Classification of Alzheimer's Disease Using the Convolutional Neural Network (CNN) with Transfer Learning and Weighted Loss

Muhammad Wildan Oktavian, Novanto Yudistira, and Achmad Ridok

Abstract—**Alzheimer's disease is a progressive neurodegenerative disorder that gradually deprives the patient of cognitive function and can end in death. With the advancement of technology today, it is possible to detect Alzheimer's disease through Magnetic Resonance Imaging (MRI) scans. With MRI technology and image recognition, early diagnosis of Alzheimer's disease can be performed automatically using machine learning. Although machine learning has many advantages, currently, deep learning is more widely applied because it has more robust learning capabilities and is suitable for solving image recognition problems. However, several challenges must be faced to implement deep learning, such as the need for large datasets, computing resources, and careful parameter setting to prevent overfitting or underfitting. In responding to the challenge of classifying Alzheimer's disease using deep learning, this study proposes the Convolutional Neural Network (CNN) with the Residual Layer (ResNet-18) architecture. Transfer learning from ImageNet and weighting the loss function are then applied so that each class has a weight depending on its size. It is proposed to overcome the need for a large and balanced dataset. Furthermore, this study experimented with changing the network activation function to a Mish activation function. From the results of the tests, the model's accuracy of 88.30% is produced using transfer learning, weighted loss, and the Mish activation function. This accuracy is better than the baseline model, which only gets an accuracy of 69.10%.**

Index Terms—**Alzheimer's Disease, Residual Network, Transfer Learning, Weighted Loss Function, Mish Activation.**

I. Introduction

Over time, human health inevitably declines, making individuals susceptible to disease. One part of the body that is particularly affected by aging is the brain, which undergoes changes in intellectual function such as difficulty with memory and slow decision-making [1]. Alzheimer's disease is a progressive neurodegenerative disorder that gradually robs patients of their cognitive function and can ultimately result in death [2]. This disease is a leading cause of dementia among the elderly, with the majority of those affected experiencing symptoms such as memory loss, changes in personality, mood swings, and difficulty with social interactions [3], which can last for three to nine years [4].

Manuscript received April 22, 2022; revised April 30, 2023. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

M. W. Oktavian is an undergraduate student of Informatics Engineering, Faculty of Computer Science, Brawijaya University, Indonesia, 65145 (e-mail: wildanokt@student.ub.ac.id)

N. Yudistira is a lecturer of the Informatics Engineering, Faculty of Computer Science, Brawijaya University, Indonesia, 65145 (e-mail: yudistira@ub.ac.id)

A. Ridok is a lecturer of the Informatics Engineering, Faculty of Computer Science, Brawijaya University, Indonesia, 65145 (e-mail: acridokb@ub.ac.id)

With the advancement of technology, it is now possible to detect Alzheimer's disease through Magnetic Resonance Imaging (MRI) scans [5]. MRI is the preferred modality for the diagnosis and monitoring of Alzheimer's disease progression [6]. Early diagnosis of Alzheimer's disease can be achieved using machine learning algorithms with MRI scans. In some cases, machine learning can even outperform medical personnel in predicting Alzheimer's disease, highlighting the need for computer-based diagnostic research [7]. Although machine learning has many advantages, it is not suitable for image recognition. Deep learning, a popular method for image recognition, offers stronger learning capabilities and is better suited to solving image recognition problems [8]. Many deep learning methods, such as Convolutional Neural Networks (CNN) and sparse autoencoder [9], outperform machine learning methods in image recognition. However, deep learning also poses some challenges, such as the requirement for large amounts of training data, which can be costly and ethically protected between organizations. Furthermore, training deep learning networks with large amounts of image data requires significant computational resources, and deep networks require careful hyperparameter settings to avoid overfitting or underfitting [10].

Numerous methods have been applied to detect the results of MRI images using deep learning, such as transfer learning and image augmentation [11], [12]. Other studies have also investigated class imbalance by using a weighted loss on the CNN architecture to reduce its effects on the dataset [13]. Building on these previous works, the proposed study utilizes a CNN model with Residual Neural Layers (ResNet), Weighted Loss, and Transfer Learning to train a classification model for Alzheimer's disease using MRI data, with the aim of classifying it into three classes.

