Identification and Prediction of Brain Tumor Using VGG-16 Empowered with Explainable Artificial Intelligence

Fahad Ahmed[1], Muhammad Asif[2,] Muhammad Saleem[3], Ume Farwa Mushtaq[4], Muhammad Imran[5]

[1, 2, 3] Department of Computer Science, National College of Business Administration economics, Lahore, 54000, Pakistan
[4]World Health Organisation (WHO)
[5]BECS- Ministry of Federal Education and Professional Training
[*]Corresponding Author: Fahad Ahmed. Email: fahad.ahmed@ncbae.edu.pk

Abstract: The abnormal development of cells is what causes brain tumors. It is one of the world's leading causes of mortality among adults. Brain tumor detection in a timely manner can prevent millions of deaths. Earlier detection of brain tumors using Magnetic Resonance Imaging (MRI) may increase the patient's chance of survival. MRI is the most prevalent diagnostic technique for brain tumors. The enhanced visibility of tumors on MRI facilitates subsequent treatment. Identification and prediction of brain tumors are essential to their diagnosis and treatment. This article presents a study that utilizes the VGG16 deep learning model to classify brain MRI images obtained from a dataset sourced from Kaggle, comprising two classes: normal and tumor. The dataset is separated into training and testing sets, and the VGG16 model is trained to achieve a testing accuracy of 97.33%. Despite the high accuracy achieved, deep learning models like VGG16 are often perceived as "black boxes," providing predictions without clear explanations. To address this limitation, Layer-wise Relevance Propagation (LRP) is applied to the VGG16 predictions to shed light on the decision-making process and provide interpretability.

Keywords: Brain tumor; Transfer learning; VGG16; Explainable artificial intelligence; Layer-wise relevance propagation

1 Introduction

Numerous types of cells constitute the human body. Every single cell has a distinct purpose. The body's cells proliferate, divide, and produce new cells in an orderly fashion. These newly formed cells aid in maintaining the health and functionality of the human body. When a cell loses the ability to regulate its own growth, it grows disorderly. The mass of tissue produced by the accumulated cells is known as a tumor. A tumor is significantly different from cancer [1].

Brain tumors have enduring and catastrophic mental and physical effects, which can have a significant impact on the patient's quality of life and overall existence. If left untreated, brain tumors can cause death [2]. According to the National Brain Tumor Foundation (NBTF), the number of individuals who have died from brain tumors has increased by 300 % during the past thirty years [3].

Numerous common imaging methods, like computed tomography (CT), X-ray, ultrasonography, and MRI, are employed in medical imaging; however, they are unable to display the complete and detailed aspects and areas of brain tumors; they do, however, improve doctors' estimations of tumor growth [4]. MRI is an extremely effective and prevalent technique for diagnosing brain tumors [5]. MRI is utilized in medical imaging to demonstrate abnormal body tissues. MRI is becoming increasingly popular for diagnosing brain tumors in clinical settings [6]. Doctors may identify the progression of the disease using a series of MRI images at various levels. However, this approach can be time-consuming and potentially lead to missed or incorrect diagnoses.

Deep learning (DL), a subfield of machine learning (ML) [7], has demonstrated remarkable capabilities in various domains, particularly in image recognition and analysis. Its potential to significantly reduce human effort and automate complex tasks has led to its widespread adoption and revolutionized numerous sectors, including healthcare. In the specific context of brain tumor detection using MRI, DL systems have shown great promise in accurately identifying and segmenting tumors, assisting medical professionals in making informed decisions.

However, one of the primary challenges encountered when employing DL for brain tumor detection is the black box problem. Not surprisingly, medical experts have voiced their concern about the black box nature of DL [8]. DL models often operate as complex, highly interconnected networks of artificial neurons, making it difficult to comprehend how they arrive at their predictions. This lack of interpretability hinders the understanding of the underlying decision-making process, posing challenges for clinicians and researchers who require transparency and explanations to trust and validate the DL system's outputs.

To address the black box issue, researchers have turned to the concept of explainable artificial intelligence (XAI). XAI aims to bridge the gap between the inherent complexity of DL models and the need for interpretable decision-making. It encompasses various techniques and methodologies designed to provide insights and explanations regarding the predictions made by DL models. In this article, a brain MRI image dataset is used for the identification and prediction of brain tumor using transfer learning empowered with explainable artificial intelligence.

