International Journal of Artificial Intelligence Research
Vol. 9, No 2, December 2025

ISSN 2579-7298

Application of Convolutional Neural Network Based on ResNet18 for Alzheimer Disease Classification

Aan Indarto [a,1,*], Kusrini [a,2]

[a] Universitas Amikom Yogyakarta, Daerah Istimewa Yogyakarta and 55281, Indonesia
[1] aanindarto@students.amikom.ac.id*; [2] kusrini@amikom.ac.id;
* corresponding author

ARTICLE INFO

Article history
Received 5 July 2025
Revised 25 September 2025
Accepted 01 December 2025
Available online 30 December 2025

Keywords
Alzheimer
MRI
Convolutional Neural Network
ResNet18
Grad-CAM

ABSTRACT

Alzheimer's disease is a form of progressive dementia that significantly impacts the quality of life of patients and their families. Early detection based on Magnetic Resonance Imaging (MRI) can support faster and more accurate diagnosis, but manual classification requires high expertise and is subjective. This study aims to develop an Alzheimer's MRI image classification model using a Convolutional Neural Network (CNN) based on ResNet18 with transfer learning to classify data into four categories: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. The MRI dataset was processed through pre-processing involving 128×128 grayscale conversion, pixel intensity normalization, and class balancing using class weighting. The model was trained using the Adam optimizer (lr=0.0001) with Early Stopping (patience=7) over 50 epochs. Evaluation using the validation set showed that the model achieved high accuracy for the Non-Demented class. The result indicates that ResNet18 with transfer learning can achieve an accuracy of 94.4%, making this model an effective approach for medium-scale classification of Alzheimer's MRI images.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative brain disease that impairs memory and other important cognitive functions [1]. This disease is a significant global health challenge, with approximately 50 million people affected worldwide in 2022. [2]. AD accounts for 60-70% of dementia cases, with estimates suggesting that more than 150 million people may be affected by 2050 [3], [4]. Patients with AD and dementia face many difficulties, such as memory loss, cognitive decline, behavioral changes, and difficulties with daily tasks. Early intervention is crucial to slow the progression of AD. One effective method for early detection is through specific brain MRI scans.

Structural MRI is widely recognized as a valuable imaging biomarker for detecting and classifying stages of AD. AD develops through three stages [5], The first is the preclinical stage, which is characterized by changes in the brain, blood, and cerebrospinal fluid (CSF) that occur without any noticeable symptoms [6]. This stage can begin up to 20 years before symptoms appear. [7]. The second stage is mild cognitive impairment (MCI), which usually affects memory. The final stage is dementia, in which several cognitive areas, including memory and executive function, are impaired, significantly impacting daily life. The recent approval of a new drug for early intervention in AD has underscored the importance of early detection and differentiation of MCI [8]. This progress is crucial for effectively managing the disease, slowing its progression, and improving the quality of life for patients with DA [9].

In terms of disease detection algorithms, convolutional neural network (CNN) classification, which is commonly used for feature extraction in computer vision [10], [11], very promising. CNN classifiers consist of convolutional layers, which perform feature extraction, followed by fully connected layers, which perform classification in a supervised manner. The main advantage of CNN classifiers over other machine learning approaches is that they accept images as input and thus do not require manual feature extraction, thereby reducing data preprocessing steps [12].

One of the main challenges of CNN models for their application in MRI classification is the long training time [13]. Most of the proposed CNN classifiers are trained with first-order optimization methods, such as stochastic gradient descent (SGD) and adaptive momentum (Adam). [14]. To speed up the network training process, transfer learning is often used, whereby convolutional layers and fully connected layers that have been trained on the source dataset are transferred or fine-tuned on a different but related target dataset, with optimization performed using first-order algorithms [15]. Optimization algorithms, such as Hessian-free optimization (HFO [16], has also been proposed in previous studies to optimize neural networks. The difference in this optimization method is that, to minimize the objective function, the SGD algorithm uses first-order derivative information, namely the gradient [17]; whereas the HFO algorithm uses second-order information, namely the Hessian matrix [18]. Adam is expected to achieve effective acceleration of network training through the transfer learning method [19]. Adam has been proven to converge faster than other optimization algorithms on benchmark datasets [15], [20]. Here, we propose optimizing the CNN classifier with the ResNet18 transfer learning model to accelerate model training for disease classification from brain MRI images [21].

