TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 6, December 2026, pp. 2048~2059
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i6.28015

Lightweight explainable CNN for brain tumor detection from MRI with grad-CAM and visual analytics

Dipta Paul[1], Ekramul Hasan[2], Md Al Amin[3], Md Tuhin Mia[3], Md Musa Haque[3]
[1]Department of Electrical and Electronic Engineering, American International University - Bangladesh (AIUB), Dhaka, Bangladesh
[2]College of Technology and Engineering, Westcliff University, California, United States
[3]School of Business, International American University, California, United States

Article Info

Article history:

Received Apr 30, 2026
Revised Aug 12, 2026
Accepted Sep 8, 2026

Keywords:

Brain tumor detection
Convolutional neural network
Deep learning
Explainable artificial intelligence
Grad-class activation mapping
Magnetic resonance imaging
Medical image analysis

ABSTRACT

Brain tumor detection from magnetic resonance imaging (MRI) is important for clinical diagnosis, requiring accurate and interpretable automated systems. However, many convolutional neural network (CNN)-based approaches provide limited insight into their predictions, restricting their potential use in clinical decision support. This study presents an explainable CNN framework for binary brain tumor classification, integrating CNN-based feature extraction with visual analytics. Evaluated on 253 MRI images (155 tumor and 98 non-tumor) using an 80:20 training-validation split, the model achieved 78.43% accuracy across 51 validation samples. The results reveal differences in class-wise performance, highlighting the need for further improvement. Despite the potential of visual analytics to support interpretability, the small dataset and limited validation constrain generalizability. Further evaluation using larger datasets, robust validation protocols, and verified explainability methods is needed to establish the framework's reliability for clinical decision support.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Dipta Paul
Department of Electrical and Electronic Engineering
American International University - Bangladesh (AIUB)
Dhaka, Bangladesh
Email: diptapal4@gmail.com

1. INTRODUCTION

Brain tumors are abnormal growths of cells within the brain that can be either benign or malignant and may significantly affect neurological function. Early and accurate diagnosis is essential for effective treatment planning and improved patient survival [1]. Magnetic resonance imaging (MRI) is widely recognized as a non-invasive imaging modality because of its high spatial resolution and ability to provide detailed visualization of brain structures [2]-[4]. Despite these advantages, most convolutional neural network (CNN)-based approaches emphasize classification accuracy while providing limited interpretability, restricting their adoption in real-world clinical decision support systems.

The lack of transparent decision-making remains a major challenge in deploying deep learning models for medical diagnosis. Manual interpretation of MRI scans by radiologists is time-consuming and depends heavily on clinical expertise, which may result in variability and delayed diagnosis, particularly during the early stages of disease [5]. Furthermore, the similarity between tumor and non-tumor tissues makes automated classification challenging. Although deep learning techniques have significantly improved diagnostic accuracy, many existing models function as "black boxes," offering limited explanation of their predictions [6]. Consequently, there is an increasing demand for explainable artificial intelligence (XAI) techniques that improve transparency while maintaining reliable classification performance. Recent advances

in artificial intelligence (AI) have established CNNs as one of the most effective approaches for medical image analysis. CNN-based and hybrid deep learning models have demonstrated promising performance in brain tumor detection as well as other healthcare applications, including skin cancer, breast cancer, and neurological disease diagnosis [7]-[17]. In particular, explainability methods such as gradient-weighted class activation mapping (Grad-CAM) have been introduced to visualize the image regions influencing model predictions, thereby increasing clinician confidence in AI-assisted diagnosis [15]-[17].

However, existing CNN-based approaches often prioritize classification accuracy while offering limited interpretability and requiring relatively high computational resources, restricting their deployment in resource-constrained healthcare environments. To address these challenges, this paper proposes a lightweight explainable CNN framework that combines efficient feature extraction with a unified visual analytics framework to improve computational efficiency, model interpretability, and practical applicability. Unlike conventional explainable CNN models that rely on a single visualization technique, the proposed framework enables complementary analysis of learned features, prediction confidence, and model attention, providing more comprehensive model interpretation and greater transparency for computer-aided diagnosis.

The main contributions of this work are summarized as: (i) a lightweight CNN framework is proposed for efficient binary brain tumor classification from MRI images, providing a balance between classification performance and computational efficiency; (ii) grad-CAM is integrated with the proposed CNN to generate interpretable visual explanations, enhancing the transparency of the model's prediction process; and (iii) an integrated visual analytics framework is incorporated to complement Grad-CAM by providing additional insights into feature representation and model behavior. The effectiveness of the proposed framework is validated through experimental evaluation using standard performance metrics and qualitative visual analysis.

2. RELATED RESEARCH WORK

Recent advancements in deep learning have significantly improved brain tumor detection from MRI images. Several studies have explored CNN, hybrid models, and transfer learning to enhance classification accuracy, robustness, and diagnostic reliability [18]-[20]. Lightweight deep learning architectures have also gained attention for reducing computational complexity while maintaining competitive performance in resource-constrained environments.

Furthermore, XAI has become increasingly important for improving transparency and trust in medical AI systems. Techniques such as Grad-CAM have been widely adopted to visualize regions influencing model predictions [21]-[24]. Despite these advances, existing approaches continue to face challenges related to interpretability, dataset limitations, and the lack of comprehensive visual analysis [25]. While lightweight CNNs improve computational efficiency and XAI enhances transparency, most studies primarily emphasize either classification performance or single-method visualization, providing limited insight into feature representation and model behavior. Consequently, there remains a need for an integrated framework that combines efficient classification with complementary visual analysis to improve both interpretability and practical applicability. In contrast, the proposed work addresses these limitations by integrating a lightweight CNN architecture with explainable AI techniques and a comprehensive visual analysis framework, thereby improving both performance and interpretability [26], [27].

