

A comparative MRI-based study of ResNet-152 and novel deep learning approaches for early Alzheimer's disease classification

Kelvin Leonardi Kohsasih, Octara Pribadi, Andy, Daniel Smith Sunario

Department of Informatics Engineering, STMIK TIME, Medan, Indonesia

Article Info

Article history:

Received Jul 10, 2025

Revised Dec 11, 2025

Accepted Jan 30, 2026

Keywords:

Convolutional neural network
Deep learning
Medical image classification
Mutli-class classification
Transfer learning

ABSTRACT

Alzheimer's disease (AD) is the leading cause of dementia, making early-stage detection essential for timely intervention. Most prior studies have focused on binary AD classification, which limits sensitivity to disease progression. This study addressed this gap by evaluating whether tailored convolutional neural network (CNN) architectures could improve stage-aware classification using a publicly available magnetic resonance imaging (MRI) dataset containing 35,984 images across four diagnostic categories. The dataset underwent grayscale conversion, resizing, contrast enhancement, normalization, and class balancing prior to model development. Four models were trained and compared: ResNet-152, a custom multiclass CNN, a one-vs-one (OvO) model, and a one-vs-rest (OvR) model. Performance was measured using accuracy, precision, recall, F1 score, and confusion-matrix-based metrics. The custom multiclass CNN achieved the strongest performance, yielding the highest accuracy and balanced results across all evaluation metrics. These findings demonstrate the value of systematically comparing decomposition strategies for multi-stage Alzheimer's detection and highlight the potential of the proposed approach to enhance early diagnostic support. Future work may incorporate multimodal inputs or hybrid architectures to improve sensitivity to subtle structural changes and further strengthen clinical applicability.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Kelvin Leonardi Kohsasih
Department of Informatics Engineering, STMIK TIME
Merbabu Street, Medan City, North Sumatra 20212, Indonesia
Email: kelvinleonardi@stmik-time.ac.id

1. INTRODUCTION

Advances in information technology and computer science have accelerated progress in medical data processing and the application of deep learning (DL) for diagnosing neurological disorders using neuroimaging techniques [1], [2]. As DL techniques have become integrated into neuroimaging research [3], the availability of large, high-quality datasets [4] has further enabled the earlier and more accurate detection of neurodegenerative diseases. Alzheimer's disease (AD) is the most common cause of dementia, responsible for 60–80% of cases among older adults [5]. AD is a progressive and irreversible neurodegenerative disorder leading to cognitive decline and eventual mortality [6], [7]. At present, no curative treatment exists, and available therapies such as aducanumab, lecanemab, and donanemab have only been shown to slow progression when administered in the early phases, particularly mild cognitive impairment (MCI) or early MCI (EMCI) [2], [8]–[10]. Nevertheless, early symptoms are nonspecific and often mistaken for normal aging [11], so diagnosis is frequently delayed until the moderate or severe stages, when interventions are less effective [12].

Magnetic resonance imaging (MRI) is crucial for examining brain anatomy and diagnosing neurological disorders. It helps identify abnormal regions [13]. Its high resolution enables clear differentiation between gray and white matter. MRI also reveals structural alterations such as hippocampal atrophy, widened sulci, and ventricular enlargement that accompany disease progression [2], [12], [14]. These features support accurate diagnosis and effective disease monitoring [15]. Despite its diagnostic value, MRI interpretation remains inherently subjective and dependent on expert judgment, which introduces variability and potential diagnostic inconsistency [12]. These challenges highlight a growing need for automated, artificial intelligence (AI)-supported analysis that can improve accuracy and reduce observer variability.

DL systems have demonstrated increasing effectiveness across various research domains and are particularly useful for analyzing structural brain changes associated with AD [16]. DL also improves diagnostic efficiency by reducing time, cost, and manual effort [17]. However, prior research on AD detection using DL has not yet achieved optimal efficiency [18]. Comparative studies evaluating you only look once (YOLO) variants, DenseNet, residual network (ResNet), visual geometry group (VGG), EfficientNet, and vision transformers indicate that although high-capacity feature extractors achieve strong performance in binary AD classification, they often struggle in multi-stage tasks, especially at MCI or EMCI levels where inter-class differences are subtle [19]–[21]. Even models such as DenseNet-169 and ResNet-50 show reduced discrimination ability when required to distinguish closely related Alzheimer’s stages [22]. Multimodal MRI–positron emission tomography (PET) fusion approaches have demonstrated improved robustness and interpretability, particularly when combined with explainability techniques; In contrast, these methods still do not consistently outperform optimized single-modality CNNs for multi-class staging tasks [23]. Overall, reported accuracies commonly range between 70% and 86%, with most studies focusing on binary AD detection rather than fine-grained staging [19], [22]–[25]. This trend highlights a persistent gap in developing models capable of reliably differentiating early Alzheimer’s stages.

To address these limitations, this study evaluates whether tailored convolutional neural network (CNN) architectures can improve fine-grained, stage-aware AD classification, an area where existing DL approaches remain limited. Using a publicly available MRI dataset comprising 35,984 images across four diagnostic categories, the dataset was preprocessed using grayscale conversion, resizing, contrast enhancement, normalization, and class balancing. Three decomposition strategies-multiclass, one-vs-one (OvO), and one-vs-rest (OvR)-were implemented alongside a ResNet-152 transfer-learning baseline to enable a systematic and controlled comparison. The novelty of this work lies in its structured evaluation of these strategies on a large and balanced MRI corpus, aimed at improving stage-specific sensitivity while maintaining computational feasibility. Improving classification granularity is clinically important because prognosis, response to therapy, and eligibility for disease-modifying treatments depend heavily on accurate early-stage differentiation. The proposed stage-aware DL framework therefore offers enhanced diagnostic value for supporting early detection of AD.

2. METHOD

This study compares the performance of ResNet-152, applied through transfer learning, with a custom DL model that employs multiclass decomposition strategies, including OvO and OvR, for Alzheimer’s detection from MRI scans. The methodology is structured into five phases: data acquisition, preprocessing, class balancing and dataset partitioning, model architecture with training configuration, and model evaluation supported by statistical analysis, as summarized in Figure 1.