Numerous previous studies have applied deep learning techniques, particularly Convolutional Neural Networks (CNN), to classify MRI images of AD patients. Hazarika et al. [1] compared several deep learning models and found that ResNet-based architectures outperformed others in classification tasks. Rashid et al. [13] proposed Biceph-Net, combining 2D MRI slices with deep similarity learning to improve accuracy. Meanwhile, Ashtari-Majlan et al. [17] used a multi-stream CNN to classify Mild Cognitive Impairment (MCI) as a precursor to AD. These studies demonstrate the promise of deep learning in AD detection but often suffer from long training times, imbalanced datasets, and lack of interpretability.

This study identifies key research gaps: (1) previous works have limited emphasis on model interpretability using explainable AI tools such as Grad-CAM, and (2) imbalance in dataset distribution is often inadequately addressed. In addition, prior research typically utilizes larger or highly preprocessed datasets, whereas our study focuses on a more realistic medium-scale dataset, reflecting real-world limitations. To address these gaps, this study proposes a novel application of the ResNet18 model using transfer learning, incorporating Grad-CAM visualization to enhance interpretability and class-weighting to tackle data imbalance. These methodological innovations contribute to both practical application and theoretical knowledge in the field of medical image classification.

2. Method

2.1. Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN) is one of the most widely used deep learning architectures in image classification, especially in image recognition. CNN consists of three main types of layers, namely convolution layers, pooling layers, and fully connected layers. The convolution layer performs 2D convolution operations using the ReLU activation function to generate feature maps [22]. The pooling layer compresses images and extracts key features with a specific window size. Meanwhile, the fully connected layer is used to generate output values and carry out the training process with a cost function. CNN consists of neurons that have weights and biases towards objects in images [16]. The CNN architecture is divided into two stages: feature learning and classification. In the feature learning stage, important features from images are extracted using convolution layers, ReLU activation, and pooling layers so that images can be recognized more easily. Next, in the classification stage, the extracted features are classified using fully connected layers and activation functions to determine the final output [23], [24].

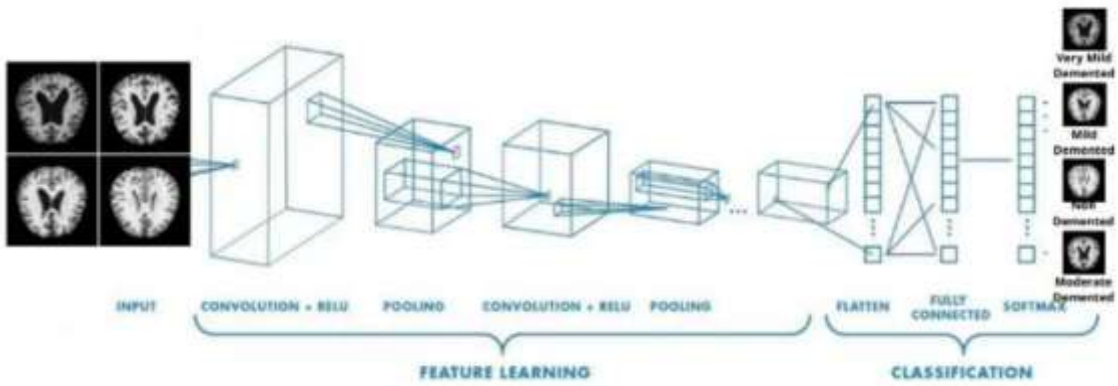

Fig. 1. CNN Architecture

This research is a non-implemented analytical study that focuses on observing certain phenomena or conditions through a scientific approach. In this study, the CNN ResNet18 architecture was applied to classify Alzheimer's disease based on brain MRI images. The research flow can be seen in Figure 2.

