Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 241, Issue 44, September/October 2025, Pages 1–16
https://doi.org/10.7250/csimq.2025-44.01

Deep Inception Based Hybrid Machine Learning Framework for Binary Classification of Brain Tumor MRI Scans

Vineet Mehan*

Nims University, Rajasthan, Jaipur-Delhi Highway (NH-11C), Jaipur, 303121, India

vineet.mehan@nimsuniversity.org

Abstract. Accurate and early detection of Brain Tumors (BT) is pivotal to improve treatment planning and increase survival rates. A significant diagnostic system for the identification of brain disorders is Magnetic Resonance Imaging (MRI). In this study, a unique hybrid framework is developed by integrating InceptionV3, a deep learning model, with three machine learning models: AdaBoost, Random Forest (RF), and Logistic Regression (LR). High-dimensional spatial characteristics are extracted from pre-processed MRI data using the deep Inception model. Binary classification is then carried out by feeding these deep features into machine learning classifiers. Two hundred MRI images were used, half of which contained tumors and the other half of which did not. To ensure the reliability of the results, 50 distinct data splits and 10-fold cross-validation were employed. With an accuracy rate of 98.2% and an Area Under Curve (AUC) of 0.999, LR was the most successful. Next was RF, which had an accuracy of 94.6% and an AUC of 0.98. AdaBoost got an AUC of 0.874 and an accuracy of 87.4%. Experimental results prove that the hybrid technique achieves better classification accuracy and fewer false positives. The proposed framework is thus appropriate for clinical decision assistance since it strikes a compromise between learning depth and decision interpretability through the combination of deep feature representations and classifiers.

Keywords: Brain Tumor Classification, MRI Imaging, InceptionV3, Machine Learning, Hybrid Framework.

1 Introduction

One of the most serious and sometimes lethal neurological disorders is Brain Tumors (BT), which develop when cells in the brain or surrounding tissues grow abnormally [1], [2]. These tumors have the potential to significantly impair motor skills, cognitive capacities, and overall quality of life, depending on their location and pace of growth. There are two types of BT, malignant and benign, each with distinct prognosis and clinical features [3], [4].

Early and accurate detection of BT is essential for deciding on the appropriate treatment option,

*Corresponding author

Reference: V. Mehan, "Deep Inception Based Hybrid Machine Learning Framework for Binary Classification of Brain Tumor MRI Scans," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 44, pp. 1–16, 2025. Available: https://doi.org/10.7250/csimq.2025-44.01

Additional information. Author's ORCID iD: V. Mehan – https://orcid.org/0000-0003-3483-0160. S225599222500241X. Received: 16 August 2025. Accepted: 27 October 2025. Available online: 31 October 2025.

which may include radiation, chemotherapy, or surgery [5]. Death or irreversible brain damage might result from a delayed diagnosis. Because of this, the treatment of BT depends on diagnostic imaging. Magnetic Resonance Imaging (MRI) [6] is currently used for detecting and monitoring BT due to its high-resolution and multi-plane imaging capabilities.

However, manually interpreting MRI scans is a time-consuming, skill-dependent process that is subject to inter-observer variability. Clinical environments with high patient load and limited access to experienced radiologists are more likely to experience misdiagnosis or delayed diagnosis. The use of AI to automatically detect and classify brain tumors using MRI is therefore becoming popular.

Automated diagnosis has been improved by the combination of Deep Learning (DL) and Machine Learning (ML) algorithms. Convolutional Neural Networks (CNNs) are particularly good at extracting features for the categorization of brain tumors [7]. CNNs, however, require a lot of labeled data and processing power. Effective feature extraction from small datasets is made possible by transfer learning with pre-trained models such as InceptionV3 [8], VGG [9], and ResNet [10], which lessens these difficulties. Although CNNs employ deep layers for classification, their clinical application is limited by their low interpretability and complexity. In contrast, more efficient and interpretable options for final classification are provided by traditional machine learning classifiers such as Random Forest (RF), AdaBoost, and Logical Regression (LR) [11], [12].

The hybrid technique combines CNN-based feature extraction with ML classifiers to improve generalization on sparse medical datasets while lowering processing costs. These technologies will help radiologists by standardizing diagnosis, uncovering ambiguous cases, and offering second opinions. In addition to improving patient outcomes and healthcare delivery by improving diagnostic efficiency and accuracy, these automated procedures expedite decision-making and reduce human error.

The objective of this study is to make and assess a hybrid framework for the binary categorization of brain MRI data that blends DL and ML. Accurately and effectively differentiating between tumor-positive and tumor-negative MRI images while maintaining computational viability and model interpretability.

The suggested method uses transfer learning to overcome data scarcity and shorten training times by extracting high-level deep features from MRI images using a pre-trained InceptionV3 model. Following that, these characteristics are categorized using three machine learning algorithms, AdaBoost, RF, and LR, that were chosen for their performance, robustness, and interpretability. Identifying the best classifier in the hybrid architecture and showcasing the framework's potential as a portable, easily comprehensible clinical decision support tool are the goals of the study.

The rest of the article is structured as follows. The literature survey is presented in Section 2. The application of InceptionV3 and ML models is explained in Section 3. The classification methodology is described in Section 4. The results are presented and discussed in Section 5. Section 6 concludes the article. Any abbreviations used in the article are amalgamated in the Appendix for the convenience of readers.

2 Literature Survey

Brain tumor detection utilizing MRI data is now much more accurate and efficient because of current advances in AI. To boost classification outcomes and aid in clinical decision-making, several studies have investigated different architectures, preprocessing methods, and hybrid models. The literature review that follows identifies significant contributions and current research gaps in this field.

The study in [13] presents a deep learning approach for brain tumor diagnosis using multimodal MRI data, including T1-weighted, T2-weighted, and diffusion-weighted scans. By integrating CNNs with RNNs and employing data augmentation, the model achieved 92.3% accuracy and an AUC of 0.95, outperforming single-modality methods. However, while data augmentation was used to

mitigate limited labeled data, the study did not explore the impact of different augmentation strategies to assess the model's generalizability across varying imaging conditions and populations.

In order to improve MRI quality and highlight tumor regions, the study in [14] suggests a deep learning method that makes use of CNNs and sophisticated image preprocessing techniques, including morphological operations, data augmentation, and histogram equalization. Transfer learning produces excellent recall and precision by further enhancing categorization accuracy. The study does not address constraints in model generalizability across varied populations and changes in MRI equipment, which may impair real-world clinical performance, even though the technique shows promise for clinical use in neuro-oncology.

