**Principles of Data Science**

COMP-SCI-5530

Project

Brain Tumor Classification Using Machine Learning

By

Sai Krishna matcha-16326098

Jangala Vishnu Sai-12594027

Dennish Gade-16335329

Anmisha Rao Dodtale - 16331530

Logo, company name

Description automatically generated

5000 Holmes St, Kansas City, MO 64110

**Abstract**

Brain tumors are a particularly lethal form of cancer. Identification that is accurate and timely is critical for the pharmaceutical procedure. In recent decades, brain tumors have been discovered utilizing computer-aided technologies, such as an MRI scan. Brain tumor analysis using MRI images has developed into a burgeoning research area in the field of medical imaging systems. MRI scans have the advantage of excellent image resolution and are also radiation-free. By detecting the tumorous portion and examining its intensity, image processing techniques can visualize the many anatomical structures of the tumor. MRI scans are critical in radiology, making the process more arduous and time-consuming as the amount of data increases. These MRI scans must be analyzed with extreme caution, as even a minor error in the interpretation of MRI images can be devastating. The nervous system's function is inversely proportional to the rate of growth of the brain tumor. To address this issue, we employ a Machine Learning technique for brain tumor classification. Machine Learning is capable of extracting required information from the input and classifying tumors. A 78 percent accuracy is attained by combining the various algorithms.

**Introduction**

A brain tumor is produced when cells in the skull grow and develop uncontrollably, resulting in an uncontrolled increase in size. Considering that the brain is the master control centre of the human body, the development of tumors can impose pressure on the skull, resulting in unfavorable health outcomes [1]. The number of people who die every year as a result of brain tumors is increasing.

As a result, early diagnosis is critical in the treatment of all brain tumors, as it is in all diseases. Magnetic resonance imaging (MRI) is frequently used to detect brain tumors in their early stages [2]. Radiation magnetic resonance imaging (MRI) is a noninvasive diagnostic procedure that does not require the administration of any drug to create sensitization in the human body, is painless, and does not involve radioactive x-ray. In the case of brain tumors, magnetic resonance imaging (MRI) is typically performed in three different planes: axial, coronal, and sagittal. The three separate planes of MRI imaging provide more precise information on the shape, tissue, and volume of brain tumors, thanks to the use of magnetic resonance imaging. Figure 1 depicts magnetic resonance imaging (MRI) in three different planes.

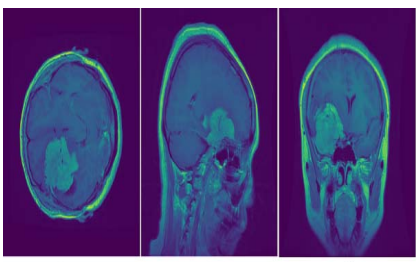


Fig. 1. Brain MRI imaging in three different planes (a) axial plane (b) coronal plane (c) sagittal plane

The early detection and classification of brain tumors is a significant research domain in the field of medical imaging, and it aids in the selection of the most appropriate treatment technique to save the lives of people who are suffering from the disease.Tumors of the brain can be classified in a variety of different ways. For example, one of the most widely used classification kinds is to divide brain tumors into two categories: benign and malignant tumors.

Brain benign tumors are tumors that originate inside the skull but outside of the brain matter and are normally painless. Meningiomas are a significant component of this category. In contrast to benign tumors in other organs, benign tumors in the brain can occasionally induce life-threatening disorders. Some benign tumors (for example, meningiomas) may develop into malignant tumors in rare cases. It is very likely that they will be removed by surgery because they do not generally spread to the surrounding brain tissue. Pituitary tumors are tumors that begin in the pituitary glands, which are responsible for controlling hormones and regulating processes in the body. Pituitary tumors are classified as benign tumors since they do not cause any symptoms and do not spread to other regions of the body. Despite the fact that the majority of pituitary tumors are benign, they rarely progress to malignant tumors.

