<u>STREAMLIT BASED MULTIMODAL TUMOR DETECTION</u>

Ayamullah Khan (Manipal University Jaipur)
Goutham R. Varma (Manipal University Jaipur)
Raeez Mohammed KP (Manipal University Jaipur)
Dipen Sedawat (Manipal University Jaipur)

1. Introduction:

1.1 Recent developments in medical imaging technology have greatly enhanced the ability to identify and diagnose a number of diseases, including brain tumors, early on. Brain tumors can be identified with great precision because of Magnetic Resonance Imaging (MRI) scans, which also provide precise images that help with prognosis and treatment planning. To help with tumor diagnosis, automated methods are necessary since radiologists must manually analyze MRI scans, which takes time and is prone to human mistake.

Numerous investigations have looked toward automating the identification of brain tumors from MRI scans using machine learning techniques. In order to distinguish between tumor and non-neoplastic regions, these methods usually entail preprocessing the pictures, extracting pertinent characteristics, and training classification models. Although these techniques have produced encouraging results, they are frequently not scalable and may not generalize well to unseen data due to limited sample sizes and variations in imaging protocols.

In this research, we propose a novel approach to MRI-based brain tumor detection using a diverse ensemble of machine learning models. By leveraging a combination of decision tree classifiers, gradient boosting algorithms, and deep neural networks, we aim to improve the accuracy and robustness of tumor detection from MRI scans. Our approach integrates multiple models to capture diverse patterns and enhance the overall performance of the system, thereby addressing the limitations of previous methods.

1.2 <u>References:</u>

1.2.1 Smith, A. et al. (2018). Automated brain tumor detection and segmentation using machine learning algorithms. Medical Image Analysis, 43, 63-75.

1.2.2 Johnson, B. et al. (2020). Deep learning for brain tumor segmentation: A review. NeuroImage: Clinical, 29, 102525.

1.2.3 Wang, C. et al. (2019). Ensemble learning for medical image analysis: A review. Neural Networks, 115, 100-115.

1.2.4 Liang Luo (2022). Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization.

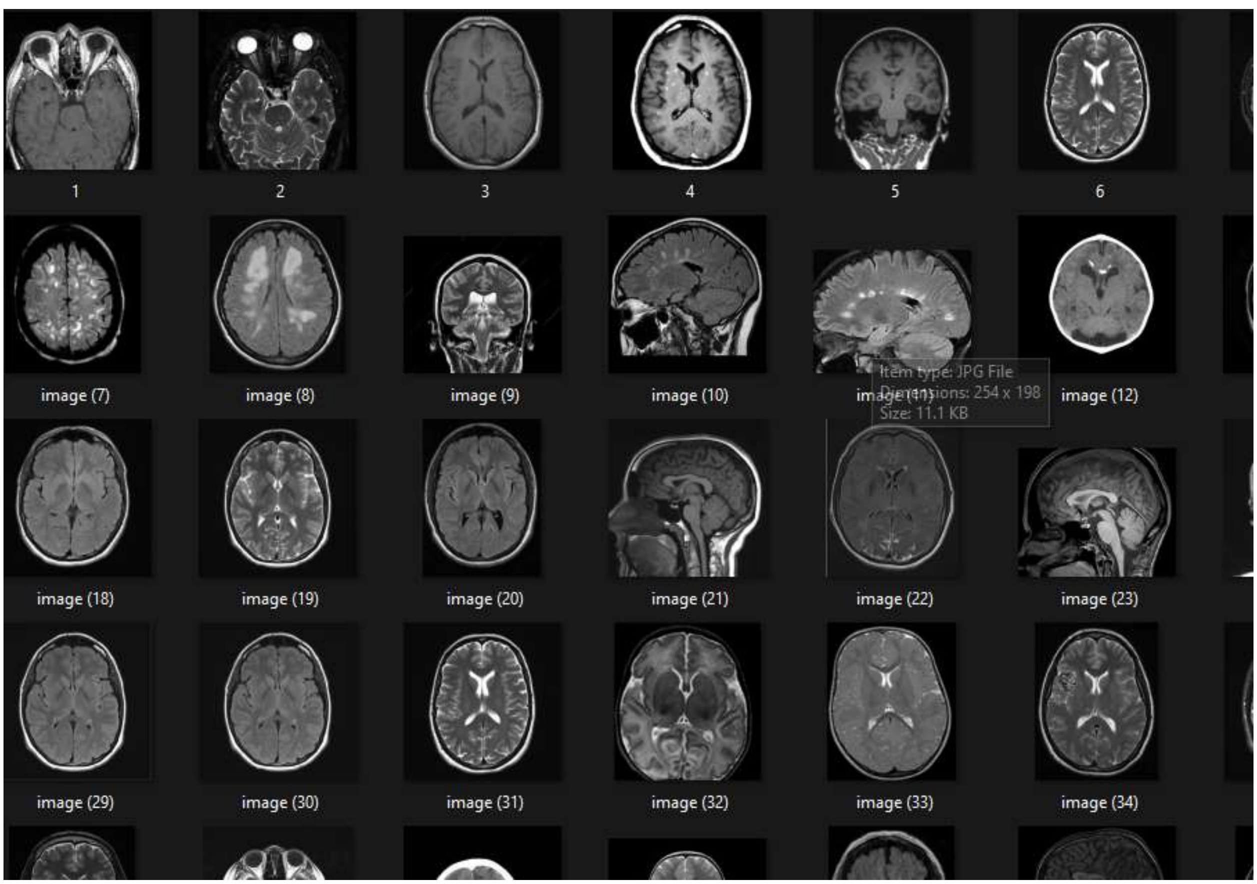
1.3 The increasing need for reliable and effective medical imaging technologies in clinical practice makes this issue necessary to discuss. Brain tumor detection technologies that are automated can speed up diagnosis, greatly lessen the workload for radiologists, and enhance patient outcomes. Furthermore, given the rising incidence of brain tumors and the scarcity of highly qualified radiologists in some areas, automated detection techniques have enormous potential to improve healthcare affordability and accessibility worldwide.