CNN-based DL for BT diagnosis utilizing MRI data is presented in [15]. The model gives information on the location, size, and type of tumors in addition to accurately detecting their presence. This strategy seeks to improve patient care in neuro-oncology by increasing diagnosis accuracy and assisting physicians in making well-informed treatment decisions.

With an accuracy of 97% to 99%, [16] presents a deep learning method for brain tumor diagnosis. The algorithm successfully differentiates between tumorous and non-tumorous areas, improving diagnostic accuracy and lowering human error. Even if the approach performs well, the learning underlines the inevitability of multi-modal imaging approaches in future research to increase the generalizability.

The area of study in [17] is to use MRI-based machine learning to improve diagnostic accuracy in brain tumor diagnosis. Using a sizable annotated dataset covering a range of tumor kinds and stages for training, validation, and testing. CNNs are used to retrieve complicated characteristics from pre-processed MRI images. By categorizing brain tumors accurately, the method advances cancer imaging. However, the study does not specifically address current research shortages in the field; instead, it focuses mostly on diagnostic improvements.

In order to detect aberrant brain cells, [18] suggests a brain tumor detection model that makes use of Improved Binomial Thresholding Segmentation (IBTBS). wavelet transform, wavelet scattering transform, and information-theoretic metrics are used to extract features. A Sparse Bayesian Extreme Learning Machine (SBELM) is used for classification after dimensionality is reduced by an Optimization-Based Feature Selection method (OBFSA). By increasing diagnostic speed and accuracy, this strategy hopes to promote early detection and perhaps improve patient survival rates.

The work in [19] investigates computer vision methods for image processing-based early brain tumor identification and classification. 253 MRI images were used as the dataset for preprocessing and data expansion. Initially, CNNs were employed for classification; however, performance was improved through transfer learning using pre-trained models. The accuracy of VGG-16 was the greatest among them, increasing from 84.61% to 92.31% after augmentation, demonstrating how well transfer learning works to improve tumor classification results.

With an emphasis on the categorization of tumor types such as gliomas, meningiomas, and metastases, the work in [20] assesses CNN-based deep learning models for automated interpretation of MRI data in brain tumor diagnosis. Evaluation metrics are used to assess model performance in order to enhance clinical diagnosis and treatment planning. An essential issue for further research is highlighted by the study's recognition of difficulties with model generalization, interpretability, and scalability across various patient groups and imaging settings.

The study in [21] integrates CNNs with hyperparameter optimization using Differential Evolution (DE), Harmony Search (HS), and Simulated Annealing (SA) to present an enhanced deep learning method for brain tumor diagnosis utilizing clinical MRI data. SA-CNN outperformed the other models in tumor detection, with the greatest classification accuracy of 99.63%. The urgent need for prompt and precise brain tumor diagnosis in clinical practice is addressed by this method.

Another study, [22], examines the area of ML and DL, namely CNNs, to improve diagnostic efficiency and emphasizes the vital significance that early brain tumor identification plays in enhancing patient survival. These techniques lessen the need for professional interpretation by

automating MRI scan processing and tumor type prediction. In order to assist efficient treatment and better clinical results, the study highlights the shortcomings of the time-consuming methods now in use and promotes automated solutions that use AI to provide prompt and precise diagnosis.

Key methods, datasets, strategies, and performance indicators from current research are compiled in Table 1.

Table 1. Summary of Recent Brain Tumor Diagnosis Studies Using ML/DL

Ref	Technique/Approach	Data Type / Modalities	Noted Limitations / Gaps
[13]	Deep Learning with CNN + RNN	Multimodal MRI (T1, T2, DWI)	No analysis on the effect of different augmentation strategies or generalizability
[14]	CNN with Advanced Image Processing	MRI (single modality)	No testing across varied MRI machines or patient populations
[15]	Deep CNN for Tumor Localization and Classification	MRI Scans	Describes output utility, lacks numerical performance comparison
[16]	CNN for Tumor Identification	Large MRI Dataset	Does not integrate multimodal imaging, a potential area for future exploration
[17]	Machine Learning (CNN-based)	Annotated MRI (Various Tumor Types/Stages)	Lacks explicit discussion on limitations or research gaps
[18]	Improved Binomial Thresholding + SBELM	MRI Images	Emphasis on speed and accuracy; lacks comparison with deep learning models
[19]	CNN + Transfer Learning (VGG-16, DenseNet, etc.)	MRI (253 images)	Highlights improvement post augmentation, but lacks a scalability discussion
[20]	CNN for Tumor Type Classification	MRI (Various Tumor Types incl. Metastasis)	Addresses generalization, interpretability, and scalability challenges
[21]	CNN with Hyperparameter Optimization	Clinical MRI	Focus on optimization; robustness across institutions has not been tested
[22]	ML + DL for Automated MRI Interpretation	MRI Reports	Addresses MRI interpretation efficiency; lacks details on real-time system integration

The reviewed research shows that deep learning and machine learning techniques, especially CNN-based architectures, have significantly improved the identification of brain tumors overall. However, even with excellent accuracy rates, many models have issues with clinical deployment, interpretability, and generalizability. Closing these gaps is crucial to creating real-time, scalable, and reliable diagnostic technologies that work in a variety of healthcare settings.

3 InceptionV3 and Machine Learning Models

InceptionV3 is a CNN that introduces the concept of Inception modules, which enable effective computation by combining several convolutional filters of diverse dimensions in parallel. This study uses transfer learning for brain MRI categorization. While the lower layers, which act as feature extractors of InceptionV3, are frozen to maintain broad image properties, the top layers are removed in order to obtain deep feature vectors. Equation 1 provides the following mathematical expression for the feature extraction process.

$$F = f\_CNN(X), \tag{1}$$

where X = Input MRI image, f_CNN = Pre-trained InceptionV3 model, and F = Extracted deep feature vector.

A powerful classifier is produced by combining many weak classifiers using the ensemble learning method AdaBoost [23], [24]. This study uses decision stumps as base estimators. Each subsequent model focuses on more samples that previous models misidentified. Equation 2 provides the final AdaBoost forecast.

$$H(x) = sign(\sum \alpha\_t * h\_t(x)), \quad (2)$$

where, α_t = Weight assigned to h_t, h_t(x) = Prediction from t^{th} weak classifier, and H(x) is the output.