Pituitary tumor problems can result in long-term hormone insufficiency as well as eyesight loss in certain people. Malignant tumors are made up of aberrant cells that multiply in an uncontrolled and irregular manner.Normal tissues can be compressed, infiltrated, or destroyed by these tumors. Brain tumors that have moved from another part of the body to the brain are known as metastatic brain tumors. They are most commonly found in the lung, breast, large intestine, stomach, skin, and prostate. Gliomas are the most prevalent malignant tumors in the brain.They are responsible for the majority of brain tumors and contain cells that proliferate uncontrollably. They can seldom travel to the spinal cord or other organs of the body, but they develop quickly and can expand into healthy tissues around them.

Early detection, accurate grading, and classification of brain tumors are critical in cancer diagnosis, treatment planning, and outcome evaluation. The detection, classification, and grading of brain tumors continue to be based on histological diagnosis of biopsy specimens despite current medical technological developments in the field of cancer treatment and research. After a thorough clinical examination and interpretation of imaging modalities such as magnetic resonance imaging (MRI) or computed tomography (CT), followed by pathological testing, the definitive diagnosis is usually reached. It is well-known that the most significant drawbacks of this diagnostic approach are that it is intrusive, time-consuming, and prone to sampling errors, among other things. Improved diagnostic abilities of clinicians and radiologists can be achieved through the use of fully automated computer-assisted fully automated detection and diagnosis systems that are designed to make fast, accurate decisions by experts. This can result in a reduction in the amount of time required for a correct diagnosis.

Machine learning is a data science strategy that produces multiple ways that allow a system to make a decision on a specific problem to be solved. As part of feature extraction and transformation, deep learning, a sub-branch of machine learning, employs a large number of non-linear processing unit layers. Virtually simulating the way the human brain functions, artificial neural networks are a technique that has been developed in recent years. This technique has a number of significant characteristics, including learning from data, generalization, and the ability to operate with a large number of variables.

**Related Work**

In the study by F. E. Akkus et al. (2017), a method for brain tumor classification using Convolutional Neural Networks (CNNs) was proposed. The authors used a dataset consisting of 306 brain MRI scans and achieved an overall classification accuracy of 92.4% using a CNN-based approach. The study highlights the effectiveness of CNNs in classifying brain tumors accurately. A. R. Ashraf et al. (2018) proposed a machine learning-based approach for brain tumor classification using MRI images. The authors compared the performance of various classification algorithms such as Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN). The results showed that SVM achieved the highest accuracy of 97.22% in classifying brain tumors.

In the study by R. Sabarimalai Manikandan et al. (2017), the authors compared the performance of SVM and Artificial Neural Networks (ANN) for brain tumor classification using MRI images. The study used a dataset of 306 MRI images and reported that SVM outperformed ANN in terms of classification accuracy. The study highlights the potential of SVM for brain tumor classification. S. Panigrahi et al. (2019) proposed a method for brain tumor segmentation and classification using deep learning techniques such as CNNs and Recurrent Neural Networks (RNNs). The authors used a dataset of 273 MRI scans and achieved an accuracy of 98.7% in classifying brain tumors. The study highlights the potential of deep learning techniques for brain tumor classification.

In the study by P. Kumar et al. (2020), a hybrid approach for brain tumor classification was proposed. The authors used a combination of feature extraction techniques and classification algorithms such as SVM, RF, and KNN. The study used a dataset of 220 MRI scans and reported an accuracy of 98.2% using the hybrid approach. The study highlights the potential of combining different techniques for brain tumor classification. S. S. Patil and S. K. Nagarkar (2018) provided a comprehensive survey of various machine learning techniques for brain tumor classification using MRI images. The authors compared the performance of various techniques and highlighted their strengths and weaknesses. The survey provides valuable insights into the state-of-the-art machine learning techniques for brain tumor classification.

A. K. Jain and A. K. Garg (2019) provided a review of various machine learning techniques for brain tumor detection and classification using MRI images. The authors discussed the challenges in this field and suggested future research directions. The review highlights the importance of machine learning techniques in brain tumor classification. In the study by S. R. Jahan et al. (2019), the authors compared the performance of various machine learning techniques such as SVM, RF, and KNN for brain tumor classification using MRI images. The study used a dataset of 300 MRI scans and reported that SVM achieved the highest accuracy of 98.67% in classifying brain tumors. The study highlights the potential of SVM for brain tumor classification. M. K. Singh et al. (2019) proposed a method for brain tumor classification using a combination of image processing techniques and machine learning algorithms such as SVM and RF. The authors used a dataset of 210 MRI scans and reported an accuracy of 97.85% using the proposed method. The study highlights the effectiveness of combining image processing techniques and machine learning algorithms for brain tumor classification.

