Type of Contribution:
Research Paper
Review Paper
Case Study

ENERGY: JURNAL ILMIAH ILMU-ILMU TEKNIK
Vol. 15, No. 2 (2025) pp 213-226
DOI: 10.51747/energy.v15i2.15208

E-ISSN: 2962-2565

This article contributes to:

Exploring 3D Convolutional Neural Network Models for Alzheimer's Disease Classification Based on 3D MRI Images

Titus Batlayeri[1], **Subairi**[1*], **Rahman Arifuddin**[1], **Bagas Martinus Rianu**[2]

[1] Department of Electrical Engineering, Merdeka University of Malang, 65146, **Indonesia**

[2] Magister Program of Electrical Engineering, Kanazawa University, 920-1192, **Japan**

*subairi@unmer.ac.id

Abstract

Alzheimer's disease is a common form of progressive dementia, especially among the elderly, and is characterized by a decline in cognitive function. Classifying this disease using 3D brain imaging through MRI is challenging due to the complexity of the data and the similarity of features across classes. This study develops a classification model based on a 3D Convolutional Neural Network (3D CNN) architecture, specifically using ResNet-18. The dataset used is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), consisting of 1,281 samples evenly distributed across three classes: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). The data undergo several preprocessing steps, including skull stripping, normalization, and augmentation. The model is tested in two configurations: without dropout and with a dropout rate of 0.3. The results show that the model with dropout performs better, achieving a classification accuracy of 62.0% and a macro F1-score of 0.604. The model outperforms ADNet and Vision Transformer, and approaches the accuracy of Vision Mamba. Nevertheless, this approach still requires further development, particularly in improving accuracy for the CN class and reducing performance imbalance across classes.

Keywords: Alzheimer; Classification; MRI; 3D CNN; ResNet-18

Article Info
Submitted: 2025-09-13
Revised: 2025-11-28
Accepted: 2025-11-29
Published: 2025-11-30

Publisher
Universitas Panca Marga

1. Introduction

Dementia is a chronic neurodegenerative disorder that progresses gradually and has a significant impact, causing a decline in cognitive abilities [1]. According to statistical data from the World Health Organization (WHO), the prevalence of

dementia is increasing globally. In 2016, dementia was the fifth leading cause of death worldwide [2]. Alzheimer's disease is the most common type of dementia found among people worldwide, particularly in the elderly population [3]. Patients experience cognitive decline caused by nerve damage in the brain, which begins in the hippocampus—the area responsible for memory formation. As the disease progresses, the damage spreads to other brain regions, resulting in substantial brain tissue shrinkage [4]. Additionally, the effects include short-term memory loss, difficulty communicating, impaired problem-solving skills, and reduced ability to care for oneself, all of which can interfere with daily activities. This disorder typically occurs in individuals over the age of 65 [5].

According to the World Alzheimer Report 2018 by Alzheimer's Disease International, a new case of dementia is estimated to occur every 3 seconds worldwide. As a result, in 2018, there were 50 million people globally living with Alzheimer's and other types of dementia, and this number is projected to triple to 152 million by 2050 [6].

In Indonesia, the Ministry of Health estimated that there were approximately 1 million individuals with Alzheimer's in 2013, a number expected to double by 2030 and quadruple by 2050 [7]. Age is the primary contributing factor, with prevalence increasing significantly in individuals over 65. Other contributing factors include education level, cardiovascular disease, mild cognitive impairment, traumatic brain injury, and limited social interaction [8]. Alzheimer's symptoms develop gradually and are difficult to diagnose; therefore, rapid and accurate detection is crucial for appropriate treatment [9]. Currently, Alzheimer's diagnosis is commonly performed using clinical tests and brain imaging analysis with Magnetic Resonance Imaging (MRI). Although effective, manual examination is time-consuming and requires medical expertise. It is also prone to inter-observer variability, particularly in assessing the severity of early-stage Alzheimer's.

With advancements in machine learning, automated image-based approaches such as the 3D Convolutional Neural Network (3D CNN) have become relevant. CNNs are known for their ability to process and analyze complex, high-dimensional data, commonly applied to 2D image inputs [10], [11]. However, for three-dimensional MRI data, a 3D CNN approach is more suitable, as it can capture complete spatial information.