A statistical model that mimics the likelihood of a binary occurrence is called logistic regression. It is simple to learn and works well with extracted attributes [25], [26]. Equation 3 illustrates how predictions are transformed into a probability between 0 and 1 via SF.

$$P(y=1|x) = 1 / (1 + e^{(-w^T x)}), \quad (3)$$

where, w = Weight vector, x = Input feature vector, and P(y=1|x) is the output.

Random Forest is a bagging-based ensemble classifier that builds several decision trees using random picks of data and attributes [27], [28]. The final forecast is made by majority voting, according to Equation 4. It resists overfitting and works well on high-dimensional feature spaces.

$$H(x) = mode\{h\_1(x), h\_2(x), ..., h\_n(x)\}, \quad (4)$$

where h_i(x) = Prediction from the i-th decision tree, and H(x) is the final strong classifier.

The effectiveness of the classification models may be evaluated using a set of metrics. These metrics give a comprehensive understanding of each model's capacity to distinguish between tumor-positive and tumor-negative MRI images. Key performance metrics include Accuracy (CA), Area under Curve (AUC), F1-score, Precision, Recall, and Matthews Correlation Coefficient (MCC).

4 Methodology

The proposed methodology classifies brain tumor MRI data in a binary format using a mix of DL and ML algorithms. The three main stages of the classification system are deep feature extraction using a pre-trained InceptionV3 model, classification using machine learning techniques, and performance evaluation utilizing robust validation approaches. The whole design leverages the powerful representation learning capabilities of CNNs while maintaining the interpretability and computational efficiency of machine learning classifiers.

The proposed algorithm is listed below and depicted in Figure 1.

1. Pre-processed MRI images are resized to 225 x 225 pixels.
2. The images are subjected to a frozen convolutional layer InceptionV3 model.
3. The feature vectors are extracted from the penultimate layer.
4. The collected features are then input into LR, RF, and AdaBoost.
5. 10-fold cross-validation and 50 stratified shuffle split repetitions are carried out.
6. ROC Analysis, Calibration Plot, CM, and Performance Curve further validate the model and provide a visual depiction.

The algorithm begins with the importing of MRI pictures, which are then tabulated using the Data Table module and visually examined using the Image Viewer. A column selection phase that follows makes sure that only pertinent data fields are kept for additional examination.

To ensure compliance with the input requirements of the InceptionV3 model, all input pictures are shrunk to a consistent dimension of 225 × 225 pixels during the preprocessing step. Consistent batch processing is another benefit of this scaling. In order to scale pixel intensities to the [−1,1] range and match the model's initial training, normalization is carried out.

After processing, the images are run through a frozen InceptionV3 network, which only uses the feature extraction layers. Deep feature representations, which are high-level embeddings of the

input data, are specifically taken out of the global average pooling layer. The downstream classifiers use these image embeddings as their input.

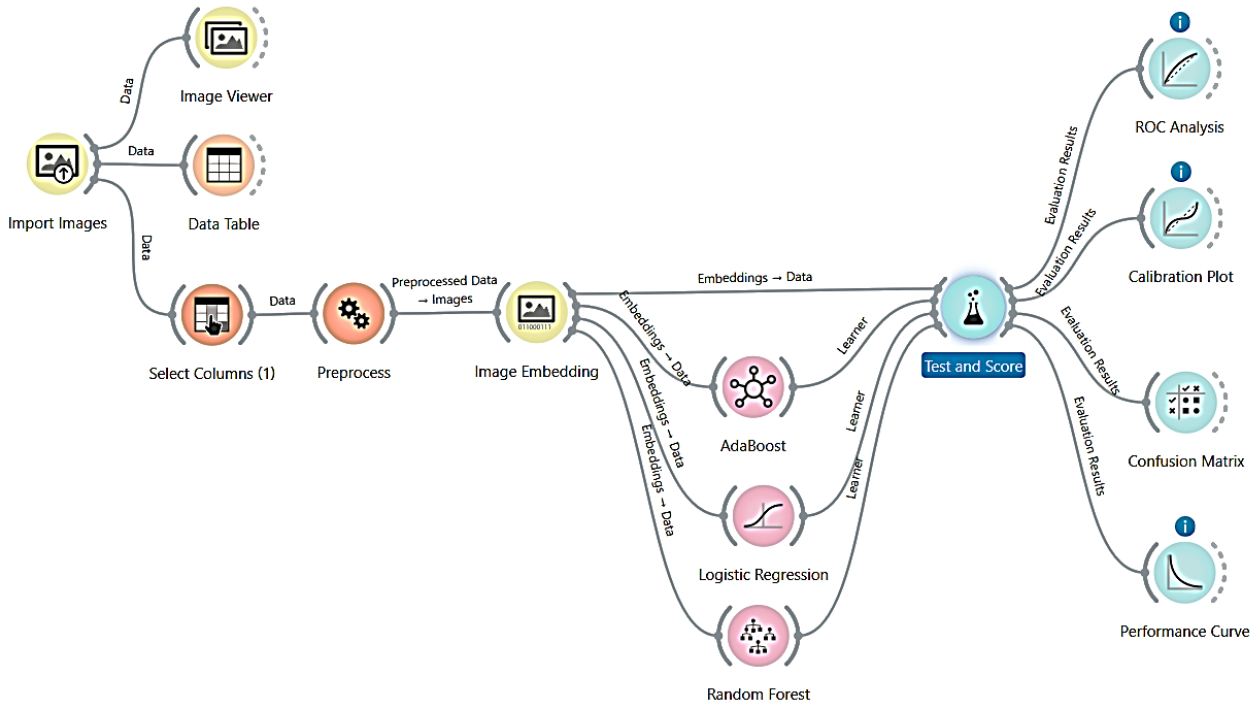


Figure 1. Proposed algorithm

Training of the retrieved feature vectors is done using LR, RF, and AdaBoost. To ensure the durability and generalizability of the model, a two-tier validation approach is employed. The creation of 10-Fold Cross-Validation divides the dataset into ten subgroups. In each iteration, one fold is used for testing and nine folds are used for training. After 10 attempts, the average performance is shown. The dataset is randomly split into a test ratio of 70:30 using repeated stratified sampling, which preserves class ratios. This process is performed fifty times to account for sample variance. The results are averaged to assess the model's stability.

The Test and Score module consolidates the evaluation results from all three classifiers. Several analytical techniques are used to evaluate performance measures both statistically and visually:

- Receiver Operating Characteristic (ROC) Analysis: Uses the Area Under the Curve (AUC) to assess discriminative capacity.
- Calibration Plot: Evaluates the degree of agreement between expected and actual probability.
- Confusion Matrix (CM): Offers information on prediction performance per class.
- Performance Curve: Provides an overview of the model's F1-score, recall, and accuracy across a range of criteria.