II. Related Work

In 2011, Lucas R. Trambaiolli et al. [14] conducted research on the classification of Alzheimer's disease using the machine learning method of Support Vector Machine (SVM) to differentiate between Alzheimer's patients and controlled patients. Their approach achieved an accuracy of 79.90% and a sensitivity of 83.20%. This study suggested using more data and reconsidering the parameters of the SVM classifier to improve performance.

In 2017, Aly Valliani and Ameet Soni [11] researched the classification of Alzheimer's disease in a study titled "Deep Residual Nets for Improved Alzheimer's Diagnosis." They focused on overcoming the problem of limited data to train the CNN model. Therefore, they used the Deep Residual Nets (ResNet) architecture, which had previously been trained with the ImageNet dataset that contains large image data. The results of this study indicated that the use of transfer learning from ImageNet and augmentation can improve classification accuracy. As a result, the model obtained a test accuracy of 81.30% for binary classification and 56.80% for multiclass or 3-way classification. This result was better than the model that did not use transfer learning and augmentation, which produced an accuracy of 78.80% for binary classification and 56.10% for 3-way classification.

In 2018, Aderghal et al. [12] studied the classification of Alzheimer's disease in a paper titled "Classification of Alzheimer's Disease on Imaging Modalities with Deep CNNs using Cross-Modal Transfer Learning." The researchers stated that public data on Alzheimer's disease is limited, leading to overfitting during training. Therefore, they proposed using transfer learning from larger datasets to improve classification accuracy. The results of this study indicated that the use of transfer learning on the CNN model can improve the model's performance, reduce the overfitting phenomenon, and increase classification accuracy. Furthermore, the classification can be leveraged to perform multi-instance learning on Alzheimer's disease datasets to localize the benign and malignant parts of the brain [27].

In 2017, Songqing Yue [13] conducted a study titled "Imbalanced Malware Images Classification: a CNN-based Approach." According to this study, CNN classification performance decreases when the dataset has an unbalanced number of classes. To overcome this problem, the loss value weighting in the last CNN layer is used. With this weighting, the misclassification of the minority class will be minimized, and the majority class's weight will be reduced so that there can be a balance among classes. Therefore, we employ the weighting term on the loss function to detect minority organelles [28].

In 2020, Krit Sriporn et al. [15] conducted a study titled "Analyzing Lung Disease Using Highly Effective Deep Learning Techniques." This study used the Mish activation function to replace the ReLU activation function in several well-known architectures, such as MobileNet, Densenet-121, and Resnet-50. The Mish activation function is the current state-of-the-art activation function. The analysis results found that the Mish activation function can increase classification accuracy to 98.88% from the model that does not use Mish activation or the baseline model with an accuracy of 97.25

III. Research Method and Materials

A. Deep Residual Network

According to Kaiming He [16], the deeper the neural network structure, the more difficult it becomes to learn. However, the Residual Neural Network (ResNet) provides a residual learning framework to simplify the training process even if a deep network structure is used. ResNet explicitly reformulates the network layer into residual learning functions that lead to the input layer. As the deeper layers of the network begin to converge, degradation problems arise. As the network depth increases, the accuracy saturates and then rapidly decreases. The degradation can be caused by vanishing gradients or overfitting as more layers are included in the model, which leads to higher training error. The architectural types of ResNet are distinguished by the number of layers in the network. The architectures used in testing for the ILSVRC competition are ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152.

B. Weighted Cross Entropy Loss

Cross Entropy is a measure in the field of information theory, which builds on entropy and is typically used to calculate the difference between two probability distributions. It is often associated with and mistaken for logistic loss, which is commonly referred to as log loss. While these two measures come from different sources, they both calculate the same quantity when used as a loss function for a classification model, and can be used interchangeably [17]. The formula for cross entropy loss can be seen in Equation 1 below.

$$-\sum_{c=1}^{M} y_{o.c} \log(P_{o.c}) \tag{1}$$

Where M is the number of existing classes. $y_{o.c}$ is a binary indicator (0 or 1) if the class c label is the correct classification for the sample o, and $p_(o.c)$ is the predicted probability sample o from class c. To address the class imbalance in the dataset, Naceur et al. [18] suggest weighting the loss function based on the number of samples from each class. The formula for calculating the weighted cross-entropy loss is shown in Equations 2 and 3 below.

$$-\sum_{c=1}^{M} W_{o.c} y_{o.c} \log(P_{o.c}) \tag{2}$$

$$W_{o.c} = 1 - \frac{x_c}{N} \tag{3}$$

Where $W_{o.c}$ is the specific weight for each class c. x_c is the number of samples in class c, and N is the total of all samples from all classes.

