IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 5106~5119
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp5106-5119

Optimizing brain tumor MRI classification using advanced preprocessing techniques and ensemble learning methods

Akim Manaor Hara Pardede[1], Ahmad Zamsuri[2], Indi Nuroini[3], Putrama Alkhairi[4]
[1]Information Systems Study Program, STMIK Kaputama, Binjai, Indonesia
[2]Master Program of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia
[3]Faculty of Law, Universitas Bhayangkara Surabaya, Surabaya, Indonesia
[4]Informatics Management, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Article Info

Article history:

Received Jan 11, 2025
Revised Oct 27, 2025
Accepted Nov 8, 2025

Keywords:

Bagging
Brain tumor classification
Ensemble learning
Medical image analysis
MRI images
ResNet

ABSTRACT

Brain tumor classification is a critical task in medical imaging that directly impacts the accuracy of diagnosis and treatment planning. However, the complexity and variability of magnetic resonance imaging (MRI) images pose significant challenges, often resulting in reduced model reliability and generalization. This study addresses these limitations by proposing a novel ResNet+Bagging model, leveraging the strengths of residual networks and ensemble learning to enhance classification performance. Using publicly available brain tumor MRI datasets, including images labeled as benign, malignant, and normal, the study employs advanced preprocessing techniques such as normalization, data augmentation, and noise reduction to ensure high-quality inputs. The proposed model demonstrated significant improvements, achieving the highest testing accuracy of 72%, outperforming other tested models such as LeNet, standard ResNet, GoogleNet, and VGGNet. Precision (0.6010), recall (0.6000), and F1-score (0.5990) metrics further highlight its superior balance in detecting positive and negative classes. The novelty of this research lies in the application of Bagging to ResNet, which effectively mitigates overfitting and enhances predictive stability in complex medical datasets. These findings underscore the proposed model's potential as a robust solution for brain tumor classification, contributing to more accurate and reliable diagnostics.

Corresponding Author:

Akim Manaor Hara Pardede
Information Systems Study Program, STMIK Kaputama
St. Veteran No. 4A-9A, Binjai-North Sumatera, Indonesia
Email: akimmhp@live.com

1. INTRODUCTION

Brain tumors are abnormal cell growths within or around brain tissue that can be either benign (non-cancerous) or malignant (cancerous) [1], [2]. This condition can interfere with vital brain functions, such as cognitive abilities, motor coordination, and sensory systems, depending on the tumor's location and size. Brain tumors can develop primarily in the brain or as metastases from cancer in other parts of the body. Symptoms vary, ranging from worsening headaches, seizures, blurred vision, to memory impairments [3], [4]. Early diagnosis is crucial for improving the chances of effective treatment, and medical imaging such as magnetic resonance imaging (MRI) is a key tool for detecting and evaluating brain tumors. With the right approach, including medical interventions and the use of artificial intelligence (AI) for image analysis, patient recovery chances can significantly improve. Brain cancer is one of the diseases with a high mortality rate, especially if not detected and treated at an early stage. Early detection of brain tumors plays an important role in increasing patients' chances of recovery. In medical practice, MRI is one of the most reliable imaging techniques for detecting and analyzing brain tumors [5]–[7]. However, manual analysis of MRI images is often time-consuming, requires high expertise, and is prone to human error. Therefore, more effective and efficient approaches are needed to improve the accuracy of brain tumor classification [8], [9].

In brain tumor classification using imaging techniques such as MRI, several deep learning architectures have proven to be reliable and widely used in research. Convolutional neural networks (CNNs), such as LeNet [10], AlexNet [11], VGG-16 [12], [13] and ResNet [14], [15] are frequently employed due to their ability to capture spatial features of images. LeNet is suitable for small datasets due to its simplicity, while AlexNet and VGG16 are better at handling more complex images due to their deeper network architecture [16], [17]. ResNet, with its skip connections, is capable of addressing the degradation problem in very deep networks, thus improving classification accuracy. Furthermore, transfer learning-based architectures, such as InceptionV3 and EfficientNet, are also popular because they allow the use of pre-trained models on large datasets, which can then be fine-tuned for brain tumor datasets. The combination of CNNs with ensemble learning methods, such as Bagging, has also proven effective in improving classification accuracy by combining predictions from multiple models.

Previous research by [18] in this article has both strengths and weaknesses. The main advantage of this article is the use of advanced deep learning models such as 3D U-Net, PSPNet, and DeepLabV3+, which have shown promising results in brain tumor segmentation from MRI images, with 3D U-Net achieving the highest dice similarity coefficient (DSC) of 0.90. Additionally, the article discusses the importance of data augmentation and transfer learning techniques in improving model accuracy, which have proven effective in enhancing model performance. However, the article's drawbacks include the use of a single dataset, namely BraTS 2018, which may limit the generalizability of the findings, and a lack of in-depth hyperparameter evaluation. Moreover, although 3D U-Net demonstrated the best performance, this model also has high computational requirements and longer training times compared to other models like ResNet50, which is easier to implement. On the other hand, the study by [19] has both significant strengths and weaknesses. The main advantage of this study is the use of transfer learning methods that have proven to improve brain tumor classification accuracy, with the VGG-16 model achieving the highest accuracy of 97% and the shortest processing time among the models tested, at 22% of the total time. Additionally, the study performs a comprehensive comparative analysis of transfer learning models such as VGG-16, MobileNet, and ResNet-50, providing valuable insights into the strengths and efficiencies of each model. However, the study's weaknesses include the limited availability of diverse datasets, which can affect classification accuracy and introduce bias, as well as challenges in generalizing the model for rare tumor subtypes.

In recent years, advancements in AI technology have made significant contributions to medical image analysis [20], [21]. Specifically, CNNs have proven their superiority in processing image data, including in the detection and classification of brain tumors. However, the accuracy of CNNs in brain tumor classification can still be improved by applying advanced preprocessing techniques and ensemble learning methods [22], [23]. Based on the literature review outlined above, this study offers a solution by optimizing the brain tumor classification process through the application of advanced preprocessing techniques and ensemble learning methods [24], [25]. Preprocessing techniques such as normalization, data augmentation, and noise removal aim to enhance the quality of MRI images before they are processed by the model. This process results in more representative features, helping the model better understand relevant patterns [19], [26]. On the other hand, ensemble learning methods such as Bagging are used to improve the performance of the classification model by combining predictions from multiple base models, thus reducing the risk of overfitting and improving overall accuracy. By integrating deep learning architectures like LeNet with ensemble techniques, this study aims to create a more accurate, reliable, and efficient classification system [27], [28].

This study aims to develop an MRI-based brain tumor classification model with high accuracy by optimizing the combination of preprocessing techniques and ensemble learning methods. Specifically, this research seeks to analyze the impact of preprocessing techniques on model performance, evaluate the effectiveness of Bagging methods in improving classification accuracy, and identify the best configuration between deep learning models and ensemble learning methods. The expected final outcome is the creation of a system that not only excels in accuracy but also can be practically implemented to support medical diagnoses, contributing to the early detection of brain tumors and improving patient care. Advanced preprocessing techniques such as image normalization, data augmentation, and noise removal can help improve the quality of input to the model, resulting in more representative features. On the other hand, ensemble learning methods like Bagging offer an approach to combine predictions from several models to improve accuracy and reduce the risk of overfitting. However, the application of these methods in brain tumor classification has not been fully optimized, particularly in combining preprocessing techniques with ensemble learning approaches [6]. Therefore, this study aims to develop a more accurate MRI-based brain tumor classification model by optimizing preprocessing techniques and ensemble learning methods. The resulting model is expected to not only improve classification accuracy but also support medical practice in providing faster and more reliable diagnoses.