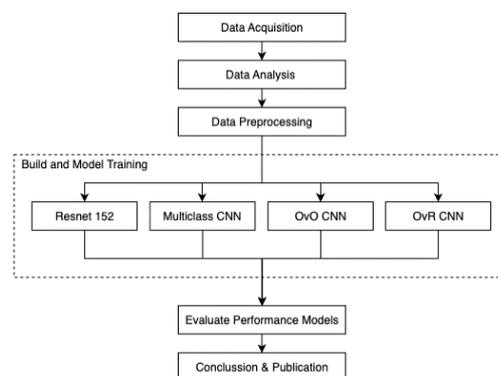


Figure 1. Method research

2.1. Data acquisition

The dataset used in this study was obtained from the publicly available Kaggle Alzheimer's Disease MRI dataset, which provides a comprehensive collection of brain MRI scans labeled according to cognitive condition. As shown in Figure 2, the dataset includes four diagnostic categories: non-demented (ND), very mild demented (VMD), mild demented (MD), and moderate demented (ModD). These categories represent progressive levels of cognitive decline, ranging from normal cognitive function to more advanced dementia symptoms, and broadly correspond to the early to moderate stages of AD progression. In total, the dataset comprises 35,984 MRI images, distributed as follows: 9,600 images (28.2%) for ND, 8,960 images (26.4%) for VMD, 8,960 images (26.4%) for MD, and 6,464 images (19.0%) for ModD. All images are provided in a standardized format and dimension, minimizing variability and ensuring seamless integration into the preprocessing pipeline. The dataset was chosen because it is openly accessible, which guarantees reproducibility and transparency. It also has a relatively balanced distribution across impairment stages, which supports robust model training and evaluation. Additionally, the dataset offers high-quality labeling, enabling reliable supervised learning. Its distribution across four diagnostic categories further allows this study to address common challenges in medical imaging research, including class imbalance and inter-class similarity, through targeted preprocessing and data augmentation strategies.



Figure 2. Representative MRI brain scan samples from the dataset

2.2. Class balancing and dataset partitioning

The initial dataset exhibited class imbalance, with ND containing 9,600 samples, MD 8,960 samples, VMD 8,960 samples, and ModD 6,464 samples, which could bias the model toward majority classes. In this approach, the majority class samples were randomly discarded until class sizes were equalized, thereby reducing bias toward dominant categories, which is consistent with the fundamental principles of re-balancing methods [26], [27]. A fixed random seed (42) was used to ensure reproducibility. The balanced dataset was subsequently divided into 80% for training and 20% for testing to support robust model learning and objective performance assessment on unseen data. The class distribution of the dataset before and after the re-balancing procedure is presented in Figure 3. Figure 3(a) shows the class distribution before the re-balancing procedure, while Figure 3(b) illustrates the distribution after the re-balancing procedure.

2.3. Data preprocessing

Before model training, all MRI images underwent a standardized preprocessing pipeline consisting of:

- Grayscale conversion to reduce computational complexity while preserving structural information.

- Resizing to 224×224 pixels to ensure uniform input dimensions for the CNN model [28].
- Contrast enhancement using contrast limited adaptive histogram equalization (CLAHE) to improve local contrast and highlight subtle anatomical features relevant to AD classification [18], [29].
- Bilateral filtering to suppress noise while preserving edge details.
- Min–max normalization to scale pixel intensities into the [0, 1] range for stable network optimization.
- Batchwise processing of 300 images per cycle to maintain memory efficiency during preprocessing.

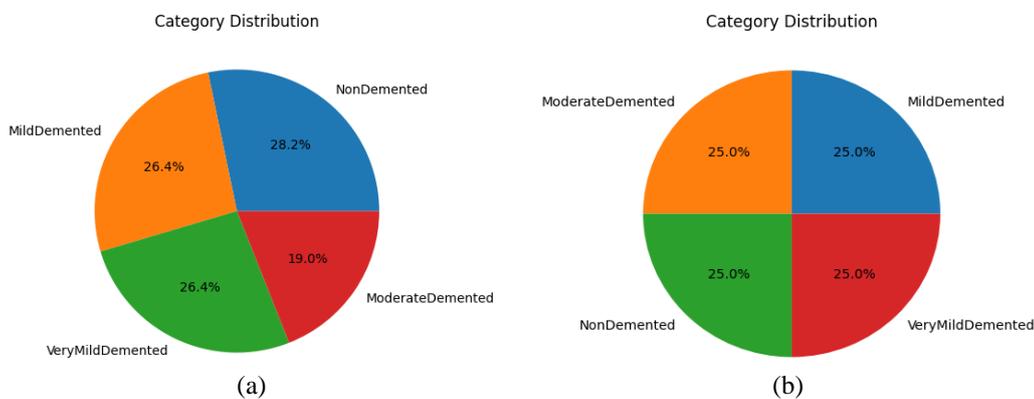


Figure 3. Initial and balanced class distributions of the AD image dataset; (a) initial imbalanced distribution and (b) balanced distribution after undersampling

2.4. Model architecture and training configuration

2.4.1. Resnet 152

ResNet-152 is a deep variant of the ResNet family that addresses gradient degradation in very deep models through residual learning with identity shortcut connections [30]. With 152 layers organized into bottleneck blocks, it provides hierarchical feature extraction that is effective for detecting subtle morphological changes in brain MRI scans associated with early AD [31], [32]. Beyond its widespread use, ResNet-152 was selected in this study because recent work on deep residual and convolutional architectures applied to structural MRI has shown that very deep networks can capture distributed and fine-grained patterns of brain atrophy that are critical for accurate AD staging [33].

In this study, ResNet-152 was implemented using transfer learning with pretrained ImageNet weights, where convolutional layers were frozen to preserve general feature representations and a task-specific classification head was trained on the Alzheimer's dataset [31], [32], [34]. The classification head consisted of a global average pooling layer, a dense layer with 1024 neurons and rectified linear unit (ReLU) activation, and a final dense layer with four neurons and softmax activation to classify the diagnostic categories. The model was trained using the adaptive moment estimation (Adam) optimizer with categorical cross-entropy loss and a batch size of 32. Early stopping was applied to prevent overfitting and to restore the best weights. This configuration aligns with established transfer learning practices in medical image analysis [32], [34].

2.4.2. Custom classifier

CNNs have demonstrated strong effectiveness in neuroimaging because they learn hierarchical spatial features directly from MRI data [2], [35], [36]. Prior studies also indicate that task-specific CNN architectures are better suited to capture disease-related morphological patterns, offering improved stability and interpretability in medical imaging applications [37]–[40]. In this study, a custom multiclass CNN was developed using stacked convolutional blocks for feature extraction and fully connected layers with dropout to enhance generalization performance [41], [42]. A final softmax layer produces predictions across four diagnostic categories based on structural MRI features [43].

A controlled comparison for multi-stage Alzheimer's classification was achieved by implementing two additional formulations. The OvO approach trains individual pairwise classifiers that refine decision boundaries between specific class pairs, while the OvR formulation trains separate models to distinguish one diagnostic category from the remaining classes. Both approaches employ compact CNN backbones and follow established multiclass learning practices in medical imaging [44]–[48]. Detailed architectural specifications, including layer composition, output dimensions, and parameter counts, are provided in Tables 1 to 4.