The remaining sections of the article are as follows: The second section describes related works, the third section explains materials and methods; the fourth section provides simulation and results, and the final section provides a conclusion.

2 Related work

ML and DL techniques have emerged as powerful tools in the medical domain, revolutionizing the field of tumor detection. By leveraging these techniques, researchers have been able to develop sophisticated algorithms capable of analyzing vast amounts of medical data with remarkable precision and speed. The ability of these algorithms to learn from patterns and extract meaningful insights has paved the way for substantial improvements in tumor detection accuracy.

Pitchai et al. [9] combined an artificial neural network (ANN) and a fuzzy k-means algorithm for the separation and detection of brain tumor. The ANN is provided with the GLCM-extracted features for a number of normal and abnormal MRI scans. Charfi et al. [10] suggested a classification technique for MRI scans of brain tumors. In his proposed ML method for image segmentation, he mentioned using histogram equalization. Then, he utilized PCA to decrease the size of the information that was acquired. And ultimately, a neural network with feed forward back propagation was utilized for. He achieved 90% accuracy when classifying images as normal or abnormal.

El Abbadi et al. [11] classified brain tumor data using Singular Value Decomposition (SVD). 50 abnormal and 20 normal data were utilized for testing their methods. They reported a 96.66% accuracy rate, 98% specificity, and a 90% sensitivity. In their study, Gupta et al. [12] conducted the categorization utilizing brain tumor MRI pictures. For classification, they utilized the Discrete Wavelet transform (DWT), PCA, and SVM. They achieved an accuracy rate of 80%, a sensitivity rate of 84%, and a specificity rate of 92%.

Narayana and Reddy [13] segmented and classified brain MRI images using a genetic algorithm, a metaheuristic optimization technique, and a support vector machine (SVM). Approximately 91% accuracy was achieved in identifying abnormal and normal brain tissues using MRI pictures. Using the Discrete Wavelet Transform (DWT), Mohsen et al. [14]

developed a deep neural networks (DNN) for the categorization and detection of brain tumors. The suggested model obtained an accuracy of 93.94 %.

Patil et al. [15] demonstrated a predictive model for brain tumor detection using a DL technique in their research This method uses a suggested convolutional neural network (CNN) model and compares the outcomes to those of the pre-trained CNN model VGG16 on the Kaggle-provided brain tumor dataset. The test accuracy and F1 score obtained by the model suggested are 80% and 0.80, respectively. In addition, the modified VGG16 demonstrated 90% accuracy and a 0.85 F1 score on the same dataset.

By adapting a DL model, Çinar et al. [16] suggested a CNN hybrid design for the detection of brain tumor. They eliminated the final five ResnNet50 layers and added eight new ones. The accuracy of their suggested hybrid ResNet50 model was 97.2%, whereas the accuracy of the single ResNet50 model was 92.53%. Waghmare et al. [17] applied various CNN architectures for the categorization and detection of brain tumors. Modified VGG-16 improved the categorization accuracy of the augmented data set to an acceptable level of 95.71 %.

Table 1: Limitations of the related work

Authors	Dataset Type	Method used	Outcomes	Limitations
Charfi et al. [10]	Brain MRI images	Neural network	Classified brain MRI into normal and abnormal with 90% accuracy.	**(i).** Accuracy can be improved **(ii).** No use of XAI
El Abbadi et al. [11]	Brain MRI images	SVD	Achieved an accuracy of 96.66%.	**(i).** Accuracy can be improved **(ii).** No use of XAI
Gupta et al. [12]	Brain MRI images	DWT, PCA and SVM	Achieved 80% accuracy.	**(i).** Low Accuracy **(ii).** No use of XAI
Narayana and Reddy [13]	Brain MRI images	Metaheuristic optimization technique	Achieved an accuracy about 91%.	**(i).** Accuracy can be improved **(ii).** No use of XAI
Mohsen et al. [14]	Brain MRI images	DNN	Proposed model achieved an accuracy of 93.94%.	**(i).** Accuracy can be improved **(ii).** No use of XAI
Patil et al. [15]	Brain MRI images	DL	90% accuracy achieved by modified VGG 16.	**(i).** Accuracy can be improved **(ii).** No use of XAI
Çinar et al. [16]	Brain MRI images	Hybrid ResNet50 and single ResNet50	Achieved accuracies of 97.2% and 92.53%.	**(i).** Accuracy can be improved **(ii).** No use of XAI
Waghmare et al. [17]	Brain MRI images	DL	95.71% accuracy achieved by modified VGG 16.	**(i).** Accuracy can be improved **(ii).** No use of XAI

Following are the two major gaps in these previous works.

