Received: 18 May 2025 | Revised: 29 April 2026 | Accepted: 7 July 2026 | Published online: 14 August 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–16
DOI: 10.47852/bonviewAIA62026182

A Novel Explainable Deep Learning Model for Early-Stage Brain Tumor Classification: Multi-Level Feature Fusion from Merged MRI Datasets

Md Sadi Al Huda[1,*], Kazi Tanvir[2], Zubaida Akhter[3], Md. Shahidul Khan Pappo[2], Md. Asraf Ali[1] and Nasim Ahmed[4]

[1] *Department of Computer Science & Engineering, Khwaja Yunus Ali University, Bangladesh*

[2] *Department of Computer Science, American International University-Bangladesh, Bangladesh*

[3] *Department of Science, Engineering and Environment, University of Salford, UK*

[4] *School of IT & Engineering, Melbourne Institute of Technology (MIT), Australia*

Abstract: Early and accurate detection of brain tumors remains a major challenge in medical imaging due to limited dataset size, trained deep learning models, patient variability, and the complexity of manual interpretation; traditional approaches sometimes rely on lower-level feature extraction, which may not apply well to medical images. To address these challenges, we propose a unique explainable deep learning model that integrates Tiny-ConvNeXt and DenseNet169 using multi-stage brain tumor classification. Two clinically approved, open-access Magnetic Resonance Imaging (MRI) datasets were combined to generate a bigger, more customized dataset, and substantial data augmentation techniques were used to improve model generalization. According to experimental results, our proposed model outperforms current state-of-the-art methods with a classification accuracy of 99.74%. Additionally, statistical significance tests confirmed the incremental contribution of each model component, and ablation studies showed the robustness of the conclusions. Additionally, explainable artificial intelligence techniques like Local Interpretable Model-agnostic Explanations and Gradient-weighted Class Activation Mapping++ were employed to enhance interpretability, enabling visual explanations of tumor localization. These results indicate that the suggested model provides a transparent and reliable framework for brain tumor detection, with potential applications in practical clinical decision support systems.

Keywords: brain tumor, MRI images, deep learning, ConvNeXt-Tiny, DenseNet169

1. Introduction

The brain is an incredibly complex organ composed of billions of cells. Brain tumors arise when certain cells begin to divide uncontrollably, forming abnormal clusters. If not addressed in time, these abnormal growths can harm healthy brain tissue and impair normal brain functions [1]. As reported by the World Health Organization in 2018, roughly 9.6 million individuals globally succumbed to diagnosed cancer. Between 30% and 50% of individuals diagnosed with primary cancer succumbed to the disease [2]. Among the various types of cancer, brain cancer stands out as particularly lethal. In 2019, statistics indicate that about 17,760 adults died due to brain tumors. Early detection of cancer with quick and cost-effective diagnostic methods can greatly decrease mortality rates. Different types of tumors necessitate varying treatment approaches, and accurate diagnosis is crucial to avoid fatal consequences [3]. In the United States, the prevalence of brain tumors surpasses that of many other nations. In 2019 alone, approximately 86,970 cases of benign and malignant brain tumors were identified through diagnosis. [4]. In recent studies by Mehndiratta et al. [5], radiologists employ various experimental techniques to diagnose brain tumors, including biopsy, cerebrospinal fluid (CSF) analysis, and X-ray examination. A small tissue sample is surgically removed for examination during a biopsy. This procedure includes significant risks, such as inflammation and excessive bleeding, and achieves an accuracy rate of only 49.1% [6]. On the other hand, the CSF is a clear fluid found within the brain. Radiologists [7] utilize CSF testing for detecting brain tumors. However, like biopsy procedures, this method carries inherent risks such as bleeding from the incision site into the bloodstream, as well as the potential for allergic

***Corresponding author:** Md Sadi Al Huda, Department of Computer Science & Engineering, Khwaja Yunus Ali University, Bangladesh. Email: sadi.cse@kyau.edu.bd

reactions following treatment [8]. Accurate detection requires special skills, and complicated cases present additional challenges. Therefore, in addition to human confirmation, the implementation of a computerized process is required for accurate detection. Nowadays, modern imaging techniques are increasingly favored by radiologists due to their higher accuracy and lower risk for patients [9].

Deep learning (DL) makes the identification task automated and meticulous, greatly speeding up the global diagnosis procedure. DL has significantly advanced various applications [10], especially in medical image processing. Automating brain tumor identification (BTI) is a crucial focus of ongoing research. Asif et al. [11] recently investigated the early classification of brain tumors utilizing Magnetic Resonance Imaging (MRI) and DL methodologies, specifically Xception and NasNet Large. Nonetheless, the Inception-ResNet-v2 demonstrated reduced sensitivity, suggesting that it may not be universally applicable to all brain tumors. Dawood et al. [12] proposed a Computed Tomography (CT)-based detection method utilizing MobileNet-V2, achieving an accuracy of 96%. They did not incorporate an attention algorithm to improve performance.

Krisnapriya and Karuna [13] researched pre-trained deep convolutional neural network (DCNN) models for the classification of brain Magnetic Resonance (MR) images, identifying difficulties in the manual assessment of MRI data. However, it lacks comprehensive examination of transfer learning's effectiveness with smaller datasets, despite the elevated accuracy attained by models such as VGG-19 and VGG-16. More research is required to improve performance with limited data and evaluate alternative models, including Inception V3. Most research studies, including Anjum et al. [14] and Alanazi et al. [15], employed transfer learning for brain tumor categorization. However, opportunities exist to surmount these challenges, improve testing precision, and diminish the time complexity of current machine learning (ML) and DL models.

The aim of our project is to create an effective BTI system using a tailored dataset for the early detection of brain tumors. This research proposes a multi-level feature extraction model that concatenates two pre-trained DL models for the early identification of brain tumors.

While previous research has indicated good performance utilizing transfer learning and pre-trained convolutional neural network (CNN) such as VGG, Inception, and ResNet, these methods frequently have disadvantages, such as reliance on limited datasets, lack of interpretability, and inadequate validation across a variety of MRI scans. On the other hand, by concatenating ConvNeXt-Tiny and DenseNet169, our proposed model leverages their complementary features to present a multi-level feature fusion architecture. While DenseNet169 uses superior gradient flow to capture richer semantic features, ConvNeXt-Tiny extracts low-level features in a lightweight and efficient manner. Furthermore, we use explainable artificial intelligence (XAI) techniques to facilitate interpretability and confidence, which is crucial for medical applications.

This research's primary contributions are as follows:

1) We have proposed an improved BTI model using a merged dataset, which offers an effective solution and enhances the ability to be generalized for precise early-stage identification.
2) Our approach combines ConvNeXt-Tiny and DenseNet169 layers for multi-level feature extraction and identification, surpassing existing state-of-the-art models.
3) We used XAI by utilizing Local Interpretable Model-agnostic Explanations (LIME) and Gradient-weighted Class Activation Mapping (Grad-CAM) to elucidate the decisions of our proposed model.

The remaining part of the paper has been organized as follows: Section 2 presents relevant work, Section 3 outlines the methodology, Section 4 details the results and findings, and Section 5 continues with a summary of future work.

2. Literature Review

Models developed using DL and ML have the capacity to address this issue. A wide range of relevant works have been thoroughly researched in this area.

2.1. Machine learning techniques

ML has become a vital component in modern medical diagnosis, particularly for identifying and classifying brain tumors through MRI scans. Manogaran et al. [16] presented an enhanced method utilizing orthogonal gamma distribution-based ML to tackle under-segmentation and over-segmentation issues in brain tumor detection. However, that may face limitations regarding the dataset's quality, handling complex tumors, requiring further validation on diverse datasets, and computational complexity for real-time clinical diagnosis.

Similarly, Khaliki et al. [17] utilized CNN and transfer learning methods like Inception-V3 and EfficientNet-B4. Their studies achieved high accuracy rates: VGG-19 at 96%, EfficientNet-B4 at 97%, Inception-V3 at 96%, and VGG-16 at 98%. This study classifies glioma, meningioma, and pituitary tumors using brain MRI, highlighting the effectiveness of these methods.

Anantharajan et al. [18] proposed the Ensemble Deep Neural Support Vector Machine (EDN-SVM) method, which achieves an accuracy of 97.93% in differentiating abnormal from normal brain tissues in MRI images. Medical imaging analysis has advanced significantly, as evidenced by the reported sensitivity and specificity values of 92% and 98%, respectively.

Saeedi et al. [19] introduced DL and ML methodologies for diagnostic purposes utilizing MRI data. The 2D CNN attained a training accuracy of 96.47%. K-nearest neighbor (KNN) exhibits the highest accuracy at 86% among ML methods, while Multilayer Perceptron (MLP) shows the lowest accuracy at 28%. This contrast highlights the potential of advanced computational methods to enhance diagnostic outcomes for brain tumors.

Mushtaq et al. [20] examined ML methods for brain tumor detection using MRI, highlighting the critical importance of imaging in diagnosis. A comparative analysis was conducted between ANN and CNN models, demonstrating the superior effectiveness of CNN. Recent studies have demonstrated that advanced DL frameworks can significantly improve the accuracy of brain tumor detection from MRI images.

2.2. Deep learning techniques

Medical image analysis utilizing DL, a subset of ML, has shown outstanding outcomes, especially in identifying brain tumors using MRI data. Archana and Komarasamy [21] proposed a novel approach that combines U-Net for segmentation with bagging ensemble and k-nearest neighbor (BKNN) for classification, achieving a classification accuracy of up to 97.7%. The method is effective; however, it may encounter scalability challenges with large datasets, indicating opportunities for enhancement via supervised techniques such as CNNs.

