Received: 21 June 2025 | Revised: 9 September 2025 | Accepted: 22 October 2025 | Published online: 4 November 2025

RESEARCH ARTICLE

Medinformatics
2026, Vol. 3(1) 56–66
DOI: 10.47852/bonviewMEDIN52026540

BON VIEW PUBLISHING

NeuroBlend-3: Hybrid Deep and Machine Learning Framework with Explainable AI for Multi-class Brain Tumor Detection Using MRI Scans

Mohammed Ibrahim Hussain[1], **Safiul Haque Chowdhury**[1,2,*], **Muhammad Minoar Hossain**[1] and **Mohammad Mamun**[1,2]

[1] *Department of Computer Science and Engineering, Bangladesh University, Bangladesh*

[2] *Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh*

Abstract: Brain tumors are complex and potentially life-threatening conditions that require accurate and timely diagnosis. This study proposes NeuroBlend-3, an explainable and hybrid artificial intelligence (AI) framework for multi-class brain tumor classification using magnetic resonance imaging scans. The framework begins with preprocessing steps, including grayscale conversion, resizing to 224 × 224 pixels, normalization, denoising, and enhancement using Contrast Limited Adaptive Histogram Equalization. To increase data variability, five augmented versions of each image are generated through horizontal flip, 15° rotation, zooming, Gaussian blur, and brightness adjustment. Deep features are then extracted using six models: HRNet, VGG16, VGG19, ResNet50, ResNet101, and Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM). These features undergo optimization using principal component analysis and recursive feature elimination (RFE) to reduce redundancy and improve performance. The optimized features train machine learning models, including XGBoost, AdaBoost, Bagging, and a custom Tree Selection and Stacking Ensemble-based Random Forest (TSRF). To ensure interpretability, explainable AI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM++, and Local Interpretable Model-Agnostic Explanations are applied to highlight the regions influencing classification decisions. The combination of CNN-LSTM, TSRF, and RFE demonstrates superior performance across all metrics through extensive experimentation. This best-performing combination is termed NeuroBlend-3. Neuro reflects the neurological focus, Blend denotes the fusion of deep and traditional learning approaches, and 3 signifies the integration of CNN-LSTM, TSRF, and RFE. NeuroBlend-3 offers a robust and interpretable solution, making it highly suitable for clinical decision-making in brain tumor diagnosis.

Keywords: brain tumor, MRI scans, machine learning, deep learning, explainable AI

1. Introduction

Brain tumors are abnormal growths of cells within the brain that can be either benign (non-cancerous) or malignant (cancerous). Benign tumors grow slowly and are less likely to invade surrounding tissues, while malignant tumors are aggressive and can disrupt normal brain functions. Among malignant brain tumors, glioblastoma is notably aggressive, with a median survival rate of approximately 12 months even with treatment, and a five-year survival rate of less than 10 [1]. Magnetic resonance imaging (MRI) is a noninvasive imaging technique that provides detailed brain images, making it a crucial tool for detecting and evaluating brain tumors. MRI scans help assess the size, location, and type of tumor, which are essential in determining the appropriate treatment strategy.

Despite advances in imaging, accurately classifying brain tumors remains clinically challenging due to their heterogeneous appearance, overlapping features among tumor types, and variations in size and location. These challenges make early and precise diagnosis difficult, yet it is critical for selecting effective treatment strategies and improving patient outcomes. Brain tumors and other central nervous system (CNS) cancers are significant contributors to cancer-related mortality worldwide. In 2016, there were approximately 227,000 deaths globally due to CNS cancers. In the United States, an estimated 18,000 people died from brain and other CNS tumors in 2021. In the United Kingdom, around 5,500 deaths occur annually due to brain, other CNS, and intracranial tumors, making it the 10th most common cause of cancer death [2]. Projections indicate that the number of deaths from brain, other CNS, and intracranial tumors in the UK will rise from around 5,700 deaths in 2023–2025 to approximately 6,600 deaths in 2038–2040. Furthermore, brain tumors are the leading cause of cancer-related death among individuals under the age of 40 [3]. These statistics underscore the urgent need for more accurate and interpretable diagnostic tools.

Recent advancements in artificial intelligence (AI) have significantly contributed to addressing the global challenge of brain tumor diagnosis. Researchers have developed innovative AI models that enhance the accuracy and speed of tumor detection. For instance, a study introduced a hybrid model combining Vision

---

[*] **Corresponding author:** Safiul Haque Chowdhury, Department of Computer Science and Engineering, Bangladesh University and Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh. Email: safiul.haque@bu.edu.bd

© The Author(s) 2025. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

Transformer and Gated Recurrent Unit (GRU) architectures, achieving precision, recall, and F1-score metrics of 97 for classifying brain tumors from MRI scans. Another approach utilized a deep reinforcement learning network to predict brain tumor locations, demonstrating the potential of reinforcement learning in medical imaging tasks [4].

In this study, we propose a comprehensive explainable AI (XAI) framework for multi-class brain tumor classification using MRI scans. We utilize a publicly available MRI dataset containing four classes: glioma, meningioma, pituitary tumor, and no tumor. The MRI images undergo extensive preprocessing, including grayscale conversion, resizing to 224 × 224 pixels, normalization, noise reduction, and enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE). To improve model generalization, data augmentation is performed by generating five transformed versions of each image through horizontal flipping, 15° rotation, zooming, Gaussian blurring, and brightness adjustment. Deep features are extracted using multiple deep learning (DL) architectures, including Long Short-Term Memory (LSTM), VGG16, VGG19, ResNet50, and ResNet101. These features are refined using feature optimization techniques such as principal component analysis (PCA) and recursive feature elimination (RFE). The optimized features are then fed into various machine learning (ML) classifiers, including Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), bootstrap aggregating (Bagging), and a custom ensemble model named Tree Selection and Stacking Ensemble-based Random Forest (TSRF).

Model performance is evaluated using accuracy, precision, recall, F1-score, and specificity, derived from confusion matrix (CM) analysis across 50 training epochs. The combination of CNN-LSTM for feature extraction and TSRF for classification yields the highest performance across all metrics. For interpretability, XAI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM++, and Local Interpretable Model-Agnostic Explanations (LIME) are employed to highlight tumor regions influencing the model's decision. Key contributions of this study include:

1) Application of grayscale conversion, resizing, normalization, denoising, CLAHE, and data augmentation to enhance MRI image quality.
2) Extraction of deep features using LSTM, VGG16, VGG19, ResNet50, and ResNet101 for robust tumor pattern recognition.
3) Optimization of features through PCA and RFE to reduce dimensionality and eliminate redundancies.
4) Application of a previously developed custom ensemble model, TSRF, combining tree selection and stacking techniques.
5) Evaluation of models based on accuracy, precision, recall, F1-score, and specificity across 50 epochs.
6) Incorporation of XAI methods like Grad-CAM, Grad-CAM++, and LIME for interpretability.
7) Identify LSTM and TSRF as the optimal combination for the highest classification performance.
8) Contribution to clinical support through an accurate and interpretable AI framework for brain tumor diagnosis.

This paper is organized into five sections. Section 2 reviews the literature, Section 3 presents the materials and methodology, Section 4 discusses the results, and Section 5 concludes the study.