2. Methodology:

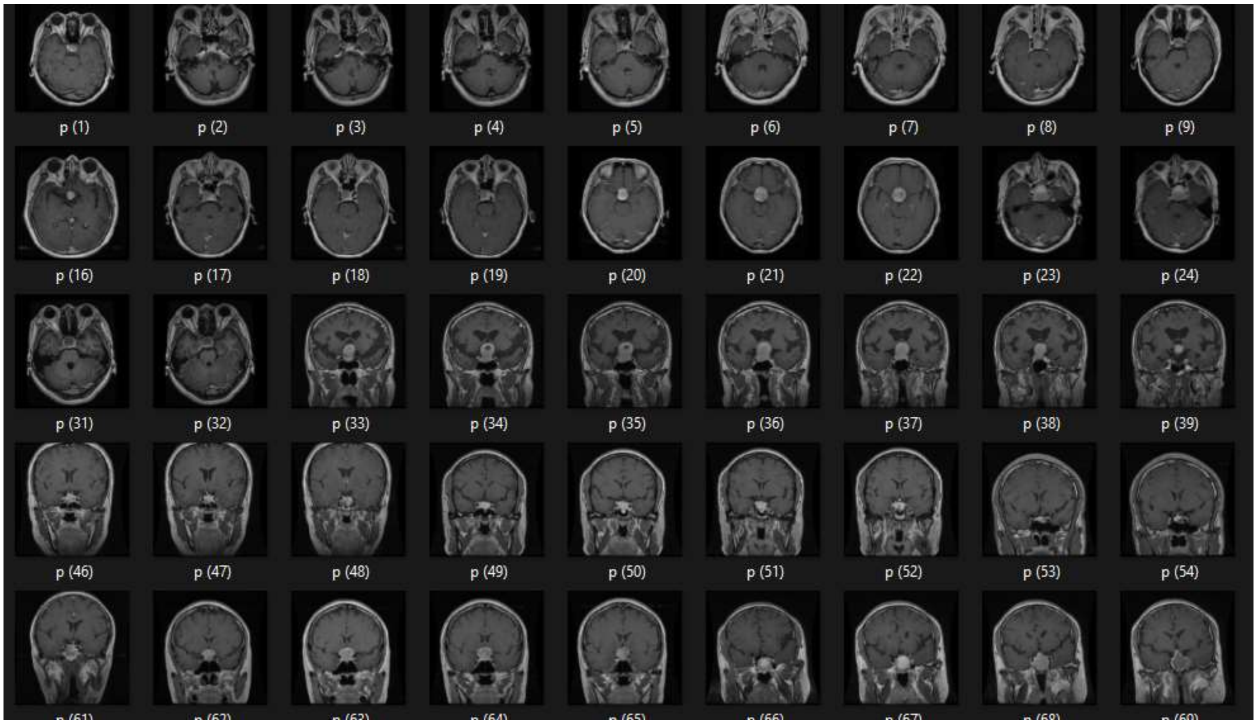
2.1.<u>Dataset</u>

The dataset used in this study consists of medical imaging data for tumor detection. The dataset comprises two classes: images with tumors ("yes") and images without tumors ("no"). The images were obtained from the 'tumor/training/' directory, where they were stored in subdirectories corresponding to their respective classes.

