JOURNAL OF
EMERGING INVESTIGATORS

Article

Assessing machine learning model efficacy for brain tumor MRI classification: a multi-model approach

Mrigank Dhingra[1], Vikas Dhingra[1]
[1] Lexington High School, Lexington, Massachusetts

SUMMARY

Brain tumors pose a significant diagnostic challenge due to their complexity and the shortage of specialized neuroradiologists, motivating the development of machine learning (ML) approaches to reduce diagnostic variability and augment clinical workflows. Here, we evaluated the effectiveness of ML models in classifying brain tumor types, addressing variability and time-intensive manual interpretation. We investigated two hypotheses. First, we hypothesized that the classification accuracy of ML models would vary across tumor types due to differences in their imaging interpretability. Second, we hypothesized that tumor boundaries and high-intensity regions in MRI scans would disproportionately influence model predictions because these regions have the most discriminative information for separating tumor classes. We used the publicly available "Brain Tumor MRI Dataset" from Kaggle, consisting of 7,023 MRI slices categorized into glioma, meningioma, pituitary, and non-tumorous cases. We evaluated five ML models: Decision Tree, k-Nearest Neighbors (k-NN), DenseNet, ResNet50, and a custom Convolutional Neural Network (CNN). DenseNet, k-NN, and ResNet50 achieved high accuracies of 0.9361, 0.9125, and 0.9112, respectively. Pituitary and non-tumorous cases reached AUC scores of 1.00, while glioma and meningioma had slightly lower AUC scores of 0.99 and 0.98, respectively. Our analyses identified tumor boundaries and high-intensity MRI regions as key features driving classification, and statistical tests confirmed no significant differences among models ($p > 0.05$). Our research highlights the potential of ML models to enhance diagnostics by improving accuracy and identifying tumor-specific performance variations, supporting future clinical integration.

INTRODUCTION

Each year, thousands of people in the U.S. are diagnosed with brain tumors, joining the hundreds of thousands of people already living with the condition (1). Accurate and timely diagnosis of brain tumors is critical for the development of effective treatment plans (2). Despite advances in imaging technology, particularly in magnetic resonance imaging (MRI), clinicians still face major challenges in reading brain scans quickly and consistently. Interpreting brain MRI scans is a complex and time-consuming task, and the increasing volume of imaging studies has led to considerable reporting backlogs (3, 4). Growing demand for radiologic exams, driven by advancing technology and an aging population, has outpaced radiologists' capacity, leading to reporting delays (4). The limited availability of specialized neuroradiologists further compounds this issue, particularly in resource-limited countries, where access to expertise is severely limited (5, 6).

The need for improved diagnostic capabilities has driven significant interest in automating MRI analysis for brain tumor detection. Branches of artificial intelligence (AI), like machine learning (ML) and deep learning, have demonstrated the potential to address these challenges by offering tools to reduce reporting times and improve diagnostic accuracy. These models leverage statistical methods and advanced algorithms to learn patterns from data, facilitating the resolution of complex real-world problems. While the field has made considerable strides, current models often focus narrowly on specific tasks, such as tumor detection or grading, without addressing broader clinical workflows that integrate multiple steps in the diagnostic process (7). Recognizing these limitations is not to discount past work, but to highlight ongoing challenges that require further exploration, such as improving generalizability and applicability across diverse clinical settings.

Key barriers to robust diagnostic tools include dataset bias, variability in MRI protocols, and the inherent complexity of brain tumors, which span over 100 distinct types (8, 9). Although recent advancements in deep learning architectures have shown promise – including Convolutional Neural Networks (CNNs), Capsule Neural Networks, and Vision Transformers – they are typically optimized for specific datasets or isolated tasks (10). This emphasizes the necessity to systematically evaluate and refine ML models to ensure that they align with real-world clinical needs, particularly in addressing diagnostic variability and enhancing workflow efficiency.

To explore the strengths of different approaches, we evaluated five ML models. The Decision Tree provides a flowchart-like structure by splitting images based on key features (11). The k-Nearest Neighbors (k-NN) classifier assigns labels based on similarity to nearby data points, making it effective for smaller, well-separated classes (12). In contrast, deep learning models can automatically extract complex features from images. We designed a custom CNN for this study. The ImageNet dataset originally trained DenseNet and ResNet50, enabling these models to serve as advanced transfer learning architectures that reuse knowledge from large-scale image collections (13). Together, these models span a spectrum from simple, transparent algorithms to sophisticated, high-capacity neural networks, allowing us to compare performance across varying levels of complexity.

Building on these advances, we evaluated multiple ML models for classifying brain tumor MRI scans into diagnostically relevant categories. We investigated two key hypotheses: 1) We hypothesized that the classification accuracy of ML models would vary significantly across tumor types, with some tumor categories, such as pituitary tumors, achieving higher accuracy and AUC due to their distinct imaging features compared to gliomas and meningiomas, and 2) We hypothesized that tumor boundaries and high-intensity

https://doi.org/10.59720/25-127

regions in MRI scans would disproportionately influence model predictions, with saliency maps and SHAP analysis identifying these features as key contributors to classification accuracy. We used a large, publicly available dataset with diverse tumor types to compare the performance of five ML models across metrics including accuracy, F1 score, and Receiver Operating Characteristic (ROC), Area Under the Curve (AUC) (14, 15). This study focused on optimizing AI-driven diagnostic tools by balancing performance and generalizability while outlining their role in addressing challenges within the diagnostic workflow.

RESULTS

Our study workflow involved preprocessing and ML (**Figure 1**). To train our ML models, we used the "Brain Tumor MRI Dataset" from Kaggle, consisting of 7,023 MRI brain scan slices from patients. Each patient contributed multiple scans representing different slices of the same tumor or brain. The dataset was categorized into four classes: glioma (23.1%, N=1,620), meningioma (23.4%, N=1,643), pituitary tumors (25.0%, N=1,756), and non-tumorous cases (28.5%, N=2,004), with the latter consisting only of healthy brain scans. To build and assess ML models, the dataset was preprocessed with an 80% training subset, 10% validation subset, and 10% testing subset.

