Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 2 (2023) 310 - 317 ISSN Media Electronic: 2580-0760

Deep Learning Implementation using Convolutional Neural Network for Alzheimer's Classification

Adhigana Priyatama[1], Zamah Sari[2], Yufis Azhar[3]
[1,2,3]Informatics, Faculty of Engineering, University of Muhammadiyah Malang
[1]adhi.gana1@gmail.com, [2]zamahsari@umm.ac.id, [3]yufis@umm.ac.id

Abstract

Alzheimer's disease is the most common cause of dementia. Dementia refers to brain symptoms such as memory loss, difficulty thinking and problem solving and even speaking. This stage of development of neuropsychiatric symptoms is usually examined using magnetic resonance images (MRI) of the brain. The detection of Alzheimer's disease from data such as MRI using machine learning has been the subject of research in recent years. This technology has facilitated the work of medical experts and accelerated the medical process. In this study we target the classification of Alzheimer's disease images using convolutional neural network (CNN) and transfer learning (VGG16 and VGG19). The objective of this study is to classify Alzheimer's disease images into four classes that are recognized by medical experts and the results of this study are several evaluation metrics. Through experiments conducted on the dataset, this research has proven that the algorithm used is able to classify MRI of Alzheimer's disease into four classes known to medical experts. The accuracy of the first CNN model is 75.01%, the second VGG16 model is 80.10% and the third VGG19 model is 80.28%.

Keywords: alzheimer's disease; image classification; deep learning; convolutional neural network; transfer learning

1. Introduction

Alzheimer's disease is the cause of 60% of dementia cases in the world. Alzheimer's disease is one of several causes of dementia. A person's mental capabilities begin to decline and reach a point where it becomes difficult for sufferers to lead a normal life. As the disease worsens, sufferers will become more dependent on their family members for a living. Sometimes people with Alzheimer's disease have problems identifying their family members. Sufferers of the mild cognitive stage can behave aggressively, but sufferers of the late cognitive stage can suffer from heart failure and respiratory system dysfunction, leading to death. An accurate early diagnosis of Alzheimer's disease cannot be made if inappropriate treatment has been done before[1]. All indicators of Alzheimer's disease usually develop slowly but are very influential over time when disturbances occur in the brain begin[2]. According to research, in 2050 every one in 85 people will be indicated with this disease and the number of sufferers will double in the next 20 years [3]. According to the International Alzheimer's Disease (AD) report. in 2019, around 95% of people believe they may suffer from Alzheimer's disease in the future[4]. The estimated annual cost of treating dementia is estimated at $1 trillion and will double by 2030[5]. Depending on age, the proportion of people affected by Alzheimer's disease varies. 5.8 million Americans in the United States (US) aged 65 years and over suffer from Alzheimer's disease in 2020. And by 2050, it is estimated to reach 13.8 million[6].

In patients with Alzheimer's disease, the size of the brain ventricles increases, and the size of the *cerebral cortex* and *hippocampus* shrinks. When the size of the *hippocampus* decreases episodically, spatial memory becomes impaired. Damage between these *neurons* leads to defective communication in planning, judgment, and short-term memory[7]. Deposition of *amyloid-β* plaques and *tau protein* concentrates is probably the first step in the development of Alzheimer's disease. In addition, neurodegeneration, associated with brain *atrophy* and *hypometabolism*, affects cognition[8].

Alzheimer's disease has had no cure until now and is considered dangerous to human health and living things. This disease has affected many people around the world. Thus, early diagnosis of this disease with computer-assisted tools is one of the most interesting and very important research interests for medical and computer science[9]. *Deep learning* methods are a

Accepted: 23-11-2022 | Received in revised: 28-02-2023 | Published: 26-03-2023

hallmark of modern artificial intelligence techniques and have been used extensively in recent studies in applications such as segmentation, classification, and natural language processing. This method can learn the robust features of the input distribution and form a high-level hierarchical path. In this era, *machine learning* (ML) methods that can calculate intercorrelation between regions have become an interesting and basic element of computer-assisted analytical techniques[10].