3. RESEARCH METHOD

The proposed methodology consists of five sequential stages: (i) MRI dataset acquisition, (ii) image preprocessing, (iii) lightweight CNN-based feature extraction and model training, (iv) brain tumor classification, and (v) explainability generation using integrated visual analytics. Following prediction, Grad-CAM and complementary visual analysis are applied to interpret the learned features and prediction outcomes. The overall workflow of the proposed framework is illustrated in Figure 1.

3.1. Dataset description

The dataset consists of 253 publicly available brain MRI images categorized into tumor and non-tumor classes [27]. It was divided into training and validation subsets using an 80:20 ratio for model evaluation, as summarized in Table 1.

3.2. Data preprocessing

All MRI images were resized to 224×224 pixels, normalized, and converted to red, green, and blue (RGB) format to ensure consistent input representation. No additional data augmentation techniques were applied, allowing the proposed framework to be evaluated using the original dataset. These preprocessing steps improve training stability while reducing computational complexity.

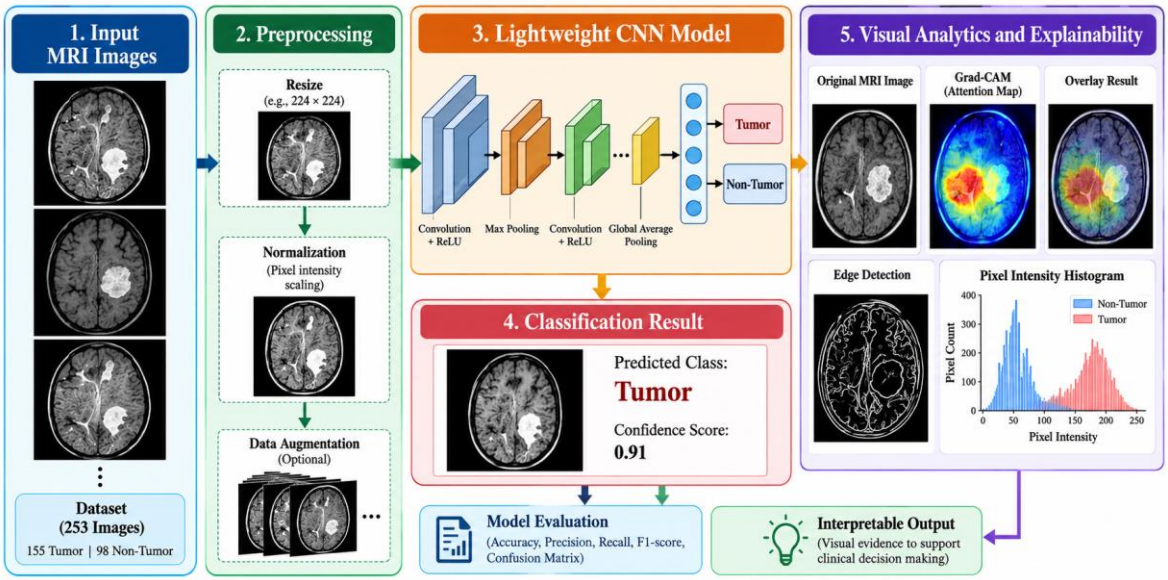


Figure 1. Overall workflow of the proposed framework illustrating dataset acquisition, preprocessing, lightweight CNN training, classification, and explainability generation using integrated visual analytics

Table 1. Dataset summary [27]

Class	Number of images
Tumor	155
Non-tumor	98
Total	253

3.3. CNN-based feature extraction

A lightweight CNN is designed to extract discriminative features from MRI images for brain tumor classification. Table 2 summarizes the proposed CNN architecture.

Table 2. Summary of proposed CNN architecture

Layer type	Output size	Description
Conv2D (32)	222×222×32	Feature extraction
MaxPooling	111×111×32	Downsampling
Conv2D (64)	109×109×64	Deeper feature learning
MaxPooling	54×54×64	Downsampling
Conv2D (128)	52×52×128	High-level features
MaxPooling	26×26×128	Downsampling
Fully connected	128	Classification layer
Output layer	2	Tumor / non-tumor

3.3.1. Training configuration and evaluation

The proposed CNN was implemented using PyTorch and trained for five epochs with a batch size of 32, using the Adam optimizer with a learning rate of 0.001 and cross-entropy loss as the optimization objective. All MRI images were resized to 224/times 224/pixels, converted to RGB format, and normalized using the implemented preprocessing pipeline. The dataset contained 253 images, comprising 155 tumor and 98 non-tumor images. It was randomly divided into training and validation subsets using an 80:20 split, resulting in 202 training images and 51 validation images. The split was generated using PyTorch's random splitting procedure without a fixed random seed or explicit class stratification. Therefore, the exact subset assignment may vary across runs, and the class distribution in the validation subset may not reproduce the overall dataset distribution. Model performance was evaluated on the held-out validation subset using the predicted class labels and corresponding ground-truth labels. Accuracy was calculated as the proportion of correctly classified images among all validation samples. Class-wise precision, recall, and F1-score were

calculated from the corresponding true-positive, false-positive, and false-negative counts. Overall precision, recall, and F1-score were reported as macro-averages across the two classes.