C. Proposed Method

Figure 1 shows that we utilized the main component of the ResNet architecture for the training process, with modifications. Specifically, we altered the activation function in the last residual block before the pooling process. We replaced the default ReLU activation function of the ResNet architecture with the Mish Activation Function. Mish is a non-monotonic activation function that is smooth, continuous, and self-regularized, inspired by the Swish Activation Function. Mish employs the Self-Gating property, where the non-modulated input is multiplied by the output of the non-linear function of the input [19]. Equation 4 below displays the formula for the Mish Activation function.

$$f(x) = x.tanh(softplus(x)) = x.tanh(ln(1 + e^x)) \tag{4}$$

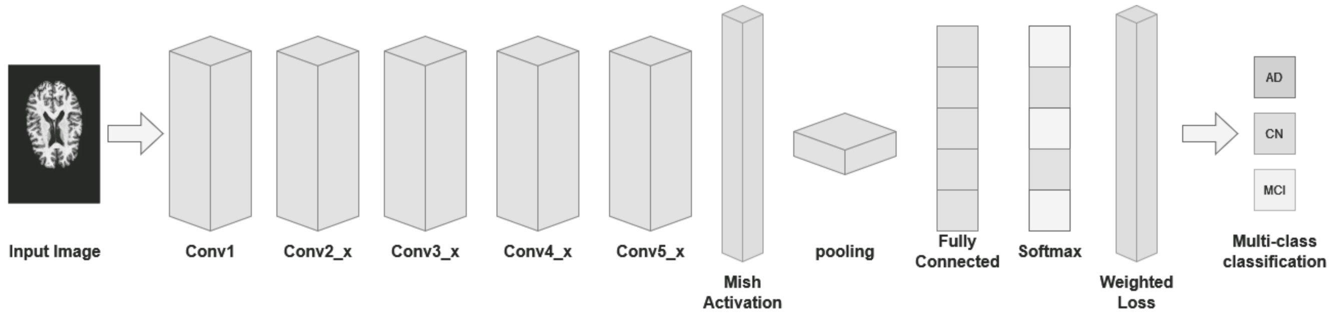
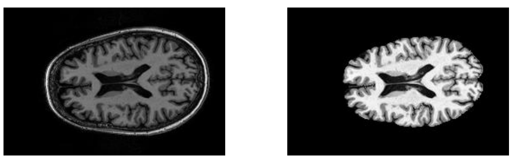


Fig. 1: An Overview of our approach

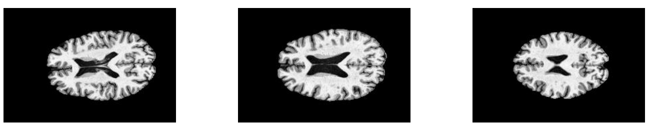
D. Data Acquisition and Preprocessing

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) [21]. The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner. MD. The primary goal of ADNI is to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). These imaging data were collected from 306 ADNI participants, including 133 with mild cognitive impairment (MCI), 58 with Alzheimer's disease (AD), and 115 with normal controls (NC). Each image has the dimension of 256 x 256 x 256, then divided into 256 slices.



(a) (b)

Fig. 2: Comparison between unpreprocessed and preprocessed data sample: (a) Unpreprocessed Data. (b) Preprocessed Data



(a) (b) (c)

Fig. 3: Preprocessed dataset samples : (a) AD Sample. (b) CN Sample. (c) MCI Sample

The dataset obtained is in the form of 3-dimensional images in nifti format. Before it can be used for training, preprocessing is necessary. The image data is first segmented to remove the skull and other parts of the head, leaving only the brain. This segmentation is essential to ensure that the model focuses solely on the brain. The DeepBrain library [22] is used for this purpose. After segmentation, the 3-dimensional images are sliced into two dimensions using the med2image library [23] to be used in network training. The difference between the image data before and after preprocessing can be observed in Figure 2.

