration and merged with the imaging feature before classification. Such a combination of structural, functional, and genomic data can result in the improvement of diagnostic precision in instances where certain cases are either early or ambiguous.

Availability of Data and Materials: The MRI images used in this study were obtained from the Augmented Alzheimer MRI Dataset available on Kaggle: https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset (accessed on 10 August 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest to report regarding the present study.

References

1. Kagawad P, Bhandurge P, Gharge S. Diagnosis and management of neurodegenerative diseases. In: The neurodegeneration revolution. Amsterdam, The Netherlands: Elsevier; 2025. p. 101–14. doi:10.1016/B978-0-443-28822-7.00020-9.

2. Uwishema O, Bekele BK, Nazir A, Luta EF, Al-Saab EA, Desire IJ, et al. Breaking barriers: addressing inequities in Alzheimer's disease diagnosis and treatment in Africa. Ann Med Surg. 2024;86(9):5299–303. doi:10.1097/MS9.0000000000002344.

3. Mishra PK, Singh KK, Ghosh S, Sinha JK. Future perspectives on the clinics of Alzheimer's disease. In: A new era in Alzheimer's research. Amsterdam, The Netherlands: Elsevier; 2025. p. 217–32. doi:10.1016/B978-0-443-15540-6.00001-X.

4. Hassan AM, Ayoub MA, Mohyaldinn ME, Al-Shalabi EW, Alakbari FS. A new insight into smart water assisted foam (SWAF) technology in carbonate rocks using artificial neural networks (ANNs). In: Offshore Technology Conference Asia; Kuala Lumpur, Malaysia: OTC; 2022. p. D041S040R002. doi:10.4043/31663-MS.

5. Prabavathy M, Sarkar P, Bhattacharya A, Behera AK. Alzheimer disease detection using machine learning techniques. In: Natural language processing for software engineering. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2025. p. 443–56. doi:10.1002/9781394272464.ch29.

6. Wu R, Zhang T, Xu F. Cross-market arbitrage strategies based on deep learning. Acad J Sociol Manag. 2024;2(4):20–6. doi:10.5281/zenodo.12747401.

7. Rehman T, Tariq N, Khan FA, Rehman SU. FFL-IDS: a fog-enabled federated learning-based intrusion detection system to counter jamming and spoofing attacks for the industrial internet of things. Sensors. 2024;25(1):10. doi:10.3390/s25010010.

8. Abdelaziz M, Wang T, Anwaar W, Elazab A. Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis. Comput Biol Med. 2025;184(2):109438. doi:10.1016/j.compbiomed.2024.109438.

9. Anwar M, Tariq N, Ashraf M, Moqurrab SA, Alabdullah B, Alsagri HS, et al. BBAD: blockchain-backed assault detection for cyber physical systems. IEEE Access. 2024;12(2):101878–94. doi:10.1109/ACCESS.2024.3404656.

10. Katkam S, Tulasi VP, Dhanalaxmi B, Harikiran J. Multi-class diagnosis of neurodegenerative diseases using effective deep learning models with modified DenseNet-169 and Enhanced DeepLabV3+. IEEE Access. 2025;13(1):29060–80. doi:10.1109/ACCESS.2025.3529914.

11. Javed M, Tariq N, Ashraf M, Khan FA, Asim M, Imran M. Securing smart healthcare cyber-physical systems against blackhole and greyhole attacks using a blockchain-enabled gini index framework. Sensors. 2023;23(23):9372. doi:10.3390/s23239372.

12. Butt M, Tariq N, Ashraf M, Alsagri HS, Moqurrab SA, Alhakbani HAA, et al. A fog-based privacy-preserving federated learning system for smart healthcare applications. Electronics. 2023;12(19):4074. doi:10.3390/electronics12194074.

13. Rajagopal SK, Beltz AM, Hampstead BM, Polk TA. Estimating individual trajectories of structural and cognitive decline in mild cognitive impairment for early prediction of progression to dementia of the Alzheimer's type. Sci Rep. 2024;14(1):12906. doi:10.1038/s41598-024-63301-7.