Initial testing of multiple ML models demonstrated varying performance levels (**Table 1**). The k-NN classifier, DenseNet, and CNN model exhibited high accuracy rates. The k-NN classifier achieved an accuracy score of 0.9125 with the hyperparameter n_neighbors set to 1 (**Figure 2**). This means the model classified each MRI slice based on its single closest match in the training set. The DenseNet model attained an accuracy score of 0.9096 after training for 10 epochs, chosen as an initial baseline. Similarly, the CNN model achieved an accuracy score of 0.8993 with 15 epochs of training, determined by incremental increases from the baseline while maintaining validation performance to avoid overfitting. In comparison, the Decision Tree classifier demonstrated moderate effectiveness with an accuracy score of 0.8413. For ResNet50, the baseline of 10 epochs initially achieved an accuracy of 0.2322. Extending training to 30 epochs improved its performance to 0.9112 (**Figure 3d**).

To refine model performance, we conducted retesting with extended training epochs of 30 for the DenseNet, ResNet50, and CNN models, and incorporated additional performance metrics, including ROC curves, AUC scores, F1 scores, and 95% confidence intervals (**Table 1**). The updated training curves showed that all three deep-learning models achieved stable convergence, with DenseNet reaching the lowest overall loss score of 0.2263 (**Figure 3a**). A loss score is a number that tells us how far off the model's predictions are from the correct labels. ROC analysis revealed clearer differences in tumor-specific discrimination. DenseNet achieved AUC values of 0.99 for gliomas, 0.96 for meningiomas, and 1.00 for pituitary and non-tumorous cases (**Figure 4a**). ResNet50 produced similarly high performance, with AUC values of 0.99 for gliomas, 0.97 for meningiomas, and 1.00 for pituitary and non-tumorous cases (**Figure 4b**). The CNN reached AUC values of 0.99 for gliomas, 0.98 for meningiomas, and 1.00 for pituitary and non-tumorous cases,

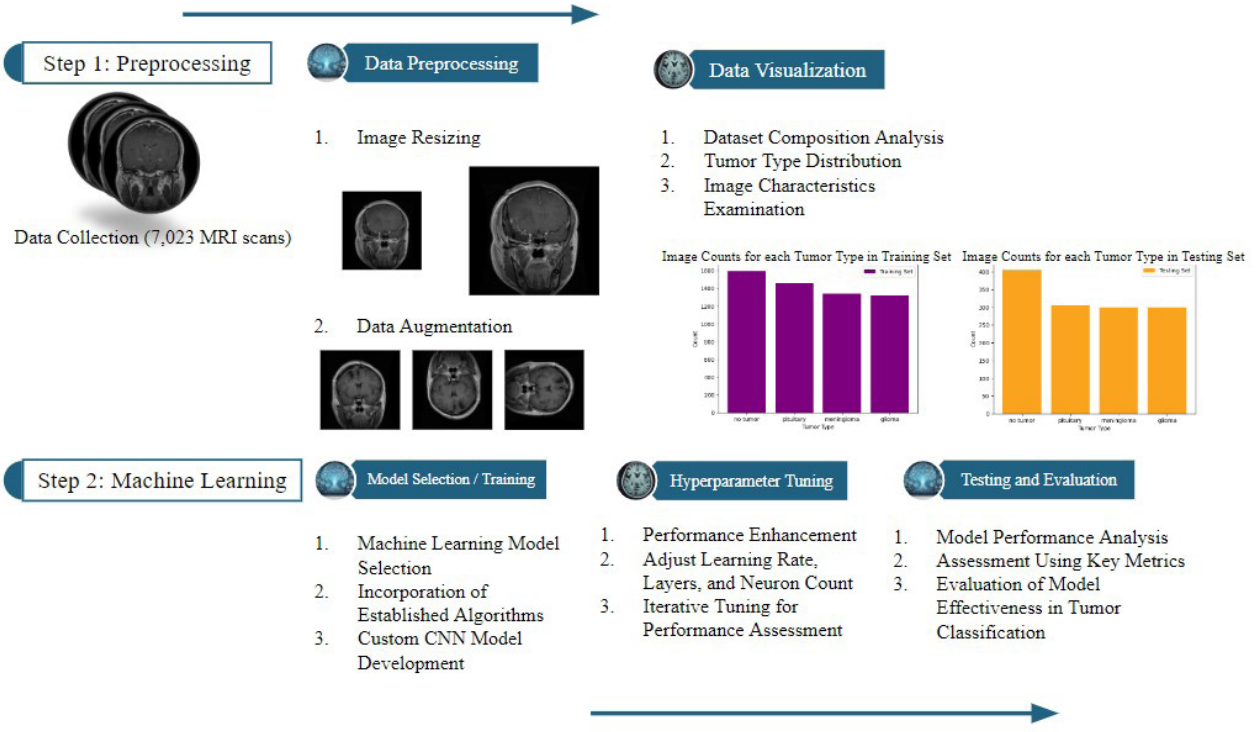

Figure 1: Workflow for brain tumor classification using machine learning (ML). The process involves two main steps: (Step 1) Data preprocessing, which includes image resizing, grayscale conversion, and data augmentation to enhance the dataset (N = 7,023 MRI images); and (Step 2) Model training and evaluation, including hyperparameter tuning and performance analysis using key metrics like accuracy, ROC AUC, and F1 scores. The workflow emphasizes the integration of advanced ML models for robust tumor classification.

https://doi.org/10.59720/25-127

Model	Accuracy score	Accuracy score 95% CI	Weighted F1 score	Weighted F1 score 95% CI	AUC score	AUC 95% CI
Decision Tree Classifier	0.8273	0.81–0.85	0.8306	0.81–0.85	0.8852	0.80–0.94
kNN Classifier	0.9125	0.90–0.93	0.9166	0.90–0.93	0.9422	0.86–0.98
DenseNet model	0.9361	0.91–0.95	0.8419	0.82–0.86	0.9858	0.96–0.99
ResNet50 model	0.9125	0.90–0.93	0.9244	0.91–0.94	0.9895	0.97–0.99
CNN model	0.8993	0.88–0.91	0.9500	0.94–0.96	0.9875	0.97–1.00

Table 1: Performance metrics for ML models evaluated on the brain tumor MRI dataset. The table compares the performance of five machine-learning models on the brain tumor MRI dataset. Metrics include accuracy, weighted F1 scores, and AUC values with 95% confidence intervals. Statistical differences between the top-performing models (k-NN, DenseNet, ResNet50, and CNN) were not significant (t = -0.243; p = 0.808, p > 0.05).