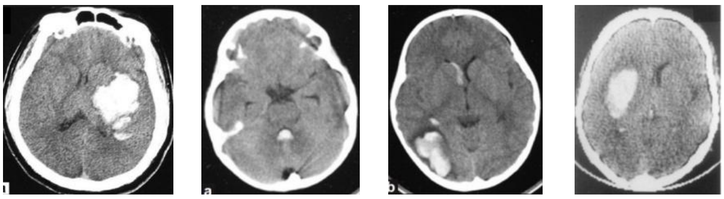
2. METHOD

This study employs a quantitative experimental approach to develop and compare the performance of brain tumor classification models using various deep learning architectures. The primary focus of the research is to compare the proposed ResNet+Bagging model with individual models such as GoogLeNet, ResNet, LeNet, and VGG-16. Experiments are conducted using a brain tumor MRI image dataset that has undergone preprocessing to ensure the quality of the input data.

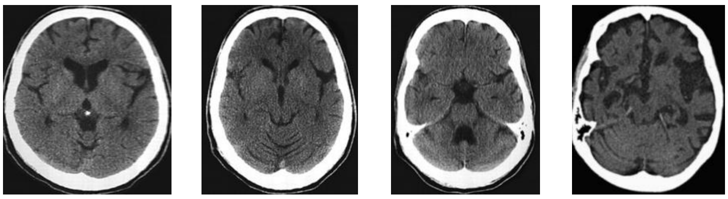
2.1. Research dataset

The dataset used in this study consists of brain tumor MRI images, which include medical images of various types of brain tumors. The data is sourced from Kaggle on the website: https://www.kaggle.com/code/guslovesmath/cnn-brain-tumor-classification-99-accuracy/input with 3 variables as the training model, with a total of 1,200 data. This dataset typically comprises a series of high-resolution images, enabling in-depth processing and analysis to identify patterns in the brain tumors. These MRI images are generated using magnetic resonance imaging, which clearly displays the detailed structure of the brain. Typically, the dataset is labeled as hemorrhage and non-hemorrhage, which facilitates the training and evaluation of classification models, as shown in Figure 1.

Figure 1 shows images in the dataset that vary in terms of tumor size, intensity, and shape, making preprocessing essential to enhance the quality of input for deep learning models. Several public datasets used in this study include the brain MRI images dataset available on Kaggle and the brain tumor segmentation challenge (BRATS) (dataset, which are frequently utilized in research related to brain tumor segmentation and classification. These datasets contain a wide range of images, covering various stages of tumor development and patient conditions, enabling the model to learn to recognize the diverse patterns present in brain tumors. The use of high-quality datasets with rich variation is crucial for improving the model's ability to accurately classify brain tumor MRI images. The MRI images show two categories of brain conditions, namely hemorrhage as shown in Figure 1(a) and non-hemorrhage as shown in Figure 1(b). In category (a) hemorrhage (top row), there is an area with bright intensity (white) indicating blood accumulation due to bleeding, usually irregular in shape according to the location and cause. Meanwhile, category (b) non-hemorrhage (bottom row) shows an MRI of the brain without indication of bleeding, with more uniform pixel intensity and normal-looking brain structures, although it may still indicate other abnormalities. These two categories play an important role in helping clinical diagnosis to determine the type of brain disorder and appropriate treatment.



(a)



(b)

Figure 1. Sample brain tumor MRI (a) hemorrhage, and (b) non-hemorrhage

2.2. Proposed model

In the proposed ResNet+Bagging model, Bagging (bootstrap aggregating) is applied to enhance model performance by combining predictions from multiple independently trained ResNet models on different subsets of the data. The core concept of Bagging is to generate several models by randomly sampling from the training dataset using bootstrap sampling, which allows each model to learn from a different subset of data. Each ResNet model provides an independent prediction, which is then combined using majority voting for classification. Hyperparameters: learning rate=0.001, optimizer=Adam, batch size=32, epochs=50. Number of ResNet base learners in Bagging=5. This approach aims to improve generalization and reduce the likelihood of overfitting by decreasing the variance of individual model predictions. Bagging applied to ResNet helps improve stability and accuracy, particularly when dealing with subtle differences between models, which is often encountered in medical image classification. Table 1 presents a comparison of the proposed model with other models.

In Table 1 of the proposed model, the data preprocessing stage plays a crucial role in ensuring the quality of the brain tumor MRI images used for model training. This preprocessing process involves several techniques, such as image normalization, which scales the pixel values to the range [0, 1] to improve model convergence, and data augmentation, including rotation, flipping, zooming, and cropping, aimed at expanding the dataset's variation and enhancing the model's ability to generalize. Additionally, the images are resized to match the model's input layer, such as 224×224 for ResNet and 28×28 for LeNet. Another step taken is noise reduction, to remove irrelevant disturbances in the images that could reduce the model's accuracy. All of these steps are designed to ensure that the data used is of high quality and ready to be processed by the model, thus improving its ability to accurately classify brain tumors.

The Bagging technique is applied by constructing multiple ResNet models that are trained in parallel on different data subsets. Each model provides its prediction, which is then combined through majority voting to determine the final outcome. This approach aims to improve generalization and reduce the risk of overfitting. Bagging enhances ResNet's ability to recognize features and patterns in brain tumor MRI images by reducing reliance on a single model that may be prone to overfitting. By combining several ResNet models, the final classification result becomes more accurate as it is less influenced by errors or noise within specific data subsets. In comparison, the ResNet model without Bagging is more likely to suffer from overfitting if the dataset is not large or diverse enough, despite the architecture's strength in capturing features from images due to its greater depth. The ResNet+Bagging model is expected to perform better than other models, such as ResNet, GoogLeNet, VGG-16, and LeNet, due to the combined strength of ResNet in extracting complex features from images and the stability provided by Bagging. By reducing prediction variance and addressing overfitting, this model is more effective in handling diverse and complex medical datasets.

Table 1. Model comparison

Model	Method	Main objective	Potential use
ResNet+Bagging	ResNet+Bagging (ensemble)	Improve accuracy and generalization by combining multiple ResNet models, reducing overfitting, and enhancing stability.	Effective for large and complex datasets such as brain tumor MRI images, offering high stability and accuracy.
ResNet	Residual networks (Deep CNN)	Address the vanishing gradient problem and recognize complex features in images.	Suitable for large datasets and medical images requiring in-depth analysis, but prone to overfitting with small datasets.
GoogLeNet	Inception (multi-scale convolutional filters)	Reduce the number of parameters and capture multi-scale information.	Ideal for applications with computational limitations, such as medical images, requiring efficiency without sacrificing accuracy.
VGG-16	Deep CNN (multiple convolutional layers)	Recognize visual patterns through consecutive convolutional layers.	Suitable for smaller or medium-sized datasets but can suffer from overfitting without proper optimization techniques.
LeNet	Simple CNN (Convolutional and subsampling layers)	A lightweight and simple model suitable for small datasets and applications that don't require high complexity.	Ideal for small datasets and applications with limited resources, but less effective for complex medical images.

