Alzheimer's Disease Classification using Lightweight Network MobileNet-V3

M Sadewa Wicaksana Wibowo [1)]*, M Cahyo Kriswantoro[2)], Neni Purwati [3)]

[1,2,3] Department of Faculty of Health Science, University of Muhammadiyah Lamongan
* Corresponding author : sadewawicaksana@umla.ac.id

Abstract

Alzheimer's disease is a major public health concern characterized by progressive cognitive decline due to irreversible neuronal damage. Alzheimer's disease represents a major global health concern in the twenty-first century. Although magnetic resonance imaging (MRI) is widely used for early diagnosis, manual interpretation is time-consuming and subject to variability. This study proposes an automated classification system based on the lightweight MobileNetV3 architecture to improve diagnostic efficiency. The model leverages depthwise separable convolutions to reduce computational complexity while maintaining high performance. MobileNetV3 models is then evaluated using appropriate metrics to assess the effectiveness of the proposed classification approach. Data augmentation techniques, including random rotation and flipping, are applied to enhance model generalization. Experimental results demonstrate that the MobileNetV3 Small model achieves superior performance, with an accuracy and F1-score of approximately 0.94, compared to 0.90 for the MobileNetV3 Large model. These findings indicate that the compact architecture provides better efficiency and reliability for Alzheimer's disease classification. The proposed approach is suitable for deployment in resource-constrained medical environments.

Keywords- Alzheimer's Disease, Classification, Deep Learning, Convolutional Neural Networks

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by a gradual decline in cognitive function due to irreversible neuronal damage. The earliest symptom typically involves memory loss related to recent events, followed by language impairment, disorientation, mood changes, reduced motivation, and behavioral disturbances as the disease progresses. As the most common form of dementia, AD represents a major global health concern in the twenty-first century. Early symptoms are often misinterpreted as normal aging, making accurate diagnosis challenging. Although definitive diagnosis requires postmortem brain examination [1], early detection remains critical due to the progressive nature of the disease. Globally, AD affects approximately 24 million individuals, including over 5 million in the United States [2]. In Indonesia, the prevalence of dementia among individuals aged 60 years and older varies considerably. Alzheimer's Disease International estimated that 1.2 million Indonesians (approximately 4% of the elderly population) were living with dementia in 2015, while other studies report prevalence rates of up to 9% or higher [3], [4]. These figures exceed global averages and highlight the need for further investigation [5]. In addition to its health impact, AD imposes a significant economic burden. Global healthcare costs are projected to increase from USD 226 billion in 2015 to USD 1.1 trillion by 2050 [6]. Various diagnostic techniques have been developed for AD detection. However, many remain limited in sensitivity, particularly for early or moderate cognitive impairment [7]. Advanced neuroimaging modalities, such as positron emission tomography (PET) and magnetic resonance imaging (MRI), enable the detection of disease-related biomarkers [8]. Despite its widespread use due to its non-invasive nature, high resolution, and relatively low cost, MRI has several limitations. Structural changes observed in MRI typically appear in later stages of the disease, reducing its effectiveness for early diagnosis. Additionally, similarities with other forms of dementia may lead to misclassification. Variations in imaging technology can affect consistency in longitudinal and cross-study analyses [9]. Therefore, although MRI remains one of the most commonly used imaging modalities for AD diagnosis [10], its limitations necessitate the development of more robust and automated diagnostic approaches.

Artificial Intelligence (AI), machine learning (ML), and deep learning (DL) are closely related yet distinct paradigms in intelligent system development. AI broadly refers to the creation of systems capable of performing tasks that typically require human intelligence, while ML enables systems to learn from data without explicit programming. DL, a subset of ML, utilizes multi-layer neural networks to model

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

complex enhance clinical decision-making, automate patient assessment, and improve disease diagnosis and management [11]. The application of AI in medical imaging continues to expand significantly. For instance, Yun Jian *et al.* proposed a multi-scale attention convolutional neural network for automated skin lesion segmentation, incorporating a Channel and Spatial Attention Residual Module (CSARM) to refine feature representation and suppress noise, thereby outperforming conventional U-Net architectures [12]. Similarly, Feng *et al.* developed a deep learning framework for early Alzheimer's disease diagnosis by combining 3D convolutional neural networks (3D CNNs) with Fully Stacked Bidirectional Long Short-Term Memory (FSBi-LSTM) networks, enabling effective extraction of spatial and volumetric features from MRI and PET data [13]. On the other hand, Smriti Regmi *et al.* explored Vision Transformers (ViT) for chest X-ray classification, demonstrating that self-attention mechanism can capture long-range dependencies more effectively than traditional CNNs [14]. In addition, Nilesh Shelke *et al.* introduced an ensemble transfer learning approach for diabetic retinopathy grading by integrating multiple fine-tuned models, such as EfficientNet-B0 and DenseNet121, resulting in improved diagnostic accuracy in resource-constrained environments [15]. In the context of Alzheimer's disease classification, several approaches have been proposed. Baglat *et al.* employed hybrid machine learning techniques, including Support Vector Machines, Random Forests, and Logistic Regression, to improve classification performance [16]. El-Assy *et al.* developed a modified convolutional neural networks (CNN) architecture for Alzheimer's detection. However, their evaluation was limited to overall accuracy without detailed analysis using confusion matrices or class-specific metrics [17]. Furthermore, Austin *et al.* utilized the ConvNeXT architecture for MRI-based Alzheimer's classification, achieving an accuracy of 75%, but at the cost of high computational complexity [18], [19].

