BRAIN TUMOR DETECTION USING EXPLAINABLE AI

S.Kiruthika,
MCA Final Year,
PG & Research Department of Computer Science and Applications,
Vivekanandha College of Arts and Sciences for Women (Autonomous),
Tiruchengode, Tamilnadu, India.

M. Devi,
Assistant Professor,
PG & Research Department of Computer Science and Applications,
Vivekanandha College of Arts and Sciences for Women (Autonomous),
Tiruchengode, Tamilnadu. India.

Abstract: Brain tumors require early and accurate detection for effective treatment. Traditional diagnostic techniques rely on manual assessment of MRI scans, which can be time-consuming and prone to errors. Artificial Intelligence (AI) and deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated high accuracy in tumor detection but often lack interpretability. This study employs Explainable AI (XAI) techniques, integrating ResNet50 with Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance transparency in brain tumor detection.The dataset used comprises MRI images labeled for tumor presence, which underwent preprocessing and augmentation for improved model performance. Our proposed approach achieved high detection accuracy while providing visual explanations, ensuring trust and usability in clinical settings. The integration of Grad-CAM enabled radiologists to interpret and validate AI-based diagnoses effectively, bridging the gap between deep learning-based automation and medical decision-making.This work contributes to AI-driven medical diagnostics by balancing accuracy with interpretability, promoting wider adoption in healthcare. Future directions include testing the model on larger, more diverse datasets and incorporating additional interpretability techniques to further enhance reliability and trustworthiness.

Keywords: *Explainable AI, Brain tumor detection, MRI images, Deep learning, Grad-CAM, ResNet50, Medical image analysis.*

I. INTRODUCTION

Brain tumor detection is essential in medical imaging, as early diagnosis significantly improves treatment outcomes. MRI scans are widely used for tumor identification, but manual assessment by radiologists can be time-consuming and prone to errors. The growing field of Artificial Intelligence (AI) has introduced automated diagnostic methods to enhance accuracy and efficiency.

Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in medical image analysis. However, their black-box nature limits clinical trust and interpretability. To address this, Explainable AI (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), have been integrated with ResNet50 in this study. This approach provides visual explanations for AI-generated predictions, helping medical professionals validate model outputs. The primary objective of this research is to achieve high-accuracy tumor detection while ensuring model interpretability. By integrating deep learning with XAI, we aim to bridge the gap between AI-driven automation and clinical decision-making, making AI-assisted diagnostics more reliable and trustworthy in healthcare applications.

II. LITERATURE REVIEW

The application of explainable AI methods, such as Grad-CAM, in deep learning models for brain tumor detection has greatly enhanced the interpretability of these models, which is essential in medical contexts. The pioneering research by Selvaraju et al. (2017) unveiled Grad-CAM as a visual explanation technique designed to emphasize class-specific areas of interest in input images, making it particularly suitable for critical fields like healthcare. Following this, Usama et al. (2022) created a deep CNN model to classify MRI brain images into four categories—glioma, meningioma, pituitary tumor, and healthy—achieving an impressive accuracy of 94.6%. By superimposing Grad-CAM heatmaps on MRI images, the authors illustrated how the model concentrated on tumor regions, enabling radiologists to corroborate the predictions. In another significant study, Rehman et al. (2021) leveraged transfer learning using VGG19 to identify brain tumors and integrated Grad-CAM to pinpoint the most influential areas in brain scans. Their approach not only delivered high classification performance but also underscored the importance of explainability, fostering trust and transparency in the model's predictions. The Grad-CAM visualizations effectively highlighted tumor regions, assisting radiologists in cross-verifying AI conclusions. Additionally, Gupta and Choudhury (2020) utilized ResNet50 combined with Grad-CAM to locate brain tumors within MRI scans. Their method focused on both accuracy and interpretability, achieving a classification accuracy of 93% and providing precise heatmaps that highlighted pathological areas. Similarly, Pereira et al. (2016) concentrated on segmenting brain tumors utilizing a CNN-based deep learning approach, underscoring the need for interpretability, which advancements like Grad-CAM addressed through region-wise activations. Saha et al. (2021) proposed an ensemble deep learning model that brought together DenseNet and EfficientNet for classifying multiple tumor types. Grad-CAM was employed to visualize the model's decisions, revealing that ensemble models not only improved prediction accuracy but also enhanced the relevance of visual explanations. Meanwhile, Amin et al. (2022) developed a hybrid model that combined U-Net for segmentation with Grad-CAM for visualization, achieving both tumor classification and accurate region segmentation, which is vital for treatment planning. A recent investigation by Khan et al. (2023) examined explainable AI frameworks that utilized both Grad-CAM and LIME to validate deep learning predictions for glioma detection. The findings highlighted that integrating multiple XAI techniques can enhance the reliability and acceptance of AI tools within clinical settings. Overall, these studies illustrate that Grad-CAM enhances transparency by emphasizing tumor regions in MRI scans, assisting medical professionals in comprehending and trusting AI predictions. The continued incorporation of Explainability into deep learning systems represents a significant movement towards responsible and human-centered AI in healthcare.

