

# Classification of Intracranial Brain Tumor from MRI Images using Ensemble Machine Learning Models

Syed Ayaz Imam

School of Computer Science & Engineering  
Vellore Institute of Technology, Vellore  
Tamil Nadu, India – 632014

Archit Aggarwal

School of Computer Science & Engineering  
Vellore Institute of Technology, Vellore  
Tamil Nadu, India - 632014

Vandit Jain

School of Computer Science & Engineering  
Vellore Institute of Technology, Vellore  
Tamil Nadu, India - 632014

**Abstract**— Automated defect detection has become an emerging field in various medical diagnostic applications. Computerized detection of a brain tumor in Magnetic resonating imaging (MRI) is crucial as it provides evidence and information about the tissues which require treatment and planning. The traditional method used to detect any defect in magnetic resonance brain images involves using the help of a doctor's inspection, which consumes a lot of time and data. Henceforth, automated classification programs are necessary to improve the death rate. So, different machine learning models are under development, which aims to provide greater accuracy and performance in the classification of brain tumors. Inspecting MRI of brain tumors is already a complicated task due to the wide variety and complexity of tumors. Ensemble machine learning models could be used to reliably detect and classify cancer cells in the brain through magnetic resonating imaging.

**Keywords**— MRI; Tumor; Machine Learning; Deep Learning; Random Forest; SVM; Gradient Boosting; Accuracy; Cross-Validation.

## I. INTRODUCTION

The brain is the most complex part and plays the most significant role in the human body. It controls various functions as motor skills, feeling emotions, thought process and any other processes in our body. Brain tumor consists of cell growth in our brain in an abnormal manner which could or could not be cancerous. Brain tumors could be of different forms. Brain tumors can be of various forms. They could start in the brain as a primary tumor and then spread to the other parts of the body, or they could start as cancer in different parts of the body and spread to the brain as a secondary tumor. The rate at which it will grow varies from person to person. The location of the tumor and the growth rate determines how it will affect the functioning of the nervous system.

A brain tumor was discovered in 1884 by Mr. Rickman Godlee when he had the performed first-ever recognized resection of a primary brain tumor. In modern times a neurological exam is conducted to diagnose a brain tumor. The exam might consist of a vision check, hearing, coordination, muscle reflex, and strength. Any abnormality in parts of the brain can act as evidence for brain tumor infection. Other tests

are also conducted, such as magnetic resonance imaging (MRI) scans, which help doctors evaluate the treatment of the tumor; computerized tomography and positron emission tomography may be recommended in some difficult situations.

Automated defect detection in every field is emerging with several purposeful applications. It is also being used in medical applications using machine learning algorithms for a significant number of medical diagnostic applications. These algorithms play a crucial role in detecting a brain tumor in magnetic resonance imaging (MRI) as it uses these images and provides specific information about abnormal tissues in the brain, which is essential for determining the treatment plan. In recent studies, it was observed that automated detection and diagnosis of the diseases, which may be based upon medical reports, might provide an adequate alternative compared to the current methods, as it would provide greater accuracy and save some time for the radiologist. Additionally, if machine learning algorithms can provide accurate results of tumor depiction, these results would help the clinical management of brain tumors by liberating the physicians from the weight of the manual depiction of tumors.

The document starts with a description of the dataset and then moves on to the machine learning techniques used. Following that, we went over the implementation and approach, including the pre-processing procedures. Finally, we compared the results of the models to one other and came to a conclusion on the work's findings.

## II. DATASET

The Brain-MRI images data set is made up of multiple different image records consisting of brain MRI images taken from different angles. It consists of a total of 253 images and is widely used for tumor detection because of its authenticity of images and no duplicates in images of the scan.

The dataset images are classified into two i.e. *Yes (Tumour Present)* and *No (Not present)*. However, the images need to be resized into the same dimensions for processing.

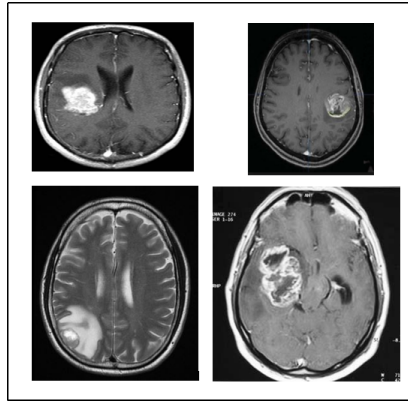


Figure 1. Tumor *present* Sample Images

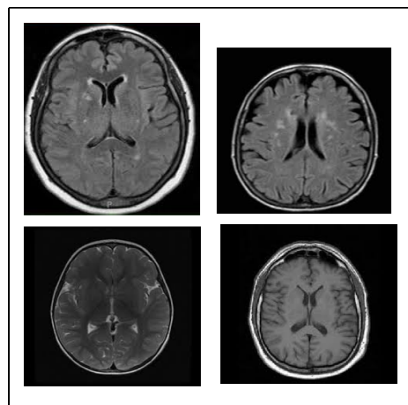


Figure 2. Tumor *not present* Sample Images

### III. ALGORITHMS USED

Support Vector Machine (SVM) is a machine learning model that uses a supervised learning method for classification. The SVM's purpose is to find the optimal decision boundary for classifying an n-dimensional space to classify data points into different, correct categories. The decision boundary that exhibits the most accuracy is known as the hyperplane.