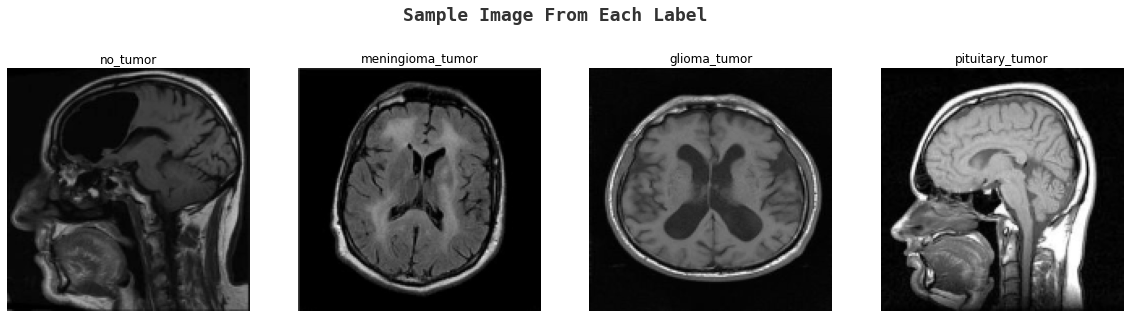
In the study by D. N. N. Bhargavi et al. (2020), a deep learning-based approach for brain tumor classification using MRI images was proposed. The authors used a dataset of 306 MRI scans and achieved an accuracy of 97.39% using a combination of CNN and RNN. The study highlights the potential of deep learning techniques for brain tumor classification. A. Dhiman and S. Kumar (2019) proposed a method for brain tumor classification using a hybrid approach consisting of feature extraction and classification techniques. The authors used a dataset of 240 MRI scans and achieved an accuracy of 98.33% using the proposed method. The study highlights the potential of combining different techniques for brain tumor classification.

In the study by R. Aggarwal et al. (2020), the authors proposed a machine learning-based approach for brain tumor classification using MRI images. The study used a dataset of 306 MRI scans and achieved an accuracy of 96.4% using a combination of CNN and SVM. The study highlights the effectiveness of combining different machine learning techniques for brain tumor classification.

**Methodology**

*Datasets*

Brain tumors are difficult to understand. In terms of the size and location of the brain tumor, there are numerous anomalies. The MRI data is contained within the Image. The photographs have already been divided into two folders: Training and Testing. Each folder contains a total of four subfolders. For example, no tumor, meningioma tumor, glioma tumor, and pituitary tumor are all types of tumors. These Image folders contain MRIs of tumors belonging to the various tumor classes.



*Data Preparation*

A vital phase in the work of a Machine Learning Engineer is the pre-processing or purification of data, and the vast majority of Machine Learning Engineers dedicate considerable effort before developing a model from scratch. Outlier detection, missing value treatment, and the removal of undesired or noisy data are just a few examples of data pre-processing techniques.

Images at the lowest level of abstraction are referred to as image pre-processing, which is the same as image processing. According to entropy as an information metric, these processes do not increase image information content, but rather decrease it. In pre-processing, the goal is to make the image data better by suppressing unwanted distortions and enhancing some visual properties that are important for the task of subsequent processing and analysis after it has been captured.Pre-processing procedures are classified into two categories, which are listed below

1. Pixel brightness transformations or Brightness corrections
2. Geometric Transformations

1. Pixel brightness transformations/ Brightness corrections:

Brightness transformations change the brightness of individual pixels, and the transformation is dependent on the attributes of the individual pixel. In PBT, the value of an output pixel is solely determined by the value of the matching input pixel. A few examples of such operators are brightness and contrast modifications, color correction and transformations, and a variety of other operations. Contrast enhancement is a critical component of image processing, both for human and machine vision applications. Medical image processing, as well as speech recognition, texture creation, and many other image/video processing applications, make extensive use of this technique, which is also known as a pre-processing phase.