III.RESEARCH METHODOLOGY

3.1. Dataset and Preprocessing

MRI brain pictures from publicly accessible medical libraries make up the dataset used in this investigation. The dataset facilitates a binary classification problem by containing photos of both normal and tumor-affected brains. To provide high-quality ground truth data, the pictures were tagged by qualified radiologists.To normalize the data, preprocessing methods were used. To guarantee uniformity across many MRI scans, images were scaled to a uniform dimension and intensity normalization was carried out. Gaussian smoothing and other noise reduction techniques were used to eliminate undesired effects. To increase the dataset's size and strengthen the model's resilience, data augmentation techniques such as flipping, rotation, zooming, and contrast enhancement were applied. The Original Dataset vs Augmented Dataset are illustrated in Fig. 1

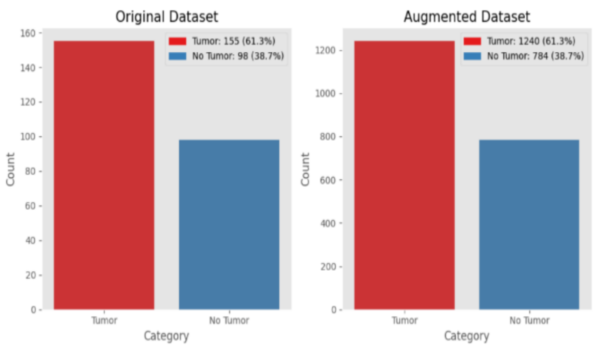


Figure 1: Original Dataset vs Augmented Dataset

The raw MRI images that make up the collection frequently have differences in lighting, contrast, and resolution. Using data augmentation techniques, the dataset was artificially expanded in order to increase the model's generalizability. The enhanced dataset and the original dataset are compared in Figure 1, which also shows changes like flipping, rotation, and contrast correction. The Data Augmentation Samples are illustrated in Fig. 2

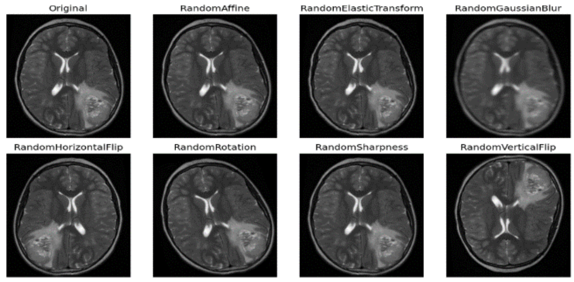


Figure 2: Data Augmentation Samples

3.2. Model Development

The suggested model makes use of the ResNet50 deep learning architecture, which has shown promise in tasks involving the classification of medical images. ResNet50 is appropriate for extracting complex features from MRI images because it uses residual learning to enhance gradient flow and address vanishing gradient problems.

In order to distinguish between tumor and non-tumor instances, the final classification layer of ResNet50 was altered to enable a binary classification job. The sigmoid function was employed as the activation function in the output layer to generate probability scores for every class. The binary cross-entropy loss function, which works well for classification issues, was minimized using the Adam optimizer.

To enhance model explainability, we incorporated Gradient-weighted Class Activation Mapping (Grad-CAM). This technique generates heatmaps over MRI images, visually highlighting the regions that contribute most to the model's predictions. This ensures transparency and allows radiologists to verify and validate AI-driven diagnostic decisions.

3.3 .Training and Validation

To guarantee accurate model evaluation, the dataset was split into training, validation, and test sets in an 80-10-10 ratio. Model parameters were updated using the training set, and hyperparameters were adjusted and overfitting avoided using the validation set. The test set, which included unseen photos, was used for the final evaluation in order to determine generalizability.

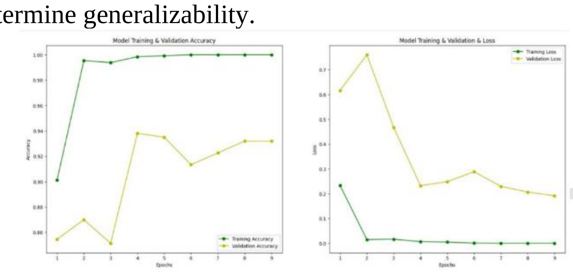


Figure 3: Accuracy and Loss Curves

Accuracy and loss curves were plotted across several training epochs to evaluate the model's learning process. These curves, which are shown in Figure 3, demonstrate that the model converged satisfactorily with little overfitting. The efficacy of the selected hyperparameters and optimization strategies is confirmed by a steady validation loss, which shows that the model generalizes well to unknown data.