3.3.2. Architecture design considerations

The proposed lightweight CNN employs a three-layer convolutional architecture with 32, 64, and 128 filters, respectively, to balance classification performance and computational cost. Based on preliminary experimentation, this architecture was selected because increasing network depth provided limited performance improvements relative to the additional computational requirements. The model was designed to support brain tumor classification while maintaining a relatively compact convolutional structure and enabling visual interpretation of its predictions through explainability techniques.

The computational characteristics of the proposed lightweight CNN were evaluated in terms of trainable parameters, model checkpoint size, theoretical floating-point operations, and single-image inference latency. The network contains 11,169,218 trainable parameters, and the saved model checkpoint occupies 42.61 MiB. Its theoretical computational cost is approximately 0.944 GFLOPs per 224/times 224/input image, assuming two floating-point operations per multiply–accumulate operation (MAC). Inference latency was measured in a CPU-only environment using PyTorch 2.10.0 and five CPU threads. Following 20 warm-up iterations, 100 inference runs yielded a median latency of 12.53 ms, a mean latency of 16.38 ms, and a 95th-percentile latency of 36.06 ms. These measurements characterize the computational requirements of the proposed model under the specified benchmarking conditions. Although the model was designed with computational efficiency in mind, further comparisons with established lightweight CNN architectures under identical hardware and evaluation conditions are required to quantify its relative efficiency.

3.4. Model evaluation

The performance of the proposed model is evaluated using standard metrics including accuracy, precision, recall, and F1-score. Additionally, a confusion matrix is used to visualize the classification performance and identify misclassification patterns.

3.5. Explainability and visual analysis

To improve model interpretability, Grad-CAM is employed to visualize the image regions contributing most to each prediction. Rather than relying on a single explainability technique, the proposed framework integrates complementary visual analytics, including feature-space visualization, edge-based analysis, confidence distribution, and attention mapping. These analyses provide multiple perspectives on model behavior, allowing a more comprehensive interpretation of learned representations, prediction confidence, and structural characteristics of MRI images. Consequently, the framework enhances transparency beyond conventional Grad-CAM-based visualization alone. In summary, the proposed methodology integrates lightweight CNN-based feature extraction with Grad-CAM and complementary visual analytics to achieve efficient and interpretable brain tumor classification.

4. RESULT ANALYSIS

The experimental results of the proposed CNN-based brain tumor detection framework are presented and analyzed using multiple evaluation metrics and visual outputs.

4.1. Training performance analysis

The training performance of the model is illustrated in Figure 2. The recorded training-loss values decreased across the five epochs. Because the implementation records the accumulated batch loss rather than a separately calculated mean epoch loss, the curve is interpreted only as a trend in the recorded training loss; validation loss was not recorded. Table 3 summarizes the primary classification performance metrics, including accuracy, precision, recall, and F1-score, used to evaluate the proposed framework.

4.2. Classification performance

The confusion matrix in Figure 3 shows that 14 of 20 tumor images and 26 of 31 non-tumor images were correctly classified. The model misclassified six tumor images as non-tumor and five non-tumor images as tumor. Overall accuracy was 78.43% (40/51). The tumor-class precision, recall, and F1-score were 0.74, 0.70, and 0.72, respectively, while the corresponding non-tumor metrics were 0.81, 0.84, and 0.83. These results indicate that performance differs between classes and that missed tumor cases remain a concern. Because evaluation used only 51 validation images, the results should be interpreted cautiously.

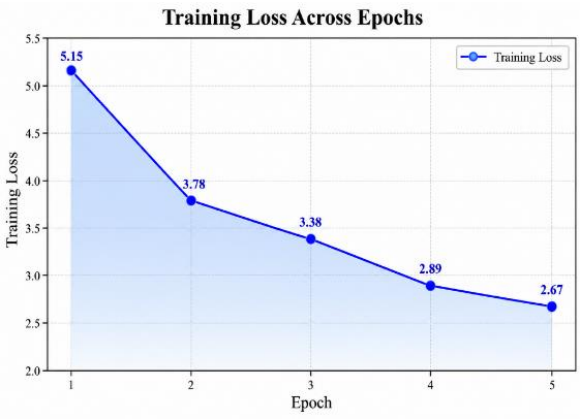


Figure 2. Training loss curve showing stable convergence of the proposed model during training

Table 3. Performance metrics of the proposed model

Metric	Tumor	Non-tumor	Overall
Precision	0.74	0.81	0.77
Recall	0.70	0.84	0.77
F1-score	0.72	0.83	0.77
Accuracy	-	-	0.78

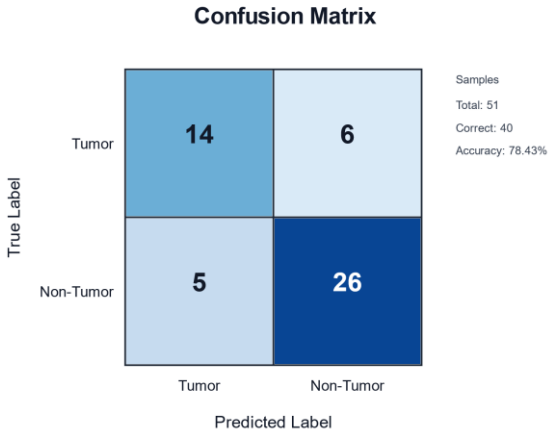


Figure 3. Confusion matrix of the proposed lightweight CNN

As shown in Table 4, the model correctly classified 14 tumor images and 26 non-tumor images, while misclassifying 6 tumor images as non-tumor and 5 non-tumor images as tumor, achieving an overall accuracy of 78.43%.