Noreen et al. [22] proposed a method for the early diagnosis of brain tumors through multi-level feature extraction and concatenation utilizing pre-trained DL models, specifically Inception-V3 and DenseNet201. The method, tested on a publicly available brain tumor dataset, achieved a testing accuracy of 99.34%, which is higher than existing methods. This study emphasizes the efficacy of DL models, specifically DenseNet201, in accurately detecting brain tumors, and recommends further investigation into fine-tuning techniques and ensemble methods for enhanced performance. Abdusalomov et al. [23] proposed a refined YOLOv7 model for brain tumor detection in MRI scans, achieving an accuracy of 99.5% through transfer learning and data augmentation. The addition of the Convolutional Block Attention Module and Spatial Pyramid Pooling Fast + layers improves the model's ability to extract features. We need to do research to make sure this method really works because we still need to find better ways to identify brain tumors so we can diagnose them correctly and help people get better.

Selvy et al. [24] utilized a depthwise separable CNN for MRI-based tumor detection, demonstrating superior performance compared to traditional methods such as support vector machines (SVMs) and KNN. DL, specifically CNN and depthwise separable methods, improves tumor detection accuracy, with a depthwise separable CNN achieving an accuracy of 92%. Although effective in tumor detection, future improvements should concentrate on identifying specific tumor types to facilitate tailored treatment recommendations, thereby advancing healthcare. Shoaib et al. [25] investigated CNN-based models for brain tumor classification, attaining an accuracy of 93.15% for transfer learning and 91.24% for CNN-based models. The proposed models utilize MR images to classify glioma, meningioma, pituitary tumors, and normal cases effectively. The transfer learning model demonstrates superior performance compared to other CNN-based methods, yielding favorable outcomes for practical applications and potential improvements in real-time detection through Field-Programmable Gate Array (FPGA) or mobile platforms. Future research will focus on expanding the database, refining model structures, and enhancing the accuracy of tumor localization in brain surgery.

Zeineldin et al. [26] presented NeuroXAI, an XAI framework designed for DL networks in the context of medical image analysis, with a specific emphasis on brain imaging. NeuroXAI utilizes seven advanced explanation methods to generate visualization maps, thereby improving the transparency of DL models. NeuroXAI's experiments in image classification and segmentation tasks illustrate its capability to aid radiologists and medical professionals in the detection and diagnosis of brain tumors. In the future, research intends to quantitatively assess the quality of explanation methods and investigate the extraction of quantitative features from these explanations, thereby enabling their incorporation into clinical practice. Abdalla et al. [27] examined the effects of data augmentation techniques on the efficacy of deep transfer learning networks (Inception-V3, VGG-16, and DenseNet169) in identifying brain tumors from MRI scans. The results indicate notable enhancements in accuracy following augmentation, with VGG-16, Inception-V3, and DenseNet169 attaining accuracy rates of 96.88%, 98.44%, and 96.88%, respectively. The findings highlight the significant impact of data augmentation on improving the reliability of MRI-based brain tumor diagnoses, thereby benefiting patient care. The study utilizes different artificial intelligence (AI) models to make a better system for finding brain tumors, and it shows that adding more information to the pictures is a really good way to make the system more accurate.

2.3. Hybrid techniques

Hybrid techniques that include multiple ML and DL models have drawn interest in their ability to improve brain tumor detection efficiency. Biswas and Islam [28] introduced a hybrid model that combines DCNN and SVM for the classification of brain tumors. Preprocessing techniques and data augmentation are utilized on MRI images to improve feature extraction. The proposed model demonstrates enhanced performance relative to transfer learning models (AlexNet, GoogLeNet, VGG-16), achieving an accuracy of 96.0%, specificity of 98.0%, and sensitivity of 95.71%. Furthermore, it exhibited reduced processing time and efficient feature extraction without overfitting, suggesting its capability for precise and swift brain tumor classification. Khalid Alduraibi [29] examined the effectiveness of deep features derived from 18 pretrained CNNs for the classification of brain tumors in MRI images. The proposed approach utilizes a combination of DenseNet201, EfficientNet-B0, and DarkNet-53 deep features alongside SVM, resulting in an accuracy of 97.1%, which exceeds that of other models. ReliefF feature selection combined with particle swarm optimization (PSO) improves performance, indicating its applicability in real-time clinical settings. The incorporation of larger datasets and data augmentation may enhance the model's reliability and accuracy in subsequent research endeavors.

Agarwal et al. [30] proposed an advanced method employing DL for the detection of brain tumors from MRI images, tackling the difficulties associated with manual detection and the necessity for automated diagnostic systems. The fine-tuned EfficientNet-B0 model demonstrates enhanced performance, achieving an accuracy of 98.87%, exceeding that of other leading CNN models. Future research will investigate further deep CNN models, refine segmentation techniques, and extend applications to various medical imaging modalities to improve accuracy and increase utility in clinical practice. Islam et al. [31] proposed a novel ensemble model that integrates three components: a 2D CNN with 13 layers, a 2D CNN with 9 layers, and a 2D CNN-LSTM, aimed at classifying brain tumors from MRI images. Augmenting the ensemble architecture with additional fully connected layers results in a notable performance increase, achieving an accuracy prediction of 98.82%. The results highlight the effectiveness of ensemble learning in the classification of medical images, presenting favorable opportunities for precise diagnosis of brain tumors. They could improve model interpretability and refine region of interest detection to enhance efficacy in clinical applications.

Abbas et al. [32] introduced an automated method for brain tumor classification utilizing a HYBRID CNN-LSTM approach. This method includes preprocessing steps such as rotation, cropping, and zooming, followed by tumor segmentation and classification based on PSO with the hybrid CNN-LSTM model. The method attains a peak accuracy of 92.03% through the application of deep transfer learning and data augmentation. This research emphasizes the efficacy of integrating CNN and Long Short-Term Memory (LSTM) architectures for effective brain tumor classification, providing valuable insights for the development of future Computer-Aided Diagnosis (CAD) tools in medical imaging.

Recent works regarding brain tumor detection are summarized in Table 1, considering DL, classical ML, and ensemble approaches.

Table 1
Recent works on brain tumor identification based on deep learning, classical ML, and ensemble approach

Refs.	Used technique	Dataset	Accuracy	Limitation
[16]	OGDMLA: Edge-based segmentation	994 MRI images	99.55%	Segmentation limitations: complex shapes, imaging variations, dataset specificity
[17]	CNN-based Inception-V3, EfcientNetB4	2870 MRI images	98%	Lack of essential preprocessing steps impacts accuracy
[18]	EDN-SVM model	253 MRI images	97.93%	Limited applicability to gray-scale images and 2D brain scans only
[19]	2D CNN-auto-encoder network	275 MRI images	96.47%	Limited evaluation on diverse MRI datasets and tumor types
[20]	ANN and CNN-based models	2020 images	88.70%	Preprocessing effect on imbalanced dataset accuracy unexplored
[21]	BKNN	3,064 MRI images	97.7%	BKNN's inefficiency hampers real-time diagnosis scalability
[22]	Inception-V3 and DenseNet201	3064 T1-weighted contrast MRI images	99.34%	Dependence on pre-trained models hampers adaptability
[23]	YOLOv7 model	8232 MRI images	99.5%	More validation needed for real-world effectiveness
[24]	KNN-CNN	253 MRI images	92%	Future enhancements for tumor-specific treatment suggestions
[25]	CNN, Inception-ResNet-v2, Inception-V3	826 MRI images	93.15%, 86.80%, 85.34%	Database size limits findings, requiring expansion
[26]	CNN-NeuroXAI framework	1251 MRI images	–	Quantitative XAI evaluation and feature extraction planned
[27]	Inception-V3, VGG-16, and DenseNet169	253 MRI images	96.88%, 98.44%, 96.88%	Small datasets necessitate frequent data augmentation techniques
[28]	CNN-SVM	2957 MRI images	96.0%	Study lacks exploration of MRI protocol impact
[29]	DenseNet201 and DarkNet-53	2870 MRI images	97.1%	Study lacks exploration of MRI protocol impact on classification
[30]	Decision Tree (DT), Naïve Bayes (NB), AdaBoost, KNN, SVM-feature extraction	4850 MRI images	99.58%, 99.59%, 99.58%, 99.58%	Study lacks consideration of dataset biases for generalization
[31]	13-layer CNN	10470 MRI images	98.82%	Ensemble learning's interpretability and biases not addressed
[32]	CNN-LSTM and CNN-BiLSTM	5712 MRI images	CNN-LSTM: 97.86%, CNN-BiLSTM: 99.77%	The small dataset limits generalizability; future work will address this with larger datasets

3. Methodology

An overview of the proposed approach is shown in Figure 1. Two MRI datasets that were previously made public were combined to make a new, more extensive dataset. Next, it was passed into the feature extractor after being preprocessed. The next step is to evaluate and fit our BTI model.

3.1. Classification dataset

In this paper, two datasets were used from Kaggle in the public domain. Both datasets were clinically approved.