2.3. Research framework

The research framework in this study is crucial as it provides a systematic guide for each stage of the research, from data processing to result evaluation. In this study, the framework helps organize the process of comparing the performance of various CNN model architectures, such as LeNet, ResNet, GoogLeNet, VGGNet, and the proposed model (ResNet+Bagging). With a clear framework, the researcher can perform comparisons transparently and accurately, enabling a better understanding of each research phase. Furthermore, this framework minimizes errors and ensures a comprehensive evaluation by using various performance metrics, making the research results more valid and reliable.

The research framework illustrated in Figure 2 outlines the workflow, beginning with brain MRI images dataset, which consists of two classes: hemorrhage and non-hemorrhage. The process starts with data preprocessing, where image data is enhanced through data augmentation techniques such as flipping, rotating, shearing, and rescaling. These techniques aim to increase the variety of the training data, ultimately improving the model's ability to recognize a broader range of patterns. Following preprocessing, the data splitting stage divides the dataset into three subsets: 80% for training, 10% for validation to assess the model's performance during training, and 10% for testing, used to evaluate the model's final performance after training. At this stage, data visualization is also performed to verify the distribution of data across the different classes. In the model training and classification phase, the preprocessed data is used to train various model architectures, including LeNet, ResNet, GoogLeNet, VGGNet, as well as the proposed modified model (ResNet+Bagging).

Figure 2 shows the research design based on all models trained on the same dataset to create a brain tumor classification model. After training, the models undergo a model comparison, where the performance of LeNet, ResNet, GoogLeNet, and VGGNet is compared with the proposed model. This comparison serves to evaluate the impact of architectural modifications and hyperparameter tuning in improving classification accuracy. Finally, in the evaluation and analysis stage, the classification results of each model are assessed using performance metrics such as accuracy, precision, recall, F1-score, loss, specificity, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC-ROC). The results are then presented visually to facilitate interpretation and further analysis.

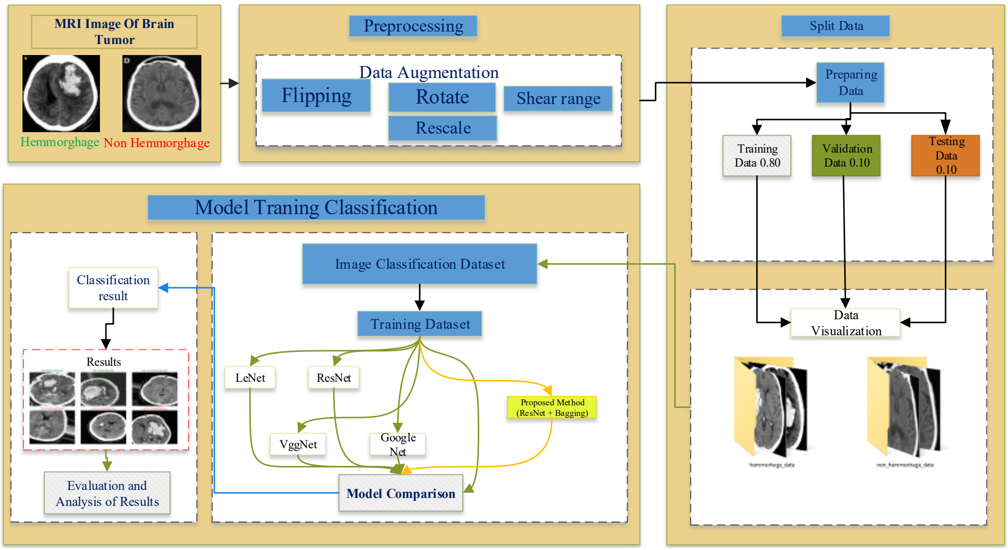


Figure 2. Research framework

3. RESULTS AND DISCUSSION

3.1. Data preprocessing and data augmentation process

In the preprocessing stage, the brain tumor MRI images undergo a series of data enhancement processes. These processes include: flipping–flipping the image horizontally or vertically to create variations in orientation; rotation–rotating the image at specific angles to increase the diversity of the dataset; shear–applying a shear transformation to create slight distortions in the image; and rescaling–normalizing the pixel values to a [0, 1] range to help the model better recognize patterns in the data. The results of the augmentation process can be seen in Figure 3.

Figure 3 shows the results of the data augmentation process applied to the brain tumor MRI scan images. The augmentation techniques, such as flipping, rotating, and rescaling, significantly enrich the dataset with varied samples. These techniques effectively enhance the model's generalization ability. The primary goal of augmentation is to ensure that the model does not become overly dependent on specific patterns, making it more adaptable to diverse test data. In this study, the experimental results were obtained from classifying MRI brain tumor images, which were divided into three categories: hemorrhage and non-hemorrhage. The dataset underwent preprocessing, including data augmentation with techniques like flipping, rotating, shearing, and rescaling. After preprocessing, the dataset was split into training data (80%), validation data (10%), and testing data (10%). This process ensures that the model evaluation is performed on unseen data, providing objective and realistic results.

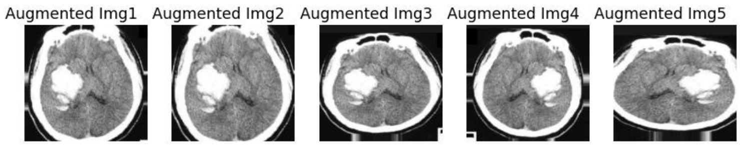
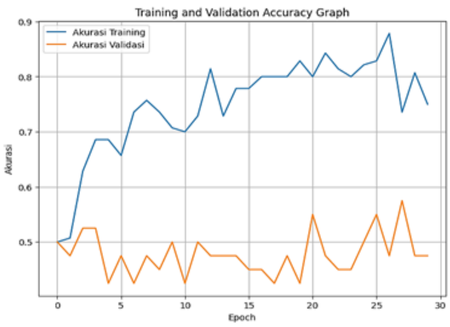


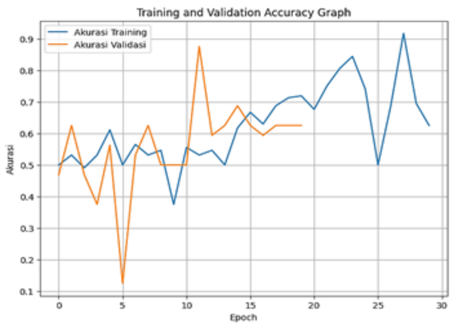
Figure 3. Augmentation of brain tumor MRI image dataset

3.2. Model training and classification

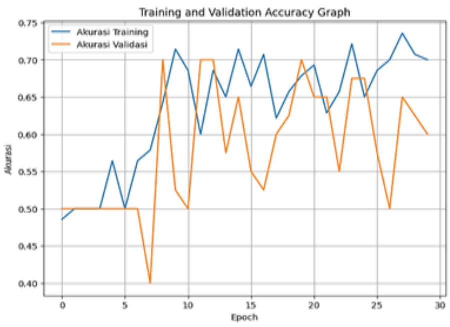
Several CNN models–LeNet, ResNet, GoogLeNet, VGGNet, and the proposed model, were evaluated for brain tumor classification using MRI images. Each model exhibited different levels of accuracy and learning behavior due to their architectural differences. While simpler models like LeNet showed limited performance, deeper models such as ResNet, GoogLeNet, and VGGNet achieved better results owing to their advanced structures. The proposed model, designed specifically for this task, aimed to optimize both accuracy and computational efficiency. Figure 4 shows the training accuracy curves, illustrating the performance trends of each model during the training process starting from the LeNet (Figure 4(a)), ResNet (Figure 4(b)), GoogleNet (Figure 4(c)), VGGNet (Figure 4(d)), and proposed model (ResNet+Bagging) models (Figure 4(e)).



