

RESEARCH

Open Access



Hybrid 2D/3D CNN and radiomics model for brain tumor classification using EfficientNetB0 and ResNet-18

Milad Taleb¹ and Sanaz Alibabaei^{2*}

Abstract

Brain tumors, caused by rapid and uncontrolled cellular proliferation, pose significant health risks worldwide. Accurate and early classification of tumor types such as Glioblastoma, IDH-wildtype, grade 4, Meningioma, and Brain Metastasis is crucial for improving treatment outcomes. This study combines Radiomic features with advanced deep learning models, including EfficientNet-B0 and ResNet-18, to classify 2D and 3D brain tumor images effectively. The classifiers, particularly MLP and CatBoost, demonstrated robust performance, achieving macro F1-scores of 86% and 84%, respectively. However, the imbalanced dataset posed challenges in classification accuracy for less represented classes, emphasizing the need for comprehensive evaluation using AUC and ROC analyses. Additionally, SHAP analysis revealed that texture heterogeneity features, especially original_gldm_DependenceNonUniformityNormalized, consistently contributed significantly across tumor types. Model uncertainty assessment showed that brain metastasis was the most confidently predicted class in both models, meningioma classification had higher confidence in MLP, and Glioblastoma, IDH-wildtype was predicted more reliably by CatBoost. These findings highlight the complementary strengths of combining Radiomics and deep learning for enhanced brain tumor classification and underscore the value of uncertainty quantification in model evaluation.

Clinical trial number

Not applicable.

Keywords Glioblastoma, IDH-wildtype, Meningioma, Brain metastatic, Deep learning, Machine learning, Classification

*Correspondence:

Sanaz Alibabaei
sanazalibabaei994@gmail.com

¹Department of Computer Engineering, K.N. Toosi University of Technology, Tehran, Iran

²Medical Physics Department, Faculty of Medicine, Semnan University of Medical Science, Semnan, Iran



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

A brain tumor is a serious condition caused by abnormal cell growth in the brain, classified as either benign or malignant [1]. Common types include gliomas, originating from neuroglial cells, and meningiomas, arising from brain membranes. High-grade gliomas (e.g., glioblastoma, IDH-wildtype, grade 4) are often malignant, while meningiomas are usually benign (grade 1) but can rarely progress to higher grades. Gliomas are more common in men, whereas meningiomas are more frequent in women [2]. Glioblastoma, IDH-wildtype and an Astrocytoma, IDH-mutant, grade 2/3, are the most aggressive and lethal primary brain tumors [3, 4], accounting for 12–15% of all cases and most common in U.S. adults, with an incidence of 3.21 per 100,000. Their invasive nature and the blood–brain barrier limit treatment effectiveness. Despite advances with temozolomide and radiotherapy, median survival is only 9–16 months [5]. The prognosis for glioblastoma, IDH-wildtype, grade 4 is poor, with a 1-year survival rate of 41.4% and a 5-year rate of 5.8%. Factors like advanced age, incomplete resection, and poor mental performance worsen outcomes, while their opposites slightly improve prognosis [6]. IDH1 and IDH2 mutations are associated with longer survival [7]. Its treatment is complicated by genetic mutations and epigenetic disruptions [4]. Gliomas and brain metastases are the most common brain cancers, contributing significantly to cancer-related deaths. Brain metastases occur in 2% of cancer patients at diagnosis, 12.1% of those with metastatic disease, and up to 15% of patients with unknown primary tumors at the time of cerebral metastasis presentation [8]. Meningiomas, the most common primary brain and spinal tumors, account for 20% of all intracranial tumors. According to WHO 2021, most meningiomas are benign (grade 1). High-grade meningiomas (grades 2 and 3) account for 6%–18% of cases and are associated with higher recurrence rates (30%–80%) and poorer survival outcomes, with a 5-year survival of 78.4% for grade 2 and 44.0% for grade 3 tumors [9]. Meningiomas typically appear on MRI as isointense or hypo intense to gray matter on T1, and isointense or hyperintense on proton density and T2-weighted images, with strong, homogeneous enhancement after gadolinium. They often exhibit a characteristic “tail” sign with peripheral dural thickening [10]. However, distinguishing between atypical or malignant and benign meningiomas via imaging is challenging. Advanced prediction of histopathology is valuable for guiding treatment and prognosis [9, 10]. Artificial intelligence (AI), particularly machine learning, has become crucial in medical analysis and decision-making [11]. Current research focuses on non-omics data, like MRI, CT, and PET scans. However, incorporating omics data, such as gene expression from transcriptomics and epigenomics, provides valuable

insights that enhance diagnosis, treatment planning, and decision-making [4]. Malignant tumors spread rapidly, while benign tumors grow slowly and remain localized, making the latter easier to remove surgically. Effective treatment and follow-up differ, highlighting the importance of classifying brain tumors alongside diagnosis [2]. Magnetic resonance imaging (MRI) is the gold standard for assessing cerebral malignancies, treatment response, and disease progression [8]. Advanced MRI techniques, such as perfusion imaging, MR spectroscopy, and diffusion tensor imaging, help differentiate Glioblastoma, IDH-wildtype, grade 4 from brain metastasis. However, these methods require longer scan times, and their quantitative results can vary with imaging parameters, complicating practical use [12]. A tissue sample is essential to distinguish between benign and malignant brain tumors, but biopsies are not routinely performed on the brain. Accurate diagnosis is crucial, as misdiagnosis can delay treatment and worsen outcomes. Therefore, new diagnostic tools, including machine learning, are needed for pre-surgical tumor classification [2]. In recent times, Radiomics analyzes medical images to predict disease outcomes but requires careful preprocessing. In contrast, deep learning automates feature extraction, simplifying decision-making. In Glioblastoma, IDH-wildtype, grade 4 imaging, deep learning has shown promise for classifying tumors based on genetic mutations and clinical outcomes [12]. In the analysis of medical images, identifying and examining the characteristics of tumors plays an important role in predicting their behavior. These traits can provide valuable information about the nature, type and intensity of the tumor that will help doctors make better decision-making [13]. The integration of hybrid deep learning and transfer learning methodologies has proven to be pivotal in advancing the accuracy of MRI-based classification and segmentation of brain tumors, as demonstrated by recent research [14]. Brain tumors exhibit significant heterogeneity across subtypes, presenting unique challenges in their automated classification. While conventional deep learning models offer strong performance, they often lack interpretability and may fail to capture subtle radiological variations across tumor types [15]. To address these limitations, we propose a hybrid framework that integrates hand-crafted radiomic features with deep neural representations, enabling more comprehensive and interpretable feature learning. Specifically, ResNet18 is employed in combination with 3D radiomics for metastatic tumors, while EfficientNetB0 is used alongside 2D radiomics for meningioma and Glioblastoma, IDH-wildtype, grade 4. This dual-stream, subtype-specific design facilitates tailored feature extraction and fusion. Unlike classical machine learning models such as AdaBoost, CatBoost, and XGBoost, our hybrid approach combining radiomics

features and deep learning representations followed by classification using CatBoost and MLP—achieves improved robustness, particularly for underrepresented tumor types such as brain metastases.

Materials and methods

The ethics committee of the Semnan University of Medical Sciences and Health Services approved the study protocol the ethics committee (Ref. No. 330099540, Ethics code: IR.SEMUMS.REC.1403.329). The MR images of glioblastoma, IDH-wildtype, grade 4, meningioma, and brain metastases (Brain-METs) patients under treatment in the radiation oncology ward of Kosar Hospital (Semnan, Iran) from 2023 to 2025 were used. Tumor types were defined according to the WHO 2021 [16] classification of central nervous system tumors. Patients' informed consent was obtained, and all images were anonymized before use. All patients' data were kept confidential in compliance with the principles of the Declaration of Helsinki.

Brain tumor dataset

Three tumor types—meningioma, glioblastoma (IDH-wildtype, grade 4), and metastatic brain tumors—were retrospectively collected from Kosar Hospital, Semnan (50, 106, and 33 cases, respectively). Preoperative contrast-enhanced T1-weighted MRIs were acquired on a 1.5T Siemens scanner with a 16-channel head and neck coil. Publicly available datasets were included in the training/testing process to increase data diversity, reduce overfitting, and ultimately improve generalizability: 2D meningioma and glioblastoma scans from SartajBhuvaji [17], and 3D metastatic scans from the Stanford [18] and BCBM-RadioGenomics (TCIA) [19] cohorts. Accordingly, our in-house meningioma and glioblastoma scans were reformatted into 2D slices (234 and 625 images), while metastatic scans were retained in 3D (33 cases). Tumor classification was established based on radiotherapy records and imaging characteristics, with confirmation provided through consensus between an oncologist and a neurologist who had direct clinical familiarity with the patients. For the majority of cases, histopathological reports were available in the medical records; however, pathology was not accessible for all patients. Public datasets were similarly reviewed to ensure diagnostic consistency.

Preprocessing

Initially, noise reduction and skull stripping were performed on all images using FSL-BET software (Version 6.0.7.4, developed by the analysis group, FMRIB, Oxford University, England) and Python (3.10). Also, using Python, the images were resized to 224×224 and normalized with MinMaxScaler.

Segmentation

Image segmentation of our Data from Kosar hospital was carried out manually by a radiologist and an experienced neurologist using the open-source software ITK-Snap (version 3.8.0, Jun 12, 2019). Image segmentation was manually conducted with the open-source software ITK-Snap, employing a comprehensive segmentation approach based on the Brain Tumor Segmentation (BRATS) guidelines [20]. 105 ready-made Stanford data [18] and 247 TCIA [19] that had ready-made segments on the site and the SartajBhuvaji data [17] were segmented by the nnU-NetV2 [21] algorithm in such a way that 225 meningioma images and 225 Glioblastoma, IDH-wildtype, grade 4 images were randomly shuffled from the images of Kosar Hospital used for train nnU-Net and 45 images were used for validation. Finally, the tumoral region of 337 Glioblastoma, IDH-wildtype, grade 4 images and 225 meningioma images from the SartajBhuvaji data [17] were segmented with 91% accuracy, and approximately 50 images that were not automatically segmented by the nnU-Net al.gorithm had to be segmented manually by a radiologist using ITK-Snap software. The segmentation and radiomics/deep feature extraction workflow is illustrated in Fig. 1.

Train-test-split

A total of 850 Glioblastoma (IDH-wildtype, grade 4) images, 571 meningioma images (2D), and 385 metastatic brain tumor images (3D) were collected. Table 1 presents the train–test–validation split of the dataset across the three tumor types, showing the number of samples allocated to each subset for every group.

