

Decision Tree Classifier Evaluation and Hyper-parameter Tuning Report

- **Introduction:**

In this report, we evaluate the performance of a Decision Tree Classifier on the Wisconsin Breast Cancer dataset. We explore various evaluation metrics and conduct hyper-parameter tuning to optimize the model's performance.

- **Dataset:**

The dataset used in this analysis is the Wisconsin Breast Cancer dataset, which contains 569 data points in the dataset: 212 – Malignant, and 357 – Benign. In this dataset, features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the characteristics of the cell nuclei present in the image below:

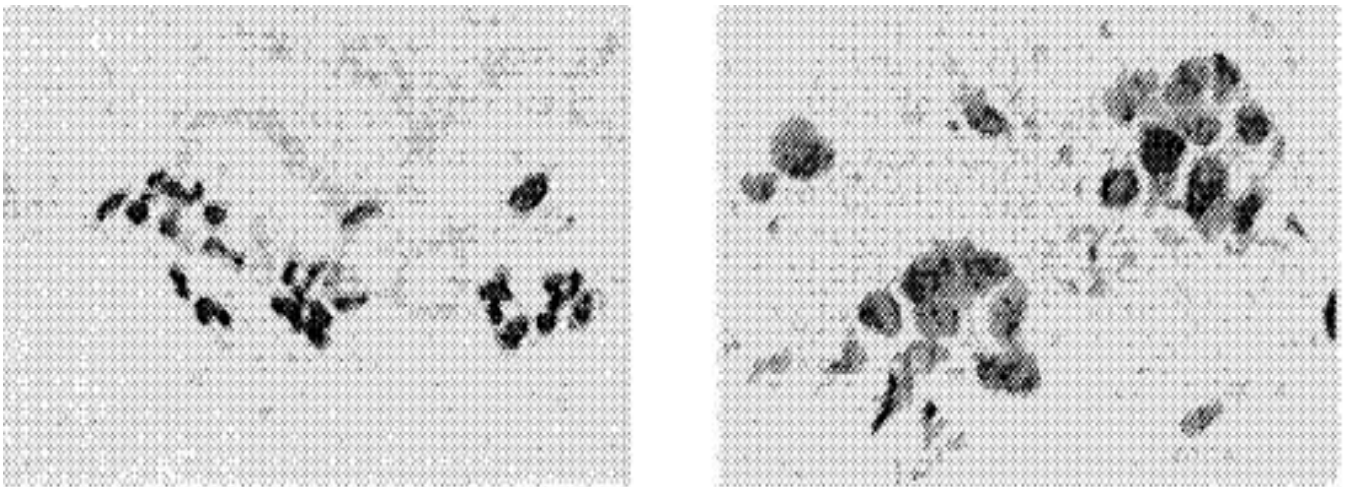


Figure 1: Images taken using the FNA test: (a) Benign, (b) Malign

- **Model Training:**

We split the dataset into training and testing sets using a 70-30 split ratio. The Decision Tree Classifier is trained on the training set with a custom hyper-parameter of entropy.

- **Evaluation Metrics:**

We have evaluated our training model on the basis of four metrics:

1. **Accuracy:** The proportion of correctly classified instances.
2. **Precision:** The ratio of correctly predicted positive observations to the total predicted positives.
3. **Recall:** The ratio of correctly predicted positive observations to the all observations in actual class.
4. **F1 Score:** The weighted average of Precision and Recall.

