JPPIPA 12(1) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

MRI Image Classification of Brain Tumors Using VGG16-Based Transfer Learning and Data Augmentation as a Medical Diagnosis Support System

Nadia Sri Aulia Noprianti[1*], Leni Aziyus Fitri[1]

[1] Master of Physics, Universitas Negeri Padang, Padang, Indonesia.

Received: September 03, 2025
Revised: December 01, 2025
Accepted: January 25, 2026
Published: January 31, 2026

Corresponding Author:
Nadia Sri Aulia Noprianti
Nadiasriaulia08@gmail.com

DOI: 10.29303/jppipa.v12i1.14218

Abstract: Brain tumors are diseases that require early detection and accurate diagnosis. Various studies have applied deep learning methods to classify MRI images of brain tumors, but they still face dataset limitations and imbalanced class distributions that impact model performance. This study aims to evaluate the performance of the transfer learning-based VGG16 model in classifying brain tumors using MRI images. The study used 7,023 MRI images, including glioma, meningioma, pituitary, and no tumor, with a balanced training data distribution. Pre-processing included resizing, data splitting, and augmentation in the form of rotation, width shift, height shift, and zoom to increase data diversity and reduce the impact of class imbalance. The model was trained using several training-validation data splits (70:30, 80:20, and 90:10) with variations of the Adam, RMSprop, and AdamW optimizers and learning rates between 0.1 and 0.0001. The best configuration was obtained in the 80:20 scenario with the Adam optimizer and a learning rate of 0.0001, which was used in the final testing stage using test data that were never used during training and validation. The results showed the highest validation accuracy of 99.89% and a testing accuracy of 98.00%. Confusion matrix analysis showed that all classes could be classified well without prediction bias.

Keywords: Brain Tumor; Data Augmentation; MRI; Transfer Learning; VGG16

Introduction

The brain plays a vital role in regulating bodily functions, such as heart rate, blood pressure, fluid balance, body temperature, and the functions of recognition, emotion, memory, motor learning, and various other forms of learning (Kusuma, 2025). Brain damage can trigger disorders, such as brain tumors (Ilawanda & Atsani, 2021). Brain tumors are abnormal cell growths in the brain, which are divided into two types: primary tumors and secondary tumors (Ardan & Indraswari, 2024). According to data from the WHO's Global Health Observatory (GHO), the number of brain tumor cases continues to increase annually, including in Indonesia. In 2021, more than 5.900.000 cases of brain tumors were recorded in Indonesia, with the mortality rate continuing to rise (Tyas et al., 2025). Early detection of brain tumors is crucial to increase the chances of recovery for brain tumor patients (Ijaz et al., 2025).

Several methods can be used to detect and diagnose tumors, one of which is Magnetic Resonance Imaging (MRI) (Abd-Ellah et al., 2019). Although MRI is more effective, the current process of interpreting MRI images still relies on the expertise of specialist doctors or radiologists, which can be time-consuming and prone to subjective errors, especially if the tumor is in the early stages of growth and does not have a clear shape (Essianda et al., 2023). Misdiagnosis can be fatal for patients with tumors, given the importance of early treatment in brain diseases (Susilo et al., 2025). Therefore, technology is needed to help improve the accuracy and efficiency of diagnosing brain diseases.

Along with the development of technology, the application of Artificial Intelligence (AI) technology in

How to Cite:
Noprianti, N. S. A. N., & Fitri, L. A. MRI Image Classification of Brain Tumors Using VGG16-Based Transfer Learning and Data Augmentation as a Medical Diagnosis Support System. *Jurnal Penelitian Pendidikan IPA*, *12*(1), 200-209. https://doi.org/10.29303/jppipa.v12i1.14218

deep learning methods, especially Convolutional Neural Networks (CNN), has proven effective in detecting brain tumors in MRI images (Tyas et al., 2025). CNN has been widely used in classification tasks because of its ability to automatically extract spatial features and produce better performance than conventional classification methods (Sumardi et al., 2025). CNN is also widely applied in medical image analysis because of its ability to automatically extract features and produce good performance in detecting image-based diseases (Islami & Defit, 2024).

CNN is a model designed to process data with a grid structure such as images and classify them based on learning from training data (Kurniawan et al., 2023; Nuresa, 2024). CNN has the advantage of capturing patterns in images, making it very effective in detecting and classifying brain tumors based on MRI imaging results (Nurhidayat et al., 2025; Sari et al., 2025). Digital images contain visual information that can be used as data for automatic analysis and classification, including in multiclass classification problems (Ariessaputra et al., 2024; Ependi & Ahmad, 2024). However, the application of CNN to MRI images of brain tumors still faces several challenges, such as limited training data, model complexity, and imbalance in the amount of data between classes.

One approach that can be used to address these challenges is transfer learning, which allows models to leverage knowledge learned from previous datasets to improve performance on new tasks. This can be a solution to overcome challenges such as large data requirements, long training times, and limited high-quality data (Nurlaela et al., 2025). Transfer learning can improve efficiency and accuracy in medical data analysis, particularly in detecting chronic diseases and classifying brain tumors (Velden et al., 2022). Therefore, transfer learning is a relevant and strategic approach in developing an MRI-based brain tumor classification system.