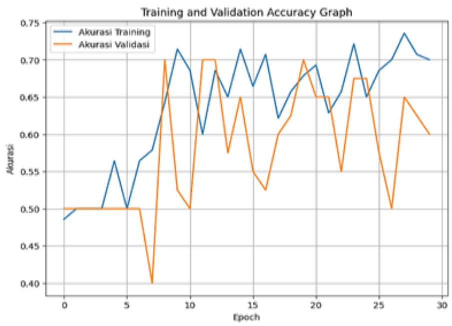
(a)



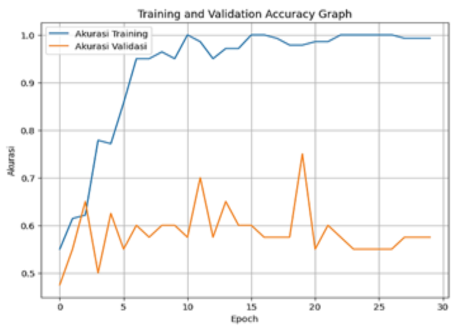
(b)



(c)



(d)



(e)

Figure 4. Training accuracy curve of (a) LeNet, (b) ResNet, (c) GoogleNet, (d) VGGNet, and (e) proposed model (ResNet+Bagging)

Figure 4 presents the training results, where the proposed ResNet+Bagging model demonstrates the best performance among all the tested models. The training and validation accuracy curves for this model show significant stability, with the validation accuracy nearly matching the training accuracy. This indicates that the ensemble learning method with Bagging effectively improved ResNet's generalization ability, minimizing the risk of overfitting. With this approach, ResNet leveraged its strength in capturing complex features from brain tumor MRI images while reducing sensitivity to noise in the data. The high stability and accuracy achieved confirm the superiority of the proposed model for medical image classification, where high precision is crucial. In comparison, other models like LeNet, VGGNet, and GoogLeNet also showed good results, though with certain limitations. LeNet, due to its simpler architecture, achieved good training accuracy but struggled to reach optimal validation accuracy. Meanwhile, GoogLeNet and VGGNet exhibited more stable performance compared to LeNet, but their validation accuracy fluctuations, though smaller than LeNet's, were still more pronounced than those of ResNet+Bagging, indicating some limitations in handling data variations. Therefore, by combining the strengths of ResNet with the stability provided by Bagging, the proposed model emerges as the best solution for brain tumor MRI classification. Losses in each training can be seen in Figure 5, where Figures 5(a) to 5(d) show the loss curves of LeNet, ResNet, GoogLeNet, and VGGNet, respectively, and Figure 5(e) presents the proposed ResNet+Bagging model.