The Decision Trees technique is also used for classification and regression problems that also uses a supervised learning method. The internal nodes of the tree are used to represent attributes of the dataset, while decision rules govern the branches. The leaf node at the end gives the outcome. There are two types of nodes in decision trees, namely- decision node and leaf node.

Random Forests are built on ensemble learning methods wherein multiple classifiers improve accuracy and get better results for complex problems. As the name suggests, Random Forest uses multiple decision trees on multiple subsets of data to give results. The increase in number allows the algorithm to collect the predicted outputs of multiple trees and average them out instead of relying solely on a singular decision tree. The method also diminishes the overfitting problem.

Gradient Boosting, like Random Forest, uses multiple decision trees for classification. However, unlike the Random Forests, Gradient Boosting builds trees for classification one at a time by optimizing the performance of the previous weaker trees. If parameters are finely tuned, gradient boosting can result in very high accuracy and make accurate predictions.

### IV. METHODOLOGY & IMPLEMENTATION

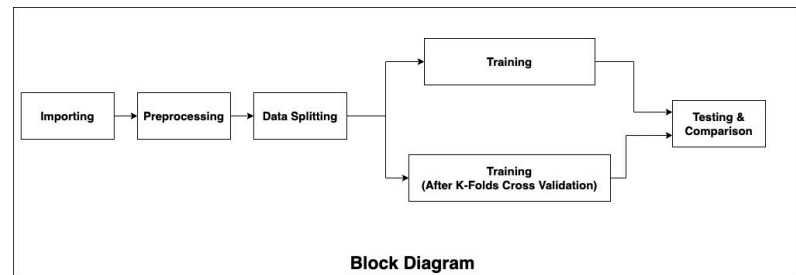


Figure 3. Methodology Block Diagram

#### A. Importing & Preprocessing

The tumor in the brain could be malignant or a non-cancerous mass of abnormal cell development. The cause of brain tumors is attributed to abnormal cell growth in the brain and significantly impacts a person's life. The patient's medical recovery can be aided by early and accurate identification of such conditions. We started by preparing the images. The actions taken to format images before they are utilized in model training and inference are known as image pre-processing. This encompasses resizing, orienting, and color corrections, among other techniques.

In the pre-processing technique, we converted the images into grayscale since grayscale representations are frequently employed to extract descriptors. Rather than operating directly on color images, grayscale simplifies the technique and reduces computational requirements.

#### B. Data Splitting

The risk of overfitting your model to the training set exists during the training process. For example, the model may learn an extremely particular function that works well on your training data but does not generalize to new images. This suggests that the model is not learning correctly and is simply memorizing the training data. This indicates that the model will struggle with fresh images that it has never seen before. Overfitting is combated by the train, validation, and testing splits. We have split our data into a training set and test set in a 70:30 ratio.

#### C. Training

We trained the SVM, Random Forest, Decision Trees, and Gradient Boosting models after pre-processing and partitioning the data into a 70:30 ratio. The output's results are discussed in the following section. We also used K-fold cross-validation (K=11) to repeat the training process and compare it to the one done without it.