This study aims to explore the use of a 3D Convolutional Neural Network (3D CNN) model to classify Alzheimer's disease based on 3D MRI images. The study compares the performance of the model in differentiating between three diagnostic classes: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Cognitively Normal (CN). It also evaluates the impact of data quantity, augmentation techniques, and regularization methods such as dropout on the

model's performance. The goal is to contribute to the development of a more accurate and efficient classification and detection system for Alzheimer's disease.

2. Methods

2.1 Preparation and Preprocessing

This study uses three-dimensional brain imaging data (3D MRI) obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database through the Image and Data Archive (IDA) system. The dataset consists of three diagnostic categories: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Cognitively Normal (CN). The MRI samples shown in **Figure 1** are anatomical scans acquired using the T1-weighted method, with an original resolution of 256 × 256 × 170 voxels.

This study includes two experiments that use the same dataset configuration, comprising a total of 1,281 samples evenly distributed across the three categories, with 427 MRI images for each class: AD, MCI, and CN. Before being used to train the 3D CNN model, all MRI data undergo preprocessing steps to standardize image quality and resolution, ensuring consistency across the dataset.

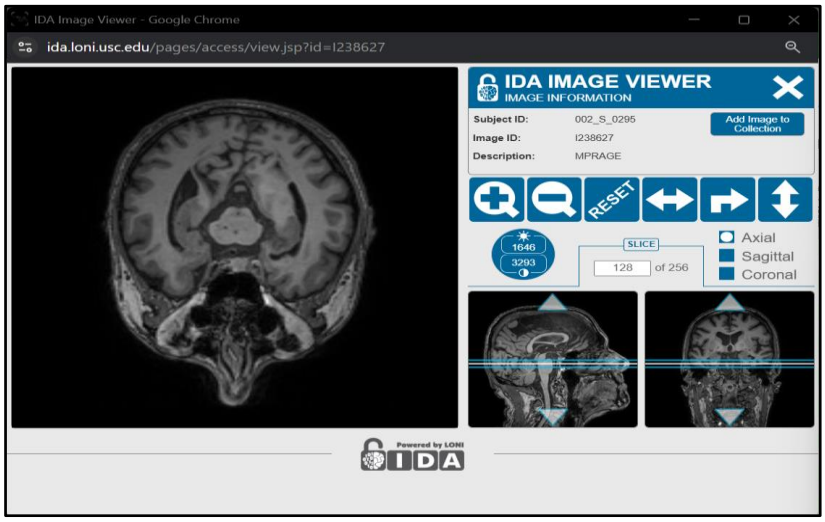

Figure 1. a 3D brain MRI scan acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) using T1-weighted imaging.

In the preprocessing stage, several methods are used to prepare the MRI images before being input into the model. The initial and crucial step is skull stripping. Skull stripping aims to remove non-brain components such as the skull and outer soft tissues, which are not needed for the classification process. This process is performed using FSL-BET (Brain Extraction Tool), which is part of the FMRIB Software Library (FSL) package.

As shown in **Figure 2(a)**, the original MRI image is processed using FSL-BET with the threshold value set to 0.5. This threshold functions as a boundary to

separate brain tissue from non-brain tissue. Additionally, the robust brain center estimation mode is enabled to improve the tool's adaptability to asymmetrical or tilted head positions. Furthermore, the parameter mask=True is applied to generate a binary mask, which is useful for visualization or further segmentation of the extracted brain area. The result of the skull stripping process is shown in **Figure 2(b)**.

