Trends in Health Informatics

www.thi.reapress.com

THI. Vol. 2, No. 2 (2025) 108–115.

Paper Type: Original Article

A Comparative Analysis of Lightweight Convolutional Neural Network Models for Brain Tumor Detection in MRI Images with an Explainable AI Approach

Fereshteh Alimahmoudi*

Department of Computer Engineering, Faculty of Engineering and Technology, Ayandegan University, Tonekabon, Iran; fereshteh.allimahmoudi@gmail.com.

Citation:

Received: 13 September 2024 *Revised: 03 December 2024* *Accepted: 27 January 2025*	Alimahmoudi, F. (2025). A comparative analysis of lightweight convolutional neural network models for brain tumor detection in MRI images with an explainable AI approach. *Trends in Health Informatics*, 2(2), 108-115.

Abstract

Accurate and timely detection of brain tumors from Magnetic Resonance Imaging (MRI) images plays a critical role in improving treatment outcomes and increasing patient survival rates. In recent years, models based on Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in medical image analysis; however, many of these models are computationally complex and thus unsuitable for practical applications, particularly in resource-constrained environments. In this study, a comparative analysis of several lightweight CNN models, including MobileNet, ShuffleNet, and EfficientNet, is presented for brain tumor detection from MRI images. A publicly available brain MRI dataset was utilized, and a preprocessing pipeline comprising normalization, data augmentation, and image quality enhancement was applied. The performance of the models was evaluated using metrics such as accuracy, sensitivity, specificity, and F1-score. In addition, to enhance model reliability, explainability techniques based on Grad-CAM were employed to analyze the regions influencing the decision-making process. The results indicate that lightweight CNN models, while significantly reducing computational complexity, are capable of achieving competitive accuracy in brain tumor detection. Furthermore, explainability analysis reveals that the models predominantly focus on tumor-relevant regions in most cases. The findings of this study can contribute to the development of fast, reliable, and clinically applicable diagnostic systems, particularly in resource-limited settings.

Keywords: Brain tumor detection, MRI images, Convolutional neural networks, Lightweight models, Explainable AI, Grad-CAM.

1 | Introduction

Brain tumors are among the most serious disorders of the central nervous system, where accurate and timely diagnosis plays a decisive role in selecting appropriate treatment strategies and improving patient survival

Corresponding Author: fereshteh.allimahmoudi@gmail.com

https://doi.org/10.22105/thi.vi.42

rates. Magnetic Resonance Imaging (MRI), as one of the most important non-invasive diagnostic tools, provides valuable structural information about brain tissues. However, manual analysis of MRI images is a time-consuming process, highly dependent on clinical expertise, and prone to human error, particularly as data volume increases and disease patterns become more complex [1].

In recent years, significant advances in deep learning, particularly Convolutional Neural Networks (CNNs), have led to a fundamental transformation in automated medical image analysis. Owing to their ability to extract hierarchical features, these models have demonstrated remarkable performance in tasks such as classification, detection, and segmentation of medical images. Nevertheless, many of the proposed models are characterized by complex architectures and a large number of parameters, resulting in high computational costs and limiting their applicability in real-world scenarios, especially in resource-constrained clinical environments [2].

In this context, the use of lightweight models (lightweight CNNs) has emerged as an effective solution for reducing computational complexity while maintaining acceptable accuracy. Models such as MobileNet, ShuffleNet, and EfficientNet, through optimized architectural design, enable deployment on systems with limited computational resources. However, a comprehensive comparative evaluation of these models in the context of brain tumor detection from MRI images still requires systematic and rigorous investigation [3].

On the other hand, one of the major challenges in applying deep learning models in the medical domain is expainability. Many of these models are considered black-box systems, and the lack of transparency in decision-making processes can pose a significant barrier to their adoption in clinical practice. In this regard, techniques such as Grad-CAM provide the capability to identify the regions that most influence model predictions, thereby enhancing the interpretability and trustworthiness of intelligent systems [4].

Considering these challenges, the present study conducts a comparative analysis of several lightweight CNN models for brain tumor detection from MRI images. In addition to quantitative performance evaluation, explainability methods are employed to analyze model behavior. The primary objective of this research is to investigate the trade-off between accuracy, computational efficiency, and interpretability, with the aim of developing practical and reliable solutions suitable for real-world clinical environments.

2 | Related Work

In recent years, the application of deep learning-based methods in medical image analysis, particularly for brain tumor detection from MRI images, has attracted considerable attention. CNNs, due to their strong capability in automatically extracting meaningful features, have demonstrated successful performance across numerous studies [5].