Table 4. Confusion matrix representation

Actual/predicted	Tumor	Non-tumor
Tumor	14	6
Non-tumor	5	26

To further evaluate the discriminative capability of the learned deep features, t-distributed stochastic neighbor embedding (t-SNE) visualization was performed. As shown in Figure 4, the extracted embeddings exhibit partial class-wise separation, indicating that the proposed CNN successfully learns meaningful representations for classification. Figure 5 illustrates the hierarchical feature extraction behavior of the CNN model. Early layers capture low-level edges and contours, while deeper layers learn more abstract tumor-related patterns.

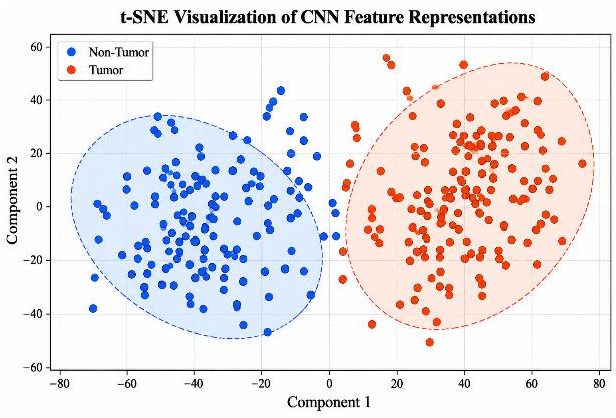


Figure 4. t-SNE feature space visualization of extracted CNN embeddings showing class-wise clustering of tumor and non-tumor MRI samples

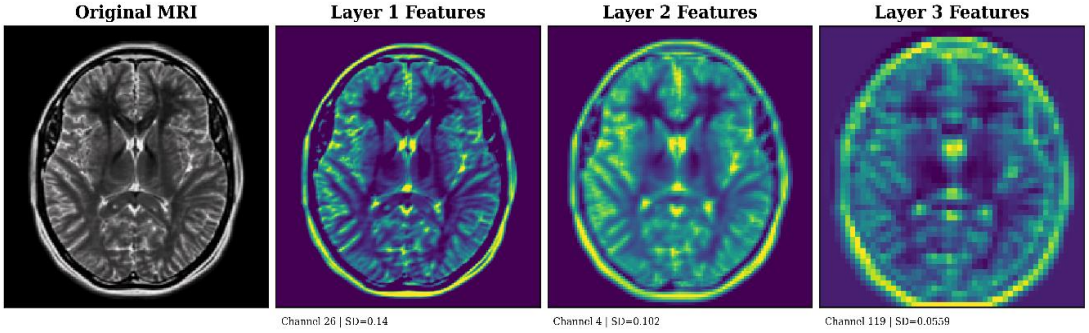


Figure 5. Multi-layer CNN feature map visualization demonstrating progressive extraction of spatial and structural features from brain MRI images

A representative misclassified case is presented in Figure 6. The model predicted the sample with high confidence despite incorrect classification, suggesting structural similarities and intensity overlap between classes.

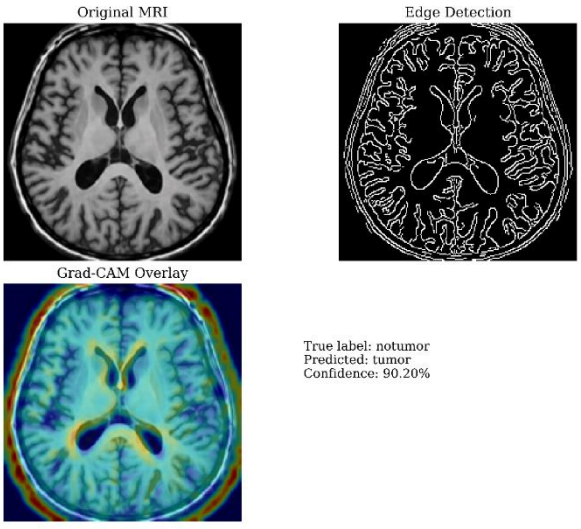


Figure 6. Visual analysis of a misclassified MRI sample including original image, edge detection, attention map, and confidence score

4.3. Misclassification dashboard

Figure 7 presents prediction confidence scores across multiple samples. Most correctly classified samples show high confidence, indicating stable predictions. The red regions correspond to the suspected tumor area, whereas the blue and green regions indicate lower model attention.