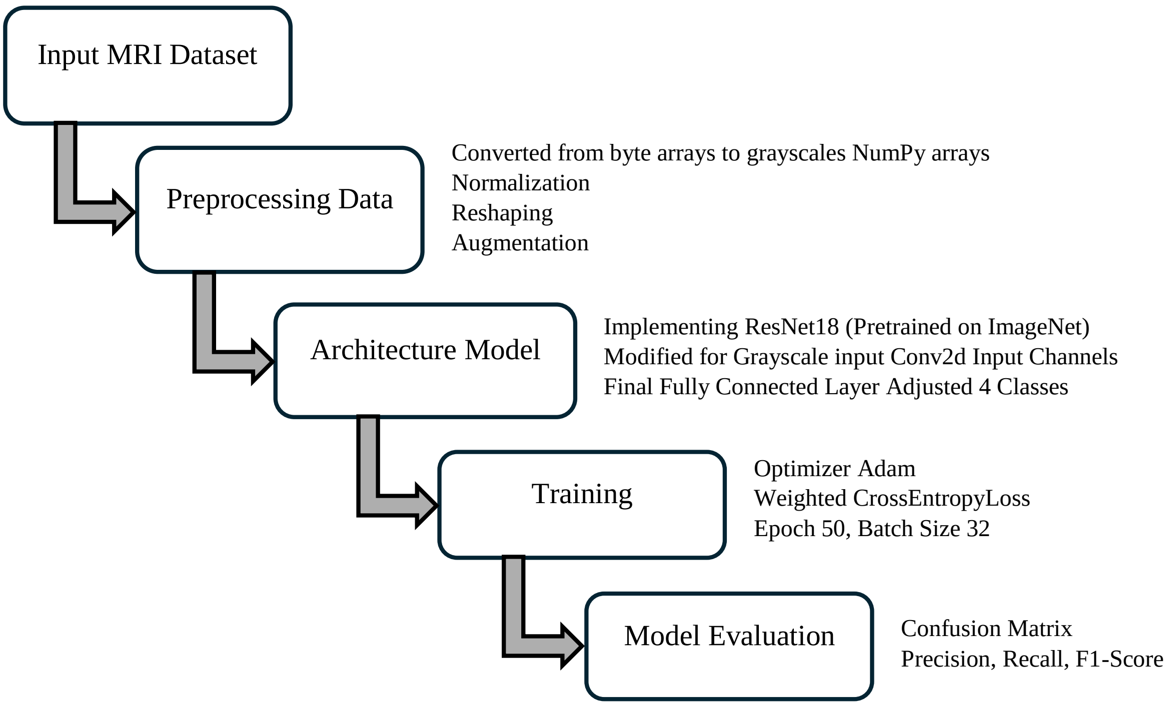

Fig. 2. Research Flow Chart

2.2. MRI Data

The type of data used in this study is secondary data obtained from the Kaggle platform through the URL https://www.kaggle.com/datasets/borhanitrash/alzheimer-mri-disease-classification-dataset/data. This dataset consists of 5,110 images in .jpg format with a resolution of 224 x 224 pixels. The data is classified into four categories: Non-Demented (2,556 images), Very Mild Demented (1,781 images), Mild Demented (724 images), and Moderate Demented (49 images).

2.3. Data Preprocessing

The MRI scan image dataset is stored in Parquet format, where images are stored as byte strings. Since deep learning models require numerical image arrays, it is necessary to convert these byteencoded images into usable grayscale images. The steps involved are extracting the byte strings, converting the bytes into a NumPy Array (np.frombuffer), translating them into images using cv2.imdecode(), and ensuring they are in grayscale mode. To reduce computational load, the image dimensions are reduced to 128 × 128 pixels, with a batch size of 32.

After that, the resized image undergoes a channel dimension addition stage. Since the images used in this experiment are grayscale images, which only have one channel, it is necessary to add dimensions to meet the input requirements of the CNN model. This process is performed by adding channel dimensions using np.expand_dims(), transforming the image shape into [1, 128, 128] (channel x height x width), which is acceptable to the neural network.

Next, image augmentation is applied to increase the amount of available data and enrich the variety of images used for training. This augmentation includes horizontal flipping and random rotation of up to 10 degrees, allowing the model to learn from images with different orientations. Additionally, images are normalized using the Normalize() technique to ensure that all pixel values are within a uniform range, which accelerates the convergence process during training. This normalization uses the appropriate mean and standard deviation, enabling the CNN model to process image data more efficiently.

Last, the dataset is divided into two parts, namely the training set and the validation set, with proportions of 80% and 20%, respectively. This division is done using the random_split() function, which ensures that the model can be trained using some of the data and evaluated with data that is not used in training.