2. Literature Review

Recent advancements in DL have significantly enhanced the accuracy and efficiency of brain tumor detection and segmentation in MRI. Several studies have introduced innovative models that leverage hybrid architectures, attention mechanisms, and advanced feature extraction techniques. This literature review explores these contributions, highlighting their methodologies, performance metrics, and clinical implications. Hosny et al. [5] proposed an ensemble model combining DenseNet121 and InceptionV3 architectures for brain tumor detection and classification. This hybrid approach achieved an accuracy of 99.02, emphasizing the effectiveness of integrating multiple DL models to enhance classification performance while reducing computational complexity. Mugdha and Uddin [6] introduced NeuroSight, a DL framework utilizing the VGG-16 architecture for brain tumor classification. The model demonstrated a test accuracy of 95.52, underscoring the importance of optimizing DL models for generalization across diverse datasets to improve diagnostic reliability. Abraham et al. [7] developed the DC-YOLOv8FEN model, incorporating dilated convolution techniques with the YOLOv8 architecture for enhanced feature extraction in brain tumor detection. Achieving an accuracy of 99.5, the study highlights the model's resilience in low-light conditions, making it suitable for practical clinical applications. Joshi et al. [8] introduced a novel approach combining a Variational Spatial Attention Graph Convolutional Neural Network (VSA-GCNN) with bidirectional GRU for brain tumor segmentation and classification. The model achieved a high accuracy of 99.98, effectively handling complex spatial interactions and addressing class imbalance issues in brain tumor datasets. Preetha et al. [9] proposed a multi-scale attention U-Net model employing the EfficientNetB4 encoder for improved brain tumor segmentation. The model achieved an accuracy of 99.79 and a Dice coefficient of 0.9339, demonstrating the efficacy of integrating advanced attention mechanisms to enhance segmentation performance.

In addition to the above, several newer works have further strengthened the evidence base. Anantharajan et al. [10] proposed a hybrid Ensemble Deep Neural Support Vector Machine (EDN-SVM) that combined deep neural networks with SVMs after Adaptive Contrast Enhancement Algorithm (ACEA)/median-filter preprocessing, Fuzzy C-means (FCM) segmentation, and Gray level Cooccurrence matrix (GLCM) feature extraction; their best model achieved 97.93 accuracy (sensitivity 92, specificity 98), and their conclusion identified limitations in using only grayscale images and called for future work with color and 3D MRI, as well as clinical software integration. Agarwal et al. [11] presented a two-phase system that first enhanced contrast via ODTWCHE and then used a modified Inception-V3–based transfer-learning classifier, yielding a best accuracy of 98.89 on a public dataset; they noted practical limitations related to potential bias and robustness due to dependence on training data and highlighted future work to optimize latency for large-scale screening, improve transparency (e.g., XAI), and consider cloud deployment. Mathivanan et al. [12] evaluated multiple transfer-learning backbones for multi-class MRI classification and reported MobileNetV3 as the best model with 99.75 accuracy (with ResNet152 also performing strongly); their conclusion acknowledged limited real-world validation, possible dataset bias, and unaddressed computational cost and outlined future work to test across diverse datasets and modalities (e.g., CT/PET/ultrasound) to strengthen generalizability. Khaliki and Başarslan [13] compared several Convolutional Neural Network (CNN)/transfer-learning models and found VGG16 to be the best performer at 98 accuracy (Area Under the Curve (AUC) 99, precision/recall 98); no specific limitations or future directions were stated. Finally, Asiri et al. [14] introduced a dual-module pipeline (ICA-NN-SVM) that enhanced images (adaptive Wiener filtering + neural networks + independent component analysis) before SVM-based segmentation/classification; the best configuration achieved 98.9 accuracy (sensitivity/specificity ≈0.99, Dice 0.981) with a mean processing time of 0.43 s, and their conclusion proposed future work to standardize classifiers for robustness, integrate advanced deep models, and explore multimodal imaging to improve generalizability.

By integrating these advanced preprocessing and feature optimization techniques, our study aims to address the limitations observed in previous research, thereby contributing to developing a more robust and accurate model for brain tumor detection and segmentation.

3. Research Methodology

This study uses an Advanced Micro Devices (AMD)-powered system with a Ryzen 9 7900X processor, 32GB of DDR5 6000 MHz RAM, and an NVIDIA GeForce RTX 2080 Ti GPU. The system is utilized for all stages of DL and ML workflows, including model creation, data preprocessing, and fine-tuning. The primary objective of this study is the transparent identification of brain tumors. Figure 1 illustrates the complete workflow of the study from start to finish, while Sections 3.1 to 3.8 provide a concise explanation of each step and describe how each is carried out within the context of this research.

3.1. Dataset

We retrieve a dataset from an online source (Kaggle), namely, PMRAM: Bangladeshi Brain Cancer – MRI Dataset, which is collected from a number of hospitals in Bangladesh [15]. The four main categories in this dataset are glioma, meningioma, pituitary, and no tumor. It consists of raw as well as pre-augmented images (1,502 raw images and 6,000 augmented images). Nonetheless, in this work, we only consider the raw images and use our augmentation methods instead of pre-augmented ones. Thus, all experimentation and processing occur on the 1,502 raw images. The sample distribution of the two classes is shown in Table 1.

3.2. Preprocessing and augmentation

The present study makes use of multiple preprocessing steps to improve image quality and to prepare the data for model training. First, a grayscale transform is performed on the original images, which can help decrease the computation cost by removing color channels, and it concentrates on structural characteristics that are valuable for medical imaging jobs [16]. Next, all the footprints are resized to 224 × 224 pixels, which enables a standard input size for the DL models across the dataset. Normalization is conducted to standardize pixel intensity values to a certain range, for example, from 0 to 1, which can facilitate faster convergence of the model during training [17]. Gaussian blur is employed to reduce noise in MRI images [18]. This smoothing reduces additive random noise while preserving important structural features in the image by averaging pixel values in the image with neighboring pixels based on a Gaussian kernel. Lastly, we increase the contrast of images using CLAHE. CLAHE enhances local contrast and exposes subtle details present in medical images by equalizing histogram levels in the local parts of the image in such a way that the significant features are more distinctive to the model [19]. The effect caused by each preprocessing method used in this work can be visualized in Figure 2, where (A) shows the original MRI image, (B) presents the converted grayscale image, (C) displays the resized image, (D) illustrates the normalized image, (E) depicts the denoised image, and (F) shows the final image after CLAHE enhancement, demonstrating the iterative improvement of image quality.