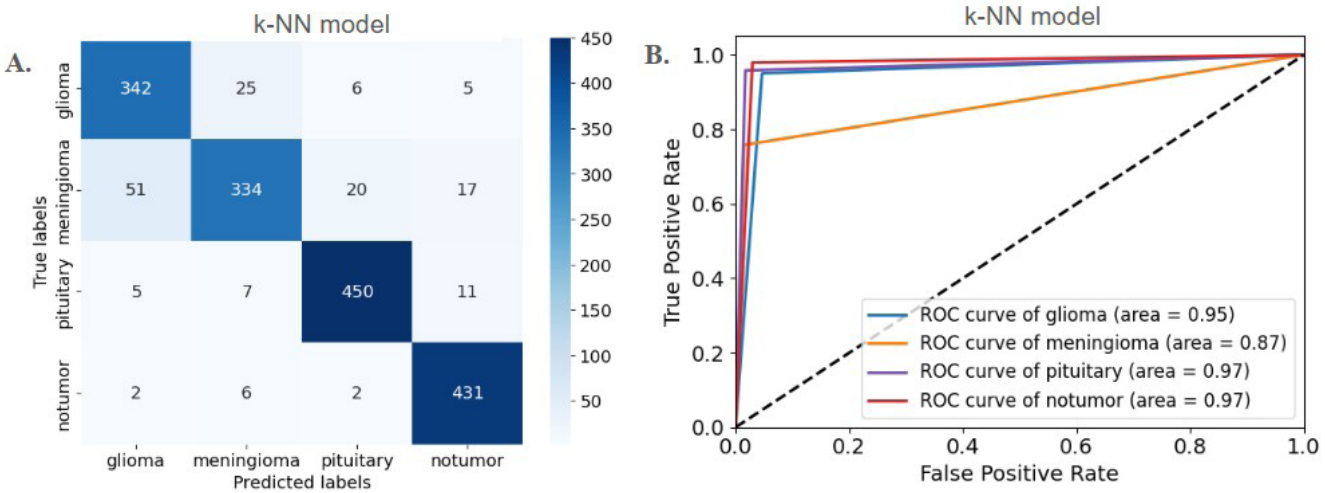

Figure 2: Confusion matrix and ROC curves for the k-NN–based brain tumor classification. (A) Confusion matrix showing true versus predicted labels for glioma, meningioma, pituitary tumors, and non-tumorous cases. (B) ROC curves demonstrate AUC values of 0.95 for gliomas, 0.87 for meningiomas, 0.97 for pituitary tumors, and 0.97 for non-tumorous cases. Data represent results from the k-NN classifier trained on MRI brain scan data, with a sample size of N = 7,023 images.

consistent with its weighted F1 score of 0.95 (**Figure 5**). These analyses provided a comprehensive evaluation of the models' capabilities.

Model performance varied across tumor types. The CNN model achieved a weighted F1 score of 0.95 (95% CI: 0.94–0.96) and ROC/AUC scores of 0.99 for gliomas, 0.98 for meningiomas, and 1.00 for pituitary and non-tumorous cases, demonstrating strong discrimination across all classes (**Figures 3e,3f,5**). The Decision Tree classifier had a weighted F1 score of 0.831 (95% CI: 0.81–0.85) and an AUC of 0.885, with higher accuracy in predicting pituitary tumors (**Figures 6a,6b**). The k-NN classifier showed a weighted F1 score of 0.917 (95% CI: 0.90–0.93) and an AUC of 0.942. These results provide a comparative analysis of model performance across tumor types (**Table 1**).

As another metric for analyzing interpretability, we generated a saliency map using the DenseNet model to visualize which regions of each MRI slice contributed most to the model's predictions. Saliency maps provide an additional layer of analysis by showing whether the model focuses on anatomically meaningful structures and not just pixel patterns. The DenseNet saliency map highlighted tumor boundaries and high-intensity regions as critical features influencing classification decisions (**Figure 7**). The map indicated that DenseNet concentrated on tumor-associated regions rather than irrelevant background areas, supporting its strong classification performance.