Research on MRI-based brain tumor classification has been extensively conducted and has significantly contributed to the development of artificial intelligence technology in the medical field. A previous study by Gayathri & Kumar (2024) demonstrated that the application of transfer learning based on the VGG-16 architecture was able to overcome overfitting issues and significantly improve the classification accuracy of brain tumor MRI images compared to conventional CNN methods. This demonstrates that the use of transfer learning is an important method used in the development of deep learning-based medical image classification systems.

Another study conducted by Rudiansyah & Husein (2024) used the Xception, VGG16, and MobileNet architectures and trained them in 50 epochs with a batch size of 32. In this study, the accuracy results were 95.00% in the Xception model, 92.00% in VGG-16, and 90.00% in MobileNet. Another study conducted by Wicaksono & Hartato (2025) used a dataset from Kaggle consisting of 253 MRI images. The study was conducted by comparing the InceptionV3 and VGG-16 models to classify brain cancer or not, and the accuracy results of the InceptionV3 model were 91.67% and the VGG-16 model had an accuracy of 82.14%. Meanwhile, research conducted by Arwinto et al. (2024) tested the CNN model based on VGG-16 and ResNet showed an accuracy of 96.07% and 87.52%. Although the VGG16 model produced high accuracy values, confusion matrix analysis showed that the model's performance was still relatively low for some classes. This indicates that high accuracy values do not fully reflect the model's ability to classify all classes equally, especially minority classes (Chen et al., 2024).

This situation indicates that using accuracy metrics alone is insufficient to comprehensively describe performance, especially in multi-class classification cases with imbalanced data (Sujon et al., 2025). In class imbalance conditions, metrics such as recall and F1-score become more relevant because they can represent the model's ability to fairly recognize each class, including minority classes that are often more clinically critical (Salmi et al., 2024). Therefore, model performance improvement should be measured not only by overall accuracy but also by the distribution of performance across classes. Based on this research gap, the novelty of this study lies in the application of a data augmentation strategy specifically designed to address class imbalance in MRI images of brain tumors. This study not only applies conventional augmentation such as rotation, flipping, and zooming, but also emphasizes improving the representation of minority classes through medically relevant spatial augmentation. This approach is expected to improve recall and F1-score for minority classes, resulting in more balanced performance compared to previous studies that focused solely on improving accuracy.

Based on the description, this research is entitled " MRI Image Classification of Brain Tumors Using VGG16 Based on Transfer Learning and Data Augmentation as a Medical Diagnosis Support System". This research aims to improve the model's generalization ability to new data, reduce the risk of overfitting and class imbalance problems, and produce more even and accurate performance on various types of brain tumors, so that the model does not only tend to recognize the majority class. This research contributes to the development of an AI-based medical decision support

system for faster, more accurate, and objective brain tumor diagnosis

Method

This research is a quantitative study with an experimental approach. Quantitative research was chosen because the data used were MRI images of brain tumors that were analyzed numerically to evaluate the classification model using computational methods and confusion matrices, such as accuracy, precision, recall, and F1-score to obtain an accurate classification model that is able to recognize and differentiate each type of brain tumor accurately (Aamir et al., 2025; Azhmy et al., 2024).

An overview of the brain tumor classification system in this study can be seen in Figure 1. Starting from the data collection stage, data processing (preprocessing) includes resizing, splitting, and data augmentation, model testing with certain hyperparameter and optimizer configurations, model testing using new data, to model performance evaluation based on classification results using accuracy, precision, recall, and F1-score metrics.

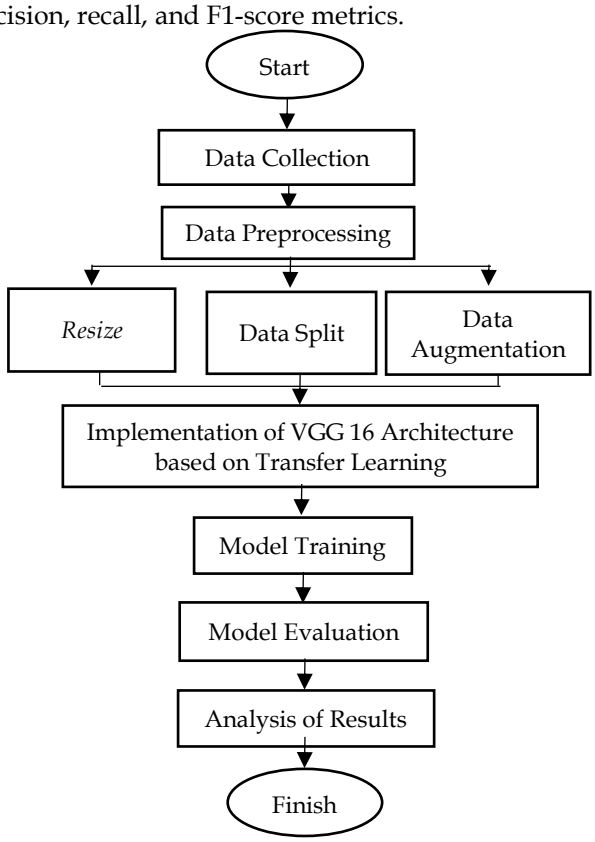


Figure 1. Research Flowchart

Data Collection

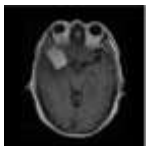
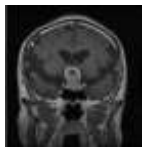
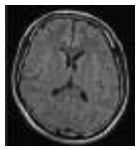
The data used in this study were obtained from the public Brain Tumor MRI dataset available on Kaggle and developed by Masoud Nickpravar. This dataset consists of four classifications: glioma, meningioma, pituitary, and no tumor. In total, 7.023.00 brain MRI images were obtained, structured into two directories: training and testing, eliminating the need for additional manual labeling. The number of datasets for each class of brain tumor disease can be seen in Table 1 as follows.

