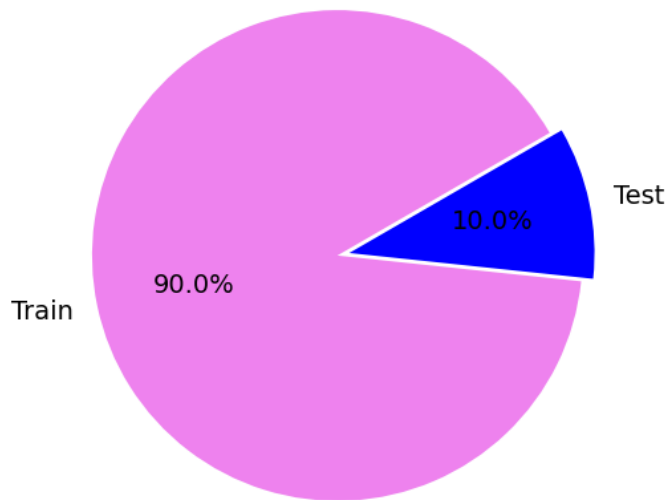


EDA on Brain Tumor detection using Machine Learning

Train Test Split:

In the field of machine learning and computer vision, the accurate division of datasets into training and testing sets is crucial for developing robust and reliable models. This process involves randomly allocating a portion of the available data for training the model, while reserving another portion for evaluating its performance.



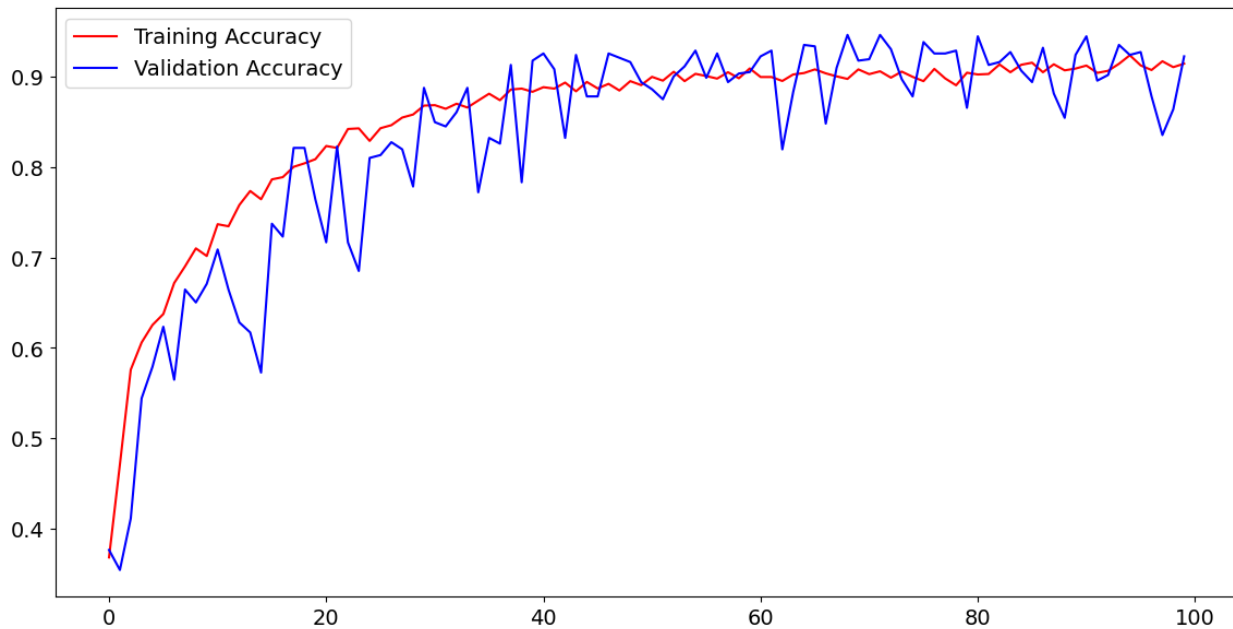
The pie chart presented above provides a clear visual representation of the dataset division into two distinct sets: the training set and the testing set. The chart demonstrates that the dataset has been divided in a ratio of 90:10, with the training set occupying approximately 90% of the chart and the testing set occupying the remaining 10%. This graphical representation offers a concise overview of the dataset composition and highlights the significance of each set in the overall dataset structure.