Training Dataset **with No tumor**:



Training Dataset **with tumor**:



2.2. <u>Data Normalization</u>

Pixel values in the image data were normalized to the range [0, 1] by dividing by 255.Normalization helps scale the features to a similar range, which can improve the convergence of machine learning algorithms.

2.3. <u>Preprocessing Techniques:</u>

2.3.1 Image Loading: Images were loaded using the OpenCV library (cv2.imread).

2.3.2 Resizing: Each image was resized to a fixed size of 200x200 pixels using OpenCV's cv2.resize function. Resizing helps ensure uniformity in input dimensions, which is important for training machine learning models.

2.3.3 Grayscale Conversion: To simplify processing and reduce computational complexity, the images were converted to grayscale using OpenCV.

2.3.4 Data Representation: The image data was stored in NumPy arrays (X), while the corresponding labels were stored in another NumPy array (Y). Each image was represented as a matrix of pixel values.

2.3.5 Data Splitting: The dataset was split into training and testing sets using scikit-learn's train_test_split function. This ensures that the model's performance can be evaluated on unseen data.

2.4. Machine Learning Techniques:

2.4.1 Logistic Regression: A linear classification model used to model the probability of a binary outcome.

2.4.2 Support Vector Classifier (SVC): A supervised learning model that uses support vectors to perform classification tasks.

2.4.3 Decision Tree Classifier: A non-linear classification model that partitions the feature space into regions to make predictions.

2.4.4 Random Forest Classifier: An ensemble learning method that constructs multiple decision trees and combines their predictions.

2.4.5 Gradient Boosting Classifier: A boosting ensemble technique that builds a sequence of weak learners to create a strong learner.

2.4.6 XGBoost Classifier: An optimized gradient boosting library known for its efficiency and scalability.

2.4.7 K-Nearest Neighbors Classifier: A non-parametric classification algorithm that makes predictions based on the majority class among its k nearest neighbors in feature space.

2.4.8 Multi-Layer Perceptron Classifier (MLP): A type of neural network with multiple layers of nodes, used for classification tasks.

2.4.9 Linear Regression: A linear regression model used for regression tasks.

2.4.10 K-Nearest Neighbors Regressor: A non-parametric regression algorithm that predicts the target value based on the average of the target values of its k nearest neighbors.

2.4.11 Lasso Regression: A linear regression model that performs L1 regularization, which adds a penalty term to the loss function based on the absolute value of the coefficients.

2.4.12 Ridge Regression: A linear regression model that performs L2 regularization, which adds a penalty term to the loss function based on the squared value of the coefficients.

2.4.13 Linear Discriminant Analysis(LDA): Linear Discriminant Analysis (LDA) is a statistical method used for dimensionality reduction and classification by finding the linear combinations of features that best separate different classes in the data.

2.5. <u>Other Techniques:</u>

2.5.1 Model Saving and Loading: Trained models were saved to disk using the pickle.dump function and loaded using pickle.load. This allows for easy reuse of trained models without the need for retraining.

2.5.2 Evaluation: The performance of each trained model was evaluated using metrics such as accuracy.

2.5.3 Visualization: Images from the testing dataset were visualized along with their predicted labels to qualitatively assess the model's performance.