We used SHAP analysis to identify which pixel-level

https://doi.org/10.59720/25-127

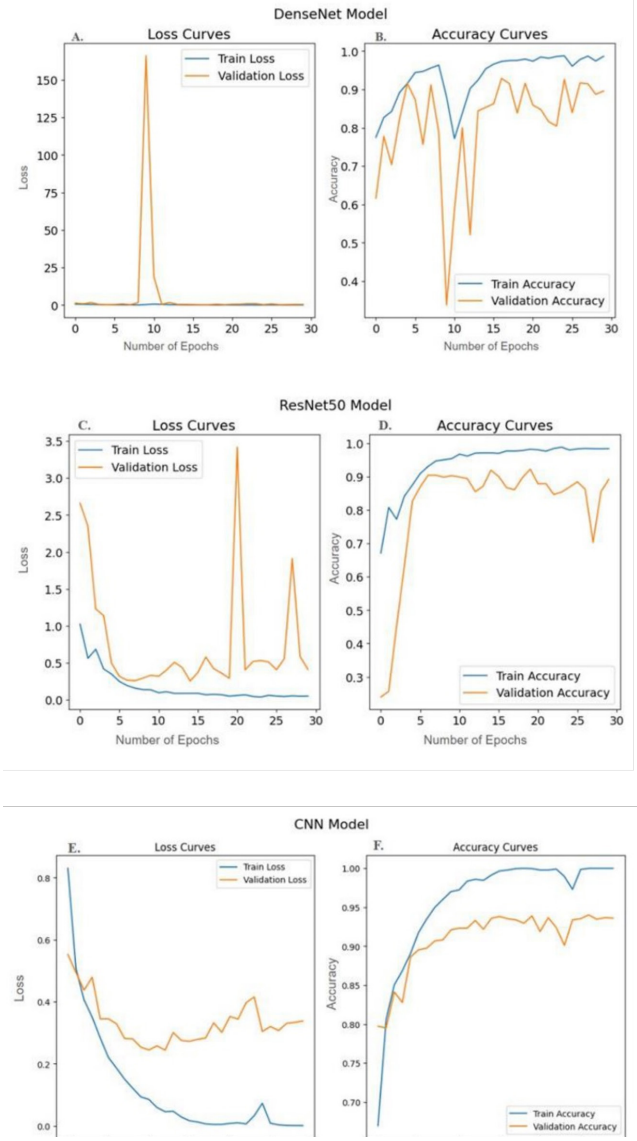

Figure 3: Training and validation loss/accuracy curves for ML models. (A) DenseNet loss shows steady training decrease with validation loss stabilizing after 10 epochs; (B) DenseNet accuracy trends upward and stabilizes; (C) ResNet50 loss decreases in training but fluctuates in validation; (D) ResNet50 accuracy increases with minor fluctuations; (E) CNN loss decreases consistently and stabilizes around epoch 20; (F) CNN accuracy rises steadily. Metrics calculated via cross-validation with N = 7,023 images split 80% training, 10% validation, 10% testing.

features contributed most to the Decision Tree's predictions. In this representation, each SHAP feature has a number, which refers to the intensity of a specific pixel location in the 32x32 MRI image. The SHAP summary plot showed that a small subset of high-intensity pixels had the greatest impact on classification outcomes, with Features 897 and 490 contributing the strongest effects across tumor types (**Figure 6c**). The plot also revealed that several of the most influential pixels contributed disproportionately to non-tumorous and pituitary tumor predictions.

We performed a paired t-test to compare model accuracies. The Decision Tree classifier showed significantly lower accuracy and AUC values than the other models, indicating that it did not perform at a comparable level. Because of this large performance gap, including it in the paired comparison with the other models would have skewed the analysis. When comparing the other models, the t-test showed no statistically significant differences (t = -0.243; p = 0.808), suggesting that the observed performance variations between the k-NN, DenseNet, ResNet50, and CNN were likely due to random chance. These findings emphasize that, under our study conditions, the four models demonstrated comparable reliability, while the Decision Tree consistently underperformed relative to them.

DISCUSSION

This study showcases the capability of ML models, particularly simpler models like k-NN and transfer learning–based approaches such as DenseNet and CNN, to enhance brain tumor diagnostics by accurately classifying tumor types using MRI scans. By leveraging a large, balanced dataset of gliomas, meningiomas, pituitary tumors, and non-tumorous cases, our models achieved strong performance metrics, including 0.9125 accuracy for k-NN, and ROC/AUC scores of 0.99 for gliomas and 1.00 for pituitary tumors.

The balanced class distribution helped minimize bias during model training, supporting model differentiation between various tumors and healthy tissue. The large scale and diversity of the dataset make it suitable for evaluating ML models in brain tumor classification. However, the non-tumorous category consisted solely of healthy brain scans, which may limit the model's ability to distinguish tumors from other brain abnormalities with similar imaging patterns. We partitioned the dataset into training and testing subsets using a stratified split to preserve class proportions. This division ensures that a substantial portion of the data is available for model training, while an appropriate amount is reserved for unbiased evaluation during testing.

Statistical analysis showed that while model performance varied, the differences in accuracy among models were not statistically significant (p > 0.05). Practical differences, however, were notable. For example, DenseNet and ResNet50 excelled in handling complex features and interpretability, while k-NN offered simplicity and computational efficiency, making it suitable for scenarios requiring rapid deployment with minimal resources. If a higher image resolution was used, performance might improve for models like the Decision Tree Classifier and k-NN, which rely on the raw details in each image. This could help with tumor types like meningioma, where important distinctions disappear when images are down-sampled too much. Downsampling is the process of reducing the data resolution to make it more compact, while still preserving its characteristics. However, using larger images would also need more computing power and could make some models over-fit to the training data, so the resolution would need to be increased carefully.

The results showed similar AUC values across all tumor categories. This suggests that the models learned useful features for all tumor types, possibly helped by the balanced dataset and data augmentation. Although small differences in performance remain, larger datasets and higher-quality images might reveal clearer distinctions. Our results indicate that further work is needed to improve tumor-specific classification, potentially by adding more data types or clinical information.