1. Overall accuracy can be improved [10]-[17].

2. No use of XAI to explain the decision making in these previous works [10]-[17].

The following are significant contributions to this article:

1. Employing the proposed model, normal and tumor brain MRI images have been identified and predicted.

2. Accuracy, misclassification rate, precision, specificity, sensitivity, false negative rate (FNR), false positive rate (FPR), and F1 score are some of the evaluation criteria that have demonstrated excellent outcomes.

3. The suggested model shows better overall accuracy as compared to previous literatures.

4. The proposed model explains the predictions made by deep learning model by using LRP and provides insight view on what grounds decision making was made.

3 Materials and methods

Different DL approaches for image evaluation and recognition have become prevalent in image processing in recent years [18]. The proposed model architecture, as depicted in Figure 1, comprises two distinct phases: the training phase and the validation phase. This architecture consists of five key layers: the data acquisition layer, the data pre-processing layer, the deep learning layer, the explainable artificial intelligence (XAI) layer, and the validation layer. During the training phase, the data acquisition layer is responsible for acquiring the raw dataset of brain MRI images from the Kaggle repository. This dataset includes two classes: "normal" and " tumor." The raw dataset of brain MRI images is subsequently pre-processed in the data pre-processing layer, adhering to the requirements of the DL model employed in the architecture. The DL layer shows a crucial part in generating predictions based on the pre-processed data. These predictions are then compared with the pre-processing data within the explainable artificial intelligence layer. The XAI layer utilizes these comparisons to provide explanations for the predictions made by the deep learning model. If the explanations provided by the XAI layer are deemed satisfactory, indicating fair reasoning behind the predictions, the trained model is stored in the cloud. However, if the explanations are deemed inadequate, the model undergoes retraining to improve its performance.

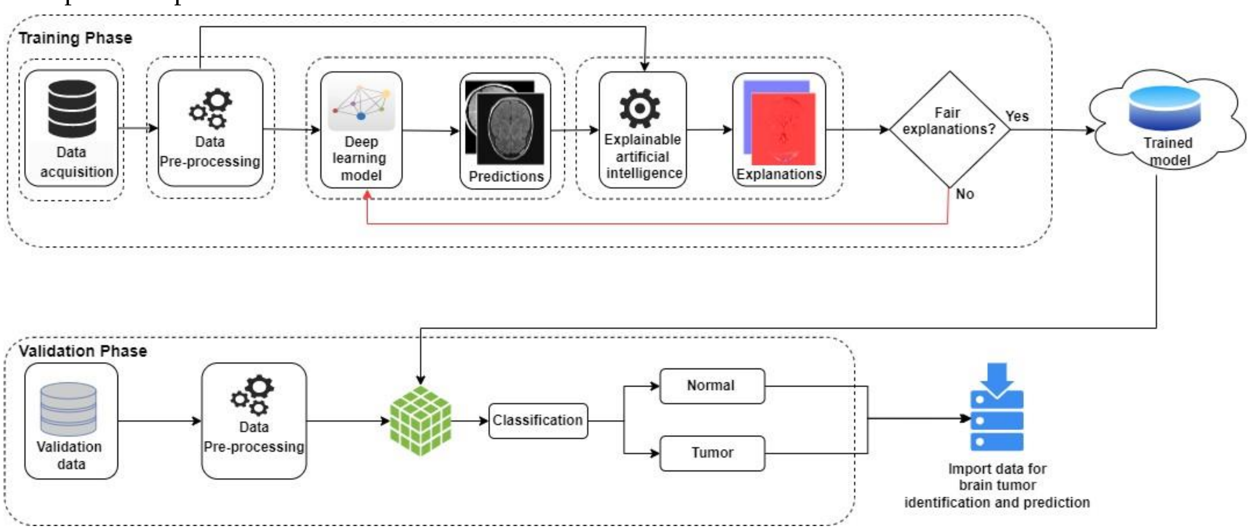

Figure1: Architecture of the proposed methodology

In the validation phase, the validation layer acquires raw data from MRI brain tumors and applies pre-processing techniques to prepare it for evaluation. The trained model, stored in the cloud, is then imported to classify the validation data, consisting of brain MRI scans, into "normal" and "tumor" categories. Lastly, the identified and predicted data related to brain tumors is imported and made available for further analysis and utilization.