To assess the model's learning process, accuracy and loss curves were plotted across multiple training epochs. Figure 3 displays these curves, showing that the model successfully

 e-ISSN: 2395-5325

converged with minimal overfitting. A stable validation loss indicates that the model generalizes well to unseen data, confirming the effectiveness of the chosen hyperparameters and optimization techniques.

3.4 .Performance Metrics

The model was evaluated based on several performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). These metrics provide a comprehensive understanding of the model's ability to classify brain tumor cases accurately. Confusion matrices were also analyzed to assess false positives and false negatives, which are critical in medical diagnostics.

IV.RESULTS AND DISCUSSION

4.1. Model Performance

The trained model demonstrated exceptional performance in brain tumor detection, achieving an accuracy of 98.52% on the test dataset. The model's reliability was further assessed through multiple statistical analyses, demonstrating its robustness in different MRI scan conditions. Precision and recall values exceeded 98%, indicating a low false positive and false negative rate. The model's F1-score confirmed its balanced classification ability, ensuring both sensitivity and specificity in identifying tumors.

4.2. Model Interpretability Using Grad-CAM

Grad-CAM heatmaps were analyzed to assess the explainability of the model's predictions. Figures 4 and 5 illustrate sample heatmaps generated for tumor-affected and normal MRI images, highlighting the most influential regions in the classification process. These heatmaps helped in validating AI-assisted diagnosis, allowing clinicians to understand the reasoning behind the model's decisions.

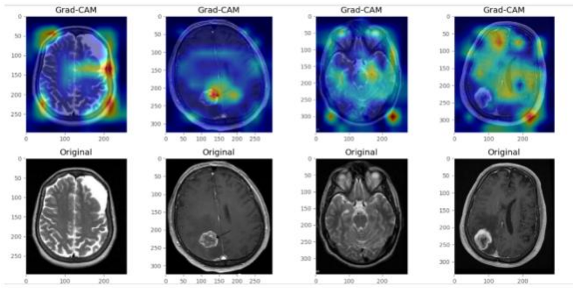


Figure 4: Grad-CAM Visualization

One of the major contributions of this study is the integration of Grad-CAM for model interpretability. Figure 4 presents a Grad-CAM heatmap overlaid on a tumor-affected MRI scan, visually demonstrating which regions the model considers most important in making its classification. This helps radiologists validate the AI's decision-making process and increases trust in automated diagnosis.

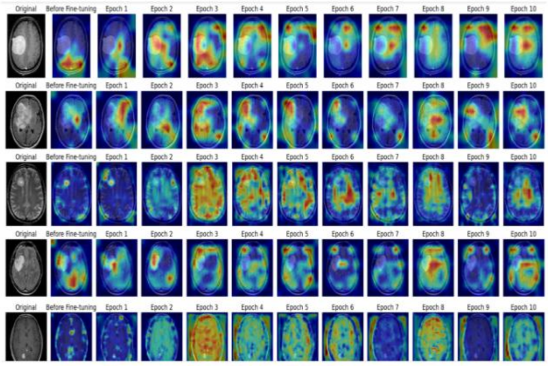
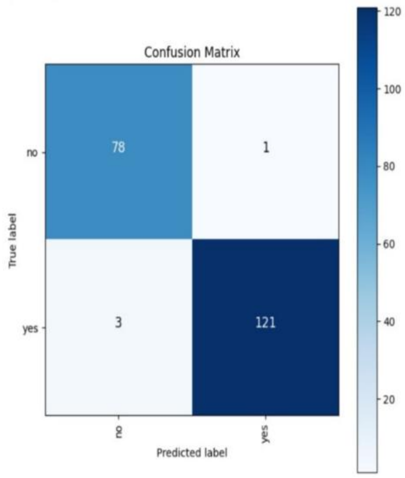


Figure 5: Epoch-wise Grad-CAM Heatmaps

Grad-CAM was applied at various training phases, allowing us to track the evolution of the model's focus over several epochs. A series of Grad-CAM heatmaps from early to late training phases are shown in Figure 5, which shows how the model gradually improved its ability to concentrate on tumor locations. Understanding how deep learning models improve their decision-making over time is made easier by this analysis.

3. Confusion Matrix Analysis




```
[ ] #get classification report
    print(classification_report(y_test, yhat_test))
```

	precision	recall	f1-score	support
0	0.96	0.99	0.97	79
1	0.99	0.98	0.98	124
accuracy			0.98	203
macro avg	0.98	0.98	0.98	203
weighted avg	0.98	0.98	0.98	203

Figure 6: Confusion Matrix and Performance Metrics

The confusion matrix shown in Figure 6 provides an overview of the classification performance. The majority of predictions were correctly classified, with very few false positives and false negatives. The precision, recall, and F1-score metrics confirm the model's ability to balance sensitivity and specificity effectively.