Data augmentation

Image augmentation is a powerful method utilized in Brain tumor detection to improve the accuracy and resilience of machine learning models. The objective is to apply various transformations to the source images to produce new synthetic data samples with realistic variations [14]. To balance the dataset and prevent overfitting, we applied several data augmentation techniques, including random flipping, 90-degree rotations, and intensity shifting, using the MONAI framework [22]. Ultimately, as shown in Fig. 2, these transformations increased the number of training images in each tumor group to 640.

Feature extraction, standardization, and selection

In the group of meningioma and Glioblastoma, IDH-wildtype, grade 4 patients, which consisted of two-dimensional images, feature extraction was performed using radiomics and EfficientNet-B0, while in the group of metastatic brain tumors, which consisted of three-dimensional images, feature extraction was performed using radiomics and ResNet-18. We selected ResNet-18

Tumor Type	Meningioma	Glioblastoma, IDH-wildtype, grade 4	Brain_MET
Actual Image			
Segmentation: nnU-Netv2 and ITK-SNAP			
Feature Extraction Method	EfficientNetB0 & Radiomics (2D Images)	EfficientNetB0 & Radiomics (2D Images)	ResNet18 & Radiomics (3D Images)

Fig. 1 Illustration of the segmentation and radiomics/deep feature extraction process for Meningioma, Glioma, and Brain metastases

Table 1 Train-test split of the dataset across the three tumor types: Meningioma, Glioma, and brain metastases

Tumor Type	All	Train (75%)	Test (12.5%)	Validation (12.5%)
Meningioma	571	429	71	71
Glioblastoma, IDH-wildtype, grade 4	850	640	105	105
Brain-Met	385	289	48	48

The table shows the number of samples allocated to training, testing and validation sets for each group

and EfficientNet-B0 due to their efficiency and suitability for limited data and standard GPU resources. Although GAN-based or transformer models may provide additional performance for larger datasets, they were not feasible in this study and are suggested as future work. A total of 1280 features were extracted by EfficientNet-B0 in the meningioma and Glioblastoma, IDH-wildtype, grade 4 groups and 512 features were extracted by ResNet-18 in the brain metastatic tumors; this number was increased to 1280 features by adding a linear layer. All deep features were extracted from the penultimate

fully connected layer of each network. Additionally, radiomics features were extracted from the segmented tumor regions using the PyRadiomics library, yielding 101 radiomics features per tumor group, including first-order, GLCM, GLDM, GLRLM, GLSZM, and NGTDM categories. The features extracted from deep learning and radiomics algorithms were then merged for each tumor type. Feature standardization was performed using the MinMaxScaler function, resulting in 1381 features from 688 patients per tumor category. To ensure balanced representation across all tumor types and avoid class bias during training, the dataset was balanced before model training, as summarized in Table 2.

To reduce redundancy and enhance the discriminative power, SelectKBest, Sequential Feature Selector (LASSO estimator, alpha = 0.1), and mRMR (Minimum Redundancy Maximum Relevance) algorithms were employed sequentially, and ultimately 88 most relevant features were selected for the classification task, of which 20 features were extracted from the radiomics algorithm and the rest from the Deep algorithm. Among the radiomics

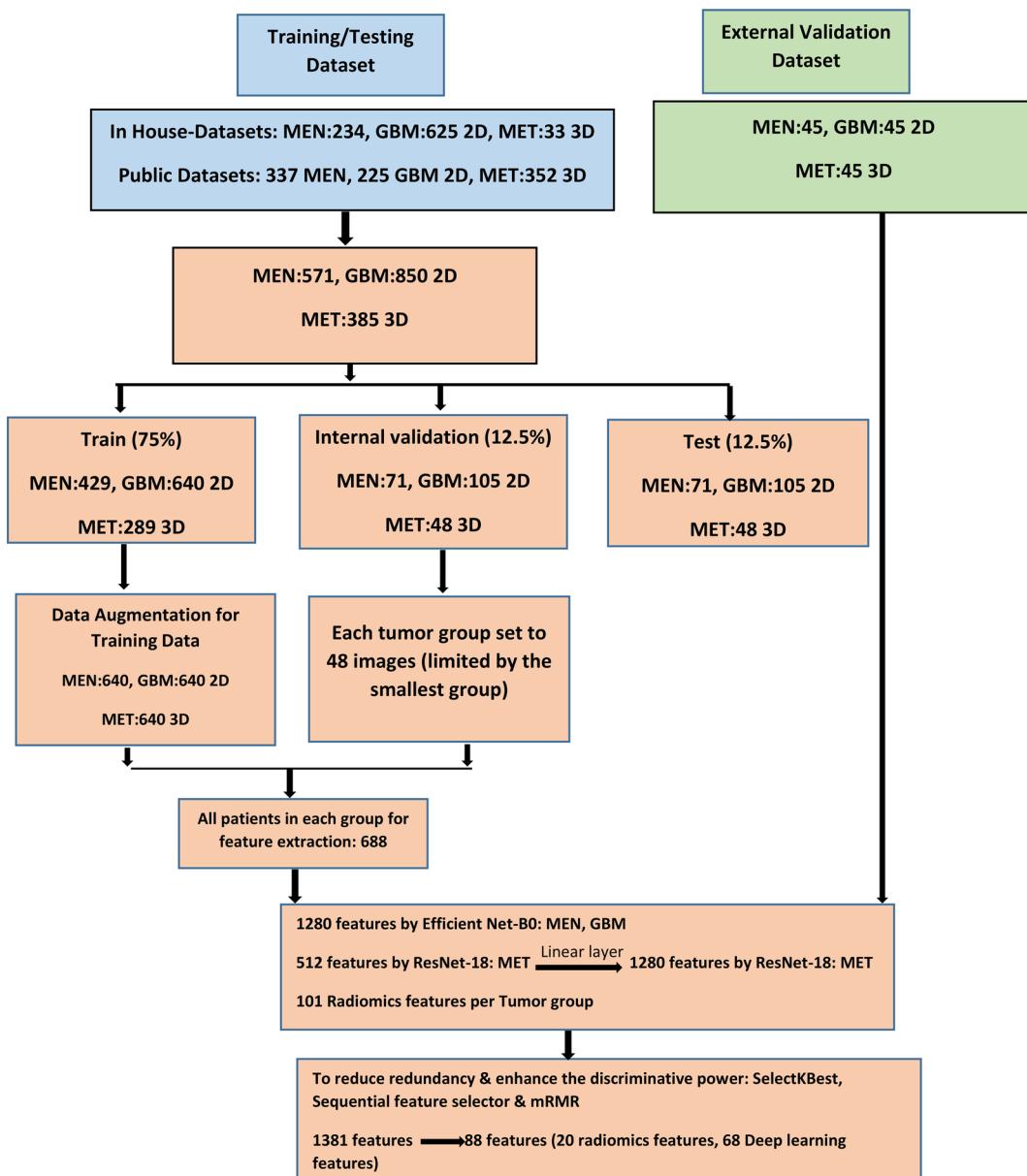


Fig. 2 Data flow diagram showing the number of patients and extracted features for each tumor type

Table 2 Number of samples in each tumor group after balancing the dataset to ensure equal representation of Meningioma, Glioma, and brain metastases

Tumor Type	Label	Train	Validation	All patients in each group for feature extraction
Meningioma	0	640	48	688
Glioblastoma, IDH-wildtype, grade 4	1	640	48	688
Brain-Met	2	640	48	688

features, 20 were selected as the most informative ones, as summarized in Table 3.

Machine learning model training and performance evaluation

Hybrid method combines machine learning (XGBoost, Adaboost, LightGBM, CatBoost) with MLP (Multi Layer Perceptron) [23] model for better classification accuracy. Ensemble methods, which aggregate multiple model predictions, also improve accuracy and reduce variance, benefiting clinical applications [14]. Tuning of the models was done on Adaboost, CatBoost, XGBoost, LightGBM using grid search CV based validation [24] (with 88 features). Each classifier was trained for 10 epochs with data

Table 3 The 20 selected radiomics features categorized into four main types: First-order, GLDM, GLRLM, and diagnostic features

No.	Feature Categories	Features
1	first order	original_firstrder_Maximum/ original_firstrder_RootMeanSquared original_firstrder_Skewness/original_ firstrder_90Percentile/ original_firstrder_Mean/ original_firstrder_Range/original_firstrder_Me dian/ original_firstrder_MeanAbsoluteDeviation
2	GLDM	original_gldm_DependenceEntropy/ original_gldm_DependenceVariance/ original_gldm_LargeDependenceEmphasis/ original_gldm_LargeDependenceLowGray- LevelEmphasis/ original_gldm_DependenceNon- UniformityNormalized/ original_grlm_ShortRunLowGrayLevelEmphasis
3	GLRLM	original_grlm_ShortRunEmpha sis/original_grlm_RunPercentage/ original_grlm_ShortRunHighGrayLevelEmphasis
4	Diagnostic Features	Diagnostic image interpolated maximum/ Diagnostic mask interpolated maximum/ diagnostics_Mask-interpolated_Mean

Table 4 Optimized hyperparameters for each ensemble model obtained through grid search during cross-validation

Model	Hyperparameters
AdaBoost	estimator:DecisionTreeClassifier(max_depth=1) n_estimators: 300 learning_rate: 0.1
XGBoost	n_estimators: 100 learning_rate: 0.3 max_depth: 2
LightGBM	n_estimators: 300 learning_rate: 0.1 max_depth: 10
CatBoost	iterations: 300 learning_rate: 0.05 depth: 6

shuffling applied in each epoch. To ensure optimal model performance, a grid search strategy was applied during cross-validation to tune the hyperparameters of each ensemble model. The optimized parameters obtained for each model are summarized in Table 4.

The architecture consisted of Input layer with 88 feature 4 hidden layers, LeakyReLU and ReLU activation functions [25], a dropout layer (dropout rate = 0.3 and 0.2) after each hidden layer and Output layer with 3 neurons. The model was trained using the AdamW optimizer [26] (learning rate = 0.001) and categorical cross-entropy loss Data. 100epochs were used for MLP training. The evaluation metrics included the ROC curve, confusion matrix, and F1-score report. An overview of the proposed framework, including data preprocessing, segmentation, feature extraction, and classification, is presented in Fig. 3.

Statistical comparison of model performance

Table 5 presents the pairwise comparison of different machine learning models. P-values were calculated using DeLong's test to assess whether the differences in AUC between two models are statistically significant. Values smaller than 0.001 are reported as < 0.001 to highlight highly significant differences. Most comparisons are statistically significant (P-value < 0.05), except for AdaBoost vs. LightGBM and XGBoost vs. LightGBM, where no significant difference was observed.

