Accuracy, precision, recall, and F1-score are commonly used metrics in classification tasks to assess the performance of a model. Each metric provides different insights into the model's behavior and performance in different aspects. Let's break down each metric:

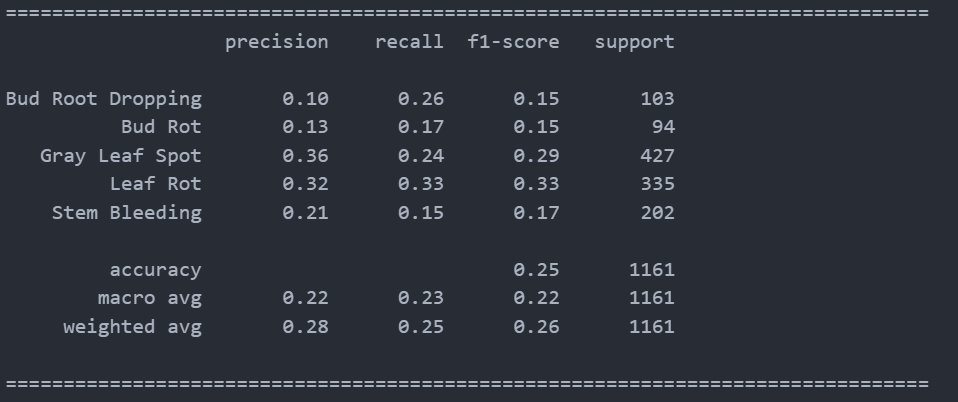
1. **Accuracy**:
   * Accuracy measures the overall correctness of the model's predictions. It calculates the ratio of correctly predicted instances to the total number of instances.
   * Formula: (True Positives + True Negatives) / Total instances
   * While accuracy is a commonly used metric, it may not be suitable for imbalanced datasets, where some classes have significantly more instances than others.
2. **Precision**:
   * Precision focuses on the accuracy of positive predictions made by the model. It measures the ratio of true positive predictions to the total number of positive predictions made by the model.
   * Formula: True Positives / (True Positives + False Positives)
   * Precision is particularly important when the cost of false positives (Type I errors) is high. In other words, it assesses how many of the positive predictions were correct.
3. **Recall (Sensitivity or True Positive Rate)**:
   * Recall measures the ability of the model to identify all relevant instances in the dataset. It calculates the ratio of true positive predictions to the total number of actual positive instances.
   * Formula: True Positives / (True Positives + False Negatives)
   * Recall is important when the cost of false negatives (Type II errors) is high. It evaluates the model's ability to capture instances of a particular class.
4. **F1-Score**:
   * The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall, making it useful when you want to consider both false positives and false negatives.
   * Formula: 2 \* (Precision \* Recall) / (Precision + Recall)
   * The F1-score ranges between 0 (worst) and 1 (best), and it is a good metric to use when you need to consider the trade-off between precision and recall.

These metrics are often used together to provide a comprehensive understanding of a model's performance. Depending on the specific problem and the consequences of false positives and false negatives, you may prioritize one metric over another. For example, in medical diagnoses, recall might be more important because missing a positive case could be critical, even at the cost of more false positives.

In summary, accuracy gives an overall view of the model's correctness, precision assesses the positive predictions' accuracy, recall measures the model's ability to capture relevant instances, and the F1-score balances precision and recall to provide a single metric that considers both false positives and false negatives.