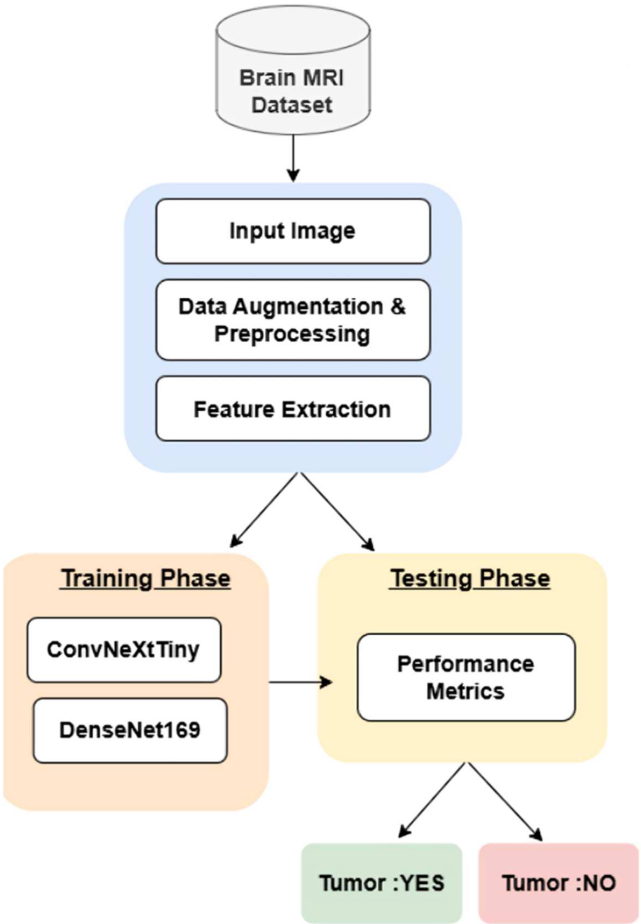
3.1.1. Dataset a

The first set of data was the Br35H: Brain Tumor Detection 2020, which addresses the complexities of BTI due to variations in size and location [33]. This dataset comprises a well-curated selection of tumors, organized into three folders labeled "yes," "no," and "pred," containing a total of 3060 brain MRI images.

3.1.2. Dataset b

The second dataset was sourced from the Kaggle Brain MRI Images for Brain Tumor Detection [34]. With advancements in medical technology, MRI scans have emerged as a dependable resource. This dataset comprises 253 images, classified into "yes"

Figure 1
Methodology block diagram



and "no" files, denoting the presence or absence of a tumor in patients.

3.1.3. Merged dataset

The two publicly available MRI datasets were merged to create a customized binary classification dataset consisting of two classes: tumor ("yes") and non-tumor ("no"). The final merged dataset contained 3253 MRI images, including 1655 tumor images and 1598 non-tumor images. The dataset was randomly divided into training (2603 images, 80%), validation (325 images, 10%), and testing (325 images, 10%) subsets, as summarized in Table 2.

Table 2
Overview of dataset distribution for training, test, and validation

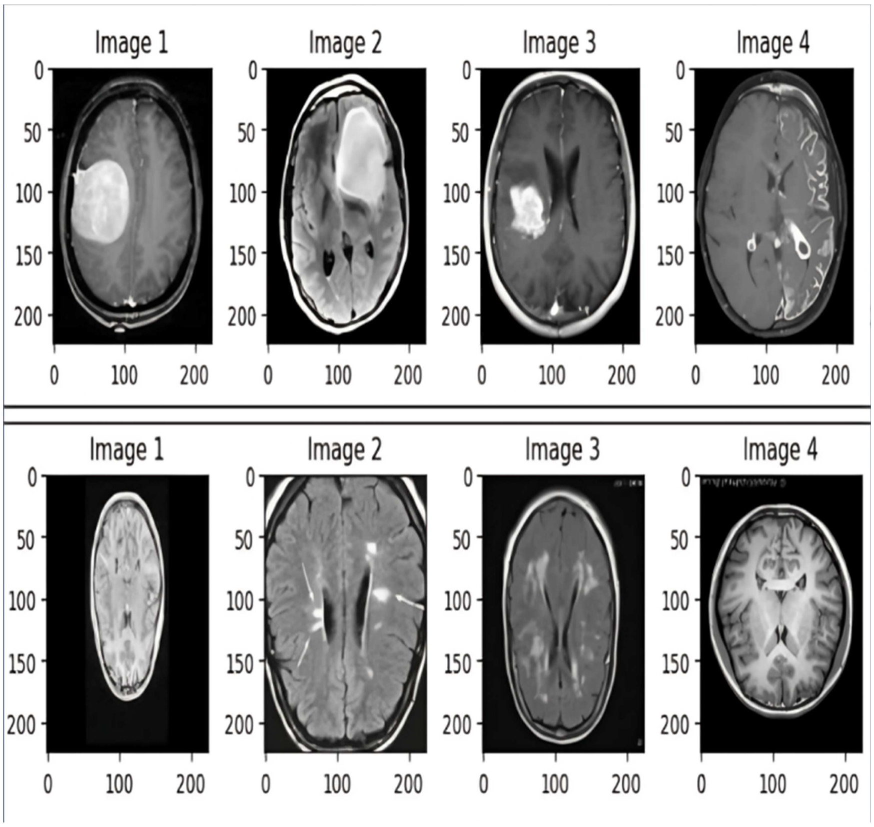
Class	Training	Validation	Test	Total
Yes	1325	165	165	1655
No	1278	160	160	1598
Total	2603	325	325	3253

Besides, some random sample images of a merged dataset of brain tumors are given in Figure 2. The upper four images have brain tumors, and the lower four images are normal brain images.

3.2. Data preprocessing

1) **Image augmentation:** Image augmentation is an important preprocessing method used to artificially increase the size and diversity of the dataset. This procedure involves performing a sequence of arbitrary transformations to the original images, including rotations, translations, zooming, shearing, flipping, and further modifications. These changes are implemented in real time during model training, ensuring the model encounters a broader range of visual circumstances. The objective of image augmentation is to mitigate overfitting by enhancing the model's generalization capabilities.
2) **Image rotation:** We used images captured from different orientations due to variations in patient positioning or scanner settings. To account for these differences, we applied image rotation as a form of augmentation. Specifically, the images were rotated 90 degrees clockwise using the rotation_range parameter in the Keras ImageDataGenerator. This augmentation simulates a variety of orientations, enabling our proposed model to learn and recognize tumors from different angles and perspectives. By training on rotated images, the model improves its robustness to variations in brain structural orientation, resulting in an augmentation of its capacity to generalize across diverse scanning settings and patient locations.
3) **Image shifting:** This augmentation simulates minor movements or misalignments of the MRI images, which can occur during the imaging process or due to patient motion. Both horizontal and vertical shifting are applied to the images within the range defined by the "width_shift_range" and "height_shift_range" parameters, respectively. This helps the proposed model to be more effective in the detection of tumors regardless of their position within the image. By applying horizontal and vertical shifts to the images, the proposed model is able to concentrate on different areas of the brain for the detection of tumors regardless of their position within the image.
4) **Image shearing:** Shearing transformations cause distortions to the shape of the images, simulating possible deformations that may be present in the images due to factors such as the positioning of the patients, the equipment, and possible artifacts that may be present during the scanning process. The "shear_range" parameter was used to have a better understanding of the images. This data augmentation technique allows the model to learn from images that are distorted, hence helping the model to recognize the presence of the tumor even if the images are distorted to some extent. By allowing the model to learn from sheared images, the proposed model is able to detect the tumor even if the images are distorted to some extent.
5) **Image zooming:** Zooming magnifies or shrinks specific parts of the MRI images, allowing our proposed model to focus on different regions with varying levels of detail. We have used image zooms by scaling their dimensions by a factor of 1.5 in both the x and y directions ($f_x = 1.5$, $f_y = 1.5$), a parameter that controls the image. This helps our proposed model learn to detect tumors of different sizes and ensures robustness to variations in image resolution.
6) **Image flipping:** Horizontal flipping method creates mirror images of the original MRI image, effectively doubling the size of the training dataset. This data augmentation method enables our proposed model to generalize even better, as the model is exposed to images with reversed left-right orientation with a certain probability defined by the "horizontal_flip"

Figure 2
Brain tumor "yes" and "no" images



parameter, thereby enabling the model to detect tumors regardless of orientation. This is particularly useful when dealing with asymmetrical structures in the brain, where flipping enables the model to detect tumors regardless of whether they are on the left or right side of the brain. Table 3 shows the result of data augmentation.

Table 3
Number of data in datasets 1 and 2 before and after augmentation

Dataset	Yes	No	Total images
Br35H:2020	1500	1500	3000
Brain MRI images	155	98	253
Before augmentation	1655	1598	3253
After augmentation	3153	3058	6211

The original merged dataset contained 3253 MRI images. Data augmentation was applied only to the training dataset to improve model generalization, whereas the validation and testing datasets were kept unchanged to ensure an unbiased evaluation of the proposed model.

7) **Outlier handling:** Outliers refer to elements that vary significantly from the majority. We have used the interquartile range method to detect outliers based on the spread of the pixel intensity values within the image. Then, finally, the outliers are replaced with a more representative value, which then reduces their impact on the model's training. This ensures that the proposed model does not get affected by the outlier pixel intensity values present within the MRI images.

These preprocessing techniques not only increase the diversity of the training data but also ensure that the model is not misled by extreme pixel intensity values that could skew its learning process. This contributes to the effectiveness and robustness of the model.

3.3. Feature extraction

Two pre-trained models were evaluated for feature extraction. All the feature extractors utilized in this study are illustrated in Figures 3 and 4. Transfer learning is a methodology in which a pre-trained model, originally developed for one task, is utilized for a related one. This method decreases training duration as the model has acquired knowledge from a comparable issue. In image classification, multiple CNN models that achieved prominence in the ImageNet challenge are commonly utilized via transfer learning to tackle diverse image classification tasks. This study employs two pre-trained CNN models for feature extraction: ConvNeXt-Tiny [35] and DenseNet169 [29]. We have utilized all these pre-trained models on the three aforementioned datasets.