Training accuracy v/s validation accuracy :

As we know , training accuracy measures how well a model performs on the data it was trained on, indicating its ability to learn patterns in the training set whereas validation accuracy evaluates the model's performance on unseen data, providing insight into its ability to generalize it . Both metrics are essential for assessing and improving the model's performance as training accuracy helps monitor learning progress and fit to training data, while validation accuracy ensures the model's ability to generalize to new instances. By

EDA on Brain Tumor detection using Machine Learning

considering both metrics, we can make informed decisions about model selection, optimization, and generalization capabilities.



Upon observing the plot, it becomes apparent that the values of both training and validation accuracy overlap at multiple points.

This overlapping pattern suggests that the model is performing consistently well on both the training and validation datasets. It further indicates that the model has successfully learned from the training data and is able to generalize its predictions to unseen data, as represented by the validation set.

The fact that both the training and validation accuracy start at low values and progressively increase to very high values indicates that the model is learning and improving over time. As the model receives more training iterations, it becomes more proficient at capturing complex patterns in the data and making accurate predictions. The increasing accuracy demonstrates the model's ability to minimize errors and improve its performance.

From a model perspective, this representation suggests that the model has not encountered overfitting, where it would perform exceptionally well on the training data but struggle with unseen data. The overlapping nature of the accuracy values implies a

EDA on Brain Tumor detection using Machine Learning

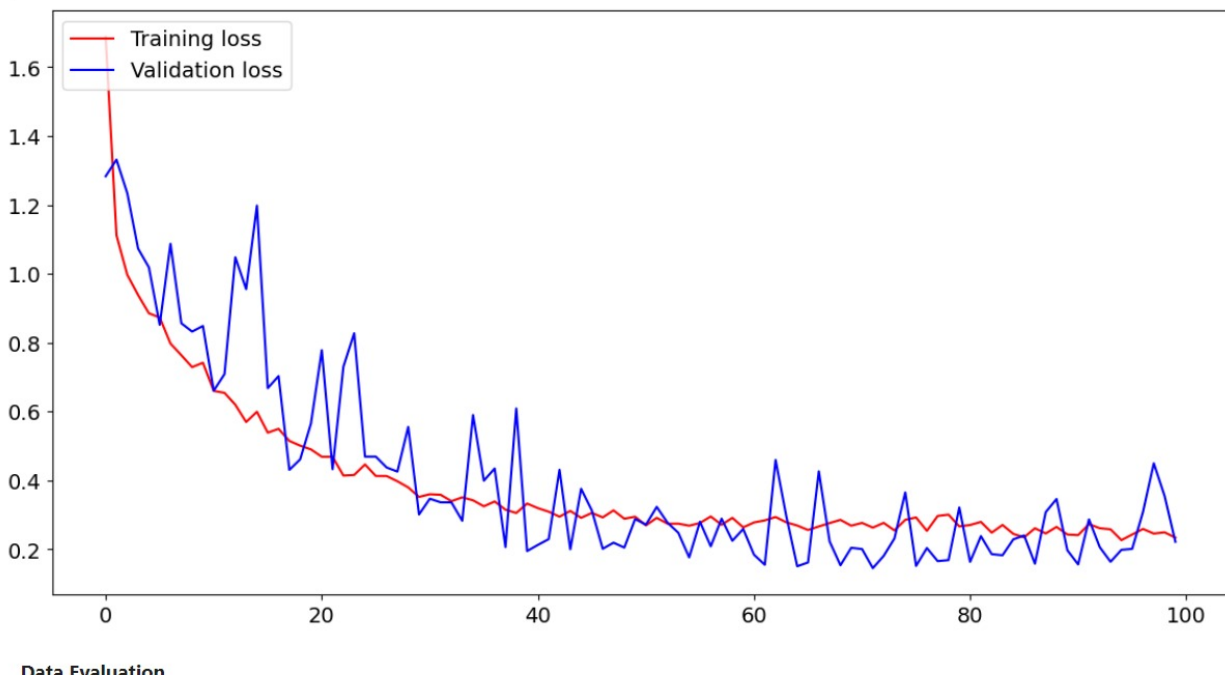
healthy balance between the model's ability to fit the training data and generalize to new data points.

