International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 12; Issue: 12; December 2025
Website: www.ijrrjournal.com
Research Paper E-ISSN: 2349-9788; P-ISSN: 2454-2237

Optimization of Convolutional Neural Network-Based Classification Using EfficientNet-B1

Muhammad Fachri Mahyudin[1], Maman Somantri[1], Oky Dwi Nurhayati[2]

[1]Department of Electrical Engineering, [2]Department of Computer Engineering,
Universitas Diponegoro, Semarang, Indonesia.

Corresponding Author: Muhammad Fachri Mahyudin

DOI: https://doi.org/10.52403/ijrr.20251270

ABSTRACT

Brain cancer is a life-threatening disease with a global mortality count reaching 241,037 cases, with Asia recording the highest number of deaths. Advances in artificial intelligence (AI) and machine learning offer significant opportunities to improve the accuracy and consistency of diagnosis through MRI image analysis. Convolutional Neural Networks (CNNs) have been widely used in cancer detection due to their ability to automatically extract features and perform high-accuracy image classification. This thesis employs the EfficientNet B1 model because its compound scaling architecture optimally balances network depth, width, and resolution. This design enables the model to achieve high computational efficiency, operate smoothly on various hardware systems, and still maintain strong accuracy performance. These characteristics make EfficientNet B1 particularly suitable for identifying complex patterns in brain MRI images. This research focuses on optimizing and evaluating EfficientNet B1 for brain cancer detection tasks, emphasizing both accuracy and computational efficiency. The experimental results show that the model achieved an accuracy of 0.9734, confirming its effectiveness in brain cancer classification. These findings highlight the potential of EfficientNet B1 as a fast, accurate, and practical model for AI-based diagnostic support systems.

Keywords: *Brain Cancer, MRI, CNN, EFFICIENTNET-B1*

INTRODUCTION

Brain cancer is a life-threatening disease caused by the uncontrolled growth of cells within brain tissue. According to the 2018 Cancer Registry, there were 296,851 brain cancer cases worldwide, with the highest proportion reported in Asia (52.6%) and the lowest in Oceania (0.82%). The number of deaths reached 241,037 cases, following a similar distribution pattern [1]. Early detection is essential for improving patient recovery outcomes. MRI is the primary method for identifying brain tumors; however, manual analysis by medical professionals often requires considerable time and is susceptible to subjective errors. Therefore, the use of artificial intelligence (AI) and machine learning can help enhance consistency and accuracy in the diagnosis of brain cancer [2].

In this study, the EfficientNet-B1 model was selected due to its strong performance in medical image classification. EfficientNet was chosen because it offers an optimal combination of computational efficiency and high accuracy. By applying the Compound Scaling approach, EfficientNet proportionally adjusts its depth, width, and resolution, resulting in a lighter model with significantly fewer parameters without compromising accuracy. Another advantage of EfficientNet is the use of the Swish activation function and Drop Connect regularization, which enhance the model's generalization ability and make it more resistant to overfitting, particularly when applied to relatively small datasets such as brain MRI images. In addition, EfficientNet has been pretrained on the ImageNet dataset, making it highly suitable for transfer learning [3]. By leveraging pretrained weights, the model can quickly adapt to brain cancer classification tasks using MRI images, improving training efficiency while maintaining high accuracy.

Among the EfficientNet variants, B1 offers a practical balance between model size and accuracy, whereas higher variants such as EfficientNet-B7 require significantly greater computational resources. Previous studies have also explored the use of CNN architectures for brain tumor identification. For example, the study "Brain Tumor Diagnosis and Classification via Pre-Trained Convolutional Neural Networks" [4] reported that EfficientNet-B1 achieved a validation accuracy of 89.55% in detecting brain tumors.

In this study, the term optimization refers to performance-enhancing strategies such as data preprocessing, transfer learning, and hyperparameter tuning. The objective of this research is to optimize a CNN-based classification model using EfficientNet-B1 for brain cancer detection through MRI images and to compare its performance across different configurations. The goal is to improve the model's accuracy while reducing computational overhead, thereby providing a more effective solution to support medical diagnosis.

LITERATURE REVIEW

EfficientNet

EfficientNet is a Convolutional Neural Network (CNN) architecture designed to address the challenge of scaling neural networks efficiently without excessively increasing computational costs. Its core approach is compound scaling, a method that proportionally and uniformly scales the network's depth, width, and input resolution [5]. This strategy produces a balanced and efficient network, making it more optimal compared to conventional scaling methods. The design enables high computational efficiency, minimizing resource usage while maintaining excellent accuracy [6].