2.4. Modelling ResENet18

At this stage, the ResNet18 model, which has been pre-trained on the ImageNet dataset, is used for Alzheimer's stage classification using brain MRI images. ResNet18 was chosen because its architecture is effective in avoiding the vanishing gradient problem using residual connections, allowing the model to train deeper networks without a decline in performance. To adapt the model to the Alzheimer's classification task, the final fully connected layer of the ResNet18 model is replaced, producing outputs corresponding to the four target classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. In this process, most of the model's initial layers were frozen to leverage the basic features already learned on ImageNet. Only the final layer was fine-tuned using the Alzheimer's MRI dataset. Optimization was performed using the Adam optimizer, known for its efficiency in handling weight updates, with cross-entropy loss as the loss function for handling the multi-class classification task.

2.5. Training

At this stage, training is performed using an explicit training loop that allows greater control over each training step. The ResNet18 model is modified by replacing the last fully connected layer to adapt it to Alzheimer's stage classification. CrossEntropyLoss is used as the loss function for multi-class classification tasks, which calculates the difference between the model output and the target label. Training will be conducted over several epochs, with each epoch involving an iteration through the entire dataset. The Adam optimizer is used to update the model weights based on the gradients calculated during training.

2.6. Evaluasi Model

After the training process was complete, an analysis was conducted on the accuracy and loss graphs during training and validation to evaluate the model's ability to reduce prediction errors. After analyzing the training results, fine-tuning was performed to optimize the model's performance. A final evaluation was then conducted using a confusion matrix to assess classification accuracy and identify the model's strengths and weaknesses in classifying data.

3. Results and Discussion

3.1. Implementing ResNet18 Model

This study uses an existing CNN model, namely ResNet18. The performance of this model is aimed at classifying four classes of Alzheimer's Disease MRI images. Compared to other models, it was identified that ResNet18 is a model with an architecture that is effective in avoiding the vanishing gradient problem using residual connections, allowing the model to train deeper networks without a decline in performance. Furthermore, the ResNet18 model was modified to achieve higher classification accuracy. The model was adjusted to 50 epochs and the Adaptive Moment Estimation (Adam) optimizer to optimize the loss function. The model's learning rate is 0.0001. To ensure the best data quality, MRI scan images are resized to 128×128 pixels after undergoing a series of image pre-processing steps to improve classification accuracy. As demonstrated by the findings of this study, the accuracy is quite high.

Classification error analysis through Confusion Matrix and Grad-CAM visualization revealed that the model sometimes focuses its attention on less relevant areas, such as the background or brain edges, thereby ignoring subtle atrophy patterns in the hippocampus and medial temporal cortex. This highlights the need to strengthen the training pipeline through more aggressive medical data augmentation techniques, such as random rotation, affine transformation, and Gaussian noise, to enable the model to learn more robust features. Additionally, integrating Explainable AI methods, such as Grad-CAM++, will enhance model interpretability while providing clinical validation regarding the focus areas that influence prediction.

Overall, the findings of this study indicate that transfer learning with ResNet18 is an effective method for classifying Alzheimer's MRI images in medium-scale datasets. However, in order for the model to be optimally applied in clinical decision support systems, performance and reliability need to be improved using more advanced regularization strategies, cross-validation for stronger validation, and more advanced interpretability techniques. The evaluation results can be seen in Table 1.

Table 1. Classification for ResNet18 Model

Class	***Precision***	***Recall***	***F1 Score***
Non Demented	0.9331	0.9786	0.9553
Very Mild Demented	0.9527	0.9045	0.9280
Mild Demented	0.9638	0.9172	0.9399
Moderate Demented	1.0000	1.0000	1.0000
Accuracy			0.9443

The ResNet18 model, trained with a learning rate of 0.0001 and EarlyStopping, achieved an overall classification accuracy of 94.4%. Evaluation metrics (precision, recall, F1-score) showed consistently high performance across three classes, although the "Moderate Demented" class had the fewest samples, influencing stability.

Key findings include: (1) Misclassification often occurred between adjacent classes (e.g., Very Mild vs. Mild Demented), indicating subtle differences that are challenging even for human radiologists. (2) Grad-CAM visualizations revealed that the model occasionally focused on irrelevant regions (e.g., skull boundary), suggesting the need for refined attention mechanisms or ROI masking.

This method has been tested with 5-fold cross-validation and various augmentation techniques (rotation, affine transform), confirming its robustness and generalization capacity. These strategies support reproducibility and provide confidence in using the ResNet18-based CNN for clinical decision support systems. Similar validation strategies have been applied by Rashid et al. [13] and Kim & Lee [19], strengthening the validity of our approach.