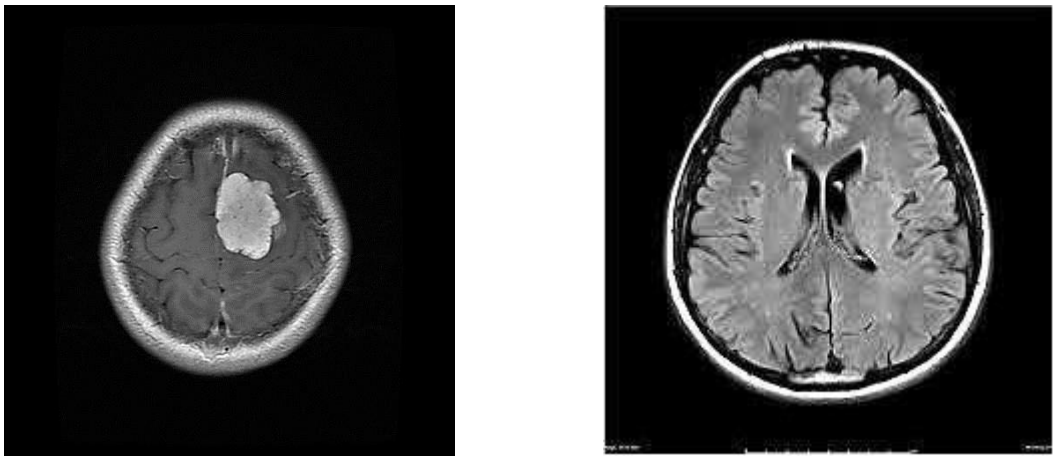
This extensive pipeline is appropriate for developing AI-driven diagnostic tools in medical imaging as it guarantees both thorough feature extraction and exact performance evaluation.

5 Results and Discussion

5.1 Dataset Description

The 200 MRI images in the dataset used for this study were specifically selected to facilitate the binary classification of brain tumors. With 100 pictures classified as tumor-positive and another 100 as tumor-negative, the dataset's class distribution is completely balanced. Each image displays a T1-weighted contrast-enhanced MRI scan of the human brain. This common imaging modality is used

extensively in clinical diagnostics due to its ability to accurately depict anatomical features and pathological abnormalities. Figure 2 shows a sample of Positive and Negative tumor images used.



a) b)

Figure 2. Tumor images: a) tumor-positive, b) tumor-negative

The images were taken from a publicly available, ethically approved source [29] that values unrestricted access for educational and research purposes. These datasets are becoming increasingly significant to the medical imaging and AI sectors because they enable experimentation and the development of new algorithms without the moral and regulatory issues that come with handling confidential healthcare data.

In order to depict the variety characteristic of scans from several institutions, the dataset consists of grayscale medical pictures with different resolutions. For evaluating the generalizability of machine learning models, this variation in picture quality, orientation, and contrast provides a reliable testbed. Rigid cross-validation and transfer learning using a pre-trained deep learning model were used to guarantee dependable performance.

Raw medical imaging data usually has to go through a lot of preprocessing before it can be effectively used in deep learning or machine learning workflows. This study employed a variety of preprocessing methods to standardize and enhance the quality of the MRI images.

To make sure that all input images were compatible with the specifications of the InceptionV3 architecture, the first preprocessing step was to standardize their spatial dimensions. A fixed size of 225 × 225 pixels was applied to each image due to the variations in the original resolutions and aspect ratios. This scaling allowed for effective batch processing throughout the training and inference stages in addition to meeting the model's input limitations. Using cropping and zero-padding techniques, the original aspect ratios were maintained in order to reduce geometric distortion and retain anatomical authenticity.

The pixel intensities of every image were normalized based on the deep learning architecture's specifications to guarantee uniformity and model compatibility. Normalization to the interval [-1, 1] was done for InceptionV3. The effectiveness and performance of transfer learning-based techniques are improved by this preprocessing step, which lowers internal covariate shift, stabilizes the learning process, and encourages faster convergence.

5.2 Analysis of Classifier Performance

The classification performance was evaluated for LR, RF, and AdaBoost. A summary of the average outcomes from these iterations is shown below in Table 2.

Table 2. Performance parameters

Classifier	Accuracy (%)	AUC
InceptionV3 + LR	98.2	0.999
InceptionV3 + RF	94.6	0.988
InceptionV3 + AdaBoost	87.4	0.874

LR outperformed all the classifiers evaluated in terms of both classification accuracy (98.2%) and AUC (0.999). The simplicity and effectiveness of this method make it suitable for binary classification on high-quality feature vectors extracted from InceptionV3.

RF performed well, managing feature variability and exhibiting resilience to overfitting with an accuracy of 94.6% and an Area Under Curve (AUC) of 0.988. Its ensemble nature allows for better generalization, especially in noisy data conditions.

With an accuracy of 87.4% and an AUC of 0.874, AdaBoost did not do sufficiently well, even if it was successful at fortifying weak classifiers. Its noise sensitivity and overfitting on cases that were improperly detected account for its performance trade-offs in this scenario.

5.3 Comparison with the Results in Related Work

Our suggested classification system, which combines InceptionV3 for deep feature extraction with Logistic Regression (LR) for classification, shows significant gains when compared to the method described in [19]. Our system produced a considerably greater accuracy of 98.2% than the technique in [19], which used VGG-16 after data augmentation to get its greatest accuracy of 92.31%. This indicates an improvement in categorization accuracy of around 6.4%. In addition, our model achieved an outstanding AUC of 0.999, which indicates nearly perfect discriminatory power, in contrast to the highest AUC documented in [19], which was not specifically stated but may be deduced to be lower given the model complexity and accuracy levels.

In addition to accuracy, our methodology provides a more streamlined and computationally efficient pipeline. We lessen reliance on computationally costly data augmentation approaches and do away with the requirement to retrain deep CNN layers by employing a frozen InceptionV3 model only for feature extraction and combining it with a lightweight LR classifier. This reduces the possibility of overfitting and expedites training and inference, which is pivotal for medical datasets with sparse sample sizes.