External validation

For external validation, we used a total of 135 images, comprising 45 2D images each of Meningioma and glioblastoma, IDH-wildtype, grade 4 from the Mendeley dataset [27], and 45 3D images of metastatic tumors from The University of California San Francisco Brain Metastases Stereotactic Radiosurgery (UCSF-BMSR) MRI Dataset [28]. All preprocessing steps applied to the training data were similarly applied to this independent cohort to ensure consistency. This external validation set was held out entirely from the training and testing process and used solely to evaluate model generalizability.

Model uncertainty and explainability

To better understand the decision-making process and reliability of our classification models, we conducted an explainability analysis using SHAP [29] (SHapley Additive exPlanations) and an uncertainty analysis using entropy-based [30] measures.

SHAP-based explainability

To investigate the contribution of each radiomic feature to the classification decisions, we employed SHapley Additive exPlanations (SHAP), a game-theoretic approach to model interpretation. SHAP values quantify the marginal contribution of each feature by approximating Shapley values from cooperative game theory [31].

For the CatBoost model, TreeSHAP was used to compute exact SHAP values based on the model structure [32]. For the MLP model, KernelSHAP was employed due to its model-agnostic nature, allowing estimation of feature importance for neural networks [33]. SHAP summary plots were generated for each model to highlight the most influential features contributing to the classification of glioma, meningioma, and brain metastases.

Entropy-based uncertainty

To quantify model uncertainty, we utilized Shannon entropy as a measure of prediction confidence. The entropy of the predicted class probabilities was calculated for each test sample as follows Eq. 1 [34].

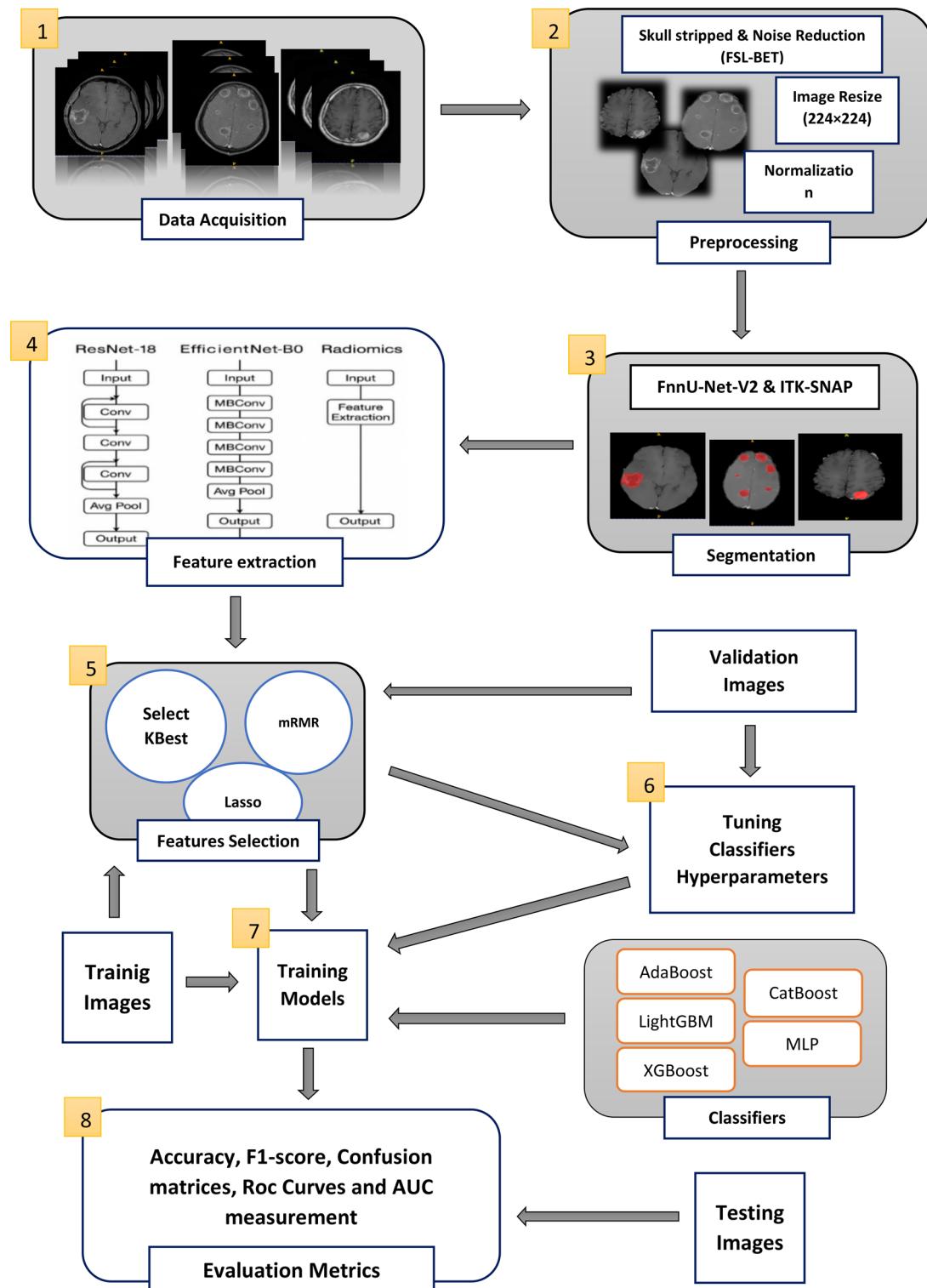


Fig. 3 Complete workflow of the proposed framework, illustrating all major steps from data preprocessing and tumor segmentation to feature extraction and final classification

Table 5 Pairwise comparison of machine learning models using AUC with delong's test for statistical significance

Model1	Model2	p-value (DeLong)
MLP	CatBoost	0.008
MLP	AdaBoost	<0.001
MLP	XGBoost	<0.001
MLP	LightGBM	<0.001
CatBoost	AdaBoost	<0.001
CatBoost	XGBoost	<0.001
CatBoost	LightGBM	<0.001
AdaBoost	XGBoost	0.044
AdaBoost	LightGBM	0.598
XGBoost	LightGBM	0.152

$$H(x) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

where $p(x_i)$ represents the predicted probability for class i, and (n) is the total number of tumor classes. A lower entropy value indicates higher model confidence. Entropy distributions were computed separately for the MLP and CatBoost models, and visualized to compare the uncertainty patterns across the three tumor types.

Results

A comparative evaluation of classification models based on the macro F1-score revealed that the Multi-Layer Perceptron (MLP) and CatBoost outperformed the other approaches, achieving scores of 86% and 84%, respectively. LightGBM demonstrated moderate performance with a macro F1-score of 70%, while AdaBoost and XGBoost exhibited the lowest scores, at 65% and 61%, respectively. The macro F1-score, which calculates the unweighted mean of F1-scores across all classes, is particularly appropriate for multi-class problems, as it ensures balanced assessment regardless of class distribution.

To further assess the discriminatory power of the models, Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) values were analyzed (Fig. 4). Both the MLP and CatBoost classifiers exhibited outstanding ability in differentiating brain tumor subtypes, with AUC values of 1.00 for the Brain Metastasis class. Interestingly, despite LightGBM also achieving an AUC of 1.00 for this class, it misclassified 11 out of 48 Brain Metastasis samples. This highlights a critical nuance in model evaluation: AUC, being threshold-independent, measures the model's ranking capability rather than its actual classification accuracy at a fixed threshold. Consequently, a perfect AUC may still

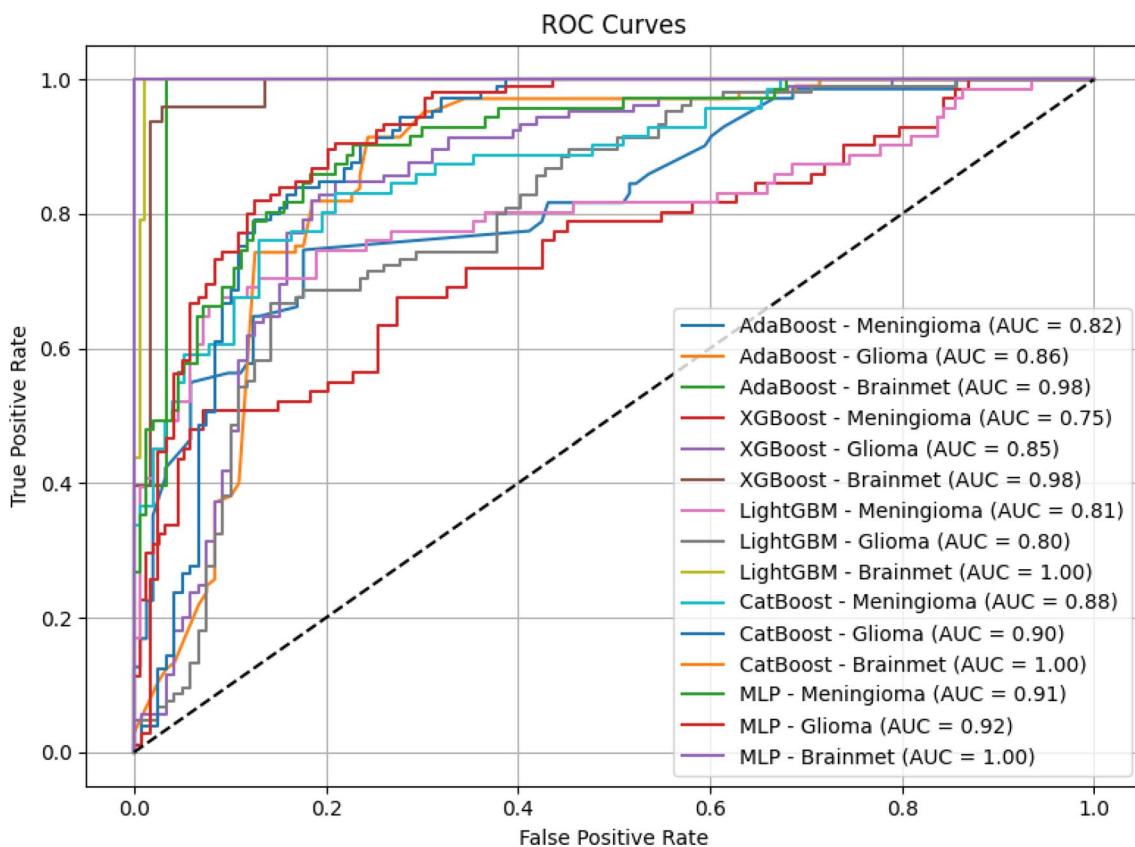


Fig. 4 ROC curves and corresponding AUC values for each classification model across the three tumor classes: Meningioma, Glioma, and Brain Metastases

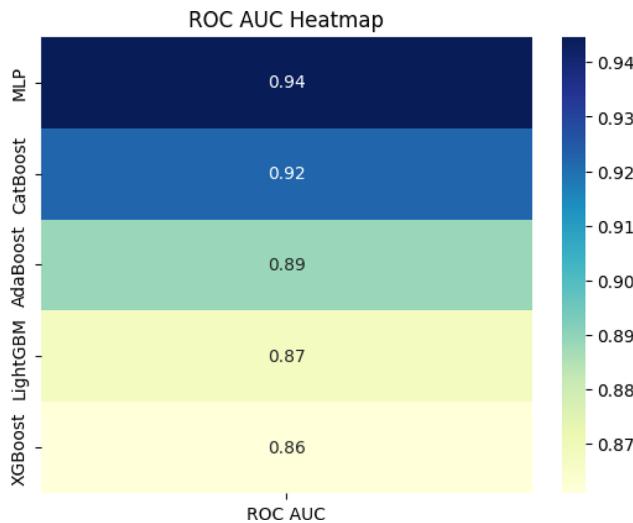


Fig. 5 Comparison of ROC AUC scores among all evaluated models using a heatmap. Higher AUC values indicate better overall classification performance

coexist with considerable misclassification under a specific decision threshold.