3.1 Dataset

Dataset is acquired from kaggle respiratory [19]. Dataset includes 3000 brain MRI images of 2 classes. 1500 normal brain MRI images and 1500 tumor brain MRI images. Figure 2 shows the samples of normal and tumor brain MRI images.

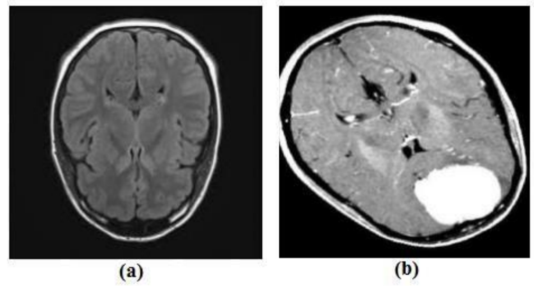

Figure 2: Samples of brain MRI images of (a) Normal; (b) Tumor

3.2 Transfer Learning

Transfer learning (TL) is an effective DL method that facilitates the re-use of previously trained models and their learned approximations for new purposes. It accelerates training, enhances generalization, and improves performance, making it a valuable approach in various domains where data availability and computational resources are limited. TL additionally enables data training with reduced model construction costs [20]. In this article, pre-trained VGG 16 model is utilized.

3.2.1 VGG16

Visual Geometry Group 16 (abbreviated as VGG16) is a CNN design that was presented by the Visual Geometry Group at the University of Oxford in 2014. It is extensively used for several computer vision functions, including image segmentation, classification, and object detection, due to its deep architecture.

Figure 3 shows architecture of the VGG 16 consisting of 1000 classes. Architecture of the VGG 16 is modified according to this article for 2 classes. The input image size for VGG16 is 224 × 224 × 3. 224 x224 represents the length and width while 3 represents the number of channels. The model consists of 138 million parameters [21].

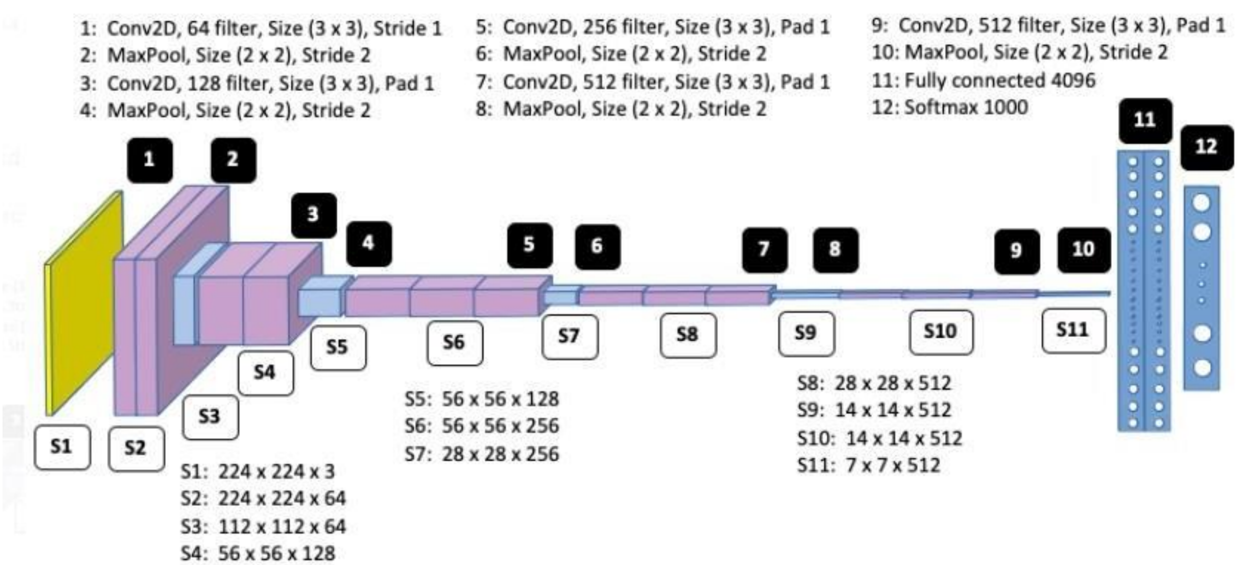

Figure 3: Architecture of VGG16 [22]