Despite these promising advancements, many existing deep learning models rely on computationally intensive architectures, limiting their applicability in resource-constrained environments. To address this challenge, lightweight deep learning has emerged as an alternative paradigm that emphasizes reduced model parameters, lower latency, minimal memory usage, and improved energy efficiency while maintaining competitive performance [19]. One prominent lightweight architecture is MobileNetV3, which is specifically designed for mobile and embedded applications through a combination of hardware-aware Neural Architecture Search (NAS) and the NetAdapt optimization algorithm [20]. Several studies have demonstrated the effectiveness of this architecture. For example, L. Zhao *et al.* improved the MobileNetV3 structure by optimizing partial residual blocks, resulting in reduced parameters and floating-point operations (FLOPs), as well as faster inference speed [21]. Similarly, Liang *et al.* applied MobileNetV3 to facial expression recognition, achieving accuracies of 87.5% and 86.6% on the FERPlus and RAF-DB datasets, respectively, while reducing computational costs [22]. In the medical domain, Majeed *et al.* utilized MobileNetV3 with transfer learning for multiclass brain lesion classification, achieving an accuracy of 91% across multiple datasets [23]. Motivated by these findings, this study proposes the use of the MobileNetV3 architecture for Alzheimer's disease classification based on MRI data. Specifically, two variants, MobileNetV3-Small and MobileNetV3-Large, are implemented and evaluated. Furthermore, the performance of these models is compared with other deep learning architectures to identify the most effective and efficient approach for this task.

2. METHODS AND DATA

This study adopts a methodological approach that begins with a comprehensive review of previous research, followed by the analysis of relevant architectural models. The proposed approach consists of several stages, including dataset preparation, data preprocessing, data augmentation, model fine-tuning, training, and performance evaluation. During the preprocessing stage, the dataset is prepared and transformed to ensure consistency and suitability for model training. Data augmentation techniques, such as rotation, scaling, and flipping, are applied to increase data diversity and improve model generalization. Subsequently, the models and their corresponding hyperparameters are fine-tuned before the training process is conducted. The performance of the MobileNetV3 models is then evaluated using appropriate metrics to assess the effectiveness of the proposed classification approach. Figure 1 illustrates the overall workflow of the research methodology.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

Figure 1. Research Stages

2.1. Preparation Datasets

The Alzheimer's MRI dataset (4-class images), licensed by the Alzheimer's Disease Neuroimaging Initiative (ADNI), was utilized in this study [24], [25]. ADNI is a global research initiative that collects longitudinal data from individuals with Alzheimer's disease, Mild Cognitive Impairment (MCI), and normal cognitive function to support early detection using neuroimaging techniques. The datasets is publicly available on Kaggle (https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset/data). Previous studies have also employed this dataset for Alzheimer's disease classification tasks [26]. The dataset consists of four class: non-demented, very mild dementia, mild dementia, and moderate dementia. All MRI brain images are provided as 2D slices. The class labels are based on the Clinical Dementia Rating (CDR) scale, which quantifies the severity of cognitive impairment. In total, the dataset contains 6,400 images with a uniform resolution of 208x176 pixels. The class distribution includes 3,200 non-demented images, 2,240 very slightly demented images, 896 mild dementia images, and 64 moderate dementia images. **Error! Reference source not found.** illustrates sample images from each class. The dataset is divided into three subsets: training, validation, and testing. A split ration of 60:20:20 is applied, resulting 3,840 training images, 1,280 validation images, and 1,280 testing images. The splitting process is performed using the scikit-learn library with a fixed random state of 42 to ensure reproducibility. One-hot encoding is applied to the class labels to facilitate multi-class classification. To improve model generalization and mitigate overfitting, data augmentation techniques are applied after dataset partitioning. This step is particularly important due to the limited size and class imbalance of MRI datasets. The augmentation pipeline includes horizontal and vertical flipping to increase data diversity. Given that Alzheimer's disease exhibits relatively symmetric structural changes in the brain, flip-based augmentation is considered appropriate for this task.

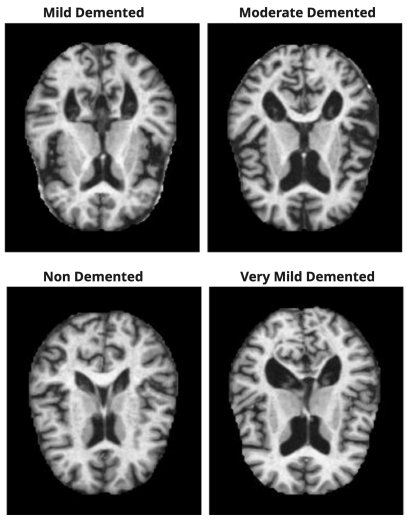


Figure 2. Alzheimer Disease Image

In addition, geometric transformations such as random rotation (±15°) and affine scaling (0.9-1.1) are applied while preserving anatomical consistency. All input images are normalized using a mean of 0.5 and a standard deviation of 0.5 to ensure stable training. The overall augmentation process includes random rotation, affine transformation, horizontal and vertical flipping ($p = 0.5$), and normalization. **Error! Reference source not found.** presents examples of the augmented images.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