The key difference between EfficientNet and traditional CNNs lies in its use of the Mobile Inverted Bottleneck (MBConv) as the primary convolutional block. This block is engineered to be more computationally efficient and to require fewer parameters than standard convolutional layers. EfficientNet incorporates a combination of depthwise convolution and pointwise convolution, which significantly reduces the number of computational operations as well as the overall parameter count [7]. This technique allows the network to enhance computational efficiency while maintaining high performance with fewer parameters.

Furthermore, the EfficientNet architecture integrates Squeeze-and-Excitation (SE) blocks, which enable the network to reweight feature channels based on their importance. This mechanism allows the model to focus more effectively on the most relevant features, further improving both efficiency and accuracy [8]. Figure 1 illustrates the architecture of EfficientNet-B0.

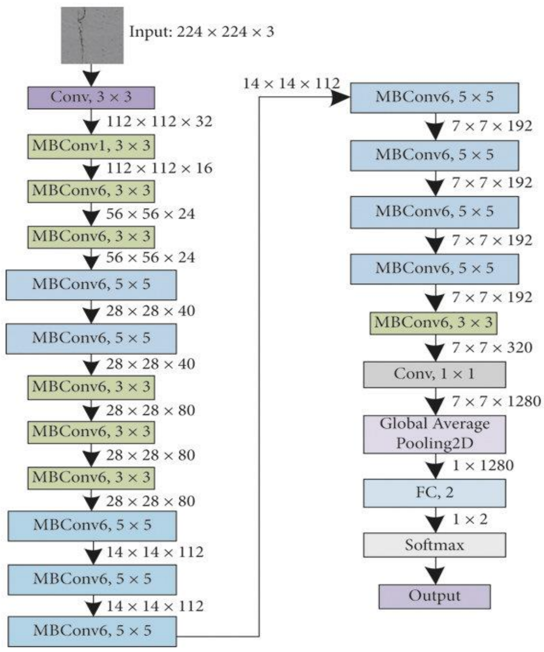

Figure 1. Arsitektur EfficientNet-B0.

EfficientNet consists of several variants, ranging from EfficientNet-B0 to EfficientNet-B7. The differences among these variants lie in model size, the number of parameters, and the corresponding performance improvements. Nevertheless, the core architecture remains the same, as illustrated in Figure 1. The primary distinction is the variation in the number of feature maps (channels), which leads to an overall increase in the number of parameters. This results in improved accuracy but also introduces higher computational complexity. Compound scaling serves as the foundation of the EfficientNet architecture, where three key dimensions are scaled simultaneously: depth, width, and image resolution[9]. The formula for compound scaling is presented in Equation (1).

$$d = \alpha^p, \omega = \beta^p, r = \beta\gamma^p \qquad (1)$$

d represents the depth scaling factor used to determine the number of MBConv blocks.
ω denotes the width scaling factor, which determines the number of filters in each convolutional layer.
r is the resolution scaling factor applied to the initial input layer.
Meanwhile, α, β, and γ are scaling coefficients used in the compound scaling process.

MATERIALS & METHODS

Dataset Description

The dataset used in this study is a brain MRI dataset obtained from the Kaggle platform, titled "Brain Tumor Classification (MRI)", which is open source. This dataset is a combination of three primary sources: Figshare, SARTAJ, and Br35H. It consists of four classes, namely glioma, meningioma, pituitary tumor, and normal brain MRI images. A total of 7,024 images were used in this study. To support optimal model training, the dataset was divided into three subsets: training, validation, and testing, with a consistent class-wise ratio of approximately 70%: 15%: 15%. Table 1 presents the data distribution for each class.