Furthermore, the preprocessing results for each class can be seen in Figure 3. Figures 2 and 3 demonstrate that the preprocessed image data has brighter and sharper colors. Subsequently, the preprocessed dataset is divided into training, validation, and test data. This dataset division process uses K-Fold with K totaling 5. First, the dataset is split into two parts with 20% test data and 80% training data. The 80% training data is further divided into 80% training data and 20% validation data.

IV. Result and Discussion

A. ResNet Architecture Comparison

The capacity of a neural network to learn is determined by its architecture. Thus, an appropriate architecture allows the network model to learn data patterns better. The testing scenario for the model's architecture can be seen in Figure 4 below.

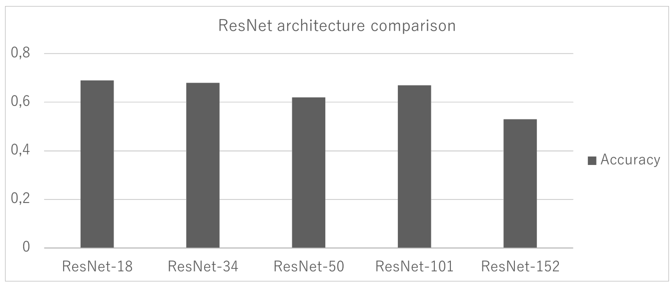


Fig. 4: ResNet architecture comparison by total layers

The results of the architecture tests on the accuracy of the testing data show that as the network layer gets deeper, the accuracy tends to decrease, although the decrease is not significant. This is because of overfitting, as the architecture becomes increasingly complex while the size and variation of the training dataset remain the same, causing the accuracy to decrease. The results of this test indicate that the ResNet-18 architecture performs better in this study.

B. Optimizer Comparison

The optimization algorithm, or optimizer, was tested three times using Stochastic Gradient Descent (SGD), Root Mean Square Propagation (RMSprop), and Adam Optimizer. For the SGD optimization, we set the learning rate parameter to 0.001 and the momentum parameter to 0.9. For RMSprop optimization, we initialized the learning rate parameter to

0.01, alpha to 0.99, epsilon to 1e-08, and zero weight decay and momentum. Finally, we used a learning rate of 0.001, a beta of 0.9-0.999, epsilon of 1e-08, and zero weight decay for Adam optimization.

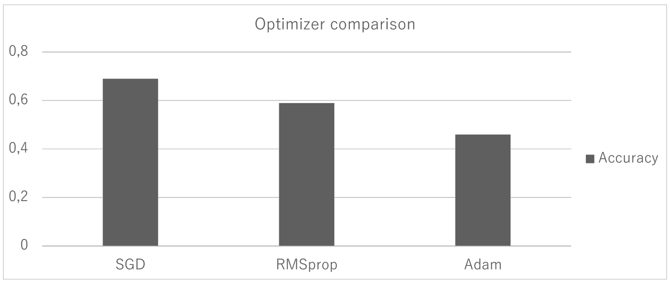


Fig. 5: Optimizer comparison between SGD, RMSprop, and Adam

The tests shown in Figure 5 indicate that the SGD optimization algorithm performs the best compared to other optimization algorithms on the test data. However, for the validation data, the highest accuracy is obtained by the Adam optimization algorithm. Since the reference for this study was the test data, SGD will be used as the optimization algorithm in the next test.

C. Weighted Loss Function

The accuracy and loss results of the class weighting test can be seen in Figures 6 and 7 respectively. The results show that the network model with class weighting in the loss function has better accuracy and lower loss values than the network model without class weighting. This indicates that the class weighting technique effectively handles the problem of imbalanced datasets, resulting in better classification performance. Therefore, in the final test, the ResNet-18 architecture with class weighting in the loss function using the SGD optimizer will be used for evaluation.