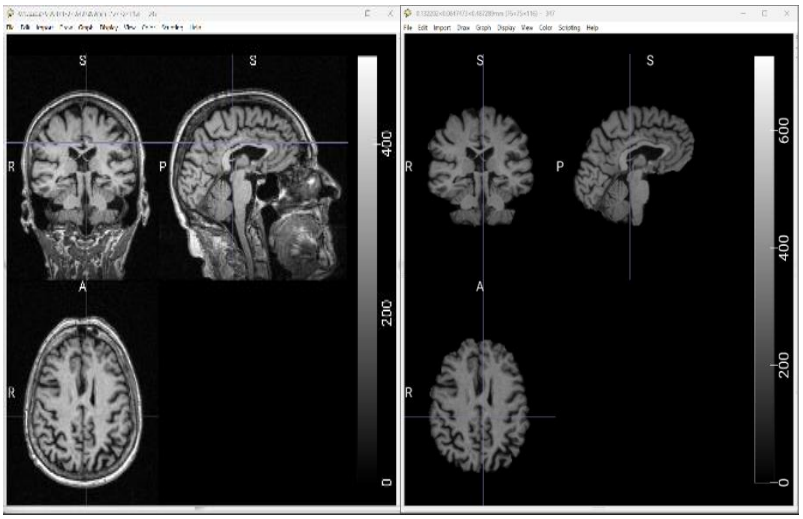

Figure 2. (a) Original brain MRI image before skull stripping using FSL-BET; (b) Brain MRI image after skull stripping processed with FSL-BET

After the skull-stripping process, the preprocessing stage continues with intensity normalization of the voxels into the range [0, 1]. This step aims to ensure a uniform scale of intensity values across all volumes in the dataset. Subsequently, the MRI images are resized to a dimension of 160×160×160 voxels. This adjustment is performed to standardize the input dimensions across samples while also reducing the computational load during the training process. Additionally, the channel dimension is adjusted by adding a single channel, resulting in a final format for each volume of (160, 160, 160, 1). The results of the preprocessing can be seen in **Figure 3**.

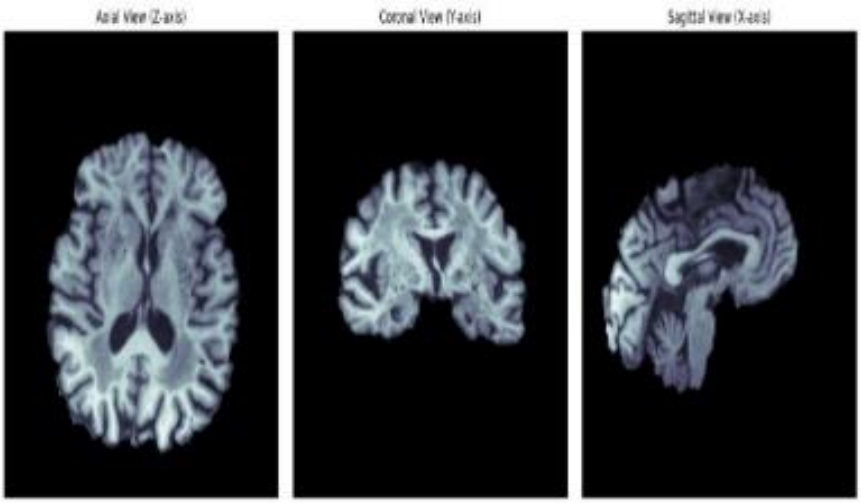

Figure 3. Example of a brain MRI image after preprocessing steps

The dataset is subsequently split into three subsets: 70% for training, 15% for validation, and 15% for testing. The distribution of samples across each diagnostic class—Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Cognitively Normal (CN)—is shown in **Tabel 1**.

Tabel 1. Data Allocation for Training, Validation, and Testing in This Study

Amount of data from Experiments 1 and 2				
Class	**Number Of Subjects**	**Training**	**Validation**	**Testing**
AD	427	299	64	64
MCI	427	299	64	64
CN	427	299	64	64
Total	1281	897	192	192

In addition to data splitting, an augmentation process is also applied, specifically to the training subset. The purpose of augmentation is to increase the spatial variability of brain images used by the model, allowing it to learn from a wider range of possible shapes and structures. The augmentation method used in this study is random rotation applied to the 3D MRI images. The rotation is performed on the transverse plane (XY plane), with rotation angles randomly selected from the following set of values: {−20°, −10°, −5°, 5°, 10°, 20°}.

2.2 Model Architecture

This study employs the 3D CNN ResNet-18 architecture, which is a three-dimensional variant of the Residual Network (ResNet-18). This model is well-suited for processing volumetric data such as MRI because it preserves spatial information across all three dimensions of brain images, while also offering advantages in training stability. ResNet utilizes residual connections that allow gradients to flow directly to the deeper layers, thereby minimizing the risk of vanishing gradients and accelerating convergence during deep network training [13].