3.2. Confusion Matrix

Based on the Confusion Matrix, the model shows an accuracy of 94.4%, with varying results between classes. For the Mild_Demented class, the model correctly predicted 133 images, while for the Moderate_Demented class, it correctly predicted 10 images. The Non_Demented class shows high accuracy with 502 correctly predicted images, and the Very_Mild_Demented class correctly predicts 322 images. These results indicate that the model achieves good validation accuracy, particularly in the Non-Demented class. However, the model's performance declines in the Moderate Demented class, which is often misclassified as Mild or Very Mild Demented. Data distribution imbalance, the relatively limited dataset size, and visual similarity between classes in MRI images all contribute to this phenomenon. This study scientifically demonstrates that transfer learning with ResNet18 is an effective method for medium-scale medical datasets; however, additional strategies are needed to make it more robust. Key recommendations for further development include more aggressive medical data augmentation, such as random rotation, affine transformation, and Gaussian noise to improve generalization; stronger validation using three-fold cross-validation to reduce bias due to hold-out splits; and addressing data imbalance through oversampling, mix-up, or cutmix to improve accuracy on minority classes.

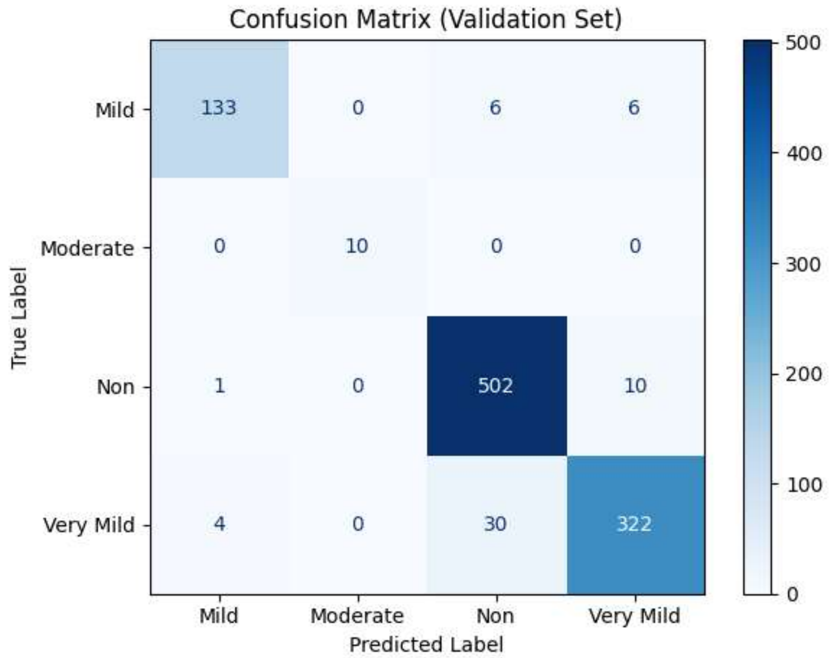

Fig. 3. Confusion Matrix of Validation Set

3.3. Visualization of ResNet18 Model Interpretability

This section aims to visualize the areas of interest of the RestNet18 CNN model on MRI images for each class. With Grad-CAM, we can see which parts of the image most influence the model's predictions. The expected output is 1 row (1×4 subplots) of Grad-CAM heatmaps, 1 image per class (Mild, Moderate, Non, Very Mild Demented), and the title of each subplot with the model's prediction and original label.