Table 1. Layers of custom multiclass CNN

Type	Description	Parameters	Output shape
Input layer	Input_1	0	(None, 224, 224, 1)
Conv block 1	2×Conv (32) + BatchNorm + MaxPool	9,824	(None, 110, 110, 32)
Conv block 2	2×Conv (64) + BatchNorm + MaxPool	55,936	(None, 53, 53, 64)
Conv block 3	2×Conv (128) + BatchNorm + MaxPool	222,464	(None, 24, 24, 128)
Conv block 4	1×Conv (256) + BatchNorm + MaxPool	296,192	(None, 11, 11, 256)
Flatten	Flatten	0	(None, 30976)
Dense (fc1)	Dense (512)	15,860,224	(None, 512)
Dropout	Dropout (0.5)	0	(None, 512)
Dense (fc2)	Dense (256)	131,328	(None, 256)
Dropout	Dropout (0.5)	0	(None, 256)
Output layer	Dense (4, softmax)	1,028	(None, 4)

Table 2. Layers of custom OvO baseline-CNN

Type	Description	Parameters	Output shape
Input layer	Input_1	0	(None, 224, 224, 1)
Conv2D	Conv2D (32, 3×3, ReLU)	320	(None, 222, 222, 32)
MaxPooling2D	MaxPool (2×2)	0	(None, 111, 111, 32)
Conv2D	Conv2D (64, 3×3, ReLU)	18,496	(None, 109, 109, 64)
MaxPooling2D	MaxPool (2×2)	0	(None, 54, 54, 64)
Conv2D	Conv2D (128, 3×3, ReLU)	73,856	(None, 52, 52, 128)
MaxPooling2D	MaxPool (2×2)	0	(None, 26, 26, 128)
Flatten	Flatten	0	(None, 86,528)
Dense (fc)	Dense (64, ReLU)	5,537,856	(None, 64)
Dropout	Dropout (0.5)	0	(None, 64)
Output layer	Dense (1, sigmoid)	65	(None, 1)

Table 3. Layers of custom OvO deep-batch normalization (BN)-CNN

Type	Description	Parameters	Output shape
Input layer	Input_1	0	(None, 224, 224, 1)
Conv2D	Conv2D (32, 3×3, ReLU)	320	(None, 222, 222, 32)
BatchNorm	BatchNormalization (32)	128	(None, 222, 222, 32)
MaxPooling2D	MaxPool (2×2)	0	(None, 111, 111, 32)
Conv2D	Conv2D (64, 3×3, ReLU)	18,496	(None, 109, 109, 64)
BatchNorm	BatchNormalization (64)	256	(None, 109, 109, 64)
MaxPooling2D	MaxPool (2×2)	0	(None, 54, 54, 64)
Conv2D	Conv2D (128, 3×3, ReLU)	73,856	(None, 52, 52, 128)
BatchNorm	BatchNormalization (128)	512	(None, 52, 52, 128)
MaxPooling2D	MaxPool (2×2)	0	(None, 26, 26, 128)
Conv2D	Conv2D (256, 3×3, ReLU) (extra convolutional block)	295,168	(None, 24, 24, 256)
MaxPooling2D	MaxPool (2×2)	0	(None, 12, 12, 256)
Flatten	Flatten	0	(None, 36,864)
Dense (fc1)	Dense (128, ReLU)	4,718,720	(None, 128)
BatchNorm	BatchNormalization (128)	512	(None, 128)
Dropout	Dropout (0.3)	0	(None, 128)
Dense (fc2)	Dense (64, ReLU)	8,256	(None, 64)
BatchNorm	BatchNormalization (64)	256	(None, 64)
Dropout	Dropout (0.3)	0	(None, 64)
Dense (fc3)	Dense (32, ReLU)	2,080	(None, 32)
Output layer	Dense (1, sigmoid)	33	(None, 1)

Table 4. Layers of custom OvR CNN

Type	Description	Parameters	Output shape
Input layer	Input_1	0	(None, 224, 224, 1)
Conv2D	Conv2D (32, 3×3)	320	(None, 222, 222, 32)
MaxPooling2D	MaxPool (2×2)	0	(None, 111, 111, 32)
Conv2D	Conv2D (64, 3×3)	18,496	(None, 109, 109, 64)
MaxPooling2D	MaxPool (2×2)	0	(None, 54, 54, 64)
Conv2D	Conv2D (128, 3×3)	73,856	(None, 52, 52, 128)
MaxPooling2D	MaxPool (2×2)	0	(None, 26, 26, 128)
Flatten	Flatten	0	(None, 86,528)
Dense (fc1)	Dense (512)	44,302,848	(None, 512)
Dropout	Dropout (0.5)	0	(None, 512)
Output layer	Dense (1, sigmoid)	513	(None, 1)

Reproducibility was ensured by standardizing the primary training configurations. A dropout rate of 0.5 was applied to reduce overfitting, and early stopping with a patience of ten epochs was used to promote stable convergence. Training was conducted using the Adam optimizer with a learning rate of 0.0001 and a batch size of 32, settings commonly reported as effective for MRI-based DL tasks. Hyperparameters were refined through exploratory grid searches that varied learning rate, dropout magnitude, and batch size to ensure stable optimization across all classification strategies.

2.5. Model evaluation

This study is evaluated using a set of confusion-matrix-based metrics that capture complementary aspects of classification performance. These include accuracy, precision (positive predictive value, PPV), recall (sensitivity), F1-score, specificity, false positive rate (FPR), false negative rate (FNR), and negative predictive value (NPV) [8]. Since the class distribution is balanced, overall accuracy is considered a reliable indicator of performance. In addition, we report precision, recall, F1-score, and specificity to provide complementary perspectives. In the multiclass setting, per-class precision, recall, and F1-score are computed by treating each class against the rest and then summarized using macro, weighted, or micro averages, consistent with standard classification reports and confusion-matrix analyses [49], [50]. The formulas used to compute these metrics are as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision (PPV) = \frac{TP}{TP+FP} \quad (2)$$

$$Recall / Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

$$FNR = \frac{FN}{FN+TP} \quad (7)$$

$$NPV = \frac{TN}{TN+FN} \quad (8)$$

3. RESULTS AND DISCUSSION

The dataset was first balanced through random undersampling, resulting in 6,464 images per class and a total of 25,856 samples. It was then split into 80% for training and 20% for testing. All images were processed using a standardized pipeline that consisted of grayscale conversion, resizing to 224×224 pixels, CLAHE-based contrast enhancement, bilateral filtering, and min–max normalization to the [0, 1] range. This procedure produced clearer and more consistent MRI images, with enhanced visibility of sulci and ventricles, reduced noise, and standardized dimensions. Representative examples of the preprocessed data are shown in Figure 4.