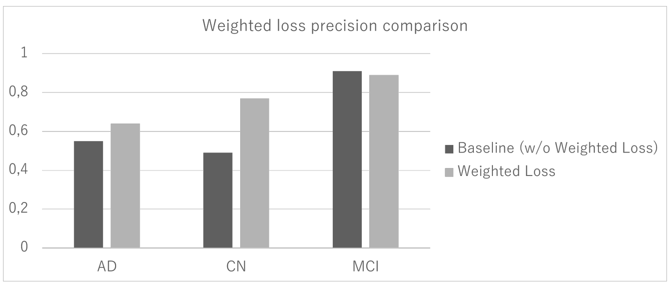


Fig. 6: Comparison between baseline and weighted loss usage on AD, CN, and MCI classes in terms of precision

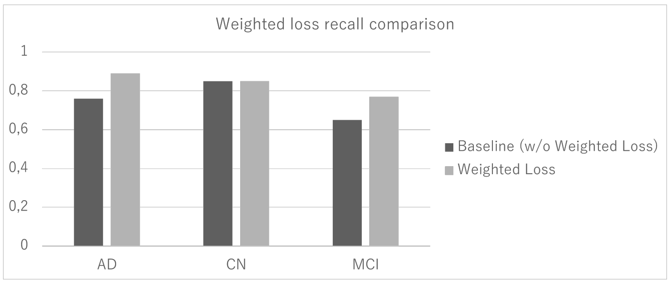


Fig. 7: Comparison between baseline and weighted loss usage on AD, CN, and MCI classes in terms of recall

The results of the tests in Figures 6 and 7 show that applying class weighting results in better performance. The increased precision values indicate that the model can better distinguish between AD and CN classes. The F1 score also shows that the number of correct predictions in each class increases, indicating that class weighting improves the model's classification performance.

D. Transfer Learning

The results of the tests shown in Figure 8 and 9 indicate that using transfer learning with pre-trained models can improve classification performance. The accuracy, precision, recall, and F1 score values are higher than when training from scratch. This indicates that the model can better recognize AD and CN classes, which is important in diagnosing Alzheimer's disease. Therefore, transfer learning can be considered as an effective technique in building neural network models for medical image classification.

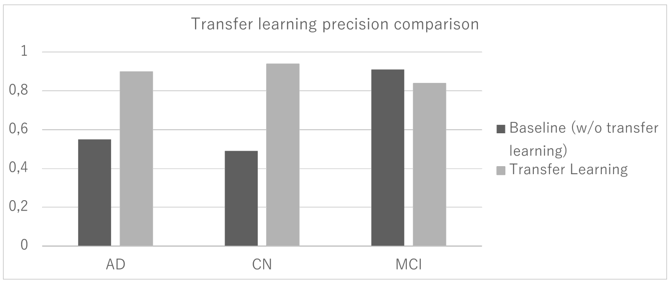


Fig. 8: Comparison between baseline and transfer learning usage on AD, CN, and MCI classes in terms of precision

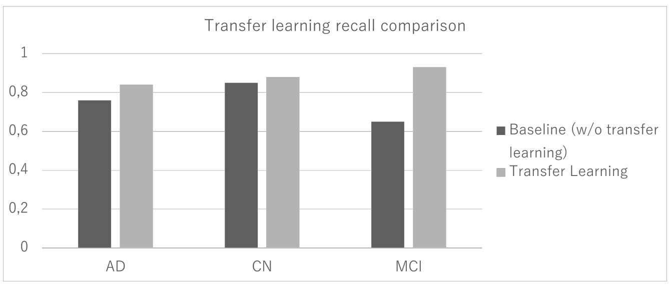


Fig. 9: Comparison between baseline and transfer learning usage on AD, CN, and MCI classes in terms of recall

The results of the tests carried out in Figure 8 and 9 above show that transfer learning significantly increases the accuracy, precision, recall, and F1 score obtained.

E. Mish Activation

In this section, a test is performed by changing the activation function. The ReLU function used previously is replaced with the Mish function. The network layer whose activation function is changed is divided into several conditions. The first change is applied to the last convolution layer before entering the fully connected layer. The second change is applied to all activation functions in the network. These two conditions will be compared with the initial architecture that uses the ReLU activation function as the baseline condition. The scenario for testing the effect of using the Mish activation function can be seen in Figures 10 and 11 below.