This area of research is becoming important as machine learning methods such as *deep convolutional network networks* (DCNN) [11] have shown earlier success in the automatic detection of Alzheimer's disease. Wang et al. introduced the fundamentals of DCNN for image classification. Further, they introduced the application of *deep learning* to classify focal liver lesions on a *computerized tomography scan* (CT-scan) modality[12].

The traditional method relies on manual feature extraction, which relies heavily on technical experience and repeated efforts, which appear to be time-consuming and subjective. As a result, *deep learning*, especially *convolutional neural networks* (CNN), is an effective way to overcome these problems[13]. CNN can improve efficiency further by demonstrating great success in disease diagnosis, and there is no need to perform hand-made feature extraction when extracting features automatically [14].

In previous studies for the classification of Alzheimer's disease in the *dataset* from ADNI, using CNNs and ICAE *transfer learning* obtained different accuracy values, CNNs (61.05%) and ICAE *transfer learning* (77.46%) [15]. From this study, the CNNs model obtained lower accuracy compared to other methods used in this study.

In another study, in the classification of Alzheimer's disease in the ADNI *dataset* using the CNN and *Alexnet* models, the accuracy values were respectively CNN (64.76%) and *Alexnet* (67.62%) [16]. This study suggests modifying the CNN architecture and fine-tuning it to improve the model performance. Other studies carried out several classifications of Alzheimer's disease using the CNN, VGG16, and VGG19 models to obtain accuracy values, including CNN (71%), VGG16 (77%), and VGG19 (77%) [11]. An unbalanced data composition could cause the model performance results from this study.

In some previous studies, there are several advantages and disadvantages. From these advantages and disadvantages, we conduct research to get the best performance. *Data augmentation* and *pooling layers* are used to overcome the conditions in solving problems from previous studies, such as the CNN *handcraft model*. Changing the architecture and Image data manipulation is done at the *pooling layer*.

Based on the research above, this study's main objective is to surpass previous studies' performance results. In this study, we built three models: the CNN model, the VGG16 *Transfer learning* model, and the VGG19 *Transfer learning* model. The three models used have better performance results than the other models. The built model also undergoes data augmentation. Data augmentation is done by changing the model architecture to match the *dataset*.

2. Research Methods

The research method consists of several stages: *dataset* collection, *data preprocessing*, *dataset* division, *data augmentation*, *model training*, and *model evaluation*. The following is the flow of research stages which can be seen in Figure 1.

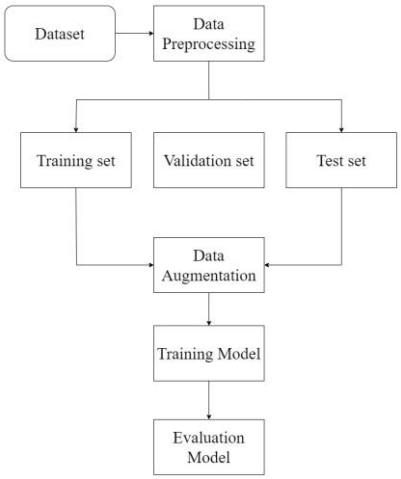

Figure 1. Research stages

Figure 1 illustrates the research flow from start to finish. The research began by collecting *datasets* from the *Kaggle* website. The next stage is *data preprocessing*, such as rescaled image data and dividing the Alzheimer's disease MRI *dataset* into three parts: *training*, *test*, and *validation data*.

Data dividing aims to train the *dataset* so that the built model can learn according to the parameters built in the model. Then the data is augmented to provide modifications to the Alzheimer's disease MRI images. Furthermore, the model is trained using modified data so that the final result of the model that has been built is the result of an evaluation of the model in the form of the value of each metric evaluated using *model evaluate*. *Model evaluate* is a function in the *Keras* library in the *Python* programming language that is useful for evaluating models trained using *data validation* or *test data* and appropriate labels.

