

MRI-Based Alzheimer's Disease Classification Using ResNeXt50 with Label Smoothing

Mengyan Xia *

School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, 211166, China

* Corresponding author Email: xiamengyan2004@163.com

Abstract: Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive cognitive decline. Identifying the condition at an early stage is essential for timely intervention and diagnosis. To improve MRI-based four-class AD classification, this study employs ResNeXt50 with Label Smoothing (LS) for four-class AD recognition. Public Magnetic Resonance Imaging (MRI) images were preprocessed and augmented during training using random horizontal flipping, slight rotation, translation, and scaling. The images were then classified by ResNeXt50, where LS was added to the cross-entropy (CE) loss to make it more generalizable. ResNeXt50 uses ResNeXt Blocks with grouped convolution to learn multi-branch features while controlling model complexity. The accuracy rate of this model is 99.50%, while the precision rate, recall rate and F1 score all exceed 99.5%, and the Kappa coefficient is 0.9936. The ablation experiments proved that LS could enhance performance, but other measures such as class weighting and extra attention modules did not produce any further improvements. The research results indicate that ResNeXt50 with LS can effectively enhance the AD classification performance of MRI images. It offers a simple and effective solution for intelligent-assisted diagnosis of AD.

Keywords: Alzheimer's Disease; MRI; ResNeXt50; Deep Learning; Label Smoothing.

1. Introduction

Alzheimer's disease (AD) is a common neurodegenerative condition that mainly entails memory degradation and cognitive dysfunction[1]. With rapid population aging, the burden of AD continues to increase and has turned into a significant public health challenge [2]. The existing AD diagnosis and other forms of dementias mainly based on clinical presentation, neuropsychological testing, functional testing, and neuroimaging results[3]. These diagnostic methods are difficult to achieve early identification and are prone to be influenced by the experience of doctors. Thus, an exact and effective diagnosis using the computer assistance is of practical significance.

Magnetic resonance imaging (MRI) is widely used in AD research, as it can noninvasively provide structural brain information[4]. Deep learning has become more popular in recent years to analyze AD based on MRI. Convolutional

neural networks (CNNs) can automatically learn local imaging features and capture subtle morphological changes related to AD[5].

ResNeXt50 is a type of CNN that adds on grouped convolutions and aggregated residual transformations to enhance multi-branch feature representation while maintaining comparable model complexity[6].

However, AD progression is continuous, and MRI differences between adjacent stages are often subtle. Hard one-hot label training could lead to overconfident predictions and hinder generalization.

To address this issue, this study introduces the Label Smoothing (LS) method to improve the discrimination of adjacent AD stages and the generalization performance.

2. Methods

2.1. Framework

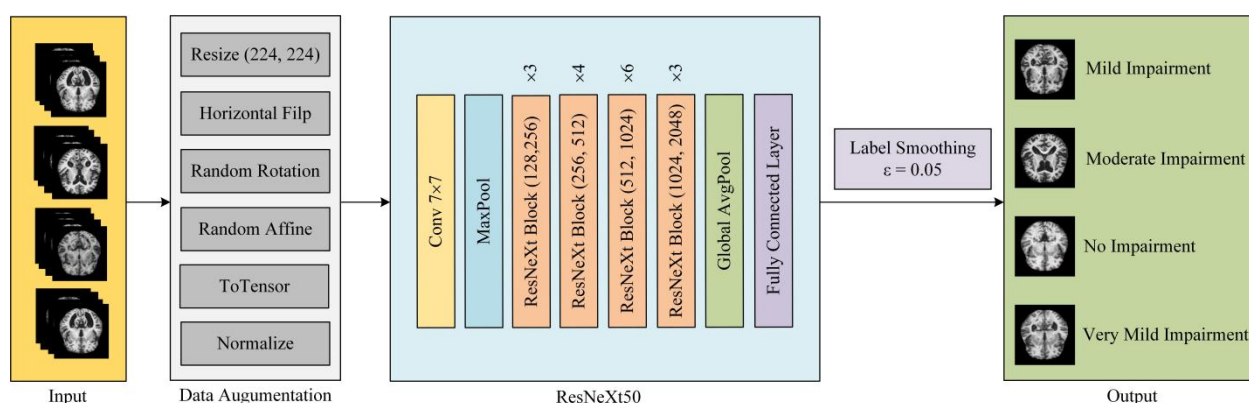


Fig 1. Overall framework of the proposed method

As shown in Fig. 1, ResNeXt50 was used as the backbone network, and LS with $\epsilon = 0.05$ was applied during the training process. The input images included four AD-related stages.