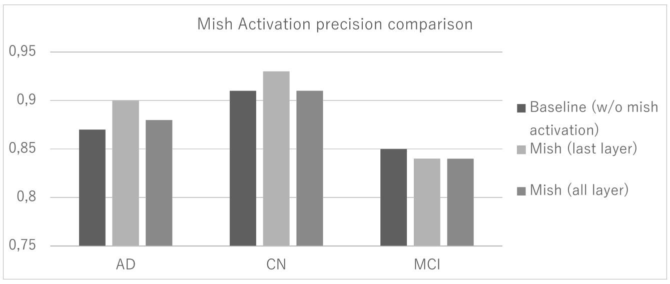


Fig. 10: Comparison between baseline and mish activation usage on AD, CN, and MCI classes in terms of precision

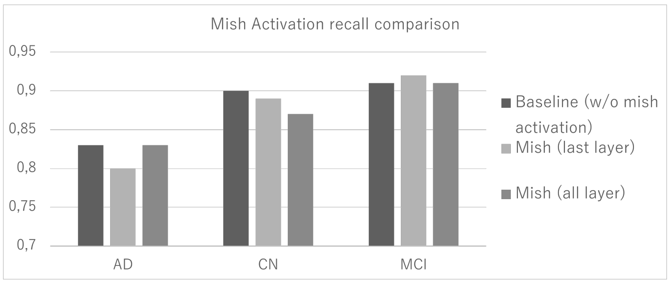


Fig. 11: Comparison between baseline and mish activation usage on AD, CN, and MCI classes in terms of recall

The results shown in Figure 10 and 11 above indicate that the model combining weighted loss and transfer learning with the Mish activation function on the last layer performs better than the others in terms of precision. This means that the model with the Mish activation function in the network's last layer improves the precision of the minority classes of the AD and CN class.

F. Best Scenario Result

For the best scenario, we have chosen the pre-trained ResNet-18 model with Weighted Loss and Mish Activation on the last layer of ResNet-18. Meanwhile, the baseline model is vanilla ResNet-18 without Weighted Loss and Mish activation. To comprehensively evaluate this scenario, we have used the confusion matrix of the baseline and best scenario model, as well as Grad-CAM, to present the results.

	AD	CN	MCI
AD	47	0	340
CN	8	174	608
MCI	17	38	939

0 – 1000

Fig. 12: Baseline Model's Confusion Matrix

	AD	CN	MCI
AD	345	17	25
CN	9	755	14
MCI	42	133	819

0 – 1000

Fig. 13: Best Scenario's Confusion Matrix

The confusion matrix consists of four types of values: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). TP or TN represents the number of correctly classified positive or negative data, while FP and FN represent the number of misclassified data [20]. As shown in Figure 13, the best scenario model can accurately detect each label in 2159 test images. In comparison, as shown in Figure 12, most of the data is classified as the MCI class. Most of the misclassified data occurred for the MCI label because it is biased towards the other two classes, which increases the challenges [21]. However, the proposed model detects more than 80% of MCI images correctly.

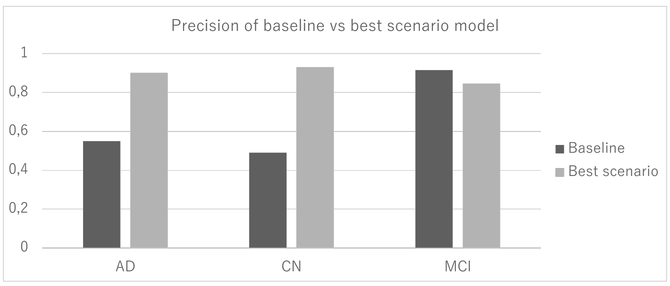


Fig. 14: Comparison between baseline and best scenario on AD, CN, and MCI classes in terms of precision

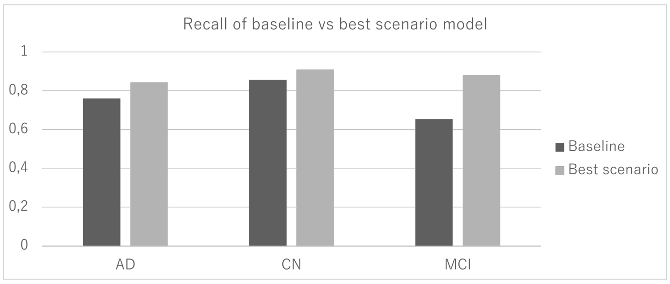
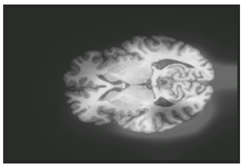
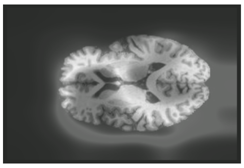