Once the final CLAHE-enhanced images are obtained, we use five augmentation procedures to further increase dataset variety and model generalization. These are horizontal flip (for orientation), 15-degree rotation (to be invariant to position), scaling (to get internal structures), Gaussian blur (to model variance of image quality), and brightness adjustment (illumination) [20]. Every method creates one additional version of our data per image, resulting in five augmented

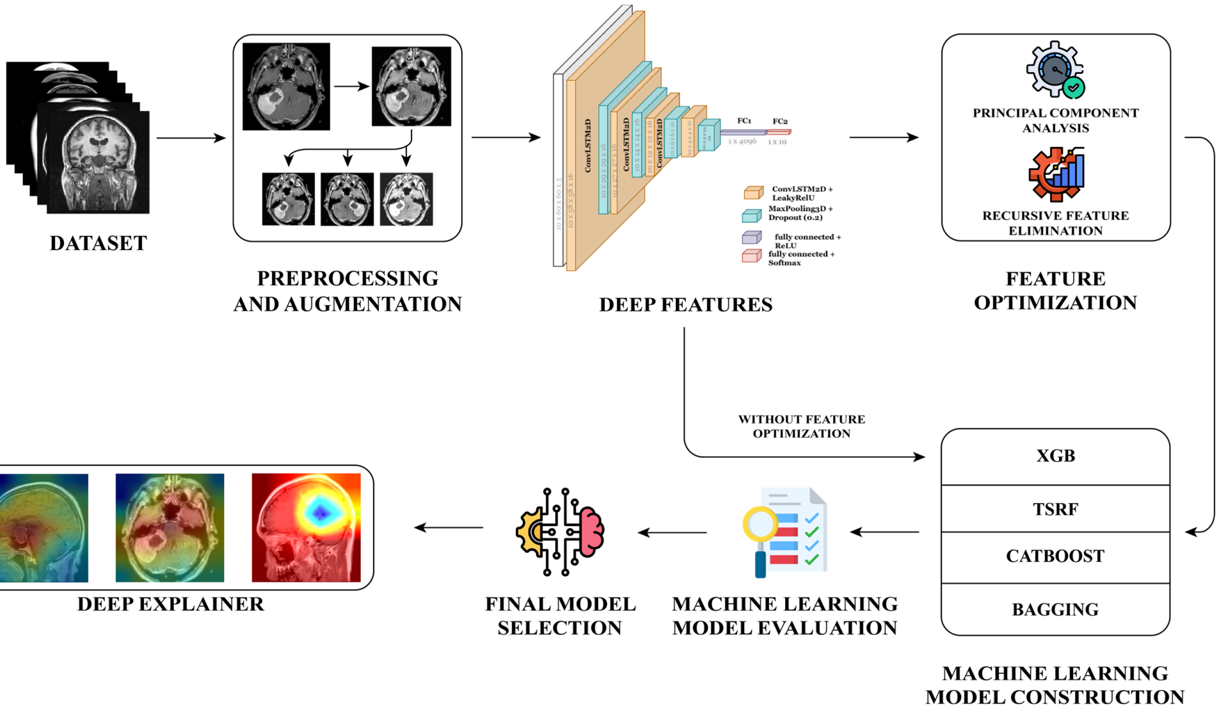
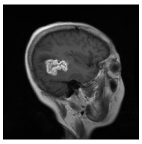
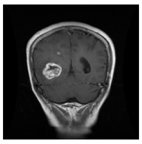
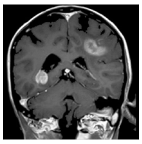
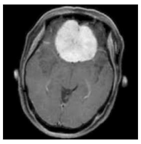
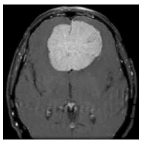
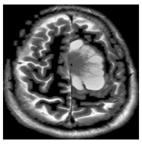
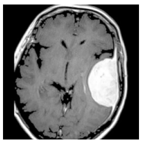
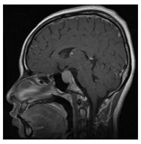
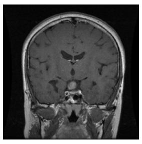
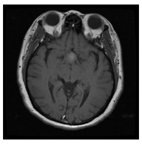
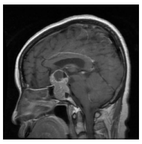
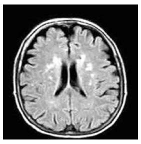
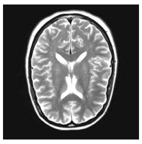
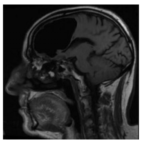
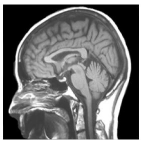


Figure 1. Methodological framework for brain tumor detection

Table 1. Class-wise distribution of raw MRI images in the PMRAM dataset

Categories	Total Sample	Sample 1	Sample 2	Sample 3	Sample 4
Glioma	373				
Meningioma	363				
Pituitary	373				
No Tumor	396				

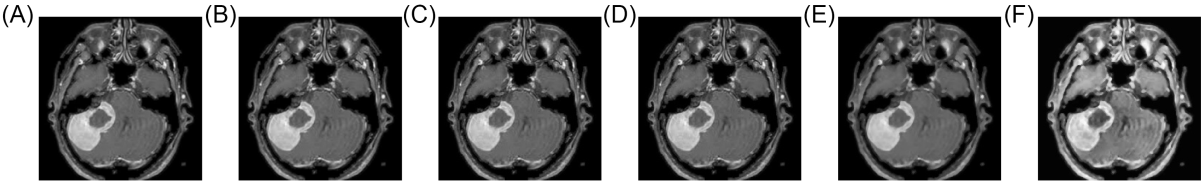


Figure 2. Visual impact of preprocessing techniques on MRI images. (A) Original image. (B) Grayscale image. (C) Resized image. (D) Normalized image. (E) Denoised image. (F) CLAHE-enhanced final image.

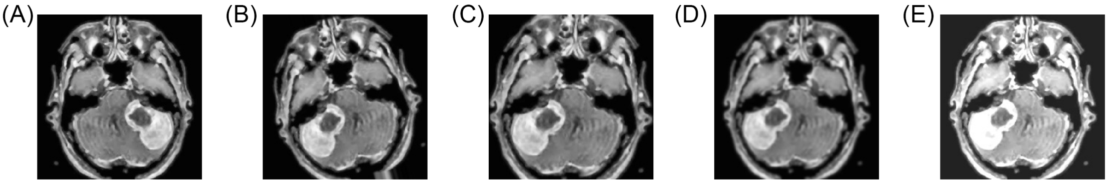


Figure 3. Sample outputs of the five augmentation techniques. (A) Horizontal flip. (B) 15° rotation. (C) Zoom. (D) Gaussian blur. (E) Brightness adjustment.

images for each original sample. The outputs of these augmentation techniques are illustrated in Figure 3, where (A) shows the horizontally flipped image, (B) presents the 15° rotated image, (C) displays the zoomed image, (D) depicts the Gaussian-blurred image, and (E) illustrates the brightness-adjusted image.

By applying the preprocessing techniques to the 1,502 raw MRI images, we create five augmented versions for each image and also the last CLAHE enhancement. This amounts to 7,510 images, leading to a larger and more diverse dataset. The images are divided into training, testing, and validating sets in a 70:20:10 ratio for balanced evaluation and model robustness.

3.3. Deep features

To search for a more powerful deep feature extractor for brain tumor analysis, we independently evaluate five mainstream CNN architectures and a hybrid CNN-LSTM framework by independently feeding each of them to process the preprocessed and augmented MRI images and judge their capability to extract discriminative features. High-Resolution Network (HRNet) can keep high-resolution feature maps, which can retain spatial information and rules in medical imaging (i.e., to find out some minor structure patterns in brain tumors), and is a powerful cooperating first-level feature extractor for two-color image features and also the pixel-level fusion in patches from the whole input image [21]. A deep CNN VGG16 has a simple and uniform architecture and captures hierarchical information from low-level edge information to high-level discriminative classes that are compatible with the complex MRI pattern [22]. VGG19 is similar to VGG16 but with more convolutional layers for extracting deeper features and potentially achieving better classification accuracy [23]. ResNet50 uses the techniques of residual learning using shortcut connections, which allows the relative ease of training deeper networks and learning rich and abstract features commonly used in medical image analysis [24]. ResNet101 further extends this by being highly deep, so that it can learn more interesting nonlinear patterns that are useful for discriminating tumor types [25]. Finally, the hybrid architecture of CNN-LSTM adopted the spatial feature extraction of the CNN and the exploration of contextual and temporal relations along the feature dimensions to increase the network's understanding of spatial dependencies and improve the classification performance [26].