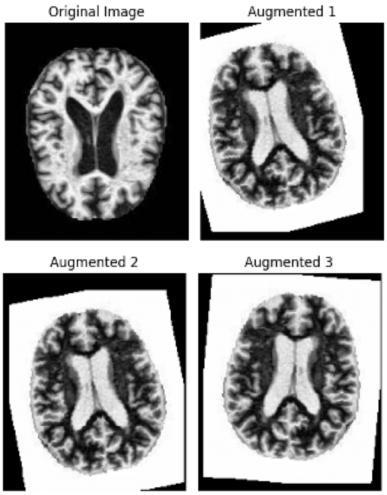
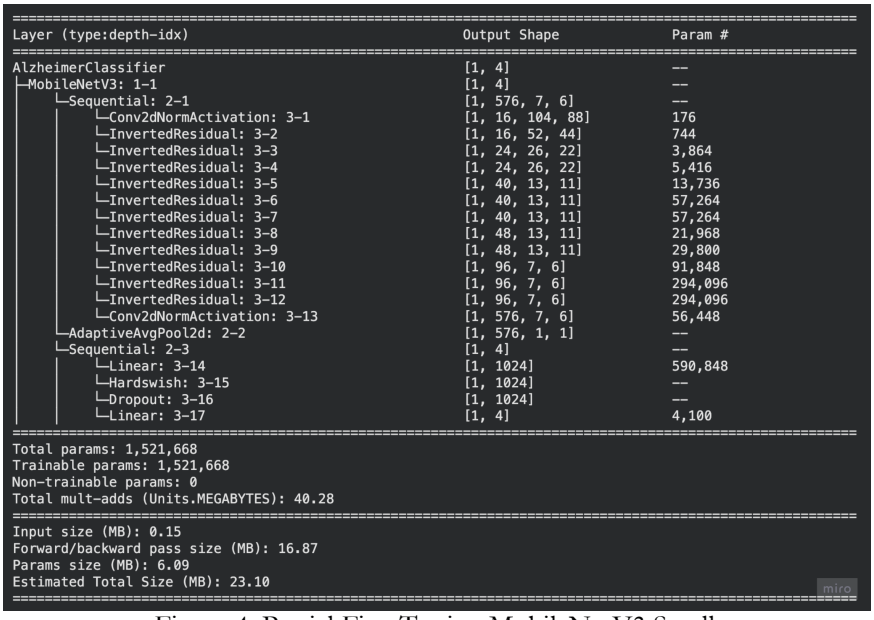