This study includes two separate experiments using different model configurations, specifically in the application of dropout regularization, namely:

1. The first experiment uses the standard 3D ResNet-18 architecture without the addition of dropout. The architecture used in this experiment is shown in **Figure 4**. This configuration serves as the baseline to observe the model's performance without explicit regularization.

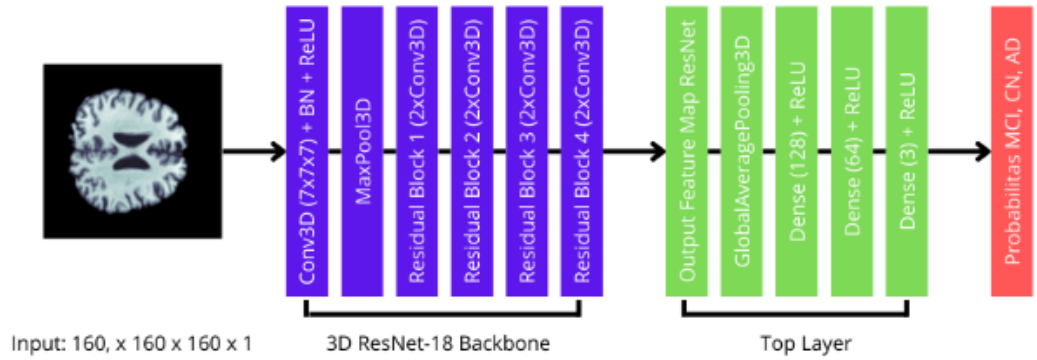

Figure 4. Example of a brain MRI image after preprocessing steps

2. The second experiment uses a 3D ResNet-18 architecture, as shown in **Figure 5**, which is modified by adding dropout layers. A dropout rate of 0.3 is applied after the GlobalAveragePooling3D layer and before the dense layer. Additionally, dropout is also inserted between the first and second dense layers. This modification is intended to reduce overfitting and improve model accuracy through stronger regularization, particularly considering the large number of parameters and the complexity of 3D MRI data.

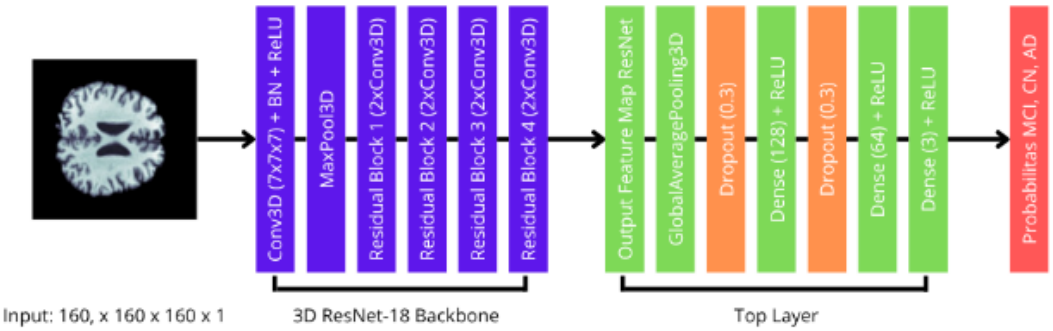

Figure 5. Architecture of the 3D CNN ResNet-18 Model with Dropout

Tabel 2. Training Hyperparameter Settings

Parameter	Mark
Optimizer	Adam
Learning Rate	1e-5 (with a minimum of 1e-8)
Loss Function	Categorical Crossentropy
Batch Size	1
Epoch Maksimum	100
Input Shape	(160, 160, 160, 1)
Number of output classes	3 (MCI, CN, AD)

2.3 Training Configuration

After designing the model architecture, the next step is to carry out the training process. The training is conducted using the preprocessed 3D MRI data. The model is trained using the TensorFlow and Keras frameworks, with hyperparameter configurations designed to support optimization stability and prevent overfitting. Both experiments in this study use the same training configuration. The detailed

training settings, including the initial learning rate, optimizer used, and maximum number of epochs, are presented in **Tabel 2**.

2.4 Model Validation and Evaluation

The model validation and evaluation process aims to measure how well the trained model can generalize to unseen data. Evaluation is performed on the test data subset after the training process is completed. All experiments in this study use the same evaluation approach, both in terms of the test data and the performance metrics applied. The model's performance is assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and the confusion matrix.