3.3 Explainable artificial intelligence

XAI aims to address the inherent "black box" nature of many deep learning models, where it can be challenging to understand how the models arrive at some particular predictions. With explainability, researchers hope to achieve the objectives of XAI methods, which include improved discovery, control, development, and justification. In this article, XAI technique LRP is utilized to address the black box issue.

3.3.1 Layer-wise relevance propagation

LRP enables users to go beyond the "black box" nature of deep learning model and provides a more comprehensive understanding of how the model reaches its conclusions. The primary goal of LRP is to explain how the DL model predictions reaches its output by propagating the relevance or importance of the output back to the input layer. This backward propagation process helps identify which specific features or neurons in the network have the most significant influence on the final prediction.

The LRP architecture was discovered to be effective at delivering significant sense and quantifiable quantities characterizing the feature processing and decision making of a network [23], [24].The benefits of LRP include increased transparency, interpretability, and trustworthiness of deep learning model.

4 Simulation and Results

Google Coolab and Pytorch are utilized for simulation and results. The dataset is separated into training and testing sets. 80% of the dataset is separated into training while 20% of the dataset is separated into testing. The evaluation metrics utilized to assess the effectiveness of the suggested approach are presented in Equations 1-8 [25].

$$\text{Accuracy}=\frac{\text{Tp+Tn}}{\text{Tp+Fp+Tn+Fn}} * 100 \quad (1)$$

$$\text{Misclassification rate}=\frac{\text{Fp+Fn}}{\text{Tp+Fp+Tn+Fn}} * 100 \quad (2)$$

$$\text{Precision}=\frac{\text{Tp}}{\text{Tp+Fp}} * 100 \quad (3)$$

$$\text{Specificity}=\frac{\text{Tn}}{\text{Tn+Fp}} * 100 \quad (4)$$

$$\text{Sensitivity}=\frac{\text{Tp}}{\text{Tp+Fn}} * 100 \quad (5)$$

$$\text{FNR}=\frac{\text{Fn}}{\text{Fn+Tp}} * 100 \quad (6)$$

$$\text{FPR}=\frac{\text{Fp}}{\text{Fp+Tn}} * 100 \quad (7)$$

$$\text{F1 score}=\frac{2*(\text{Precision}*\text{Sensitivity})}{\text{Precision+Sensitivity}} \quad (8)$$

Where Tp stands for true positive, Tn for true negative, Fp for false positive, and Fn for false negative.

The suggested model classifies brain MRI images into normal and tumor. The training parameters for the suggested model, including the number of epochs, optimization algorithm, input image size, batch size, and learning rate, are detailed in Table 2.

Table 2: Training parameters for the proposed model

Training parameters	Value
No. of epochs	10
Optimization algorithm	Adam
Input image size	224x224x3
Batch size	16
Learning rate	0.0001

Figure 4 shows the proposed model's training confusion matrix. In case of normal, 1198 images are correctly classified as normal while 2 images are misclassified as tumor. In case of tumor, 1199 images are correctly classified as tumor while 1 image is misclassified as normal.