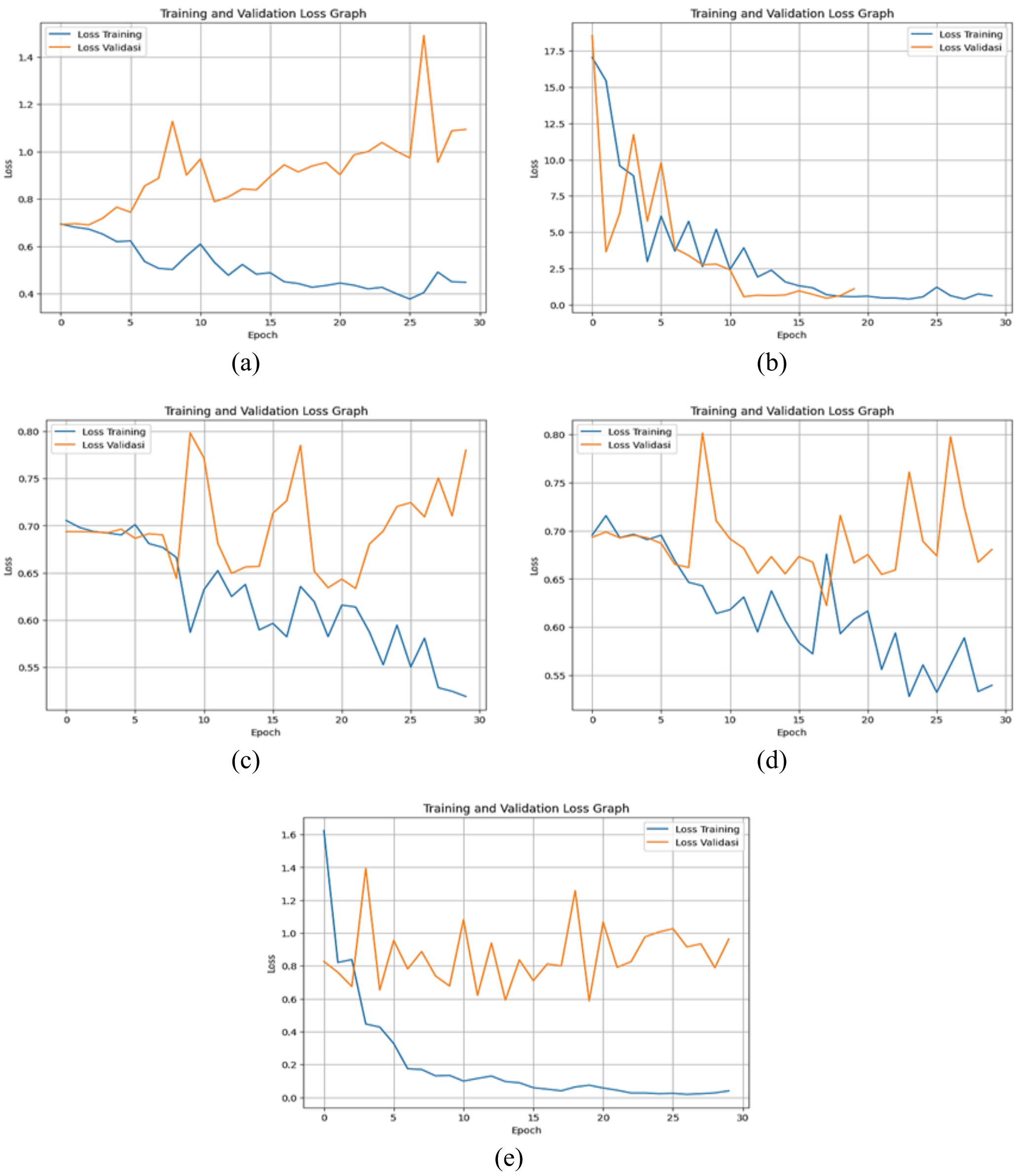
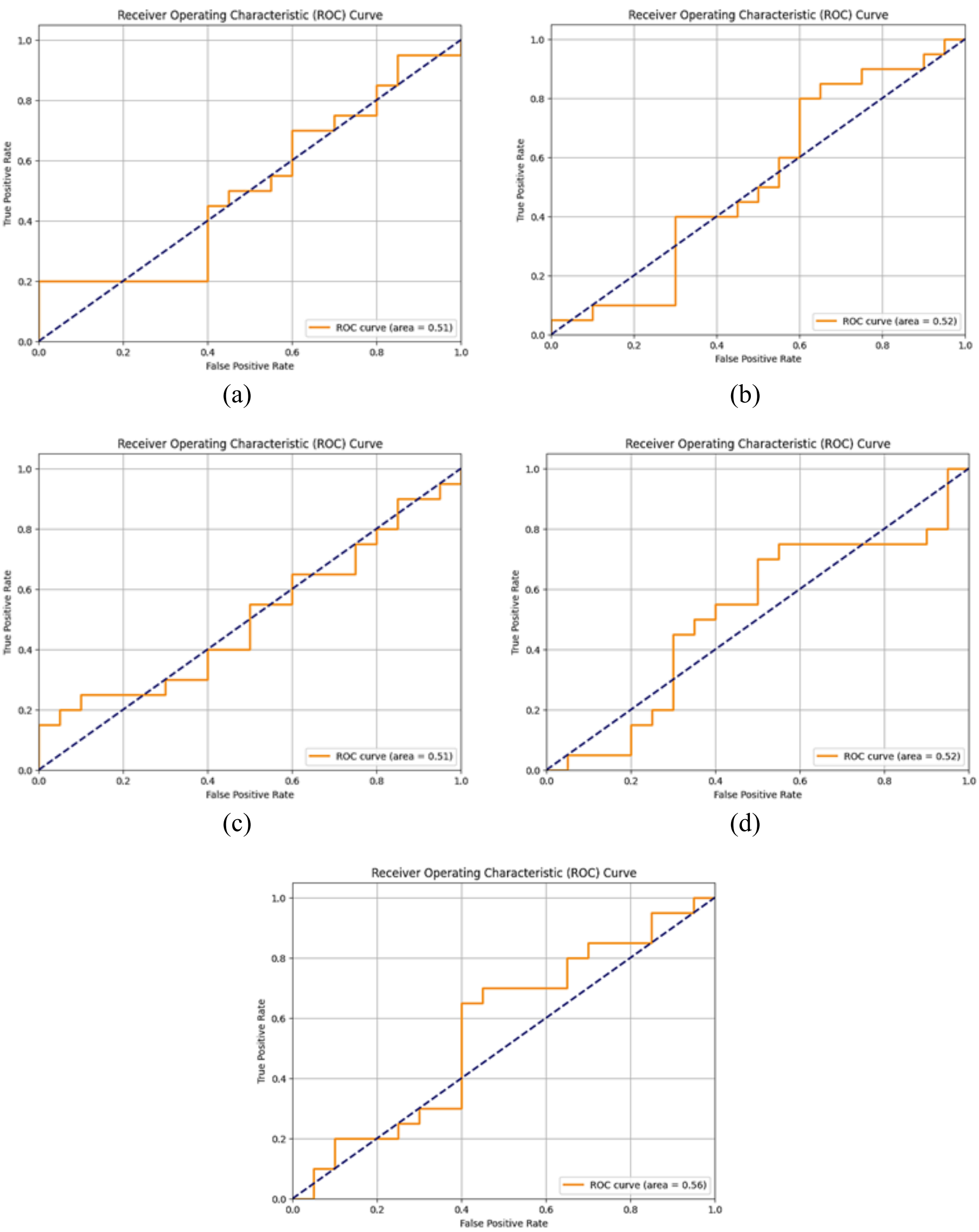


Figure 5. Training loss curve of (a) LeNet, (b) ResNet, (c) GoogleNet, (d) VGGNet, and (e) proposed model (ResNet+Bagging)

Figure 5 illustrates the training and validation loss curves, highlighting the performance differences between the LeNet, ResNet, GoogLeNet, VGGNet, and the proposed model. The loss graphs for all five models show varying patterns in reducing loss during training and validation. LeNet exhibits fluctuating and unstable validation loss, indicating the difficulty the model has in learning complex patterns from the data. ResNet demonstrates a consistent decline in loss for both training and validation, although there is a slight fluctuation in the validation loss, suggesting the model may need additional regularization. GoogLeNet performs well, with a stable reduction in loss, though its validation loss is slightly higher than the training loss. VGGNet shows a significant drop in training loss, but its validation loss remains unstable, hinting at mild overfitting. In contrast, the proposed ResNet+Bagging model exhibits a very stable loss trend, with validation loss lower than that of the other models. This indicates that the Bagging method successfully enhanced the model's generalization and stability. As a result, the proposed model proves to be the most effective in handling complex data, such as brain tumor MRI images. Figure 6 shows the ROC curve, where Figures 6(a) to 6(d) show the ROC curves of LeNet, ResNet, GoogLeNet, and VGGNet, respectively, and Figure 6(e) presents the proposed ResNet+Bagging model.



(e)

Figure 6. ROC curve of (a) LeNet, (b) ResNet, (c) GoogleNet, (d) VGGNet, and (e) proposed model (ResNet+Bagging)

Figure 6 shows the ROC curve for all five models, reflecting their overall performance in distinguishing between brain tumor classes. LeNet exhibits a ROC curve that closely aligns with the diagonal line, indicating its limited classification ability, with a low area AUC. ResNet presents a higher ROC curve compared to LeNet, demonstrating its better ability to distinguish between classes. GoogLeNet also shows a decent ROC curve, with an AUC similar to ResNet, but it is slightly less optimal in some classification areas. VGGNet's ROC performance is comparable to GoogLeNet's, though it is somewhat less stable. In contrast, the proposed ResNet+Bagging model shows the highest and most stable ROC curve among all the models, indicating superior classification ability with an AUC nearing the maximum value. This highlights that the use of Bagging with ResNet significantly enhances classification performance. The image displayed provides an example of classification results from the proposed CNN model for detecting brain tumor MRI scans. A detailed explanation of the classification results shown in Figure 7.

Figure 7 displays the classification results of the proposed CNN model, showcasing example predictions on a brain tumor MRI dataset. Nine images are arranged in a 3×3 matrix, with each image labeled to show both the true class (true) and the predicted class (pred) for the categories of malignant (cancer) and normal (no cancer). Several predictions are correct, such as the Malignant cases correctly predicted as Malignant, and the Normal cases accurately predicted as normal. However, there are also some misclassifications, such as a normal case incorrectly predicted as Malignant. The pixelated appearance of the images reflects the preprocessing steps applied to enhance the model's ability to detect key visual patterns. While the model performs well in identifying most patterns, the misclassifications highlight the need for further refinement, such as adjustments to parameter selection or architectural structure, to improve the accuracy and reliability of brain tumor MRI classification.