2.5 Software and Hardware

The programming language used in this study is Python, due to its flexibility, open-source nature, and broad support for libraries relevant to data processing and deep learning implementation. Several libraries used in this research include os, time, itertools, numpy, pandas, matplotlib, scikit-learn, and TensorFlow. TensorFlow is integrated with Keras as the API for building and training the deep learning model. TensorFlow also provides GPU support and tools for monitoring the training process.

All experiments in this study were conducted on the researcher's personal computer with the following hardware specifications: Intel Core i7-12700F processor, 16 GB RAM, and an NVIDIA RTX 3060 GPU with 16 GB of memory, supporting CUDA acceleration. This hardware configuration enables stable and efficient training of the 3D CNN model on volumetric MRI data, although the training process is relatively time-consuming due to the complexity of the data.

3.Results and Discussion

3.1 Experimental Results

This section presents the results of two experiments on Alzheimer's disease classification using the 3D CNN ResNet-18 model on 3D MRI images. The first experiment uses the baseline architecture without dropout, while the second experiment incorporates a dropout rate of 0.3 as a regularization technique.

3.1.1 First Experiment Results

The first experiment uses the baseline 3D ResNet-18 architecture without the addition of dropout, but is conducted on a balanced dataset consisting of 1,281 samples, with 427 samples for each class (CN, MCI, AD). The objective of this experiment is to evaluate the impact of increased data quantity and class balance on classification performance.

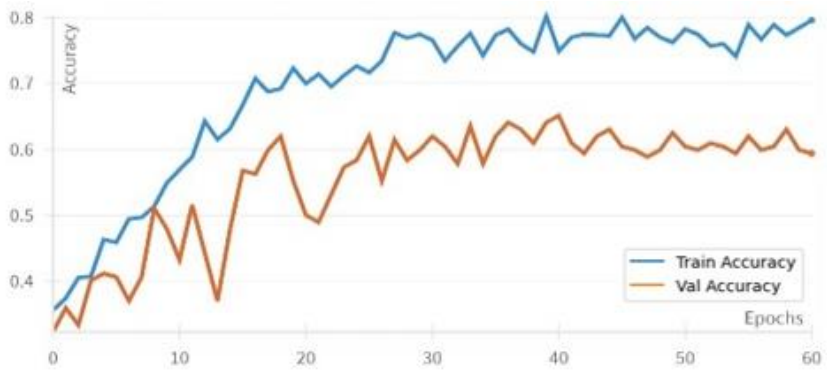

Figure 6. Training and Validation Accuracy Curve for the First Experiment

Accuracy and Loss

The results shown in Figure 6 indicate that the training accuracy increased steadily, approaching 80%, with the highest validation accuracy reaching 79%. Although there were some fluctuations, the validation accuracy remained relatively stable and did not experience a significant drop at the end of training. Meanwhile, as shown in **Figure 7**, the training loss consistently decreased, while the lowest validation loss was recorded at 0.86, indicating improved stability in the validation process compared to earlier trials. These findings suggest that increasing the quantity of data has a positive impact in reducing overfitting.

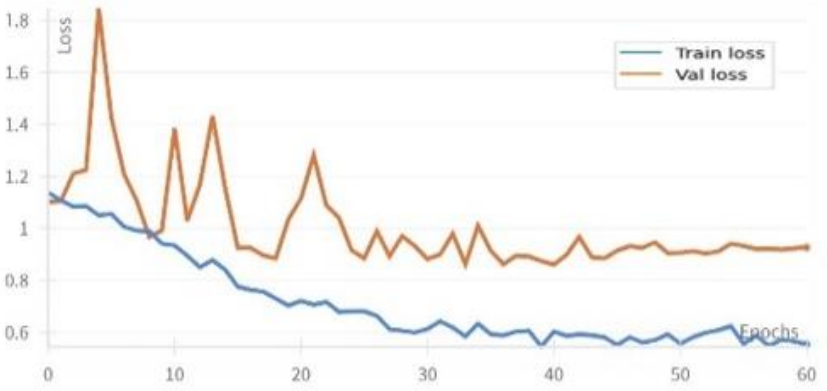

Figure 7. Training and Validation Loss Curve for the First Experiment

Confusion Matrix for the First Experiment

Based on **Tabel 3**, the model achieved an overall accuracy of 55.2%, with a macro F1-score of 0.550. The best performance was obtained in the MCI class (label 1), with a recall value of 0.641, while the CN class (label 0) showed the lowest performance, with a recall of 0.469. These results indicate that the model is more sensitive in recognizing the MCI class but still struggles to distinguish CN images from the other classes.