3.4. Feature optimization

We perform feature optimization to improve the performance and the correctness of ML classifiers. The feature vectors obtained by DL models are of high dimensionality, possibly responding to non-informative or redundant dimensions. Such redundant features may cause overfitting, increase the computational cost, and weaken the generalization ability of the model. In order to deal with this issue, we introduce the two most commonly used feature selection and dimensionality reduction methods: PCA and RFE.

PCA is a statistical process to take the original high-dimensional feature space and reduce it to a lower-dimensional space by projecting data along directions that capture the highest variance (principal components). PCA can keep the most important and informative features but discard the noise and redundancy within image features. This is useful for increasing model efficiency and sometimes also for improving accuracy [27].

In contrast, RFE is a wrapper method and is heavily based on a wrapper feature selection method of ranking different features by recursively removing the least important feature, based on an estimator's weight coefficients, like Support Vector Classifier (SVC) or logistic regression. When adapted to the deep pattern learned from an image, RFE built on deep differential chooses the most useful features that dominate the final prediction, leading to model simplification while preserving accuracy [28].

3.5. Machine learning model construction

To evaluate the effectiveness of our approach in identifying brain tumors, we apply a set of ML models both with and without feature optimization. This allows us to observe how feature optimization influences performance and helps identify the most effective combination for accurate prediction. The models used for classification include XGBoost [29], AdaBoost [30], Bagging [31], and TSRF [32]. These ML models are tested on deep features extracted from six different architectures: HRNet, VGG16, VGG19, ResNet50, ResNet101, and CNN-LSTM. For every deep feature set, we perform classification twice – once using the original unoptimized features and again using the features optimized through PCA and RFE. This comparison helps us assess the importance of reducing feature dimensionality and selecting the most relevant attributes for classification.

We aim to determine the most effective pipeline for brain tumor detection by combining each DL model with each ML classifier, both with and without feature optimization. This comprehensive evaluation ensures that we select the best arrangement regarding accuracy and efficiency.

Table 2 presents the key hyperparameter settings used for the ML classifiers (XGBoost, AdaBoost, Bagging, and TSRF), providing transparency and supporting the reproducibility of our experimental framework.

Table 2. Hyperparameter settings used for ML classifiers

Model	Hyperparameter	Value/Range
XGBoost [24]	Learning Rate (η)	0.01–0.3
	Max Depth	3–10
	Subsample	0.6–1.0
	Estimators	100–500
AdaBoost [25]	Estimators	50–300
	Learning Rate	0.1–1.0
	Base Estimator Depth	1–5
Bagging [26]	Estimators	10–200
	Max Samples	0.5–1.0
	Max Features	0.5–1.0
TSRF [27]	Total Trees	57
	Subforest Size	5–15
	Max Depth	4–12

3.6. Machine learning model evaluation

To identify the best combination of DL model, ML classifier, and feature optimization technique, we evaluate the performance of each arrangement using a comprehensive set of metrics. The metrics used in this study include accuracy, precision, recall, F1-score, specificity, and Cohen's kappa. These evaluation metrics provide a balanced understanding of how well each model performs across different aspects of classification.

All metrics are calculated based on a 4 × 4 CM, which summarizes the prediction results by comparing the actual and predicted classes [33]. In a multi-class setting like ours, the CM shows how many samples from each actual class are correctly or incorrectly classified into the four categories: glioma, meningioma, pituitary, and no tumor.

The CM helps identify the overall correctness of a model and where it tends to make specific errors. From this matrix, we compute the following performance metrics by deriving the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for each class. These values form the foundation for metric calculations [34]. Table 3 presents the calculation process of these performance metrics using the TP, TN, FP, and FN values for each class in the multi-class classification setting.

After identifying the best-performing model based on evaluation metrics, we further analyze the final models using statistical measures to ensure the reliability of the results. The standard deviation quantifies the amount of variation or dispersion in the model's performance scores, indicating how consistently the model performs across different test samples [35]. To assess the range within which the true performance metric is likely to

Table 3. Computation of evaluation metrics from CM components

Metric	Formula	Description
Accuracy	$\frac{TP+}{TP+TN+FP+FN} \times 100$	Correct brain tumor predictions
Precision	$\frac{TP}{TP+FP} \times 100$	Of predicted tumors
Recall	$\frac{TP}{TP+FN} \times 100$	Of actual tumors
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100$	Balance between precision and recall
Cohen's Kappa	$p_0 = \frac{TP+TN}{TP+TN+FP+FN}$ $p_{positive} = \frac{(TP+FP)(TP+FN)}{(TP+TN+FP+FN)^2}$ $p_{negative} = \frac{(FN+TN)(FP+TN)}{(TP+TN+FP+FP)^2}$ $p_e = p_{positive} + p_{negative}$ $K_e = \frac{p_0 - p_e}{1 - p_e}$	Prediction match beyond random chance

fall, we calculate the confidence interval, where the lower and upper bounds provide the minimum and maximum values that capture the metric with a given level of confidence [36]. In addition, we apply the bootstrap mean, which involves repeatedly resampling the data with replacement and calculating the mean performance across these resamples, thereby providing a robust estimate of the central tendency while accounting for sampling variability [37]. Together, these measures present a comprehensive statistical view of the model's stability and reliability.

3.7. Final model selection

After applying each ML model in combination with the DL-based feature extractors, we evaluate the performance of all configurations using the selected evaluation metrics. Through this analysis, we identify that the CNN-LSTM model, when paired with the TSRF classifier, consistently outperforms other combinations in accuracy, precision, recall, F1-score, specificity, and kappa.

Based on these results, we selected the CNN-LSTM + TSRF pipeline as the best and final model for brain tumor identification in this study. This configuration provides the most reliable and robust performance, making it the most suitable choice for practical implementation.

3.8. Deep features

To improve the interpretability and trustworthiness of our final CNN-LSTM + TSRF model, we use XAI methodologies, which are indispensable in medical tasks, such as for brain tumor detection, in the sense that the interpretations are as important as the accuracy [38]. To interpret the model decision, we adopt three popular XAI methods, which are LIME, Grad-CAM, and Grad-CAM++. LIME, a technique that is model agnostic and based on local surrogate models, provides importance weights to features by approximating the model locally and computing their contribution to predictions and helps us identify which regions in the extracted MRI features have an impact on the classification. Our method could identify these crucial Regions of Interest (ROIs) [39]. Grad-CAM produces the heatmap by utilizing gradients derived from the last layer of the convolutional layer to highlight those locations in the image that are most important for a model's decision, providing spatial reasoning for the model. Grad-CAM [40], an enhanced version of Grad-CAM, could offer more accurate and detailed visual explanations by recording pixel-level contributions and is warranted in finding small or subtle tumor regions. Such interpretability aids confirmatory (= diagnostic) trust and accountability for the clinician in automated diagnostic systems.