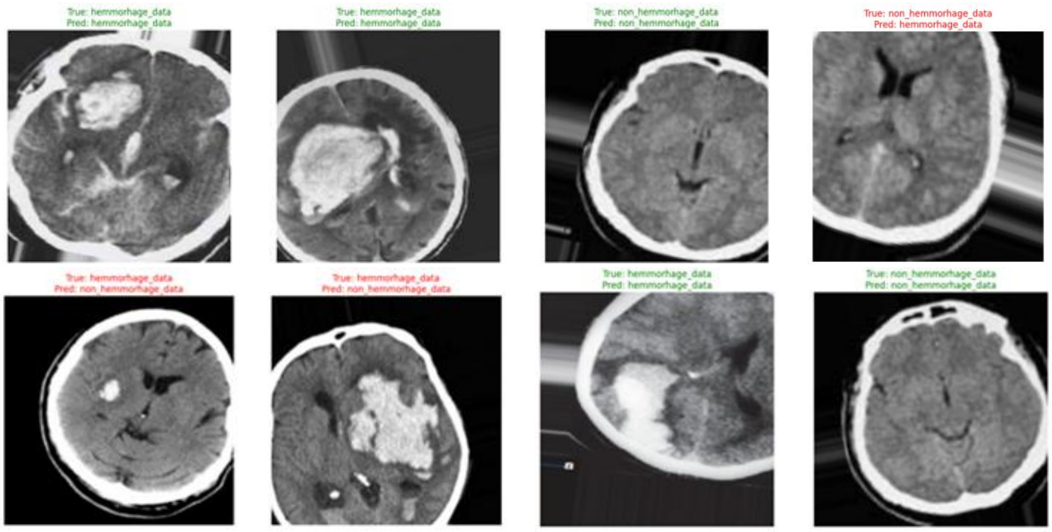


Figure 7. Classification results of the proposed model

3.3. Discussion

The training and testing results include confusion matrices for each CNN model, highlighting their classification performance across three categories: benign, malignant, and normal brain tumors. These matrices provide a clear view of each model’s strengths and weaknesses in distinguishing between the classes. Figure 8 illustrates the confusion matrices, allowing for direct comparison of classification accuracy and error distribution among the evaluated models.

Figure 8 illustrates the confusion matrices for the five models, showing varying performance in classifying hemorrhage and non-hemorrhage data. LeNet as shown in Figure 8(a) demonstrates moderate performance, with several prediction errors in both classes. ResNet as shown in Figure 8(b) performs slightly better than LeNet, correctly predicting more instances of the hemorrhage class. GoogleNet as shown in Figure 8(c) shows some imbalance in its predictions, performing less optimally for the non-hemorrhage class. VGGNet as shown in Figure 8(d) shows an improvement in accuracy over GoogleNet, with a more balanced prediction distribution across both classes. The proposed model (ResNet+Bagging) as shown in Figure 8(e) delivers the best performance, with the fewest prediction errors among all models, reflecting the effectiveness of ensemble learning in enhancing classification accuracy. The results are presented in Table 2.

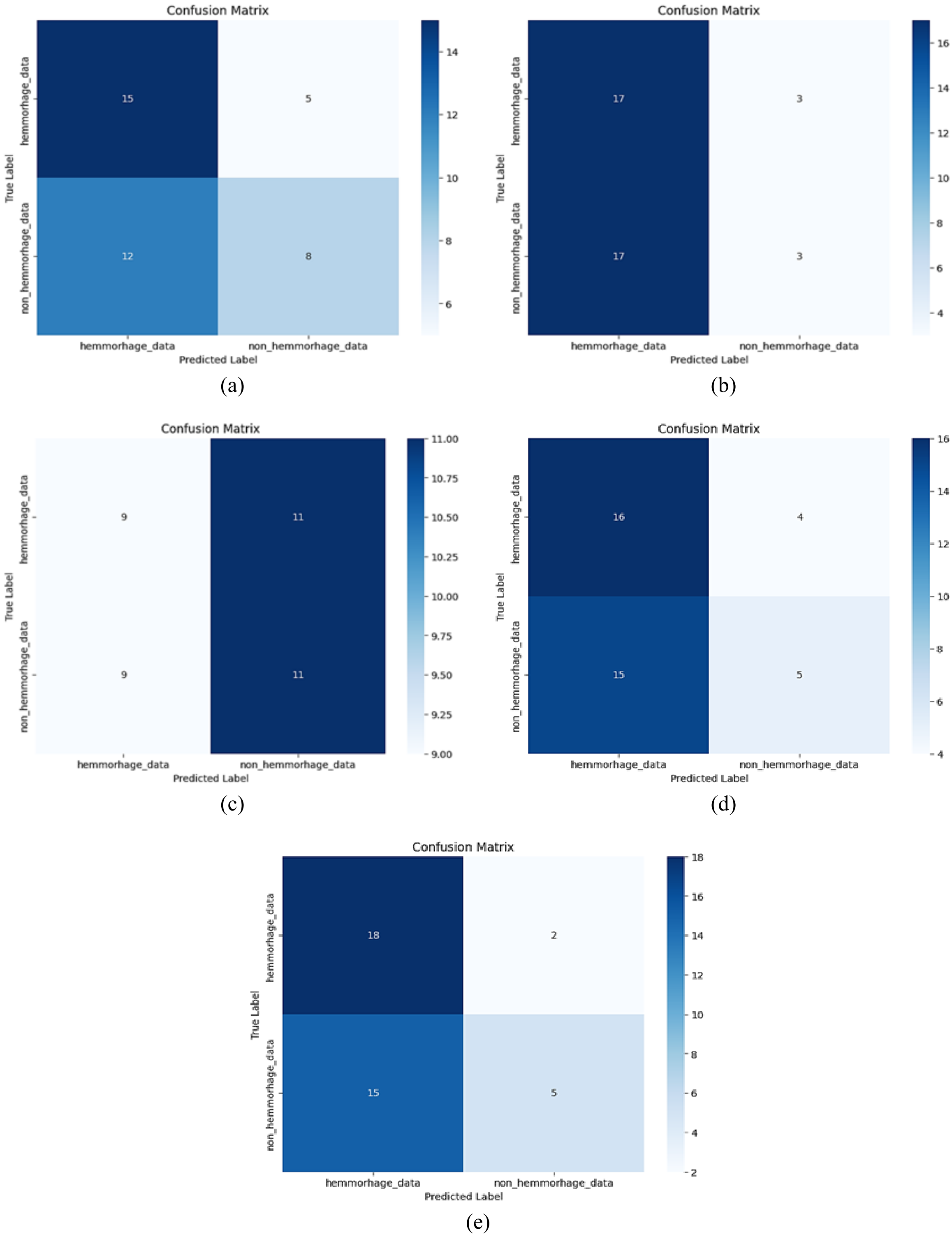


Figure 8. Confusion matrix of (a) LeNet, (b) ResNet, (c) GoogleNet, (d) VGGNet, and (e) proposed model (ResNet+Bagging)

Table 2 shows that the proposed model (ResNet+Bagging) outperforms all other models tested. With a testing accuracy of 72%, this model demonstrates significant generalization capabilities, far surpassing other models such as LeNet (45%), standard ResNet (57%), GoogleNet (45%), and VGGNet (50%). Furthermore, the proposed model also records the highest precision (0.6010), recall (0.6000), and F1-score (0.5990) among the other models, indicating a good balance between detecting positive and negative classes. Its specificity (0.5500) and MCC (0.2010) also show an advantage in handling imbalanced data. Although the AUC-ROC of the proposed model (0.5625) is not significantly higher than the other models, it still supports the overall better performance. These results indicate that the application of Bagging to ResNet significantly enhances the stability and generalization ability of the model, making it the most effective model for brain tumor MRI classification in this study. In Figure 9, you can see that the training and testing result curves are displayed.