DOI: https://doi.org/10.29207/resti.v7iX2.4707

2.1. Dataset

The *dataset* used in this study is the Alzheimer's Dataset (4 classes of images). The *dataset* was obtained from the *Kaggle* website, which provides online *datasets* for research and analysis in various areas, www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images.

The *dataset* obtained contains 6400 images with a size of 176 x 208 pixels which are partitioned into; *Train set* with 5121 pictures and four classes and a *Test set* with 1279 pictures and four classes. The four classes in the dataset are *Mild demented*, *Moderate demented*, *Non-demented*, and *Very mild demented*. The following are examples of images in each class, as shown in Figure 2 [17].

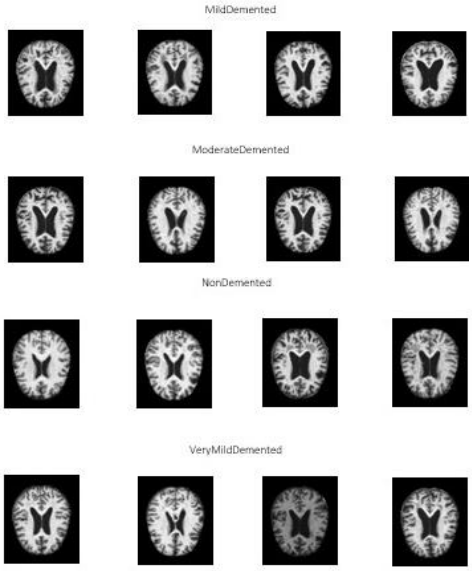

Figure 2. Dataset MRI Alzheimer

2.2. Model Architecture

This study proposes three models, which are then used as comparisons. The first architecture used is the CNN model, as shown in Figure 3, using an *input layer* of 224x224 pixels. It can assist in image recognition and speed up the computation to distinguish four image classes consisting of *Mild demented*, *Moderate demented*, *Non-demented*, and *Very mild demented*.

As shown in Figure 3, the first model uses three *convolutional layers* and three *pooling layers* by implementing *max pooling* with a 2 x 2 filter. This study also uses three *convolutional layers* with filters 64, 32, and 16, which use a 3 x 3 kernel and use *relu* activation. Furthermore, the *fully connected layer* is used, which has a *flatten layer* and a *dense layer* followed by a *dropout layer* 0.2 which uses *relu* activation. In the last process, a *dense layer* is used with *softmax* activation.

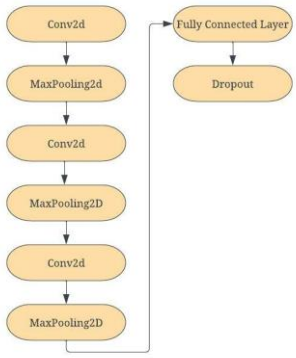

Figure 3. CNN Model Architecture

As shown in Figures 4 and 5, the second and third model architectures use *Visual Geometry Group* 16 (VGG16) and *Visual Geometry Group* 19 (VGG19) transfer learning. In the manufacturing model, there are several special attentions, namely, *Categorical Cross Entropy* as a form of *cross-entropy loss* or *log loss* used in measuring the performance of several classification problems that give more than two outputs (*softmax*), and *batch size*, which describes the number of samples worked before the internal model parameters are updated.