Tabel 3. Classification Performance Metrics for the First Experiment

Class	Precision	Recall	F1-Score	Support
CN (0)	0.508	0.469	0.488	64
MCI (1)	0.569	0.641	0.603	64
AD (2)	0.574	0.574	0.56	64
Macro Avg	0.551	0.552	0.55	192
Weighted Avg	0.551	0.552	0.55	192
Accuracy			0.552	192

Confusion Matrix Analysis

During the field trial, several issues commonly faced by local farmers were confirmed, including water turbidity, fluctuating pH, and low oxygen levels caused by temperature changes. These findings validate the system's practical relevance for real-world aquaculture operations.

Figure 8. Confusion Matrix for the First Experiment

Based on the results in **Figure 8**, the model correctly identified 41 out of 64 MCI samples, indicating improved performance in this class. Meanwhile, 35 Alzheimer's Disease (AD) samples and 30 Cognitively Normal (CN) samples were correctly classified. However, misclassifications still frequently occurred between the CN and MCI classes.

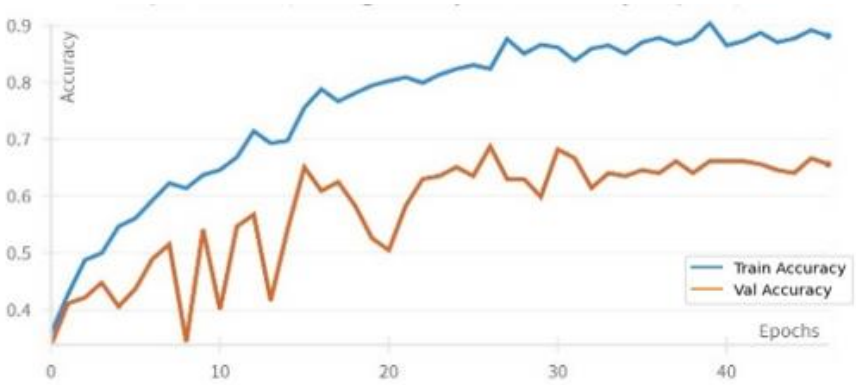

Figure 9. Training and Validation Accuracy Curve for the Second Experiment

3.1.2 Second Experiment Results

The second experiment uses the baseline 3D ResNet-18 architecture with the addition of dropout, conducted on a balanced dataset consisting of 1,281 samples, with 427 samples for each class (CN, MCI, and AD). The purpose of this experiment is to evaluate the impact of increased data quantity and class balance on classification performance

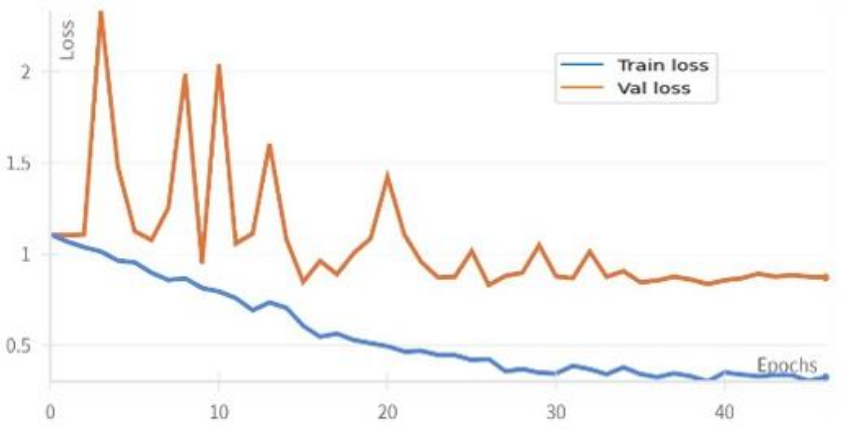

Figure 10. Training and Validation Loss Curve for the Second Experiment

Training and Validation Accuracy and Loss

Figure 9 shows a well-structured training graph, with training accuracy steadily increasing to nearly 90%, and a peak validation accuracy reaching 79%, which is the highest achieved compared to the previous experiment. The training loss decreased significantly, while the validation loss demonstrated greater stability, with relatively smaller fluctuations than those observed in the previous experiment, as shown in **Figure 10**.