https://doi.org/10.59720/25-127

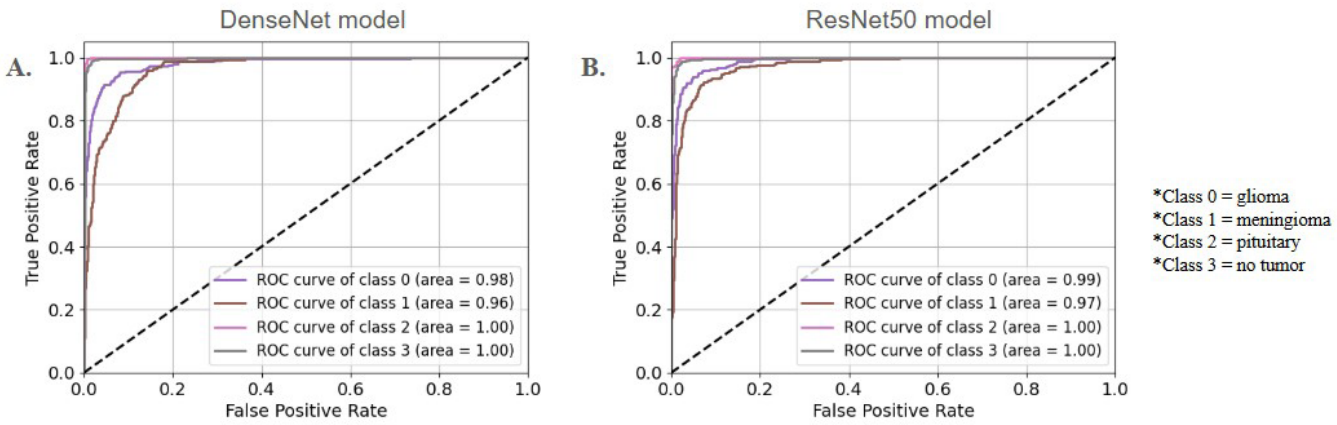

Figure 4: ROC curves for DenseNet and ResNet50 models. (A) ROC curves for the DenseNet model with AUC values of 1.00 for pituitary tumors and non-tumorous cases, 0.98 for gliomas, and 0.96 for meningiomas. (B) ResNet50 model demonstrates high AUC values across all tumor types, with AUC = 0.99 for gliomas, 0.97 for meningiomas, and 1.00 for pituitary tumors and non-tumorous cases. The diagonal dashed line represents an AUC score of 0.50 with no discriminative ability, included as a baseline for comparison. Results are based on the evaluation of the models on the testing subset (N = 703 images).

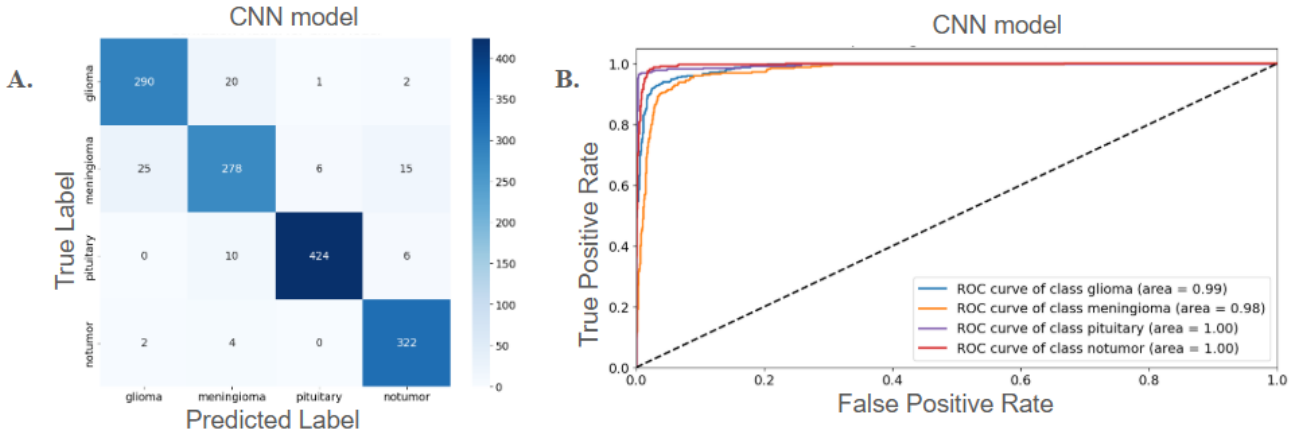

Figure 5: Performance evaluation of the CNN model. (A) Confusion matrix illustrating true versus predicted labels for glioma, meningioma, pituitary tumors, and non-tumorous cases. Pituitary tumors and non-tumorous cases achieved the highest classification accuracy, with 424 and 322 correct predictions, respectively. (B) ROC curves demonstrate excellent model discrimination, with AUC values of 0.99 for gliomas, 0.98 for meningiomas, and 1.00 for both pituitary tumors and non-tumorous cases. The analysis includes N = 7,023 MRI images.

Detailed evaluation of ML models revealed that classification accuracy differs by tumor type, highlighting the importance of metrics such as ROC/AUC, F1 scores, and confusion matrices for robust assessment. Beyond achieving high accuracy, the models demonstrated the capability to distinguish specific tumor types rather than merely classifying healthy versus tumorous cases. This addresses the critical need for precise, multi-class tumor classification, reinforcing the value of developing this method of classification for distinguishing specific brain tumor types.

The use of saliency maps was a key contribution of this study, consistently highlighting tumor boundaries and high-intensity regions critical for classification. Prior studies have shown that interpretability tools, such as saliency maps, improve clinician trust in AI tools by offering insights into the decision-making process (3). These findings align with the broader goal of bridging technical performance with clinical usability. Our analysis of DenseNet's performance, including its ability to achieve a loss score of 0.2263 and near-perfect AUC scores for pituitary tumors and non-tumorous cases, demonstrates the value of transfer learning as a key enabler for brain tumor classification. DenseNet and ResNet50 models demonstrated strong performance metrics. These results validate the approach of adapting pre-trained models to medical imaging tasks. However, there are still significant challenges for real-world implementation, such as variability in imaging protocols and patient populations. Additionally, we observed spikes in validation loss for some pre-trained models, which may be due to the small batch sizes used in training. This instability can lead to sudden jumps in validation performance. Future efforts should prioritize multi-institutional datasets to better reflect clinical imaging diversity and ensure generalizability, alongside optimizing training settings to improve model stability. Detailed dataset descriptions in public repositories could further support research and clinical applications.