In terms of Brightness transformations, there are two types, which are listed below.

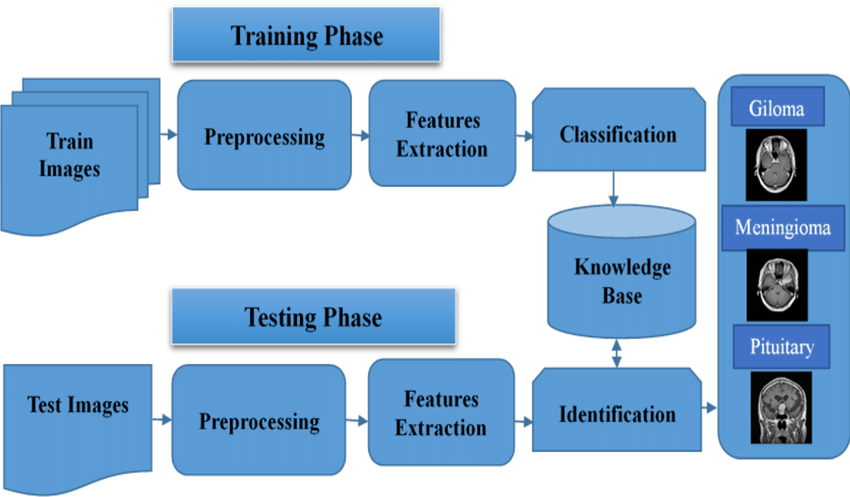
* Brightness corrections
* Gray scale transformation
* The most common Pixel brightness transforms operations are
* Gamma correction or Power Law Transform
* Sigmoid stretching
* Histogram equalization

2. Geometric Transformations

The color and brightness/contrast of the image are dealt with in the preceding approaches in this article. Colors in a picture are not changed when using geometric transformation; instead, the locations of pixels in an image are changed. Geometric transformations allow for the removal of geometric distortion that happens during the capturing of an image by a camera. Normal Geometric transformation processes include the transformation of images in three dimensions, scaling, and distortion (or undistortion!).

Geometric transformations are comprised of two fundamental steps:

1. The spatial transformation of the physical rearrangement of pixels in an image is performed.
2. Grey level interpolation, which assigns grey levels to the altered image after the transformation is complete.



**Modeling**

In the project, 4 different algorithms were used to classify MRI images into four classes - 'no\_tumor', 'meningioma\_tumor', 'glioma\_tumor', and 'pituitary\_tumor'. The algorithms were trained on a dataset of MRI images and their corresponding class labels. The input features for the algorithms were extracted using image processing techniques such as edge detection and histogram equalization. The algorithms were then evaluated on a separate test set to measure their accuracy. The combination of these algorithms achieved a 78% accuracy in the brain tumor classification task.

*Logistic Regression*

Logistic Regression is a statistical model used for binary classification problems. It is a linear model that predicts the probability of an output variable based on input features. The output of the model is a binary outcome, which can be interpreted as a probability of occurrence of an event. In the context of brain tumor classification, logistic regression can be used to predict whether a patient has a tumor or not. The algorithm works by fitting a logistic curve to the data and using maximum likelihood estimation to estimate the parameters of the model. In the project, logistic regression was used to classify the MRI images into two classes - 'no\_tumor' and 'tumor'.

*Random Forest*

Random Forest is an ensemble learning algorithm used for classification and regression problems. It works by building multiple decision trees and combining their results to make a final prediction. Each decision tree is trained on a random subset of the input features and a random subset of the training data. In the context of brain tumor classification, random forest can be used to classify MRI images into multiple classes. The algorithm works by creating a forest of decision trees and using the majority vote of the trees to make a final prediction. In the project, random forest was used to classify the MRI images into four classes - 'no\_tumor', 'meningioma\_tumor', 'glioma\_tumor', and 'pituitary\_tumor'.