4. Results and Discussion

In this study, we evaluate a variety of DL and ML model combinations to identify the most effective configuration for brain tumor classification. Each combination is tested with and without feature optimization to understand the impact of dimensionality reduction and feature selection on model performance. The outcomes of these experiments are summarized in Table 4, which

Table 4. Comprehensive performance comparison of DL-ML model combinations with and without feature optimization

	DL Model	ML Model	Accuracy	Precision	Recall	F1 Score	Kappa
No Feature Optimization	HRNet	XGBoost	75.80	79.88	79.95	76.88	0.74
		AdaBoost	74.13	76.63	76.80	78.53	0.23
		Bagging	74.79	75.66	75.22	76.47	0.52
		TSRF	77.02	76.30	79.17	79.46	0.47
	VGG 16	XGBoost	76.17	77.50	77.76	78.54	0.62
		AdaBoost	76.24	75.60	75.65	74.36	0.49
		Bagging	77.09	79.00	78.65	79.40	0.61
		TSRF	79.33	76.58	79.81	75.88	0.70
	VGG 19	XGBoost	78.43	77.91	76.78	76.52	0.18
		AdaBoost	75.88	75.43	66.47	75.33	0.51
		Bagging	75.72	77.77	78.60	75.09	0.09
		TSRF	77.59	78.18	78.10	75.05	0.51
	ResNet50	XGBoost	75.24	76.25	75.21	75.12	0.10
		AdaBoost	79.38	77.82	77.39	55.38	0.46
		Bagging	79.14	79.89	77.86	77.02	0.97
		TSRF	74.20	75.39	78.79	74.17	0.50
	ResNet101	XGBoost	75.64	75.28	78.99	74.03	0.80
		AdaBoost	76.48	77.45	76.04	77.93	0.02
		Bagging	76.26	65.42	77.48	76.56	0.27
		TSRF	75.68	79.53	76.41	78.53	0.63
	CNN-LSTM	XGBoost	79.06	79.35	76.80	77.33	0.16
		AdaBoost	65.00	59.99	64.33	55.37	0.58
		Bagging	78.26	74.10	75.29	76.06	0.81
		TSRF	79.82	76.07	74.83	79.47	0.89
PCA	HRNet	XGBoost	80.43	82.61	81.39	83.11	0.89
		AdaBoost	81.15	80.94	82.77	84.00	0.91
		Bagging	82.76	83.15	80.06	81.22	0.90
		TSRF	85.12	84.42	83.65	84.88	0.95
	VGG 16	XGBoost	83.01	82.57	84.11	83.69	0.92
		AdaBoost	81.79	81.44	82.03	81.93	0.88

(*Continued*)

Table 4. (*Continued*)

	DL Model	ML Model	Accuracy	Precision	Recall	F1 Score	Kappa
		Bagging	84.26	83.88	85.36	85.00	0.97
		TSRF	86.44	85.02	83.19	84.60	0.93
	VGG 19	XGBoost	84.93	83.40	84.25	84.17	0.91
		AdaBoost	80.67	/81.77	80.98	80.85	0.87
		Bagging	82.33	82.16	81.25	81.83	0.89
		TSRF	85.18	84.26	83.84	84.79	0.94
	ResNet50	XGBoost	83.89	84.08	82.67	83.34	0.92
		AdaBoost	81.72	83.52	81.02	82.21	0.90
		Bagging	84.11	84.31	83.13	84.02	0.91
		TSRF	85.76	85.63	84.74	85.34	0.96
	ResNet101	XGBoost	83.38	82.41	83.57	83.72	0.88
		AdaBoost	82.56	82.02	81.45	81.86	0.89
		Bagging	84.48	85.01	83.76	84.73	0.94
		TSRF	86.70	86.22	85.35	85.96	0.98
	CNN-LSTM	XGBoost	84.21	83.18	82.43	83.05	0.93
		AdaBoost	82.67	82.88	81.77	82.03	0.90
		Bagging	85.84	86.13	84.92	85.73	0.97
		TSRF	86.93	85.69	86.11	86.41	0.99
RFE	HRNet	XGBoost	86.71	87.05	85.92	86.47	0.93
		AdaBoost	85.28	85.67	86.21	85.94	0.89
		Bagging	88.34	88.01	87.49	87.75	0.96
		TSRF	89.21	88.87	88.74	89.06	0.98
	VGG 16	XGBoost	87.56	86.45	86.81	86.62	0.92
		AdaBoost	86.38	87.92	86.20	87.05	0.91
		Bagging	88.67	89.04	88.35	88.66	0.97
		TSRF	89.42	89.21	88.92	89.06	0.99
	VGG 19	XGBoost	86.11	86.72	87.48	87.10	0.90
		AdaBoost	85.73	86.01	85.64	85.82	0.88
		Bagging	88.12	87.56	88.22	87.89	0.95
		TSRF	89.85	89.43	89.06	89.24	0.99
	ResNet50	XGBoost	87.14	87.91	86.82	87.36	0.93
		AdaBoost	86.83	85.76	86.58	86.16	0.91
		Bagging	89.04	88.52	89.09	88.80	0.96
		TSRF	89.77	89.92	88.90	89.41	0.98
	ResNet101	XGBoost	85.92	86.20	85.84	86.01	0.89
		AdaBoost	86.79	85.92	86.11	86.01	0.90
		Bagging	88.93	88.47	88.60	88.53	0.95
		TSRF	89.68	89.55	89.16	89.35	0.98
	CNN-LSTM	XGBoost	88.26	88.31	87.44	87.87	0.96
		AdaBoost	87.34	87.02	86.23	86.62	0.92
		Bagging	89.57	89.04	88.76	88.89	0.97
		TSRF	99.88	99.85	99.84	98.89	0.99

presents the performance metrics for all configurations and serves as the basis for the following analysis and discussion.

As inferred from the results presented in Table 3, the optimal feature combination for detecting brain tumors is CNN-LSTM with TSRF, generated using RFE. This configuration, named NeuroBlend-3, produces good results with 99.88 accuracy, 99.85 precision, 99.84 recall, 98.89 F1 score, and 0.99 Kappa score. The spatial and temporal dependency between coordinates is learned by the CNN-LSTM model, and the TSRF classifier provides a good generalization performance through robust ensemble decision. RFE also provides efficiency by selecting the most discriminative features. Lastly, a number of XO explainability tools applied to ones and zeros are used to attribute the decision to the machine to improve transparency and trust.

Table 5 presents the performance comparison of the TSRF model using CNN-LSTM features for brain tumor detection, along with significant statistical analysis. We see similar results across the five folds, with accuracy in the range of 99.81 to 99.93, precision between 99.81 and 99.91, recall of 99.79–99.89, F1-score of approximately 98.86 to 98.95, and Cohen's kappa that remains close to 1, indicating substantial predictive agreement. The average values – 99.88 accuracy, 99.85 precision, 99.84 recall, 98.89 F1-score, and 0.99 kappa – reflect the model's strength. The standard deviation is still very small (0.03–0.04), which implies almost no variation among folds and demonstrates the robustness of the performance. The low Confidence Interval Lower (CIL) and high Confidence Interval Upper (CIU) values from the 95 confidence intervals show the robustness and the statistical significance of the results as well. An additional strength of the TSRF is that the bootstrap mean matches the fold-wise means exactly: this highlights the reliability of the prediction of the TSRF model. All the above statistical indices confirm that the TSRF model is accurate and reliable for brain tumor detection.