Our SHAP analysis confirmed that the Decision Tree model heavily relied on pixels located near tumor boundaries and other high-intensity regions. These pixels produced the strongest contributions to class predictions,

https://doi.org/10.59720/25-127

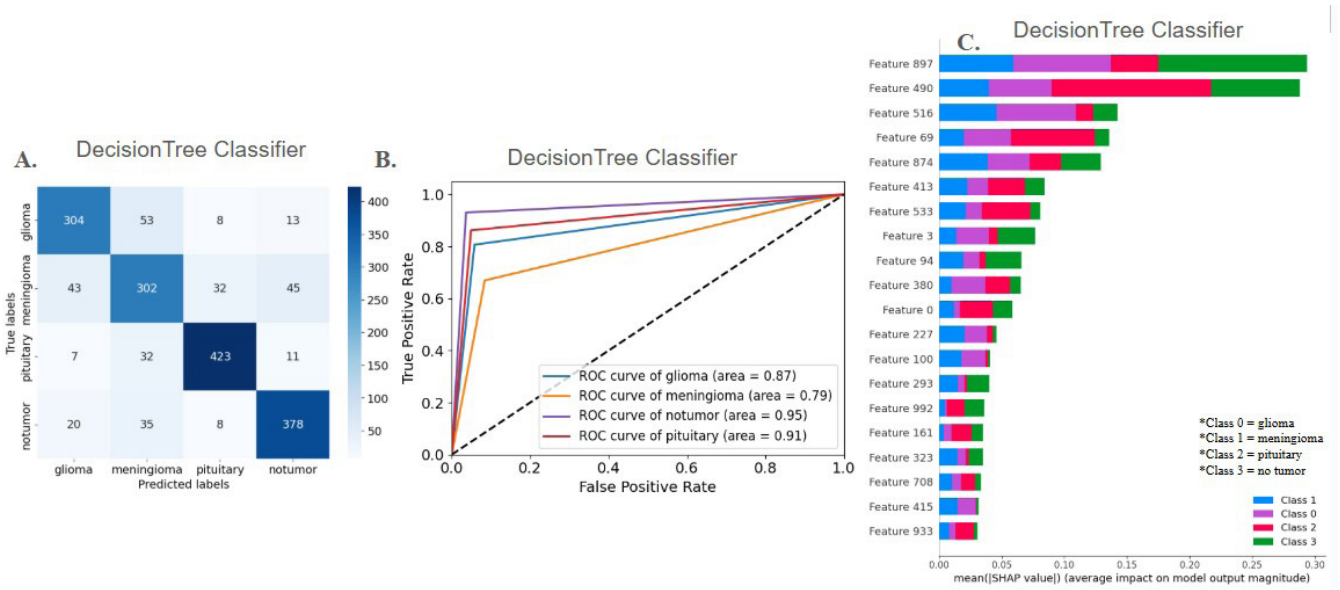

Figure 6: Performance Evaluation and Feature Importance for the Decision Tree Classifier. (A) Confusion matrix showing true versus predicted labels for glioma, meningioma, pituitary tumors, and non-tumorous cases. The matrix displays the number of correctly and incorrectly classified images for each category. (B) ROC curves demonstrate varying classification accuracy, with AUC values of 0.87 for gliomas, 0.79 for meningiomas, 0.91 for pituitary tumors, and 0.95 for non-tumorous cases. (C) SHAP summary plot for feature importance, where each horizontal bar represents the mean absolute SHAP value (average impact on model output magnitude) for a given feature. Bars are color-coded by class: Class 0 = glioma, Class 1 = meningioma, Class 2 = pituitary, Class 3 = non-tumor. The top features correspond to pixel intensity values at specific MRI image coordinates, with higher-ranked features contributing more strongly to classification decisions.

particularly for pituitary tumors, and aligned closely with the patterns highlighted in our saliency maps. Together, these interpretability tools show that the models learned clinically relevant features rather than relying on background noise.

Benchmarking against expert neuroradiologists is a critical step for clinical adoption. MRI sensitivity can range from 80–95%, influenced by tumor type and size, establishing a performance standard that ML models must meet or exceed to be considered viable decision-support tools (16). In this study, DenseNet and CNN models demonstrated high accuracy and AUC scores, underscoring their potential to complement human expertise. Validation on diverse, real-world datasets remains essential to confirm these findings and ensure clinical-grade reliability.

Beyond classification, ML models have the potential to predict clinically relevant variables such as prognosis, treatment stratification, recurrence, and survival. While this study focused on tumor classification, its findings demonstrate that ML models can serve as foundational tools for broader diagnostic workflows. Collaborative efforts to create richer, annotated datasets, such as biobanking initiatives, could help generate diverse datasets that support the development of clinically robust models.

In conclusion, this study highlights the potential of ML models, particularly simpler methods like k-NN, and transfer learning-based approaches, like DenseNet and CNN, to improve the accuracy and efficiency of brain tumor classification. By aligning our findings with the initial hypotheses, the models demonstrated precise multi-class classification capabilities and reliance on critical image features. While challenges such as dataset diversity and clinical validation remain, these findings underscore the promise of ML tools in reducing diagnostic variability and improving patient outcomes. By enabling earlier and more reliable tumor detection, these models have the potential to significantly impact cancer care, reduce diagnostic errors, and improve survival rates.

MATERIALS AND METHODS

Study dataset

In this study, a dataset of 7,023 T1-weighted contrast-enhanced axial MRI images was used. The dataset was categorized into four balanced classes: glioma (23.1%; 1,620 images), meningioma (23.4%; 1,643 images), pituitary tumor (25.0%; 1,756 images), and non-tumorous cases (28.5%; 2,004 images) (15). The dataset included pre-segmented brain slices focusing on tumor regions, making it suitable for classification tasks.

The images were preprocessed by resizing to 32x32 pixels, flattening to grayscale, and normalizing pixel intensity values to a range of 0–1 by dividing each pixel value by 255. Memory usage and training efficiency were improved through this preprocessing. The preprocessed images were stored in numpy arrays for computational efficiency. Data visualization, including bar graphs and charts, was used to examine the distribution of tumor types. Data augmentation, including horizontal flips and rotations, was applied to the training set, effectively doubling the dataset to 11,418 images and reducing the risk of overfitting. The dataset was divided into 80% for training, 10% for validation, and 10% for testing. The scans were randomly assigned to these groups using stratified sampling to maintain the proportions of glioma, meningioma, pituitary tumors, and non-tumorous cases across each subset. Specifically, 80% of the images were allocated for training (5,618 images), 10% for validation (702 images), and 10% for testing (703 images).