Moreover, the imbalance in the test dataset (48 Brain Metastasis, 105 Glioblastoma, IDH-wildtype, grade 4, and 71 Meningioma)—in contrast to the balanced training set—likely impacted threshold-dependent metrics such as accuracy and F1-score. This emphasizes the importance of employing robust, threshold-independent metrics like AUC, especially in real-world medical datasets where perfect class balance is rarely achievable.

As shown in the ROC heatmap (Fig. 5), MLP and CatBoost consistently yield high AUC scores across all tumor subtypes, confirming their robustness. In contrast, XGBoost showed lower discriminative ability for

Meningioma ($AUC = 0.75$), indicating potential limitations in generalizability.

The uncertainty plots (Fig. 6) reveal that MLP provides more confident predictions with lower entropy across most instances, particularly for the Brain Metastasis class, suggesting potential for reliable clinical deployment in high-risk scenarios.

Overall, MLP and CatBoost models demonstrated the highest performance across all evaluation metrics, including AUC, accuracy, F1-score, and uncertainty estimates. Furthermore, their feature interpretability via SHAP values suggests strong potential for deployment in clinical diagnostic workflows.

To gain deeper insights into the instance-level classification performance, Fig. 7 presents the confusion matrices for all models.

To provide an overall comparative visualization of AUC scores across models and classes, a ROC heatmap was generated.

The following results, summarized in Table 6, report the performance of the proposed models on the independent external validation cohort, which was completely excluded from the training and internal testing process.

As summarized in Table 6 MLP and CatBoost achieved the highest performance on the independent external validation cohort, with Accuracies of 84% and 80%, respectively, and the highest Macro F1-scores among all models. These results are further supported by the corresponding confusion matrices and ROC curves (Figs. 8 and 9), demonstrating more reliable classification and better generalizability compared to the other models.

To analyze the confidence levels and predictive uncertainty of each model, entropy-based uncertainty estimates were computed for the internally trained models.

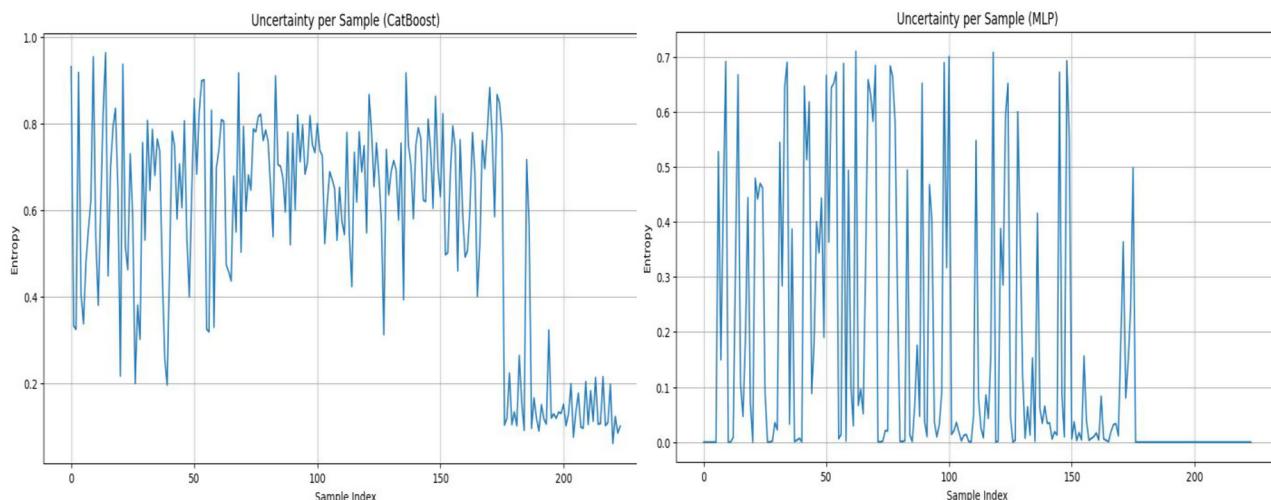


Fig. 6 Entropy-based uncertainty per sample for the test set. The left plot corresponds to the CatBoost model, and the right plot to the MLP model. Each point represents a sample's entropy score, reflecting the model's prediction confidence. Samples are grouped by tumor type: meningioma, glioma, and brain metastasis (from left to right along the x-axis)

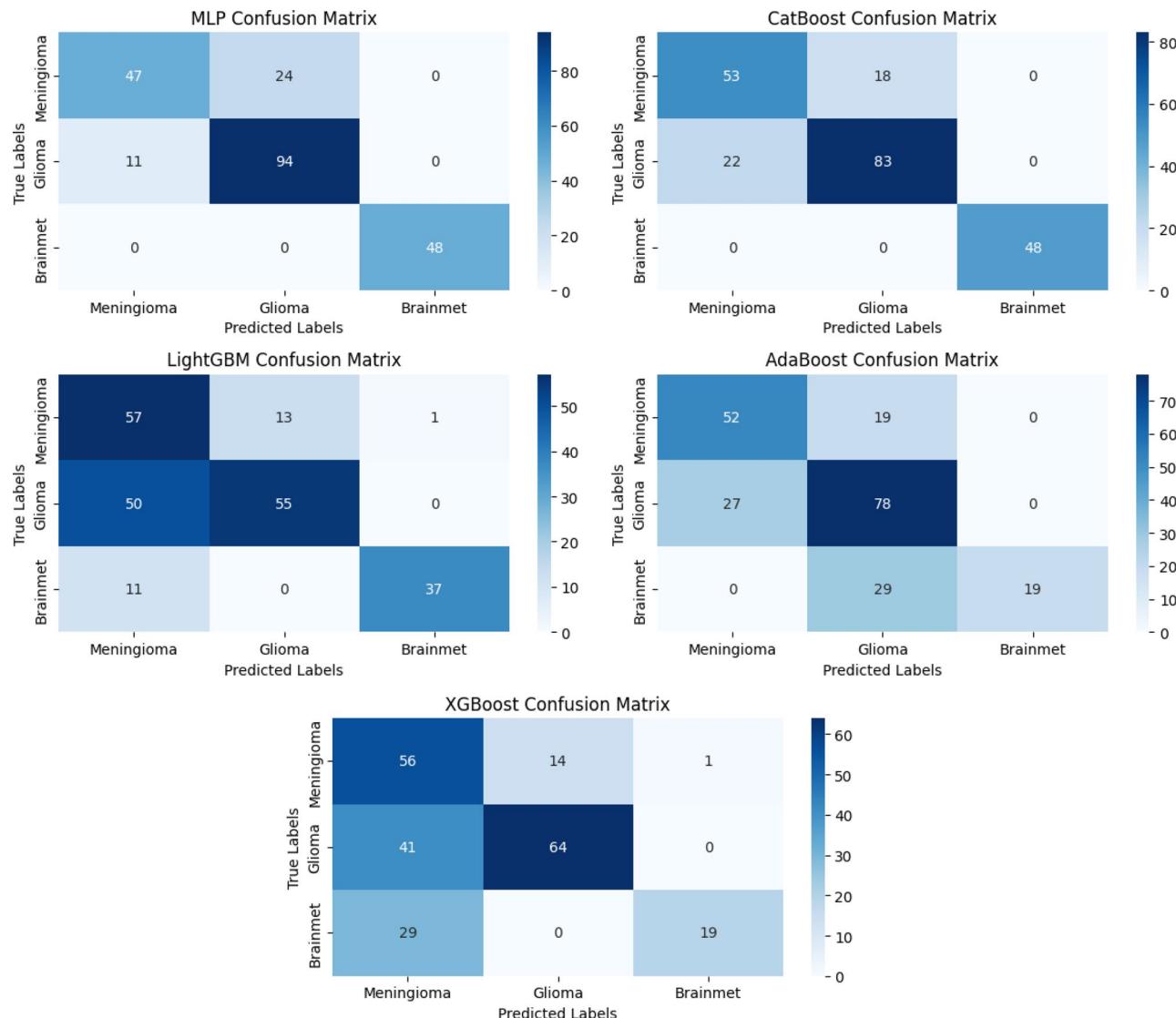


Fig. 7 Confusion matrices showing the performance of five different models in classifying meningioma, glioma, and brain metastases. Diagonal elements indicate the number of correctly classified cases for each tumor type, while off-diagonal elements represent misclassifications

Table 6 Performance comparison of classification models on the external validation cohort

Models	Accuracy	Macro F1-score
AdaBoost	79	76
XGBoost	71	64
LightGBM	77	75
CatBoost	80	78
MLP	84	83

To interpret the contribution of individual radiomic features in the classification decision, SHAP (SHapley Additive exPlanations) values were analyzed for the internally trained models, as illustrated in Fig. 10.