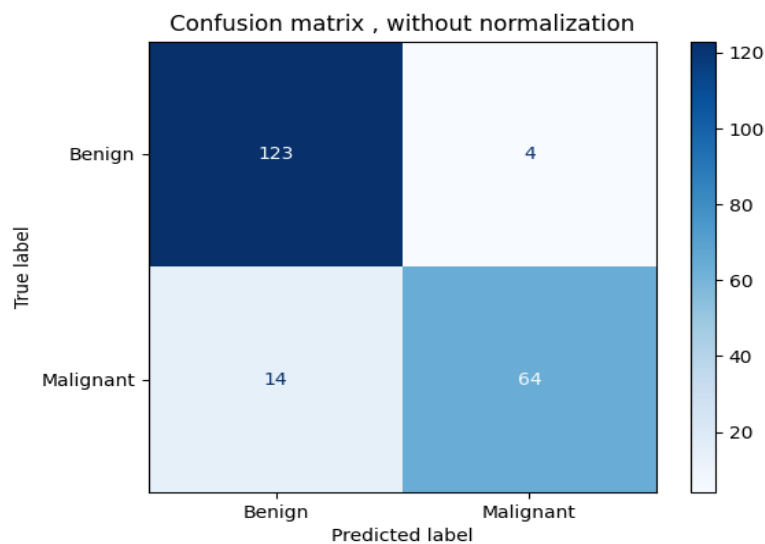
- **Results:**

1. **Accuracy:** The accuracy on the test set is approximately **93.66%**, indicating that about 93.66% of the instances in the test data were classified correctly by the model. On the other hand, the accuracy on the train set is 100%, suggesting that the model perfectly fits the training data, possibly indicating overfitting.
2. **Precision:** Precision on the test set is approximately **93.77%**, which means that out of all the instances predicted as positive (either benign or malignant), around 93.77% were correctly classified. Similarly, precision on the train set is 100%, indicating perfect precision on the training data.
3. **Recall:** Recall on the test set is approximately **93.66%**, implying that the model correctly identified around 93.66% of all actual positive instances (either benign or malignant). Recall on the train set is 100%, indicating that the model captured all positive instances in the training data.
4. **F1 Score:** The F1 score on the test set is approximately **93.71%**, which is the harmonic mean of precision and recall. It provides a balance between precision and recall. The F1 score on the train set is 100%.

Overall, the model demonstrates strong performance on both the training and test data, with high accuracy, precision, recall, and F1 score. However, the perfect scores on the training data raise concerns about potential overfitting. Further investigation into model generalization and validation on unseen data may be necessary to ensure the model's reliability in real-world scenarios.

- **Confusion Matrix:**

Confusion matrices are plotted for both the test and train sets, providing insights into the model's performance in terms of true positives, true negatives, false positives, and false negatives which can be seen in Figure 2:



In the confusion matrix provided:

True Positives (TP): There are 124 instances correctly classified as benign (negative) cases.

False Negatives (FN): There are 3 instances incorrectly classified as benign cases when they were actually malignant (positive) cases.

False Positives (FP): There are 10 instances incorrectly classified as malignant cases when they were actually benign cases.

True Negatives (TN): There are 68 instances correctly classified as malignant cases.

This information provides insights into the model's performance in terms of correctly and incorrectly classified instances, which can be further analyzed to improve the model's effectiveness in identifying benign and malignant cases.

• Hyper-parameter Tuning:

We performed hyper-parameter tuning on the Decision Tree Classifier by varying the criterion (splitting strategy) and maximum depth of the tree. The following results were obtained:

1. The best result in terms of F1 Score: {0.947007}

```
The best result in terms of F1 is: Criterion      entropy
max_depth      12
Accuracy      0.946341
Precision      0.947673
Recall        0.946341
F1            0.947007
Name: 11, dtype: object
```

2. The best result in terms of Accuracy: {0.946341}

```
The best result in terms of Accuracy is: Criterion      entropy
max_depth      12
Accuracy        0.946341
Precision        0.947673
Recall          0.946341
F1              0.947007
Name: 11, dtype: object
```

3. The best result in terms of Precision: {0.947673}

```
The best result in terms of Precision is: Criterion      entropy
max_depth      12
Accuracy        0.946341
Precision        0.947673
Recall          0.946341
F1              0.947007
Name: 11, dtype: object
```

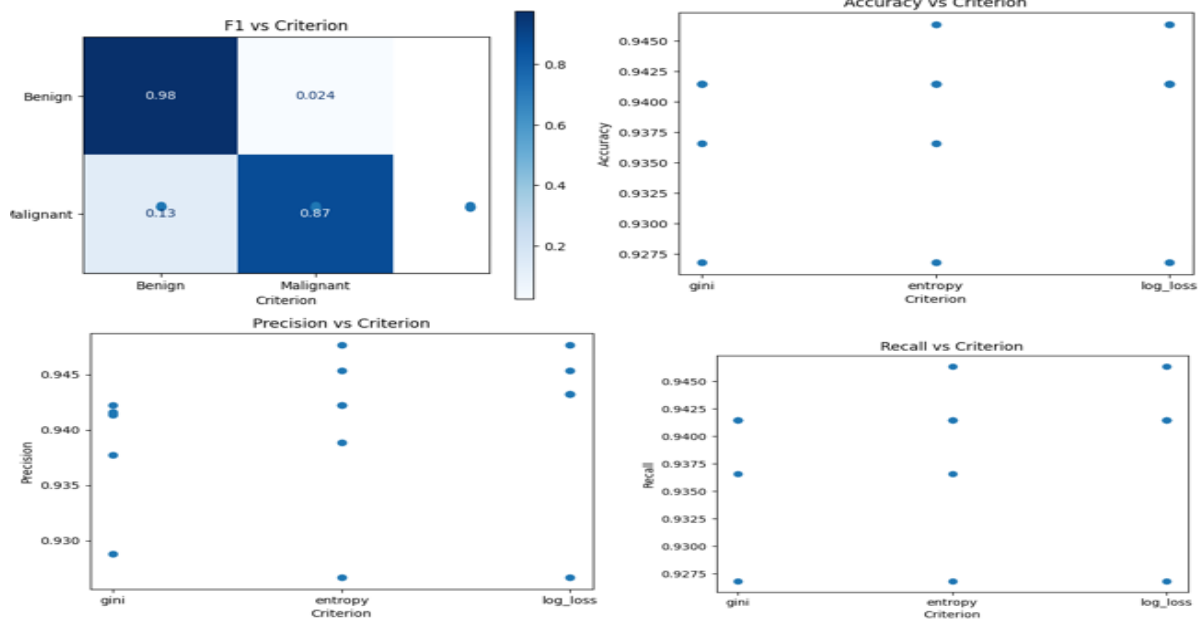
4. The best result in terms of Recall: {0.946341}

```
The best result in terms of Recall is: Criterion      entropy
max_depth      12
Accuracy        0.946341
Precision        0.947673
Recall          0.946341
F1              0.947007
Name: 11, dtype: object
```