Each MRI image was scaled to 224 x224 pixels. Random horizontal flipping, slightly rotating, translating and scaling were applied to data augmentation in training. To achieve

consistent assessment during testing, only resizing, tensor conversion as well as normalization (using mean and standard deviation) with a value of 0.5 was done.

The preprocessed images were then fed into ResNeXt50 for feature extraction. Eventually, the four-class prediction was formed by using the global average pooling and a fully connected layer, and LS was included in the cross-entropy (CE) loss of the form $\varepsilon = 0.05$.

2.2. ResNeXt Block

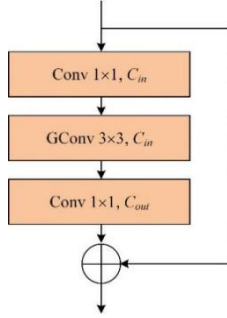


Fig 2. Structure of the ResNeXt Block

The ResNeXt Block is the core unit of ResNeXt50. It adds the concept of grouped convolution and aggregated transformations to a residual framework. Since they all have the same topology, it is structurally regular and simple to stack. The residual function is defined as follows for input feature x :

$$F(x) = \sum_{i=1}^C T_i(x) \quad (1)$$

where $T_i(x)$ denotes the i -th transformation branch. C denotes cardinality, which means the number of parallel branches. The aggregated feature is then added to the input through a residual connection:

$$y = x + \sum_{i=1}^C T_i(x) \quad (2)$$

where y denotes the block output. With this architecture, residual learning is beneficial for the gradient propagation property. The multi-branch aggregation further enhances the feature representation capability, enabling the model to capture the local texture and brain morphology differences of MRI images in different feature subspaces.

The block contains a 1×1 convolution, a 3×3 grouped convolution (GConv) and yet another 1×1 convolution as depicted in Fig. 2. The first 1×1 convolution compresses the channels, the grouped convolution extracts local spatial features within channel groups, and the final 1×1 convolution restores the channel dimension and fuses grouped features for residual addition.

2.3. ResNeXt50

ResNeXt50 is a deep CNN composed of stacked ResNeXt Blocks. It follows the stage-wise design of ResNet50 but replaces standard residual blocks with blocks based on grouped convolution and aggregated transformations.

As shown in Fig. 1, the input MRI image first passes through a 7×7 convolutional layer and a max-pooling layer. The feature maps then enter four stages containing 3, 4, 6, and 3 ResNeXt Blocks, respectively, according to the ResNeXt50 $32 \times 4d$ configuration. As the network deepens, spatial resolution decreases while channel depth increases, enabling the model to learn discriminative structural and semantic features for AD classification.

The resulting high-dimensional feature maps are compressed by global average pooling and passed to a fully connected layer followed by Softmax to predict the probabilities of No Impairment, Very Mild Impairment, Mild

Impairment, and Moderate Impairment.

2.4. Label Smoothing Strategy

Label Smoothing (LS) is a label regularization strategy that smooths target labels to mitigate overconfident predictions and improve generalization [7]. It has also been applied in medical image analysis to alleviate the effects of label uncertainty and overfitting [7, 8].

Conventional CE loss uses one-hot labels, assigning 1 to the true class and 0 to all other classes. However, adjacent AD stages present continuous progression and subtle MRI differences; consequently, hard labels may lead to overconfidence. Therefore, LS was introduced during training.

For a K -class task, let t denote the true class and ε denote the smoothing coefficient. The smoothed label distribution is:

$$y_i^{LS} = \begin{cases} 1-\varepsilon, & i=t \\ \frac{\varepsilon}{K-1}, & i \neq t \end{cases} \quad (3)$$

Here, ε controls the smoothing strength. By assigning part of the probability mass to non-target classes, LS reduces overfitting to hard labels and improves the smoothness of the output distribution. In this study, LS was implemented using PyTorch CrossEntropyLoss, and $\varepsilon = 0.05$ was selected according to the ablation results.

3. Experiments

3.1. Dataset

This study used the public Kaggle Alzheimer's Dataset [9], which contains real axial brain MRI images and WGAN-GP-generated synthetic images to mitigate data scarcity and class imbalance. A total of 11,519 MRI images were included, covering No Impairment, Very Mild Impairment, Mild Impairment, and Moderate Impairment.