Table 1. Number of Datasets for Each Class in Brain Tumor Disease

Tumor Types	Number of Data
Glioma	1.621.00
Meningioma	1.645.00
Pituitary	1.445.00
No Tumor	2.312.00
Total	7.023.00

The dataset classes for brain tumor disease can be seen in Table 2.

Table 2. Dataset Classes

Glioma	Meningioma	Pituitary	No Tumor
			

Data Preprocessing

The resizing process was performed by resizing the original MRI image to 224 x 224 pixels to match the input size of the CNN model and the pre-trained VGG-16 model (Sadr et al., 2025). This size was chosen because it is the standard input size for the architecture used, simplifying the model training process using transfer learning on VGG-16. This is because the models already have the same size, thus maintaining data suitability and consistency without burdening the computational process (Deb & Rahman, 2025).

The dataset was divided using three training and testing data scenarios: 70:30, 80:20, and 90:10. The training data was used to build and train the model, the validation data was used to evaluate the model's performance during training to prevent overfitting, and the testing data was used to assess the model's performance on new data (Pratama et al., 2025). The use of several data ratios aimed to analyze the effect of different amounts of training data on the performance of the transfer learning-based VGG16 model. By comparing these three scenarios, the data sharing configuration that produces the most optimal brain tumor classification performance was determined (Bichri et al., 2024).

Data augmentation is a technique used to increase the diversity and quantity of training data, thereby

reducing the risk of overfitting and improving the generalizability of deep learning models (Islam et al., 2024). The augmentation process was applied to the training data in real time using the Image Data Generator module. In this study, the augmentation techniques used were random rotation up to 15.00°, width shift and height shift of 10.00%, and zoom range of up to 10.00%. This was done so that the model can recognize tumor patterns from various positions and understand various data variations (Shorten & Khoshgoftaar, 2019). Visual patterns such as texture and pixel intensity can be utilized as important information in the process of analyzing and classifying digital images (Rizki & Defit, 2024).

Implementation of CNN Architecture with Transfer Learning

This study uses the Visual Geometry Group 16 (VGG16) architecture as the primary architecture in the transfer learning approach. VGG16 is a Convolutional Neural Network (CNN) architecture consisting of 16 layers specifically designed for image classification (Umar, 2025). CNN-based deep learning models have the potential to support computer-aided diagnosis systems in improving the accuracy and efficiency of disease diagnosis (Mirza et al., 2025; Muttaqin & Sudiana, 2024). In this study, the initial layers of the VGG16 architecture were frozen to maintain the weights of the initial training results that can extract general features such as edges, shapes, and textures. Next, the final layers were modified and adjusted to the number of classes in this study, namely four classes of brain tumors, so that the model can perform multi-class classification optimally.

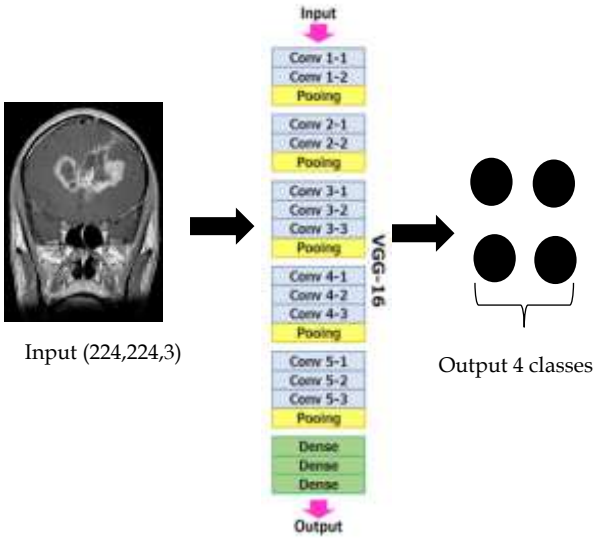


Figure 2. VGG16 Architecture

Model Training

The modified VGG16 model was then trained using MRI brain tumor image data. The training process utilized 5.712.00 training images and 1.311.00 testing images. Model training was conducted in several experimental scenarios, testing various optimizers, namely Adam, AdamW, and RMSprop, and varying learning rates of 0.1, 0.01, 0.001, and 0.0001. The model was trained for 50 epochs with a batch size of 32 for each optimizer and learning rate combination to obtain the training configuration that yielded the best performance. To avoid overfitting, an early stopping technique was applied, which terminates the training process if validation performance does not improve, and dropout was applied at certain layers to improve the model's generalization ability (Anam et al., 2024).

After the initial training was complete, fine-tuning was performed to improve the model's capabilities. The goal of fine-tuning is to allow the model to adapt to the characteristics of brain tumor MRI images, which differ from the general data in ImageNet (Davila et al., 2024). This allows the model to recognize tumor patterns more accurately and produce better classification results.