Tabel 4. Classification Performance Metrics for the Second Experiment

Class	Precision	Recall	F1-Score	Support
CN (0)	0.508	0.469	0.488	64
MCI (1)	0.569	0.641	0.603	64
AD (2)	0.574	0.574	0.56	64
Macro Avg	0.551	0.552	0.55	192
Weighted Avg	0.551	0.552	0.55	192
Accuracy			0.552	192

Classification Performance Metrics

Based on **Tabel 4**, the evaluation on the test data shows that the model achieved an overall accuracy of 62%, with a macro F1-score of 0.604. The best performance was achieved in the MCI class (label 1), with a recall of 0.797 and the highest F1-score among the three classes. In contrast, the CN class (label 0) had the lowest recall, at only 0.359, although its precision was relatively high.

Figure 11. Training and Validation Loss Curve for the Second Experiment

Classification Performance Metrics

Based on **Figure 11**, the model successfully classified 51 out of 64 MCI samples and 45 out of 64 AD samples correctly. Performance on the CN class remains a challenge, with only 23 samples correctly classified, while the rest were misclassified into the MCI and AD classes. This indicates that although dropout improves the model's overall generalization ability, the identification of CN images still requires further attention, both in terms of model architecture and data representation. In general, this experiment demonstrates that the addition of dropout enhances validation stability and test accuracy, making it the best configuration among the two experiments conducted.

3.2 Comparative Analysis of Experimental Results

Based on **Tabel 5**, it can be concluded that the addition of dropout had a positive impact on the overall performance of the model. The total accuracy increased from 55.2% to 62.0%, and the macro F1-score also improved from 55.0% to 60.4%, indicating that the model in Experiment 2 achieved better and more balanced generalization across all classes. Additionally, the model's performance for each individual class also showed significant changes.

Tabel 5 Performance Comparison of Experiment 1 and Experiment 2

Aspect	Experiment 1	Experiment 2
Total Accuracy	55.20%	62.00%
Macro F1-Score	55.00%	60.40%
MCI (1) Recall	64.10%	79.70%
CN (0) Recall	46.90%	35.90%
AD (2) Recall	54.70%	70.30%

In the MCI class (class 1), the recall increased dramatically from 64.1% to 79.7%, indicating that the addition of dropout made the model more sensitive in recognizing MCI symptoms, which is the most challenging category in Alzheimer's classification. An improvement was also observed in the AD class (class 2), where recall increased from 54.7% to 70.3%, demonstrating that the model became more capable of accurately detecting Alzheimer's cases after applying regularization.

However, the model's performance in the CN class (class 0) experienced a decline in recall, dropping from 46.9% to 35.9%. This may be due to the model redistributing its attention more toward the MCI and AD classes, making the CN class features less dominant. Nevertheless, this trade-off is still acceptable considering the significant improvements in the other two classes, which are critical for early Alzheimer's diagnosis.

Overall, these results indicate that dropout plays an important role in improving model stability and generalization, especially for classes that were previously difficult to identify. The second experiment can therefore be considered the best-performing model configuration in this study.

3.3 Comparative Analysis with Prior Research Findings

A comparison with previous research is conducted to provide insight into the performance of the model developed in this study. The comparison is performed indirectly, based on the accuracy values reported in the literature using similar methods and the same dataset (ADNI).

Tabel 6. Model Comparison

Metode	Accuracy	Dataset
Vision Mamba	0.65	ADNI
Visions Transformer	0.46	ADNI
ADNet	0.52	ADNI
3DCNN ResNet-18 (ours)	0.62	ADNI

As shown in **Tabel 6**, the 3D CNN ResNet-18 model with the addition of dropout in this study outperformed the Vision Transformer model from the study "Classification of Alzheimer's Disease using Vision Transformers with MRI and FDG-PET" [14], as well as the ADNet model from the study "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI" [11]. Furthermore, its performance approached that of the Vision Mamba model presented in the study "Vision Mamba: Cutting-Edge Classification of Alzheimer's Disease with 3D MRI Scans" [12].