The training history of ResNet-152 is presented in Figure 5. Although the model was allocated 2000 epochs, training converged early at epoch 81, as the early stopping criterion was met. The model reached 72% training accuracy at the final epoch, while validation accuracy followed a similar trend with moderate fluctuations, reflecting sensitivity to data variability but no severe overfitting. The corresponding loss curves show a consistent decrease in training loss, with validation loss exhibiting higher variance but overall convergence with training loss.

The second model trained was the custom CNN, which employed a multiclass classification strategy. This model achieved the best performance with a test accuracy of 90%, confirming its effectiveness in handling the four-class Alzheimer's classification task. The training history shown in Figure 6 indicates that both training and validation accuracy stabilized at around 90%, while the loss curves converged smoothly without signs of overfitting.



Figure 4. Sample images after preprocessing steps

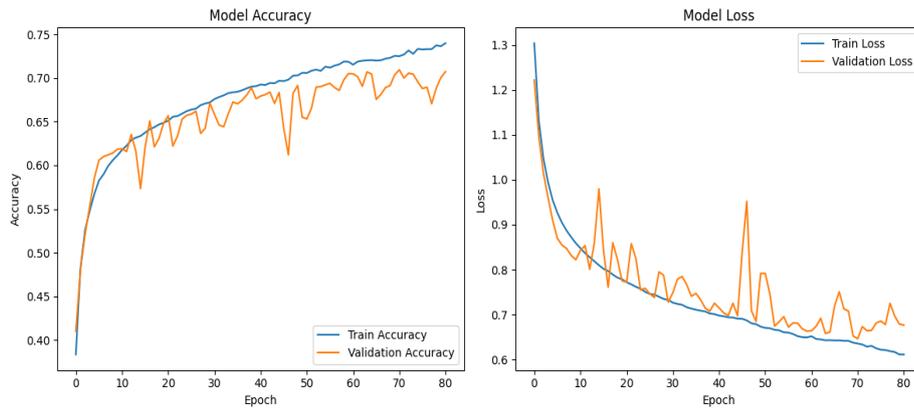


Figure 5. Training accuracy and loss curves of ResNet-152

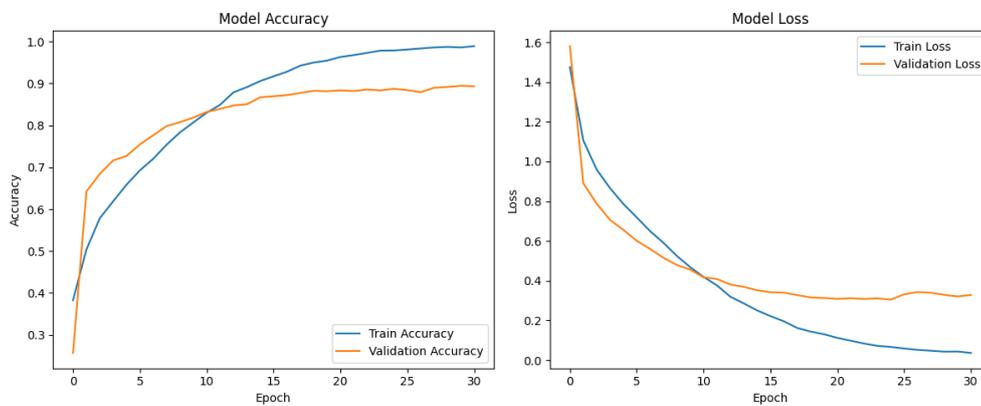


Figure 6. Training accuracy and loss curves of custom CNN multiclass

The third model trained was the custom CNN with an OvO classification strategy. In this approach, six independent binary CNNs were trained, each focusing on a specific pair of classes. The training histories of the six classifiers are shown in Figures 7(a)–(f): training accuracy and loss curves of custom CNN OvO. Most classifiers achieved stable convergence, with some pairs, such as ND vs. ModD and MD vs. ModD, reaching validation accuracies above 90%. However, classifiers involving early-stage categories, such as ND vs VMD or VMD vs MD, exhibited lower and more fluctuating validation accuracy, reflecting the inherent difficulty in distinguishing subtle changes at these stages. The classification report further supports these findings, with an overall accuracy of 85%. ND reached an F1-score of 93%, VMD 89%, MD 82%, and ModD 99%. The macro-average F1 was 73%, indicating variability across classes, while the weighted-average F1 reached 91%, confirming the strong overall performance of the OvO approach.