3.2. Experimental Setup

The dataset was divided into training and testing sets at a ratio of 80:20. Images were resized to 224×224 pixels. Training images were augmented by random horizontal flipping, slight rotation, translation, and scaling, whereas testing images underwent only by resizing, tensor conversion, and normalization. The batch size was 16, the maximum number of epochs was 100, the initial learning rate was 0.001, and Adam was used as the optimizer. Exponential learning-rate decay with a factor of 0.96 was applied. The loss function was CE with LS ($\varepsilon = 0.05$). A random seed of 42 was fixed, and all models were compared under the same split. The best model based on testing accuracy was saved, and confusion matrices, ROC curves, and training curves were recorded.

3.3. Evaluation Measures

To comprehensively evaluate the model's performance in the MRI-based four-class AD classification task, Accuracy, Precision, Recall, F1-score, and Cohen's Kappa coefficient were used as the main evaluation metrics. Accuracy reflects the overall classification correctness. Precision measures the proportion of true positives in positive predictions, while Recall measures the proportion of true positives in actual positive samples. F1-score provides a combined measure of Precision and Recall. Cohen's Kappa coefficient evaluates the agreement between predicted and true labels while accounting for chance agreement. In addition, confusion matrices, ROC curves, and training curves were used to further analyze the model's class discrimination ability and training stability.

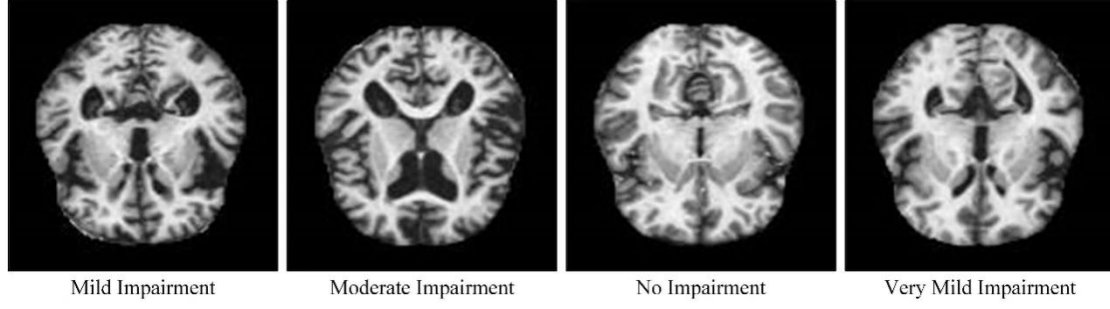


Fig 3. Representative samples from the AD MRI dataset

3.4. Comparison

In order to determine whether ResNeXt50 is suitable for the classification of AD MRI, it was compared with other representative models, such as ResNet50, MobileNetV3, ViT and EfficientFormer[10-13]. Each model was trained on the same proportion of data, number of epochs, without pretrained mode, and ordinary CE loss. To examine the influence of the training strategy, ResNeXt50 with LS (epsilon=0.05) was also considered.

As illustrated in Table 1, CNN-based models typically performed better than ViT. MobileNetV3, ResNet50, and ResNeXt50 all achieved accuracies above 98%, whereas ViT

reached only 48.17%. This may be because MRI classification relies heavily on local texture and brain morphological features, which are effectively captured by convolutional operations. By contrast, ViT typically requires larger datasets and sufficient pretraining.

Among all models, ResNeXt50 achieved the best performance, probably due to its grouped convolution and aggregated residual transformations. After introducing LS, accuracy increased to 99.50%, and Precision, Recall, and F1-score all exceeded 99.5%. This result indicates that LS can further enhance performance without changing the network architecture.

Table 1. Performance comparison of ResNeXt50 and representative models

Models	Accuracy (%)	precision(%)	recall(%)	F1-score (%)	Kappa
MobileNetV3[10]	99.00	99.06	99.05	99.05	0.9866
EfficientFormer[11]	80.95	81.87	81.69	81.73	0.7457
ViT[12]	48.17	46.85	48.73	46.16	0.3098
Resnet50[13]	98.74	98.79	98.79	98.79	0.9831
ResNeXt50[6]	99.39	99.42	99.40	99.41	0.9919
ResNeXt50 with LS $\epsilon=0.05$	99.50	99.56	99.54	99.55	0.9936