## V. RESULTS & CONCLUSION

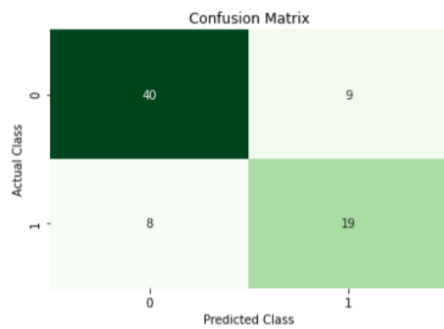


Figure 4. (a) SVM Confusion Matrix

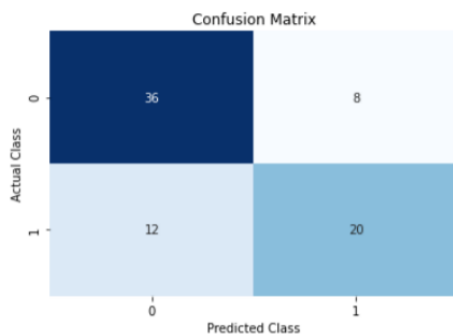


Figure 4. (b) Decision Tree Confusion Matrix



Figure 4. (c) Random Forest Confusion Matrix

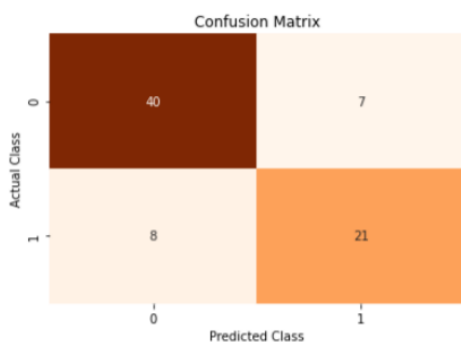


Figure 4. (d) Gradient Boosting Confusion Matrix

The above graphs depict the confusion matrices of all the algorithms used in our study. Confusion Matrices are of great significance as they summarize the performance of the classifiers. It gives a clearer sense of how the classifier is performing when compared to other metrics like accuracy. Accuracy can give ambiguous results if there are unequal observations for a class or in the presence of multiple classes in the dataset. The below table presents the results of all the classifiers of all the categories in the matrix.

Table 1. Confusion Matrix Results Summary

Classifiers	(False Positive) FP	(True Positive) TP	(False Negative) FN	(True Negative) TN
SVM	8	40	9	19
Random Forest	7	41	6	22
Decision Trees	12	36	8	20
Gradient Boosting	8	40	7	21

Random State plays a crucial role in determining the accuracy of the classifier. Figure x provides a boxplot depicting the variation in accuracy for changing random states of each classifier. Performance metrics depend greatly on random states. Random states are used to split the dataset into training and test sets. Some algorithms use the random state to select a subset of features, initialize weights, etc.

Algorithms like SVM use it for initial probability estimations, while random forests use the random state to select features. The classifier thus provides huge variations due to such high significance of the random state.

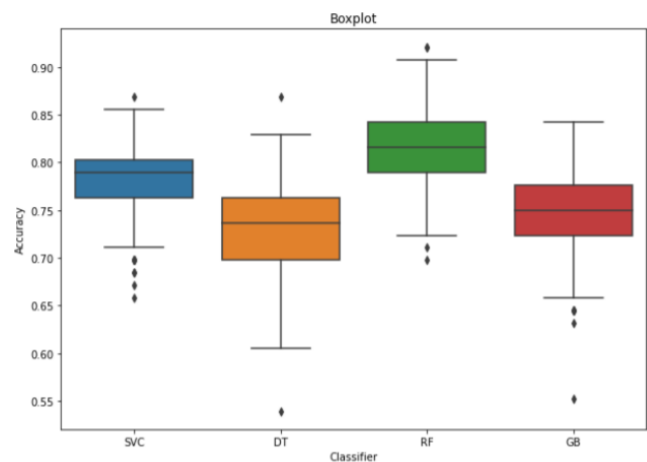


Figure 5. Boxplot for the Various Random State of Classifiers

Cross-validation is a process to judge machine learning models on a dataset. K- Fold Cross-Validation makes use of a single parameter named k, which is used for cross-validation. The value of k gives the number of groups to split the dataset. The value of k influences the accuracy by a great measure. The figures below depict the changes in accuracy of the classifiers after K-cross validation for **K=11**.

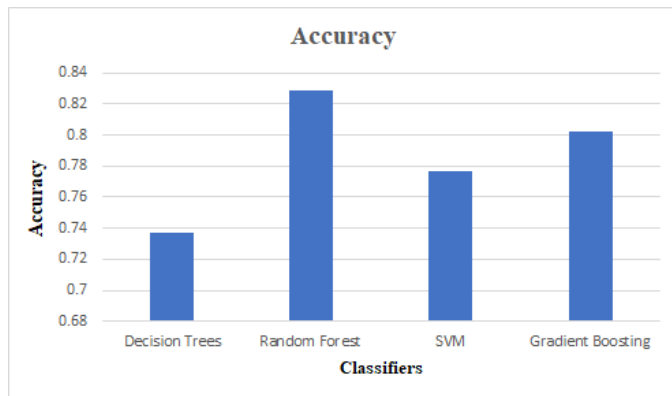


Figure 6. (a) Accuracy of Classifiers

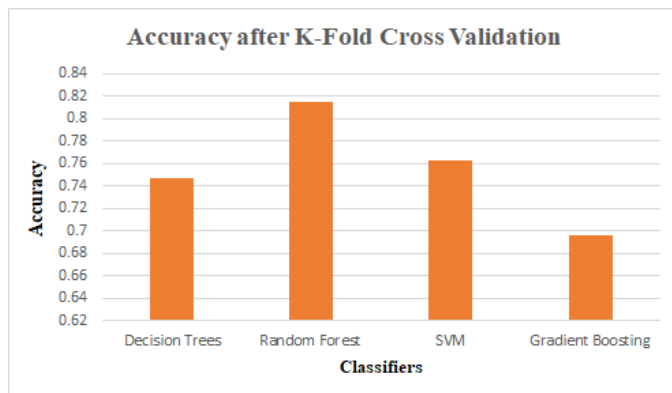


Figure 6. (b) Accuracy of Classifiers after K-Fold

**Table 2.** Accuracy of Classifiers before and after cross validation

Performance Metrics	SVM	Decision Tree	Random Forest	Gradient Boosting
Accuracy before K-Fold	77.63	73.68	82.89	80.26
Accuracy after K-Fold	76.28	74.70	81.42	69.56