In early research, deep and complex architectures such as AlexNet, VGG, and ResNet were employed for MRI image classification. These models, leveraging multi-layered structures, achieved high accuracy in brain tumor detection. However, their high computational complexity, reliance on powerful hardware resources, and long training times are considered major limitations of these approaches [6].

Subsequently, researchers adopted transfer learning techniques to mitigate the need for large-scale datasets and extensive training time by utilizing pre-trained models. In this context, architectures such as VGG16, ResNet50, and Inception have been widely used, yielding satisfactory results in automated brain tumor detection [7].

With the growing emphasis on practical applications and hardware constraints, lightweight CNN models have been introduced as efficient alternatives to heavy architectures. Models such as MobileNet, ShuffleNet, and EfficientNet are specifically designed to reduce the number of parameters and optimize computational efficiency. Several studies have demonstrated that these models can significantly lower computational costs while maintaining acceptable accuracy, making them particularly suitable for deployment in clinical environments and edge-based systems [8].

On the other hand, the concept of Explainable AI (XAI) has gained increasing importance in deep learning-based systems, especially in the medical domain. Due to the black-box nature of many deep models, the trust of clinicians in these systems remains a significant challenge. In this regard, methods such as Grad-CAM and other heatmap-based techniques have been utilized to identify the regions that influence model decisions. These approaches enable visual interpretation of model behavior and can play a crucial role in validating predictions and enhancing transparency [9].

Despite these advancements, several gaps remain in the existing literature. Many studies primarily focus on achieving maximum accuracy, with less attention given to the balance between accuracy, computational complexity, and explainability. Moreover, comprehensive and systematic comparisons of lightweight models within a unified framework, particularly for brain tumor detection from MRI images, have been relatively underexplored [10].

Accordingly, the present study aims to address these gaps by providing a comparative analysis of multiple lightweight CNN models while incorporating explainability techniques. The goal is to establish a practical framework that balances performance, efficiency, and interpretability for real-world clinical applications.

3 | Methodology

In this study, a deep learning framework is designed and implemented for brain tumor detection from MRI images, with the objective of providing a consistent and fair comparison of several lightweight CNN models in terms of detection accuracy, computational efficiency, and explainability. The overall structure of the proposed method consists of four main stages: data preparation, image preprocessing, model training, and performance evaluation along with explainability analysis.

Fig. 1 illustrates the flowchart of the proposed method for brain tumor detection using lightweight CNN models. As shown in *Fig. 1*, the framework comprises a sequence of stages ranging from data acquisition to explainability analysis. In this pipeline, MRI images are first subjected to preprocessing and data augmentation, after which they are fed into lightweight CNN models. Subsequently, model performance is evaluated using multiple metrics, and finally, Grad-CAM is employed to analyze the regions influencing model decisions. This framework enables simultaneous assessment of accuracy, computational efficiency, and interpretability.

In the first stage, a publicly available brain MRI dataset containing both tumor and non-tumor samples is utilized. In this study, the BRTS dataset was employed to conduct experiments and evaluate the proposed models. The collected data are divided into training and testing sets to ensure a fair evaluation of model performance. To improve model robustness and prevent overfitting, data augmentation techniques such as rotation, scaling, translation, and brightness adjustment are applied. This process increases the diversity of training samples and enhances the generalization capability of the models in real-world scenarios.

During the preprocessing stage, MRI images are resized to a uniform dimension, followed by pixel intensity normalization. This ensures that all images fall within a standardized numerical range, allowing the models to learn features more effectively and consistently. Additionally, in the presence of noise, simple filtering techniques are applied to reduce noise effects and improve input quality.

In the training stage, three lightweight CNN models, namely MobileNet, ShuffleNet, and EfficientNet, are selected as baseline architectures and trained on the prepared dataset. These models are employed using a transfer learning approach, where their initial weights are derived from pre-training on large-scale datasets such as ImageNet and subsequently fine-tuned for the brain tumor detection task. All models are trained under identical conditions, including the same number of epochs, a fixed learning rate, and a consistent batch size, to ensure a fair comparison.

To optimize the learning process, the cross-entropy loss function is used as the objective function, and the Adam optimizer is employed for parameter updates. Furthermore, to mitigate overfitting, techniques such as Dropout and Early Stopping are incorporated during training.