4. Conclusion

This study developed an Alzheimer's classification model based on 3D MRI images using the 3D ResNet-18 architecture, targeting three classes: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). Two experiments were conducted with different configurations—one without dropout (Experiment 1) and one with a dropout rate of 0.3 (Experiment 2). Experiment 1 achieved a test accuracy of 55.2% and a macro F1-score of 0.550 but still showed signs of overfitting, particularly in the CN class. In contrast, Experiment 2 demonstrated improved performance after applying dropout, achieving an accuracy of 62.0%, a macro F1-score of 0.604, and the highest recall in the MCI (79.7%) and AD (70.3%) classes.

These results confirm that the addition of dropout improves the model's generalization ability, especially in distinguishing classes with complex patterns such as MCI and AD. Although challenges remain in accurately identifying the CN class, the model configuration used in Experiment 2 can be considered the most optimal in this study.

Authors' Declaration

Authors' contributions and responsibilities - The authors made substantial contributions to the conception and design of the study. The authors took responsibility for data analysis, interpretation, and discussion of results. The authors read and approved the final manuscript.

Funding - No funding information from the authors.

Availability of data and materials - All data is available from the authors.

Competing interests - The authors declare no competing interest.

Additional information - No additional information from the authors.

References

[1] H. N. Kasprata and H. S. Harahap, "Literature Review: Diagnosis and Management of Alzheimer's Dementia," Jurnal Ilmu Kedokteran dan Kesehatan, vol. 10, no. 12, pp. 3471–3479, 2023.

[2] G. H. E., "Deaths by cause, age, sex, by country and by region, 2000–2016," 2018.

[3] J. S. Birks and R. J. Harvey, "Donepezil for dementia due to Alzheimer's disease," Cochrane Database of Systematic Reviews, no. 6, 2018.

[4] R. M. Gemiralda and M. Marlaokta, "Neuroprotective effect of turmeric in Alzheimer's patients," Jurnal Ilmu Keperawatan Jiwa, vol. 2, no. 3, pp. 171–178, 2019.

[5] C. Patterson, "World Alzheimer Report 2018," The State of the Art of Dementia Research: New Frontiers, 2018.

[6] Indonesian Ministry of Health, "Healthy Elderly, Elderly Free from Dementia," 2016. [Online]. Available: http://www.depkes.go.id/article/print/16031000003

[7] Alzheimer's Association, "2016 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 12, no. 4, pp. 459–509, 2016.

[8] G. Mirzaei and H. Adeli, "Machine learning techniques for diagnosis of Alzheimer's disease, mild cognitive disorder, and other types of dementia," Biomedical Signal Processing and Control, vol. 72, p. 103293, 2022.

[9] M. A. Wirya, "Alzheimer's Disease Detection on Magnetic Resonance Imaging Using Machine Learning with Convolutional Neural Network Method," Undergraduate Thesis, UIN Syarif Hidayatullah Jakarta, 2023.

[10] M. R. Efrian and U. Latifa, "Image recognition using convolutional neural network (CNN) for human skin disease detection," Power Elektronik: Jurnal Orang Elektro, vol. 11, no. 2, pp. 276–282, 2022.

[11] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI," Frontiers in Bioengineering and Biotechnology, vol. 8, Oct. 2020, doi: 10.3389/fbioe.2020.534592.

[12] M. K. A., A. Gurung, and P. Ranjan, "Vision Mamba: Cutting-Edge Classification of Alzheimer's Disease with 3D MRI Scans," Jun. 2024.

[13] F. Ramzan et al., "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks," Journal of Medical Systems, vol. 44, no. 2, p. 37, Feb. 2020, doi: 10.1007/s10916-019-1475-2.

[14] H. Shin, S. Jeon, Y. Seol, S. Kim, and D. Kang, "Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images," Applied Sciences, vol. 13, no. 6, p. 3453, Mar. 2023, doi: 10.3390/app13063453.