The above table summarizes the findings of our study. We can expect less overfitting and a more accurate proxy for accuracy when we utilize cross-validation. The accuracy percentages should not be used as a benchmark for model accuracy because the numbers could be skewed by overfitting. K-fold cross validation does not reduce accuracy; rather, it provides a better estimate of accuracy, with less overfitting. As a result, we may

infer that the Random Forest ensemble model outperformed the others in our tests. Furthermore, we used a sample of 253 images to train our model, thus the accuracy can be increased by collecting more image samples.

## REFERENCES

- [1]. Wu, M. N., Lin, C. C., & Chang, C. C. (2007, November). Brain tumor detection using color-based k-means clustering segmentation. In *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)* (Vol. 2, pp. 245-250). IEEE.
- [2]. Mustaqeem, A., Javed, A., & Fatima, T. (2012). An efficient brain tumor detection algorithm using watershed & thresholding based segmentation. *International Journal of Image, Graphics and Signal Processing*, 4(10), 34.
- [3]. Akram, M. U., & Usman, A. (2011, July). Computer aided system for brain tumor detection and segmentation. In *International conference on Computer networks and information technology* (pp. 299-302). IEEE.
- [4]. Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2020). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognition Letters*, 139, 118-127.
- [5]. Suján, M., Alam, N., Noman, S. A., & Islam, M. J. (2016). A segmentation based automated system for brain tumor detection. *International Journal of Computer Applications*, 153(10), 41-49.
- [6]. Devkota, B., Alsadoon, A., Prasad, P. W. C., Singh, A. K., & Elchouemi, A. (2018). Image segmentation for early stage brain tumor detection using mathematical morphological reconstruction. *Procedia Computer Science*, 125, 115-123.
- [7]. Praveen, G. B., & Agrawal, A. (2015, November). Hybrid approach for brain tumor detection and classification in magnetic resonance images. In *2015 Communication, Control and Intelligent Systems (CCIS)* (pp. 162-166). IEEE.
- [8]. Borole, V. Y., Nimbhore, S. S., & Kawthekar, D. S. S. (2015). Image processing techniques for brain tumor detection: A review. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 4(5), 2.
- [9]. Zhang, S., & Xu, G. (2016). A novel approach for brain tumor detection using MRI Images. *Journal of Biomedical Science and Engineering*, 9(10), 44-52.
- [10]. Shanthakumar, P., & Ganeshkumar, P. (2015). Performance analysis of classifier for brain tumor detection and diagnosis. *Computers & Electrical Engineering*, 45, 302-311.
- [11]. Sharif, M., Amin, J., Raza, M., Anjum, M. A., Afzal, H., & Shad, S. A. (2020). Brain tumor detection based on extreme learning. *Neural Computing and Applications*, 1-13.
- [12]. Singh, A., Bajpai, S., Karanam, S., Choubey, A., & Raviteja, T. (2012). Malignant brain tumor detection. *International Journal of Computer Theory and Engineering*, 4(6), 1002.
- [13]. Gondal, A. H., & Khan, M. N. A. (2013). A review of fully automated techniques for brain tumor detection from MR images. *International Journal of Modern Education and Computer Science*, 5(2), 55.
- [14]. Anitha, R., & Siva Sundhara Raja, D. (2018). Development of computer-aided approach for brain tumor detection using random forest classifier. *International Journal of Imaging Systems and Technology*, 28(1), 48-53.
- [15]. Iftekharuddin, K. M., Zheng, J., Islam, M. A., Ogg, R. J., & Lanningham, F. (2006, October). Brain tumor detection in MRI: technique and statistical validation. In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers* (pp. 1983-1987). IEEE.
- [16]. Zeljkovic, V., Druzgalski, C., Zhang, Y., Zhu, Z., Xu, Z., Zhang, D., & Mayorga, P. (2014, April). Automatic brain tumor detection and segmentation in MR images. In *2014 Pan American Health Care Exchanges (PAHCE)* (pp. 1-1). IEEE.
- [17]. Banerjee, S., Mitra, S., Masulli, F., & Rovetta, S. (2019). Deep radiomics for brain tumor detection and classification from multi-sequence MRI. *arXiv preprint arXiv:1903.09240*.