In the evaluation stage, model performance is assessed using standard metrics, including accuracy, sensitivity, specificity, and F1-score. In addition, to evaluate computational efficiency, the number of parameters and inference time are measured, enabling analysis of the trade-off between accuracy and computational complexity.

Finally, to enhance the interpretability of the models, the Grad-CAM method ([11]) is employed. This technique generates heatmaps of the most activated regions in the input image, indicating the areas on which the model focuses during its decision-making process. Such analysis helps determine whether the model is truly attending to tumor-relevant regions, thereby improving the overall trustworthiness of the system in medical applications.

Overall, the proposed method aims to provide an integrated framework for comparing lightweight CNN models in brain tumor detection, in which, beyond accuracy, computational efficiency and explainability are also taken into consideration. This approach can be utilized in the development of fast, reliable, and practical diagnostic systems, particularly in resource-constrained clinical environments.

The pseudocode of the proposed approach is provided in the following section.

Table 1 presents a comparative evaluation of the models' performance. As shown, EfficientNet-B0 achieves the highest accuracy and F1-score, while ShuffleNet demonstrates the lowest number of parameters and the fastest inference time. These results indicate that model selection is directly dependent on the trade-off between predictive performance and computational efficiency.

Table 1. Comparative analysis of model performance.

Params (Millions)	F1-Score (%)	Specificity (%)	Sensitivity (%)	Accuracy (%)	Model
3.4	93.9	94.8	93.5	94.2	MobileNet
2.3	92.0	93.1	91.6	92.8	ShuffleNet
5.3	95.3	96.0	95.1	95.6	EfficientNet-B0