SHAP analysis was employed to interpret the contribution of radiomic features in classifying Glioblastoma, IDH-wildtype, grade 4, Meningioma, and Metastatic

brain tumors using both CatBoost and MLP models. In both models, original_gldm_DependenceNonUniformityNormalized consistently emerged as one of the most impactful features across all tumor types, highlighting the importance of texture heterogeneity. For Glioblastoma, IDH-wildtype, grade 4, the CatBoost model emphasized first-order features like Median, Skewness, and Mean, while the MLP model reinforced the role of similar intensity-based metrics along with DependenceVariance and LargeDependenceEmphasis, suggesting complementary reliance on both intensity and texture information. In Meningioma, both models highlighted Skewness, MeanAbsoluteDeviation, and Mean, with MLP additionally showing influence from RootMeanSquared and several diagnostic mask features, implying texture and intensity spread are critical. For Metastatic tumors,

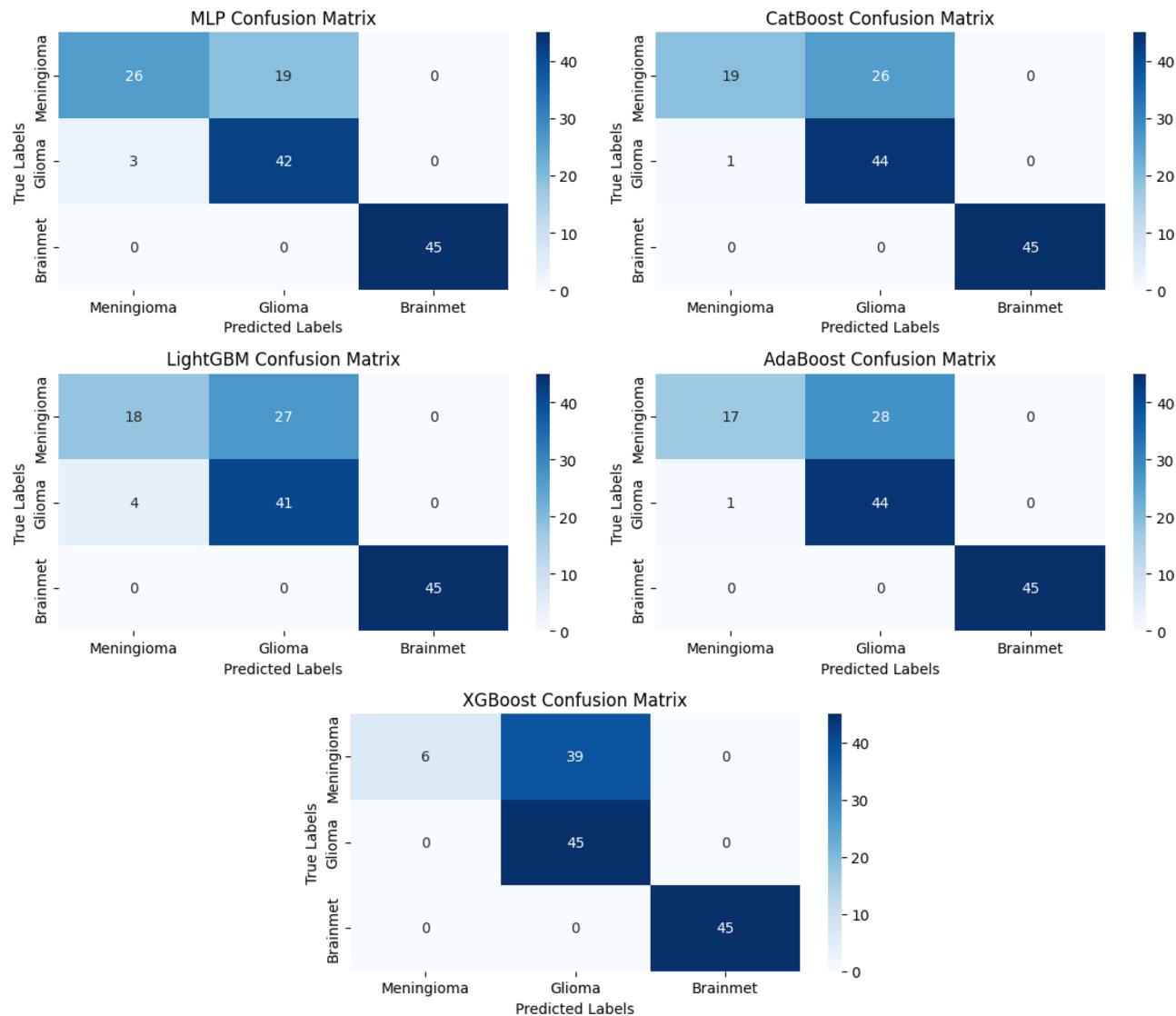


Fig. 8 Confusion matrices for the independent external validation cohort

CatBoost focused on a mix of Median, Skewness, and RunPercentage, while MLP revealed a strong impact from both first-order (Mean, MeanAbsoluteDeviation) and high-variance texture features like DependenceEntropy. These consistent yet slightly differing patterns across models confirm the robustness of key radiomic features, and suggest that combining classifiers may enhance diagnostic performance in brain tumor classification.

Discussion

In this study, our goal was to develop a hybrid deep learning model for classifying brain metastases, meningiomas, and glioblastoma (IDH-wildtype, grade 4). We employed EfficientNet-B0, ResNet-18, and radiomics features to extract critical information, while adopting a hybrid design of 3D CNNs for brain metastases and 2D CNNs for meningiomas and glioblastoma. The rationale behind

this choice lies in the structural and clinical characteristics of these tumors as well as data availability. Brain metastases often exhibit complex three-dimensional shapes and variable spatial distribution, making 3D feature extraction more effective. In contrast, meningiomas and glioblastomas typically present with more localized patterns, and the limited sample size favors the use of 2D CNNs to reduce the risk of overfitting. Although differentiating meningiomas from glioblastomas is generally straightforward for experienced radiologists, certain cases with atypical imaging appearances or overlapping features may still pose diagnostic challenges. Our approach is not intended to replace expert judgment but rather to provide an automated decision-support tool that can reinforce diagnostic confidence in borderline or complex cases and streamline the clinical workflow. Considering the significant advancements in brain

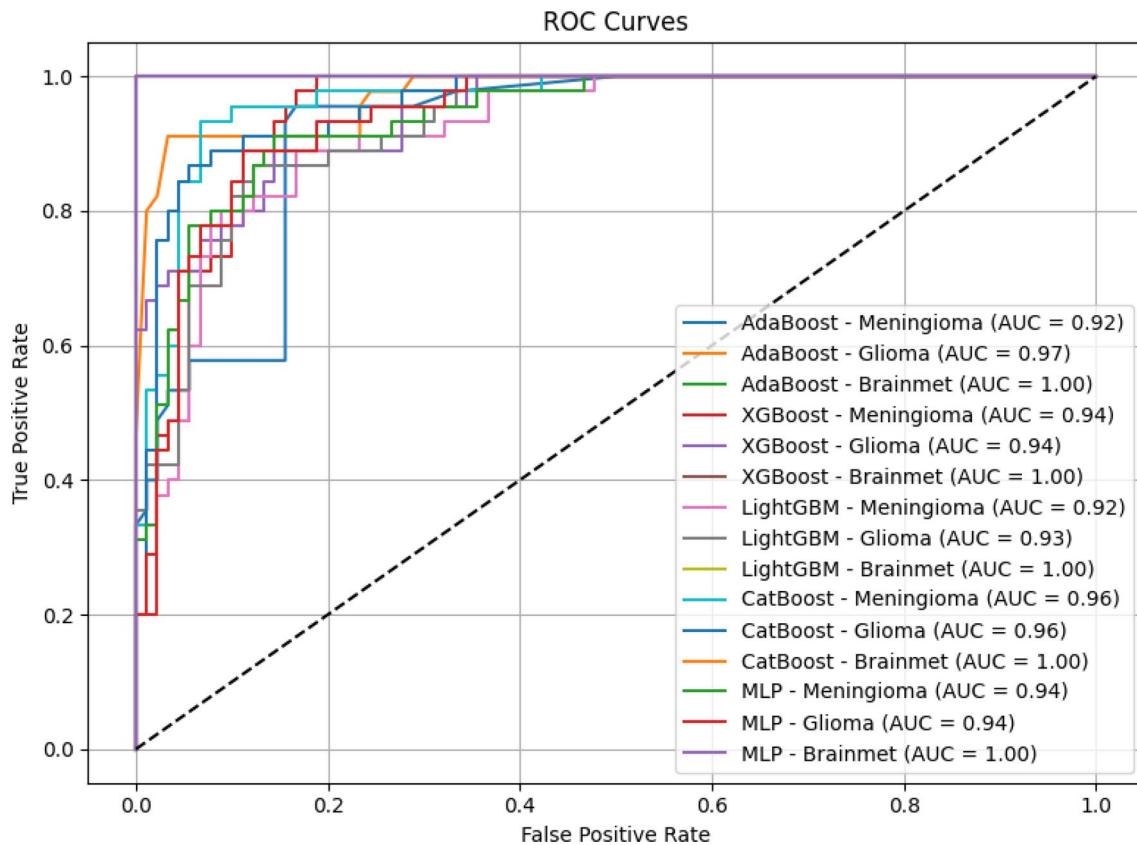


Fig. 9 ROC curves and corresponding AUC values for the independent external validation cohort

tumor analysis driven by deep learning, we further compared our findings with existing studies to highlight the improvements achieved in our work.

Ahsan et al. (2025) [1], showed that YOLOv5 achieved the best detection and segmentation performance for brain tumors, but its applicability may be dataset-specific, which limits broader use. Akter et al. (2024) [35], proposed a deep CNN-based model for automatic brain tumor classification and segmentation using MRI images. Tested on six datasets, their model outperformed pre-trained models, achieving up to 98.8% accuracy with segmentation. This framework shows promise for clinical applications in tumor identification and segmentation.

In the study of Amin et al. (2024) [36] An unsupervised clustering approach for tumor segmentation was introduced, using a fused feature vector of GWF, HOG, LBP, and SFTA features. A Random Forest classifier differentiates tumor regions (complete, enhancing, and non-enhancing). Cross-validation techniques (five-fold and 0.5 holdout) were employed to prevent overfitting, demonstrating the approach's promising detection efficiency.

Sandhiya and Kanaga Suba Raja (2024) [37] introduced an improved PSO-KELM model for brain tumor classification by integrating deep learning (Inception V3, DenseNet201) with radiomic features, achieving

high accuracies of 96.17% and 97.92% on two datasets. In contrast, our hybrid method combines radiomic and deep learning features with classifiers such as MLP and CatBoost, reaching promising macro F1-scores of 0.86 and 0.84. Although slightly lower in performance, our approach emphasizes the potential for further optimization. Unlike complex, resource-intensive models, our lightweight integration of EfficientNet-B0 and ResNet-18 with radiomics offers a practical balance between accuracy and computational efficiency, making it suitable for clinical applications.

Karthik et al. (2025) [38] proposed a model integrating Attention-Augmented CNN, Random Forest, and U-Net to combine attention mechanisms, ensemble learning, and semantic segmentation. It achieves high classification accuracy and precise tumor segmentation in MRI brain images. Experimental results demonstrate strong performance metrics (recall, precision, accuracy, F1-score), showcasing its potential for enhancing brain tumor analysis in medical imaging.