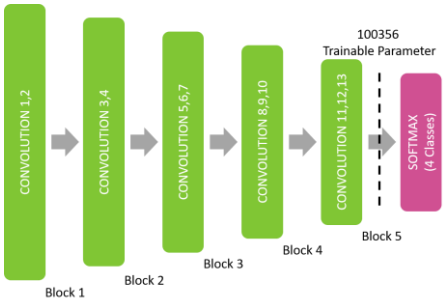

Figure 4. VGG16 Model Architecture

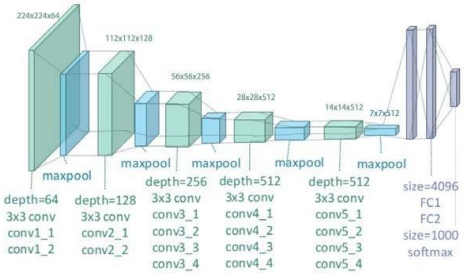

Figure 5. VGG19 Model Architecture

2.3. Data Augmentation

Augmentation is a technique used to change Alzheimer's disease images into several image variants so that they can help increase the value of model accuracy. The augmentation parameters used in this study are as follows*; rescale = 1/255, width_shift_range = 0.2, height_shift_range = 0.2, shear_range = 0.2, rotation_range = 5,*

DOI: https://doi.org/10.29207/resti.v7iX2.4707

horizontal_flip = True, vertical_flip = True, fill_mode = 'nearest'.

2.4. Testing Scenario

The experiments in Model 1, Model 2, and Model 3 involved four classes labeled *MildDemented*, *ModerateDemented*, *NonDemented*, and *VeryMildDemented*. Furthermore, the dataset is divided into three parts: *data train*, *data test* and *data validation*. The *test data* takes 20% of the entire *dataset*, while the *train data* takes 80% of the entire *dataset*, but the *train data* will be divided into 20% *validation data* and 80% *train data*.

In this study, each experiment used 20 *epochs*. The evaluation of the model used to measure the model's performance in this study is to find *accuracy*, *precision, recall, f1-score, and auc*. *Accuracy* is the percentage of the data that is correctly predicted in the total amount of data. *Precision* is the true positive prediction of the overall true positive prediction results. The *recall* is a comparison between true positives and the amount of data that is actually positive. The *f1-score* is calculated by comparing the average *precision* and *recall* with a weighted value. *Auc* (*Area Under Curve*) is the area under the curve that measures the two-dimensional area under the entire *ROC* curve from (0,0) to (1,1). Using more than one metric is very important because model evaluation cannot only be fixed on one metric, especially on metrics sensitive to unbalanced data [18].

3. Results and Discussions

The results of this study are the results of the stages that have been carried out based on the arrangement of the test scenarios. Then the researcher tested the three proposed model scenarios and the performance results were compared.

3.1. Scenario 1 (Model CNN *Handcraft Model*)

The *dataset* is tested using the Convolutional Neural Network (CNN) model that has been built. In the training process that was carried out with the first model, graph plots were obtained showing accuracy, loss, AUC, precision, and f1 score graphs, as shown in Figures 6, 7, 8, 9, and Figure 10.

The graph of the CNN model evaluation metric plot above, listed in Figures 6, 7, 8, 9, and 10, shows that the model was trained using 20 *epochs*. From *epochs* 1 to 20, the graphs show the instability of model training on several evaluation metrics. This instability can be caused by a lack of compatibility between the model created and the *dataset*.

Once the plot graphs of the CNN model are known, the experiment is then evaluated using five different metrics to ensure the performance of the CNN model that has been built. Performance testing is done using *model evaluate*.