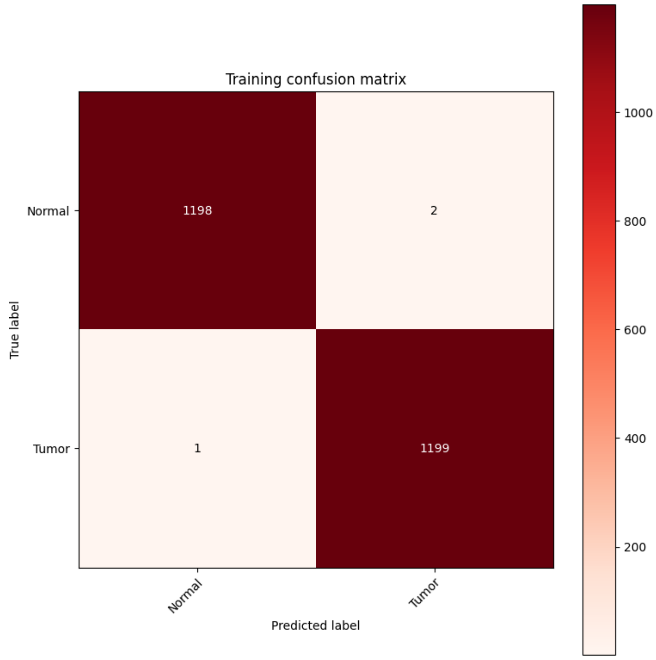

Figure 4: The proposed model's training confusion matrix

Figure 5 shows the proposed model's testing confusion matrix. In case of normal, 297 images are correctly classified as normal while 3 images are misclassified as tumor. In case of tumor, 287 images are correctly classified as tumor while 13 images are misclassified as normal.

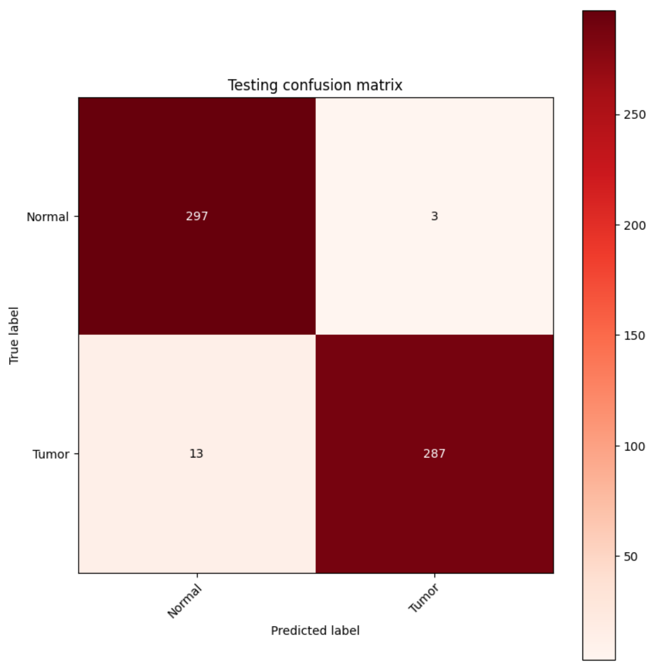

Figure 5: The proposed model's testing confusion matrix

Table 3 displays the performance parameters of the proposed model for both training and testing.

Table 3: Proposed model's performance parameters

Performance Parameters	Training	Testing
Accuracy	99.88%	97.33%
Misclassification rate	0.12%	2.67%
Precision	99.83%	99%
Specificity	99.83%	98.97%
Sensitivity	99.92%	95.81%
FNR	0.08%	4.19%
FPR	0.17%	1.03%
F1 score	0.998	0.974

Figure 6 shows VGG 16 predictions empowered with LRP. The application of LRP on the predictions generated by the VGG 16 model provides insight into the basis of decision-making for distinguishing between normal and tumor brain MRI images. Upon analyzing the LRP-generated results, it is evident that the normal brain MRI image exhibits no indications of a tumor. In contrast, the tumor brain MRI image highlights specific regions that correspond to the presence of a tumor. This observation serves as evidence for the efficacy of LRP in identifying and localizing tumor regions within brain MRI images, enabling informed decision-making regarding the presence or absence of a tumor.