Table 1. Distribution by Class

Class	Number of Images
Glioma	1726
Meningioma	1818
Pituitary Tumor	1736
Normal	1744
Total	7024

Architectural Model

This study employs a Convolutional Neural Network (CNN) architecture that has been widely recognized for its high effectiveness in medical image classification. The model used in this research is EfficientNet-B1. EfficientNet was selected because it offers an optimal combination of computational efficiency and high accuracy. By implementing the Compound Scaling approach, EfficientNet proportionally adjusts the network's depth, width, and resolution, resulting in a lighter model with significantly fewer parameters without compromising accuracy.
Another advantage of EfficientNet is the use of the Swish activation function and Drop Connect, both of which enhance the model's generalization capability and make it more robust against overfitting, particularly when working with relatively small datasets such as brain MRI images. The model applied in this study is the baseline version of EfficientNet-B1, which serves to provide an initial overview of the network's fundamental ability to recognize visual patterns in medical images. This baseline architecture uses the default configuration without applying pretrained ImageNet weights or additional optimization techniques such as fine-tuning, preprocessing, or hyperparameter adjustments. The use of this baseline model serves as a reference point, enabling a more systematic evaluation of performance improvements produced in subsequent optimization stages. The baseline architecture of EfficientNet-B1 is presented in Table 2.

Table 2. Efficientnet-B1

Model: "sequential"		
Layer (type)	**Output Shape**	**Param #**
efficientnetb1 (Functional)	(None, 9, 9, 1280)	6,575,239
Batch normalization (Batch Normalization)	(None, 1280)	5,120
global_average_pooling2d (Batch Normalization)	(None, 1280)	0
dense_1 (Dense)	(None, 4)	1,028
Total params: 6,585,483 (25.12 MB)		
Trainable params: 7,684 (30.02 KB)		
Non-trainable params: 6,577,799 (25.09 MB)		

Model Optimization Strategy

This optimization process is carried out to enhance the model's ability to recognize complex visual features, address the limitations of the dataset size, and minimize the risk of overfitting during training. The main approaches applied in the model optimization process include data preprocessing, transfer learning, and hyperparameter configuration. These approaches are selected to support the model in extracting features more effectively, improving the learning process, and reducing the risk of overfitting during training. The details are as follows:

1. Data preprocessing is used to adapt MRI images to the EfficientNet-B1 architecture and stabilize pixel values using the preprocessing input function. At this stage, the model is trained from scratch without pretrained ImageNet weights.
2. Transfer learning utilizes pretrained ImageNet weights to accelerate training. After the initial feature extraction, several layers are unfrozen for fine-tuning so the model can better adapt to the characteristics of MRI images[10].
3. Hyperparameters control the learning process and influence the model's accuracy, computational efficiency, and generalization capability[11].

Experimental Scenario

To achieve optimal performance in detecting brain cancer from MRI images, several experimental scenarios were implemented, focusing on improving accuracy while maintaining computational efficiency. These scenarios were developed through three main approaches: preprocessing optimization, fine-tuning of pretrained layers, and hyperparameter configuration. Each scenario was designed to analyze the contribution of each method and identify the most effective combination for brain cancer detection. The experimental scenarios are presented in Table 3.

Table 3. Eksperimental Scenarios

Group	**Number of Scenarios**	**purpose**
Effiicentnet Baseline	1	Serves as a benchmark for assessment.
Augmentasi	3	Assessing the influence of data preprocessing intensity
Hyperparameter	3	Assessing optimization of image size, batch size, LR, optimizer, and layer
Transfer Learning	3	Assessing the contribution of the number of trained layers

The preprocessing evaluation used three configurations with varied augmentation and image processing settings to examine their impact on brain tumor classification performance. Detailed parameters for each scenario are shown in Table 4.

Table 4. Preprocessing scenarios.

Parameter	Scenario 1	Scenario 2	Scenario 3
Preprocessing function	Preprocess input	Preprocess input	Preprocess input
Fill mode	-	nearest	nearest
Rotation range	10	20	60
Width shift range	-	0,2	0,6
Height shift range	-	0,2	0,6
Shear range	-	0,2	0,6
Zoom range	0.1	0,2	0,6
Horizontal flip	*True*	*True*	*True*
Vertical flip	-	-	*True*

Transfer learning was evaluated using three fine-tuning strategies with progressively unfrozen layers. This approach assesses both accuracy improvement and computational efficiency. The scenario configurations are shown in Table 5.

Table 5. Transfer Learning Scenarios

Parameter	Scenario 1	Scenario 2	Scenario 3
Pretrained Weight	*ImageNet*	*ImageNet*	*ImageNet*
Frozen Layers	All except the last 10	All except the last 25	50% of total layers
Unfrozen Layers	Last 10 layers	Last 25 layers	50% of layers unfrozen

Finally, hyperparameter testing was conducted across three scenarios that varied the batch size, input size, learning rate, optimizer, and classification layer design to identify the most efficient configuration. The scenario details are provided in Table 6.