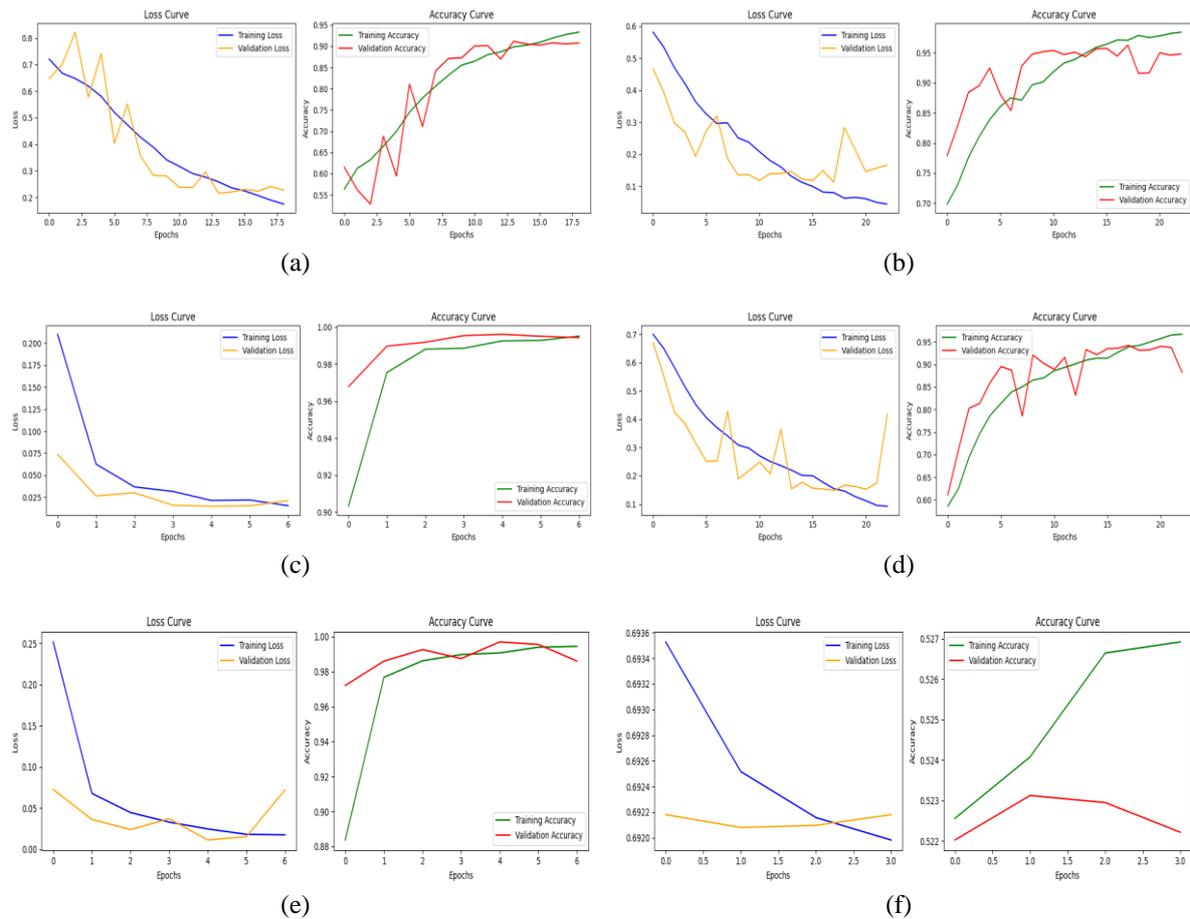


Figure 7. Training accuracy and loss curves of the six custom CNN OvO binary classifiers; (a) ND vs VMD, (b) ND vs ModD, (c) VMD vs ModD, (d) ND vs MD, (e) VMD vs MD, and (f) MD vs ModD

The final model developed was the custom CNN using an OvR strategy, where four binary classifiers were trained independently, each isolating one target class from the remaining categories. The training behaviors of these models are illustrated in Figures 8(a)–(d). Figures 8(a) and (b), representing the ND vs rest and ModD vs rest classifiers, show the most stable convergence, with training and validation accuracies consistently exceeding 85%. In contrast, Figure 8(c) for the MD vs rest classifier exhibits greater variability, with noticeable fluctuations in validation accuracy, indicating the challenges in correctly separating MDs from neighboring classes. Meanwhile, Figure 8(d), corresponding to the VMD vs Rest classifier, demonstrates moderate performance and reveals sensitivity to class imbalance and overlapping early-stage features. These observations highlight the varying complexity across diagnostic boundaries within the OvR framework.

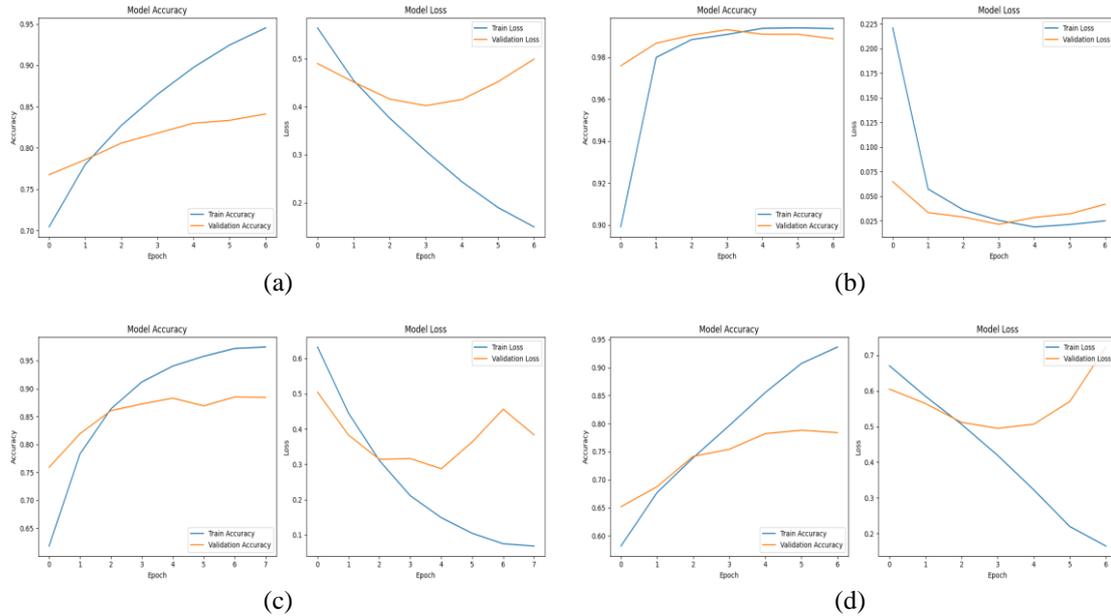


Figure 8. Training accuracy and loss curves of the four custom CNN OvR classifiers; (a) ND vs rest, (b) ModD vs rest, (c) MD vs rest, and (d) VMD vs rest

Although the custom multiclass CNN showed the strongest and most consistent performance, the confusion matrices reveal why the OvO and OvR models performed poorly on early-stage classes. Both approaches frequently misclassified VMD and MD, which reflects the subtle morphological differences between these adjacent stages and their overlap with normal aging. The OvR models tended to favor the dominant “Rest” class, reducing sensitivity to early-stage cases, while the OvO models produced inconsistent pairwise outputs that accumulated during voting and increased the likelihood of error propagation. These observations indicate that decomposition-based strategies are more vulnerable to boundary ambiguity than a unified multiclass approach. Applying statistical significance tests such as McNemar’s test or paired bootstrap comparisons would help confirm whether the performance differences between models are statistically meaningful.

A consolidated comparison across all models is presented to evaluate overall diagnostic performance and error characteristics. Table 5 summarizes the results and highlights the differences among the multiclass, OvO, OvR, and ResNet-152 approaches.

Table 5. Results performance evaluation

Model	Approach	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	FPR (%)	FNR (%)	NPV (%)
ResNet-152	Multiclass	72	72	72	71	91	9	28	91
Custom CNN	Multiclass	90	90	90	90	97	3	10	97
Custom CNN	OvO	85	78	85	90	97	3	15	96
Custom CNN	OvR	81	71	81	84	95	5	19	95

This study presented a stage-aware DL pipeline for AD classification on structural MRI, comparing a transfer-learning baseline ResNet-152 with three proposed classifiers: custom CNN multiclass, custom CNN OvO, and custom CNN OvR. On the test set, the custom multiclass CNN delivered the strongest and most balanced performance, achieving 90% accuracy with consistent precision, recall, and F1 scores across the non-demented, very mild, mild, and moderate stages. This result surpassed the performance of OvO at 85%, OvR at 81%, and the baseline ResNet-152 at 72%. Model reliability was further supported by stable validation trends, with fluctuations of approximately $\pm 2\%$ for the custom multiclass CNN, $\pm 3\%$ for the OvO model, $\pm 4\%$ for the OvR model, and $\pm 5\%$ for ResNet-152. Collectively, these findings indicate that the multiclass formulation is particularly effective for early-stage Alzheimer’s detection, where subtle structural differences often resemble normal aging and introduce subjectivity in clinical assessment.