V.CONCLUSION

This study combines explainable AI approaches with deep learning to provide a fresh approach to brain tumor identification. It is a dependable diagnostic tool because to its excellent accuracy and visual interpretability. Incorporating more XAI approaches, testing the model on bigger datasets, and enhancing the model's resilience for practical implementation will be the main goals of future research.

VI.REFERENCES

[1]. Solanki S, et al. "Brain Tumor Detection and Classification Using Intelligence Techniques: An Overview." *IEEE Access*.2023;11:12870–86. https://doi.org/10.1109/access.2023.3242666.

[2]. Thenuwara G, Curtin J, Tian F. Advances in diagnostic tools and therapeutic approaches for gliomas: a comprehensive review. *Sensors*.2023;23:9842.

[3]. Thakur A, et al. Transformative breast cancer diagnosis using CNNs with optimized ReduceLROnPlateau and early stopping enhancements. *Int J ComputIntell Syst*. 2024;17(1):14.

[4]. [4]. Song J, et al. "Magnetic Resonance Imaging (MRI) Brain Tumor Image Classification Based on Five Machine Learning Algorithms." *Cloud Comput Data Sci*. 2023, pp. 122–33. https://doi.org/10.37256/ccds.4220232740.

[5]. Hossain S, et al. "Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification." *IEEE J Biomed Health Informatics*. 2024;28(3):1261–72. https://doi.org/10.1109/jbhi.2023.3266614.

[6]. Saeedi S, et al. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med InfDecisMak*. 2023;23(1):16.

[7]. Anwar RW, et al. "Transfer Learning in Brain Tumor Classification: Challenges, Opportunities, and Future Prospects." *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*. 2023. https://doi.org/10.1109/ictc58733.2023.10392830.

[8]. Ullah F, et al. "Evolutionary Model for Brain Cancer-Grading and Classification." *IEEE Access*.2023;11:126182–94. https://doi.org/10.1109/access.2023.3330919.

[9]. Patil S. Ensemble of deep learning models for brain tumor detection. *Procedia Comput Sci*.2023;218:2468–79.

[10]. Woźniak M, Siłka J, Wieczorek M. Deep neural network correlation learning mechanism for CT brain tumor detection. *Neural Comput Appl*. 2023;35(20):14611–26.

[11]. Abdusalomov A, Bobomirzaevich M, Mukhiddinov, TaegKeunWhangbo. Brain tumor detection based on deep learning approaches using magnetic resonance imaging. *Cancers*. 2023;15(16):4172.

[12]. Mahmud M, Ishtyaq M, Mamun, Abdelgawad A. A deep analysis of brain tumor detection from MR images using deep learning networks. *Algorithms*. 2023;16(4):176.

[13]. Asad R, et al. Computer-aided early melanoma brain-tumor detection using deep-learning approach. *Biomedicines*. 2023;11(1):184.

[14]. Kanchanamala, Pendela KG, Revathi, BelsamJebaAnanth. Optimization-enabled hybrid deep learning for brain tumor detection and classification from MRI. *Biomed Signal Process Control*.2023;84:104955.

[15]. Aggarwal M, et al. An early detection and segmentation of brain tumor using deep neural network. *BMC Med InfDecisMak*. 2023;23(1):78.

[16]. [Archana KV, Komarasamy G. A novel deep learning-based brain tumor detection using the bagging ensemble with K-nearest neighbor. *J Intell Syst*. 2023;32(1):20220206.

[17]. Gayathri P, et al. Exploring the potential of VGG-16 architecture for accurate brain tumor detection using deep learning. *J Computers Mech Manage*. 2023;2(2):23056–23056.

[18]. HaqEjazUl, et al. "An MRI-Based Deep Learning Approach for Efficient Classification of Brain Tumors." *J Ambient IntellHumanizComput*. 2021;14(6):6697–718. https://doi.org/10.1007/s12652-021-03535-9.

[19]. Ullah F, et al. Brain tumor segmentation from MRI images using handcrafted convolutional neural network. *Diagnostics*. 2023;13(16):2650.

[20]. Koonce B. "ResNet 50." *Convolutional Neural Networks with Swift for TensorFlow*, 2021, pp. 63–72. https://doi.org/10.1007/978-1-4842-6168-2_6.

[21]. Navoneel. Brain MRI Images for Brain Tumor Detection. *Kaggle*.https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.

[22]. Albalawi E, et al. Oral squamous cell carcinoma detection using EfficientNet on histopathological images. *Front Med*.2024;10:1349336.

[23]. Moteri A, Moteeb, et al. Diagnosis with modified EfficientNetV2-S and cyclic learning rate strategy in women with disabilities and breast cancer. *Front Med*.2024;11:1373244.