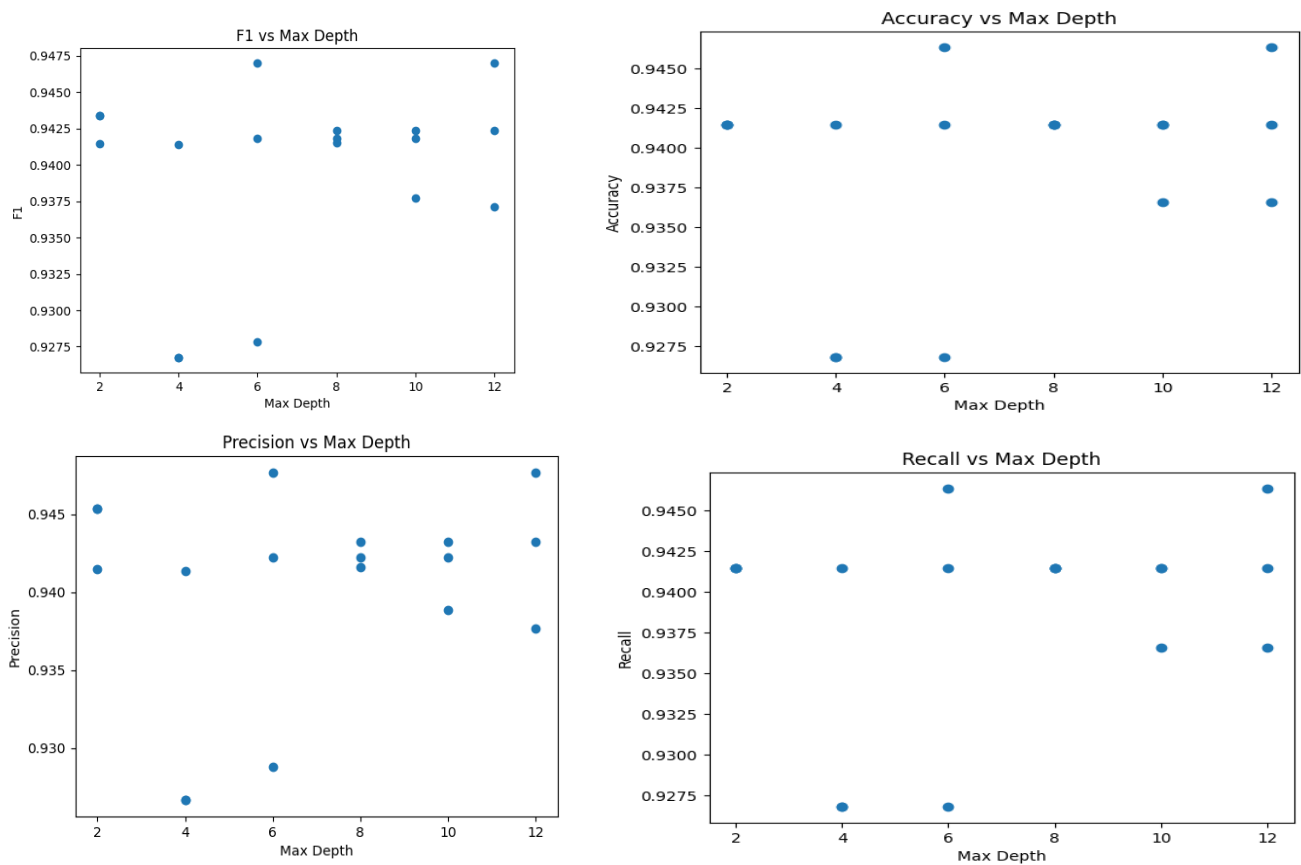
- **Plots:**

Scatter plots are provided to visualize the relationship between F1 Score, Accuracy, Precision, Recall, and the selected hyper-parameters (criterion and max depth).

1. Relationship between F1 Score, Accuracy, Precision, Recall and Criterion (Entropy, Log Loss and Gini)



2. Relationship between F1 Score, Accuracy, Precision, Recall and Max Depth (2,4,6,8,10,12)



- **Conclusion:**

The Decision Tree Classifier demonstrates promising performance in classifying breast cancer instances based on the provided features. Hyper-parameter tuning has further optimized the model's performance, yielding improved results in terms of various evaluation metrics. Further exploration and fine-tuning could potentially enhance the model's performance.