5.4 Confusion Matrix

According to Figure 3a, the confusion matrix of the AdaBoost classifier [30] shows a somewhat worse performance when compared to the other models. A considerable number of misclassifications are seen, particularly in the prediction of tumor-positive occurrences. Despite its ability to detect many tumor-negative instances properly, AdaBoost's sensitivity is lower, leading to a higher proportion of false negatives. This is consistent with an overall accuracy of 87.4% and an AUC of 0.874 reported in the results. Perhaps the algorithm overfitted on certain noisy patterns in the data, which reduced its ability to generalize, due to its repeated focus on examples that were erroneously classified. This problem is especially significant in medical diagnostics, where a false negative result from failing to detect a tumor case might have disastrous consequences.

Logistic Regression has exceptional classification capabilities, as seen by its confusion matrix in Figure 3b. With few to no misclassifications, the performance of both the tumor-positive and tumor-negative groups is nearly perfect. This result is in line with its outstanding 0.999 AUC and 98.2% accuracy. FP and FN are decreased by the framework's ability to distinguish between the two categories. Its linear decision boundary appears to be a suitable fit for InceptionV3's deep features, allowing it to achieve optimal classification performance. Because of this, LR is the most accurate, consistent, and clinically reliable of the three models that were studied.

The RF classifier also does acceptably, as seen in Figure 3c, where a confusion matrix shows balanced sensitivity and specificity and high accuracy. The majority of tumor-positive and tumor-negative cases are successfully detected, although a few more misclassifications than with LR. This corresponds to an AUC of 0.988 and an accuracy rate of 94.6%. Its tendency to produce false positives or false negatives, despite its tiny size, suggests that tree selection involves a trade-off between depth and diversity. RF is a very dependable and broadly applicable solution for this classification task.

		Predicted		
		Negative	**Positive**	**∑**
Actual	**Negative**	86.7 %	11.9 %	**1,700**
	Positive	13.3 %	88.1 %	**1,700**
	∑	**1,732**	**1,668**	**3,400**

a)

		Predicted		
		Negative	**Positive**	**∑**
Actual	**Negative**	97.7 %	1.3 %	**1,700**
	Positive	2.3 %	98.7 %	**1,700**
	∑	**1,717**	**1,683**	**3,400**

b)

		Predicted		
		Negative	**Positive**	**∑**
Actual	**Negative**	92.1 %	3.6 %	**1,700**
	Positive	7.9 %	96.4 %	**1,700**
	∑	**1,782**	**1,618**	**3,400**

c)

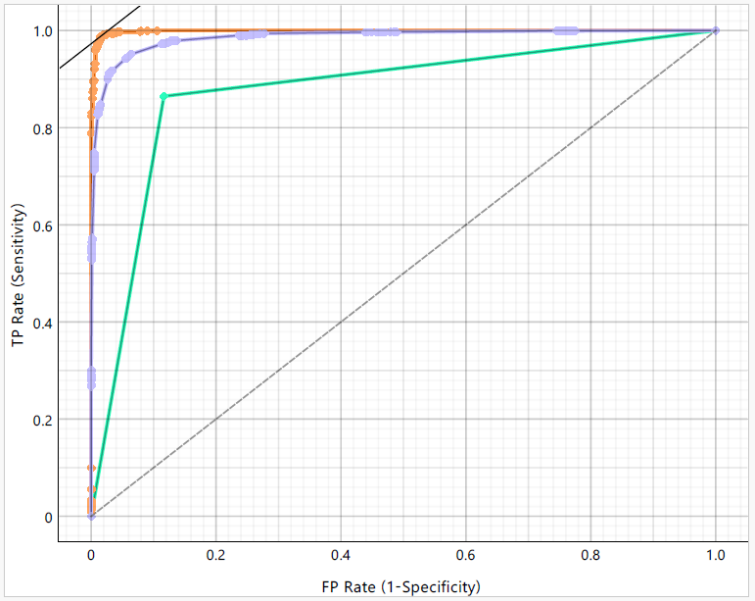
Figure 3. Confusion matrix for a) AdaBoost, b) Logistic Regression, c) Random Forest

5.5 ROC Analysis

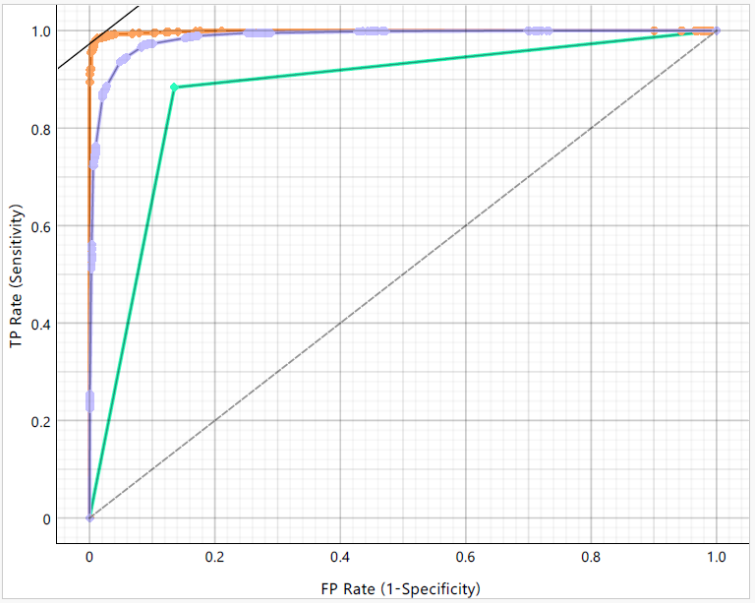
The model's capacity to accurately detect tumor-present MRI images is demonstrated by the ROC curve [31] for tumor-positive situations, which is displayed in Figure 4a. In the figure orange curve corresponds to LR, while the blue represents RF, and the green curve represents the AdaBoost classifier. When the classifiers have a high true positive rate (sensitivity) throughout a range of false positive rates, they are highly successful in differentiating between tumor-positive and tumor-negative cases. The curve's quick initial ascent, which approaches the graph's upper-left corner, demonstrates exceptional performance, with very few real tumor instances misclassified. This result aligns with the model's nearly perfect discriminating ability to identify true positive cases, which is reinforced by the high AUC values for classifiers like LR (AUC = 0.999). This is an important discovery, particularly in the field of medicine, where failing to identify a tumor might have disastrous results.