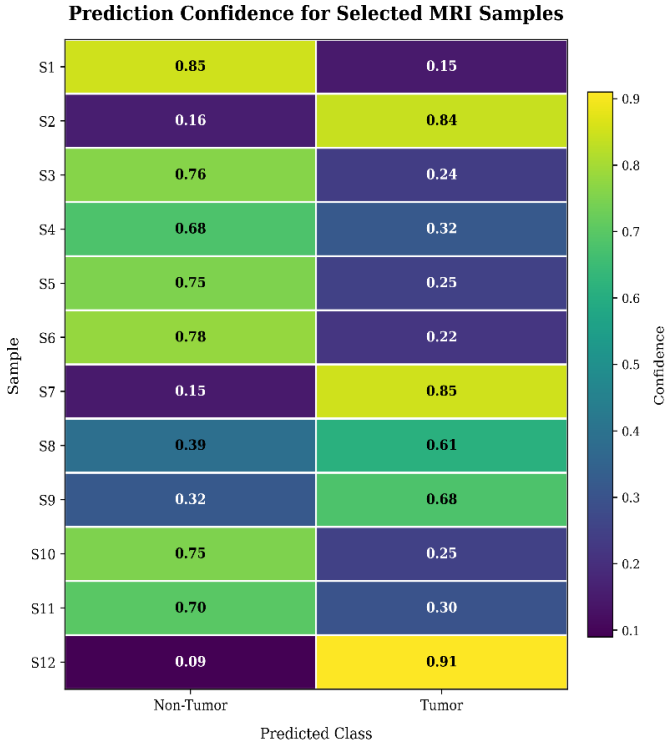


Figure 7. Prediction confidence distribution across selected MRI samples for tumor and non-tumor classes

4.4. Prediction results

Figure 8 presents sample classification results, highlighting the capability of the proposed lightweight CNN to distinguish between tumor and non-tumor MRI images.

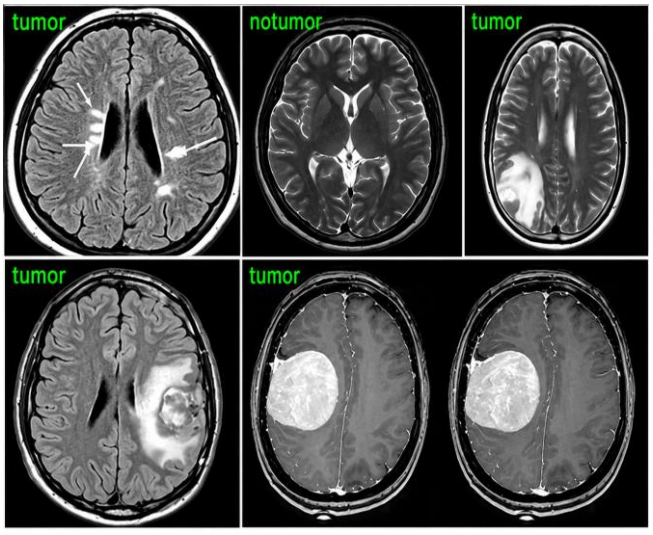


Figure 8. Sample MRI classification results generated by the proposed lightweight CNN

4.5. Grad-CAM visualization

To enhance interpretability, Grad-CAM visualizations are shown in Figure 9. The highlighted regions represent the areas that contribute most to the model's decision-making process. It is evident that the model focuses on the tumor-affected regions of the brain, confirming that the learned features are clinically meaningful and relevant.

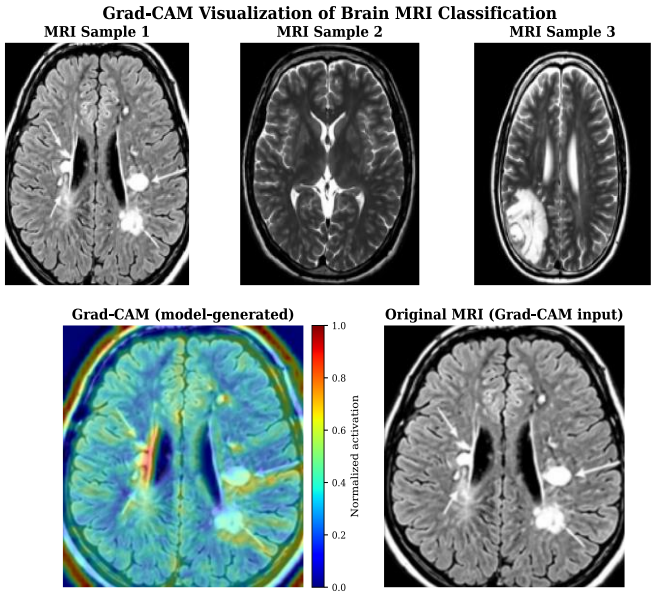


Figure 9. Grad-CAM visualization highlighting tumor-associated regions

4.6. Comprehensive visual analysis

Figure 10 presents a unified visualization that combines multiple analytical perspectives, including the original MRI image, prediction output, edge detection, and attention mapping. The edge detection highlights structural boundaries within the brain, while the attention map emphasizes important regions influencing classification.