Fig. 15: Comparison between baseline and best scenario on AD, CN, and MCI classes in terms of recall

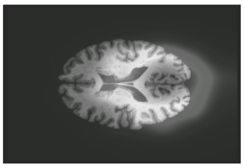
Figure 14 shows the precision of the baseline and the best scenario model classification result. Significant increases in precision are achieved in the minority classes of AD and CN detection. It means that the best scenario model can mostly classify input images into the correct class, even though there is an insignificant decrease in the majority class of MCI. Figure 15 shows a significant increase in recall for all classes, with particular attention to the MCI class, which achieved a considerable increase in recall compared to the baseline. It means that the dominance of MCI is reduced by minimizing the incorrect classification of AD and CN as MCI.



(a)



(b)



(c)

Fig. 16: Model visualization with Grad-CAM : (a) AD Sample. (b) CN Sample. (c) MCI Sample

Figure 16 presents a visual explanation of the classification results using Grad-CAM, which is used to check the areas of the brain that the model focuses on during classification. Grad-CAM, which stands for Gradient-weighted Class Activation Mapping, produces a coarse localization map highlighting the important regions in the image for predicting the concept by using the gradients of any target concept flowing into the final convolutional layer [24]. In this case, the focus area of classification in brain images shows the exact location in each class because the characteristics of each class can be in the same part of the brain. For instance, the model in the figure above can focus on three brain areas in classifying Alzheimer's disease, namely the hippocampus, ventricles, and cortex, which are the locations of the most common symptoms of Alzheimer's disease [25]. Therefore, the model can recognize and classify Alzheimer's disease accurately. Grad-CAM is typically used to distinguish between different objects, but in this study, it helps to identify specific brain areas and their importance in the classification process.

G. Comparison with previous study

TABLE I: Comparison with previous study

Author	Architecture	Accuracy
Trambaiolli et. al.. 2011	Support Vector Machine (SVM)	79.90% (multi-class)
Valliani. 2017	ResNet-18 + Pretrain + augmentation	56.80% (multiclass). and 81.30% (binary)
Acharya et. al. 2021	VGG-16. ResNet-50. Modified AlexNet	75.25% (ResNet-50). 85.07% (VGG-16). 95.70% (Alexnet) (multiclass)
Proposed method	ResNet-18 + Weighted Loss + Transfer Learning + Mish Activation	88.30% (multi-class)

The research conducted by Trambaiolli et al. [14] was used to compare experimental results between traditional machine learning and deep learning. Moreover, we managed to achieve a test accuracy of 79.90%. Although the proposed research has succeeded in providing higher test accuracy, the data used is not the same. Then, the research conducted by Acharya et al. [26] used three models, namely VGG-16, ResNet-50, and Modified AlexNet, with an accuracy of 85.07%, 75.25%, and 95.70%, respectively. The proposed research outperforms two out of three architectures which are VGG-16 and ResNet-50. The AlexNet architecture gets a higher accuracy by making modifications such as using only two of the five convolution layers and the Adam optimizer. This study also uses a different dataset contained in the Kaggle repository. Valliani's research [11], which used ResNet-18 enhanced by pretraining and augmentation, achieved 56.80% accuracy for multiclass and 81.30% for binary classification. The dataset and model settings used are the same as the proposed research. Thus, using weighted loss and Mish activation improves the model performance.

V. Conclusion

In this study, we were able to classify AD, CN, and MCI data with an accuracy of 88.30%, and precision of 90% and 93% for AD and CN, respectively. Interestingly, increasing the number of layers in the ResNet network did not improve the model's performance, and even slightly lowered the accuracy. This can be attributed to the small amount of data used, with only 10,794 images extracted from 306 subjects. Using weighted loss, transfer learning, and Mish activation function individually in the network model improved the model's performance by increasing the precision of the AD and CN classes. However, when combining transfer learning with weighted loss and Mish activation function, there was no significant increase in accuracy due to the different modalities between MRI and ImageNet.

References

[1] Sidiarto, L. (1999). Management and Care System for Alzheimer's/Dementia Disease (Tatalaksana dan Sistem Asuhan pada Penyakit Alzheimer/Demensia). Berkala Neuro Sains, 1(1), 31-35.