3.3.1. ConvNeXt-Tiny model

ConvNeXt, proposed by Facebook AI Research (FAIR) in 2022, makes operations such as window shifting and relative position bias, resulting in improved performance and reduced computational requirements relative to transformer networks. It is important to note the significant role of this model in comparison to other pre-trained alternatives. ConvNeXt, inspired by the ResNet architecture, utilizes a comparable structure that incorporates residual blocks. Furthermore, it incorporates multiple advanced network design

Figure 3
Pre-trained ConvNeXt-Tiny model architecture

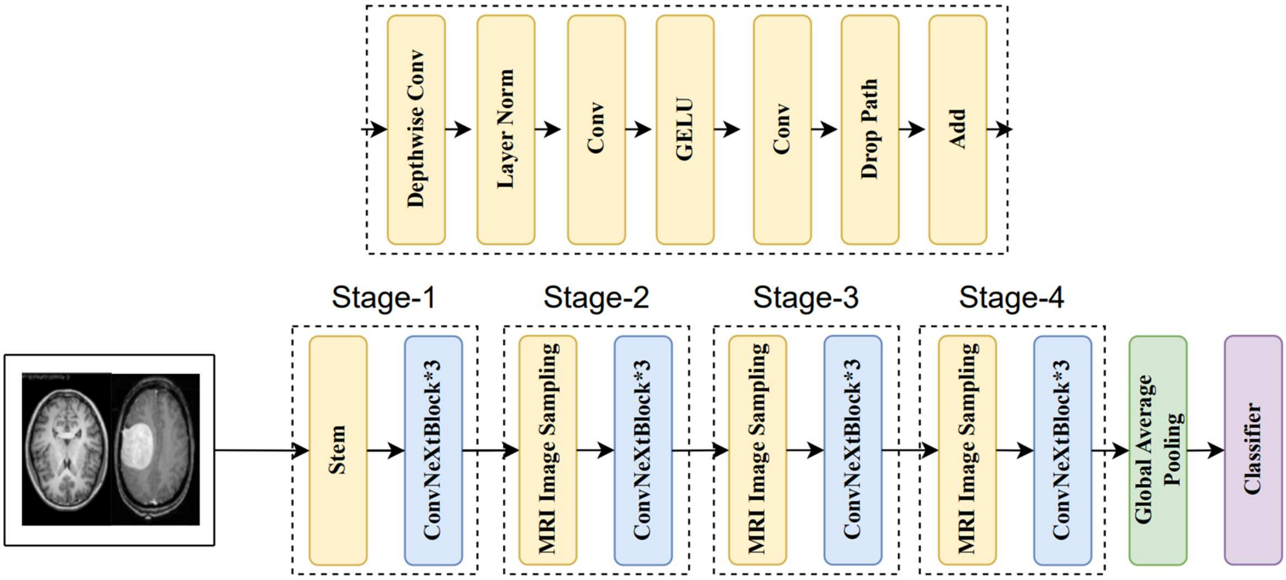
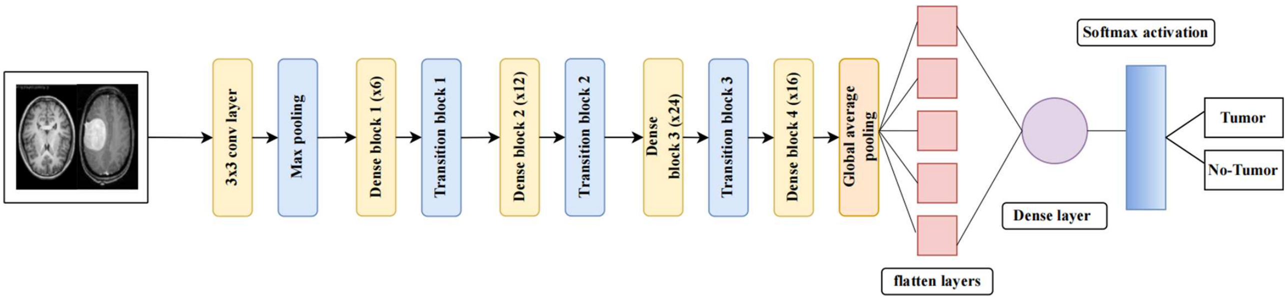


Figure 4
Structure of DenseNet169 model



strategies to improve overall performance. Figure 3 illustrates the detailed structure of ConvNeXt-Tiny and the ConvNeXt Block.

3.3.2. DenseNet169 model

DenseNet169 is a powerful CNN architecture. It tackles vanishing gradients, a training hurdle, by connecting each layer to all prior layers. This "dense connectivity" improves feature reuse and learning efficiency. DenseNet169's depth and design make it effective for complex image classification tasks, influencing future network designs [30]. It is important to highlight the significant role of this model compared to other pre-trained CNNs. DenseNet-based architectures are widely used in fields such as image classification, Natural Language Processing (NLP), autonomous vehicles, earthquake prediction, object detection, and speech recognition [36]. Figure 4 shows the architecture of the DenseNet169 model.

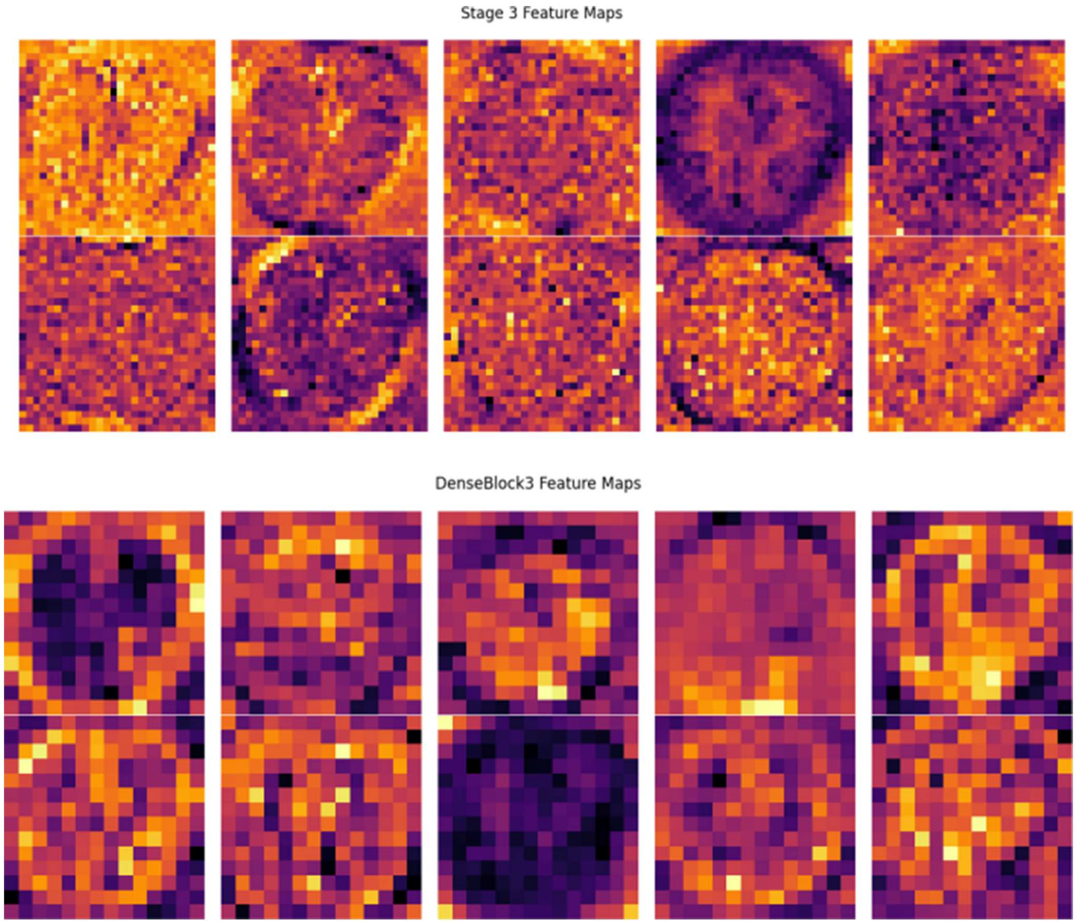
These visualizations illustrate the intermediate feature maps extracted from the ConvNeXt-Tiny and DenseNet169 models during the processing of brain MRI images.

The ConvNeXt-Tiny feature maps in Figure 5(a) represent early-stage activations from the model's convolutional layers, which are responsible for low-level spatial feature representations such as edge, contour, and textural features. These types of feature representations are essential for developing structural awareness within the image, enabling the model to differentiate between normal and abnormal brain regions based on visual characteristics such as irregularities in shape and boundary disruption. As a result of its lightweight design, ConvNeXt-Tiny is computationally efficient in representing basic feature representations.

In contrast, the DenseNet169 feature maps in Figure 5(b) display deeper-level representations, where each layer is connected to all previous layers. These maps reflect more semantic and abstract features associated with the presence of tumors, such as variations in intensity, tissue density, and irregular anatomical structures. DenseNet's architecture facilitates more efficient feature reuse, which is particularly important when examining complex and subtle images, such as MRI scans of the human brain.

Together, these visualizations underscore the complementary roles of both models in the proposed ensemble framework. ConvNeXt-Tiny provides fast, efficient extraction of basic features, while DenseNet169 contributes deeper interpretive power and enhanced sensitivity to fine-grained tumor characteristics. This multi-level feature fusion enriches the model's overall representation capability, leading to improved accuracy, robustness,

Figure 5
ConvNeXt-Tiny feature heat map and (b) DenseNet feature heat map



and generalization in classifying both tumor and non-tumor cases.

3.3.3. Feature extractor selection

Analyzing the pre-trained models, we select ConvNeXt-Tiny and DenseNet169 for their performance in datasets a and b and the merged dataset.

3.4. Proposed ensemble model architecture

After selecting the feature extractor, we ensemble two pre-trained base models, ConvNeXt-Tiny and DenseNet169. By combining the strengths of these two models, we concentrate on and use techniques aiming to extract features from MRI images and classify the accuracies. The models are used by freezing the weights and enabling the model's performance and generalization ability. The proposed model architecture is given in Figure 5 and described each layer below.

3.4.1. Pre-trained base models

As seen in Figure 6, the model's architecture demonstrates how we integrated two pre-trained base models—ConvNeXt-Tiny and DenseNet169—to build an efficient brain tumor classification model.