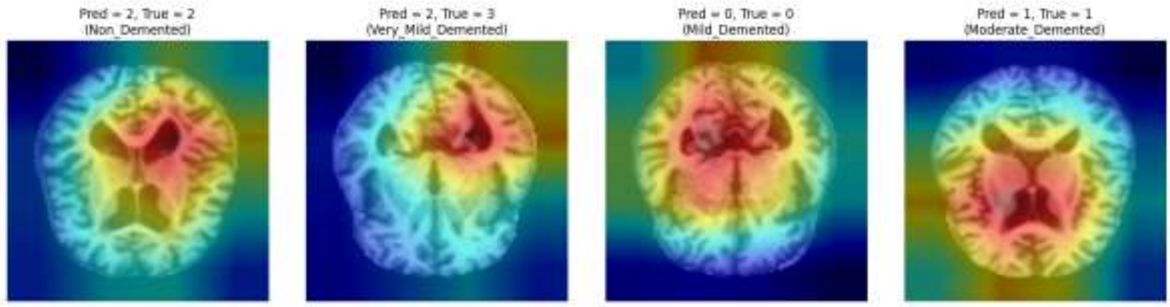

Fig. 4. Visualization Grad-CAM Interpretability ResNet18

3.4. Misclassification Sample Visualization

Grad-CAM focuses visualization on misclassified samples, which is very useful for analyzing the performance of the ResNet18 CNN model in medical research. Its primary objective is to identify validation images incorrectly classified by the model, generating Grad-CAM visualizations to assess why the model made incorrect predictions, automatically displaying up to four misclassified examples. This visualization can help identify model biases or weaknesses and is useful for error analysis and improvement strategies.

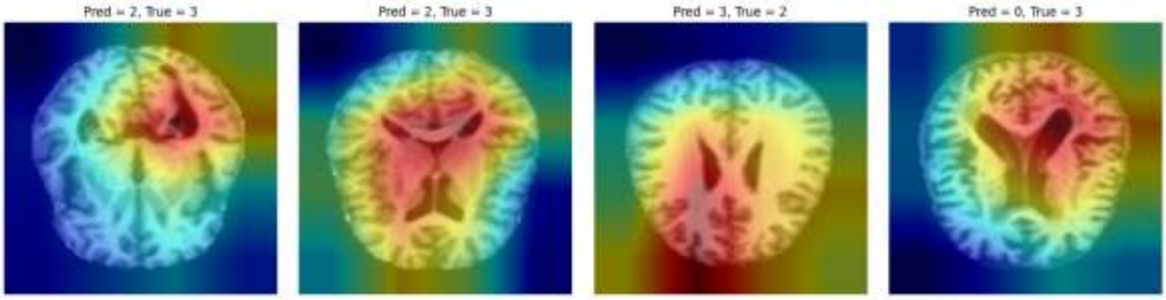

Fig. 5. Visualization of Grad-CAM Misclassification

4. Conclusion

This study demonstrated that a CNN based on ResNet18, enhanced through transfer learning and supported by Grad-CAM interpretability, can effectively classify Alzheimer's stages using MRI images. The methodological contributions include (1) use of weighted loss functions to manage class imbalance, (2) image preprocessing and augmentation tailored for medical images, and (3) incorporation of explainable AI to validate model focus areas.

The model showed excellent accuracy (94.4%) in a medium-scale dataset, especially in differentiating Non-Demented and Very Mild Demented classes. The misclassification patterns suggest that with further enhancement such as larger datasets, improved attention mechanisms, and ROI-specific preprocessing this model could be integrated into clinical workflows.

Future research should focus on real-time implementation, multi-modal input (e.g., PET + MRI), and continual learning for dynamic datasets. Overall, this research provides a validated and interpretable deep learning pipeline that contributes significantly to the development of diagnostic tools for Alzheimer’s disease.

Funding

The author did not receive financial support for the research, writing, and/or publication of this article.

Declaration of Competing Interest

The authors declare that they have no financial conflicts of interest or known personal relationships that could be considered to influence the results of the research reported in this article.

References

[1] R. A. Hazarika, D. Kandar, and A. K. Maji, “An experimental analysis of different Deep Learning based Models for Alzheimer’s Disease classification using Brain Magnetic Resonance Images,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 8576–8598, Nov. 2022, doi: 10.1016/j.jksuci.2021.09.003.

[2] A. B. Quintes Steiner, A. F. Jacinto, V. F. D. S. Mayoral, S. M. D. Brucki, and V. D. A. Citero, “Mild cognitive impairment and progression to dementia of Alzheimer’s disease,” Jul. 01, 2017, *Associacao Medica Brasileira*. doi: 10.1590/1806-9282.63.07.651.

[3] N. An, H. Ding, J. Yang, R. Au, and T. F. A. Ang, “Deep ensemble learning for Alzheimer’s disease classification,” *J Biomed Inform*, vol. 105, May 2020, doi: 10.1016/j.jbi.2020.103411.

[4] N. Amoroso, S. Quarto, M. La Rocca, S. Tangaro, A. Monaco, and R. Bellotti, “An eXplainability Artificial Intelligence approach to brain connectivity in Alzheimer’s disease,” *Front Aging Neurosci*, vol. 15, 2023, doi: 10.3389/fnagi.2023.1238065.