Saeed et al. (2024) introduced a three-phase strategy for brain tumor segmentation and classification. Tumors were first segmented using the DeeplabV3+ model, with hyperparameters optimized via Bayesian optimization. For classification, features were extracted from

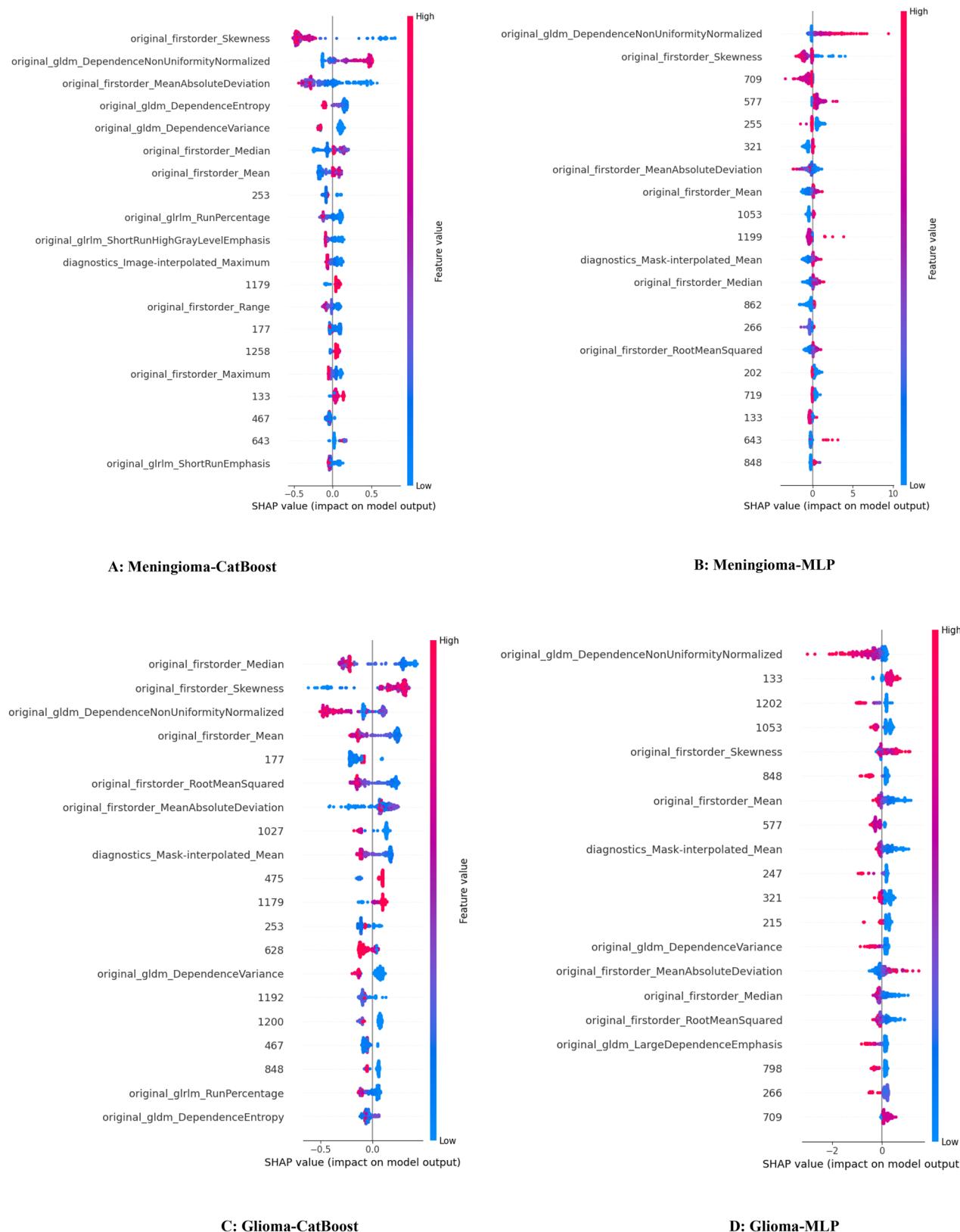


Fig. 10 SHAP summary plots illustrating the contribution of deep learning and radiomic features in classifying (a) Glioma, (b) Meningioma, and (c) Metastatic brain tumors using CatBoost and MLP models

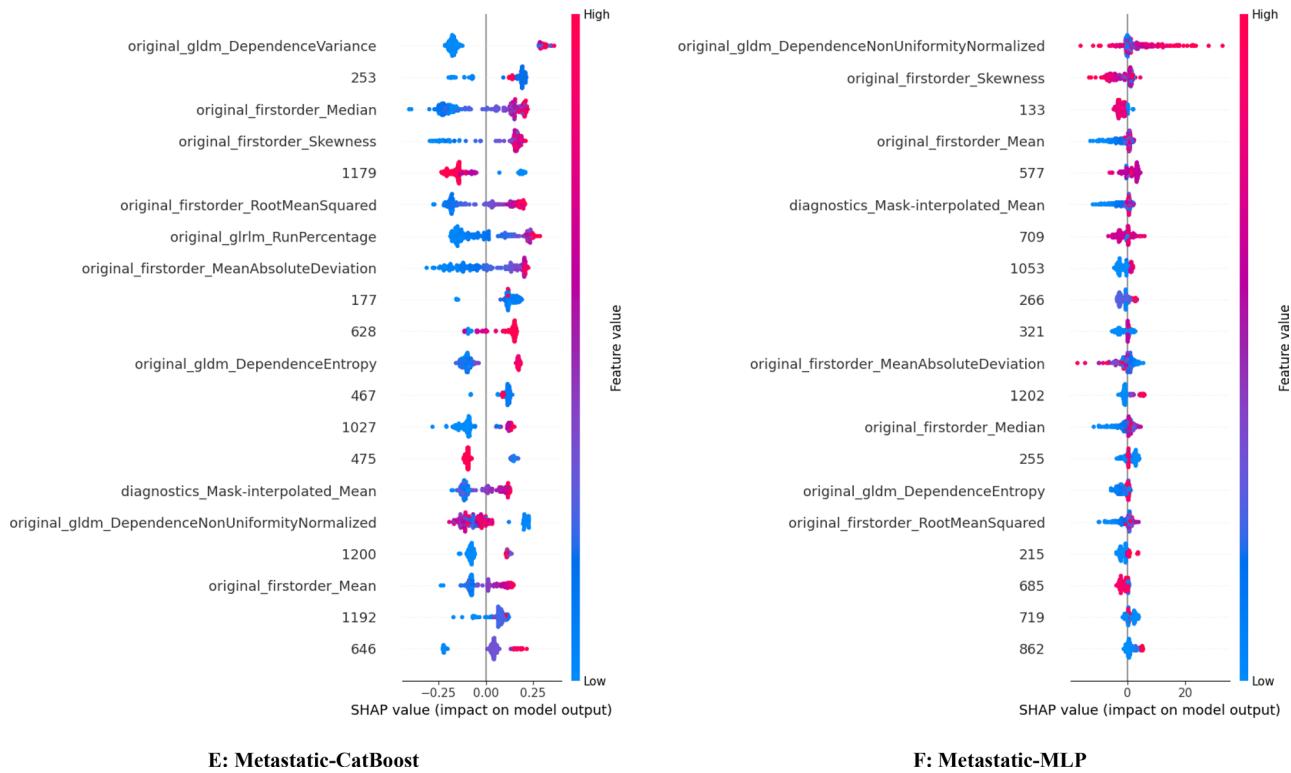


Fig. 10 (continued)

Darknet53 and MobileNetV2 and classified by an SVM, with hyperparameter optimization. Explainable AI methods were applied to interpret the model's decisions, and uncertainty estimation was performed to improve trust in predictions. This framework achieved 97% classification accuracy and outperformed previous methods.

Cengil et al. (2023) [39] developed an AI-based method for the automatic classification and localization of meningiomas and gliomas in brain MRI images. Using a public dataset, they labeled tumor regions and applied EfficientNet for feature extraction, followed by a Path Aggregation Network (PANet) for feature pyramid generation, and YOLO for object detection. Their model achieved strong performance, with a precision of 0.885, recall of 1.0, and mAP of 0.856, highlighting the potential clinical utility of AI-driven tumor detection.

Shin et al. (2021) [12] proposed a deep learning-based model using ResNet-50 for differentiating glioblastoma from solitary brain metastasis in conventional MRI scans. Trained on 498 patients, the model achieved AUCs of 0.889 and 0.835 on internal and external test sets, respectively. Compared to two neuroradiologists, the model outperformed one in external validation, indicating its potential as a supplementary tool for preoperative diagnosis.

Yildirim et al. (2023) [2] developed a hybrid convolutional neural network model for classifying gliomas, meningiomas, pituitary tumors, and normal brain tissues

in MRI scans. The model leverages features extracted from pre-trained EfficientNet-B0 and ShuffleNet architectures. Image enhancement techniques were applied to the dataset, and features were concatenated and reduced using the mRMR method before classification with an SVM. The proposed method outperformed the individual architectures, achieving an accuracy of 95.4%.

Karabacak et al. (2023) [40] used machine learning models to predict postoperative outcomes after intracranial meningioma resections. They found that LightGBM and Random Forest algorithms performed best, with LightGBM achieving AUROCs of 0.842 and 0.846 for prolonged LOS and nonhome discharges, respectively. This study highlights the potential of machine learning in predicting short-term adverse outcomes after surgery.

The XAISS-BMLBT technique introduced in the study by Lakshmi et al. 2025 [41] combines semantic segmentation and Bayesian machine learning for brain tumor detection in MRI images. It involves preprocessing with bilateral filtering, MEDU-Net + for segmentation, and feature extraction using ResNet50. The Bayesian regularized artificial neural network (BRANN) model is employed for tumor identification, with hyperparameter tuning via an improved radial movement optimization model. Experimental results showed an accuracy of 97.75%, outperforming existing models.