Model Evaluation

After the training process using various optimizers and learning rate combinations was completed, the best model was selected based on the hyperparameter configuration that produced the most optimal performance in classifying brain tumor types, as indicated by the highest accuracy value in the validation data. The selected model was then evaluated using a testing dataset that was not used in the training process, with the aim of assessing the model's generalization ability to new data (Xu & Goodacre, 2018). Model performance evaluation was carried out by calculating classification metrics such as accuracy, precision, recall, and F1-score for each brain tumor class (Prasetyo & Nabiilah, 2023). The use of these per-class metrics aims to analyze the effect of class imbalance, so that it can be determined whether the model has balanced performance across classes or tends to be better in the majority class than in the minority class (Farhadpour et al., 2024).

To monitor and identify overfitting, graphical visualizations of accuracy and loss on the training and validation data at each epoch were used. A model is considered overfitting if there is a significant difference between its performance on the training and validation data (Maspaeni et al., 2025). Through this analysis, the effectiveness of the data augmentation, regularization techniques, and hyperparameter configurations can be comprehensively evaluated. Furthermore, a confusion matrix was used to analyze the distribution of model predictions relative to the actual labels for each brain tumor. Confusion matrix analysis helps identify patterns of misclassification between classes, particularly in

minority classes, thus determining the extent to which the model can classify each type of brain tumor equally.

The calculation of these evaluation metrics is based on the following standard formulas:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (4)$$

TP is the number of positive data items correctly classified, TN is the number of negative data items correctly classified, FP is the number of positive data items incorrectly classified, and FN is the number of negative data items incorrectly classified. These definitions are crucial for evaluating the performance and accuracy of predictive models in various research fields and applications.

Table 3. Accuracy Level (Gorunescu, 2011)

Category	Accuracy Level %
Very Good	90.00-100.00
Good	80.00-90.00
Pretty Good	70.00-80.00
Not Good	60.00-70.00
Fail	50.00-60.00

Result and Discussion

Preprocessing Result

The initial preprocessing stage involves dataset exploration to check the amount of data between classes in both the training and testing data. This stage aims to identify data imbalances between classes. The dataset exploration results indicate an imbalance in the amount of data between classes, especially in the training data. Therefore, a data augmentation stage is necessary to balance the number of images in each class and increase the variety of training data to avoid model bias during the training process.

Next, all images were resized to the same size of 224 x 224 pixels to meet the standard input requirements of the VGG16 architecture. The resizing process was performed automatically using ImageDataGenerator with the parameter image_size=(224,224). Furthermore, pixel values were normalized across all images by dividing each pixel value by 255, resulting in values within the [0.1] range. This normalization process has proven to be a consistent and effective normalization method for modern CNN architectures because it maintains gradient stability during the optimization process. This normalization aims to accelerate model convergence and improve learning stability during the training process.

Augmentation Result

To increase the amount and variety of data in the dataset, address class imbalance, and prevent overfitting, data augmentation was performed on the training data. In this study, augmentation was applied to MRI images of brain tumors to balance the distribution of data between glioma, meningioma, pituitary, and non-tumor brain tumors. The augmentation process was only performed on the training data because the testing data must reflect real data that has never been seen by the model before (Gholamy et al., 2018). The augmentation technique was performed by modifying the original image through several parameters such as rotation, translation, shear, zoom, and flip (Shorten & Khoshgoftaar, 2019).

This data augmentation was applied using ImageDataGenerator to balance the amount of data in each class according to the amount of data in the largest class, thus achieving a balanced data distribution. The results of this argument yielded an equal amount of data in each class, as can be seen in Table 4.

Table 4. Number of Datasets Before and After Augmentation (Training Data)

Training Data	Before Augmentation	After Augmentation
Glioma	1.321.00	1.595.00
Meningioma	1.339.00	1.595.00
Pituitary	1.457.00	1.595.00
No tumor	1.595.00	1.595.00

Based on Table 4, after the augmentation process, the amount of data in each class was balanced. This aims to avoid potential model bias towards certain classes during the training process. To balance the minority class, images from the glioma, meningioma, and pituitary classes, which are fewer in number than the no-tumor class, were then augmented so that their number was closer to the majority class. Each augmented image was then combined back with the original dataset to balance the distribution. An imbalance in the amount of data between classes can affect model performance, as the model tends to learn patterns from the majority class more easily than the minority class (Chazar et al., 2025).

Meanwhile, no augmentation was performed on the testing data to ensure that the model performance evaluation reflects real-world conditions. The number of testing datasets for each class is shown in Table 5.

Table 5. Number of Datasets Before and After Augmentation (Testing Data)

Testing Data	Before Augmentation	After Augmentation
Glioma	300.00	Not Augmented
Meningioma	306.00	Not Augmented
Pituitary	300.00	Not Augmented
No tumor	405.00	Not Augmented

After the augmentation process was complete, the training data were split again into training data and validation data using the train-validation split method with several ratio variations, namely 80:20, 70:30, and 90:10. The training data split scenarios can be seen in Table 6. Testing of several data split ratios aims to find the configuration that produces the best model performance. The data split process is carried out randomly but remains consistent with the random_state setting of 42. Validation data are used to monitor model performance during training, while testing data are used separately for the final evaluation of the best model.