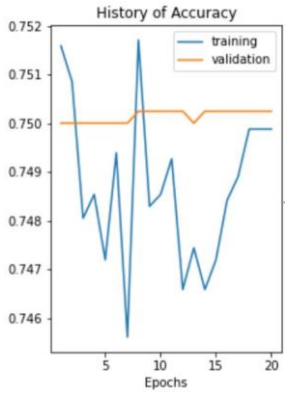

Figure 6. CNN Model Accuracy Chart

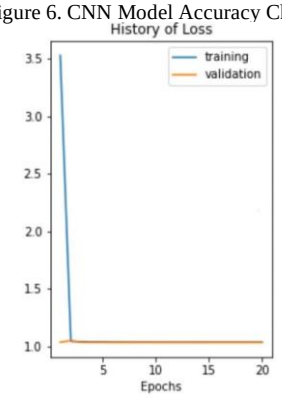

Figure 7. CNN Model Loss Chart

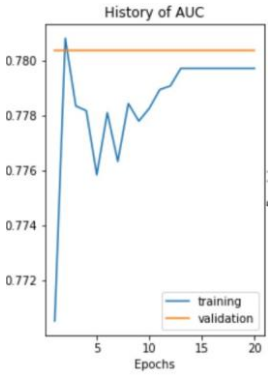

Figure 8. CNN Model AUC Chart

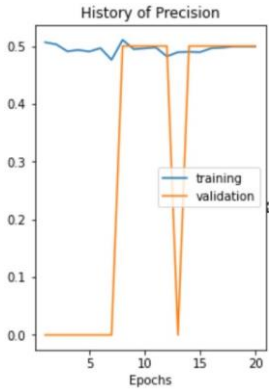

Figure 9. CNN Model Precision Chart

DOI: https://doi.org/10.29207/resti.v7iX2.4707

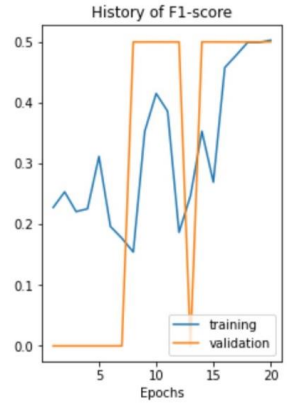

Figure 10. CNN Model F1-score Chart

```
Accuracy =  0.7501954436302185
Precision =  0.5003909468650818
Recall =  0.5003909468650818
AUC =  0.7805577516555786
F1_score =  0.5003023743629456
```

Figure 11. CNN Model Evaluation Results

The performance results from CNN based on figure 11 include; an *accuracy* of 75.01%, the *precision* of 50.04%, *recall* of 50.04%, an *auc* of 78.05%, and a *f1 score* of 50.03%.

3.2. Scenario 2 (VGG16 Model)

The dataset is tested using the VGG16 model that has been built. In the training process carried out with the second model, graph plots were obtained showing *accuracy*, *loss*, *auc*, *precision*, and *f1-score* graphs, as shown in Figures 12, 13, 14, 15, and Figure 16.

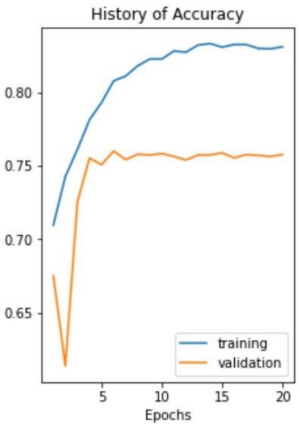

Figure 12. VGG16 Model Accuracy Chart

The graph of the metric evaluation plot of the VGG16 model is listed in Figures 12, 13, 14, 15, and Figure 16. It can be seen that the model was trained using 20 *epochs*. From *epochs* 1 to 5, the graph shows the instability of model training on several evaluation metrics. However, *epochs* 5 to 20 show the stability of the model training that occurs because the model has started to learn the *dataset*.