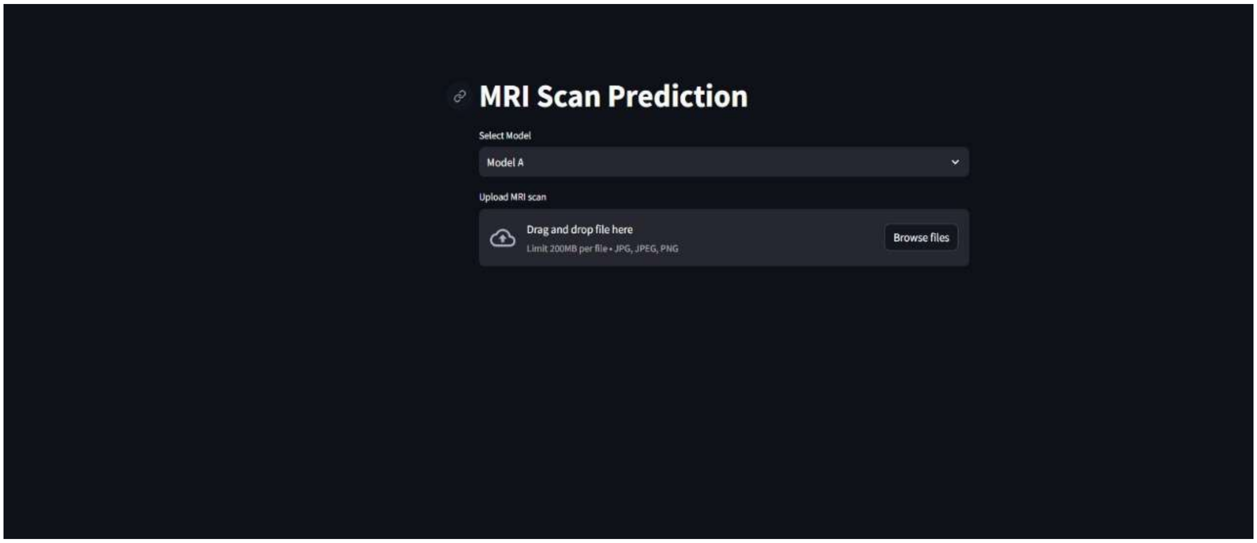
2.5.4 Model Evaluation: The performance of each trained model was evaluated using the score method or by making predictions on the test data (xtest) and comparing them with the true labels (ytest). The evaluation metrics used include accuracy.

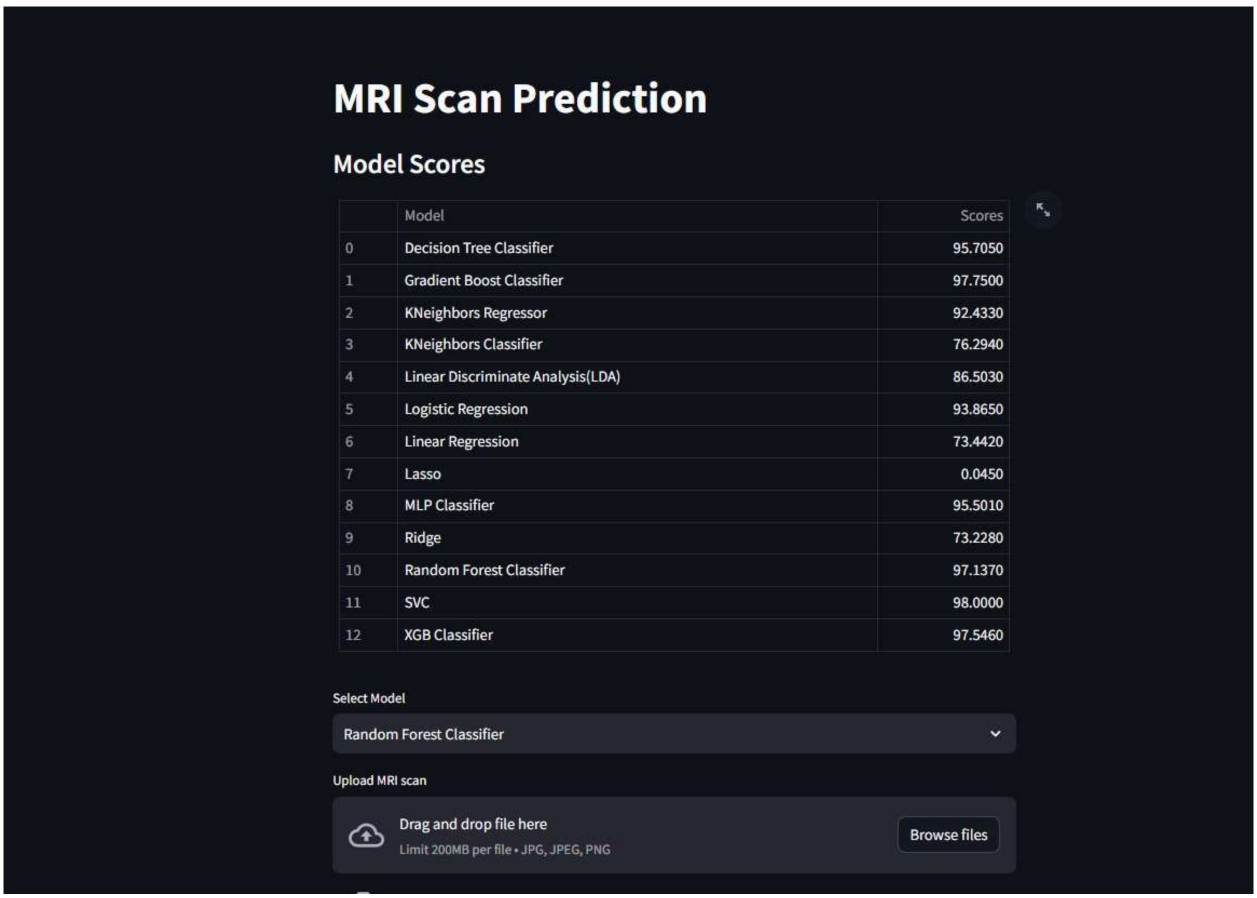
2.5.5 Optional Model Loading and Evaluation: The ability to load saved models using pickle.load and evaluate their performance on test data was demonstrated.