Table 6. Hyperparameter

Parameter	Scenario 1	Scenario 2	Scenario 3
Batch Size	16	64	32
Input Size	250	224	240
Learning Rate	1e-4	0,001	0.0001
Optimizer	Adam	SGD	Adam
Epoch	50	50	50
Early Stop	patience=5	patience=3	patience=3
Dense	512	128	Conv2D + Dense256 + Dense128
Dropout	0.6	0.3	0.5 + 0.3
Pooling	GAP + BN	GAP + BN	GAP + BN

RESULT

Experimental Results

This section presents the results of a series of experiments conducted to evaluate the model's performance across various training scenarios. All optimization experiment results are summarized in Table 7.

Table 7. Experimental Results

Experiment Name	Validation Accuracy	Training Time
Baseline EfficientNet	0.2849	4 hours 12 minutes 45 seconds
Preprocessing Scenario 1	0.2849	4 hours 12 minutes 45 seconds
Preprocessing Scenario 2	0.2887	3 hours 52 minutes 8 seconds
Preprocessing Scenario 3	0.2849	3 hours 54 minutes 38 seconds
Transfer Learning Scenario 1	0.9630	3 hours 28 minutes 48 seconds
Transfer Learning Scenario 2	0.9734	3 hours 44 minutes 4 seconds
Transfer Learning Scenario 3	0.9924	5 hours 53 minutes 31 seconds
Hyperparameter Scenario 1	0.9563	5 hours 22 minutes 54 seconds
Hyperparameter Scenario 2	0.8955	2 hours 46 minutes 14 seconds
Hyperparameter Scenario 3	0.9668	3 hours 7 minutes 39 seconds

Comparison with Previous Methods

To demonstrate the effectiveness of the optimization strategies applied in this study, the EfficientNet-B1 model was compared with results from previous research that used the same dataset and architecture. The dataset, compiled from Figshare, SARTAJ, and Br35H via the Kaggle platform, ensures a consistent basis for comparison. This comparison highlights that the combination of preprocessing, transfer learning, and hyperparameter tuning techniques can substantially enhance model performance. The accuracy comparison is presented in Table 8.

Table 8. Comparison with Previous Methods

Author	Dataset	Method	Accuracy
Shafayat Bin S. M. et al. [8]	7023 Images	VGG16	95.52%
Dmytro Filatov et al. [12]	7023 Images	EfficientNet-B1	89.55%
This Study	7023 Images	EfficientNet-B1	97.34%

Based on Table 8, the optimized EfficientNet-B1 model in this study achieved a validation accuracy of 97.34%, outperforming two previous studies that used the same dataset of 7,023 brain tumor MRI images. The study by Shafayat Bin S. M. et al. using VGG-16 reported an accuracy of 95.52%, while the study by Dmytro Filatov et al. using EfficientNet-B1 achieved 89.55%. This comparison, conducted on an identical dataset, provides objective evidence that the optimization strategies implemented in this research significantly improve model performance without data-related bias. The performance gain is attributed to the combination of optimization methods applied, including data augmentation, fine-tuning of selected pretrained layers, and hyperparameter adjustments, all of which enhance the model’s ability to extract critical patterns from MRI images. This improvement is particularly important in medical applications, where high accuracy is essential to reduce the risk of misdiagnosis and to support more reliable clinical decision-making.

DISCUSSION

The experimental results demonstrate that the optimization strategies applied in this study have a substantial impact on the performance of the EfficientNet-B1 model for brain tumor MRI classification. The baseline and preprocessing stages, despite applying basic input normalization, produced very low accuracy (below 0.29), indicating that training the model entirely from scratch limits its ability to extract meaningful features and leads to poor generalization.

A significant performance improvement emerged when transfer learning with ImageNet-pretrained weights was applied, with accuracy rising sharply to above 0.96. This confirms that features learned from large-scale natural image datasets remain relevant for medical imaging tasks, especially during early feature extraction. Gradual fine-tuning further enhanced performance, where Scenario 2 (unfreezing the last 25 layers) achieved the highest validation accuracy while maintaining efficient training time.

The hyperparameter experiments also showed that training configuration strongly influences prediction quality. The best performance was achieved using an input size of 240, a batch size of 32, and the Adam optimizer with a 0.0001 learning rate, yielding a validation accuracy of 0.9668 with the fastest training time. Additional classification layers and dropout contributed to improved generalization and reduced overfitting.