The radar charts (Figure 9) reveal the performance of five models—LeNet, ResNet, GoogleNet, VGGNet, and the proposed model (ResNet+Bagging)—across multiple metrics, including training accuracy, precision, recall, F1-score, testing accuracy, MCC, and AUC-ROC. Among these, the proposed model demonstrates the best overall performance, achieving high precision (~60%), recall (~60%), F1-score (~60%), and testing accuracy (~72%), alongside the highest MCC (~20%) and AUC-ROC (~56%). This indicates its strong ability to generalize and handle imbalanced datasets, thanks to the stability provided by the Bagging technique. In contrast, simpler models like LeNet and GoogleNet show limitations in generalization, with low metrics such as precision, recall, and MCC, making them less suitable for complex tasks like brain tumor MRI classification. ResNet and VGGNet perform moderately well, with improved precision, recall, and F1-score compared to LeNet and GoogleNet, but still fall short in testing accuracy and MCC compared to the proposed model. Notably, while all models exhibit reasonable training accuracy, specificity remains a common challenge, reflecting room for improvement in reducing false positives. The results highlight the effectiveness of the proposed model for applications demanding high precision and robustness, particularly in medical imaging tasks, while simpler models may be more suitable for resource-constrained environments despite their limited performance. Further optimization, especially in specificity, could enhance the applicability of these models for diverse datasets.

Table 2. Testing results of the compared models

Model	Training results	Training accuracy	Precision	Recall	F1-score	Specificity	MCC	AUC-ROC	Testing accuracy
LeNet	0.8212	0.8212	0.5	0.5	0.4505	0.8	0.0	0.51	0.45
ResNet	0.6663	0.6663	0.5285	0.525	0.51	0.7	0.0534	0.52	0.57
GoogleNet	0.7614	0.7614	0.5	0.5	0.455	0.8	0.01	0.515	0.45
VGGNet	0.8609	0.8609	0.5767	0.575	0.5726	0.65	0.1517	0.515	0.5
Proposed model (ResNet+Bagging)	0.8443	0.8443	0.601	0.6	0.599	0.55	0.201	0.5625	0.72

Figure 9. Training and testing results curve

4. CONCLUSION

Based on the research findings, it can be concluded that the proposed ResNet+Bagging model outperforms LeNet, standard ResNet, GoogleNet, and VGGNet in the classification of brain tumor MRI images. The proposed model achieved the highest testing accuracy of 72%, indicating significant generalization capability on validation data. Moreover, the model recorded higher precision (0.6010), recall (0.6000), and F1-Score (0.5990) compared to the other models, reflecting a well-balanced ability to detect both positive and negative classes. The application of the ensemble learning method with Bagging on ResNet significantly enhanced the overall performance of the model, particularly in mitigating the risk of overfitting and improving the stability of predictions. This improvement highlights the effectiveness of combining ResNet's architectural strengths with the stabilizing effects of Bagging, which collectively address the challenges associated with complex datasets. Such a methodology not only boosts predictive reliability but also demonstrates adaptability to diverse patterns within medical image data. Consequently, this approach can be recommended as an efficient solution for complex medical image classification tasks, such as brain tumor identification, supporting more accurate and reliable diagnostic processes. The results affirm that ResNet+Bagging is a promising strategy to enhance model robustness and precision, making it a valuable tool in advancing medical image analysis. And this research can be continued by validating on larger multi-center datasets, exploring scaling/stacking strategies, and adopting modern architectures such as vision transformers.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Akim Manaor Hara Pardede	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Ahmad Zamsuri	✓	✓		✓	✓	✓			✓	✓	✓	✓		✓
Indi Nuroini		✓					✓		✓	✓	✓	✓	✓	
Putrama Alkhairi	✓		✓	✓	✓	✓		✓	✓	✓				

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/code/guslovesmath/cnn-brain-tumor-classification-99-accuracy/.

REFERENCES

[1] G. G. Nalina and C. Raju, "Bayesian weighted convolutional neural network for brain tumor classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 4, pp. 674–681, 2023.

[2] A. Sarkar, M. Maniruzzaman, M. A. Alahe, and M. Ahmad, "An effective and novel approach for brain tumor classification using AlexNet CNN feature extractor and multiple eminent machine learning classifiers in MRIs," *Journal of Sensors*, vol. 2023, 2023, doi: 10.1155/2023/1224619.

[3] P. Thamilselvan, "Lung cancer prediction and classification using Adaboost data mining algorithm," *International Journal of Computer Theory and Engineering*, vol. 14, no. 4, pp. 149–154, 2022, doi: 10.7763/IJCTE.2022.V14.1322.

[4] D. Li, G. Li, S. Li, and A. Bang, "Classification prediction of lung cancer based on machine learning method," *International Journal of Healthcare Information Systems and Informatics*, vol. 19, no. 1, 2023, doi: 10.4018/IJHISI.333631.

[5] N. Remzan, K. Tahiry, and A. Farchi, "Automatic Classification of Preprocessed MRI brain tumors images using deep convolutional neural network," *International Journal on Technical and Physical Problems of Engineering*, vol. 15, no. 1, pp. 68–73, 2023.

[6] W. Baccouch, S. Oueslati, B. Solaiman, and S. Labidi, "A comparative study of CNN and U-Net performance for automatic segmentation of medical images: application to cardiac MRI," *Procedia Computer Science*, vol. 219, pp. 1089–1096, 2023, doi: 10.1016/j.procs.2023.01.388.

[7] G. V. Sivanarayana and K. N. Kumar, "MRI based brain tumor classification using modified convolutional neural network model," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 8s, pp. 450–457, 2024.

[8] J. A. Jeba, S. N. Devi, and M. Meena, "Modified CNN architecture for efficient classification of glioma brain tumour," *IETE Journal of Research*, vol. 69, no. 12, pp. 9310–9323, 2023, doi: 10.1080/03772063.2022.2101553.

[9] B. Sterniczuk and M. Charytanowicz, "An ensemble transfer learning model for brain tumors classification using convolutional neural networks," *Advances in Science and Technology Research Journal*, vol. 18, no. 8, pp. 204–216, 2024, doi: 10.12913/22998624/193627.

[10] V. G. Krishnan, P. V. Rao, M. V. V. Saradhi, K. Sathyamoorthy, and V. Vijayaraja, "Prediction of Alzheimer disease using LeNet-CNN model with optimal adaptive bilateral filtering," *International Journal of Communication Networks and Information Security*, vol. 15, no. 1, pp. 12–23, 2023, doi: 10.17762/ijcnis.v15i1.5706.