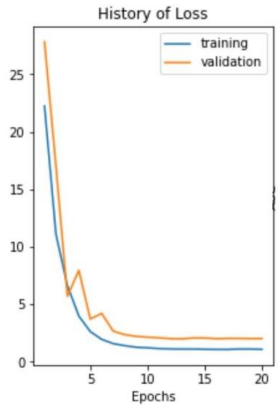

Figure 13. VGG16 Model Loss Chart

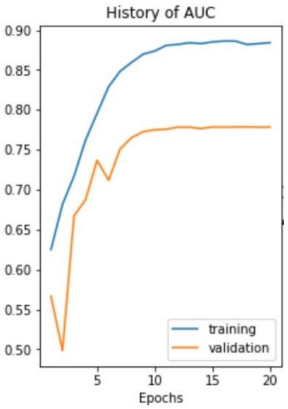

Figure 14. VGG16 Model AUC Chart

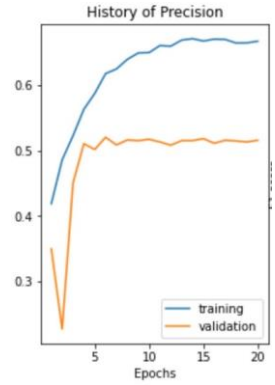

Figure 15. VGG16 Model Precision Chart

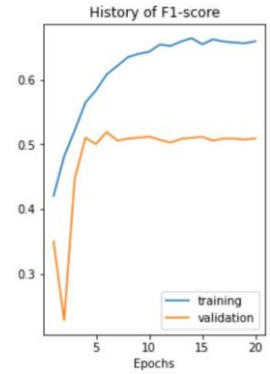

Figure 16. VGG16 Model F1-score Chart

DOI: https://doi.org/10.29207/resti.v7iX2.4707

After knowing the plot graph of the VGG16 model, the experiment was then evaluated using five different metrics to ensure the performance of the VGG16 model that had been built. Performance testing is done using *model evaluate*.

```
Accuracy =  0.8010163903236389
Precision =  0.6044836044311523
Recall =  0.5903049111366272
AUC =  0.8382682204246521
F1_score =  0.5967336297035217
```

Figure 17. VGG16 Model Evaluation Results

The performance results of VGG16 based on Figure 17 include; an *accuracy* of 80.10%, a *precision* of 60.45%, a *recall* of 59.03%, an *auc* of 83.83%, and a *f1-score* of 59.67%.

3.3. Scenario 3 (VGG19 Model)

The dataset is tested using the VGG19 model that has been built. In the *training* process with the third model, graph plots were obtained showing *accuracy*, *loss*, *auc*, *precision*, and *f1-score* graphs, as shown in figures 18, 19, 20, 21, and 22.