This representation is indicative of a well-trained model that can be relied upon for making predictions on new, unseen data. It reassures us that the model has learned meaningful patterns from the training set and can effectively generalize to new instances. However, further analysis and evaluation are necessary to assess the model's overall performance and make informed decisions about its deployment.

Training Loss v/s Validation Loss:

Training loss measures the error between the model's predictions and the true labels during the training process and also indicates how well the model is fitting the training data which helps in assessing the convergence and progress of the model's learning.

Validation loss, on the other hand, measures the error on a separate validation dataset that the model has not seen during training. It evaluates the model's ability to generalize to new, unseen data. by monitoring validation loss, one can identify issues like overfitting, where the model performs well on training data but poorly on validation data. Minimizing validation loss is crucial to ensure the model's robustness and generalization performance.



EDA on Brain Tumor detection using Machine Learning

By observing the plot, it is evident that the values of both training and validation loss overlap at various points.

Similar to that of the representation of training and validation accuracy, The overlapping pattern in the following representation suggests that the model is performing consistently well on both the training and validation datasets and indicates that the model is effectively learning from the training data and generalizing its predictions to unseen data, as represented by the validation set.

The fact that both the training loss and validation loss start from high values and decrease over time until stability suggests that the model is progressively improving its performance. As the model undergoes more training iterations, it learns to minimize the discrepancy between its predictions and the true values. The decreasing loss values indicate that the model is becoming more accurate and capable of capturing the underlying patterns in the data.