Table 6. Training Data Sharing Scenario

Scenario	Training	Validation
70:30	70.00%	30.00%
80:20	80.00%	20.00%
90:10	90.00%	10.00%

Hyperparameters Testing

The model was trained using the categorical cross-entropy loss function, which is suitable for multi-class classification problems, and accuracy was used as an evaluation metric. The batch size used was 32, with a maximum number of epochs of 50 epochs. During the training process, model performance was monitored using validation data to prevent overfitting. In addition, several callback processes were implemented during the training process, such as early stopping, model checkpoints, and Reduce Learning Rate on Plateau to improve efficiency and maintain model quality during training. Early stopping was used to stop the training process when the model showed no change in seven consecutive epochs, model checkpoints functioned to store the best model weights based on the highest validation accuracy value, and ReduceLROnPlateau was used to automatically reduce the learning rate when the validation loss value showed stagnation.

The results of model training with various optimizers, learning rates, and data splits are shown in Table 6. In test 16 with an 80:20 data split, the Adam optimizer and a learning rate of 0.0001 provided the best results with an accuracy of 99.89%. This trial aimed to find the best hyperparameter configuration that performed best in classifying brain tumor types based on the accuracy values of each configuration.

Table 7. Model Training Result

Data Split	Optimizer	Learning Rate	Batch Size	Accuracy (%)
90:10	Adam	0.1	32	24.61
90:10	Adam	0.01	32	24.95
90:10	Adam	0.001	32	99.36
90:10	Adam	0.0001	32	99.73
90:10	AdamW	0.1	32	25.72
90:10	AdamW	0.01	32	72.68
90:10	AdamW	0.001	32	99.81
90:10	AdamW	0.0001	32	99.83
90:10	RMSprop	0.1	32	25.45
90:10	RMSprop	0.01	32	25.73
90:10	RMSprop	0.001	32	99.69
90:10	RMSprop	0.0001	32	99.85
80:20	Adam	0.1	32	25.44
80:20	Adam	0.01	32	24.02
80:20	Adam	0.001	32	99.75
80:20	Adam	0.0001	32	99.89
80:20	AdamW	0.1	32	24.63
80:20	AdamW	0.01	32	24.92
80:20	AdamW	0.001	32	99.78
80:20	AdamW	0.0001	32	99.75
80:20	RMSprop	0.1	32	25.44
80:20	RMSprop	0.01	32	25.41
80:20	RMSprop	0.001	32	99.43
80:20	RMSprop	0.0001	32	99.63
70:30	Adam	0.1	32	24.70
70:30	Adam	0.01	32	25.21
70:30	Adam	0.001	32	99.62
70:30	Adam	0.0001	32	99.73
70:30	AdamW	0.1	32	24.74
70:30	AdamW	0.01	32	24.94
70:30	AdamW	0.001	32	99.35
70:30	AdamW	0.0001	32	99.76
70:30	RMSprop	0.1	32	24.75
70:30	RMSprop	0.01	32	25.07
70:30	RMSprop	0.001	32	99.03
70:30	RMSprop	0.0001	32	99.54

Evaluation Result

Based on previous hyperparameter testing results, the transfer learning-based VGG16 architecture with the Adam optimizer, a learning rate of 0.0001, and an 80:20 data split demonstrated the best accurate performance. The test results showed that this configuration produced the highest training accuracy of 99.89%, with the process steps described below.

Training Result

The training and validation results are displayed graphically in Figure 4, which shows the accuracy and loss of the VGG16 model based on the hyperparameter configurations applied during the training process. Training accuracy consistently increased from the initial epoch to nearly 100%, indicating that the model was able to learn the training data patterns very well.

Meanwhile, validation accuracy also steadily increased, reaching a maximum value of 99.89% during the training process, despite minor fluctuations in some epochs. This trend indicates that the model has good generalization ability to the validation data.

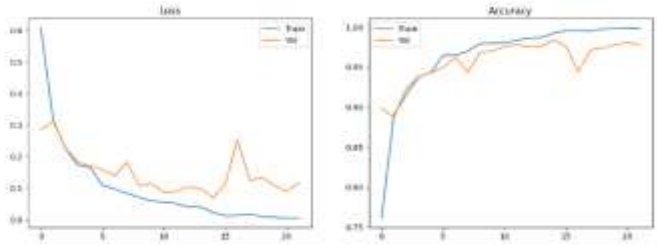


Figure 3. VGG16 Architectural Graphics

The loss graph shows a significant decrease in training loss, from around 0.60 to near zero as the number of epochs increases. Validation loss also tends to decrease and remains at a relatively low value, despite experiencing slight fluctuations. The difference between training loss and validation loss remains within reasonable limits, thus not indicating significant overfitting. Thus, the VGG16 model has good generalization capabilities and is suitable for use in brain tumor classification.

Testing Result

The testing process was conducted to evaluate the model's accuracy in classifying the testing data, which consisted of four classes and 1,311,000 images. To measure the accuracy level, a Confusion Matrix was used to calculate the accuracy, precision, recall, and F1-score values, as shown in Table 8 and Figure 4.