The ROC analysis for tumor-negative cases shows that the model can accurately identify tumor-absent MRI images (Figure 4b). The curve remains high and to the left when there are few false positives and a high true negative rate (specificity). To avoid unnecessary anxiety and invasive treatments, the classifiers must be able to successfully avoid incorrectly identifying normal brain scans as tumor-positive. The findings demonstrate that the models, in particular RF and LR, successfully balance sensitivity and specificity, and their performance aligns with the high

specificity indicated by evaluation measures. The ROC curve illustrates the classifiers' competence in both detecting and accurately certifying the absence of malignancies.



a)



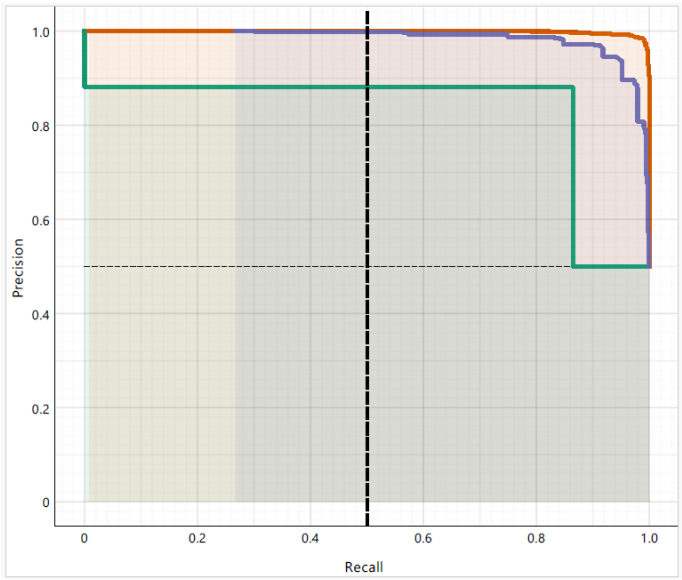
b)

Figure 4. ROC analysis: a) tumor-positive cases, b) tumor-negative cases

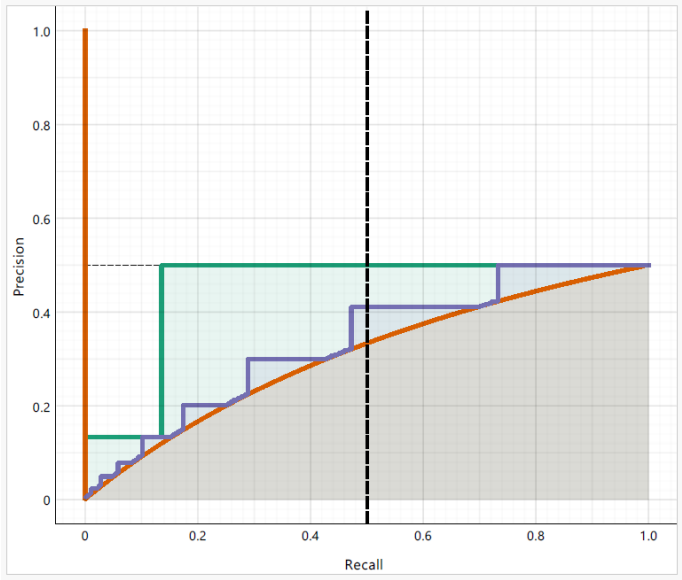
5.6 Performance Curve

The performance curve [32] for positive cases, which displays a consistently high accuracy and recall (Figure 5a), demonstrates the classifiers' remarkable ability to recognize tumor-present MRI images. A high area under the precision-recall curve indicates that the model successfully captures a significant portion of actual tumor cases, indicating that most instances classified as positive are, in fact, true positives. This is particularly significant for clinical applications since it shows that the model can correctly identify images with cancers without oversaturating the system with false positives. Moreover, the stability of the curve across a range of thresholds shows that the classifiers – in particular, LR – continue to function reliably even in the presence of competing decision limits.

As seen in Figure 5b, the performance curve for negative cases also displays acceptable behavior, with high accuracy values held throughout a broad range of recall levels. This demonstrates how effectively the system detects scans devoid of malignancies, ensuring that the vast majority of predictions labeled as negative are, in fact, negative. Reducing false alarms and unnecessary treatments in healthy individuals requires maintaining this level of proficiency. The constantly well-balanced precision-recall trade-off further supports the classifiers' capacity to decrease false positives and false negatives within the negative class. By verifying that the models are well-generalized and do not favor any one class, these results ensure reliability in both diagnostic outcomes.



a)



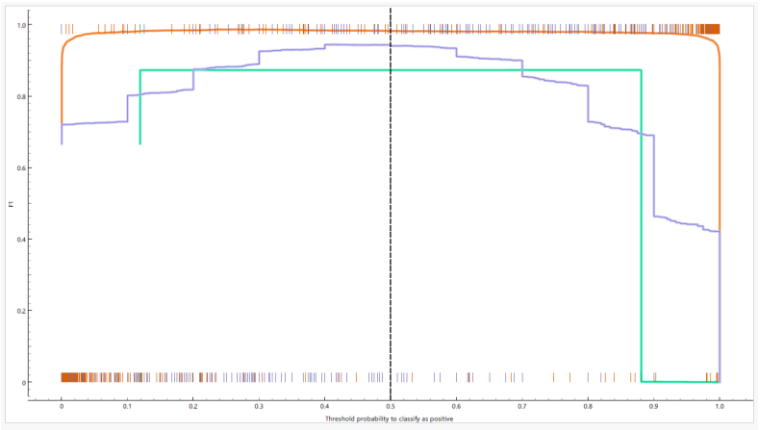
b)

Figure 5. Performance curves a) Tumor-positive cases b) Tumor-negative cases

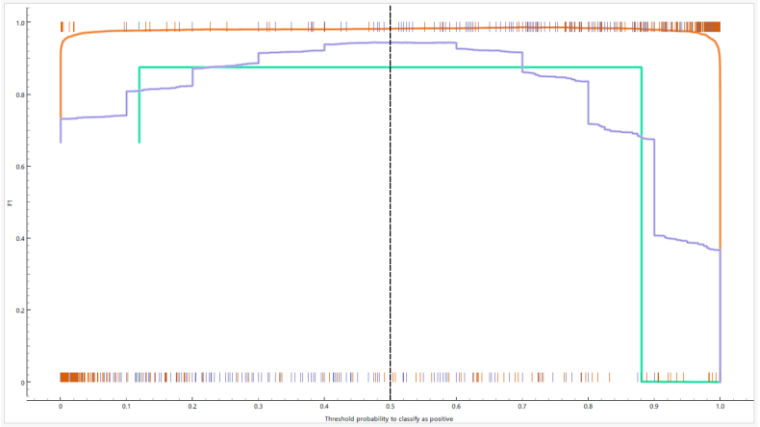
5.7 Calibration Plots

Over a variety of threshold values, the calibration plot [33] for positive cases, which is displayed in Figure 6a, demonstrates a high agreement between anticipated probability and actual findings. As the threshold probability increases, the F1-score remains high and stable over a wide range, suggesting that the classifiers – specifically, LR and RF – are well-calibrated for detecting

tumor-positive MRI data. This implies that the model is generally correct when it predicts a high probability of a positive case, which is an essential feature for clinical applications where tumor identification has to be done early and confidently. The plot's smooth trajectory and elevated F1-score curve further illustrate the model's reliability in maintaining a balance between recall and accuracy at different decision thresholds.