Beyond accuracy, computational efficiency is essential for clinical deployment. The results indicate that several optimization strategies achieve an effective balance between performance and resource requirements, making the model more feasible for real-world medical decision-support systems.

Compared to previous studies using the same dataset, the optimized model in this research outperforms both EfficientNet-B1 and VGG16 implementations reported in the literature. This confirms the effectiveness of the applied optimization strategies in enhancing brain tumor MRI classification performance.

CONCLUSION

This study demonstrates that the combination of preprocessing techniques, transfer learning, and hyperparameter tuning significantly enhances the performance of the EfficientNet-B1 model for MRI-based brain tumor detection. Transfer learning serves as the key contributor to performance improvement, with accuracy rising dramatically from below 0.29 in the baseline stage to over 0.96 after incorporating ImageNet-pretrained weights. Fine-tuning by unfreezing the last 25 layers provided the best balance between accuracy and training time, while optimal hyperparameter settings—input size of 240, batch size of 32, and the Adam optimizer with a 0.0001 learning rate—improved training stability and efficiency.

The best-performing model achieved a validation accuracy of 97.34%, surpassing previous studies using the same dataset and demonstrating stronger generalization to medical imaging. With its high accuracy and computational efficiency, the model shows strong potential for deployment as a decision-support system in brain tumor diagnosis. Future work may involve leveraging larger and more diverse datasets, as well as integrating explainable AI methods to enhance model interpretability in clinical practice.

Declaration by Authors

Acknowledgement: None

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Z. Khazaei et al., "The association between incidence and mortality of brain cancer and human development index (HDI): an ecological study," BMC Public Health, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12889-020-09838-4.
2. B. Babu Vimala, S. Srinivasan, S. K. Mathivanan, Mahalakshmi, P. Jayagopal, and G. T. Dalu, "Detection and classification of brain tumor using hybrid deep learning models," Sci Rep, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-50505-6.
3. N. Sharma, M. Kumar Jain, A. K. Agarwal, N. Sharma, M. K. Jain, and A. K. Agarwal, "BRAIN TUMOR CLASSIFICATION USING CNN," 2021.
4. E. Albalawi et al., "Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor," BMC Med Imaging, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01261-0.
5. M. Tan and Q. V Le, "EfficientNetV2: Smaller Models and Faster Training," 2021. [Online]. Available: https://github.com/google/
6. I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification," Cluster Comput, vol. 27, no. 8, pp. 11187–11212, Nov. 2024, doi: 10.1007/s10586-024-04532-1.
7. W. Zhou, J. Ji, Y. Jiang, J. Wang, Q. Qi, and Y. Yi, "EARDS: EfficientNet and attention-based residual depth-wise separable convolution for joint OD and OC segmentation," Front Neurosci, vol. 17, 2023, doi: 10.3389/fnins.2023.1139181.
8. J. Wang, Q. Liu, H. Xie, Z. Yang, and H. Zhou, "Boosted efficientnet: Detection of lymph node metastases in breast cancer using convolutional neural networks," Cancers (Basel), vol. 13, no. 4, pp. 1–14, Feb. 2021, doi: 10.3390/cancers13040661.
9. H. M. Rai, J. Yoo, and S. Dashkevych, "Two-headed UNetEfficientNets for parallel execution of segmentation and classification of brain tumors: incorporating postprocessing techniques with connected component labelling," J Cancer Res Clin Oncol, vol. 150, no. 4, Apr. 2024, doi: 10.1007/s00432-024-05718-1.
10. S. Okazaki et al., "RadImageNet and ImageNet as Datasets for Transfer Learning

in the Assessment of Dental Radiographs: A Comparative Study," Journal of Imaging Informatics in Medicine, vol. 38, no. 1, pp. 534–544, Feb. 2025, doi: 10.1007/s10278-024-01204-9.

11. O. Filatov, J. Ebert, J. Wang, and S. Kesselheim, "Time Transfer: On Optimal Learning Rate and Batch Size in The Infinite Data Limit," Jan. 2025, [Online]. Available: http://arxiv.org/abs/2410.05838

How to cite this article: Muhammad Fachri Mahyudin, Maman Somantri, Oky Dwi Nurhayati. Optimization of convolutional neural network-based classification using EfficientNet-B1. *International Journal of Research and Review*. 2025; 12(12): 684-691. DOI: *https://doi.org/10.52403/ijrr.20251270*

******