Our findings demonstrate that integrating radiomic features with deep convolutional neural networks

Table 7 Detailed classification metrics for each tumor type, including TP, FP, FN, TN, Sensitivity, Specificity, PPV, and NPV, are provided

Classifier	Tumor Type	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
AdaBoost	Meningioma	52	27	19	128	0.73	0.82	0.66	0.87
	GBM	78	48	27	71	0.74	0.60	0.62	0.72
	Brain-MET	19	0	29	176	0.40	1	1	0.86
XGBoost	Meningioma	56	70	15	83	0.79	0.54	0.44	0.85
	GBM	64	14	41	105	0.61	0.88	0.82	0.72
	Brain-MET	19	1	29	175	0.40	0.99	0.95	0.86
LightGBM	Meningioma	57	61	14	92	0.80	0.60	0.48	0.87
	GBM	55	13	50	106	0.52	0.90	0.81	0.68
	Brain-MET	37	1	11	175	0.77	0.99	0.97	0.94
CatBoost	Meningioma	53	22	18	131	0.75	0.86	0.71	0.88
	GBM	83	18	22	101	0.80	0.85	0.82	0.82
	Brain-MET	48	0	0	176	1	1	1	1
MLP	Meningioma	47	11	24	142	0.66	0.93	0.81	0.86
	GBM	94	24	11	95	0.90	0.80	0.80	0.90
	Brain-MET	48	0	0	176	1	1	1	1

Table 8 Performance comparison of classification models based on accuracy and macro F1-score metrics

Models	Accuracy	Macro F1-score
AdaBoost	67	65
XGBoost	62	61
LightGBM	67	70
CatBoost	82	84
MLP	84	86

provides an effective and reliable solution for brain tumor classification. By employing EfficientNet-B0 for 3D brain metastases and ResNet-18 for 2D glioblastomas (IDH-wildtype, grade 4) and meningiomas, our framework captures both handcrafted radiomic and deep hierarchical features, enhancing classification performance across multiple tumor types. As shown in Tables 7 and 8, neural network-based and advanced ensemble models such as CatBoost achieved high overall performance, whereas conventional ensemble methods like AdaBoost and XGBoost exhibited lower stability. Notably, our hybrid model with MLP reached the highest accuracy (85%) and macro F1-score (86%), consistently maintaining high sensitivity and specificity across tumor types. These results, summarized in Tables 7 and 8, highlight the added value of combining radiomics with deep learning, particularly in capturing spatial and textural nuances often missed by traditional models, providing a more robust approach to brain tumor classification. While recent studies have shown impressive results using complex, highly optimized deep learning pipelines including Bayesian-tuned CNNs [42] and explainable AI approaches such methods often require extensive computational resources, large-scale datasets, and intensive parameter tuning.

In contrast, our model provides a lightweight yet robust alternative, balancing accuracy and efficiency, thereby offering improved feasibility for clinical implementation.

Importantly, the robustness of the proposed approach is further demonstrated through independent external validation, where comparable performance to the internal evaluation was maintained. Moreover, the strength of our approach lies in its evaluation on diverse and heterogeneous datasets, including in-house clinical images and publicly available sources such as Stanford [18], TCIA [19] and SartajBhuvaji [17] enhances the generalizability of the proposed method and supports its applicability in real-world clinical environments.

The ROC curves in Fig. 4. visualize the classification performance of five models—AdaBoost, XGBoost, LightGBM, CatBoost, and MLP—across the three tumor subtypes: Brain Metastasis, Glioma, and Meningioma, using a one-vs-rest approach. LightGBM, CatBoost, and MLP all achieved an AUC of 1.00 for the BrainMet class, indicating perfect ranking of BrainMet samples relative to the other classes. Nevertheless, LightGBM misclassified 11 out of 48 BrainMet samples, revealing an important nuance: AUC reflects the model's ability to rank predictions, not its classification accuracy at a fixed threshold. Hence, even with misclassifications at a standard threshold (e.g., 0.5), a model can achieve a perfect AUC if it consistently assigns higher probabilities to the true class. AUC values for Glioma and Meningioma were slightly lower, ranging between 0.75 and 0.92, indicating a higher degree of overlap and greater difficulty in separating these classes. This suggests that these tumor types share more similar imaging features or patterns, making them more challenging to classify. Overall, MLP and CatBoost demonstrated the strongest performance, combining high macro F1-scores with consistently strong AUC values across all classes. LightGBM also showed competitive results, particularly for BrainMet, despite some classification errors. AdaBoost and XGBoost, while still effective,

showed lower performance, especially in distinguishing Glioblastoma, IDH-wildtype, grade 4 and Meningioma.

The ROC curves for BrainMet are sharply skewed toward the top-left corner, consistent with near-perfect sensitivity and specificity. In contrast, ROC curves for Glioblastoma, IDH-wildtype, grade 4 and Meningioma exhibit a more gradual arc, reflecting overlap in predicted probabilities and class boundaries.

To assess the confidence of our models in their predictions, we computed entropy-based uncertainty scores for each test sample. Figure 6. In the CatBoost model, meningioma samples exhibited the highest entropy values, indicating higher uncertainty in classification. Glioma samples showed moderate uncertainty, whereas brain metastasis samples had consistently low entropy values, suggesting higher model confidence for this class.

In contrast, the MLP model showed relatively low entropy for meningioma, but significantly higher and more fluctuating uncertainty for glioma, indicating that the model struggled to confidently classify glioma samples. For brain metastasis, both models demonstrated low entropy and high confidence, with the MLP model showing an even sharper drop in uncertainty compared to CatBoost.

Overall, brain metastasis was the most confidently predicted class in both models. Meningioma was better handled by the MLP model, while glioma was more confidently predicted by CatBoost.

Although the classification accuracy of our method may be slightly lower compared to state-of-the-art models, it offers a promising trade-off between performance and simplicity. Our approach requires less computational cost and is more easily deployable in clinical settings, where timely and reliable decision support is essential.

The present study has several limitations. First, the dataset size was relatively limited, and only three tumor classes were included, which may affect generalizability. Although additional public datasets were incorporated to partially address this issue, larger and more diverse cohorts would further strengthen the findings. Second, for the majority of cases, histopathological confirmation was available in the medical records. However, for a subset of patients, pathology reports were not accessible. In these cases, tumor labels were assigned based on a consensus between an experienced radiation oncologist and a neurologist with direct clinical familiarity with the patients, using radiotherapy records and imaging characteristics. While histopathology remains the gold standard, this expert-driven approach reflects real-world clinical practice and mitigates potential labeling bias. Third, although independent public datasets were incorporated into the training and evaluation process to enhance data diversity, external validation on a completely unseen cohort remains limited, and future studies

with larger independent cohorts are warranted. Fourth, the proposed model relies solely on MRI data and does not incorporate multimodal imaging (e.g., PET) or clinical variables, which could further improve performance. Additionally, experiments were conducted using Google Colab, imposing hardware constraints that limited model complexity and batch size. Future work will focus on expanding dataset size, incorporating multimodal information, ensuring comprehensive histopathological confirmation, and leveraging high-performance computing resources to enable more complex architectures and extensive hyperparameter optimization.

Conclusion

In this study, we proposed a hybrid framework for simultaneous multi-class brain tumor detection, segmentation, and classification by integrating nnU-Net and ITK-SNAP for segmentation, and combining deep learning models (EfficientNetB0 and ResNet-18) with hand-crafted radiomic features for classification. The results demonstrate that our approach effectively distinguishes between Glioblastoma, IDH-wildtype, grade 4, meningiomas, and brain metastases, achieving robust and accurate performance across tumor types. By leveraging both deep feature representations and handcrafted radiomic descriptors, and employing advanced classifiers such as CatBoost and multilayer perceptron (MLP), the proposed method enhances feature interpretability while maintaining strong classification performance. This hybrid strategy offers a reliable, scalable, and clinically applicable solution for automated brain tumor diagnosis. For future work, we aim to expand the framework by incorporating multi-modal data, including PET scans, to enrich the diagnostic information. Additionally, we plan to adopt advanced variants of U-Net to further improve segmentation accuracy and facilitate end-to-end tumor characterization.

Acknowledgements

This study is based on a verified research project at Semnan University of Medical Sciences (Ref. No. 330099540). We sincerely thank all the participants who volunteered for this study and their companions. We also extend our gratitude to the oncologists and the staff of the radiation oncology department at Kosar Hospital for their invaluable support and assistance.

Author contributions

Sanaz Alibabaei: study design, supervision, software and data analysis, initial draft, final edit, python codes, Milad Taleb: software and data analysis and python codes, and final edit.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The study uses two types of data: 1. **Publicly available datasets**. a) Datasets used for training and internal evaluation: Stanford university BrainMetShare, Datasets Available: <https://doi.org/10.71718/z66c-qr59>, BCBM-RadioGenomics Version 1, Datasets Available: <https://doi.org/10.7937/RRSE-W278> and

SartajBhuvaji Datasets Available: <https://github.com/SartajBhuvaji/Brain-Tumor-Classification-DataSet>. b) Datasets used exclusively for external validation: Mendeley Data <https://data.mendeley.com/datasets/zwr4ntf94j/4>, doi: <https://doi.org/10.17632/zwr4ntf94j.1> 2025. The University of California San Francisco Brain Metastases Stereotactic Radiosurgery (UCSF-BMSR) MRI Dataset <https://arxiv.org/abs/2304.07248>, <https://pubs.rsna.org/doi/10.1148/rayi.230126>, doi: <https://doi.org/10.58078/C24W2K>. 2. **Clinical MRI data** collected at [Kosar hospital] under ethical approval (Ref. No. 330099540, Ethics code: IR.SEMUMS.REC.1403.329). These data contain sensitive patient information and are not publicly available, but anonymized MR images and analysis code can be provided upon reasonable request to the corresponding author, subject to institutional ethical approvals and data-sharing agreements.

Declarations

Ethical approval

This study was carried out at the radiation oncology ward of Kosar Hospital (Semnan, Iran). The protocol was approved by the ethics committee of Semnan University of Medical Sciences and Health Services (Ref. No. 330099540, Ethics code: IR.SEMUMS.REC.1403.329). The MR images of Glioma, Meningioma, and Brain-METs patients under treatment in the radiation oncology ward of Kosar Hospital (Semnan, Iran) from 2023 up to 2025 were used, available in the hospital PACS system. Patients' informed consent was obtained, and all images were anonymized before use. All procedures were conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. The corresponding author, Sanaz Alibabaei, affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent to publish

Not applicable. No patient data or images are shown, and all illustrated MR images are anonymized.

Declaration of generative AI and AI-assisted technologies in the writing process

While preparing this work, the authors used chatGPT to improve the writing of some sentences, shorten the text, and check the grammar. After using this tool/service, the authors reviewed and edited the sentences and content as needed and take full responsibility for the content of the publication.

Competing interests

The authors declare no competing interests.