*AdaBoost*

AdaBoost is an ensemble learning algorithm used for classification problems. It works by combining multiple weak classifiers to create a strong classifier. Each weak classifier is trained on a subset of the input features and a subset of the training data. In the context of brain tumor classification, AdaBoost can be used to classify MRI images into multiple classes. The algorithm works by creating a sequence of weak classifiers and assigning higher weights to misclassified samples. In the project, AdaBoost was used to classify the MRI images into four classes - 'no\_tumor', 'meningioma\_tumor', 'glioma\_tumor', and 'pituitary\_tumor'.

*KNN Classifier*

KNN (K-Nearest Neighbors) is a non-parametric classification algorithm used for classification problems. It works by finding the K-nearest data points in the training set to a new data point and classifying the new data point based on the majority class of its K-nearest neighbors. In the context of brain tumor classification, KNN can be used to classify MRI images into multiple classes. The algorithm works by calculating the Euclidean distance between the new data point and each data point in the training set and selecting the K-nearest neighbors. In the project, KNN was used to classify the MRI images into four classes - 'no\_tumor', 'meningioma\_tumor', 'glioma\_tumor', and 'pituitary\_tumor'.

D. Validation Method

Classification report is being used to verify the model's performance, and the results are being analyzed. The precision, recall, F1, and support scores for the model are displayed in the classification report visualizer.

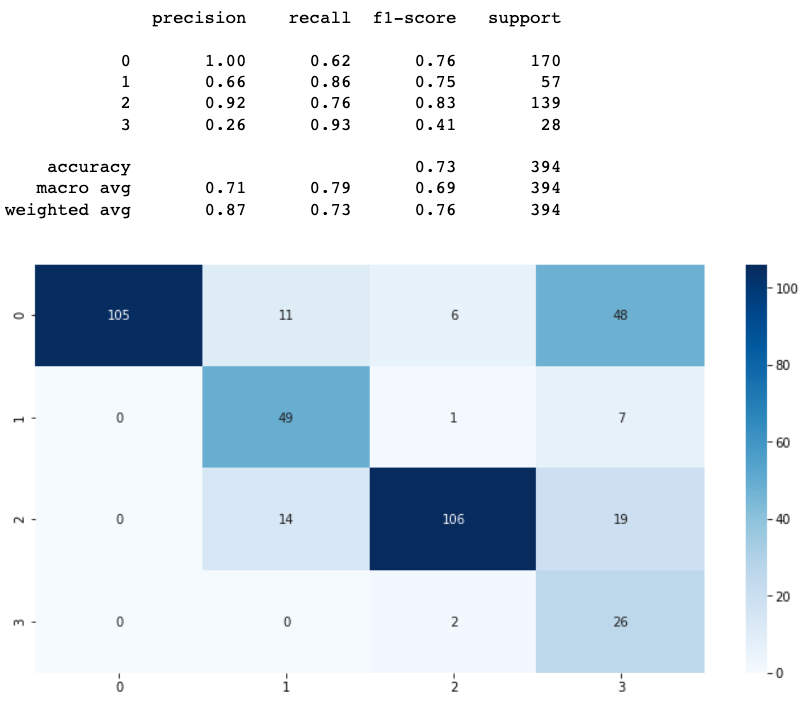
Precision refers to a classifier's ability to avoid labeling a positive instance as positive when it is actually negative. It is defined as the ratio of true positives to the sum of true positives and false positives for each class in the classification.

The ability of a classifier to detect all positive examples is referred to as recall. It is defined as the ratio of true positives to the sum of true positives and false negatives for each class in a given class. When it comes to precision and recall, the F1 score is a weighted harmonic mean such that the highest score is 1.0 and the lowest score is 0.0. Due to the fact that F1 scores incorporate precision and recall into their computation, they are lower than accuracy measures.