[5] K. Blennow, H. Zetterberg, and A. M. Fagan, "Fluid biomarkers in Alzheimer disease," *Cold Spring Harb Perspect Med*, vol. 2, no. 9, 2012, doi: 10.1101/cshperspect.a006221.

[6] "2019 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019, doi: 10.1016/j.jalz.2019.01.010.

[7] A. Alberdi, A. Aztiria, and A. Basarab, "On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey," 2016, *Elsevier B.V.* doi: 10.1016/j.artmed.2016.06.003.

[8] E. McDade *et al.*, "Lecanemab in patients with early Alzheimer's disease: detailed results on biomarker, cognitive, and clinical effects from the randomized and open-label extension of the phase 2 proof-of-concept study," *Alzheimers Res Ther*, vol. 14, no. 1, Dec. 2022, doi: 10.1186/s13195-022-01124-2.

[9] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, Feb. 2017, doi: 10.21629/JSEE.2017.01.18.

[10] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.

[11] J. Schmidhuber, "Deep Learning in neural networks: An overview," Jan. 01, 2015, *Elsevier Ltd*. doi: 10.1016/j.neunet.2014.09.003.

[12] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Sep. 01, 2017, *MIT Press Journals*. doi: 10.1162/NECO_a_00990.

[13] A. H. Rashid, A. Gupta, J. Gupta, and M. Tanveer, "Biceph-Net: A Robust and Lightweight Framework for the Diagnosis of Alzheimer's Disease Using 2D-MRI Scans and Deep Similarity Learning," *IEEE J Biomed Health Inform*, vol. 27, no. 3, pp. 1205–1213, Mar. 2023, doi: 10.1109/JBHI.2022.3174033.

[14] N. Hassan, A. S. M. Miah, T. Suzuki, and J. Shin, "Gradual Variation-Based Dual-Stream Deep Learning for Spatial Feature Enhancement With Dimensionality Reduction in Early Alzheimer's Disease Detection," *IEEE Access*, vol. 13, pp. 31701–31717, 2025, doi: 10.1109/ACCESS.2025.3542458.

[15] R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul, and D. Kandar, "A Survey on Classification Algorithms of Brain Images in Alzheimer's Disease Based on Feature Extraction Techniques," 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2021.3072559.

[16] 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.

[17] M. Ashtari-Majlan, A. Seifi, and M. M. Dehshibi, "A Multi-Stream Convolutional Neural Network for Classification of Progressive MCI in Alzheimer's Disease Using Structural MRI Images," *IEEE J Biomed Health Inform*, vol. 26, no. 8, pp. 3918–3926, Aug. 2022, doi: 10.1109/JBHI.2022.3155705.

[18] Anjali, D. Singh, O. J. Pandey, and H. N. Dai, "STCNN: Combining SMOTE-TOMEK with CNN for Imbalanced Classification of Alzheimer's Disease," *IEEE Sens Lett*, vol. 8, no. 3, pp. 1–4, Mar. 2024, doi: 10.1109/LSENS.2024.3357196.

[19] C. M. Kim and W. Lee, "Classification of Alzheimer's Disease Using Ensemble Convolutional Neural Network With LFA Algorithm," *IEEE Access*, vol. 11, pp. 143004–143015, 2023, doi: 10.1109/ACCESS.2023.3342917.

[20] 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.

[21] M. Gao, P. Song, F. Wang, J. Liu, A. Mandelis, and D. Qi, "A Novel Deep Convolutional Neural Network Based on ResNet-18 and Transfer Learning for Detection of Wood Knot Defects," *J Sens*, vol. 2021, 2021, doi: 10.1155/2021/4428964.

[22] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Feb. 2019, [Online]. Available: http://arxiv.org/abs/1803.08375

[23] K. Kumar, I. Suwalka, A. Uche-Ezennia, C. Iwendi, and C. N. Biamba, "An Improved Deep Learning Unsupervised Approach for MRI Tissue Segmentation for Alzheimer's Disease Detection," *IEEE Access*, vol. 12, pp. 188114–188121, 2024, doi: 10.1109/ACCESS.2024.3510454.

[24] B. Dewi, L. Muflikhah, and B. D. Setiawan, "KLASIFIKASI PENYAKIT ALZHEIMER BERDASARKAN CITRA MRI OTAK MENGGUNAKAN RESNET50V2," 2025. [Online]. Available: http://j-ptiik.ub.ac.id