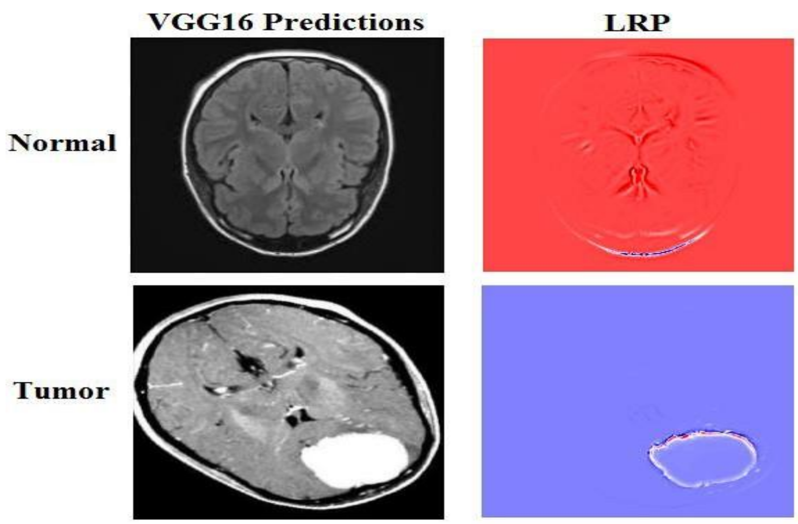

Figure 6: VGG 16 predictions empowered with LRP

Table 4 displays the comparison of the proposed model with previous related work. It is evident from the table that the proposed model has better overall accuracy as compared to previous researches. XAI also explains the decision making of the proposed model.

Table 4: Comparison of the proposed model with related work

Authors	Year	Method used	Accuracy (%)	Misclassification rate (%)	Explainable artificial intelligence
Charfi et al. [10]	2014	Neural network	90	10	No
El Abbadi et al. [11]	2016	SVD	96.66	3.34	No
Gupta et al. [12]	2017	DWT, PCA and SVM	80	20	No

Narayana and Reddy [13]	2018	Metaheuristic optimization technique	91	9	No
Mohsen et al. [14]	2018	DNN	93.94	6.06	No
Patil et al. [15]	2020	DL	90	10	No
Çinar et al. [16]	2020	Hybrid ResNet50 Single ResNet50	97.2 92.53	2.8 7.47	No
Waghmare et al. [17]	2021	DL	95.71	4.29	No
Proposed model	2023	VGG16 empowered with LRP	97.33	2.67	Yes

5 Conclusion

This study effectively trained a VGG16 model utilizing a dataset of brain images containing both normal and tumor images, obtaining remarkable training accuracy (99.88%) as well as high testing accuracy (97.33%). The model was able to precisely identify and predict brain images. The implementation of Layer-wise Relevance Propagation (LRP) provided valuable insights into the decision-making process of the model. The combination of the VGG16 model and LRP offers a promising identification and interpretation strategy for brain tumors.

This study demonstrates the significance of explainable artificial intelligence in medical applications and paves the way for future research on leveraging explainable techniques to improve the performance and reliability of deep neural networks in healthcare scenarios.

References

[1] A. Mustaqeem, A. Javed, and T. Fatima, “An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation,” *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 10, pp. 34–39, 2012, doi: 10.5815/ijigsp.2012.10.05.

[2] Y. Meng, C. Tang, J. Yu, S. Meng, and W. Zhang, “Exposure to lead increases the risk of meningioma and brain cancer: A meta-analysis,” *Journal of Trace Elements in Medicine and Biology*, vol. 60, no. December 2019, p. 126474, 2020, doi: 10.1016/j.jtemb.2020.126474.

[3] E. A. S. El-Dahshan, H. M. Mohsen, K. Revett, and A. B. M. Salem, “Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm,” *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014, doi: 10.1016/j.eswa.2014.01.021.

[4] J. Liu *et al.*, “A Survey of MRI-Based Brain Tumor Segmentation Methods,” *TSINGHUASCIENCEANDTECHNOLOGY*, vol. 19, no. 6, pp. 578–595, 2014.

[5] M. M. Badža and M. C. Barjaktarović, “Classification of brain tumors from mri images using a convolutional neural network,” *Applied Sciences (Switzerland)*, vol. 10, no. 6, 2020, doi: 10.3390/app10061999.

[6] P. Y. Wen *et al.*, “Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group,” *Journal of Clinical Oncology*, vol. 28, no. 11, pp. 1963–1972, 2010, doi: 10.1200/JCO.2009.26.3541.

[7] C. Zuo *et al.*, “Deep learning in optical metrology: a review,” *Light: Science and Applications*, vol. 11, no. 39, pp. 1–54, 2022, doi: 10.1038/s41377-022-00714-x.

[8] X. Jia, L. Ren, and J. Cai, “Clinical implementation of AI technologies will require interpretable AI models,” *Medical Physics*, vol. 47, no. 1, pp. 1–4, 2020, doi: 10.1002/mp.13891.