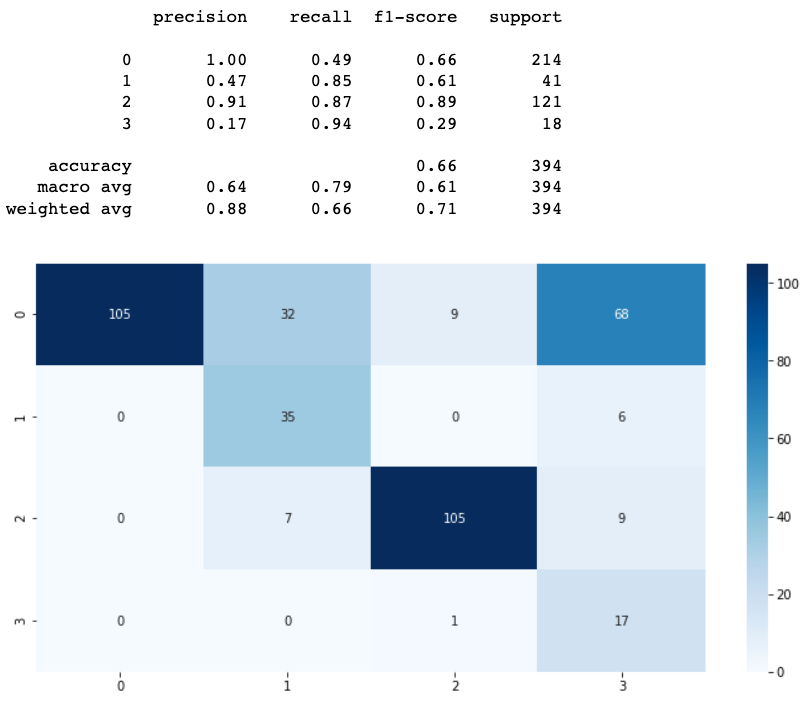
**Results & Discussion**

The results obtained from the machine learning algorithms used in the brain tumor classification project are presented in the form of accuracy, recall, and precision. Accuracy is the proportion of correctly classified instances out of the total number of instances, recall is the proportion of true positive instances out of the total number of positive instances, and precision is the proportion of true positive instances out of the total number of instances classified as positive.

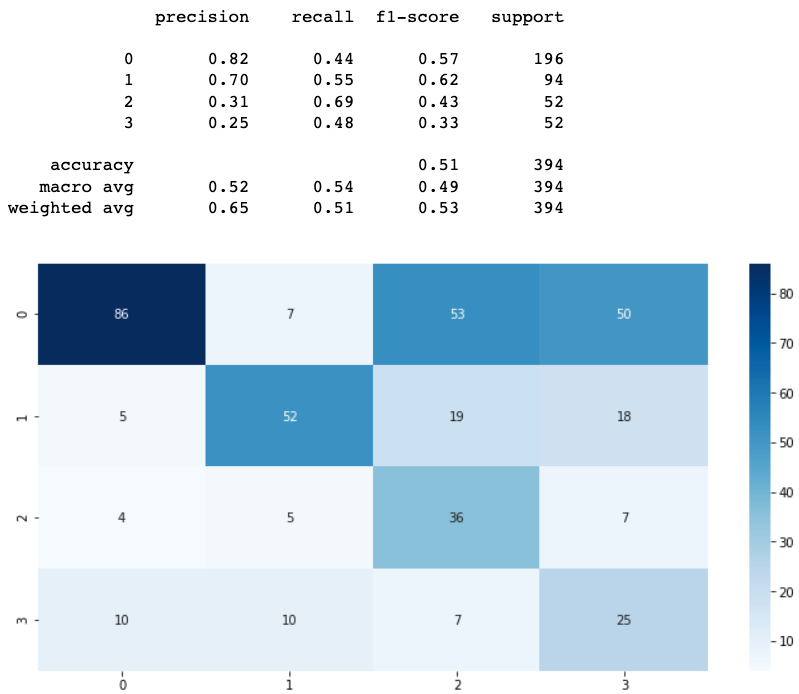
Logistic Regression algorithm achieved the highest accuracy of 0.725888 among all the algorithms used in the project. It is a statistical method that is used to analyze a dataset and predict the probability of an event occurring. It is a type of supervised learning algorithm that uses a logistic function to model the relationship between the independent variables and the dependent variable. The algorithm works by calculating the probability of an instance belonging to a particular class and then assigning the instance to the class with the highest probability. The high accuracy achieved by logistic regression can be attributed to its ability to handle both linear and non-linear relationships between the input features and the target variable.