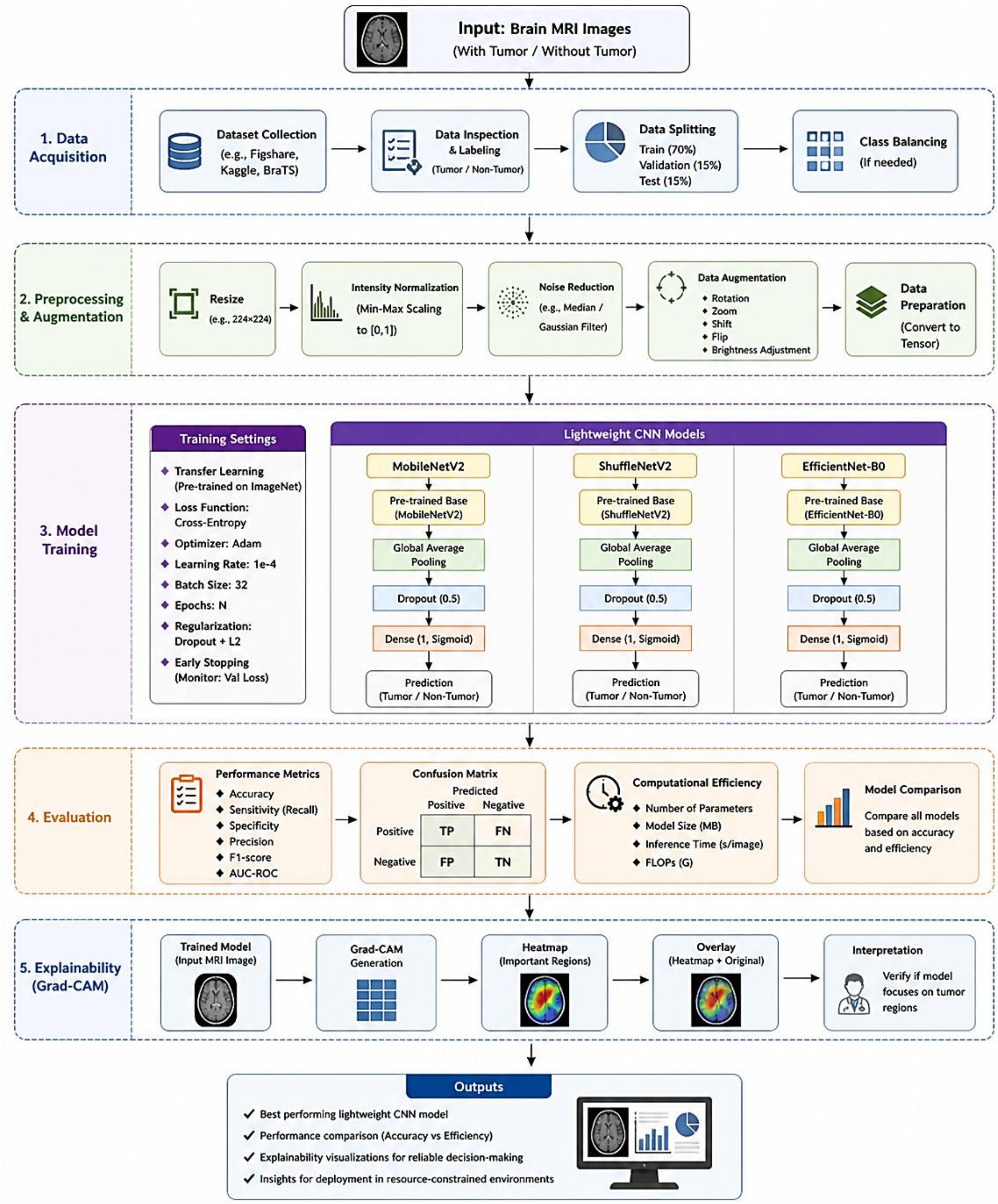

Fig. 1. Flowchart of the proposed method for brain tumor detection from MRI images using lightweight CNN models, including data acquisition, preprocessing, model training, performance evaluation, and explainability analysis.

Algorithm 1. Brain tumor detection using lightweight CNNs with explainability.

Input:

MRI Dataset D = {X, Y}, X: MRI images, Y: labels (Tumor / No Tumor)

Output:

Best performing model M_best, Performance metrics and Explainability maps

Begin

1. Load Dataset D

2. Preprocessing:

For each image x in X do:

Resize x to fixed size (e.g., 224×224)

Normalize pixel values

If noise exists: Apply denoising filter

End For

3. Data Augmentation:

For each image x in X do:

Generate augmented samples using: (Rotation, Horizontal/Vertical flipping, Scaling, Brightness, adjustment)

End For

4. Split Dataset: Divide D into: Training set (D_train), Testing set (D_test)
5. Initialize Models: Models = {MobileNet, ShuffleNet, EfficientNet}
6. For each model M in Models do:
    a. Load pre-trained weights (Transfer Learning)
    b. Modify final classification layer for binary output
    c. Train M on D_train using:

Loss function: CrossEntropy

Optimizer: Adam

Apply Dropout to reduce overfitting

Use Early Stopping

    d. Evaluate M on D_test: Compute Accuracy, Compute Sensitivity, Compute Specificity, Compute F1-score, Measure Inference Time, Count number of Parameters
    e. Generate Explainability Maps: Apply Grad-CAM on test images, Visualize important regions

End For

7. Compare all models based on: Classification performance, Computational efficiency, Explainability results
8. Select best model: M_best = argmax(performance trade-off score)

Return M_best, evaluation results, Grad-CAM visualizations

End

4 | Discussion

In this study, the performance of three lightweight CNN models, namely MobileNet, ShuffleNet, and EfficientNet-B0, was comparatively evaluated for brain tumor detection from MRI images. The results indicated that although all three models demonstrate acceptable performance in medical image classification, there are meaningful differences in terms of accuracy, computational efficiency, and explainability, highlighting the importance of selecting an appropriate model based on application-specific constraints.

Based on the obtained results, EfficientNet-B0 achieved the highest performance in terms of classification metrics, including accuracy, sensitivity, and F1-score. This superiority can be attributed to its optimized architecture and its ability to extract multi-level feature representations. However, this improved performance was accompanied by a relatively higher number of parameters and increased inference time, indicating that higher accuracy is often associated with greater computational cost.

In contrast, ShuffleNet exhibited the lowest computational complexity among the evaluated models and achieved the best performance in terms of parameter efficiency and inference speed. However, its relatively lower performance in accuracy and F1-score suggests that excessive reduction in model complexity may impair the ability to learn complex patterns in medical images. MobileNet, as an intermediate model, provided a balanced trade-off between accuracy and computational efficiency and can be considered a suitable option for practical applications in resource-constrained environments.

From a clinical perspective, the findings of this study suggest that model selection should not be based solely on classification accuracy, but should also consider factors such as processing speed, hardware limitations, and real-time requirements. In clinical settings with limited computational resources, models such as MobileNet or ShuffleNet may be more practical choices, whereas in more advanced diagnostic systems, EfficientNet-B0 may provide higher diagnostic accuracy.