3.5. Confusion Matrix and ROC Analysis

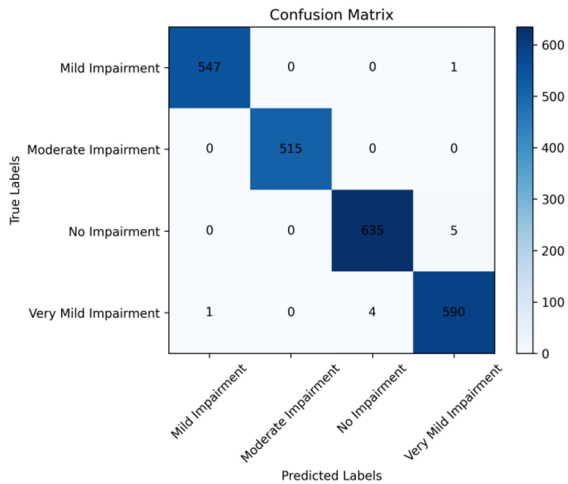


Fig 4. Confusion Matrix of ResNeXt50 with Label Smoothing ($\epsilon = 0.05$)

To further analyze the classification performance of ResNeXt50 across different AD stages, the confusion matrix and multi-class ROC curves on the test set were plotted, as shown in Fig. 4 and Fig. 5. The confusion matrix shows that most samples are concentrated along the diagonal, indicating that the model can accurately identify the four categories. Misclassifications mainly occur between No Impairment and

Very Mild Impairment, which may be attributed to the subtle imaging differences between normal cognition and very mild impairment. The ROC curves show that the AUC values of all classes are close to 1.0000, with micro-average AUC and macro-average AUC reaching 0.9999 and 0.9998, respectively, indicating strong discrimination ability across different AD stages.

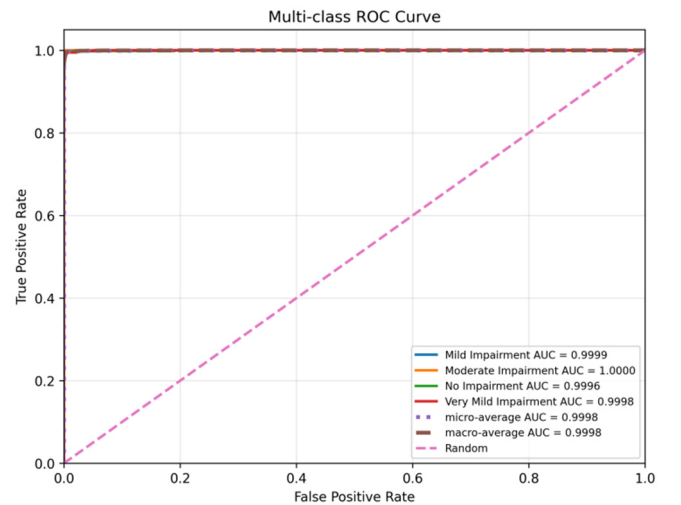


Fig 5. Multi-class ROC Curves of ResNeXt50 with Label Smoothing ($\epsilon = 0.05$)

3.6. Training Process Analysis

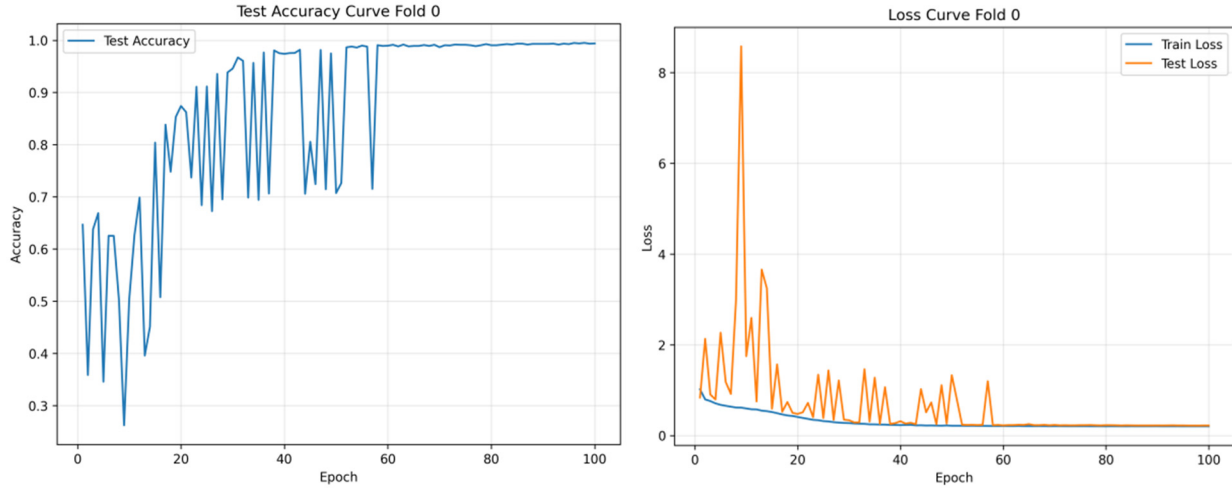


Fig 6. Training process of ResNeXt50 with Label Smoothing ($\varepsilon = 0.05$)