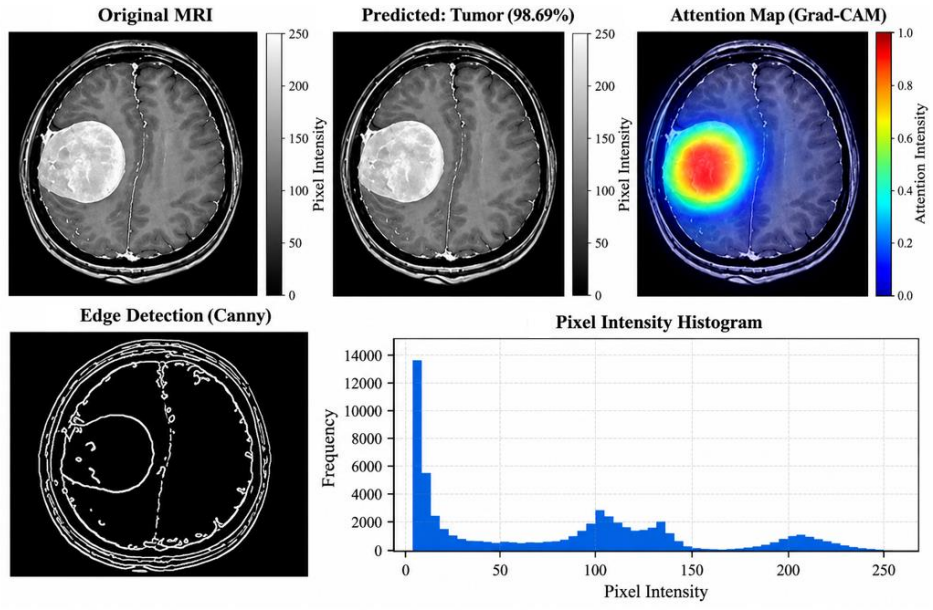


Figure 10. Comprehensive visual analysis including prediction, edge detection, and attention mapping

To provide enhanced interpretability, a three-dimensional surface representation of the tumor attention region is shown in Figure 11. Higher peaks correspond to stronger model focus areas associated with abnormal tissue structures.

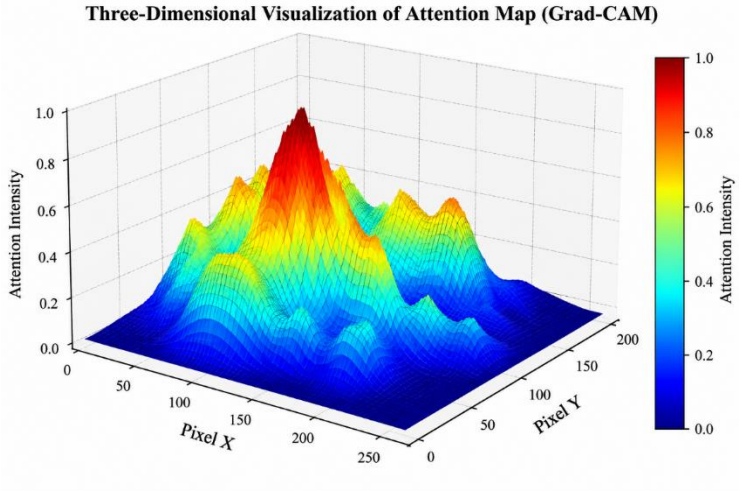


Figure 11. Three-dimensional attention surface representation of the tumor region highlighting spatial response intensity

4.7. Comparative performance analysis

The proposed model achieved an accuracy of 78.43% on the validation subset, based on the confusion matrix reported in Table 4. Recent studies have reported higher accuracy for brain-tumor classification using different MRI datasets and experimental configurations. However, direct performance comparisons should be interpreted cautiously because dataset composition, class definitions, validation protocols, and model architectures may differ. The present study focuses on binary classification using a relatively small dataset and integrates Grad-CAM with complementary visual analytics. Further evaluation using larger datasets, reproducible validation protocols, and standardized benchmarking is required to establish the model's generalizability and comparative performance.

As shown in Table 5, the proposed model achieved 78.43% accuracy on a dataset of 253 MRI images for binary brain-tumor classification. Although a higher accuracy of 99.54% was reported by AlShehri *et al*. [16] for multi-class classification, direct comparison is limited by differences in classification tasks, datasets, and evaluation settings. Furthermore, Repetto *et al*. [26] investigated the robustness of explainable AI rather than reporting a brain-tumor classification accuracy. Therefore, the reported results should be interpreted in the context of each study's specific objectives and experimental setup.

Table 5. Comparison of proposed model with recent studies

Study	Task/dataset	Reported result
AlShehri *et al*. [16]	Multi-class brain-tumor classification using MRI images	99.54% accuracy
Huda and Mahamud [25]	Review of CNN-based brain-tumor segmentation	Review study; no single model accuracy reported
Repetto *et al*. [26]	Robustness of explainable AI in medical image recognition under natural and adversarial data corruption	Evaluates the robustness of explanations; not a brain-tumor classification accuracy study
Proposed	Binary classification; 253 MRI images	78.43% accuracy*

5. LIMITATIONS AND FUTURE WORK

Despite the potential of the proposed framework, several limitations should be acknowledged. First, the dataset contains only 253 MRI images, which limits the generalizability of the findings to broader and more diverse populations. Second, the model achieved 70% recall for the tumor class, misclassifying 6 of the 20 tumor images in the validation subset. These missed cases highlight the need to improve tumor detection

sensitivity before considering practical clinical applications. Third, although the model's parameter count, theoretical computational cost, saved checkpoint size, and CPU inference latency were assessed, comparative benchmarking against established lightweight architectures under identical hardware conditions has not been performed. Peak memory consumption and edge-device performance also remain unevaluated. Finally, the reported results are based on a single random train–validation split, and no independent test set or repeated cross-validation was used. Future work will focus on evaluating the framework on larger and more diverse MRI datasets, employing repeated stratified validation or cross-validation, and comparing its performance with established CNN architectures. Further investigation will also examine the reliability of explainability methods, computational efficiency under consistent benchmarking conditions, and suitability for deployment on resource-constrained devices. These improvements are necessary to establish the robustness, generalizability, and practical applicability of the proposed framework.