Received: 10 October 2025 / Accepted: 25 December 2025

Published online: 03 January 2026

References

1. Ahsan R, Shahzadi I, Najeeb F, Omer H. Brain tumor detection and segmentation using deep learning. *Magn Reson Mater Phys Biol Med*. 2024;37:1–10. <https://doi.org/10.1007/s10334-024-01203-5>.
2. Yıldırım M, Cengil E, Ergo lu Y, Cinar A. Detection and classification of glioma, meningioma, pituitary tumor, and normal in brain magnetic resonance imaging using deep learning-based hybrid model? *Iran J Comput Sci*. 2023;6:455–64. <https://doi.org/10.1007/s42044-023-00139-8>.
3. Lu VM, O'Connor KP, Shah AH, Eichberg DG, Luther EM, Komotor RJ, et al. The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: a systematic review of the contemporary literature. *J Neurooncol*. 2020;148:221–9. <https://doi.org/10.1007/s11060-020-03528-2>.
4. El Hachimy I, Kabelma D, Echchafref C, Hassani M, Benamar N, Hajji N. A comprehensive survey on the use of deep learning techniques in glioblastoma. *Artif Intell Med*. 2024;150:102902. <https://doi.org/10.1016/j.artmed.2024.102902>.
5. Babaei Rikan S, Sorayaie Azar A, Naemi A, Bagherzadeh Mohasefi J, Pirnejad H, Wil UK. Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques. *Sci Rep*. 2024;14:2371. <https://doi.org/10.1038/s41598-024-53006-2>.
6. Neil ZD, Pierzchajlo N, Boyett C, Little O, Kuo CC, Brown NJ, et al. Assessing metabolic markers in glioblastoma using machine learning: a systematic review. *Metabolites*. 2023;13:161. <https://doi.org/10.3390/metabolites13020161>.
7. Mohammed S, Dinesan M, Ajayakumar T. Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: a retrospective study. *Rep Pract Oncol Radiother*. 2022;27:1026–36. <https://doi.org/10.5603/RPOR.a2022.0113>.
8. Jekel L, Brim WR, von Reppert M, Staib L, Cassinelli Petersen G, Merkaj S, et al. Machine learning applications for differentiation of glioma from brain metastasis—a systematic review. *Cancers*. 2022;14:1369. <https://doi.org/10.3390/cancers14061369>.
9. Zhang H, Mo J, Jiang H, Li Z, Hu W, Zhang C, et al. Deep learning model for the automated detection and histopathological prediction of meningioma. *Neuroinformatics*. 2021;19:393–402. <https://doi.org/10.1007/s12021-020-09492-6>.
10. Huang RY, Bi WL, Griffith B, Kaufmann TJ, la Fougère C, Schmidt NO, et al. Imaging and diagnostic advances for intracranial meningiomas. *Neuro Oncol*. 2019;21(Suppl 1):44–61. <https://doi.org/10.1093/neuonc/noy143>.
11. Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews. *Health Serv Res Manag Epidemiol*. 2024;11:2333928241234863. <https://doi.org/10.1177/2333928241234863>.
12. Shin I, Kim H, Ahn S, Sohn B, Bae S, Park J, et al. Development and validation of a deep learning-based model to distinguish glioblastoma from solitary brain metastasis using conventional MR images. *AJNR Am J Neuroradiol*. 2021;42:838–44. <https://doi.org/10.3174/ajnr.A7003>.
13. Sawakare S, Chaudhari D. Classification of brain tumor using discrete wavelet transform, principal component analysis and probabilistic neural network. *Int J Res Emerg Sci Technol*. 2014;1:13–9.
14. Das S, Goswami RS. Advancements in brain tumor analysis: a comprehensive review of machine learning, hybrid deep learning, and transfer learning approaches for MRI-based classification and segmentation. *Multimed Tools Appl*. 2025;84:26645–82. <https://doi.org/10.1007/s11042-024-20203-0>.
15. Zarenia E, Far AA, Rezaee K. Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping. *Sci Rep*. 2025;15:8114. <https://doi.org/10.1038/s41598-025-92776-1>.
16. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol*. 2021;23:1231–51. <https://doi.org/10.1093/neuonc/noab106>.
17. Bhuvaji S. Brain-Tumor-Classification-DataSet. GitHub repository. 2020. Available from: <https://github.com/SartajBhuvaji/Brain-Tumor-Classification-DataSet/tree/master/Training>
18. Stanford University. BrainMetShare [Dataset]. 2020. <https://doi.org/10.71718/z66cqr59>
19. Taha B, Wu D, Sabal L, Kollitz M, Venteicher A, Watanabe Y. MRI dataset of metastatic breast cancer to the brain with Expert-reviewed segmentations and Tumor-derived radiomic features (BCBM-RadioGenomics) (Version 1) [Dataset]. Cancer Imaging Archive. 2025. <https://doi.org/10.7937/RRSE-W278>.
20. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. 2014;34:1993–2024. <https://doi.org/10.1109/tmi.2014.2377694>.
21. Mol FN, van der Hoek L, Ma B, Nagam BC, Sijtsema NM, van Dijk LV, et al. MRI-based head and neck tumor segmentation using nnU-Net with 15-fold cross-validation ensemble. *Challenge on head and neck tumor segmentation for MRI-Guided applications*. Springer; 2024. pp. 179–90. <https://doi.org/10.48550/arXiv.2412.06610>.
22. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. Monai: an open-source framework for deep learning in healthcare. *ArXiv Preprint arXiv:2211.02701*. 2022. <https://doi.org/10.48550/arXiv.2211.02701>.
23. Rana A, Rawat AS, Bijalwan A, Bahuguna H, editors. Application of multi layer (perceptron) artificial neural network in the diagnosis system: a systematic review. In: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE). IEEE; 2018. <https://doi.org/10.1109/RICE.2018.85509069>
24. Ngassou M, Mwangi RW, Nyarige E. A hybrid ensemble learning approach utilizing light gradient boosting machine and category boosting model for

- lifestyle-based prediction of type-II diabetes mellitus. *J Data Anal Inf Process.* 2023;11:480–511. <https://doi.org/10.4236/jdaip.2023.114025>.
- 25. Dubey AK, Jain V, editors. Comparative study of convolution neural network's ReLU and leaky-ReLU activation functions. In: Applications of Computing, Automation and Wireless Systems in Electrical Engineering: Proceedings of MARC 2018. Springer; 2019. https://doi.org/10.1007/978-981-13-6772-4_76
 - 26. Reyad M, Sarhan AM, Arafa M. A modified Adam algorithm for deep neural network optimization. *Neural Comput Appl.* 2023;35:17095–112. <https://doi.org/10.1007/s00521-023-08568-z>.
 - 27. Kabir HI, Hossain MS, Bitheea MA, Sara U, Hasan MM, Towsif AA, Ahmed MK. Brain tumor MRI dataset (Glioma, Meningioma, Pituitary, No Tumor). Mendeley Data, V5. 2025. <https://doi.org/10.17632/zwr4ntf94j1>
 - 28. Rudie JD, Saluja R, Weiss DA, Nedelec P, Calabrese E, Colby JB, Laguna B, Mongan J, Braunstein S, Hess CP, Rauschecker AM, Sugrue LP, Villanueva-Meyer JE. The university of California San Francisco brain metastases stereotactic radiosurgery (UCSF-BMSR) MRI dataset. *Radiol Artif Intell.* 2024;6:e230126. <https://doi.org/10.1148/ryai.230126>.
 - 29. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30. <https://doi.org/10.48550/arXiv.1705.07874>.
 - 30. Ray M, Mahata N, Sing JK. Uncertainty parameter weighted entropy-based fuzzy c-means algorithm using complemented membership functions for noisy volumetric brain MR image segmentation. *Biomed Signal Process Control.* 2023;85:104925. <https://doi.org/10.1016/j.bspc.2023.104925>.
 - 31. Yang C, Guan X, Xu Q, Xing W, Chen X, Chen J, et al. How can SHAP (SHapley additive exPlanations) interpretations improve deep learning based urban cellular automata model? *Comput Environ Urban Syst.* 2024;in press. <https://doi.org/10.1016/j.compenvurbsys.2024.102133>.
 - 32. Mayer M. SHAP for additively modeled features in a boosted trees model. *ArXiv Preprint ArXiv:2207.14490.* 2022. <https://doi.org/10.48550/arXiv.2207.14490>
 - 33. Zhao C, Liu J, Parilina E. ShapG: new feature importance method based on the Shapley value. *Eng Appl Artif Intell.* 2025;148:110409. <https://doi.org/10.8550/arXiv.2407.00506>.
 - 34. Sale Y, Hofman P, Wimmer L, Hüllermeier E, Nagler T. Second-order uncertainty quantification: Variance-based measures. *ArXiv Preprint arXiv:2401.00276.* 2023. <https://doi.org/10.48550/arXiv.2401.00276>.
 - 35. Akter A, Nosheen N, Ahmed S, Hossain M, Yousuf MA, Almoyad MAA, et al. Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. *Expert Syst Appl.* 2024;238:122347. <https://doi.org/10.1016/j.eswa.2023.122347>.
 - 36. Amin J, Sharif M, Raza M, Yasmin M. Detection of brain tumor based on features fusion and machine learning. *J Ambient Intell Humaniz Comput.* 2024;1–17. <https://doi.org/10.1007/s12652-018-1092-9>.
 - 37. Sandhiya B, Raja SKS. Deep learning and optimized learning machine for brain tumor classification. *Biomed Signal Process Control.* 2024;89:105778. <https://doi.org/10.1016/j.bspc.2023.105778>.
 - 38. Karthik A, Sahoo SK, Kumar A, Patel N, Chinnaraj P, Maguluri LP, et al. Unified approach for accurate brain tumor multi-classification and segmentation through fusion of advanced methodologies. *Biomed Signal Process Control.* 2025;100:106872. <https://doi.org/10.1016/j.bspc.2024.106872>.
 - 39. Cengil E, Eroğlu Y, Çınar A, Yıldırım M. Detection and localization of glioma and meningioma tumors in brain MR images using deep learning. *Sakarya Univ J Sci.* 2023;27:550–63. <https://doi.org/10.16984/saufenbilder.1067061>.
 - 40. Karabacak M, Jagtiani P, Shrivastava RK, Margetis K. Personalized prognosis with machine learning models for predicting in-hospital outcomes following intracranial meningioma resections. *World Neurosurg.* 2024;182:e210–30. <https://doi.org/10.1016/j.wneu.2023.11.081>.
 - 41. Lakshmi K, Amaran S, Subbulakshmi G, Padmini S, Joshi GP, Cho W. Explainable artificial intelligence with UNet based segmentation and Bayesian machine learning for classification of brain tumors using MRI images. *Sci Rep.* 2025;15:690. <https://doi.org/10.1038/s41598-024-84692-7>.
 - 42. Chaddad A, Hu Y, Wu Y, Wen B, Kateb R. Generalizable and explainable deep learning for medical image computing: an overview. *Curr Opin Biomed Eng.* 2024;100:100567. <https://doi.org/10.48550/arXiv.2503.08420>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.