Figure 3. Image Augmentations

2.2. Fine Tuning and Training Architecture

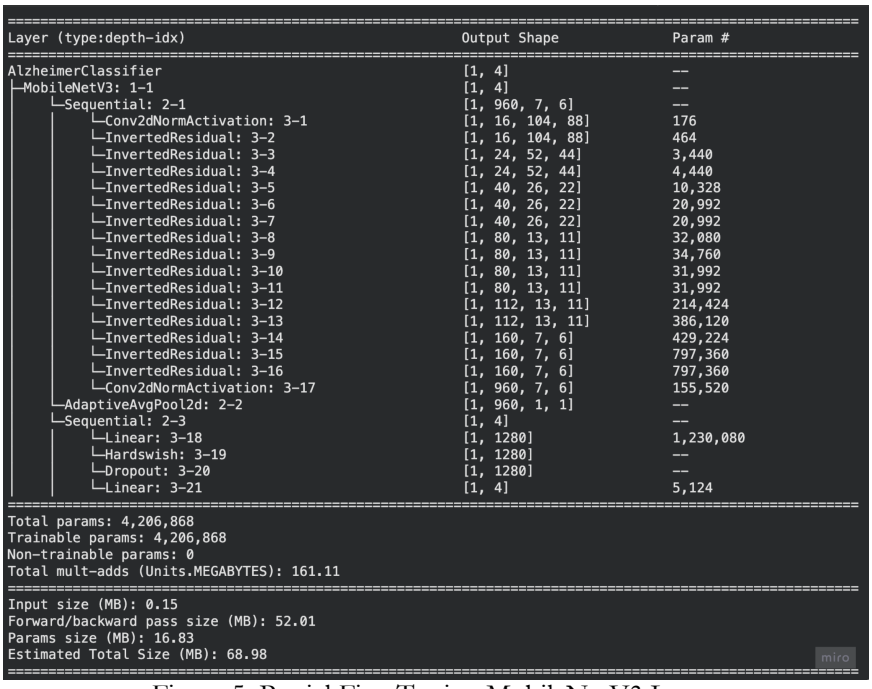
MobileNetV3 is a family of lightweight convolutional neural network architectures that introduces significant improvements over previous versions of MobileNet. These enhancements can be categorized into three main aspects. First, MobileNetV3 utilizes automated architecture design through Neural Architecture Search (NAS) and the NetAdapt algorithm. Unlike MobileNetV1 and MobileNetV2, which were primarily hand-crafted, MobileNetV3 is optimized using these techniques to achieve an optimal balance between accuracy and computational efficiency. Specifically, NAS determines the global network structure based on hardware constrains, while NetAdapt refines the network at the layer level to meet latency requirements [20]. Second, MobileNetV3 incorporates efficient architectural components to enhance performance. Squeeze-and-Excitation (SE) modules are integrated into the bottleneck structure to improve feature representation by emphasizing important channels. In addition, a novel activation function, knows as h-swish, is introduced to provide a computationally efficient alternative to traditional activation functions. Third, the architecture includes modifications to reduce computational cost. In particular, the final 1x1 convolutional layer is repositioned after the global average pooling layer, resulting in lower computational complexity without sacrificing accuracy. Furthermore, the number of filters in the initial 3x3 convolutional layer is reduced from 32 to 16, contributing to a more efficient model design. In this study, two variants of MobileNetV3 are employed, namely MobileNetV3-Large and MobileNetV3-Small, to evaluate their performance in Alzheimer's disease classification. The Large variant is designed for higher accuracy, whereas the Small variant is optimized for resource-constrained environments. Partial fine-tuning is applied by modifying the final classification layer to match the number of output classes in the dataset. **Error! Reference source not found.** and **Error! Reference source not found.** illustrate the architecture of MobileNetV3-Small and MobileNetV3-Large, respectively. The primary architectural differences between the two variants are observed in the number of layers within the Inverted Residual blocks. The MobileNetV3-Small model consists of layers up to stage 3-12, whereas the MobileNetV3-Large model extends to stage 3-16, resulting in additional depth in the Large variant. The models are implemented using the PyTorch framework and trained on Google Colaboratory with GPU acceleration. The input image size is set to 208x176 pixels, and the final classification layer is adjusted to produce four output classes. The Adam optimizer is used for training, and cross-entropy loss is employed as the objective function to optimize model performance. **Error! Reference source not found.** provides detailed information on the hyperparameters used in this study.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58



```
==========================================================================================
Layer (type:depth-idx)                                  Output Shape              Param #
==========================================================================================
AlzheimerClassifier                                     [1, 4]                    --
├─MobileNetV3: 1-1                                      [1, 4]                    --
│    └─Sequential: 2-1                                  [1, 576, 7, 6]            --
│    │    └─Conv2dNormActivation: 3-1                   [1, 16, 104, 88]          176
│    │    └─InvertedResidual: 3-2                       [1, 16, 52, 44]           744
│    │    └─InvertedResidual: 3-3                       [1, 24, 26, 22]           3,864
│    │    └─InvertedResidual: 3-4                       [1, 24, 26, 22]           5,416
│    │    └─InvertedResidual: 3-5                       [1, 40, 13, 11]           13,736
│    │    └─InvertedResidual: 3-6                       [1, 40, 13, 11]           57,264
│    │    └─InvertedResidual: 3-7                       [1, 40, 13, 11]           57,264
│    │    └─InvertedResidual: 3-8                       [1, 48, 13, 11]           21,968
│    │    └─InvertedResidual: 3-9                       [1, 48, 13, 11]           29,800
│    │    └─InvertedResidual: 3-10                      [1, 96, 7, 6]             91,848
│    │    └─InvertedResidual: 3-11                      [1, 96, 7, 6]             294,096
│    │    └─InvertedResidual: 3-12                      [1, 96, 7, 6]             294,096
│    │    └─Conv2dNormActivation: 3-13                  [1, 576, 7, 6]            56,448
│    └─AdaptiveAvgPool2d: 2-2                           [1, 576, 1, 1]            --
│    └─Sequential: 2-3                                  [1, 4]                    --
│    │    └─Linear: 3-14                                [1, 1024]                 590,848
│    │    └─Hardswish: 3-15                             [1, 1024]                 --
│    │    └─Dropout: 3-16                               [1, 1024]                 --
│    │    └─Linear: 3-17                                [1, 4]                    4,100
==========================================================================================
Total params: 1,521,668
Trainable params: 1,521,668
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 40.28
==========================================================================================
Input size (MB): 0.15
Forward/backward pass size (MB): 16.87
Params size (MB): 6.09
Estimated Total Size (MB): 23.10
==========================================================================================
```

Figure 4. Partial Fine Tuning MobileNetV3 Small



```
==========================================================================================
Layer (type:depth-idx)                                  Output Shape              Param #
==========================================================================================
AlzheimerClassifier                                     [1, 4]                    --
├─MobileNetV3: 1-1                                      [1, 4]                    --
│    └─Sequential: 2-1                                  [1, 960, 7, 6]            --
│    │    └─Conv2dNormActivation: 3-1                   [1, 16, 104, 88]          176
│    │    └─InvertedResidual: 3-2                       [1, 16, 104, 88]          464
│    │    └─InvertedResidual: 3-3                       [1, 24, 52, 44]           3,440
│    │    └─InvertedResidual: 3-4                       [1, 24, 52, 44]           4,440
│    │    └─InvertedResidual: 3-5                       [1, 40, 26, 22]           10,328
│    │    └─InvertedResidual: 3-6                       [1, 40, 26, 22]           20,992
│    │    └─InvertedResidual: 3-7                       [1, 40, 26, 22]           20,992
│    │    └─InvertedResidual: 3-8                       [1, 80, 13, 11]           32,080
│    │    └─InvertedResidual: 3-9                       [1, 80, 13, 11]           34,760
│    │    └─InvertedResidual: 3-10                      [1, 80, 13, 11]           31,992
│    │    └─InvertedResidual: 3-11                      [1, 80, 13, 11]           31,992
│    │    └─InvertedResidual: 3-12                      [1, 112, 13, 11]          214,424
│    │    └─InvertedResidual: 3-13                      [1, 112, 13, 11]          386,120
│    │    └─InvertedResidual: 3-14                      [1, 160, 7, 6]            429,224
│    │    └─InvertedResidual: 3-15                      [1, 160, 7, 6]            797,360
│    │    └─InvertedResidual: 3-16                      [1, 160, 7, 6]            797,360
│    │    └─Conv2dNormActivation: 3-17                  [1, 960, 7, 6]            155,520
│    └─AdaptiveAvgPool2d: 2-2                           [1, 960, 1, 1]            --
│    └─Sequential: 2-3                                  [1, 4]                    --
│    │    └─Linear: 3-18                                [1, 1280]                 1,230,080
│    │    └─Hardswish: 3-19                             [1, 1280]                 --
│    │    └─Dropout: 3-20                               [1, 1280]                 --
│    │    └─Linear: 3-21                                [1, 4]                    5,124
==========================================================================================
Total params: 4,206,868
Trainable params: 4,206,868
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 161.11
==========================================================================================
Input size (MB): 0.15
Forward/backward pass size (MB): 52.01
Params size (MB): 16.83
Estimated Total Size (MB): 68.98
==========================================================================================
```