a)



b)

Figure 6. Calibration plots: a) Tumor-positive cases, b) Tumor-negative cases

The calibration plot for negative scenarios also demonstrates effective model behavior; as Figure 6b illustrates, the F1-score curve performs best at moderate to high threshold probabilities. This suggests that, in addition to being successful in identifying negative cases, the classifiers are cautious in minimizing false positives by not overestimating the probability of tumor presence in healthy scans. The curve's consistent slope, which indicates constant prediction quality without abrupt drops, supports the model's careful and well-balanced decision-making.

6 Conclusion

In this work, DL and ML approaches are used to develop a hybrid framework for the binary classification of BT MRI images. By means of an InceptionV3 model and transfer learning, deep features were retrieved from pre-processed MRI images. These features were then identified using LR, RF, and AdaBoost. The interpretability and simplicity of machine learning models are maintained while utilizing CNNs' representational strength. The framework's generalizability and dependability were validated using a rigorous validation process that included 50 stratified shuffling repetitions and 10-fold cross-validation. The best-performing classifier was LR, which outperformed other models by obtaining an accuracy of 98.2% and an AUC of 0.999. RF

performed moderately well, although AdaBoost showed average results. The superiority of the system is further demonstrated by an in-depth study using calibration plots, ROC curves, confusion matrices, and performance metrics, which demonstrate the framework's ability to successfully reduce false positives and negatives. One of the study's main contributions is its high diagnosis accuracy, which makes it both computationally feasible and appropriate for use in clinic settings. In addition, the hybrid strategy guarantees radiologists quicker, scalable, and interpretable decision-making. Multimodal imaging and multi-class tumor classification can be added to this framework in future research to increase diagnosis accuracy and broaden its clinical usefulness.

Acknowledgment: The author would like to thank Nims University in Jaipur, Rajasthan, India, where the work was accomplished.

References

[1] A. S. Jadhav, R. Pawar, S. M. Hattaraki, S. Bagali, R. S. Patil, and B. Patil, "Brain Tumor Detection Using MRI Image Processing Techniques," *2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA)*, pp. 1–5, 2024. Available: https://doi.org/10.1109/innova63080.2024.10846990

[2] A. Singh and P. Chauhan, "A review on brain tumor MRI image segmentation," *International Journal of Innovative Research in Computer Science & Technology*, vol. 12, no. 1, pp. 303–306, 2024. Available: https://doi.org/10.55524/csistw.2024.12.1.53

[3] K. R. Prabha, B. Nataraj, A. Karthikeyan, N. Mukilan, and M. S. Parkavin, "MRI Imaging for the Detection and Diagnosis of Brain Tumors," *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1541–1546, 2023. Available: https://doi.org/10.1109/icosec58147.2023.10276177

[4] G. Mahesh and K. M. Yogesh, "Brain Tumor Detection and Classification Using MRI Images," *International Journal for Research in Applied Science & Engineering Technology*, vol. 12, no. 10, pp. 856–863, 2024. Available: https://doi.org/10.22214/ijraset.2024.64719

[5] S. Loeber, "Brain MRI Protocol and Systematic Approach to Interpretation of Brain Tumors on MRI," *Veterinary Clinics of North America: Small Animal Practice*, vol. 55, no.1, pp. 11–21, 2025. Available: https://doi.org/10.1016/j.cvsm.2024.07.003

[6] N. Saito, N. Hirai, Y. Koyahara, S. Sato, Y. Hiramoto, S. Fujita, H. Nakayama, M. Hayashi, and S. Iwabuchi, "Comparative Study of Postmortem MRI and Pathological Findings in Malignant Brain Tumors," *Cureus*, vol. 16, no. 3, 2024. Available: https://doi.org/10.7759/cureus.56241

[7] D. Saveetha, B. P. Prathaban, A. Rashmi, and K. E. Purushothaman, "Brain Tumor Classification Using CNNs," *2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, pp. 1–6, 2024. Available: https://doi.org/10.1109/icpects62210.2024.10780208

[8] P. Mohan and G. Ramkumar, "An Inception V3 Based Glioma Brain Tumor Detection in MRI Images," *2024 5th International Conference on Circuits, Control, Communication and Computing (I4C)*, pp. 560–564, 2024. Available: https://doi.org/10.1109/i4c62240.2024.10748505

[9] K. Rama Krishna, M. Arbaaz, S. N. C. Dhanekula, and Y. M. Vallabhaneni, "Modified VGG16 for accurate brain tumor detection in MRI imagery," *Informatics, Control, Measurement in Economy and Environment Protection*, vol. 14, no. 3, pp. 71–75, 2024. Available: https://doi.org/10.35784/iapgos.6035

[10] S. Balaji, T. Krithika, and S. Joshi, "Brain Tumor Detection Using Deep Learning: ResNet+LSTM," *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, pp. 1064–1070, 2024. Available: https://doi.org/10.1109/icicnis64247.2024.10823218

[11] K. Kumar, K. Jyoti, and K. Kumar, "Machine learning for brain tumor classification: evaluating feature extraction and algorithm efficiency," *Discover Artificial Intelligence*, vol. 4, article 112, 2024. Available: https://doi.org/10.1007/s44163-024-00214-4

[12] S. Gujala, V. Rajesh, "An MRI brain tumour detection using logistic regression-based machine learning model," *International Journal of System Assurance Engineering and Management,* vol. 15, pp. 124–134, 2024. Available: https://doi.org/10.1007/s13198-022-01680-8

[13] B. K. Saraswat and N. Tiwari, "Brain Tumor Detection," *International Journal for Research in Applied Science & Engineering Technology*, vol. 11, no. 5, pp. 5634–5640, 2023. Available: https://doi.org/10.22214/ijraset.2023.52981

[14] M. Narang and D. Narang, “Brain Tumour Detection Using Deep Learning and Image Processing,” *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 11, pp. 1–7, 2024. Available: https://doi.org/10.55041/ijsrem38964