When interpreted in the context of existing MRI-based Alzheimer’s classification research, these results show a clear advancement. Prior studies using YOLOv5, VGG16, DenseNet-169, or ResNet-50

typically report accuracies ranging from 61 to 82% [19], [22], [24]. and multi-stage models often struggle to separate adjacent stages along the clinical dementia rating scale [51]. In contrast, the proposed custom multiclass CNN demonstrates stronger stage-specific discrimination, providing empirical evidence that a tailored architecture can leverage disease-related structural patterns more effectively than conventional transfer-learning backbones. This improvement in stage-aware performance represents a key novelty of the present work, as it shows that a dedicated multiclass architecture can simultaneously enhance accuracy, reduce inter-class confusion, and improve early-stage detection sensitivity.

The clinical relevance of these gains is considerable. Early identification of very mild and MD categories is difficult even for experienced radiologists, yet the proposed model captured these borderline distinctions with greater reliability. This capability can support radiologists by reducing inter-rater variability, flagging subtle morphological changes for re-evaluation, and prioritizing patients for additional biomarker testing or timely initiation of disease-modifying therapy [41]. Nonetheless, several challenges remain for real-world deployment. Models trained on a single dataset may fail to generalize across scanners or populations due to domain shift, and reliance solely on structural MRI may overlook early functional or metabolic abnormalities detectable by PET or functional magnetic resonance imaging (fMRI) [52], [53]. Future work should therefore examine cross-dataset training, domain adaptation, and multimodal fusion, and should explore hybrid or ensemble architectures combined with explainable AI to further improve robustness and clinical safety.

4. CONCLUSION

This study introduced a stage-aware DL framework for AD classification using structural MRI, demonstrating that the proposed custom multiclass CNN achieved 90% accuracy and outperformed both the OvO, OvR models, and the ResNet-152 transfer-learning baseline. These findings highlight the model's ability to capture subtle morphological differences across Alzheimer's stages and its clinical relevance for improving early-stage detection, where symptoms often mimic normal aging. Nonetheless, several limitations remain, including reliance on a single MRI dataset, potential domain shift across scanners or populations, and the exclusive use of structural MRI, which may limit sensitivity to early pathological changes. Future research should incorporate multimodal imaging such as PET or fMRI, integrate clinical and cognitive biomarkers, perform cross-dataset or multicenter validation, and explore hybrid or ensemble architectures to enhance robustness. With further refinement and external validation, the proposed framework has the potential to serve as a reliable decision-support tool that strengthens diagnostic consistency and facilitates earlier intervention in clinical practice.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to STMIK TIME for institutional support and the research facilities provided during the completion of this study. This work was supported by the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Education, Culture, Research, and Technology of Indonesia under Research Grant No. 122/C3/DT.05.00/PL/2025 and Subcontract No. 69/SPK/LL1/AL.04.03/PL/2025.

FUNDING INFORMATION

This research received financial support from the Ministry of Education, Culture, Research, and Technology of Indonesia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kelvin Leonardi	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Kohsasih														
Octara Pribadi	✓	✓				✓		✓	✓	✓	✓	✓	✓	✓
Andy	✓		✓	✓	✓		✓		✓	✓	✓		✓	✓
Daniel Smith Sunario	✓	✓			✓	✓	✓	✓	✓	✓		✓		

C : C onceptualization	I : I nterpretation	Vi : V isualization
M : M ethodology	R : R esources	Su : S upervision
So : S oftware	D : D ata Curation	P : P roject administration
Va : V alidation	O : O riginal Draft	Fu : F unding acquisition
Fo : F ormal analysis	E : E diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [initials, KLK], upon reasonable request.