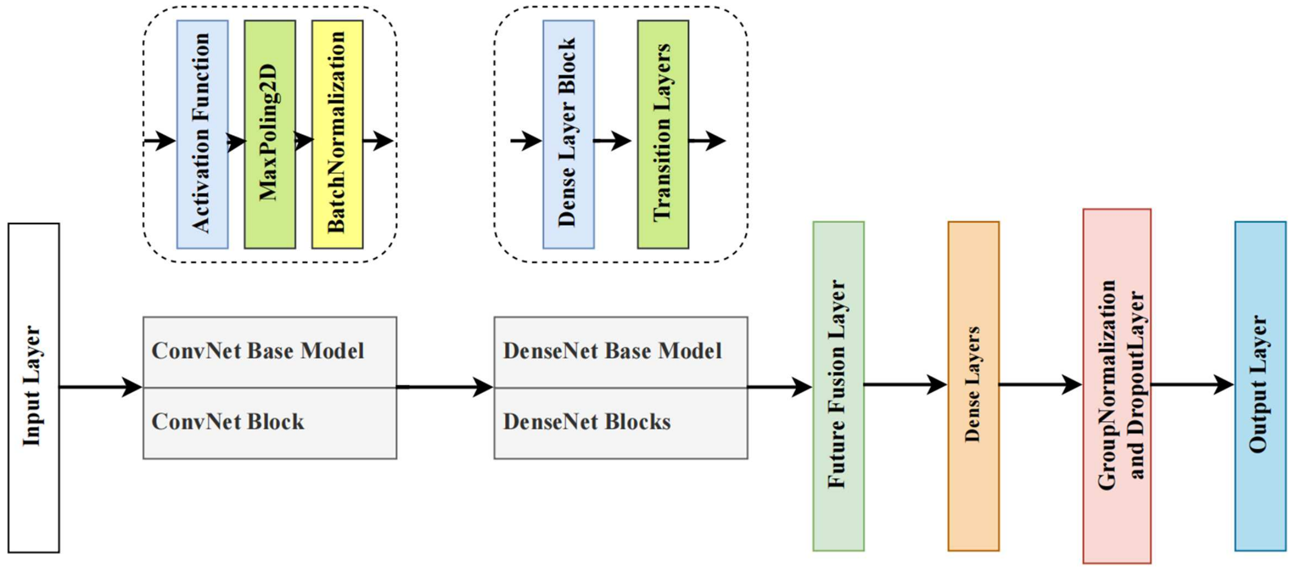
The second pre-trained base model is DenseNet169, where each layer is only connected to the next subsequent layer in a feed-forward fashion. This architecture enhances gradient flow throughout the network, improving parameter efficiency and overall performance.

The merged dataset was randomly divided into training (2603 images, 80%), validation (325 images, 10%), and testing (325 images, 10%) subsets before model training. Data augmentation was applied only to the training dataset, while the validation and testing datasets remained unchanged. All MRI images were resized to 224 × 224 pixels and normalized prior to feature extraction. The proposed model was trained using the Nadam optimizer with the categorical cross-entropy loss function and a batch size of 64. Training was performed for a maximum of 50 epochs. To improve convergence and reduce overfitting, ReduceLROnPlateau and Early Stopping (patience = 7) were employed, and the model with the lowest validation loss was selected for final evaluation.

To adapt these models for brain tumor detection, we introduced several custom layers. In ConvNeXt-Tiny, global average pooling and fully connected layers were integrated to further refine the feature maps and enable classification specific to MRI images. For DenseNet169, additional dropout and dense layers were added to prevent overfitting, particularly important when working with relatively small datasets like medical images. Batch normalization layers and ReLU activation functions were included in both models to stabilize learning, enhance convergence, and allow the networks to capture more complex patterns within the MRI data.

Both models underwent fine-tuning, where only the final layers were updated to adjust the models specifically for MRI scans.

Figure 6
Proposed model architecture



The number of epochs was determined effectively through early stopping. Training was initially intended for 50 epochs; however, validation accuracy dropped after the ninth epoch, triggering the early termination condition at epoch 16. This guaranteed that the model performed optimally, preventing overfitting and ensuring that our proposed models could generalize well to unseen data.

The combined architecture of ConvNeXt-Tiny's lightweight efficiency and DenseNet169's advanced feature reuse capabilities significantly improves detection performance, leading to a robust and accurate system for brain tumor detection across various cases.

3.4.2. Frozen weights

The weights in the models were frozen to have more effective processing of the two base models, which made the parameters stop updating the pre-trained models during the training process. We froze the weights to get the benefit and give a chance to learn the features of an extensive dataset without implementing computation functions. Some of the most crucial points of freezing the weights are discussed below.

1) Preserving learned features: freezing the lower layers of pre-trained models keeps useful basic features (like edges and shapes) intact, allowing them to serve as reliable feature extractors for new tasks.
2) Fine-tuning: the lower layers remain fixed, while the higher layers near the output are actively adjusted or retrained to suit the new task, with their weights updated based on training gradients.
3) Preventing overfitting: freezing the lower layers of a pre-trained model helps prevent overfitting, especially with limited training data.

3.4.3. Concatenation of outputs and layers

First, MRI images were taken by the input layer containing arrays. These images were then fed into the model. The ConvNeXt-Tiny model served as the foundation of the proposed concatenated model. It included layers with activation and pooling functions that extract hierarchical features, capturing patterns at various spatial scales.

Next, the DenseNet model acted as the second base in the proposed concatenated model. Its densely connected convolutional layers let each layer receive feature maps from previous layers. This dense connectivity enabled feature reuse, enhancing gradient flow and promoting robust learning.

The feature fusion step merges the distinct features learned by ConvNeXt-Tiny and DenseNet169, providing more varied and diverse input data. Then, dense layers and group normalization were applied, where concatenated features were passed through several layers. Each layer applied a linear transformation to the input features, followed by group normalization to support stable training and improved generalization.

Dropout was then applied to help prevent overfitting by randomly dropping units during training. Finally, the output layer, containing two units, classifies images into two classes. A dense layer with softmax activation generated final class probabilities for brain tumor detection, ensuring the classes were correctly classified.

3.5. Model evaluation parameters

We applied several performance evaluation metrics such as accuracy, precision, recall, and F1-score in order to validate the performance of the proposed model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1} - \text{score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{4}$$

$$\text{TPR} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{6}$$

After training, the best-performing model was evaluated using the independent testing dataset consisting of 325 MRI images that were not used during the training or validation stages. No data augmentation was applied to the testing dataset. The model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and receiver operating characteristic (ROC) curve with the corresponding area under the curve (AUC). These evaluation metrics were computed on the unseen testing dataset to ensure an objective assessment of the proposed framework.

4. Results and Findings

Our proposed ensemble model has been validated through various methods to accurately assess performance. The performance is evaluated using various performance metrics. The effects of ensemble learning are also discussed. The analysis includes the performance improvements following augmentation. The proposed model is validated to assess its generalization capability. We integrate Grad-CAM++ and LIME to elucidate the transparency of our proposed model. The proposed model is compared with existing research.

4.1. Performance metrics

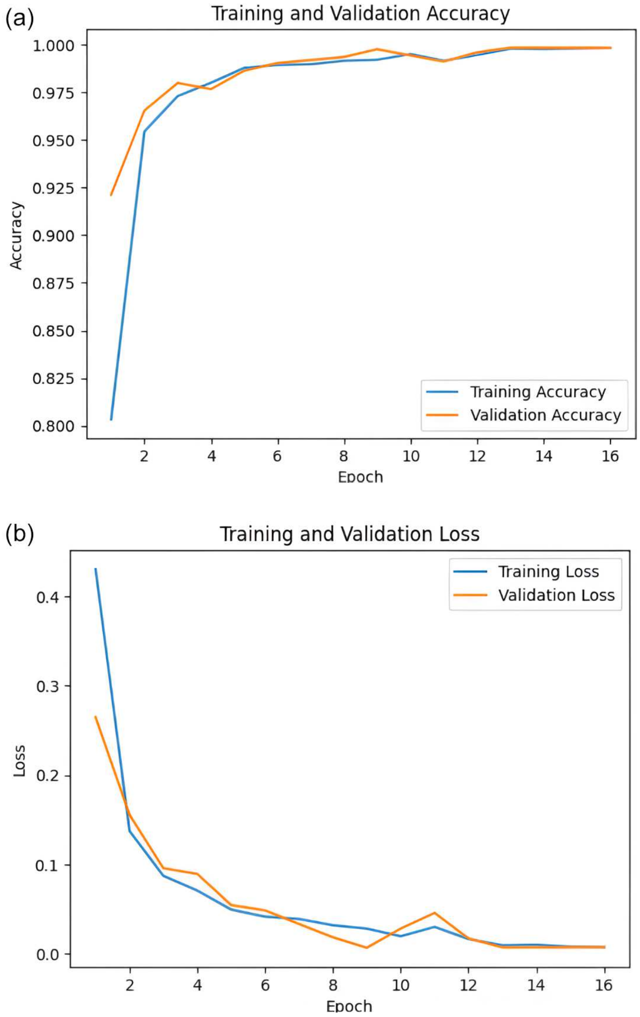
A wide range of performance indicators, including accuracy, precision, recall, F1-score, ROC, and AUC score, were used to evaluate the proposed model. Upon training our proposed model, the accuracy and loss curves for both training and validation are illustrated in Figure 6. Figure 7(a) and (b) illustrate the training and validation accuracy and loss curves, demonstrating the learning and generalization capabilities of our proposed model. Early halting was implemented at the 16th epoch to mitigate overfitting, with peak validation performance noted at approximately the 9th epoch. The progressive rise in training accuracy signifies an acquaintance with the training samples, whereas elevated validation accuracy validates the model's efficacy on novel data. The consistent decline in training and validation loss signifies effective learning, with negligible disparity between predicted and actual labels, suggesting that the model generalizes effectively without overfitting.