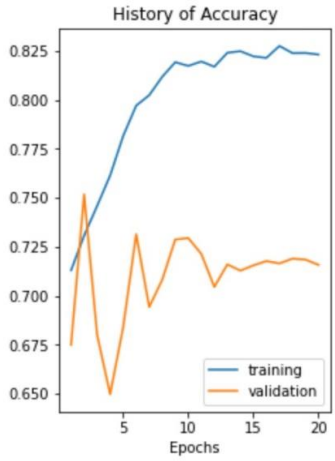

Figure 18. VGG19 Model Accuracy Chart

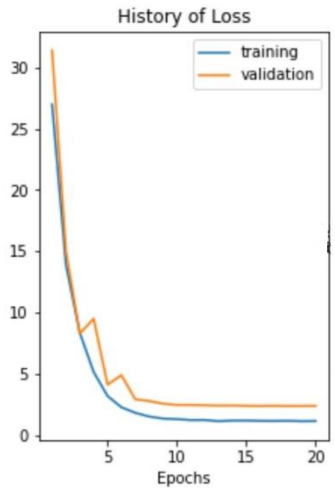

Figure 19. VGG19 Model Loss Chart

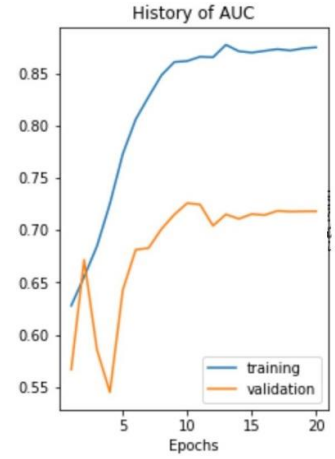

Figure 20. VGG19 Model AUC Chart

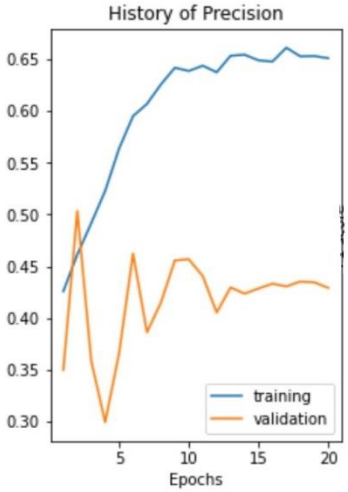

Figure 21. VGG19 Model Precision Chart

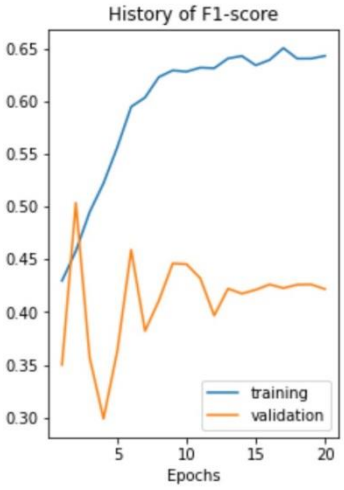

Figure 22. VGG19 Model F1-score Chart

In the graph of the metric evaluation plot of the VGG19 model above, listed in Figures 18, 19, 20, 21, and 22, it can be seen that the model was trained using 20 *epochs*. The graph shows model training instability on the *loss* metric from *epochs* 1 to 5. However, *epochs* 5 to 20 show the stability of the model training

DOI: https://doi.org/10.29207/resti.v7iX2.4707

that occurs because the model has started to learn the *dataset*. Furthermore, on other metrics, from *epochs* 1 to 13, it shows model training instability, and from *epochs* 13 to 20, it shows model training stability.

After knowing the plot graph of the VGG19 model, then the experiment is evaluated using five different metrics to ensure the performance of the VGG19 model that has been built. Performance testing is done using *model evaluate*.

```
Accuracy =  0.8027756214141846
Precision =  0.6064668893814087
Recall =  0.6012510061264038
AUC =  0.8237602710723877
F1_score =  0.6038089990615845
```

Figure 23. VGG19 Model Evaluation Results

The performance results from VGG19 based on Figure 23 include; an *accuracy* of 80.28%, a *precision* of 60.65%, a *recall* of 60.13%, an *auc* of 82.38%, and a *f1 score* of 60.38%.

3.4. Model Performance Comparison

After training the model, three model scenarios are tested for performance using the evaluate model. The following compares the performance results of the three model scenarios, summarized in table 1.

Table 1. Evaluation Results

Model	Accuracy	Precision	Recall	F1-Score	AUC
CNN	75%	50%	50%	50%	78%
VGG16	80%	60%	59%	59%	83%
VGG19	80%	60%	60%	60%	82%

After testing the performance in three scenarios using the evaluate model, it can be seen in table 1 with the VGG19 model gets the best performance results among other scenarios.

3.5. Comparison of the Best Model Performance with Previous Research

After conducting experiments using several different model scenarios, the next step is to compare the best performance between the previous research and the best performance in this study. Based on table 2, the best performance of this study was obtained in scenario 3 of VGG19 with an accuracy of 80.28%.

Table 2. Accuracy results of each model scenario

Model	Accuracy
ICAE [15]	77%
VGG19 [11]	77%
Model terbaik yang diajukan dengan VGG19	80%

4. Conclusion

This study proposes a *transfer learning* method to detect Alzheimer's disease from structured MRI data. We performed several Alzheimer's disease classifications using CNN, VGG16, and VGG19 and proved that the method allows for multiple medical image classifications that can be applied to similar fields. This study applies several algorithms (CNN, VGG16, VGG19) for the multi-classification of Alzheimer's disease *datasets*. The results of several medical image classifications are quite good, but there is still room for improvement. VGG19 gets the best performance with a value of 80% for *accuracy*, 60% for *precision*, 60% for *recall*, 60% for *f1-score*, and 82% for *auc*. On the other hand, VGG16 got better performance results than the CNN *handcraft model* on all the performance metrics used. This study also has good medical image processing by using several evaluation metrics that are relevant to reveal the limited capacity of the model.