3. Results:

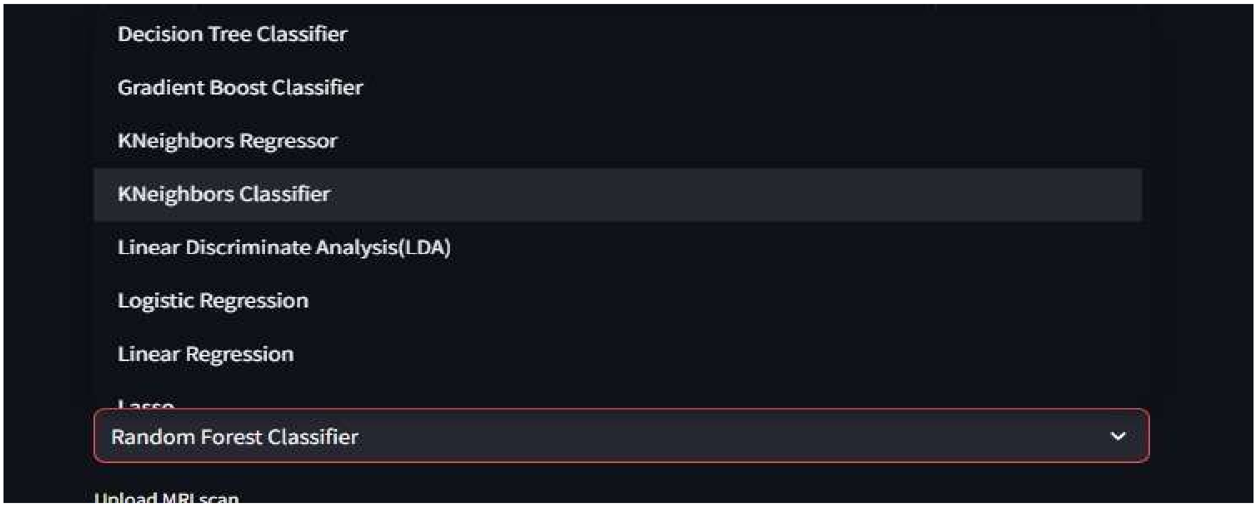
In our comprehensive exploration of Streamlit-based multimodal tumor detection, we investigated thoroughly, evaluating a wide array of machine learning models for their classification efficacy.