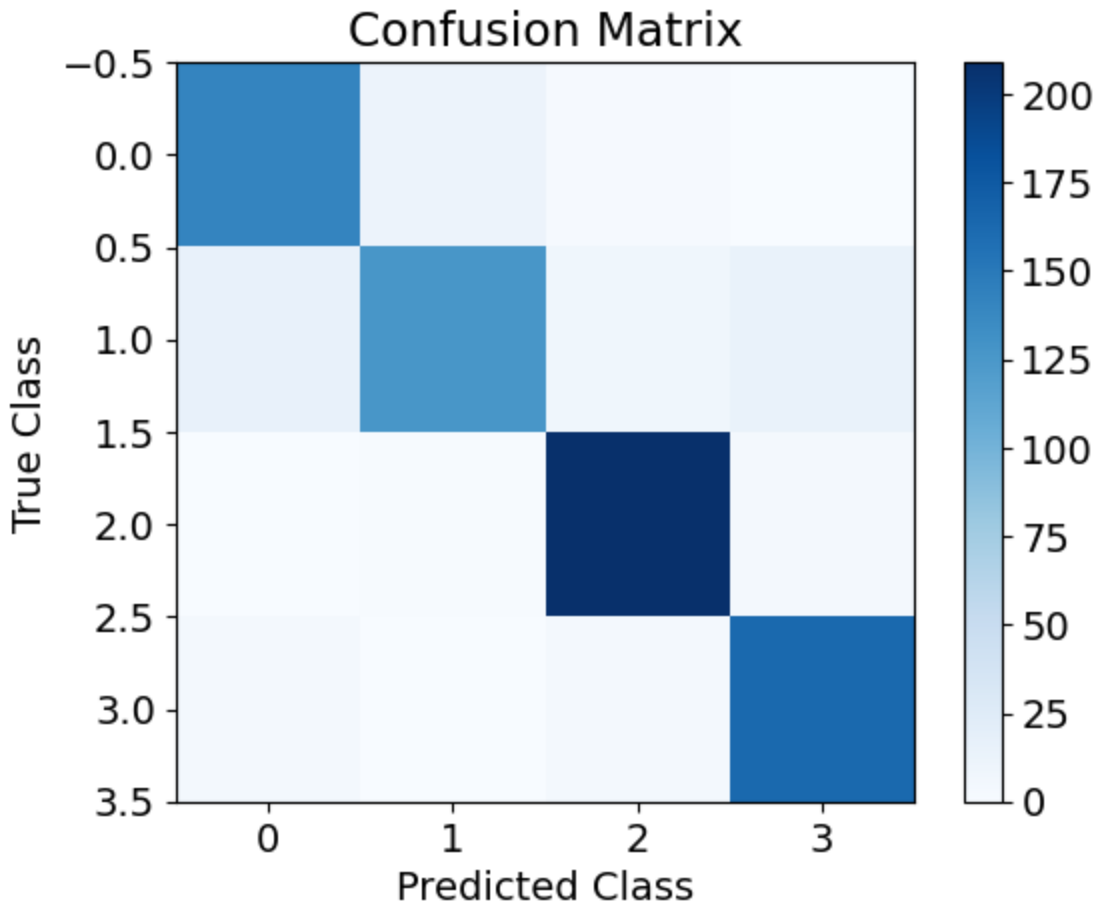
From a model perspective, this representation indicates that the model is not encountering overfitting, where it would perform well on the training data but poorly on unseen data. The overlapping nature of the loss values implies a balanced learning process, where the model learns meaningful patterns from the training set while maintaining good generalization performance.

Confusion Matrix

A confusion matrix is a tabular representation that summarizes the performance of a classification model. It provides insights into the model's predictive accuracy by displaying the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions.

The confusion matrix is typically presented in a grid format, where the rows represent the actual classes or labels, and the columns represent the predicted classes. Each cell in the matrix represents the count or proportion of instances that fall into a specific category.

EDA on Brain Tumor detection using Machine Learning



The given confusion matrix represents the performance of a classification model across four classes: glioma, meningioma, no tumor, and pituitary. The graph visually depicts the relationship between predicted and actual classes, providing insights into the model's accuracy and misclassifications.