REFERENCES

- [1] H. Zhang and Y. Qie, "Applying deep learning to medical imaging: A review," *Applied Sciences*, vol. 13, no. 18, Sep. 2023, doi: 10.3390/app131810521.
- [2] B. S. Rao and M. Aparna, "A review on Alzheimer's disease through analysis of MRI images using deep learning techniques," *IEEE Access*, vol. 11, pp. 71542–71556, 2023, doi: 10.1109/ACCESS.2023.3294981.
- [3] S. M. Plis *et al.*, "Deep learning for neuroimaging: a validation study," *Frontiers in Neuroscience*, vol. 8, Aug. 2014, doi: 10.3389/fnins.2014.00229.
- [4] W. Yan *et al.*, "Deep learning in neuroimaging: promises and challenges," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 87–98, Mar. 2022, doi: 10.1109/MSP.2021.3128348.
- [5] M. G. Alsubaie, S. Luo, and K. Shaukat, "Alzheimer's disease detection using deep learning on neuroimaging: A systematic review," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 464–505, Feb. 2024, doi: 10.3390/make6010024.
- [6] M. Al-Zharani, S. I. Ansarullah, M. S. Al-Eissa, G. M. Dar, R. A. Alqahtani, and S. Alkahtani, "Exploring the efficacy of deep learning techniques in detecting and diagnosing Alzheimer's disease: A comparative study," *Journal of Disability Research*, vol. 3, no. 6, 2024, doi: 10.57197/JDR-2024-0064.
- [7] S. Pallawi and D. K. Singh, "Review and analysis of deep neural network models for Alzheimer's disease classification using brain medical resonance imaging," *Cognitive Computation and Systems*, vol. 5, no. 1, pp. 1–13, Mar. 2023, doi: 10.1049/ccs2.12072.
- [8] M. Kaya and Y. Çetin-Kaya, "A novel deep learning architecture optimization for multiclass classification of Alzheimer's disease level," *IEEE Access*, vol. 12, pp. 46562–46581, 2024, doi: 10.1109/ACCESS.2024.3382947.
- [9] L.-K. Huang, Y.-C. Kuan, H.-W. Lin, and C.-J. Hu, "Clinical trials of new drugs for Alzheimer disease: a 2020–2023 update," *Journal of Biomedical Science*, vol. 30, no. 1, Oct. 2023, doi: 10.1186/s12929-023-00976-6.
- [10] J. R. Sims *et al.*, "Donanemab in early symptomatic Alzheimer disease," *JAMA*, vol. 330, no. 6, pp. 512–527, Aug. 2023, doi: 10.1001/jama.2023.13239.
- [11] D. V. Puri *et al.*, "LEADNet: Detection of Alzheimer's disease using spatiotemporal EEG analysis and low-complexity CNN," *IEEE Access*, vol. 12, pp. 113888–113897, 2024, doi: 10.1109/ACCESS.2024.3435768.
- [12] Z. Yao *et al.*, "Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images," *European Journal of Radiology*, vol. 165, Aug. 2023, doi: 10.1016/j.ejrad.2023.110934.
- [13] Y. Pusparani *et al.*, "Deep learning applications in MRI-based detection of the hippocampal region for Alzheimer's diagnosis," *IEEE Access*, vol. 12, pp. 103830–103838, 2024, doi: 10.1109/ACCESS.2024.3426085.
- [14] M. Pafitis, M. Constantinou, and C. Christodoulou, "Accelerating training of convolutional neural networks with hessian-free optimization for detecting Alzheimer's disease in brain MRI," *IEEE Access*, vol. 12, pp. 176184–176198, 2024, doi: 10.1109/ACCESS.2024.3487114.
- [15] S. Li, H. Qu, X. Dong, B. Dang, H. Zang, and Y. Gong, "Leveraging deep learning and xception architecture for high-accuracy MRI classification in Alzheimer diagnosis," *Prepr. arXiv.2403.16212*, Mar. 2024.
- [16] I. Sarasua, S. Pölsterl, and C. Wachinger, "Hippocampal representations for deep learning on Alzheimer's disease," *Scientific Reports*, vol. 12, no. 1, May 2022, doi: 10.1038/s41598-022-12533-6.
- [17] M. Mujahid, A. Rehman, T. Alam, F. S. Alamri, S. M. Fati, and T. Saba, "An efficient ensemble approach for Alzheimer's disease detection using an adaptive synthetic technique and deep learning," *Diagnostics*, vol. 13, no. 15, Jul. 2023, doi: 10.3390/diagnostics13152489.
- [18] G. N. Alwakid, S. Tahir, M. Humayun, and W. Gouda, "Improving Alzheimer's detection with deep learning and image processing techniques," *IEEE Access*, vol. 12, pp. 153445–153456, 2024, doi: 10.1109/ACCESS.2024.3481238.
- [19] A. Alshahrani, J. Mustafa, M. Almatrafi, L. Albaqami, R. Aljabri, and S. Almuntashri, "A comparative study of deep learning techniques for Alzheimer's disease detection in medical radiography," *IJCSNS International Journal of Computer Science and Network Security*, vol. 24, no. 5, pp. 53–63, 2024.
- [20] F. M. J. M. Shamrat *et al.*, "AlzheimerNet: An effective deep learning based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images," *IEEE Access*, vol. 11, pp. 16376–16395, 2023, doi: 10.1109/ACCESS.2023.3244952.
- [21] Y. Ding *et al.*, "A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain," *Radiology*, vol. 290, no. 2, pp. 456–464, Feb. 2019, doi: 10.1148/radiol.2018180958.
- [22] W. Al Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Computer Science*, vol. 8, Dec. 2022, doi: 10.7717/peerj-cs.1177.
- [23] M. Odusami, R. Maskeliūnas, R. Damaševičius, and S. Misra, "Explainable deep-learning-based diagnosis of alzheimer's disease using multimodal input fusion of PET and MRI images," *Journal of Medical and Biological Engineering*, vol. 43, no. 3, pp. 291–