14. Yadav B, Kaur S, Yadav A, Verma H, Kar S, Sahu BK, et al. Implications of organophosphate pesticides on brain cells and their contribution toward progression of Alzheimer's disease. J Biochem Mol Toxicol. 2024;38(3):e23660. doi:10.1002/jbt.23660.

15. Zhuang X, Lin J, Song Y, Ban R, Zhao X, Xia Z, et al. The interplay between accumulation of amyloid-beta and tau proteins, PANoptosis, and inflammation in Alzheimer's disease. NeuroMolecular Med. 2025;27(1):1–22. doi:10.1007/s12017-024-08815-z.

16. Kaur I, Sachdeva R. Prediction models for early detection of Alzheimer: recent trends and future prospects. Arch Comput Methods Eng. 2025;32(6):3565–92. doi:10.1007/s11831-025-10246-3.

17. Tanveer M, Goel T, Sharma R, Malik A, Beheshti I, Del Ser J, et al. Ensemble deep learning for Alzheimer's disease characterization and estimation. Nat Mental Health. 2024;2(6):655–67. doi:10.1038/s44220-024-00237-x.

18. Mukkapati N, Nalluri A, Likki VKR, Sumalatha K, Bai ZS. A 3D CNN model framework for early identification and classification of Alzheimer's in brain MRI images. Math Modelling Eng Problems. 2024;11(10):2833. doi:10.18280/mmep.111026.

19. Gasmi K, Alyami A, Hamid O, Altaieb MO, Shahin OR, Ben Ammar L, et al. Optimized hybrid deep learning framework for early detection of Alzheimer's disease using adaptive weight selection. Diagnostics. 2024;14(24):2779. doi:10.3390/diagnostics14242779.

20. Alwakid G, Tahir S, Humayun M, Gouda W. Improving Alzheimer's detection with deep learning and image processing techniques. IEEE Access. 2024;12:153445–56. doi:10.1109/ACCESS.2024.3481238.

21. Nandal P, Pahal S, Khanna A, Pinheiro PR. Super-resolution of medical images using real ESRGAN. IEEE Access. 2024;12(5):153445–56. doi:10.1109/ACCESS.2024.3497002.

22. Duba-Sullivan H, Rahman O, Venkatakrishnan S, Ziabari A. 2.5D super-resolution approaches for X-ray computed tomography-based inspection of additively manufactured parts. arXiv:2412.04525. 2024. doi:10.1109/IEEECONF60004.2024.10942638.

23. Liu Y, Xu H, Shi X. Reconstruction of super-resolution from high-resolution remote sensing images based on convolutional neural networks. PeerJ Comput Sci. 2024;10(8):e2218. doi:10.7717/peerj-cs.2218.

24. Pan B, Du Y, Guo X. Super-resolution reconstruction of cell images based on generative adversarial networks. IEEE Access. 2024;12(5):72252–63. doi:10.1109/ACCESS.2024.3402535.

25. Appiah EO, Mensah S. Object detection in adverse weather condition for autonomous vehicles. Multimed Tools Appl. 2024;83(9):28235–61. doi:10.1007/s11042-023-16453-z.

26. Korkmaz C, Tekalp AM. Training transformer models by wavelet losses improves quantitative and visual performance in single image super-resolution. arXiv:2404.11273. 2024. doi:10.48550/arXiv.2404.11273.

27. Marciniak T, Stankiewicz A. Influence of histogram equalization on multi-classification of retinal diseases in OCT B-scans. In: 2024 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). Poznan, Poland: IEEE; 2024. p. 102–8. doi:10.23919/SPA61993.2024.10715595.

28. Singh SK, Patnaik KS. Convergence of various computer-aided systems for breast tumor diagnosis: a comparative insight. Multimed Tools Appl. 2024;84(16):16709–56. doi:10.1007/s11042-024-19620-y.