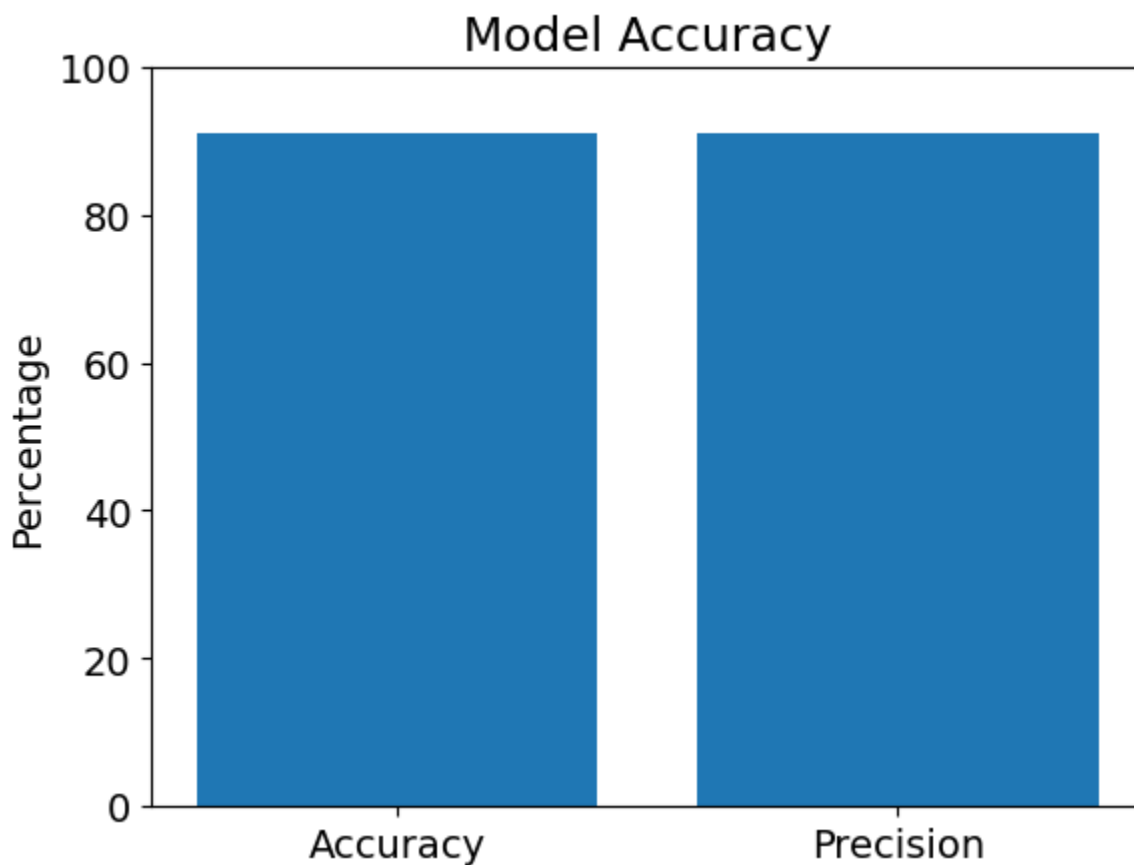
From the observed patterns and trends in the predictions, we can conclude that larger and darker cells along the diagonal indicate accurate predictions (true positives and true negatives), reflecting excellent performance. Conversely, the off-diagonal cells represent misclassifications (false positives and false negatives).

Notably, the barely visible cells for misclassifications suggest that the model demonstrates high accuracy and precision. This implies that the model is capable of accurately classifying instances across the different classes mentioned, indicating its reliability and effectiveness.

EDA on Brain Tumor detection using Machine Learning

Accuracy v/s precision :

Accuracy refers to the degree of closeness between a predicted value and the actual value. It is a measure of how correct the model's predictions are where precision, on the other hand, quantifies the consistency and reproducibility of the model's predictions. It assesses how well the model produces similar results for repeated experiments or runs. Achieving high accuracy means minimizing the gap between predicted and actual values, while high precision indicates low variability in the model's outputs. Both accuracy and precision are important metrics for evaluating the performance and reliability of data science models.