4.2. Explainable AI (XAI) techniques

We carried out a comprehensive analysis of our preprocessed MRI scans using sophisticated ML models to make predictions regarding abnormalities. To make our model's predictions more comprehensible, we utilized XAI techniques such as LIME and Grad-CAM. This involved the use of various algorithms to make accurate and reliable predictions.

LIME provides interpretable explanations by highlighting specific brain regions that significantly affected the model's predictions. On the other hand, Grad-CAM generates heatmaps that illustrate areas with the highest importance based on the model's gradient values. This method not only reveals which regions of the MRI scan the model focuses on when making predictions but also provides visual validation of the model's reasoning process. By visually depicting the relevant features that contribute to a specific prediction, Grad-CAM enhances our understanding of the model's behavior, thereby offering deeper insights into the decision-making process.

Figure 7
Training progress: (a) accuracy value during the training and validation process, (b) loss value during the training and validation process



In addition, we have also ensured the robustness of our approach by incorporating custom layers within our ML models. These custom layers have been developed to adapt to the unique characteristics of the MRI data, which often includes complex patterns and variations that standard models might overlook. By enhancing feature extraction and representation, these layers allow the model to capture more relevant information from the data.

Throughout the training process, we conducted iterative model feature updates to refine the learning process continuously. This iterative approach allowed the model to adapt dynamically to the data, improving its performance on novel instances while maintaining its ability to generalize across different cases. By integrating advanced interpretability tools and a powerful model architecture, we have been able to develop a reliable and transparent framework for predicting brain abnormalities.

The inclusion of XAI is especially important for medical applications, where black-box predictions cannot be fully reliable. LIME and Grad-CAM++ provide transparency by showing which regions of the scan influence the decision, enabling clinicians to validate the attention of the model. This can increase trust, interpretability, and the likelihood of clinical adoption of AI-assisted diagnosis systems.

In Figure 8(a) and (b), the leftmost image in both the "yes" and "no" images of tumor and non-tumor MRI scans has undergone several preprocessing steps. Then we applied Gaussian filtering to use and reduce noise, enhancing image quality. Pixel intensity normalization and segmentation removed non-brain tissues, standardizing and isolating brain structures for accurate analysis.

The middle image in Figure 8 in both tumor and non-tumor images shows the results of applying LIME to the MRI scan. LIME is a technique that we used to explain the predictions of ML models by highlighting the regions of the input image that most strongly influence the model's decision. The LIME explanation is visualized in this image as a heatmap overlay on the MRI scan. The highlighted regions indicate the areas the model found most significant when making its prediction. These regions are crucial for understanding which parts of the brain the model considers important.

The rightmost image in Figure 8 in both tumor and non-tumor images presents a Grad-CAM heatmap applied to the same MRI scan. We used Grad-CAM as a visualization technique that provides insights into the DL model's decisions by generating a heatmap that highlights the regions of the image with the highest gradient values relative to the predicted class. In this image, the Grad-CAM heatmap is overlaid on the MRI scan. The areas with the highest importance (usually in warm colors like red and yellow) indicate where the model is focusing its attention when making a prediction. This helps in understanding the model's behavior and ensuring that its decisions are based on relevant anatomical features.

4.3. Confusion matrix analysis

A confusion matrix is a table that visualizes the performance of a model by displaying the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. It also helps us to understand the types of errors made by the model and identify the classes of the model. Figure 9 shows Actual Yes (positive class): TP: 613, FN: 2 and Actual No (negative class): FP: 1, TN: 629.

The matrix shows the TP correctly predicted 613 instances as "yes" when they were actually "yes." The model incorrectly predicted two instances as "no" when actually "yes." Again, for the FP, the model incorrectly predicted one instance as "yes" when it was actually "no," and the model correctly predicted 629 instances as "no" when they were actually "no." According to the confusion matrix, we calculate the accuracy, precision, recall (sensitivity), and F1-score of the model, as given in Figure 9.

4.4. ROC curve

The ROC is a graphic representation of the True Positive Rate (TPR) against the False Positive Rate (FPR). It helps to understand the trade-off between sensitivity and specificity for different classification thresholds. The ROC curve represents the model's performance by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). Each point on the ROC curve represents the point (0,0) and ends at (1,1). In our model, the ROC curve that passes through the point (0,1), indicating a TPR of 1 (all positives correctly predicted) and an FPR of 0 (no false positives). As seen in Figure 6, it has a colored curve; the blue line represents one class, and the red line represents the other. The

Figure 8
(a) Brain tumor original image "yes," LIME explanation and Grad-CAM image, and (b) brain tumor original image "no," LIME explanation and Grad-CAM image

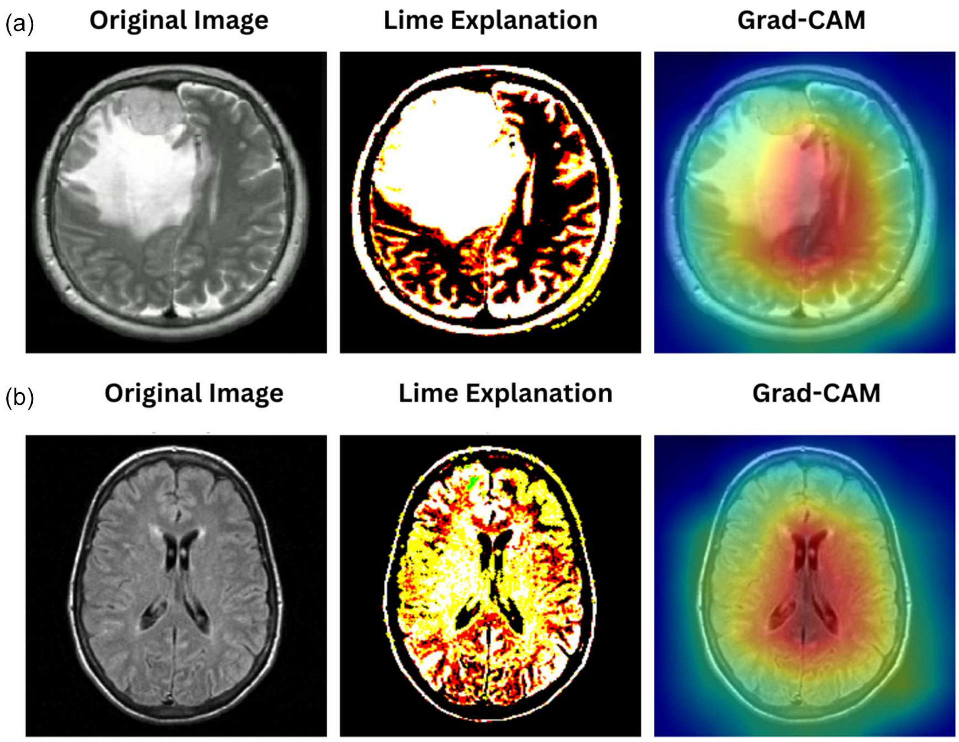
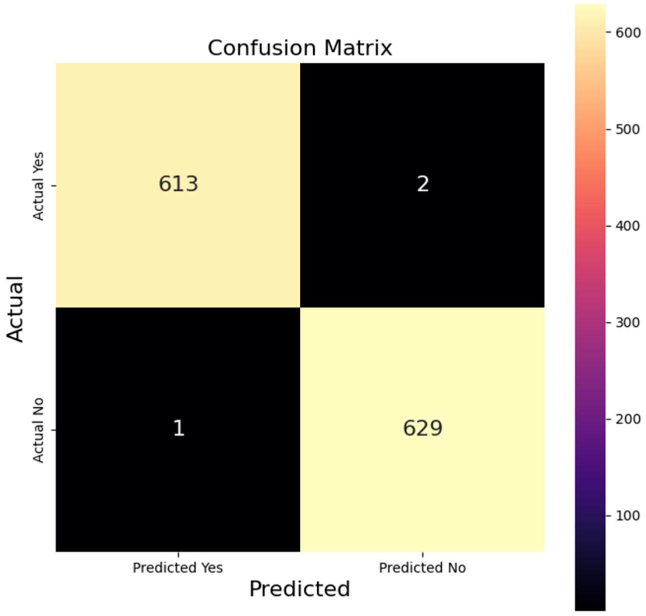


Figure 9
Confusion matrix



AUC is also displayed in the legend. A higher AUC value suggests better model performance between the positive and negative classes. In our model, both classes have high AUC values close to 1, which indicates the excellent ability of the model.

Our hybrid model outperformed DenseNet169 with a 1.9% higher accuracy, 2.3% higher recall, and 2.1% higher F1-score. When compared to ConvNeXt-Tiny, the improvements were more significant, with a 4.1% increase in accuracy and a 5.8% increase in recall. These constant increases across all performance parameters demonstrate the benefits of feature fusion and augmentation in our approach.

4.5. Computational performance

Our proposed hybrid model required approximately 3 h and 15 min of training on an NVIDIA RTX 3080 GPU (10 GB VRAM). Although the maximum number of training epochs was set to 50, the training process converged after 16 epochs due to the Early Stopping criterion. The average inference time per MRI image was 42 ms. These results demonstrate that the proposed model achieves high computational efficiency while remaining suitable for real-world clinical deployment.

5. Discussion

Among the evaluated architectures, ResNet50 V2 and DenseNet169 achieved high classification accuracies of 98.97% and 97.87%, respectively, demonstrating the effectiveness of deeper, feature-rich networks. MobileNet V2 and ConvNeXt-Tiny, both designed for computational efficiency, performed robustly with accuracy exceeding 95%, making them strong candidates for deployment in resource-constrained environments. To understand how our model and other models that use pre-trained CNN models have worked and which combination of CNN pre-trained models has given the best accuracy, we will discuss them in Figures 8 and 9.