6. CONCLUSION

This study presented a lightweight CNN framework for binary brain tumor classification from MRI images, supported by visual analytics for examining classification results. The proposed model achieved an overall accuracy of 78.43% on 51 validation images, correctly classifying 14 tumor images and 26 non-tumor images. The results demonstrate the potential of CNN-based approaches for automated MRI image classification while highlighting the need for further improvement in tumor detection performance. The integration of classification analysis and visual representations provides a basis for investigating model predictions more systematically. Overall, this study offers a preliminary framework for lightweight brain tumor classification and further research into explainable medical image analysis.

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
Dipta Paul	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ekramul Hasan		✓		✓	✓			✓	✓	✓				
Md Al Amin		✓								✓				
Md Tuhin Mia	✓								✓	✓	✓			
Md Musa Haque					✓					✓				

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

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The brain MRI image dataset used in this study is publicly available on Kaggle: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.

REFERENCES

[1] Q. T. Ostrom, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018," *Neuro-Oncology*, vol. 23, no. Supplement_3, pp. III1–III105, 2021, doi: 10.1093/neuonc/noab200.
[2] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," Physics

in Medicine and Biology, vol. 58, no. 13, pp. R97–R129, 2013, doi: 10.1088/0031-9155/58/13/R97.
[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[4] E. Hasan, D. S. Islam, D. Paul, M. Al Amin, M. T. Mia and M. Rahaman, "A CNN-Driven Deep Learning Model for Accurate Stage Detection in Medical Imaging to Enhance Stroke Diagnosis," *2025 2nd International Conference on Innovations in Engineering, Science and Technology for Sustainable Development (ICEST)*, 2025, pp. 1-6, doi: 10.1109/ICEST65883.2025.11428574.
[5] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, 2019, doi: 10.1016/j.jocs.2018.12.003.
[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy*, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
[7] O. K. Pal, D. Paul, E. Hasan, M. Mohammad, M. A. H. Bhuiyan, and F. Ahammed, "Advanced Convolutional Neural Network Model to Identify Melanoma Skin Cancer," *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, pp. 1–5, Apr. 21, 2023. doi: 10.1109/InC457730.2023.10262990.
[8] P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation," *Information Processing in Agriculture*, vol. 7, no. 4, pp. 566–574, 2020, doi: 10.1016/j.inpa.2019.11.001.
[9] S. Mohsen, S. Oraby, and M. Abdel-Aziz, "Deep LearnIng and Machine Learning for Brain Tumor Detection: A Review, Challenges, and Future Directions," *Archives of Computational Methods in Engineering*, vol. 33, no. 3, pp. 3931–3955, Oct. 2025, doi: 10.1007/s11831-025-10416-3.
[10] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020, doi: 10.3389/fbioe.2020.534592.
[11] H. C. Yoon and L. P. Lin, "Brain Tumor Classification in MRI: Insights From LIME and Grad-CAM Explainable AI Techniques," in *IEEE Access*, vol. 13, pp. 154172-154202, 2025, doi: 10.1109/ACCESS.2025.3603272.
[12] R. Singh, S. Gupta, S. Bharany, A. Almogren, A. Altameem and A. Ur Rehman, "Ensemble Deep Learning Models for Enhanced Brain Tumor Classification by Leveraging ResNet50 and EfficientNet-B7 on High-Resolution MRI Images," in *IEEE Access*, vol. 12, pp. 178623-178641, 2024, doi: 10.1109/ACCESS.2024.3494232.
[13] Md. A. Rahman *et al*., "GliomaCNN: An Effective Lightweight CNN Model in Assessment of Classifying Brain Tumor from Magnetic Resonance Images Using Explainable AI," *Computer Modeling in Engineering & Science*, vol. 140, no. 3, pp. 2425–2448, 2024, doi: 10.32604/cmes.2024.050760.
[14] M. Aly, A. Ghallab, and I. S. Fathi, "ViT-GRU: Advanced Brain Tumor Diagnosis Framework: Vision Transformer and GRU Integration for Improved MRI Analysis: A Case Study of Egypt," *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/access.2024.3513235.
[15] N. Bibi, J. Courtney, and K. McGuinness, "Enhancing Brain Disease Diagnosis with XAI: A Review of Recent Studies," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 2, pp. 1–35, Feb. 2025, doi: 10.1145/3709152.
[16] H. AlShehri, L. Almogbil, J. Alrajhi, L. Alosaimi, and J. Almarri, "An interpretable CNN framework for multi-class brain tumor classification with visual explanation in MRI scans," *Systems and Soft Computing*, vol. 8, p. 200493, Jun. 2026, doi: 10.1016/j.sasc.2026.200493.
[17] V. Viswan, N. Shaffi, M. Mahmud, K. Subramanian, and F. Hajamohideen, "Explainable Artificial Intelligence in Alzheimer's Disease Classification: A Systematic Review," *Cognitive Computation*, vol. 16, no. 1, pp. 1–44, Nov. 2023, doi: 10.1007/s12559-023-10192-x.
[18] G. M. Salama, M. K. Abd-Ellah, A. I. Hussein, E. S. Bayoumi, S. Ashraf, and M. Safy, "Brain tumor diagnosis techniques key achievements, lessons learned, and a new CNN architecture," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 56, no. 1, Sep. 2025, doi: 10.1186/s43055-025-01567-1.
[19] N. Ghaffar Nia, R. Manwar, and K. Avanaki, "Overview of State-of-the-Art Learning-Based Classification Methods in Medical Imaging," *Annals of Biomedical Engineering*, Jun. 2026, doi: 10.1007/s10439-026-04241-3.
[20] Z. Gan, Z. Xie, Y. Zhang, W. Han, B. Zhang, and X. Chai, "WTCM-UNet: A hybrid CNN-SSM framework combining wavelet transform for medical image segmentation," *Biomedical Signal Processing and Control*, vol. 112, p. 108525, Feb. 2026, doi: 10.1016/j.bspc.2025.108525..
[21] I. Ahmed, M. Ahmad, A. Chehri, and G. Jeon, "From data to diagnosis: AI-driven multi-modal fusion and generative AI-enhanced GAN-based MRI for brain tumour detection," *Information Fusion*, vol. 126, p. 103527, Feb. 2026, doi: 10.1016/j.inffus.2025.103527..
[22] J. Patel, H. Muralikrishna, K. Chadaga, A. Thalengala, and N. Sampathila, "Sickle cell disease detection in low-resource conditions using transfer-learning and contrastive-learning coupled with XAI," *Scientific Reports*, vol. 16, no. 1, Jan. 2026, doi: 10.1038/s41598-026-35831-9..
[23] R. Deepa, M. P. Mathew, S. Baskar, and A. K. M, "L-Net: a lightweight CNN framework for sustainable multicrop leaf disease detection and classification on edge devices," *Sustainable Food Technology*, vol. 4, no. 1, pp. 985–1003, 2026, doi: 10.1039/d5fb00191a.
[24] S. U. Rehman Khan, S. Asif, M. Zhao, W. Zou, Y. Li, and C. Xiao, "ShallowMRI: A novel lightweight CNN with novel attention mechanism for Multi brain tumor classification in MRI images," *Biomedical Signal Processing and Control*, vol. 111, p. 108425, Jan. 2026, doi: 10.1016/j.bspc.2025.108425.
[25] N. Huda and K. R. Ku-Mahamud, "CNN-Based Image Segmentation Approach in Brain Tumor Classification: A Review," *The 8th Mechanical Engineering, Science and Technology International Conference*, p. 66, Feb. 17, 2025. doi: 10.3390/engproc2025084066..
[26] S. Repetto, I. Maljkovic, M. Lotto, A. E. Cinà, S. Vascon, and F. Roli, "Evaluating the robustness of explainable AI in medical image recognition under natural and adversarial data corruption," *Machine Learning*, vol. 115, no. 1, Dec. 2025, doi: 10.1007/s10994-025-06919-6.
[27] N. Chakrabarty, "Brain MRI images for brain tumor detection." Accessed: Jul. 04, 2026. [Online]. Available: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