For further development of this research, the researcher suggests using a more balanced *dataset*, experimenting on each layer of the CNN model architecture, and experimenting with similar or other *transfer learning* methods such as *Inception V4*.

References

[1] A. Association, "2019 ALZHEIMER'S DISEASE FACTS AND FIGURES Includes a Special Report on Alzheimer's Detection in the Primary Care Setting: Connecting Patients and Physicians," *Alzheimer's Dement. Vol. 15, Issue 3*, pp. 321–387, 2019, [Online]. Available: https://www.alz.org/media/Documents/alzheimers-facts-and-figures-2019-r.pdf%0Ahttps://alz.org/media/Documents/alzheimers-facts-and-figures-2019-r.pdf.

[2] E. Picón *et al.*, "Does empirically derived classification of individuals with subjective cognitive complaints predict dementia?," *Brain Sci.*, vol. 9, no. 11, pp. 1–18, 2019, doi: 10.3390/brainsci9110314.

[3] S. Sarraf, D. D. Desouza, J. A. E. Anderson, and C. Saverino, "MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models," *IEEE Access*, vol. 7, no. Mci, pp. 155584–155600, 2019, doi: 10.1109/ACCESS.2019.2949577.

[4] Alzheimer's Disease International, "The costs of dementia: advocacy, media and stigma.," *World Alzheimer Rep. 2019 Attitudes to Dement.*, pp. 100–101, 2019, [Online]. Available: www.daviddesigns.co.uk.

[5] C. Lynch, "World Alzheimer Report 2019: Attitudes to dementia, a global survey," *Alzheimer's Dement.*, vol. 16, no. S10, p. 38255, 2020, doi: 10.1002/alz.038255.

[6] "2020 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 16, no. 3, pp. 391–460, 2020, doi: 10.1002/alz.12068.

[7] G. Umbach *et al.*, "Time cells in the human hippocampus and entorhinal cortex support episodic memory," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 45, pp. 28463–28474, 2020, doi: 10.1073/pnas.2013250117.

[8] S. Shantanam and MUELLER, "乳鼠心肌提取 HHS Public Access," *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2018, doi: 10.1038/s41583-018-0067-3.Imaging.

[9] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cogn. Syst. Res.*, vol. 57, pp. 147–159, 2019, doi: 10.1016/j.cogsys.2018.12.015.

[10] K. Amit, *Machine Learning Techniques*. 2019.

[11] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, O. F. AFE, and K. I. Okesola, "Multi-classification of alzheimer disease

DOI: https://doi.org/10.29207/resti.v7iX2.4707

on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches," *Int. J. Adv. Comput. Res.*, vol. 11, no. 53, pp. 51–60, 2021, doi: 10.19101/ijacr.2021.1152001.

[12] D. Kaul, H. Raju, and B. K. Tripathy, *Deep Learning in Healthcare BT - Deep Learning in Data Analytics: Recent Techniques, Practices and Applications*. 2022.

[13] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, 2020, doi: 10.1186/s40708-020-00112-2.

[14] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/7068349.

[15] K. Oh, Y. C. Chung, K. W. Kim, W. S. Kim, and I. S. Oh, "Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, 2019, doi: 10.1038/s41598-019-54548-6.

[16] B. Khagi, C. G. Lee, and G. R. Kwon, "Alzheimer's disease Classification from Brain MRI based on transfer learning from CNN," *BMEiCON 2018 - 11th Biomed. Eng. Int. Conf.*, pp. 1–4, 2019, doi: 10.1109/BMEiCON.2018.8609974.

[17] SARVESH DUBEY, "Alzheimer's Dataset (4 class of Images) Images of MRI Segementation." 2020, [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images.

[18] N. W. S. Wardhani, M. Y. Rochayani, A. Iriany, A. D. Sulistyono, and P. Lestantyo, "Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data," *2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019*, pp. 14–18, 2019, doi: 10.1109/IC3INA48034.2019.8949568.

DOI: https://doi.org/10.29207/resti.v7iX2.4707