Five ML models were implemented: Decision Tree, k-Nearest Neighbors (k-NN), DenseNet, ResNet50, and a

https://doi.org/10.59720/25-127

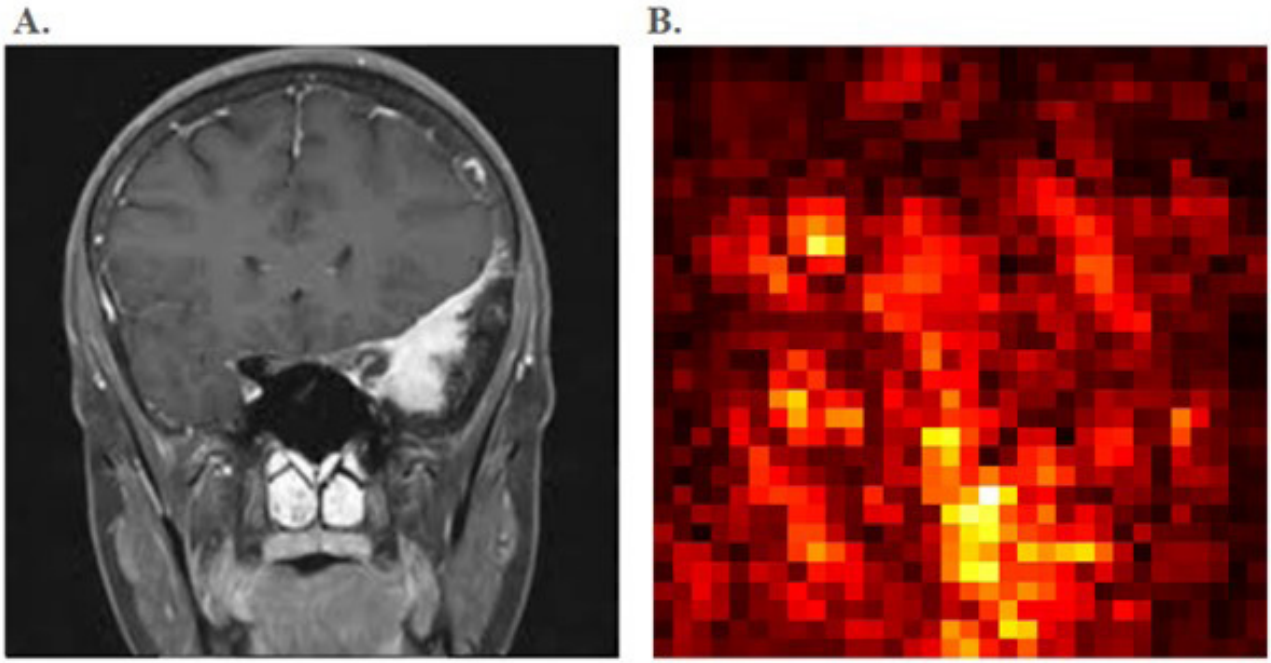

Figure 7: Saliency Map of MRI Brain Scan Generated Using the DenseNet Model. (A) Original MRI brain scan and (B) the corresponding saliency map. Brighter regions on the saliency map highlight tumor boundaries and high-intensity areas identified as critical features for classification. The DenseNet model was evaluated on a subset of N = 1,317 images.

custom Convolutional Neural Network (CNN). The Decision Tree model was selected for its simplicity and interpretability, as data were split based on key features to form a tree-like structure. Hyperparameters such as splitting criteria, splitters, and maximum tree depths were optimized using GridSearchCV (17). A k-NN classifier was used for its strong performance in classifying images based on pixel values. In this model, labels were assigned by considering the majority vote among the closest data points, making it a straightforward and intuitive approach. To identify the optimal number of neighbors, a grid search was conducted over values ranging from 1 to 9 (18, 19). For each value, the model was assessed on the validation set, and the one with the highest accuracy was selected.

DenseNet and ResNet50 models, pre-trained on ImageNet, were adapted for multi-class classification by repeating grayscale images across three channels to match the input requirements. The top layers were replaced with a Global Average Pooling layer and a Dense layer using a SoftMax activation function. The custom CNN was constructed with two convolutional layers using ReLU activation and max pooling, followed by a flattening layer, a dense layer, and an output layer tailored for this dataset (18, 19). All transfer-based learning models were trained for 30 epochs with a batch size of 32, using the Adam optimizer at a learning rate of 0.001. Categorical cross-entropy was used as the loss function, measuring the difference between the predicted class probabilities and the true class labels, guiding the models to improve prediction accuracy during training.

Model performance was assessed using accuracy, confusion matrices, ROC curves, AUC scores, and F1 scores. Accuracy was defined as the proportion of correctly classified images, while confusion matrices were used to quantify misclassification patterns across tumor types. These metrics were computed from counts of true positives, true negatives, false positives, and false negatives. ROC curves evaluated sensitivity versus specificity across threshold values. The F1 score provided a single measure of classification performance by incorporating precision and recall for uneven class sizes. It was calculated as twice the product of precision and recall divided by the sum of precision and recall. Precision was defined as the proportion of correct positive predictions out of all positive predictions made by the model, while recall was defined as the proportion of actual positive cases that the model correctly identified (20). Additionally, the interpretability of model predictions was explored through saliency maps, aiming to provide transparency in decision-making and insights into key features driving model outputs. Bootstrap sampling with 1,000 iterations was performed to establish 95% confidence intervals for AUC and F1 scores, ensuring robust evaluation. Pairwise statistical comparisons of model performance were conducted using statistical tests to determine if observed differences were significant.

For interpretability, SHAP values were calculated for the Decision Tree model, visualized with SHAPley plots, while saliency maps were generated for DenseNet, ResNet50, and the custom CNN models to highlight key image features influencing classification decisions. Because each MRI slice was resized to 32x32 and flattened into a one-dimensional array, each feature in the SHAP analysis corresponds to a single pixel intensity value. The model therefore treats each pixel position as an individual input feature (0 to 1023).