The CM of NeuroBlend-3 is given in Figure 4. The matrix performs consistently well in all four classes: undifferentiated (glioma), normal, pituitary, and meningioma. It correctly predicts

Table 5. 5-fold cross-validation performance of the TSRF model with CNN-LSTM feature extraction for brain tumor

Folds	Accuracy	Precision	Recall	F1 Score	Kappa
Fold 1	99.87	99.82	99.85	98.92	0.97
Fold 2	99.93	99.82	99.86	98.86	0.99
Fold 3	99.90	99.81	99.79	98.86	0.98
Fold 4	99.89	99.89	99.89	98.86	0.97
Fold 5	99.81	99.91	99.81	98.95	1.04
Average	99.88	99.85	99.84	98.89	0.99
Std. Dev	0.04	0.04	0.04	0.04	0.03
95 CIU	99.84	99.81	99.80	98.85	0.96
95 CIL	99.92	99.89	99.88	98.93	1.01
Boot. Mean	99.88	99.85	99.84	98.89	0.99

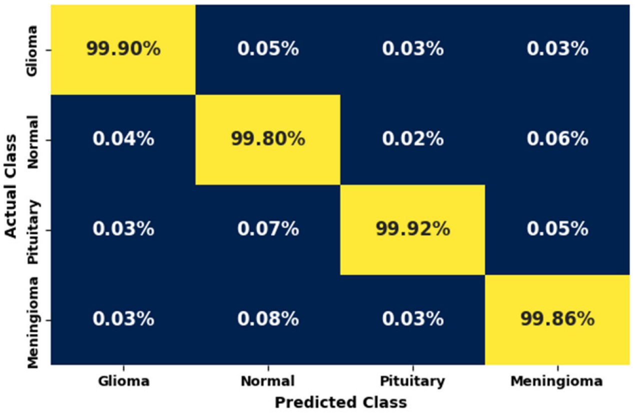


Figure 4. Visual CM (percentage) for NeuroBlend-3

glioma with an accuracy of 99.90, normal with 99.80, pituitary with 99.92 and meningioma with 99.86. The number of misclassifications is low and less than 1 in most of the off-diagonal cells. This shows the capability of the model to discriminate between the types of tumors and the accuracy and reliability of the model for multi-class classification.

The learning curves of the proposed NeuroBlend3 pipeline demonstrate stable and balanced model behavior across 100 epochs. The training and validation accuracy curves (Figure 5(A)) both converge to a high final accuracy of approximately 99.88, with only minimal fluctuations between them. This close alignment indicates that the model generalizes well to unseen data without exhibiting signs of overfitting (where validation accuracy would diverge downward from training accuracy) or underfitting (where both accuracies would plateau at a low level). Similarly, the training and validation loss curves (Figure 5(B)) decrease smoothly and consistently, maintaining a close overlap throughout training. The absence of divergence between the two loss curves further confirms that the model has achieved an optimal balance between learning from the training set and maintaining strong generalization on the validation set. Overall, these results validate the robustness of NeuroBlend3 in effectively learning from the dataset without performance degradation.

Table 6 presents an instance-wise visual interpretation of NeuroBlend-3 predictions (i.e., visualization of the relevant regions in the input image) for three brain tumor classes (glioma, meningioma, and pituitary) based on different heatmap explanation methods: Grad-CAM, Grad-CAM++, and LIME. Each row corresponds to a particular tumor type and presents a depiction of the various interpretability methods and what regions are magnified by the method the most in terms of contributing toward the model's prediction. Examples of MRI images are depicted in Table 6, where the sagittal slice of the brain for the glioma case is illustrated. Grad-CAM directs attention toward a red area of apparently high concentration in the vicinity of the tumor, representing the model's focus. Grad-CAM++ improves this by centering the attention more precisely around the lesion border. LIME, in contrast, conveys a partitioned binary region heat map, which is in accordance with the significance of its surrounding part – that is, people have reasons to believe that the localization of the tumor is highly relevant to the prediction. In

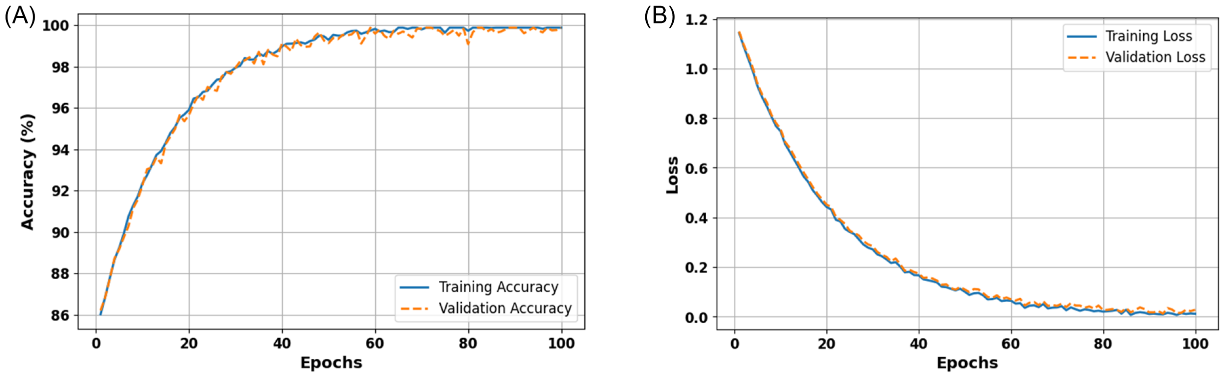


Figure 5. Learning curves of the proposed NeuroBlend-3 pipeline. (A) Training and validation accuracy over 100 epochs. (B) Training and validation loss over 100 epochs.

Table 6. Visual interpretability for NeuroBlend-3

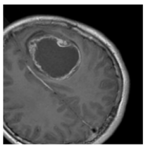
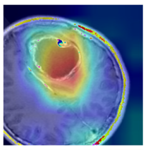
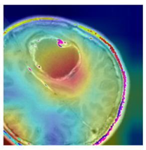
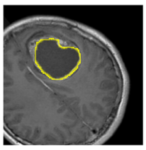
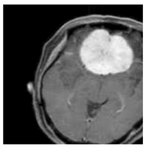
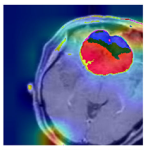
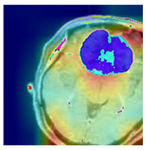
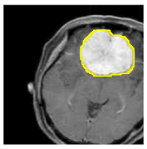
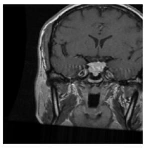
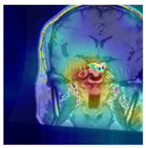
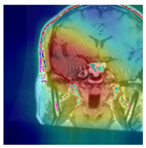
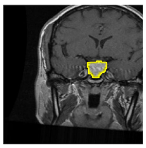
Class	Original	Grad-CAM	Grad-CAM++	Lime
Glioma				
Meningioma				
Pituitary				

Table 7. Benchmarking existing methods against the proposed NeuroBlend-3 framework