Figure 5. Partial Fine Tuning MobileNetV3 Large

Table 1. Hyperparameter

Hyperparameter	Values
Epoch	128
Batch Size	8
Learning Rate	0.001

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

2.3. Evaluation Architecture

The performance of the proposed architecture is evaluated using a confusion matrix, which represents the relationship between the predicted labels and ground truth. The confusion matrix categorizes the results into four components: True Positive (TP) and True Negative (TN), which indicate correct classifications, as well as False Positive (FP) and False Negative (FN), which represent misclassifications. These four components from the basis for three key performance metrics: Precision, Recall, and F1-Score. Precision measures the accuracy of positive predictions by calculating the proportion of correctly predicted positive instances. Recall evaluates the model's ability to identify all relevant instances by measuring the proportion of actual positives that are correctly classified. The F1-Score provides a balanced measure by computing the harmonic mean of Precision and Recall.

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall} \quad (3)$$

3. RESULTS AND DISCUSSION

3.1 Validation Loss Performance

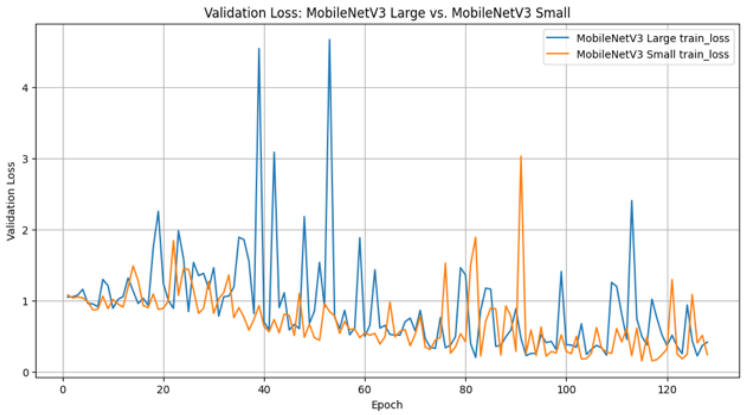


Figure 6. Validation Loss

Error! Reference source not found. illustrates the validation loss trajectories of both MobileNetV3-Large and MobileNetV3-Small over 128 training epochs. Both models demonstrate a clear learning trend, with loss values decreasing significantly from the initial epochs to the later stages of training. However, the MobileNetV3-Large model exhibits noticeable instability, suggesting higher sensitivity to specific training hyperparameters on this dataset. Although it achieves a final loss comparable to that of the Small variant, the observed fluctuations indicate potential optimization instability or transient overfitting. In contrast, the MobileNetV3-Small model demonstrates a more stable and consistent training trajectory, indicating better reliability for this classification task.

3.2 Performance Validation, Precision, Recall, and F1-Score

Figure 7 presents the progression of validation precision for both models. A strong relationship between decreasing loss and increasing precision is observed, with performance improving from an initial value of approximately 0.15 to peak values exceeding 0.90. The precision results further support the stability trends identified in the loss analysis, where performance drops in the MobileNetV3-Large model correspond to spikes in validation loss. Figure 8 illustrates the validation recall performance of both

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

architectures. The MobileNetV3-Small model demonstrates superior training stability and faster convergence. During the early stages of training, the Small variant consistently achieves higher recall values, whereas the MobileNetV3-Large model exhibits noticeable fluctuations. Figure 9 shows the progression of the validation F1-Score, which provides a balanced evaluation of classification performance by combining precision and recall. The MobileNetV3-Small model maintains stable and efficient performance throughout the training process. In contrast, the MobileNetV3-Large model continues to exhibit variability, consistent with the trends observed in previous metrics. The F1-Score results indicate that the Small variant achieves faster convergence toward an optimal balance between false positives and false negatives, resulting in a more robust and reliable performance during validation.