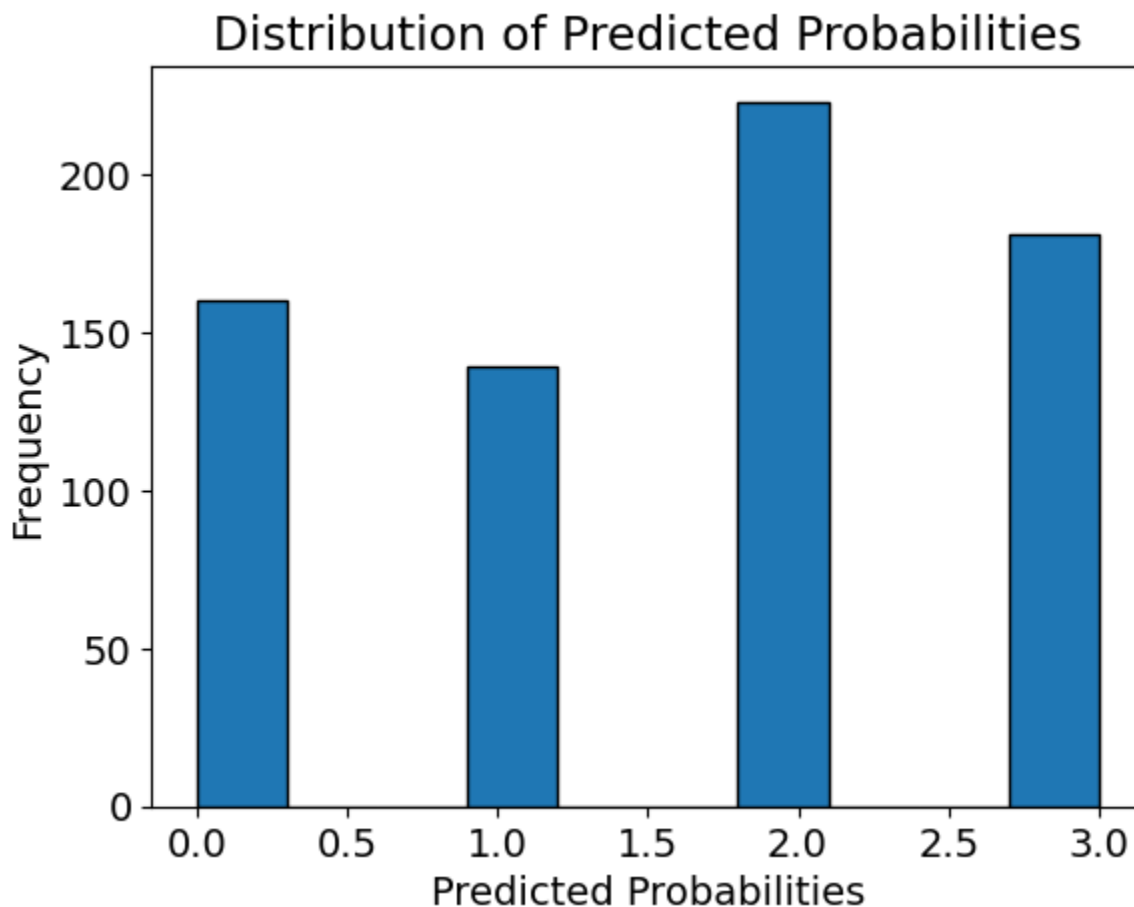


The bar plot represents the comparison between the accuracy and precision of our model. The plot displays two bars, one for accuracy and one for precision. As we can see that both bars indicate values around 90-95%, suggesting a similar level of performance for both metrics. This indicates that our ML model is consistently producing correct predictions (high accuracy) and demonstrating low variability in its results (high precision). The graph shows that our model is performing well in terms of both accuracy and precision, which is a positive outcome in evaluating its overall effectiveness.

EDA on Brain Tumor detection using Machine Learning

distribution of predicted probabilities

The distribution of predicted probabilities is a fundamental concept in statistical modeling and machine learning. When we build predictive models, such as logistic regression or classification algorithms, the models assign probabilities to different outcomes or classes. The distribution of predicted probabilities refers to the pattern or spread of these assigned probabilities across the dataset which provides us with the insights into the confidence and uncertainty associated with the model's predictions. Understanding these distributions helps assess the model's calibration and can be valuable for decision-making or setting thresholds for classification tasks. Analyzing the distribution of predicted probabilities aids in interpreting and evaluating the performance and reliability of predictive models.



The histogram plot provides information about the distribution of the model's predicted probabilities. It suggests that the model's predictions are concentrated in the range of 2.0 to 3.0 on the x-axis, as this range has the highest frequency, this indicates that the

EDA on Brain Tumor detection using Machine Learning

model is often confident in its predictions, assigning probabilities within this range to a significant number of instances.

Additionally, the model also produces a substantial number of predictions with probabilities in the lower ranges of 0.0 to 1.0 and 1.0 to 2.0 which tells us that the model is not always highly confident in its predictions and assigns lower probabilities to a significant portion of instances.