[11] E. P. Hemand, G. Mohandass, F. H. Shajin, and D. Kirubakaran, "AlexNet-based deep convolutional neural network optimized with group teaching optimization algorithm (GTOA) for paediatric bone age assessment from hand X-ray images," *The Imaging Science Journal*, vol. 72, no. 3, pp. 336–348, 2024, doi: 10.1080/13682199.2023.2204669.

[12] T. Tang, T. Chen, X. Chen, and G. Gui, "Integration of VGGNet-driven precipitation similarity analysis and hydrological modeling for enhanced streamflow forecasting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 14543–14555, 2024, doi: 10.1109/JSTARS.2024.3441550.

[13] S. Amutha, "VGGNet-CNN based classification of white blood cell leukemia with efficient salp swarm optimization algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 44, no. 4, pp. 6973–6989, 2023, doi: 10.3233/JIFS-221302.

[14] A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet," *BMC Medical Imaging*, vol. 21, no. 1, 2021, doi: 10.1186/s12880-020-00529-5.

[15] B. Akhila, K. Parichitha, M. Thrisha, B. Deepthi, and G. Manjula, "Deep learning-based detection of rice leaf diseases for smart agriculture," *Journal of Science Engineering Technology and Management Sciences*, vol. 2, no. 4, pp. 111–122, 2025, doi: 10.63590/jsetms.2025.v02.i04.pp111-122.

[16] Y. Zhang, Y. Wang, L. Xu, Y. Yao, W. Qian, and L. Qi, "ST-GAN: a swin transformer-based generative adversarial network for unsupervised domain adaptation of cross-modality cardiac segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 893–904, 2024, doi: 10.1109/JBHI.2023.3336965.

[17] Z. Xiao and H. Ou, "Research on data generation based on the combination of growing-pruning GAN and intelligent parameter optimization," *International Journal of Cooperative Information Systems*, vol. 33, no. 4, pp. 149–161, 2024, doi: 10.1142/S0218843023500235.

[18] A. S. Chauhan, J. Singh, S. Kumar, N. Saxena, M. Gupta, and P. Verma, "Design and assessment of improved convolutional neural network based brain tumor segmentation and classification system," *Journal of Integrated Science and Technology*, vol. 12, no. 4, Feb. 2024, doi: 10.62110/sciencein.jist.2024.v12.793.

[19] S. S. Dip, M. H. Rahman, N. Islam, M. E. Arafat, P. K. Bhowmick, and M. A. Yousuf, "Enhancing brain tumor classification in MRI: leveraging deep convolutional neural networks for improved accuracy," *International Journal of Information Technology and Computer Science*, vol. 16, no. 3, pp. 12–21, 2024, doi: 10.5815/ijitcs.2024.03.02.

[20] S. Mouloodi, H. Rahmanpanah, S. Gohery, C. Burvill, and H. M. S. Davies, "Feedforward backpropagation artificial neural networks for predicting mechanical responses in complex nonlinear structures: a study on a long bone," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 128, 2022, doi: 10.1016/j.jmbbm.2022.105079.

[21] Z. Mohammed, *Artificial intelligence definition, ethics and standards*. The British University in Egypt, 2019.

[22] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00299-5.

[23] S. Y. Prasetyo, G. Z. Nabiilah, Z. N. Izdihar, and S. M. Isa, "Pneumonia detection on x-ray imaging using softmax output in multilevel meta ensemble algorithm of deep convolutional neural network transfer learning models," *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 2, pp. 319–330, 2023, doi: 10.26555/ijain.v9i2.884.

[24] Y. J. Luwe, C. P. Lee, and K. M. Lim, "Wearable sensor-based human activity recognition with ensemble learning: a comparison study," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4029–4040, 2023, doi: 10.11591/ijece.v13i4.pp4029-4040.

[25] S. Luo, "Lung cancer classification using reinforcement learning-based ensemble learning," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 1112–1122, 2023, doi: 10.14569/IJACSA.2023.01408120.

[26] A. Sekhar, S. Biswas, R. Hazra, A. K. Sunaniya, A. Mukherjee, and L. Yang, "Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 983–991, 2022, doi: 10.1109/JBHI.2021.3100758.

[27] R. Zaitoon and H. Syed, "Enhanced brain tumor detection and classification in MRI scans using convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 266–275, 2023, doi: 10.14569/IJACSA.2023.0140929.

[28] M. P. Potadar and R. S. Holambe, "automatic multi-class brain tumor classification using residual network-152 based deep convolutional neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 37, no. 4, 2023, doi: 10.1142/S0218001423560013.

BIOGRAPHIES OF AUTHORS

Akim Manaor Hara Pardede completed his Bachelor's degree in Informatics Engineering from the Indonesian College of Science and Technology in Bandung (ST-INTEN) in 2003, completed his Master's degree in Informatics Engineering from the University of North Sumatra in 2013, and completed his Doctoral degree in Computer Science at the University of North Sumatra in 2019. He currently serves as the Head of the Research and Community Service Institute at STMIK Kaputama. In addition to teaching, he is also active in conducting research/community service and winning research/community service awards funded by the Ministry of Research and Technology/National Research and Innovation Agency, and is active in writing scientific articles in the form of seminars or international journals indexed by Scopus. In addition, the author is also active as a reviewer at international seminars and reviewer in national and international journals including as a journal reviewer at Elsevier. He can be contacted at email: akimmhp@live.com.

Ahmad Zamsuri completed his Master's degree in Computer Science and Doctoral (Ph.D.) degree at Universitas Putra Indonesia 'YPTK' Padang in 2013 and 2024. Currently, he is an active lecturer at Universitas Lancang Kuning Pekanbaru since 2014, teaching in the Master's Program in Computer Science. His research focuses on artificial intelligence, including decision support systems, expert systems, data mining, neural networks, fuzzy logic, machine learning, and deep learning. Additionally, he serves as a reviewer for various nationally accredited journals (SINTA 3 to SINTA 6) and manages the "Rumah Jurnal" community at the Faculty of Computer Science, Universitas Lancang Kuning. Ahmad Zamsuri has received several research grants at both local and DIKTI levels. He can be contacted at email: ahmadzamsuri@unilak.ac.id.

Indi Nuroini an academic in the field of law who has taught industrial relations law courses at the Faculty of Law, Bhayangkara University, Surabaya since 2013. In order to support professionalism as an academic in the field of law, the author is also a legal consultant in several companies. Apart from that, he also actively preaches as a scholar who is respected by his congregation. He can be contacted at email: nuroini.indi@gmail.com.

Putrama Alkhairi completed a master's degree in Computer Science in 2023 at the Main Potential University (UPU) Medan. The author has been an active lecturer at STIKOM Tunas Bangsa since 2021, teaching in the Information Systems program. Their research focuses on artificial intelligence (decision support systems, data mining, artificial neural networks, deep learning, and genetic algorithms). Apart from that, the author has also been active as a member of a community called "Journal Hunters" at STIKOM Tunas Bangsa as a student who took part in joint research with lecturers and has won several awards as the best article writer at the national seminar held by SENARIS 2018. He can be contacted at email: putrama@amiktunasbangsa.ac.id.