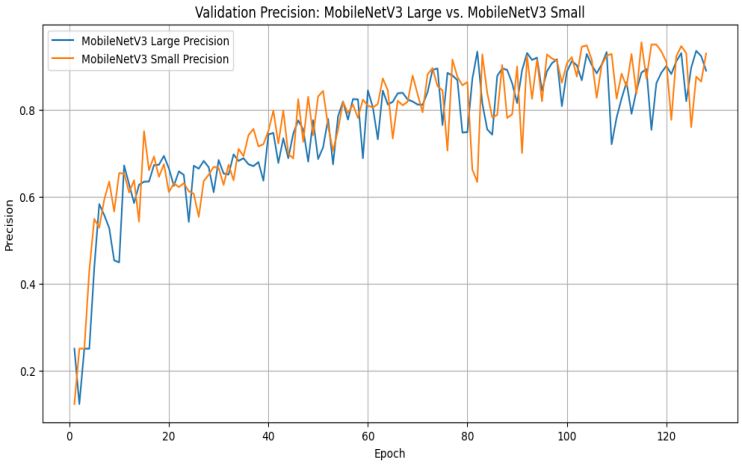


Figure 7. Validation Precision Performance

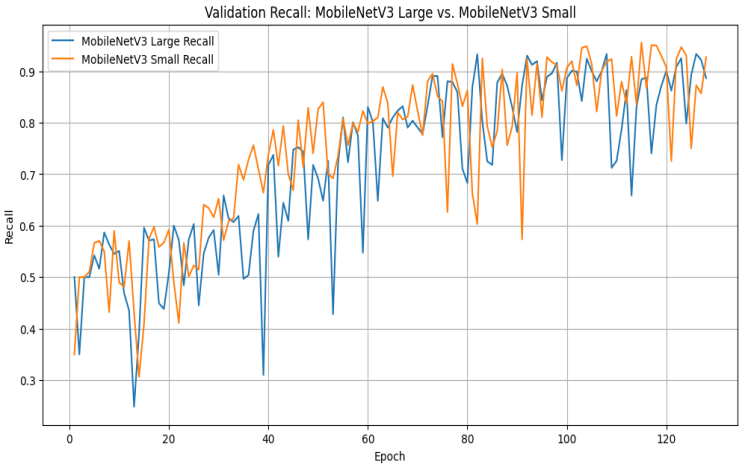


Figure 8. Validation Recall Performance

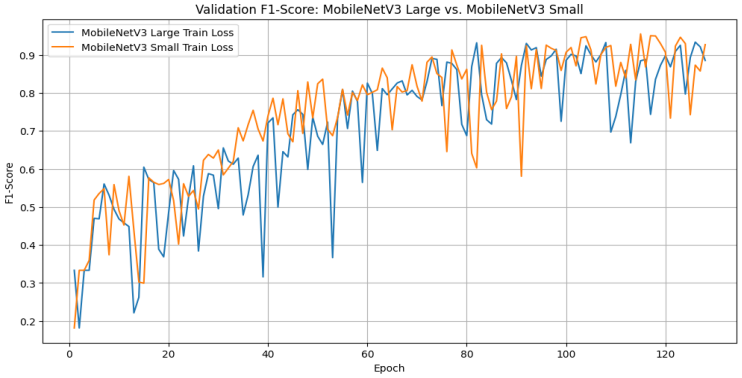


Figure 9. Validation F1-Score Performance

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

3.3 Performance Testing Precision, Recall, F1-Score, and Accuracy

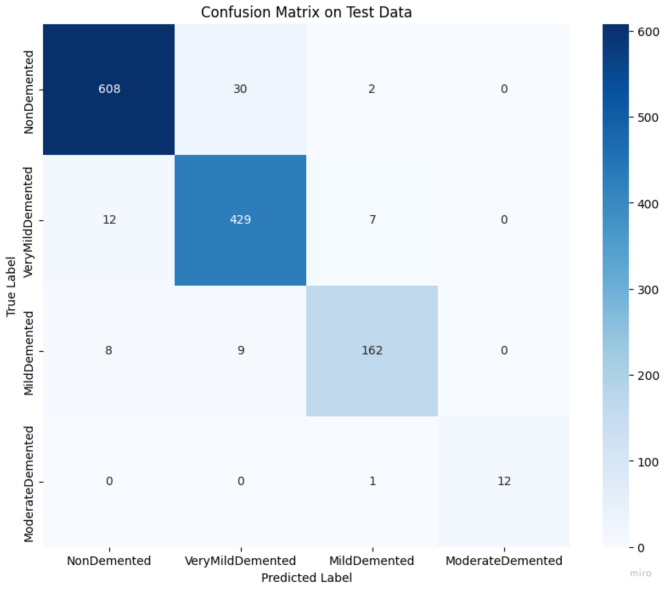


Figure 10. Confusion Matrix MobileNetV3 Small

The MobileNetV3-Small model demonstrates high effectiveness in classifying the four dementia classes, as shown in Figure 10. Out of 1,280 test samples, the model correctly classified 1,211 instances. The largest class is Non-Demented, with 640 samples. In contrast, Moderate Demented represents the minority class, consisting of 13 samples, of which 12 were correctly classified and one was misclassified as Mild Demented. Figure 11 shows that the MobileNetV3-Large model correctly classified 1,158 samples. However, this model exhibits a stronger bias toward the majority class (Non-Demented), resulting in a higher number of correctly classified samples in this category compared to the Small variant. A significant limitation is observed in the Very Mild Demented class, where the model fails to identify 58 early-stage cases and misclassifies them as Non-Demented. This indicates that, although the Large model provides slightly better specificity for non-demented cases, it underperforms in detecting early-stage dementia compared to MobileNetV3-Small. Table 2 summarizes the performance metrics, showing that MobileNetV3-Small outperforms MobileNetV3-Large across all evaluation metrics, including Precision, Recall, and F1-Score.