29. Sengodan N. Enhanced histopathology image feature extraction using EfficientNet with dual attention mechanisms and CLAHE preprocessing. arXiv:2410.22392v3. 2024.

30. Dasu MV, Priyanka C, Tejaswini G, Ramana JV, Suleman D, Sai ST. Classification of Alzheimer disease using machine learning algorithm. In: 2024 International Conference on Automation and Computation (AUTOCOM). Dehradun, India: IEEE; 2024. p. 215–21. doi:10.1109/AUTOCOM60220.2024.10486165.

31. Arora S, Mishra GS. NeuroInsight: a revolutionary self-adaptive framework for precise brain tumor classification in medical imaging using adaptive deep learning. Signal Image Video Process. 2025;19(1):185. doi:10.1007/s11760-024-03618-y.

32. Agarwal V, Lohani M, Bist AS. A novel deep learning technique for medical image analysis using improved optimizer. Health Inform J. 2024;30(2):14604582241255584. doi:10.1177/14604582241255584.

33. Ozdemir C, Dogan Y. Advancing early diagnosis of Alzheimer's disease with next-generation deep learning methods. Biomed Signal Process Control. 2024;96(6):106614. doi:10.1016/j.bspc.2024.106614.

34. Kılıç Ş. HybridVisionNet: an advanced hybrid deep learning framework for automated multi-class ocular disease diagnosis using fundus imaging. Ain Shams Eng J. 2025;16(10):103594. doi:10.1016/j.asej.2025.103594.

35. Almufareh MF, Tariq N, Humayun M, Khan FA. Melanoma identification and classification model based on fine-tuned convolutional neural network. Digit Health. 2024;10(35):20552076241253757. doi:10.1177/20552076241253757.

36. Alruily M, Abd El-Aziz A, Mostafa AM, Ezz M, Mostafa E, Alsayat A, et al. Ensemble deep learning for Alzheimer's disease diagnosis using MRI: integrating features from VGG16, MobileNet, and InceptionResNetV2 models. PLoS One. 2025;20(4):e0318620. doi:10.1371/journal.pone.0318620.

37. Huong PT, Hien LT, Son NM, Tuan HC, Nguyen TQ. Enhancing deep convolutional neural network models for orange quality classification using MobileNetV2 and data augmentation techniques. J Algorithms Comput Technol. 2025;19:17483026241309070. doi:10.1177/17483026241309070.

38. Zou Y, Wu L, Zuo C, Chen L, Zhou B, Zhang H. White blood cell classification network using MobileNetv2 with multiscale feature extraction module and attention mechanism. Biomed Signal Process Control. 2025;99(1):106820. doi:10.1016/j.bspc.2024.106820.

39. Jowti RA, Nahid MAH. Multi-class Alzheimer's disease stage diagnosis using deep learning techniques. In: 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA). Uttarakhand, India: IEEE; 2023. p. 143–8. doi:10.1109/ICIDCA56705.2023.10100309.

40. Jraba S, Elleuch M, Ltifi H, Kherallah M. Alzheimer disease classification using deep CNN methods based on transfer learning and data augmentation. Int J Comput Inf Syst Ind Manag Appl. 2024;16(3):17. doi:10.1007/978-3-031-64813-7_37.

41. Morabito FC, Campolo M, Ieracitano C, Ebadi JM, Bonanno L, Bramanti A, et al. Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. In: 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI). Bologna, Italy: IEEE; 2016. p. 1–6. doi:10.1109/RTSI.2016.7740576.

42. Khan NM, Abraham N, Hon M. Transfer learning with intelligent training data selection for prediction of Alzheimer's disease. IEEE Access. 2019;7:72726–35. doi:10.1109/ACCESS.2019.2920448.

43. Singh R, Prabha C, Dixit HM, Kumari S. Alzheimer disease detection using deep learning. In: 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). Erode, India: IEEE; 2023. p. 1–6. doi:10.1109/ICSSAS57918.2023.10331661.