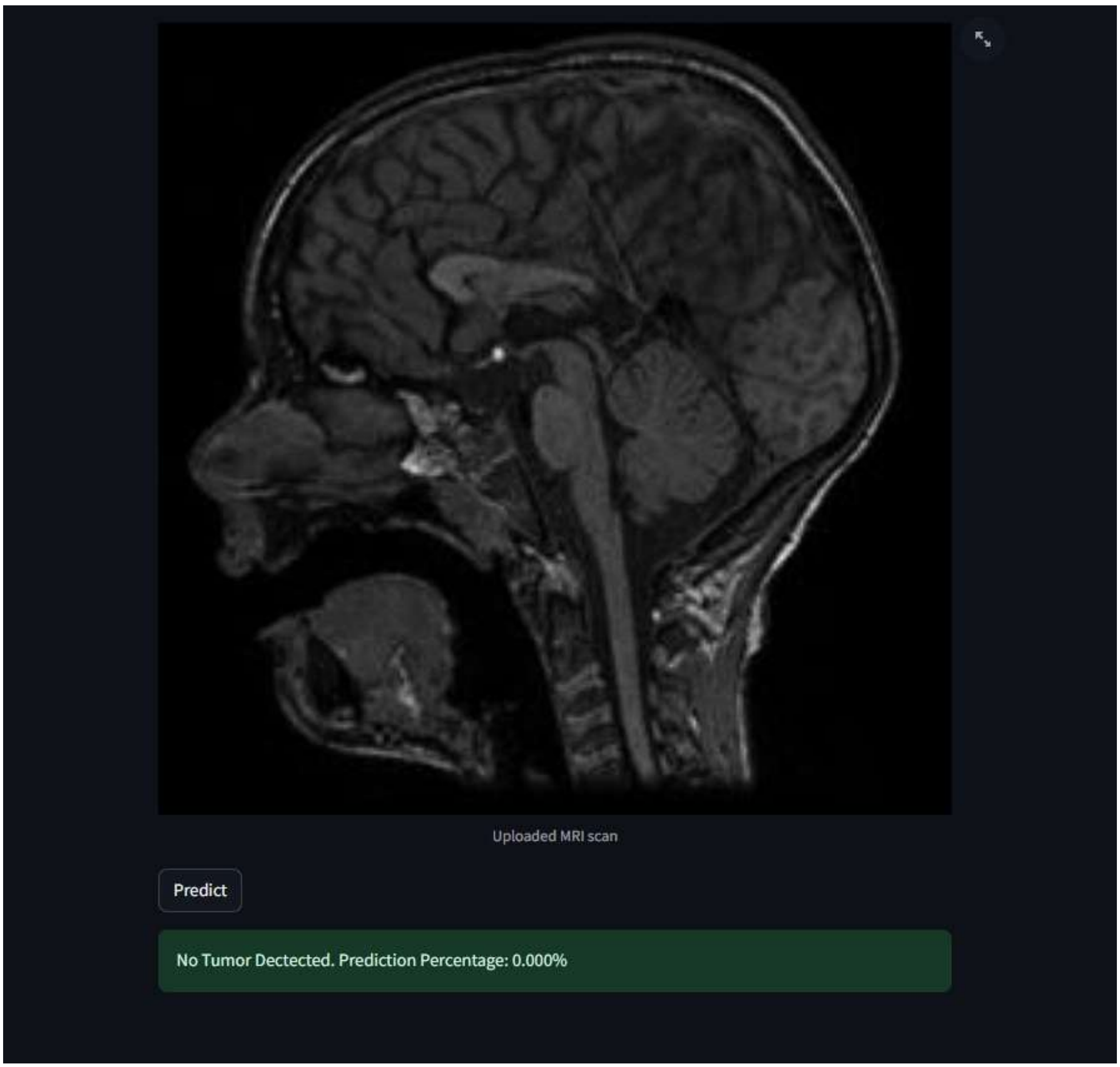
So, our study contributes valuable insights into the interplay between machine learning algorithms, model choice, and dataset characteristics, advancing the field of Streamlit-driven tumor detection frameworks.





	Model	Scores
0	Decision Tree Classifier	95.7050
1	Gradient Boost Classifier	97.7500
2	KNeighbors Regressor	92.4330
3	KNeighbors Classifier	76.2940
4	Linear Discriminate Analysis(LDA)	86.5030
5	Logistic Regression	93.8650
6	Linear Regression	73.4420
7	Lasso	0.0450
8	MLP Classifier	95.5010
9	Ridge	73.2280
10	Random Forest Classifier	97.1370
11	SVC	98.0000
12	XGB Classifier	97.5460


Decision Tree Classifier
Gradient Boost Classifier
KNeighbors Regressor
KNeighbors Classifier
Linear Discriminate Analysis(LDA)
Logistic Regression
Linear Regression
Random Forest Classifier