Table 8. VGG16 Classification Results

Tumor Types	Precision	Recall	F1-score
Glioma	0.99	0.95	0.97
Meningioma	0.94	0.97	0.95
Pituitary	0.99	0.99	0.99
No Tumor	0.99	1.00	0.99
Accuracy			0.98
Macro Avg	0.98	0.98	0.98
Weight Avg	0.98	0.98	0.98

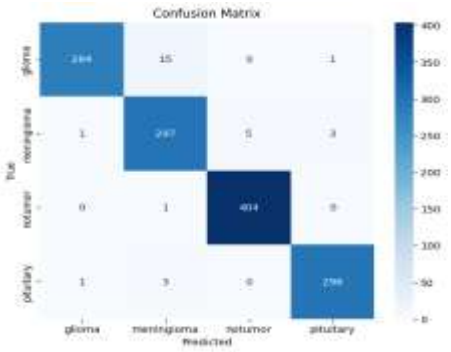


Figure 4. Confusion Matrix VGG16

Based on Table 8 and Figure 4, the model demonstrated consistent performance across all classes. In the glioma class, 284.00 out of 300.00 images were correctly classified (recall 0.95), indicating that most of the key characteristics of gliomas were accurately recognized by the model. Small misclassification errors occurred due to visual similarities between tumor types, not due to the model's inability to extract important features. The meningioma class showed stable performance with 297.00 out of 306.00 images correctly classified (recall 0.97). The pituitary class successfully classified 296.00 out of 300.00 images with precision and recall values of 0.99 each, while the no tumor class achieved the highest performance with 404.00 out of 405.00 images correctly classified (recall 1.00). Overall, the model achieved accuracy, macro average precision, recall and F1-score all reaching 98.00% during the testing process.

These results confirm that the proposed transfer learning-based VGG16 approach can produce balanced and stable classification performance on datasets with imbalanced class distributions. The high recall achieved across all classes, including glioma, meningioma, pituitary, and no tumor, indicates that the model is effective in recognizing diverse visual characteristics, both in classes with complex tumor morphology variations and in non-tumor classes. The consistent performance across classes indicates that the model does not show bias towards the majority or minority class and is able to comprehensively learn important spatial features, making it relevant to support the classification of brain tumor MRI images in the medical field.

Conclusion

This study successfully applied the transfer learning-based VGG16 architecture combined with data augmentation for multi-class classification of brain tumor MRI images. With the best configuration using the Adam optimizer, a learning rate of 0.0001, and a data split of 80:20, the model demonstrated high and stable performance. The final evaluation on 1,311,000 MRI images of test data that were never used in the training process resulted in an accuracy of 98.00%, with macro average precision, recall, and F1-score values also reaching 98.00%, indicating good model generalization ability to new data. The application of data augmentation has been proven to improve model stability and reduce the impact of class imbalance by enriching the variety of training data and balancing the class distribution. This is indicated by an increase in the recall value in the minority class (glioma) to 0.95, which is comparable to the majority class, thereby reducing prediction bias. However, this study is limited to the use

of the VGG16 architecture, so further research is recommended to conduct testing and comparisons with other architectures such as ResNet, EfficientNet, or Xception, as well as using clinical datasets from different sources to improve the validity and applicability of the model in real medical contexts.

Acknowledgments
The author would like to thank all parties who have helped and supported the implementation of this research.

Author Contributions
Conceptualization, provision of resources, data curation, and initial writing were performed by N.S.A.N. Methodology, formal analysis, review, and editing were performed by N.S.A.N. and L.A.F.

Funding
This research did not receive funding from any external party.

Conflicts of Interest
All authors declare that there is no conflict in this research.

References

Aamir, M., Rahman, Z., Bhatti, U. A., Abro, W. A., Bhutto, J. A., & He, Z. (2025). An automated deep learning framework for brain tumor classification using MRI imagery. *Scientific Reports*, *15*(1), 17593. https://doi.org/10.1038/s41598-025-02209-2

Abd-Ellah, M. K., Awad, A. I., Khalaf, A. A., & Hamed, H. F. (2019). A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. *Magnetic Resonance Imaging*, *61*, 300–318. https://doi.org/10.1016/j.mri.2019.05.028

Anam, M. K., Defit, S., Haviluddin, H., Efrizoni, L., & Firdaus, M. B. (2024). Early stopping on CNN-LSTM development to improve classification performance. *Journal of Applied Data Sciences*, *5*(3), 1175–1188. https://doi.org/10.47738/jads.v5i3.312

Ardan, I. S., & Indraswari, R. (2024). Sistem Berbasis Deep Learning untuk Segmentasi dan Klasifikasi Tingkat Keganasan Tumor Otak Menggunakan Citra MRI 3D. *ILKOMNIKA*, *6*(2), 117–126. https://doi.org/10.28926/ilkomnika.v6i2.643

Ariessaputra, S., Vidiasari, V. H., Al Sasongko, S. M., Darmawan, B., & Nababan, S. (2024). Classification of Lombok Songket and Sasambo Batik Motifs Using the Convolution Neural Network (CNN) Algorithm. *JOIV: International Journal on Informatics Visualization*, *8*(1), 38–44. https://doi.org/10.62527/joiv.8.1.1386

Azhmy, M. F., Suyar, A. S., & Setiamy, A. A. (2024). Digital HRM with Machine Learning Approach. *Jurnal Penelitian Pendidikan IPA*, *10*(SpecialIssue), 18–23. https://doi.org/10.29303/jppipa.v10iSpecialIssue.7213