To examine the stability of the training process, the test accuracy curve and the loss curves of the training and test sets were further plotted, as shown in Fig. 6. As the number of training epochs increased, the test accuracy generally improved and remained stable at a high level in the later stage. The training loss showed an overall downward trend, while

the test loss fluctuated in the early stage and then gradually decreased and stabilized. These results indicate that the model converged progressively and showed no obvious overfitting under the current training settings.

3.7. Loss Function Strategy Analysis

Table 2. Effects of Different Loss Function Strategies on ResNeXt50 Performance

Settings	Accuracy(%)	Precision(%)	Recall(%)	F1-score (%)	Kappa
ResNeXt50 with CE	99.39	99.42	99.40	99.41	0.9919
ResNeXt50 with Class Weight CE	99.05	99.08	99.07	99.08	0.9872
ResNeXt50 with LS $\varepsilon=0.03$	99.50	99.53	99.55	99.54	0.9936
ResNeXt50 with LS $\varepsilon=0.05$	99.50	99.56	99.54	99.55	0.9936
ResNeXt50 with LS $\varepsilon=0.10$	99.52	99.56	99.54	99.55	0.9936

After selecting ResNeXt50 as the backbone, different loss strategies were compared, including ordinary CE, class-weighted CE, and LS with different smoothing coefficients. This experience was conducted to analyze the influence of training objective settings on classification performance and to provide a basis for selecting the Label Smoothing coefficient in the final model.

As shown in Table 2, compared with standard CE, Label Smoothing with different smoothing coefficients led to slight performance improvements, indicating that moderate label softening can reduce the model's excessive reliance on one-hot labels and improve generalization. Class Weight CE performed worse than standard CE, possibly because the class distribution in the current combined dataset was already relatively balanced, and further introducing class weights may have caused overcorrection. Although $\varepsilon = 0.10$ achieved a slightly higher Accuracy in this single split, its difference from $\varepsilon = 0.05$ was minimal, and its training process was less stable. Considering that $\varepsilon = 0.05$ weakens the supervision signal of the true class to a smaller extent, Label Smoothing

with $\varepsilon = 0.05$ was finally adopted as the core model setting.

3.8. Additional Evaluation of Attention Modules

To further investigate the effect of attention mechanisms on model performance, SE, ECA, CBAM, and MHSA modules were separately incorporated into ResNeXt50 for comparison. For consistency, all attention modules were inserted between the final feature extraction stage of ResNeXt50, namely layer4, and the global average pooling layer. All attention-based models were trained with standard cross-entropy loss, without introducing Label Smoothing. This insertion position is located after the output of high-level semantic features, allowing the extracted discriminative features to be further recalibrated or enhanced before classification. Attention mechanisms can strengthen informative responses in the channel or spatial dimension, and are theoretically expected to help the model focus on more discriminative brain structures and local imaging features.

Table 3. Performance Comparison of ResNeXt50 with Different Attention Modules

Settings	Accuracy(%)	Precision(%)	Recall(%)	F1-score (%)	Kappa
ResNeXt50 with SE[14]	99.00	99.03	99.06	99.04	0.9866
ResNeXt50 with CBAM[15]	99.05	99.09	99.09	99.09	0.9872
ResNeXt50 with ECA[16]	98.96	99.00	99.01	99.00	0.9860
ResNeXt50 with MHSA[17]	99.22	99.23	99.26	99.24	0.9895
ResNeXt50	99.39	99.42	99.40	99.41	0.9919

As shown in Table 3, adding SE, ECA, CBAM, or MHSA did not further improve performance over the baseline

ResNeXt50. One possible explanation is that the dataset is relatively balanced and the preprocessed MRI images show

consistent structural patterns. The grouped convolution and aggregated residual structure of ResNeXt50 may already extract sufficient local texture and morphological information, whereas additional attention modules increase model complexity and may not yield stable gains under the current sample size and training setting. Therefore, ResNeXt50 with LS ($\epsilon = 0.05$) was retained as the final model.