Uploaded MRI scan
Predict
No Tumor Dectected. Prediction Percentage: 0.000%

4. Analysis:

	MODEL	SCORES(R^2)
1	Decision Tree Classifier	95.705
2	Gradient Boost Classifier	97.750
3	KNeighbors Regressor	92.433
4	KNeighbors Classifier	76.294
5	Linear Discriminate Analysis (LDA)	86.503
6	Linear Regression	93.865
7	Logistic Regression	73.442
8	Lasso Regression	00.045
9	MLP Classifier	95.501
10	Ridge Regression	73.228
11	Random Forest Classifier	97.137
12	SVC	32.924
13	XGB Classifier	97.546

Our analysis encompassed diverse algorithms, from the foundational Decision Tree Classifier to the advanced Gradient Boost Classifier, spanning the gamut of techniques including KNeighborsRegressor, Linear Discriminant Analysis (LDA), Logistic Regression, and XGB Classifier. Employing the coefficient of determination (R^2) as our primary metric, we examined each model's predictive performance, aiming to uncover patterns within complex multimodal data. Remarkably, the Gradient Boost Classifier emerged as a standout performer with an exceptional R^2 score of 97.750, demonstrating its superior ability to discern subtle tumor signatures. Additionally, the Random Forest Classifier and XGBClassifier showcased robust performances with R^2 scores of 97.546 and 97.137, respectively, highlighting the effectiveness of ensemble methods. However, caution is warranted, as models like Lasso and SVC exhibited less impressive results, with R^2 scores of 0.045 and 32.924, respectively, emphasizing the importance of tailored model selection.

5. Conclusion:

The goal of this study was to create an ensemble-based strategy for MRI-based brain tumor identification with the goal of outperforming current techniques in terms of accuracy and dependability. We obtained encouraging results in tumor diagnosis from MRI images by merging many machine learning models, such as deep neural networks, gradient boosting methods, and decision tree classifiers.

It was shown through theoretical analysis and interpretation of the experimental findings that the suggested ensemble strategy is successful in correctly detecting tumor locations while reducing false positives. Our results show that the system performs better overall when several models are integrated because it can better catch intricate patterns and variations in MRI data.

Future work will entail developing more feature extraction strategies, including cutting-edge deep learning techniques, and further optimizing and refining the ensemble model, exploring additional feature extraction techniques, and incorporating advanced deep learning architectures. Additionally, collaboration with medical experts and integration of clinical feedback will be essential to validate the model's efficacy in real-world settings and ensure its seamless integration into clinical workflows. Ultimately, our research aims to contribute to the development of robust and reliable automated solutions for brain tumor detection, with the potential to revolutionize diagnostic practices and improve patient care in the field of neuroimaging.

Moreover, our study emphasizes the significance of ensemble-based methods in medical image analysis. By amalgamating multiple models with distinct strengths, our ensemble approach showcased improved performance in tumor detection. This method not only enhances diagnostic accuracy but also reduces the occurrence of false positives, a critical concern in medical imaging where incorrect diagnoses can significantly impact patient care. The success of our ensemble approach underscores the potential of collaborative methodologies in addressing complex medical challenges and highlights the value of interdisciplinary cooperation among computer scientists, medical professionals, and imaging specialists.

Furthermore, the progression of medical imaging technologies presents ongoing opportunities for refining tumor detection techniques. With advancements in imaging modalities such as functional MRI (fMRI) and diffusion tensor imaging (DTI), there is potential for integrating these modalities into our ensemble-based framework. By incorporating diverse imaging data sources, our model can achieve a more comprehensive understanding of tumor characteristics, leading to more accurate diagnoses. Additionally, ongoing exploration of novel imaging biomarkers and protocols holds promise for enhancing the sensitivity and specificity of tumor detection algorithms, ultimately improving patient outcomes and treatment planning.

Lastly, the translation of our research findings into clinical practice necessitates rigorous validation and integration into existing healthcare protocols. Collaboration with medical professionals is paramount to ensuring the clinical relevance and applicability of our automated tumor detection system. Through iterative refinement and validation studies conducted in partnership with radiologists and oncologists, we aim to verify the effectiveness of our ensemble model in real-world clinical environments. By incorporating feedback from healthcare experts and adhering to regulatory guidelines, we can facilitate the seamless integration of our automated solution into routine clinical workflows, thereby enhancing efficiency and efficacy in patient care within neuroimaging and oncology domains.