Table *3* further presents a comparison with other architectures. MobileNetV3-Small achieves the highest overall accuracy of 0.94, while MobileNetV3-Large and EfficientNetV2-s both achieve an accuracy of 0.90. In contrast, heavier architectures such as ConvNeXT and EfficientNetV2-m demonstrate lower performance, with accuracies of 0.75 and 0.81, respectively. These results suggest that the compact architecture of MobileNetV3-Small, which leverages depthwise separable convolutions and squeeze-and-excitation modules, is more effective in capturing relevant features in this dataset compared to deeper and more complex models. Table 4 presents the class-wise performance metrics for both model variants. Overall, both models demonstrate strong performance, with most F1-Scores exceeding 0.86. However, MobileNetV3-Small exhibits superior consistency and a better balance between precision and recall, particularly for more challenging classes. The Small variant achieves F1-Scores of 0.92 or higher across all classes, whereas the Large variant falls below 0.90 in the Moderate Demented and Non-Demented classes.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

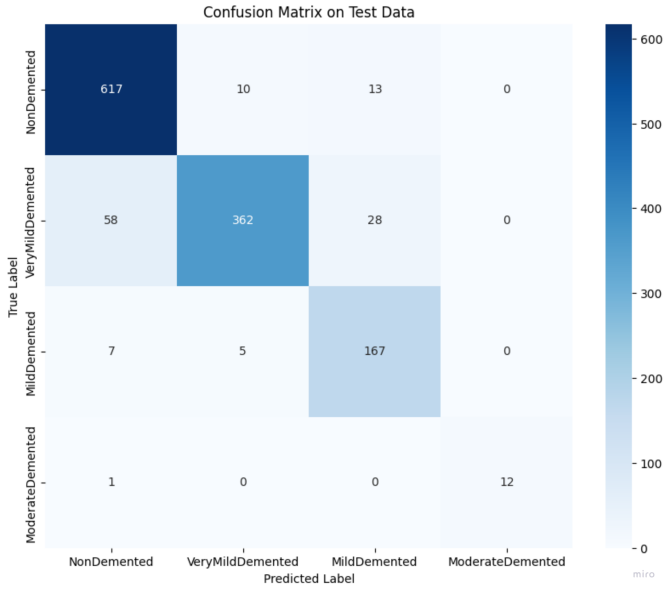


Figure 11. Confusion Matrix MobileNetV3 Large

This balanced performance is particularly important in medical applications, where both false negatives and false positives have significant clinical implications. Although the MobileNetV3-Large model has a more complex architecture and a higher number of parameters, which theoretically should provide greater representational capacity, it struggles with more challenging classes in this dataset. This further supports the effectiveness of the MobileNetV3-Small model as a more reliable and efficient solution for Alzheimer's disease classification.

Table 2. Results Precision, Recall, and F1-Score

Metrics	MobileNetV3 Large	MobileNetV3 Small
Precision	0.91	0.94
Recall	0.90	0.94
F1-Score	0.90	0.94

Table 3. Comparison Accuracy Models

Model	Accuracy
MobileNetV3 Small	0.94
MobileNetV3 Large	0.90
EfficientNetV2-s	0.90
EfficientNetV2-m	0.81
ConvNeXT [18]	0.75

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

Table 4. Precision, Recall, and F1-Score by Class

Class	MobileNetV3 Large			MobileNetV3 Small		
	P	R	F1	P	R	F1
Mild Demented	0.90	0.96	0.93	0.96	0.95	0.95
Moderate Demented	0.96	0.80	0.87	0.91	0.95	0.93
Non Demented	0.80	0.93	0.86	0.94	0.90	0.92
Very Mild Demented	1	0.92	0.96	1	0.92	0.96

3.4 Efficiency Model Comparisons

To assess architectural efficiency, this study evaluates the number of parameters and floating-point operations (FLOPs) across the compared models, as summarized in the Table 5. The MobileNetV3 family serves as a baseline for edge deployment due to its lightweight design. Notably, although the MobileNetV3-Large model contains more than twice the number of parameters compared to the Small variant, it exhibits a significantly lower FLOP count. This observation indicates that a higher number of parameters does not necessarily correspond to improved computational efficiency or classification performance.

Table 5. Parameter Size and FLOPs

Model	Parameters	FLOPs
MobileNetV3 Small	2.5 M	0.6 G
MobileNetV3 Large	5.4 M	0.22 G
EfficientNetV2-s	22 M	8.8 G
EfficientNetV2-m	54 M	24 G
ConvNeXT Tiny	28 M	4.5 G
ConvNeXT Small	50 M	8.7 G
ConvNeXT Base	89 M	15.4 G

4. CONCLUSION

The main findings of this study can be summarized as follows:

1. For the Alzheimer's disease dataset used in this study, increased model complexity, as reflected by a higher number of parameters and FLOPs, does not necessarily lead to improved classification performance or evaluation metrics.
2. Among the evaluated architectures, MobileNetV3-Small achieves the best overall performance, outperforming other comparison models across key evaluation metrics.

In addition, the class imbalance present in the dataset affects the distribution of performance metrics across classes, resulting in less consistent generalization, particularly for minority classes.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

REFERENCES

[1] Iman Beheshti, Norihide Maikusa, Morteza Daneshmand, Hiroshi Matsuda, Hasan Demirel, and Gholamreza Anbarjafari, "Classification of Alzheimer's Disease and Prediction of Mild Cognitive Impairment Conversion Using Histogram-Based Analysis of Patient-Specific Anatomical Brain Connectivity Networks," Journal of Alzheimer's Disease, vol. 60, no. 1, pp. 295–304, Aug. 2017, doi: 10.3233/JAD-161080.

[2] A. Association, "2014 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 10, no. 2, pp. e47–e92, Mar. 2014, doi: https://doi.org/10.1016/j.jalz.2014.02.001.

[3] E. Hogervorst, E. Schröder-Butterfill, Y. S. Handajani, P. Kreager, and T. B. W. Rahardjo, "Dementia and dependency vs. proxy indicators of the active ageing index in Indonesia," Int. J. Environ. Res. Public Health, vol. 18, no. 16, p. 8235, 2021.