Authors	Method	Accuracy	Feature Optimization	XAI	Statistical Analysis
Hosny et al. [5]	DenseNet121 and InceptionV3	99.02	No	Yes	No
Mugdha and Uddin [6]	NeuroSight	95.52	No	No	No
Abraham et al. [7]	DC-YOLOv8FEN	99.5	Yes	No	No
Joshi et al. [8]	VSA-GCNN	99.98	No	No	No
Preetha et al. [9]	U-Net	99.79	No	No	No
Anantharajan et al. [10]	EDN-SVM	97.93	No	No	Standard Deviation
Agarwal et al. [11]	Modified Inception V3	98.89	No	No	Standard Deviation
Mathivanan et al. [12]	MobileNetV3	99.75	No	No	No
Khaliki and Başarslan [13]	VGG16	98	No	No	No
Asiri et al. [14]	ICA-NN-SVM	98.9	Yes	No	Standard Deviation
Our Study	NeuroBlend-3	99.88	Yes	Yes	Standard Deviation, Confidence Interval (Upper and Lower), Bootstrap Mean

the meningioma sample, the original image is a cross-section. Both Grad-CAM and Grad-CAM++ present a wide, ordered activation surrounding the lesion in colors, and Grad-CAM++ provides a more fine-grained level of detail. LIME segmentation on NeuroBlend-3 is spread across multiple patches, indicating that a larger contextual region around the tumor is taken into account by NeuroBlend-3 for correct classification. The original image of the pituitary tumor is also a sagittal slice. Grad-CAM indicates distracting hot spots near the base of the brain, presumably in the vicinity of the pituitary gland. In contrast, Grad-CAM++ has a more localized, sharper focus and is more confident in its original prediction. The LIME map indicates localized areas that correspond to the activated region in the heatmaps. As depicted in Table 6, NeuroBlend-3 successfully finds crucial tumor regions by employing various explainability approaches. Grad-CAM gives a general attention map. Grad-CAM++ gives an improved spatial attention map, and LIME gives model-agnostic interpretable segmentation of the region of interest. These explanations confirm the accuracy of the model predictions and provide a glimpse of potential clinical trustworthiness and transparency.

Table 7 compares the proposed NeuroBlend-3 method with some other brain tumor classification methods. Models such as DenseNet121, InceptionV3, NeuroSight DC, YOLOv8FEN, VSA-GCNN, U-Net, EDN, SVM, Modified InceptionV3, MobileNetV3, VGG16, ICA NN, and SVM are discussed to demonstrate their relative efficacy and methodological improvements. Although some approaches, like VSA-GCNN [8], obtained the highest reported accuracy of 99.98, they did not consider feature reduction, XAI, and rigorous statistical testing, which restricts their clinical interpretability and generalization. Also, methods such as DC YOLOv8FEN [7] and ICA-NN-SVM [14] also used feature optimization, yet their explainability was low, which is important for medical decisions. The proposed NeuroBlend-3, however, achieved an accuracy of 99.88, making it one of the best-performing methods and filling important voids in the existing works. Importantly, it was the first to implement feature optimization, XAI, as well as complex statistical testing (standard deviation, confidence intervals, bootstrap mean). This incorporation allowed high predictive abilities as well as improved transparency and reliability of the model, which enabled it to be more appropriate for real clinical practice. NeuroBlend-3 was thus established as an all-around and practical improvement in the research regarding brain tumor discrimination with respect to performance, interpretability, and statistical robustness.

5. Conclusion

The results of this study show that this AI technique is an efficient method for multi-class brain tumor classification with the use of MRI images, with promising performance for diagnostic accuracy and clinical decision aiding. The framework will allow an in-depth assessment of classification performance, including integration of explainability techniques so that predictions can be more trustworthy and more transparent, leading to better outcomes in neuro-oncology. Most importantly, this study adds to the scientific literature by proposing the NeuroBlend-3 pipeline combining CNN-LSTM, TSRF, and RFE in an innovative manner and shows that the hybrid integration of DL and classic methods could achieve improved diagnostic performance along with interpretability. By demonstrating how explainability can be integrated with high-performing models, the work also adds to the few existing studies on trustworthy AI in healthcare by bridging the divide between model accuracy and clinical interpretability. Yet this study has some limitations, such as using a publicly available dataset collected from different centers that may not encompass all clinical differences, the lack of external validation cohorts, and no consideration for ambiguous or overlapped tumor regions. Future directions in research may include external validation on multi-institutional datasets to increase generalizability, more clinical information, 3D imaging with more in-depth information, and real-time application in the clinical field, in which the model can be updated continually with new data.

Conflicts of Interest

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

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Mohammed Ibrahim Hussain: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing – original draft, Writing – review & editing. **Safiul Haque Chowdhury:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Muhammad Minoar Hossain:** Writing – review & editing, Supervision. **Mohammad Mamun:** Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.

References

[1] Asiri, A. A., Khan, B., Muhammad, F., Rahman, S., Alshamrani, H. A., Alshamrani, K. A., . . . , & Alqhtani, F. F. (2023). Machine learning-based models for magnetic resonance imaging (MRI)-based brain tumor classification. *Intelligent Automation & Soft Computing*, *36*(1), 299–312. https://doi.org/10.32604/iasc.2023.032426

[2] Aamir, M., Rahman, Z., Dayo, Z. A., Abro, W. A., Uddin, M. I., Khan, I., . . . , & Hu, Z. (2022). A deep learning approach for brain tumor classification using MRI images. *Computers and Electrical Engineering*, *101*, 108105. https://doi.org/10.1016/j.compeleceng.2022.108105

[3] Poon, M. T. C., Brennan, P. M., Jin, K., Sudlow, C. L. M., & Figueroa, J. D. (2021). Might changes in diagnostic practice explain increasing incidence of brain and central nervous system tumors? A population-based study in Wales (United Kingdom) and the United States. *Neuro-Oncology, 23*(6), 979–989. https://doi.org/10.1093/neuonc/noaa282

[4] Stember, J., & Shalu, H. (2020). *Deep reinforcement learning to detect brain lesions on MRI: A proof-of-concept application of reinforcement learning to medical images*. arXiv *Preprint*. https://doi.org/10.48550/arXiv.2008.02708

[5] Hosny, K. M., Mohammed, M. A., Salama, R. A., & Elshewey, A. M. (2025). Explainable ensemble deep learning-based model for brain tumor detection and classification. *Neural Computing and Applications*, *37*(3), 1289–1306. https://doi.org/10.1007/s00521-024-10401-0

[6] Mugdha, S. B. S., & Uddin, M. (2025). NeuroSight: A deep-learning integrated efficient approach to brain tumor detection. *Engineering Reports*, *7*(1), e13100. https://doi.org/10.1002/eng2.13100

[7] Abraham, L. A., Palanisamy, G., & Goutham, V. (2025). Dilated convolution and YOLOv8 feature extraction network: An improved method for MRI-based brain tumor detection. *IEEE Access*, *13*, 27238–27256. https://doi.org/10.1109/ACCESS.2025.3539924

[8] Joshi, K. P., Gowda, V. B., Bidare Divakarachari, P., Siddappa Parameshwarappa, P., & Patra, R. K. (2025). VSA-GCNN: Attention guided graph neural networks for brain tumor segmentation and classification. *Big Data and Cognitive Computing*, *9*(2), 29. https://doi.org/10.3390/bdcc9020029

[9] Preetha, R., Jasmine Pemeena Priyadarsini, M., & Nisha, J. S. (2025). Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis. *Scientific Reports*, *15*(1), 9914. https://doi.org/10.1038/s41598-025-94267-9