Bichri, H., Chergui, A., & Hain, M. (2024). Investigating the Impact of Train/Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets. *International Journal of Advanced Computer Science & Applications*, *15*(2). Retrieved from https://shorturl.asia/QpZ0n

Chazar, C., Adli, M. A., Pardede, J., & Ichwan, M. (2025). Pendekatan Augmentasi Citra Fundus pada Model EfficientNet untuk Klasifikasi Tingkat Keparahan Retinopati Diabetik dengan Dataset Tidak Seimbang. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, *10*(2), 180–194. https://doi.org/10.26760/mindjournal.v10i2.180-194

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, *57*(6), 137. https://doi.org/10.1007/s10462-024-10759-6

Davila, A., Colan, J., & Hasegawa, Y. (2024). Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image and Vision Computing*, *146*, 105012. https://doi.org/10.1016/j.imavis.2024.105012

Deb, N., & Rahman, T. (2025). An efficient VGG16-based deep learning model for automated potato pest detection. *Smart Agricultural Technology*, *101409*. https://doi.org/10.1016/j.atech.2025.101409

Ependi, U., & Ahmad, N. A. (2024). A Novel Hybrid Classification on Urban Opinion Using ROS-RF: A Machine Learning Approach. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 5816–5824. https://doi.org/10.29303/jppipa.v10i8.8042

Essianda, V., Indrasari, A. D., Widyastuti, P., Syahla, T., & Rohadi, R. (2023). Brain Tumor: Molecular Biology, Pathophysiology, and Clinical Symptoms. *Jurnal Biologi Tropis*, *23*(4), 260–269. https://doi.org/10.29303/jbt.v23i4.5585

Farhadpour, S., Warner, T. A., & Maxwell, A. E. (2024). Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices. *Remote Sensing*, *16*(3), 533. https://doi.org/10.3390/rs16030533

Gayathri, T., & Kumar, S. (2024). Brain tumor segmentation and classification using cnn pre-trained vgg-16 model in mri images. *IIUM Engineering Journal*, *25*(2), 196–211. https://doi.org/10.31436/iiumej.v25i2.2963

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons*. Retrieved from https://scholarworks.utep.edu/cs_techrephttps://scholarworks.utep.edu/cs_techrep/

Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.

Ijaz, M., Hasan, I., Aslam, B., Yan, Y., Zeng, W., Gu, J., & Guo, B. (2025). Diagnostics of brain tumor in the early stage: current status and future perspectives. *Biomaterials Science*. https://doi.org/10.1039/D4BM01503G

Ilawanda, Z. M., & Atsani, G. F. (2021). Gambaran Radiologis pada Bidang Neurologis Tumor Otak. *Jurnal Syntax Fusion*, *1*(12), 987–1001. Retrieved from https://shorturl.asia/8omJc

Islam, T., Hafiz, M. S., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2024). A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. *Healthcare Analytics*, *5*, 100340. https://doi.org/10.1016/j.health.2024.100340

Islami, F., & Defit, S. (2024). Customized Convolutional Neural Network for Glaucoma Detection in Retinal Fundus Images. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 4606–4613. https://doi.org/10.29303/jppipa.v10i8.7614

Kurniawan, M. A., Wijaya, S. K., & Hanifa, N. R. (2023). Convolutional neural network for earthquake ground motion prediction model in earthquake early warning system in West Java. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 1004–1010. https://doi.org/10.29303/jppipa.v9i11.3514

Kusuma, A. M. (2025). Musik sebagai Sarana Terapi dalam Pendidikan: Perspektif Neuroscience dan Neuroeducation. *Irama: Jurnal Seni, Desain Dan Pembelajarannya*, *7*(1), 33–41. https://doi.org/10.17509/irama.v7i1.80905

Maspaeni, M., Imran, B., Hidayat, A., & Erniwati, S. (2025). Implementasi Machine Learning untuk Mendeteksi Penyakit Katarak menggunakan Kombinasi Ekstraksi Fitur dan Neural Network Berdasarkan Citra. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, *7*(2), 232–251. https://doi.org/10.35746/jtim.v7i2.621

Mirza, A. H., Halim, R. N., & Hasbiallah, M. J. (2025). Development of an Efficient 1D-CNN Model for Myocardial Infarction Classification Using 12-Lead ECG Signals. *Jurnal Penelitian Pendidikan IPA*, *11*(3), 1167–1182. https://doi.org/10.29303/jppipa.v11i3.10238

Muttaqin, R. Z., & Sudiana, D. (2024). Design of Realtime Web Application Firewall on Deep Learning-Based to Improve Web Application Security. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 11121–11129. https://doi.org/10.29303/jppipa.v10i12.8346

Nuresa, K. (2024). Analisis Perbandingan Klasifikasi Tumor Otak Menggunakan Deep Learning. *IJITECH: Indonesian Journal of Information Technology*, *2*(1), 16–21. https://doi.org/10.71155/rdgrt890

Nurhidayat, A., Arrosyid, W. A., & Samsinar, R. (2025). Prediksi Tumor Otak Menggunakan Metode Convolutional Neural Network (CNN) dan Algoritma Decision Tree. *Seminar Nasional Teknologi & Sains*, *4*(1), 660–666. https://doi.org/10.29407/d3822f90