An important aspect of this study is the evaluation of model explainability using Grad-CAM. The results showed that EfficientNet-B0 generally focused on tumor-relevant regions, which may indicate that the model has learned meaningful features during training. However, in some cases, lighter models such as ShuffleNet exhibited more dispersed attention maps, which could partly explain their relatively lower performance. This finding highlights the importance of explainability methods in medical applications, as the trustworthiness of intelligent systems in clinical decision-making strongly depends on model transparency.

Compared to previous studies, the results of this research demonstrate that lightweight CNN models can serve as suitable alternatives for medical applications under computational constraints while maintaining acceptable accuracy. However, as reported in related literature, challenges still remain in achieving an optimal balance among accuracy, speed, and explainability, which requires further investigation.

Overall, the findings emphasize that the design of AI-based diagnostic systems for medical applications should not rely solely on performance metrics, but should adopt a multi-criteria approach involving accuracy, efficiency, and interpretability. This perspective can play a significant role in the development of reliable and practically deployable systems in real-world clinical environments.

5 | Conclusion

The results of this study indicate that all three lightweight CNN models achieve acceptable performance in brain tumor detection from MRI images. However, EfficientNet-B0 stands out in terms of accuracy, ShuffleNet in computational efficiency, and MobileNet as a balanced alternative between the two extremes. Explainability analysis further demonstrated that more accurate models tend to exhibit better focus on tumor-relevant regions. Overall, model selection should be guided by a multi-criteria perspective that considers accuracy, speed, and interpretability simultaneously.

As future directions, simultaneous improvement of these aspects can be pursued through the design of more optimized architectures, the use of hybrid or ensemble methods, the application of advanced explainability techniques, and evaluation on larger and more diverse clinical datasets, in order to enhance model generalizability and trustworthiness in real-world environments.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

All data are included in the text.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Mohsen, S., Oraby, S., & Abdel-Aziz, M. (2025). Deep learning and machine learning for brain tumor detection: A review, challenges, and future directions. *Archives of computational methods in engineering*, *33*, 3931–3955. https://doi.org/10.1007/s11831-025-10416-3

[2] Isong, N. (2025). *Building efficient lightweight CNN models*. https://arxiv.org/abs/2501.15547

[3] Dorfner, F. J., Patel, J. B., Kalpathy-Cramer, J., Gerstner, E. R., & Bridge, C. P. (2025). A review of deep learning for brain tumor analysis in MRI. *NPJ precision oncology*, *9*(1), 2. https://www.nature.com/articles/s41698-024-00789-2

[4] Mittal, P. (2025). From black box AI to XAI in neuro-oncology: a survey on MRI-based tumor detection. *Discover artificial intelligence*, *5*(1), 30. https://doi.org/10.1007/s44163-025-00247-3

[5] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). *Shufflenet: An extremely efficient convolutional neural network for mobile devices* [presentation]. Proceedings of the ieee conference on computer vision and pattern recognition (pp. 6848–6856). https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_ShuffleNet_An_Extremely_CVPR_2018_paper.html

[6] Muthulakshmi, M., Venkatesan, K., Harigaran, R., Jeevanantham, K., Vineeth, M. S., Rahayu, S. B., & Sakthivel, V. (2024). Comparative study of efficientnet and mobilenet models for lung cancer classification using chest ct scan images. *2024 second international conference on emerging trends in information technology and engineering (ICETITE)* (pp. 1–6). IEEE. https://doi.org/10.1109/ic-ETITE58242.2024.10493412

[7] Hosny, K. M., & Mohammed, M. A. (2025). Explainable AI and vision transformers for detection and classification of brain tumor: A comprehensive survey. *Artificial intelligence review*, *58*(9), 259. https://doi.org/10.1007/s10462-025-11221-x

[8] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning* (pp. 6105–6114). PMLR. https://proceedings.mlr.press/v97/tan19a.html?ref=ji

[9] Yurdakul, M., & Taşdemir, Ş. (2026). MRI-based brain tumor detection through an explainable EfficientNetV2 and MLP-mixer attention architecture. *Physical and engineering sciences in medicine*, 1–19. https://doi.org/10.1007/s13246-026-01728-0

[10] Satushe, V., Vyas, V., Metkar, S., & Singh, D. P. (2025). AI in MRI brain tumor diagnosis: A systematic review of machine learning and deep learning advances (2010--2025). *Chemometrics and intelligent laboratory systems*, *263*, 105414. https://doi.org/10.1016/j.chemolab.2025.105414

[11] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-cam: Visual explanations from deep networks via gradient-based localization* [presentation]. Proceedings of the ieee international conference on computer vision (pp. 618–626). https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html