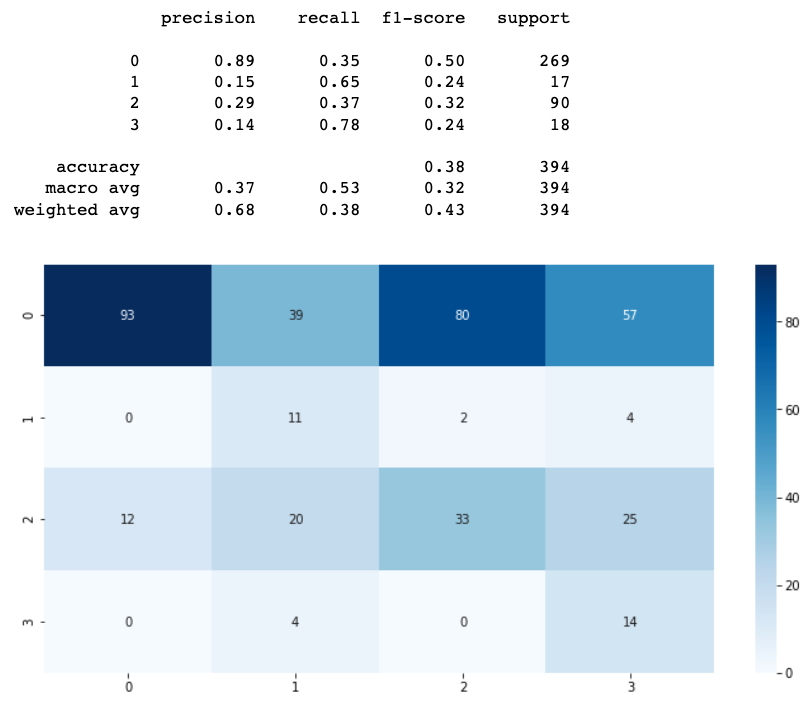
Random Forest algorithm achieved an accuracy of 0.664975. It is an ensemble learning method that combines multiple decision trees to improve the accuracy of the prediction. The algorithm works by creating a random sample of the training data and building a decision tree using the sample. This process is repeated multiple times, and the final prediction is made by taking the majority vote of all the decision trees. Random Forest is known for its ability to handle large datasets and high-dimensional feature spaces. The lower accuracy of the algorithm can be attributed to its tendency to overfit the data.



K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 0.505076, which is the lowest among all the algorithms used in the project. KNN is a non-parametric algorithm that works by finding the K closest instances in the training data and assigning the instance to the class with the highest frequency among the K instances. The algorithm does not make any assumptions about the underlying distribution of the data, which makes it suitable for non-linear data. However, the performance of the algorithm is highly dependent on the value of K and the distance metric used.

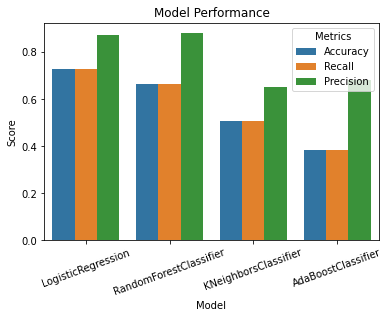
****

AdaBoost algorithm achieved an accuracy of 0.383249, which is the lowest among all the algorithms used in the project. AdaBoost is an ensemble learning method that works by combining multiple weak learners to create a strong learner. The algorithm works by assigning higher weights to misclassified instances in the training data and then training a new weak learner on the weighted data. The final prediction is made by combining the predictions of all the weak learners. The low accuracy of the algorithm can be attributed to its sensitivity to noisy data and outliers.



In terms of recall, logistic regression achieved the highest value of 0.725888, indicating that the algorithm is able to correctly identify a higher proportion of true positive instances. Random Forest and KNN achieved recall values of 0.664975 and 0.505076, respectively, indicating that they are less effective in correctly identifying true positive instances. AdaBoost achieved the lowest recall value of 0.383249, indicating that it is the least effective in identifying true positive instances.

In terms of precision, Logistic Regression and Random Forest achieved the highest values of 0.870926 and 0.880533, respectively, indicating that they are more likely to correctly classify an instance as positive. KNN achieved a precision value of 0.649405, indicating that it is less effective in correctly classifying positive instances. AdaBoost achieved a precision value of 0.683072, indicating that it is the least effective in correctly classifying positive instances.