[15] A. Sawle, S. Bhosale, S. Abhang, P. Ghagare, and D. Argade, “Brain Tumor Detection using Deep Learning,” *International Journal for Research in Applied Science & Engineering Technology*, vol. 12, no. 3, pp. 2402–2408. Available: https://doi.org/10.22214/ijraset.2024.59360

[16] S. Prudhviraj et al., “Brain tumor detection,” *International Journal of Scientific Research in Engineering and Management*, 2023. Available: https://doi.org/10.55041/ijsrem27721

[17] H. G. Mohan et al., “Brain tumor detection using machine learning,” *International Journal of Scientific Research in Engineering and Management*, 2024. Available: https://doi.org/10.55041/ijsrem34309

[18] P. Reddi, G. Srinivas, P. V. G. D. P. Reddy, and H. S. Nallagonda, “Brain Tumor Detection using Improved Binomial Thresholding Segmentation and Sparse Bayesian Extreme Learning Machine Classification,” *International Journal of Electrical & Electronics Research*, vol. 12, no. 3, pp. 1094–1100, 2024. Available: https://doi.org/10.37391/ijeer.120345

[19] H. Güven and A. Saygılı, “Computer-aided detection of brain tumors using image processing techniques,” *Kahramanmaras Sutcu Imam University Journal of Engineering Sciences*, vol. 27, no. 3, pp. 999–1018, 2024. Available: https://doi.org/10.17780/ksujes.1447899

[20] M. Sharma, M. Gahlot, and S. Chaudhary, “Brain Tumor Detection Using Deep Learning,” *International Journal of Scientific Research in Engineering and Management*, 2023. Available: https://doi.org/10.55041/ijsrem27567

[21] P. P. Debata, P. N. Pratikshya, S. Padhy, S. Agarwal, and S. Bansal, “Brain Tumor Detection using an Optimized Deep Learning Approach,” *2024 3rd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*, pp. 1–5, 2024. Available: https://doi.org/10.1109/odicon62106.2024.10797640

[22] V. V. K. Maddinala, P. V. Charan, B. C. S. Mounika, and Md. M. Younus, “Brain Tumor Detection Through Advanced Computational Methods,” *International Journal of Innovative Science and Research Technology*, vol. 9, no. 4, 2024. Available: https://doi.org/10.38124/ijisrt/ijisrt24apr569

[23] P. J. Reddy and R. Thalapathi Rajasekaran, “A Novel Approach to Improve Diabetes Prediction Accuracy using the AdaBoost Algorithm Compared with a Naive Bayes Classifier,” *2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES)*, pp. 1–5, 2024. Available: https://doi.org/10.1109/IC3TES62412.2024.10877676

[24] S. Dutta and S. K. Bandyopadhyay, “Cross-Validated AdaBoost Classifier Used for Brain Tumor Detection,” *Preprints*, 2020. Available: https://doi.org/10.20944/preprints202006.0351.v1

[25] S. Salian, S. Cherishma, and O. S. Powar, “Enhanced Brain Tumor Detection using Support Vector Classifier and Logistic Regression with Principal Component Analysis,” *2024 Control Instrumentation System Conference (CISCON)*, pp. 1–5, 2024. Available: https://doi.org/10.1109/ciscon62171.2024.10696269

[26] P. Jain, P. Mishra, K. Anand, P. Mishra, and A. Jamwal, “Predictive Precision: A Conference on Brain Tumor Detection using Logistic Regression,” *2024 Second International Conference on Advances in Information Technology (ICAIT)*, pp. 1–5, 2024. Available: https://doi.org/10.1109/icait61638.2024.10690444

[27] A. K. Soni, A. Boobalan, E. Rajesh, V. Janakiraman, and V. Kesarwani, “Hybrid Classification Framework Based on CNN and Random Forest Method to Predict Brain Tumor,” *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, pp. 1249–1255, 2024. Available: https://doi.org/10.1109/ic2pct60090.2024.10486772

[28] N. Vani and D. Vinod, “A Comparative Analysis on Random Forest Algorithm Over K-Means for Identifying the Brain Tumor Anomalies Using Novel CT Scan with MRI Scan,” *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, pp. 1–6, 2022. Available: https://doi.org/10.1109/ICBATS54253.2022.9759036

[29] Brain Tumor Dataset. Available: https://www.kaggle.com/datasets/praneet0327/brain-tumor-dataset. Accessed on July 16, 2025.

[30] K. S. Subedi et al., “Usefulness of double dose contrast-enhanced magnetic resonance imaging for clear delineation of gross tumor volume in stereotactic radiotherapy treatment planning of metastatic brain tumors: a dose comparison study,” *Journal of Radiation Research*, vol. 54, no. 1, pp. 135–139, 2013. Available: https://doi.org/10.1093/jrr/rrs053

[31] S. Bauer, C. May, D. Dionysiou, G. Stamatakos, P. Buchler, and M. Reyes, “Multiscale Modeling for Image Analysis of Brain Tumor Studies,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 25–29, 2012. Available: https://doi.org/10.1109/TBME.2011.2163406

[32] E. I. Zacharaki, C. S. Hogea, G. Biros, and C. Davatzikos, “A Comparative Study of Biomechanical Simulators in Deformable Registration of Brain Tumor Images,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1233–1236, 2008. Available: https://doi.org/10.1109/TBME.2007.905484

[33] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang, and Z. Tang, “Multimodal Brain Tumor Segmentation Boosted by Monomodal Normal Brain Images,” *IEEE Transactions on Image Processing*, vol. 33, pp. 1199–1210, 2024. Available: https://doi.org/10.1109/TIP.2024.3359815

Appendix

List of Abbreviations:

Abbreviation	Full Form	Abbreviation	Full Form
AI	Artificial Intelligence	AUC	Area Under the Curve
BT	Brain Tumor	CA	Classification Accuracy
CM	Confusion Matrix	CNN	Convolutional Neural Network
CNNs	Convolutional Neural Networks (plural)	DE	Differential Evolution
DL	Deep Learning	FN	False Negative
FP	False Positive	HS	Harmony Search
LR	Logistic Regression	MCC	Matthews Correlation Coefficient
ML	Machine Learning	MRI	Magnetic Resonance Imaging
OBFSA	Optimization-Based Feature Selection Algorithm	ResNet	Residual Neural Network
RF	Random Forest	RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic	SA	Simulated Annealing
SF	Sigmoid Function	TN	True Negative
TP	True Positive	VGG	Visual Geometry Group Network