- 302, Jun. 2023, doi: 10.1007/s40846-023-00801-3.
- [24] I. Thayumanasamy and K. Ramamurthy, "Performance analysis of machine learning and deep learning models for classification of Alzheimer's disease from brain MRI," *Traitement du Signal*, vol. 39, no. 6, pp. 1961–1970, Dec. 2022, doi: 10.18280/ts.390608.
- [25] A. Mora-Rubio, M. A. Bravo-Ortíz, S. Q. Arredondo, J. M. S. Torres, G. A. Ruz, and R. Tabares-Soto, "Classification of Alzheimer's disease stages from magnetic resonance images using deep learning," *PeerJ Computer Science*, vol. 9, Aug. 2023, doi: 10.7717/peerj-cs.1490.
- [26] B. Yousefimehr *et al.*, "Data balancing strategies: A survey of resampling and augmentation methods," *Prepr. arXiv.2505.13518*, May 2025.
- [27] X. Gao *et al.*, "A comprehensive survey on imbalanced data learning," *Prepr. arXiv.2502.08960*, Sep. 2025, doi: 10.1007/s11704-025-50274-7.
- [28] R. Kurniawan, S. Samsuryadi, and F. S. Mohamad, "Improved classification of arsenic-affected skin diseases through image processing and transfer learning," *CommIT (Communication and Information Technology) Journal*, vol. 19, no. 1, pp. 71–88, Apr. 2025, doi: 10.21512/commit.v19i1.11891.
- [29] F. Islam, M. H. Rahman, Nurjahan, M. S. Hossain, and S. Ahmed, "A novel method for diagnosing Alzheimer's disease from MRI scans using the ResNet50 feature extractor and the SVM classifier," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 1233–1242, 2023, doi: 10.14569/IJACSA.2023.01406131.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [31] S. S. Rhomadhon and D. R. Ningtias, "Developing a classification system for brain tumors using the ResNet152V2 CNN model architecture," *Journal of Soft Computing Exploration*, vol. 5, no. 2, pp. 173–182, Jun. 2024, doi: 10.52465/josce.v5i2.372.
- [32] P. K. Pandey, J. Pruthi, S. Alzahrani, A. Verma, and B. Zohra, "Enhancing healthcare recommendation: transfer learning in deep convolutional neural networks for Alzheimer disease detection," *Frontiers in Medicine*, vol. 11, Sep. 2024, doi: 10.3389/fmed.2024.1445325.
- [33] X. Yang *et al.*, "An ensemble-based 3D residual network for the classification of Alzheimer's disease," *PLOS One*, vol. 20, no. 6, Jun. 2025, doi: 10.1371/journal.pone.0324520.
- [34] G. I. Stoleru and A. Iftene, "Transfer learning for Alzheimer's disease diagnosis from MRI slices: A comparative study of deep learning models," *Procedia Computer Science*, vol. 225, pp. 2614–2623, 2023, doi: 10.1016/j.procs.2023.10.253.
- [35] D. Ozkan *et al.*, "Deep learning techniques for automated dementia diagnosis using neuroimaging modalities: A systematic review," *IEEE Access*, vol. 12, pp. 127879–127902, 2024, doi: 10.1109/ACCESS.2024.3454709.
- [36] P. Carcagni, M. Leo, M. Del Coco, C. Distanto, and A. De Salve, "Convolution neural networks and self-attention learners for Alzheimer dementia diagnosis from brain MRI," *Sensors*, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031694.
- [37] A. D. Arya *et al.*, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," *Brain Informatics*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40708-023-00195-7.
- [38] M. EL-Geneedy, H. E.-D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," *Alexandria Engineering Journal*, vol. 63, pp. 211–221, Jan. 2023, doi: 10.1016/j.aej.2022.07.062.
- [39] M. I. Nazir, A. Akter, M. A. H. Wadud, and M. A. Uddin, "Utilizing customized CNN for brain tumor prediction with explainable AI," *Heliyon*, vol. 10, no. 20, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38997.
- [40] E. Albalawi *et al.*, "Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach," *Frontiers in Computational Neuroscience*, vol. 18, Jun. 2024, doi: 10.3389/fncom.2024.1418546.
- [41] 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," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-53733-6.
- [42] S. Zolfaghari, A. Joudaki, and Y. Sarbaz, "A hybrid learning approach for MRI-based detection of alzheimer's disease stages using dual CNNs and ensemble classifier," *Scientific Reports*, vol. 15, no. 1, Jul. 2025, doi: 10.1038/s41598-025-11743-y.
- [43] S. B. Shahid *et al.*, "Novel deep learning for multi-class classification of Alzheimer's in disability using MRI datasets," *Frontiers in Bioinformatics*, vol. 5, Aug. 2025, doi: 10.3389/fbinf.2025.1567219.
- [44] G. Psaltakis, K. Rogdakis, M. Loizos, and E. Kymakis, "One-vs-one, one-vs-rest, and a novel outcome-driven one-vs-one binary classifiers enabled by optoelectronic memristors towards overcoming hardware limitations in multiclass classification," *Discover Materials*, vol. 4, no. 1, Mar. 2024, doi: 10.1007/s43939-024-00077-7.
- [45] G. R. Kumar, G. Thippanna, and M. Venkateswarlu, "Multiclass date fruit prediction using SVM and logistic regression with OVO and OVR," *Journal of Big Data Technology and Business Analytics*, vol. 3, no. 3, pp. 27–35, 2024.
- [46] Z. Lu *et al.*, "Chemistry combining elemental profile, stable isotopic ratios, and chemometrics for fine classification of a chinese herb licorice (*glycyrrhiza uralensis fisch.*) from 37 producing area," *Journal of Analytical Methods in Chemistry*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/8906305.
- [47] G. R. Kumar, G. Thippanna, and K. P. Kumar, "Unveiling the effectiveness of multi-classification algorithms: A comprehensive investigation," *Journal of Computer Science and System Software*, vol. 1, no. 2, pp. 13–17, Oct. 2024, doi: 10.48001/jocsss.2024.1213-17.
- [48] B. Şener, K. Acici, and E. Sümer, "Categorization of Alzheimer's disease stages using deep learning approaches with McNemar's test," *PeerJ Computer Science*, vol. 10, Feb. 2024, doi: 10.7717/peerj-cs.1877.
- [49] H. Yan, V. Mubonanyikuzo, T. E. Komolafe, L. Zhou, T. Wu, and N. Wang, "Hybrid-RViT: Hybridizing ResNet-50 and vision transformer for enhanced Alzheimer's disease detection," *PLOS ONE*, vol. 20, no. 2, Feb. 2025, doi: 10.1371/journal.pone.0318998.
- [50] M. A. Sahid, M. P. Uddin, H. Saha, and M. R. Islam, "Towards privacy-preserving Alzheimer's disease classification: Federated learning on T1-weighted magnetic resonance imaging data," *Digital Health*, vol. 10, Jan. 2024, doi: 10.1177/20552076241295577.
- [51] R. G. Al Rahbani, A. Ioannou, and T. Wang, "Alzheimer's disease multiclass detection through deep learning models and post-processing heuristics," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 12, no. 1, Dec. 2024, doi: 10.1080/21681163.2024.2383219.
- [52] D. Lteif, S. Sreerama, S. A. Bargal, B. A. Plummer, R. Au, and V. B. Kolachalama, "Disease-driven domain generalization for neuroimaging-based assessment of Alzheimer's disease," *Human Brain Mapping*, vol. 45, no. 8, Jun. 2024, doi: 10.1002/hbm.26707.
- [53] H. Cai, Q. Zhang, and Y. Long, "Prototype-guided multi-scale domain adaptation for Alzheimer's disease detection," *Computers in Biology and Medicine*, vol. 154, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106570.

BIOGRAPHIES OF AUTHORS

Kelvin Leonardi Kohsasih    received the B.Sc. degree in Computer Science from the College of Information Technology and Computer Management TIME (STMIK TIME), Medan, Indonesia, in 2021, and the M.Sc. degree in Computer Science from the University of Potensi Utama (UPU), Medan, Indonesia, in 2023. He is currently a permanent lecturer with the Department of Informatics, STMIK TIME, Medan, Indonesia. His current research interests include machine learning, data mining, image processing, and cloud computing. He can be contacted at email: kelvinleonardi@stmik-time.ac.id.



Octara Pribadi    received the B.Sc. degree in Informatics Engineering from the College of Information Technology and Computer Management TIME (STMIK TIME), Medan, Indonesia, in 2011, and the M.Sc. degree in Informatics Engineering from the University of Sumatera Utara (USU), Medan, Indonesia, in 2015. He is currently pursuing a Ph.D. degree in Computer Science at Dian Nuswantoro University, Semarang, Indonesia. He is a permanent lecturer with the Department of Informatics, STMIK TIME, Medan, where he also serves as the Head of the Institute for Research and Community Service (LPPM). His current research interests include computer security, data science, and machine learning. He can be contacted at email: octarapribadi@stmik-time.ac.id.



Andy    received the B.Ec. degree in Economics from Sekolah Tinggi Ilmu Ekonomi Harapan, Medan, Indonesia, in 2004, the B.Sc. degree in Computer Science from the College of Information Technology and Computer Management TIME (STMIK TIME), Medan, Indonesia, in 2015, and the M.Sc. degree in Informatics Engineering from Bina Nusantara University (Binus), Jakarta, Indonesia, in 2018. He is currently a permanent lecturer with the Department of Informatics, STMIK TIME, Medan. His current research interests include intelligent systems, artificial intelligence, and machine learning. He can be contacted at email: andy@stmik-time.ac.id.



Daniel Smith Sunario    is currently pursuing the B.Sc. degree in Informatics Engineering with the College of Information Technology and Computer Management TIME (STMIK TIME), Medan, Indonesia. His research interests include artificial intelligence, intelligent systems, and machine learning. He can be contacted at email: danielsmithsunario@stmik-time.ac.id.