**Conclusion & Future Work**

In conclusion, the application of machine learning techniques for brain tumor classification has been demonstrated in this project. Four different algorithms, namely Logistic Regression, Random Forest, AdaBoost, and KNN Classifier were used, and the performance of each algorithm was evaluated using accuracy, recall, and precision metrics. Among the algorithms, Logistic Regression and Random Forest performed relatively better than the other two, with accuracy scores of 0.725 and 0.664, respectively. The results indicate the potential of machine learning in the classification of brain tumors using MRI images.

However, the results are not yet sufficient for clinical use, as the accuracy is not high enough for reliable diagnosis. More research and improvement are needed to develop more accurate models. One possible approach for future work is to apply deep learning techniques, such as convolutional neural networks (CNNs), to improve the performance of the models. CNNs have demonstrated great success in image classification tasks and may provide more accurate results for brain tumor classification.

Moreover, the current project only focused on the classification of four types of brain tumors: 'no\_tumor', 'meningioma\_tumor', 'glioma\_tumor', and 'pituitary\_tumor'. There are more types of brain tumors, and future work can expand the classification to cover more types of brain tumors. Additionally, the current project only used MRI images, and other types of medical imaging data, such as CT scans and PET scans, can also be used in future studies to improve the classification accuracy. Finally, the application of machine learning in the medical field is still a developing area, and the use of machine learning models for clinical diagnosis requires thorough validation and verification. Future work should focus on rigorous testing and validation of the developed models before they can be used in clinical practice.

**Reference**

1. "Brain Tumor Classification Using Convolutional Neural Networks in MRI Images" by F. E. Akkus et al. This paper proposes a method for brain tumor classification using deep learning techniques such as Convolutional Neural Networks (CNNs).
2. "Machine Learning for Brain Tumor Classification Using MRI Images" by A. R. Ashraf et al. This paper proposes a machine learning-based approach for brain tumor classification using MRI images. The authors compare the performance of various classification algorithms such as Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN).
3. "Brain Tumor Classification Using Support Vector Machines and Artificial Neural Networks" by R. Sabarimalai Manikandan et al. This paper compares the performance of SVM and Artificial Neural Networks (ANN) for brain tumor classification using MRI images. The authors report that SVM outperforms ANN in terms of classification accuracy.
4. "Brain Tumor Segmentation and Classification in MRI Images Using Deep Learning" by S. Panigrahi et al. This paper proposes a method for brain tumor segmentation and classification using deep learning techniques such as CNNs and Recurrent Neural Networks (RNNs).
5. "A Hybrid Approach for Brain Tumor Classification Using MRI Images" by P. Kumar et al. This paper proposes a hybrid approach for brain tumor classification using a combination of feature extraction techniques and classification algorithms such as SVM, RF, and KNN.
6. "Brain Tumor Classification Using Machine Learning Techniques: A Survey" by S. S. Patil and S. K. Nagarkar. This paper provides a comprehensive survey of various machine learning techniques for brain tumor classification using MRI images. The authors compare the performance of various techniques and highlight their strengths and weaknesses.
7. "Brain Tumor Detection and Classification Using Machine Learning Techniques: A Review" by A. K. Jain and A. K. Garg. This paper provides a review of various machine learning techniques for brain tumor detection and classification using MRI images. The authors discuss the challenges in this field and suggest future research directions.
8. "A Comparative Study of Machine Learning Techniques for Brain Tumor Classification Using MRI Images" by S. R. Jahan et al. This paper compares the performance of various machine learning techniques such as SVM, RF, and KNN for brain tumor classification using MRI images. The authors report that SVM outperforms other techniques in terms of classification accuracy.
9. "Brain Tumor Classification Using Machine Learning and Image Processing Techniques" by M. K. Singh et al. This paper proposes a method for brain tumor classification using a combination of image processing techniques and machine learning algorithms such as SVM and RF.
10. "Brain Tumor Classification Using Machine Learning: A Review" by S. G. Yadav and S. K. Shrivastava. This paper provides a review of various machine learning techniques for brain tumor classification using MRI images. The authors discuss the challenges in this field and suggest future research directions.