Nurlaela, L., Suhanda, Y., Sopian, A., Dewi, C. S., & Syahrial, R. (2025). Pengembangan Framework Data Mining Berbasis Deep Neural Network Dengan Eksplorasi Teknik Transfer Learning Untuk Prediksi Dan Klasifikasi Data. *JRIS: Jurnal Rekayasa Informasi Swadharma*, *5*(1), 132–141. https://doi.org/10.56486/jris.vol5no1.723

Prasetyo, S. Y., & Nabiilah, G. Z. (2023). Perbandingan Model Machine Learning pada Klasifikasi Tumor Otak Menggunakan Fitur Discrete Cosine Transform. *Jurnal Teknologi Terpadu*, *9*(1), 29–34. https://doi.org/10.54914/jtt.v9i1.605

Pratama, M. H. R., Akrom, M., Santosa, A. P., Rosyid, M. R., & Mawaddah, L. (2025). Klasifikasi Otomatis Korosi Menggunakan Convolutional Neural Network dan Transfer Learning dengan Model MobileNetV2. *Jurnal Algoritma*, *22*(1), 138–148. https://doi.org/10.33364/algoritma/v.22-1.2182

Rizki, S. D., & Defit, S. (2024). Texture and Flag Color Extraction in Backpropagation Neural Network Architecture. *Jurnal Penelitian Pendidikan IPA*, *10*(5), 2191–2198. https://doi.org/10.29303/jppipa.v10i5.7146

Rudiansyah, R., & Husein, A. (2024). Klasifikasi Tumor Otak pada gambar Magnetic Resonance Images (MRI) dengan Pendekatan Pembelajaran Mendalam. *Data Sciences Indonesia (DSI)*, *4*(1), 62–68. https://doi.org/10.47709/dsi.v4i1.4265

Sadr, H., Nazari, M., Yousefzade-Chabok, S., Emami, H., Rabiei, R., & Ashraf, A. (2025). Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate classification and diagnosis. *Image and Vision Computing*, *159*. https://doi.org/10.1016/j.imavis.2025.105555

Salmi, M., Atif, D., Oliva, D., Abraham, A., & Ventura, S. (2024). Handling imbalanced medical datasets: review of a decade of research. *Artificial Intelligence Review*, *57*(10), 273. https://doi.org/10.1007/s10462-024-10884-2

Saputro, Y., Arwinto, M. R. A., Susanti, A., & Agustin, T. (2024). Analisis Performa Model Deep Learning VGG16 dan ResNet dalam Klasifikasi Jenis Tumor Otak. In *Prosiding Seminar Nasional Amikom Surakarta* (Vol. 2, pp. 136–147). Retrieved from https://ojs.amikomsolo.ac.id/index.php/semnasa/article/view/633

Sari, M. W., Sitorus, S. P., & Pane, R. (2025). Implementation of Convolutional Neural Network (CNN) Method in Determining the Level of Ripeness of Mango Fruit Based on Image. *Jurnal Penelitian Pendidikan IPA*, *11*(5), 419–428. https://doi.org/10.29303/jppipa.v11i5.11436

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Sujon, K. M., Hassan, R., Choi, K., & Samad, M. A. (2025). Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. *Journal of Big Data*, *12*(1), 268. https://doi.org/10.1186/s40537-025-01313-4

Sumardi, T., Robiyana, I., Permana, R., & Suhendra, M. A. (2025). Thermal Image-Based Classification of Okra Maturity: A Comparative Study of CNN, SVM, and LSTM. *Jurnal Penelitian Pendidikan IPA*, *11*(11), 582–589. https://doi.org/10.29303/jppipa.v11i11.12748

Susilo, R. I., Haq, I. B. I., Apriawan, T., Wahyuhadi, J., Puspikawati, S. I., & Ramadhaningtyas, E. M. (2025). Education Regarding Early Detection And Early Treatment Of Brain Tumors In Banyuwangi Regency, Indonesia. *Journal of Public Health Research & Community Health Development*, *9*(1). Retrieved from https://e-journal.unair.ac.id/JPHRECODE/article/download/68478/34156

Tyas, D. L., Rumambi, F. R., Patanduk, A., & Mailangkay, R. C. J. (2025). Klasifikasi Jenis Tumor Otak Melalui Citra MRI dengan Menggunakan Convolutional Neural Network. *Informatik: Jurnal Ilmu Komputer*, *21*(1), 26–34. https://doi.org/10.52958/iftk.v21i1.11095

Umar, R. (2025). Advanced Chicken Breed Identification Using Transfer Learning Techniques with the VGG16 Convolutional Neural Network Architecture. *Jurnal Penelitian Pendidikan IPA*, *11*(7), 152–159. https://doi.org/10.29303/jppipa.v11i7.11870

Velden, B. H., Kuijf, H. J., Gilhuijs, K. G., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis*, *79*, 102470. https://doi.org/10.1016/j.media.2022.102470

Wicaksono, A. B., & Hartato, B. P. (2025). Analisis performa arsitektur cnn inceptionv3 dan vgg16 dalam klasifikasi deteksi kanker otak. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika*, *10*(2), 938–948. https://doi.org/10.29100/jipi.v10i2.6090

Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, *2*(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2