[9] R. Pitchai, P. Supraja, A. H. Victoria, and M. Madhavi, “Brain Tumor Segmentation Using Deep Learning and Fuzzy K-Means Clustering for Magnetic Resonance Images,” *Neural Processing Letters*, vol. 53, no. 4, pp. 2519–2532, 2021, doi: 10.1007/s11063-020-10326-4.

[10] S. Charfi, R. Lahmyed, and L. Rangarajan, “A NOVEL APPROACH FOR BRAIN TUMOR DETECTION

USING NEURAL NETWORK," *International Journal of Research in Engineering & Technology*, vol. 2, no. 7, pp. 93–104, 2014, [Online]. Available: www.impactjournals.us

[11] N. K. El Abbadi and N. E. Kadhim, "Brain Tumor Classification Based on Singular Value Decomposition," *Ijarcce*, vol. 5, no. 8, pp. 553–557, 2016, doi: 10.17148/ijarcce.2016.58116.

[12] T. Gupta, T. K. Gandhi, R. K. Gupta, and B. K. Panigrahi, "Classification of patients with tumor using MR FLAIR images," *Pattern Recognition Letters*, vol. 139, pp. 112–117, 2017, doi: 10.1016/j.patrec.2017.10.037.

[13] T. Lakshmi Narayana and T. Sreenivasulu Reddy, "An Efficient optimization technique to detect brain tumor from MRI images," *Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018*, no. Icssit, pp. 168–171, 2018, doi: 10.1109/ICSSIT.2018.8748288.

[14] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018, doi: 10.1016/j.fcij.2017.12.001.

[15] S. Patil, D. K. Kirange, and V. Nemade, "Predictive Modelling of Brain Tumor Detection Using Deep Learning," *Journal of Critical Reviews*, vol. 7, no. 04, pp. 1805–1813, 2020.

[16] A. Çinar and M. Yildirim, "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture," *Medical Hypotheses*, vol. 139, no. March, p. 109684, 2020, doi: 10.1016/j.mehy.2020.109684.

[17] V. K. Waghmare and M. H. Kolekar, "Brain Tumor Classification Using Deep Learning," *Internet ofThings for Healthcare Technologies*, vol. 73, pp. 155–175, 2021, doi: 10.5812/iranjradiol.99160.

[18] F. Ahmed, W. A. Khan, M. Iqbal, A. riad A. Abazeed, H. Alrababah, and M. F. Khan, "Rock-Paper-Scissors Image Classification Using Transfer Learning," in *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, Dubai, United Arab Emirates: IEEE, 2023, pp. 1–6. doi: 10.1109/ICBATS57792.2023.10111433.

[19] "Br35H :: Brain Tumor Detection 2020 | Kaggle." https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection (accessed Jun. 25, 2023).

[20] N. Kumar, M. Gupta, D. Gupta, and S. Tiwari, "Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 1, pp. 469–478, 2023, doi: 10.1007/s12652-021-03306-6.

[21] D. M. S. Arsa and A. A. N. H. Susila, "VGG16 in Batik Classification based on Random Forest," *Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech 2019*, vol. 1, no. August, pp. 295–299, 2019, doi: 10.1109/ICIMTech.2019.8843844.

[22] "Overview of VGG16 | Mastering Computer Vision with TensorFlow 2.x." https://subscription.packtpub.com/book/data/9781838827069/7/ch07lvl1sec29/overview-of-vgg16 (accessed Jun. 25, 2023).

[23] Y. Yang, V. Tresp, M. Wunderle, and P. A. Fasching, "Explaining therapy predictions with layer-wise relevance propagation in neural networks," in *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, New York, NY, USA: IEEE, 2018, pp. 152–162. doi: 10.1109/ICHI.2018.00025.

[24] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K. R. Müller, "Unmasking Clever Hans predictors and assessing what machines really learn," *Nature Communications*, vol. 10, no. 1, pp. 1–8, 2019, doi: 10.1038/s41467-019-08987-4.

[25] N. Seliya, T. M. Khoshgoftaar, and J. Van Hulse, "A study on the relationships of classifier performance metrics," in *21st IEEE International Conference on Tools with Artificial Intelligence*, Newark, NJ, USA, 2009, pp. 59–66. doi: 10.1109/ICTAI.2009.25.