BIOGRAPHIES OF AUTHORS



Dipta Paul received the B.Sc. degree in Electrical and Electronic Engineering from American International University-Bangladesh (AIUB), Dhaka, Bangladesh. His research interests include artificial intelligence, machine learning, and deep learning, particularly for healthcare, medical image analysis, and disease prediction. His interdisciplinary research also spans explainable AI, IoT-enabled healthcare, biomedical robotics, intelligent control, embedded systems, renewable energy, and data-driven modeling, focusing on reliable and practical solutions to real-world challenges. He is a member of IEEE and has served as a reviewer for international conferences associated with IEEE and Springer, as well as for journals published by Elsevier, Frontiers, and other Scopus-indexed journals. He aims to integrate AI, biomedical engineering, and intelligent systems to address practical healthcare and engineering challenges. He can be contacted at email: diptapal4@gmail.com.



Ekramul Hasan is pursuing an M.Sc. in Engineering Management at Westcliff University. He received the B.Sc. degree in Electrical and Electronic Engineering from American International University-Bangladesh (AIUB). His research focuses on artificial intelligence, healthcare, robotics, machine learning, medical diagnosis, health informatics, and intelligent decision support systems. He has published several research articles in international journals and conference proceedings, with a particular interest in AI-driven medical technologies that improve healthcare and quality of life. He can be contacted at email: ekramhasan55@gmail.com.



Md Al Amin received the B.Tech. degree in Computer Science and Engineering from Lovely Professional University, India, and the M.B.A. degree in Management Information Systems from International American University, Los Angeles. His research focuses on artificial intelligence, machine learning, deep learning, and healthcare, particularly in medical AI, predictive modeling, and data-driven approaches for disease detection and intelligent healthcare systems. He can be contacted at email: alamin99.edu@gmail.com.



Md Tuhin Mia received the Master of Laws (LLM) degree from Jagannath University, Dhaka, Bangladesh, and the MBA degree in Business Analytics from International American University, Los Angeles. He is currently a Proficient Computational Analytics Researcher with a strong academic background. His research interests include medical artificial intelligence, multimodal deep learning, lightweight models for mobile healthcare, and explainable AI. He can be contacted at email: ttuhinmia82@gmail.com.



Md Musa Haque received the B.Eng. degree in Mechatronics Engineering from Hebei University of Technology, China and the M.B.A. degree in Management Information Systems from International American University, Los Angeles, USA. He is currently a dedicated researcher in the field of artificial intelligence and healthcare. His research explores innovative, data-driven approaches aimed at advancing early disease detection, improving patient care, and transforming healthcare outcomes. He can be contacted at email: mmhxedu@gmail.com.