[10] Anantharajan, S., Gunasekaran, S., Subramanian, T., & Venkatesh, R. (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

[11] Agarwal, M., Rani, G., Kumar, A., K, P. K., Manikandan, R., & Gandomi, A. H. (2024). Deep learning for enhanced brain tumor detection and classification. *Results in Engineering*, *22*, 102117. https://doi.org/10.1016/j.rineng.2024.102117

[12] Mathivanan, S. K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B. D., & Shah, M. A. (2024). Employing deep learning and transfer learning for accurate brain tumor detection. *Scientific Reports*, *14*(1), 7232. https://doi.org/10.1038/s41598-024-57970-7

[13] 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

[14] Asiri, A. A., Soomro, T. A., Shah, A. A., Pogrebna, G., Irfan, M., & Alqahtani, S. (2024). Optimized brain tumor detection: A dual-module approach for MRI image enhancement and tumor classification. *IEEE Access*, *12*, 42868–42887. https://doi.org/10.1109/ACCESS.2024.3379136

[15] Mannan, P. M. S., Chowdhury, M., Rahman, R., Tamim, A. U., & Rahman, M. M. (2024). *PMRAM: Bangladeshi brain cancer - MRI dataset* (Version 1) [Data set]. Mendeley Data. https://doi.org/10.17632/m7w55sw88b.1

[16] Žeger, I., Grgic, S., Vuković, J., & Šišul, G. (2021). Grayscale image colorization methods: Overview and evaluation. *IEEE Access*, *9*, 113326–113346. https://doi.org/10.1109/ACCESS.2021.3104515

[17] Delisle, P.-L., Anctil-Robitaille, B., Desrosiers, C., & Lombaert, H. (2021). Realistic image normalization for

multi-domain segmentation. *Medical Image Analysis*, *74*, 102191. https://doi.org/10.1016/j.media.2021.102191

[18] Bergstrom, A. C., Conran, D., & Messinger, D. W. (2023). *Gaussian blur and relative edge response*. arXiv *Preprint:2301.00856*. https://doi.org/10.48550/arXiv.2301.00856

[19] Liu, J., Zhou, X., Wan, Z., Yang, X., He, W., He, R., & Lin, Y. (2023). Multi-scale FPGA-based infrared image enhancement by using RGF and CLAHE. *Sensors*, *23*(19), 8101. https://doi.org/10.3390/s23198101

[20] Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2023). A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, *137*, 109347. https://doi.org/10.1016/j.patcog.2023.109347

[21] Wu, H., Liang, C., Liu, M., & Wen, Z. (2021). Optimized HRNet for image semantic segmentation. *Expert Systems with Applications*, *174*, 114532. https://doi.org/10.1016/j.eswa.2020.114532

[22] Jiang, Z.-P., Liu, Y.-Y., Shao, Z.-E., & Huang, K.-W. (2021). An improved VGG16 model for pneumonia image classification. *Applied Sciences*, *11*(23), 11185. https://doi.org/10.3390/app112311185

[23] Zhou, J., Ren, K., Wan, M., Cheng, B., Gu, G., & Chen, Q. (2021). An infrared and visible image fusion method based on VGG-19 network. *Optik*, *248*, 168084. https://doi.org/10.1016/j.ijleo.2021.168084

[24] Koonce, B. (2021). *Convolutional neural networks with swift for Tensorflow: Image recognition and dataset categorization*. USA: Apress.

[25] Suo, Y., He, Z., & Liu, Y. (2024). Deep learning CS-ResNet-101 model for diabetic retinopathy classification. *Biomedical Signal Processing and Control*, *97*, 106661. https://doi.org/10.1016/j.bspc.2024.106661

[26] Zha, W., Liu, Y., Wan, Y., Luo, R., Li, D., Yang, S., & Xu, Y. (2022). Forecasting monthly gas field production based on the CNN-LSTM model. *Energy*, *260*, 124889. https://doi.org/10.1016/j.energy.2022.124889

[27] Kurita, T. (2021). Principal Component Analysis (PCA). In K. Ikeuchi (Ed.), *Computer vision: A reference guide* (pp. 1013–1016). Springer International Publishing.

[28] Priyatno, A. M., & Widiyaningtyas, T. (2024). A systematic literature review: Recursive feature elimination algorithms. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, *9*(2), 196–207. https://doi.org/10.33480/jitk.v9i2.5015

[29] Li, J., An, X., Li, Q., Wang, C., Yu, H., Zhou, X., & Geng, Y.-A. (2022). Application of XGBoost algorithm in the optimization of pollutant concentration. *Atmospheric Research*, *276*, 106238. https://doi.org/10.1016/j.atmosres.2022.106238

[30] Favaro, P., & Vedaldi, A. (2021). AdaBoost. In K. Ikeuchi (Ed.), *Computer vision: A reference guide* (pp. 36–40). Springer International Publishing.

[31] Ali, M. M., Anwar, R., Yousef, A. F., Li, B., Luvisi, A., de Bellis, L., . . . , & Chen, F. (2021). Influence of bagging on the development and quality of fruits. *Plants*, *10*(2), 358. https://doi.org/10.3390/plants10020358

[32] Mamun, M., Chowdhury, S. H., Hossain, M. M., Khatun, M. R., & Iqbal, S. (2025). Explainability enhanced liver disease diagnosis technique using tree selection and stacking ensemble-based random forest model. *Informatics and Health*, *2*(1), 17–40. https://doi.org/10.1016/j.infoh.2025.01.001

[33] Liang, J. (2022). Confusion matrix: Machine learning. *POGIL Activity Clearinghouse*.

[34] Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

[35] Xie, J., Liu, X., & Wang, M. (2024). SFKNN-DPC: Standard deviation weighted distance based density peak clustering algorithm. *Information Sciences*, *653*, 119788. https://doi.org/10.1016/j.ins.2023.119788

[36] Riesthuis, P. (2024). Simulation-based power analyses for the smallest effect size of interest: A confidence-interval approach for minimum-effect and equivalence testing. *Advances in Methods and Practices in Psychological Science*, *7*(2), 25152459241240722. https://doi.org/10.1177/25152459241240722

[37] Kostanek, J., Karolczak, K., Kuliczkowski, W., & Watala, C. (2024). Bootstrap method as a tool for analyzing data with atypical distributions deviating from parametric assumptions: Critique and effectiveness evaluation. *Data*, *9*(8), 95. https://doi.org/10.3390/data9080095

[38] Hulsen, T. (2023). Explainable artificial intelligence (XAI): Concepts and challenges in healthcare. *AI*, *4*(3), 652–666. https://doi.org/10.3390/ai4030034

[39] Wang, S., & Zhang, Y. (2023). Grad-CAM: Understanding AI models. *Computers, Materials & Continua*, *76*(2), 1321–1324. https://doi.org/10.32604/cmc.2023.041419

[40] Gaspar, D., Silva, P., & Silva, C. (2024). Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron. *IEEE Access*, *12*, 30164–30175. https://doi.org/10.1109/ACCESS.2024.3368377

How to Cite: Hussain, M. I., Chowdhury, S. H., Hossain, M. M., & Mamun, M. (2026). NeuroBlend-3: Hybrid Deep and Machine Learning Framework with Explainable AI for Multi-class Brain Tumor Detection Using MRI Scans. *Medinformatics*, *3*(1), 56–66. https://doi.org/10.47852/bonviewMEDIN52026540