[4] I. Theresia, T. P. Sani, S. Evans-Lacko, N. Farina, L. Augustina, and Y. Turana, "Experiences of caregivers of people with dementia in Indonesia: A focus group study," Int. J. Geriatr. Psychiatry, vol. 38, no. 12, p. e6038, 2023.

[5] E. S. van der Ploeg, Y. S. Handajani, and E. Schröder-Butterfill, "Dementia Care in Indonesia: Care Networks, Awareness & Perception," J. Cross. Cult. Gerontol., vol. 40, no. 4, pp. 441–461, 2025, doi: 10.1007/s10823-025-09544-x.

[6] A. Association, "Changing the trajectory of Alzheimer's disease: How a treatment by 2025 saves lives and dollars," Chicago: Alzheimer's Association, 2015.

[7] W. Jagust, "Imaging the evolution and pathophysiology of Alzheimer disease," Nat. Rev. Neurosci., vol. 19, no. 11, pp. 687–700, 2018.

[8] K. J. Singer, "Unraveling the Neurocognitive Effects of Caregiver Stress Recall: An fNIRS Investigation," 2024.

[9] S. S. A. Ali, K. Memon, N. Yahya, and S. Khan, "Deep learning frameworks for MRI-based diagnosis of neurological disorders: A systematic review and meta-analysis," Artif. Intell. Rev., vol. 58, no. 6, p. 171, 2025.

[10] T. Tong et al., "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," IEEE Trans. Biomed. Eng., vol. 64, no. 1, pp. 155–165, 2016.

[11] A. Haleem, M. Javaid, and I. H. Khan, "Current status and applications of Artificial Intelligence (AI) in medical field: An overview," Curr. Med. Res. Pract., vol. 9, no. 6, pp. 231–237, 2019.

[12] J. Yun, S. Cao, T. Shengxin, and H. Zhang, "Skin Lesion Segmentation Based on Multi-Scale Attention Convolutional Neural Network," IEEE Access, vol. PP, p. 1, Jul. 2020, doi: 10.1109/ACCESS.2020.3007512.

[13] C. Feng et al., "Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM," IEEE Access, vol. PP, p. 1, Apr. 2019, doi: 10.1109/ACCESS.2019.2913847.

[14] S. Regmi, A. Subedi, N. K. Tomar, U. Bagci, and D. Jha, "Vision transformer for efficient chest x-ray and gastrointestinal image classification," in Medical Imaging 2025: Computer-Aided Diagnosis, SPIE, 2025, pp. 912–923.

[15] N. Shelke, R. Somkunwar, A. Pimpalkar, S. Maurya, and S. Chhabria, "Ensemble EfficientNet: a novel technique for identification, classification and prediction of diabetic retinopathy," Automatika, vol. 66, no. 3, pp. 543–558, Jul. 2025, doi: 10.1080/00051144.2025.2514884.

[16] P. Baglat, A. W. Salehi, A. Gupta, and G. Gupta, "Multiple Machine Learning Models for Detection of Alzheimer's Disease Using OASIS Dataset," in Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation, S. K. Sharma, Y. K. Dwivedi, B. Metri, and N. P. Rana, Eds., Cham: Springer International Publishing, 2020, pp. 614–622.

[17] A. M. El-Assy, H. M. Amer, H. M. Ibrahim, and M. A. Mohamed, "A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data," Sci. Rep., vol. 14, no. 1, p. 3463, 2024, doi: 10.1038/s41598-024-53733-6.

[18] "Klasifikasi Penyakit Alzheimer Dari Scan Mri Otak Menggunakan Convnext," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, no. 6, pp. 1223–1232, Dec. 2024, doi: 10.25126/jtiik.2024118117.

[19] H.-I. Liu et al., "Lightweight deep learning for resource-constrained environments: A survey," ACM Comput. Surv., vol. 56, no. 10, pp. 1–42, 2024.

[20] A. Howard et al., "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314–1324.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58

[21] L. Zhao and L. Wang, "A new lightweight network based on MobileNetV3," KSII Transactions on Internet and Information Systems (TIIS), vol. 16, no. 1, pp. 1–15, 2022.
[22] X. Liang, J. Liang, T. Yin, and X. Tang, "A lightweight method for face expression recognition based on improved MobileNetV3," IET Image Process., vol. 17, no. 8, pp. 2375–2384, Jun. 2023, doi: https://doi.org/10.1049/ipr2.12798.
[23] A. F. Majeed, P. Salehpour, L. Farzinvash, and S. Pashazadeh, "Multi-Class Brain Lesion Classification Using Deep Transfer Learning With MobileNetV3," IEEE Access, vol. 12, pp. 155295–155308, 2024, doi: 10.1109/ACCESS.2024.3413008.
[24] S. G. Mueller et al., "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)," Alzheimer's & Dementia, vol. 1, no. 1, pp. 55–66, 2005.
[25] C. R. Jack Jr. et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," Journal of Magnetic Resonance Imaging, vol. 27, no. 4, pp. 685–691, Apr. 2008, doi: https://doi.org/10.1002/jmri.21049.
[26] M. Asaduzzaman, Md. K. Alom, and Md. E. Karim, "ALZENET: Deep learning-based early prediction of Alzheimer's disease through magnetic resonance imaging analysis," Telematics and Informatics Reports, vol. 17, p. 100189, 2025, doi: https://doi.org/10.1016/j.teler.2025.100189.

DOI : https://doi.org/10.21063/jtif.2026.V14.1.47-58