[2] Cummings, J. (2002). Alzheimer Disease. JAMA, 287(18), 2335-2338.

[3] Al-Naami, B., Gharaibeh, N., & Kheshman, A. A. (2013). Automated Detection of Alzheimer Disease Using Region Growing technique and Artificial Neural Network. World Academy of Science, Engineering and Technology, 77, 1047-1052.

[4] Long, X., Chen, L., Jiang, C., Zhang, L., & Alzheimer's Disease Neuroimaging Initiative. (2017). Prediction and classification of Alzheimer disease based on quantification of MRI deformation. PloS one, 12(3), e0173372.

[5] Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 835-838.

[6] Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology, 6(2), 67-77.

[7] Kloppel, S., Stonnington, C., Barnes, J., Chen, F., Chu, C., Good, C., Mader, I., Mitchell, L., Patel, A., Roberts, C., Fox, N., Jack, C., Ashburner, J., & Frackowiak, R. (2008). Accuracy of dementia diagnosis–a direct comparison between radiologists and a computerized method. Brain, 131(11), 2969-2974.

[8] Lai, Y. (2019). A comparison of traditional machine learning and deep learning in image recognition. Journal of Physics: Conference Series, 1314, 012148.

[9] Hon, M., & Khan, N. (2017). Towards Alzheimer's disease classification through transfer learning. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1355-1360.

[10] Erhan, D., Manzagol, P. A., Bengio, Y., Bengio, S., & Vincent, P. (2009). The difficulty of training deep architectures and the effect of unsupervised pre-training. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (pp. 153-160). PMLR.

[11] Valliani, A., & Soni, A. (2017). Deep residual nets for improved Alzheimer's diagnosis. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 550-559). ACM.

[12] Aderghal, K., Khvostikov, A., Krylov, A., Benois-Pineau, J., Afdel, K., & Catheline, G. (2018). Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning. In 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS) (pp. 278-283). IEEE.

[13] Yue, S. (2017). Imbalanced Malware Images Classification: a CNN based Approach. arXiv preprint arXiv:1708.08042.

[14] Trambaiolli, L., Lorena, A. C., Fraga, F. J., Kanda, P. A., Anghinah, R., & Nitrini, R. (2011). Improving Alzheimer's disease diagnosis with machine learning techniques. Clinical EEG and neuroscience, 42(3), 160-165.

[15] Sriporn, K., Tsai, C., Tsai, C., & Wang, P. (2020). Analyzing Lung Disease Using Highly Effective Deep Learning Techniques. Healthcare, 8(2), 107.

[16] 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).

[17] Brownlee, J. (2020). A Gentle Introduction To Cross-Entropy For Machine Learning. Machine Learning Mastery. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/cross-entropy-for-machine-learning [Accessed 2 February 2021].

[18] Ben Naceur, M., Akil, M., Saouli, R., & Kachouri, R. (2020). Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. Medical Image Analysis, 63, 101692.

[19] Misra, D. (2019). Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681.

[20] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 01-11.

[21] Alzheimer's Disease Neuroimaging Initiative (ADNI). (n.d.). [online] Available at http://adni.loni.usc.edu [Accessed 11 February 2021].

[22] Deepbrain. (n.d.). Retrieved 5 june 2021, from https://github.com/iitzco/deepbrain

[23] Med2image. (n.d.). Retrieved 6 June 2021, from https://github.com/FNNDSC/med2image

[24] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618-626).

[25] Keep Memory Alive. (2021). Alzheimer's Brain. Retrieved 7 October 2021, from https://www.keepmemoryalive.org/cc-nevada/alzheimers-brain

[26] Acharya, H., Mehta, R., & Kumar Singh, D. (2021). Alzheimer Disease Classification Using Transfer Learning. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1-5). doi: 10.1109/iccmc51019.2021.9418294.

[27] Kavitha, M., Yudistira, N., & Kurita, T. (2019, November). Multi-instance learning via deep CNN for multi-class recognition of Alzheimer's disease. In 2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA) (pp. 89-94). IEEE.

[28] Yudistira, N., Kavitha, M., Itabashi, T., Iwane, A. H., & Kurita, T. (2020). Prediction of sequential organelles localization under imbalance using a balanced deep U-Net. Scientific Reports, 10(1), 1-11.