In Table 4, the highest performance was observed in the hybrid model combining DenseNet169 and ConvNeXt-Tiny, which achieved 99.74% accuracy. This suggests that integrating the strengths of multiple architectures can lead to substantial gains in model generalization and predictive accuracy. However, these findings should be interpreted within the context of the experimental design and dataset used. It is important to note that these results are not intended as a comparative study with existing state-of-the-art methods. The dataset and evaluation criteria used in this work may differ from those reported in the original papers of other methods. Therefore, unless future studies adopt a standardized benchmark or the reported methods are re-implemented under identical conditions, such comparisons would not be methodologically sound. Table 5 provides a comparative overview of selected brain tumor classification models with respect to dataset size, classification accuracy, and integration of XAI methods. Models referenced in studies [23, 24, 31, 32] demonstrate high classification accuracy ranging from 98.82% to 99.5%, utilizing substantial MRI datasets. However, none of these models incorporate XAI techniques, which are increasingly important for enhancing model transparency and clinical interpretability. In contrast, the proposed model achieves a superior accuracy of 99.74% using a dataset of 3253 MRI images and is the only approach in the comparison to integrate XAI. This highlights its potential for supporting more interpretable and trustworthy diagnostic systems in medical imaging.

To evaluate the contribution of each module in the proposed explainable hybrid DL framework, an ablation study was conducted. Different configurations of the model were tested by progressively adding or removing key components—feature fusion, data augmentation, regularization techniques, and explainability integration. The performance was evaluated using the merged MRI dataset, and the results are reported in terms of accuracy and F1-score (Table 6).

The results clearly indicate that each component contributes incrementally to the overall performance. Feature fusion alone

Table 4
Performance of individual backbone architectures on our dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG-19	93.13	93.03	92.33	93.10
ResNet50 V2	98.97	98.77	98.07	98.01
Inception-V3	92.88	92.78	92.58	92.08
MobileNet V2	95.17	95.07	95.12	95.02
ConvNeXt-Tiny	95.65	94.65	95.35	95.45
DenseNet169	97.87	97.87	97.77	97.57
Proposed model	99.74	99.64	99.59	99.60

Table 5
Comparison of the proposed framework and XAI integration with existing research

Refs.	Dataset	Accuracy	Precision (%)	Recall (%)	F1-score (%)	XAI integration
[23]	Public 3-class dataset	99.45 ± 0.05%	99.38	99.35	99.36	No
[24]	Public dataset (≈10.3k images, 4 classes)	99.5%	99.5	99.3	99.4	No
[31]	Three open benchmark datasets	99.58%	99.59	99.58	99.58	No
[32]	Three merged datasets	98.82% (Ensemble DL)	98	98	98	No
Proposed model	3253 MRI images	99.74%	99.64	99.59	99.60	Yes

Table 6
Ablation study showing the contribution of each component in the proposed hybrid model on the merged MRI dataset

Model configuration	Feature fusion	Augmentation	Regularization (Dropout + Early Stopping)	XAI	Accuracy (%)	F1-score (%)
ConvNeXt-Tiny (baseline)	✗	✗	✗	✗	96.82	96.25
DenseNet169 (baseline)	✗	✗	✗	✗	97.10	96.72
ConvNeXt-Tiny + DenseNet169 (fusion only)	✓	✗	✗	✗	98.53	98.20
Fusion + Augmentation	✓	✓	✗	✗	99.02	98.85
Fusion + Augmentation + Regularization	✓	✓	✓	✗	99.74	99.68

significantly improved accuracy over single architectures, indicating that multi-level representation learning enhances tumor feature discrimination. Data augmentation further improved generalization and robustness, while regularization (dropout and early stopping) stabilized training and prevented overfitting. Lastly, the integration of XAI modules (LIME and Grad-CAM++) not only improved interpretability but also helped in identifying meaningful feature regions, indirectly enhancing model reliability and confidence for clinical application.

Despite the fact that the suggested model showed high levels of accuracy, we closely monitored for indicators of overfitting. Early stopping and dropout were used to stabilize training, and there is no substantial divergence in the training/validation curves (Figure 7). However, given the small sample size, there is still some risk of overfitting. To further validate generalization, future research will look into additional tactics such as model pruning, adversarial training, and large-scale validation.

While we report good accuracy, precision, recall, and F1-scores, we acknowledge that this study did not include formal statistical significance assessment or confidence interval estimates. Incorporating such statistical validation would result in a more rigorous measure of resilience. In the future, we plan to provide confidence intervals and hypothesis testing to back up the observed improvements.

The improved experimental results can be directly attributed to the research contributions embedded in our proposed framework. The combination of multi-level feature fusion, data augmentation, regularization, and XAI integration significantly enhanced the model's performance.

First, the feature fusion between ConvNeXt-Tiny and DenseNet169 enabled the extraction of both low-level structural and high-level semantic features, enriching the representational capacity of the model and improving classification accuracy. Second, data augmentation through rotation, shearing, zooming, shifting, and flipping expanded dataset diversity and improved generalization on unseen MRI scans. Third, regularization techniques, such as dropout and early stopping, effectively mitigated overfitting and stabilized training convergence. Finally, the integration of XAI methods (LIME and Grad-CAM++) provided interpretability and transparency, indirectly supporting model optimization through visual insights into decision-making regions.

Collectively, these contributions explain why our hybrid model consistently outperformed baseline architectures across all performance metrics, demonstrating superior robustness, generalization, and clinical reliability.

6. Conclusion and Future Work

While our model achieved great accuracy on the merged dataset, we acknowledge that the current validation is only applicable to this dataset. External validation across independent datasets is required to evaluate real-world generalizability and diagnostic accuracy, as highlighted in recent studies [37, 38]. In the future, we intend to expand testing to include multi-institutional datasets and previously unseen individuals to assess robustness in a variety of clinical contexts.

This study focused on identifying brain tumors from MRI data via a hybrid model that integrates DenseNet169 and ConvNeXt-Tiny architectures. Although CNNs have demonstrated efficacy in medical imaging applications, such as BTI, the application of the ConvNeXt-Tiny model for this objective was unprecedented. The ConvNeXt-Tiny model attained an accuracy of 82.51% independently. Our hybrid model, which integrates DenseNet169 and ConvNeXt-Tiny, attained an exceptional accuracy of 99.74%. Furthermore, we employed a merged dataset consisting of 3253 MRI images collected from two publicly available datasets to ensure comprehensive model training, validation, and testing. Our principal objective was to develop a robust and accurate hybrid framework for reliable brain tumor classification from MRI scans while enhancing model transparency through XAI techniques.

Future research should focus on larger and more diverse datasets to improve robustness and clinical applicability of DL models [39, 40]. This expanded dataset will enhance the model's accuracy and generalization potential. Moreover, recent studies have demonstrated that ML techniques are also effective in diagnosing and assessing other neurological disorders, such as Parkinson's disease, indicating the broader potential of AI in computer-aided medical diagnosis [41].

Ethical Statement

The authors declare that this study does not involve direct human or animal participation. All data used in this research were obtained from publicly available and anonymized MRI datasets. Therefore, formal ethical approval and informed consent were not required. This exemption complies with standard research guidelines for the use of publicly accessible datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in the American Society of Clinical Oncology at https://www.cancer.org/cancer/types/brain-spinal-cord-tumors-adults/key-statistics.html, in Kaggle at https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection, and in Kaggle at https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.

Author Contribution Statement

Md Sadi Al Huda: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Kazi Tanvir:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization. **Zubaida Akhter:** Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Md. Shahidul Khan Pappo:** Validation, Writing – review & editing, Visualization. **Md. Asraf Ali:** Resources, Writing – review & editing, Supervision, Project administration. **Nasim Ahmed:** Resources, Writing – review & editing, Supervision, Project administration.

References

[1] Pandey, V., Bhatt, P., Wagh, S., Shukla, S., Shivade, S., & Dahiwalkar, N. (2024). BTD-DCNN-BiLSTM: To build and enhance the brain tumour disease prediction and classification using deep CNN and BiLSTM. In *2024 1st International Conference on Advances in Computing, Communication and Networking*, 939–944. https://doi.org/10.1109/ICAC2N63387.2024.10894950

[2] Mohamed Shakeel, P., Tobely, T. E. E., Al-Feel, H., Manogaran, G., & Baskar, S. (2019). Neural network based brain tumor detection using wireless infrared imaging sensor. *IEEE Access*, *7*, 5577–5588. https://doi.org/10.1109/ACCESS.2018.2883957

[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://doi.org/10.1038/s41698-024-00789-2

[4] Abidin, Z. U., Naqvi, R. A., Haider, A., Kim, H. S., Jeong, D., & Lee, S. W. (2024). Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: A prospective survey. *Frontiers in Bioengineering and Biotechnology*, *12*, 1392807. https://doi.org/10.3389/fbioe.2024.1392807

[5] Zahoor, M. M., Khan, S. H., Alahmadi, T. J., Alsahfi, T., Mazroa, A. S. A., Sakr, H. A., . . . , & Alshemaimri, B. K. (2024). Brain tumor MRI classification using a novel deep residual and regional CNN. *Biomedicines*, *12*(7), 1395. https://doi.org/10.3390/biomedicines12071395

[6] Ullah, M. S., Khan, M. A., Masood, A., Mzoughi, O., Saidani, O., & Alturki, N. (2024). Brain tumor classification from MRI scans: A framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm. *Frontiers in Oncology*, *14*, 1335740. https://doi.org/10.3389/fonc.2024.1335740

[7] Rasheed, Z., Ma, Y. K., Ullah, I., Al Shloul, T., Tufail, A. B., Ghadi, Y. Y., . . . , & Mohamed, H. G. (2023). Automated classification of brain tumors from magnetic resonance imaging using deep learning. *Brain Sciences*, *13*(4), 602. https://doi.org/10.3390/brainsci13040602