4. Conclusion

This study proposes an MRI-based four-class AD classification method that integrates ResNeXt50 with Label Smoothing. Specifically, ResNeXt50 is applied to extract features of MRI images of the brain, whereas Label Smoothing is presented as an optimization tool. It can train and reduce overconfident predictions due to unclear class limits between various AD stages. Experimental results show that the proposed method achieves excellent performance on the public AD MRI dataset. It achieves an accuracy of 99.50% and a Kappa coefficient of 0.9936, outperforming MobileNetV3, ResNet50, EfficientFormer, and ViT. In addition, the confusion matrix and ROC curves indicate that the method can effectively distinguish different AD stages and has potential value for computer-aided diagnosis.

Additional ablation experiments indicate that a slight Label Smoothing also enhances model generalization and that $\epsilon = 0.05$ attains a desirable overall performance. In contrast, class weighting and attention modules, including SE, ECA, CBAM, and MHSA, do not lead to clear performance gains. Such a result implies that, under the current dataset and experimental settings, ResNeXt50 is already capable of sufficiently extracting MRI features relevant to classification.

Despite the promising results, this study still has several limitations. First, the experiments were conducted on a single public dataset. So the model's generalization ability across different data sources and scanning devices requires further validation. Second, this study used a fixed Label Smoothing coefficient, whose adaptability under different class distributions and sample conditions remains to be further explored. Future work may validate the model on additional public datasets and investigate adaptive label smoothing strategies to further improve classification performance and practical applicability.

References

- [1] Zheng, Q., & Wang, X. (2025). Alzheimer's disease: insights into pathology, molecular mechanisms, and therapy. *Protein & Cell*, 16(2), 83–120. <https://doi.org/10.1093/procel/pwae026>.
- [2] Safiri, S., et al. (2024). Alzheimer's disease: a comprehensive review of epidemiology, risk factors, symptoms diagnosis, management, caregiving, advanced treatments and associated challenges. *Frontiers in Medicine*, 11. <https://doi.org/10.3389/fmed.2024.1474043>.
- [3] Xue, C., et al. (2024). AI-based differential diagnosis of dementia etiologies on multimodal data. *Nature Medicine*, 30(10), 2977–2989. <https://doi.org/10.1038/s41591-024-03118-z>.
- [4] Jia, H., & Lao, H. (2022). Deep learning and multimodal feature fusion for the aided diagnosis of Alzheimer's disease. *Neural Computing and Applications*, 34(22), 19585–19598. <https://doi.org/10.1007/s00521-022-07501-0>.
- [5] Patil, V., Madgi, M., & Kiran, A. (2022). Early prediction of Alzheimer's disease using convolutional neural network: a review. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 58(1), 130. <https://doi.org/10.1186/s41983-022-00571-w>.
- [6] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5987–5995). <https://doi.org/10.1109/CVPR.2017.634>.
- [7] Zhang, C. B., et al. (2021). Delving deep into label smoothing. *IEEE Transactions on Image Processing*, 30, 5984–5996. <https://doi.org/10.1109/TIP.2021.3089942>.
- [8] Shi, J., Zhang, K., Guo, C., Yang, Y., Xu, Y., & Wu, J. (2024). A survey of label-noise deep learning for medical image analysis. *Medical Image Analysis*, 95, 103166. <https://doi.org/10.1016/j.media.2024.103166>.
- [9] Chugh, L., Alshabrawy, O., & Shang, Y. (2023). Addressing data scarcity and class imbalance in Alzheimer's. *International Journal of Engineering Research Technology*, 1–40.
- [10] Howard, A., et al. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1314–1324). <https://doi.org/10.1109/ICCV.2019.00140>.
- [11] Li, Y., et al. (2022). Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems*, 35, 12934–12949. <https://doi.org/10.48550/arXiv.2206.01191>.
- [12] Dosovitskiy, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. <https://doi.org/10.48550/arXiv.2010.11929>.
- [13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>.
- [14] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141). <https://doi.org/10.1109/CVPR.2018.00717>.
- [15] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19). https://doi.org/10.1007/978-3-030-01234-2_1.
- [16] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11534–11542). <https://doi.org/10.1109/CVPR42600.2020.01155>.
- [17] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://doi.org/10.48550/arXiv.1706.03762>.