Ethics Statement: This study did not involve human subjects. Data were sourced from a publicly available Kaggle repository and are fully anonymized, ensuring compliance with ethical standards (21). No additional ethical approval was required.

https://doi.org/10.59720/25-127

ACKNOWLEDGMENTS

We thank Ribhav Gupta, University of Minnesota Medical School, and Michael Bryan, Harvard Medical School, for their assistance with this research project, and to Inspirit AI for providing the opportunity to do this research.

Received: January 3, 2025
Accepted: February 20, 2026
Published: March __, 2026

REFERENCES

1. Cordeiro, Brittany. "Progress to Advance Care, Treatments, and Outcomes for Brain Tumor Patients." *Cancer.gov*, 25 Apr. 2022, www.cancer.gov/rare-brain-spine-tumor/blog/2022/advances-brain-tumors.
2. Arora, Yash, et al. "Brain Tumor Classification Using Weighted Least Square Twin Support Vector Machine with Fuzzy Hyperplane." *ScienceDirect*, Dec. 2024, https://doi.org/10.1016/j.engappai.2024.109450.
3. Pereira, Sergio, et al. "Automatic Brain Tumor Grading from MRI Data Using Convolutional Neural Networks and Quality Assessment." *Lecture Notes in Computer Science*, 1 Jan. 2018, pp. 106–14. https://doi.org/10.1007/978-3-030-02628-8_12.
4. Omofoye, Toma S., et al. "Backlogs in Formal Interpretation of Radiology Examinations: A Pilot Global Survey." *Clinical Imaging*, Feb. 2024. https://doi.org/10.1016/j.clinimag.2023.110049.
5. Lui, Y. W., et al. "Artificial Intelligence in Neuroradiology: Current Status and Future Directions." *AJNR: American Journal of Neuroradiology*, vol. 41, no. 8, 1 Aug. 2020, pp. E52–E59, www.ncbi.nlm.nih.gov/pmc/articles/PMC7658873/. https://doi.org/10.3174/ajnr.A6681.
6. Trauernicht, Christoph, et al. "Medical Physics Services in Radiology and Nuclear Medicine in Africa: Challenges and Opportunities Identified Through Workforce and Infrastructure Surveys." *Health and Technology*, vol. 12, no. 4, 31 Mar. 2022, pp. 729–37. https://doi.org/10.1007/s12553-022-00663-w.
7. Cepeda, Santiago, et al. "Predicting Regions of Local Recurrence in Glioblastomas Using Voxel-Based Radiomic Features of Multiparametric Postoperative MRI." *Cancers*, vol. 15, no. 6, 1 Jan. 2023, p. 1894. https://doi.org/10.3390/cancers15061894.
8. Gupta, Jaya, et al. "Approaches to Motivate Physicians and Nurses in Low- and Middle-Income Countries: A Systematic Literature Review." *Human Resources for Health*, vol. 19, no. 1, 6 Jan. 2021. https://doi.org/10.1186/s12960-020-00522-7.
9. Rigsby, R. K., et al. "Newly Recognized CNS Tumors in the 2021 World Health Organization Classification: Imaging Overview with Histopathologic and Genetic Correlation." *American Journal of Neuroradiology*, vol. 44, no. 4, 1 Apr. 2023, pp. 367–80. https://doi.org/10.3174/ajnr.A7827.
10. Chen, Jieneng, et al. "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation." *arXiv*, 8 Feb. 2021, https://doi.org/10.48550/arXiv.2102.04306.
11. Panhalkar, Archana R., and Dharmpal D. Doye. "A Novel Approach to Build Accurate and Diverse Decision Tree Forest." Evolutionary Intelligence, vol. 15, no. 1, 3 Jan. 2021, pp. 439–453, https://doi.org/10.1007/s12065-020-00519-0.
12. Hu, Li-Yu, et al. "The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets." SpringerPlus, vol. 5, no. 1, 9 Aug. 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC4978658/, https://doi.org/10.1186/s40064-016-2941-7.
13. Kensert, Alexander, et al. "Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes." SLAS DISCOVERY: Advancing the Science of Drug Discovery, vol. 24, no. 4, 14 Jan. 2019, pp. 466–475, https://doi.org/10.1177/2472555218818756.
14. Amin, Mohammad, et al. "DieT Transformer model with PCA-ADE integration for advanced multi-class brain tumor classification." *ScienceDirect*, 2025, https://doi.org/10.1016/j.ibmed.2024.100192.
15. Mandadapu, Pranay. "Automated Human Activity Recognition from Controlled Environment Videos" *Minds at UW*, 1 Dec. 2023, https://digital.library.wisc.edu/1793/93472.
16. "Brain MRI and Tumor Detection: Accuracy, Limitations, and Alternatives" *Neurolaunch*, 30 Sep. 2024, https://neurolaunch.com/can-an-mri-detect-a-brain-tumor/.
17. Liashchynskyi, Petro, and Pavlo Liashchynskyi. "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS." *ArXiv preprint arXiv:1912.06059*, 12 Dec. 2019, https://arxiv.org/abs/1912.06059.
18. Simoes, Rodrigo A.S. "Network Resilient Federated Machine Learning." *Ciencias ULisboa*, 2023, https://www.di.fc.ul.pt/~nuno/THESIS/RodrigoSimoes_MSI24.pdf.
19. Ramalho, Gabriel M.F. "Structural Health Monitoring of Adhesive Joints Using Lamb Waves: A Review." *ResearchGate*, Sept. 2021, https://onlinelibrary.wiley.com/doi/epdf/10.1002/stc.2849.
20. Hicks, Steven, et al. "On Evaluation Metrics for Medical Applications of Artificial Intelligence" *Scientific Reports*, 8 Apr. 2022, https://pmc.ncbi.nlm.nih.gov/articles/PMC8993826/.
21. Nickparvar, Masoud. "Brain Tumor MRI Dataset." *Kaggle*, 2021, https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.