[8] Rahman, A., Hayat, M., Iqbal, N., Alarfaj, F. K., Alkhalaf, S., & Alturise, F. (2025). Enhanced MRI brain tumor detection using deep learning in conjunction with explainable AI SHAP based diverse and multi feature analysis. *Scientific Reports*, *15*(1), 29411. https://doi.org/10.1038/s41598-025-14901-4

[9] Kimberly, W. T., Sorby-Adams, A. J., Webb, A. G., Wu, E. X., Beekman, R., Bowry, R., . . . , & Sheth, K. N. (2023). Brain imaging with portable low-field MRI. *Nature Reviews Bioengineering*, *1*(9), 617–630. https://doi.org/10.1038/s44222-023-00086-w

[10] Lamba, R., Bhatt, I., Madan, S., & Suman. (2023). Advancements in brain tumor detection: A deep learning approach. *Journal of Pharmaceutical Negative Results*, *14*(2), 2941–2948. https://doi.org/10.47750/0esx8677

[11] Asif, S., Yi, W., Ain, Q. U., Hou, J., Yi, T., & Si, J. (2022). Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images. *IEEE Access*, *10*, 34716–34730. https://doi.org/10.1109/ACCESS.2022.3153306

[12] Dawood, N. M., AbouEl-Magd, L. M., Abdel-Aty, A.-H., & Awad, W. S. (2023). Brain tumors detection using computed tomography scans based on deep neural networks. *Information Sciences Letters*, *12*(4), 33. https://doi.org/10.18576/isl/120433

[13] Krishnapriya, S., & Karuna, Y. (2023). Pre-trained deep learning models for brain MRI image classification. *Frontiers in Human Neuroscience*, *17*, 1150120. https://doi.org/10.3389/fnhum.2023.1150120

[14] Anjum, S., Hussain, L., Ali, M., Alkinani, M. H., Aziz, W., Gheller, S., . . . , & Duong, T. Q. (2022). Detecting brain tumors using deep learning convolutional neural network with transfer learning approach. *International Journal of Imaging Systems and Technology*, *32*(1), 307–323. https://doi.org/10.1002/ima.22641

[15] Alanazi, M. F., Ali, M. U., Hussain, S. J., Zafar, A., Mohatram, M., Irfan, M., . . . , & Albarrak, A. M. (2022). Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors*, *22*(1), 372. https://doi.org/10.3390/s22010372

[16] Manogaran, G., Shakeel, P. M., Hassanein, A. S., Malarvizhi Kumar, P., & Chandra Babu, G. (2019). Machine learning approach-based gamma distribution for brain tumor detection and data sample imbalance analysis. *IEEE Access*, *7*, 12–19. https://doi.org/10.1109/ACCESS.2018.2878276

[17] Khaliki, M. Z., & Başarslan, M. S. (2024). Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. *Scientific Reports*, *14*(1), 2664. https://doi.org/10.1038/s41598-024-52823-9

[18] Anantharajan, S., Gunasekaran, S., Subramanian, T., & R, V. (2024). MRI brain tumor detection using deep learning and machine learning approaches. *Measurement: Sensors*, *31*, 101026. https://doi.org/10.1016/j.measen.2024.101026

[19] Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making*, *23*(1), 16. https://doi.org/10.1186/s12911-023-02114-6

[20] Aamir, M., Rahman, Z., Bhatti, U. A., Abro, W. A., Bhutto, J. A., & He, Z. (2025). An automated deep learning framework for brain tumor classification using MRI imagery. *Scientific Reports*, *15*(1), 17593. https://doi.org/10.1038/s41598-025-02209-2

[21] Archana, K. V., & Komarasamy, G. (2023). A novel deep learning-based brain tumor detection using the Bagging ensemble with K-nearest neighbor. *Journal of Intelligent Systems*, *32*(1), 20220206. https://doi.org/10.1515/jisys-2022-0206

[22] Ahmed, M. M., Hossain, M. M., Islam, M. R., Ali, M. S., Nafi, A. A. N., Ahmed, M. F., . . . , & Islam, M. K. (2024). Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh. *Scientific Reports*, *14*(1), 22797. https://doi.org/10.1038/s41598-024-71893-3

[23] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers*, *15*(16), 4172. https://doi.org/10.3390/cancers15164172

[24] Vamsidhar, D., Desai, P., Joshi, S., Kolhar, S., Deshpande, N., & Gite, S. (2025). Hybrid model integration with explainable AI for brain tumor diagnosis: A unified approach to MRI analysis and prediction. *Scientific Reports*, *15*(1), 20542. https://doi.org/10.1038/s41598-025-06455-2

[25] Shoaib, M. R., Elshamy, M. R., Taha, T. E., El-Fishawy, A. S., & Abd El-Samie, F. E. (2021). Efficient brain tumor detection based on deep learning models. *Journal of Physics: Conference Series*, *2128*(1), 012012. https://doi.org/10.1088/1742-6596/2128/1/012012

[26] Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2022). Explainability of deep neural networks for MRI analysis of brain tumors. *International Journal of Computer Assisted Radiology and Surgery*, *17*(9), 1673–1683. https://doi.org/10.1007/s11548-022-02619-x

[27] Abdalla, P. A., Mohammed, B. A., & Saeed, A. M. (2023). The impact of image augmentation techniques of MRI patients in deep transfer learning networks for brain tumor detection. *Journal of Electrical Systems and Information Technology*, *10*(1), 51. https://doi.org/10.1186/s43067-023-00119-9

[28] Biswas, A., & Islam, M. S. (2023). A Hybrid deep CNN-SVM approach for brain tumor classification. *Journal of Information Systems Engineering & Business Intelligence*, *9*(1), 1–15. https://doi.org/10.20473/jisebi.9.1.1-15

[29] Khalid Alduraibi, A. (2022). A hybrid deep features PSO-ReliefF based classification of brain tumor. *Intelligent Automation & Soft Computing*, *34*(2), 1295–1309. https://doi.org/10.32604/iasc.2022.026601

[30] Agarwal, R., Pande, S. D., Mohanty, S. N., & Panda, S. K. (2023). A novel hybrid system of detecting brain tumors in MRI. *IEEE Access*, *11*, 118372–118385. https://doi.org/10.1109/ACCESS.2023.3326447

[31] Islam, M. N., Azam, M. S., Islam, M. S., Kanchan, M. H., Parvez, A. H. M. S., & Islam, M. M. (2024). An improved deep learning-based hybrid model with ensemble techniques for brain tumor detection from MRI image. *Informatics in Medicine Unlocked*, *47*, 101483. https://doi.org/10.1016/j.imu.2024.101483v

[32] Abbas, Z. K., Alsarray, Z. A., Al-obeidi, A. H. H., & Mutashar, M. R. (2024). Enhancing brain tumor detection: Integrating CNN-LSTM and CNN-BiLSTM models for efficient classification in MRI images. *International Journal of Advanced Technology and Engineering Exploration*, *11*(115), 888–898. https://doi.org/10.19101/IJATEE.2024.111100084

[33] Ojha, B., Maharjan, R., & Acharya, T. (2023). Brain tumor detection using convolutional neural networks: A comparative study. *International Journal on Engineering Technology*, *1*(1), 11–26. https://doi.org/10.3126/injet.v1i1.60896

[34] Meshram, P., Barai, T., Tahir, M., & Bodhe, K. (2023). The brain tumors identification, detection, and classification with AI/ML algorithm with certainty of operations. In *International Conference on Image Processing and Capsule Networks*, 611–628. https://doi.org/10.1007/978-981-99-7093-3_41

[35] Shah, H. A., Saeed, F., Yun, S., Park, J.-H., Paul, A., & Kang, J.-M. (2022). A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet. *IEEE Access*, *10*, 65426–65438. https://doi.org/10.1109/ACCESS.2022.3184113

[36] Farooq, A., Anwar, S., Awais, M., & Rehman, S. (2017). A deep CNN based multi-class classification of Alzheimer's disease using MRI. In *2017 IEEE International Conference on Imaging Systems and Techniques*, 1–6. https://doi.org/10.1109/IST.2017.8261460

[37] Khushi, H. M. T., Masood, T., Jaffar, A., Rashid, M., & Akram, S. (2023). Improved multiclass brain tumor detection via customized pretrained EfficientNetB7 model. *IEEE Access*, *11*, 117210–117230. https://doi.org/10.1109/ACCESS.2023.3325883

[38] Raza, A., Ayub, H., Khan, J. A., Ahmad, I. S., Salama, S., Daradkeh, Y I., . . . , & Hamam, H. (2022). A hybrid deep learning-based approach for brain tumor

classification. *Electronics*, *11*(7), 1146. https://doi.org/10.3390/electronics11071146

[39] Gurunathan, A., & Krishnan, B. (2022). A hybrid CNN-GLCM classifier for detection and grade classification of brain tumor. *Brain Imaging and Behavior*, *16*(3), 1410–1427. https://doi.org/10.1007/s11682-021-00598-2

[40] Mohammed, B. A., Senan, E. M., Alshammari, T. S., Alreshidi, A., Alayba, A. M., Alazmi, M., & Alsagri, A. N. (2023). Hybrid techniques of analyzing MRI images for early diagnosis of brain tumours based on hybrid features. *Processes*, *11*(1), 212. https://doi.org/10.3390/pr11010212

[41] Ali, M. A., Morol, M. K., Mridha, M. F., Fahad, N., Al Huda, M. S., & Ahmed, N. (2023). Exploring a novel machine learning approach for evaluating Parkinson's disease, duration, and vitamin D level. *International Journal of Advanced Computer Science & Applications*, *14*(12). http://dx.doi.org/10.14569/IJACSA.2023.0141265

How to Cite: Al Huda, M. S., Tanvir, K., Akhter, Z., Khan Pappo, M. S., Ali, M. A., & Ahmed, N. (2026). A Novel Explainable Deep Learning Model for Early-Stage Brain Tumor Classification: Multi-Level Feature Fusion from Merged MRI Datasets. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62026182