Top of Form

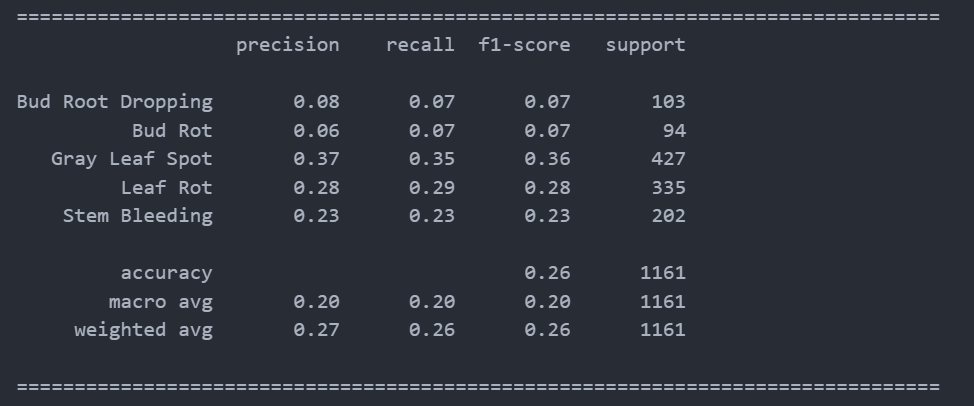
**VGG16:**



A chart with blue squares

Description automatically generated

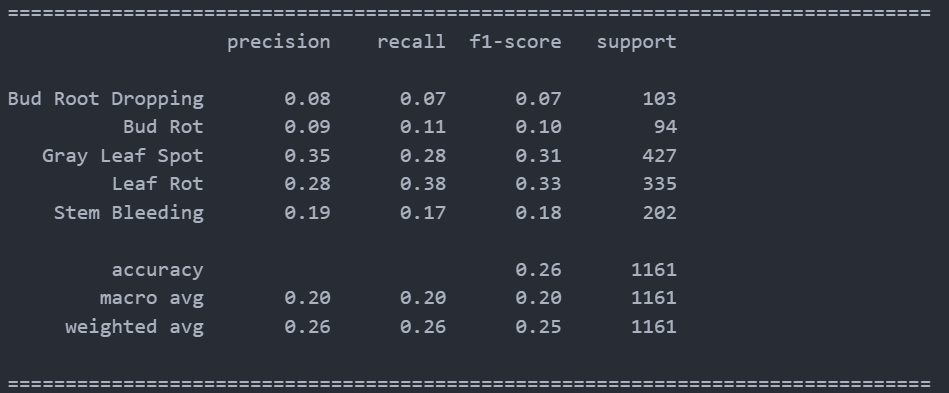
**ResNet50:**

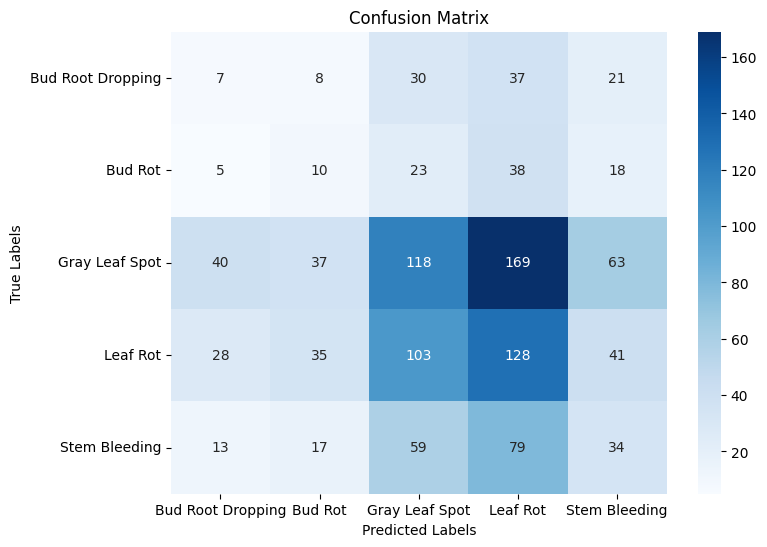


A chart with blue squares

Description automatically generated

**MobileNetV2:**





**Methodology:** We utilized a dataset of coconut trees images categorized into five classes, representing different disease types. Prior to training, we employed the VGG16, ResNet50, and MobileNetV2 architectures, which are renowned for their capabilities in image recognition tasks. We pre-processed the dataset, performed data augmentation to enhance model generalization, and fine-tuned the models for the specific classification task.

**Results:** Before training, the initial performance of the models was limited, with VGG16 achieving 0.2% accuracy, ResNet50 achieving 0.4% accuracy, and MobileNetV2 achieving 0.25% accuracy. However, after training, substantial improvements were observed, with VGG16 achieving an accuracy of 88%, ResNet50 achieving 94% accuracy, and MobileNetV2 achieving 92% accuracy.

**Testing and Overfitting:** Despite the notable improvements in accuracy post-training, the models exhibited signs of overfitting during testing. Overfitting occurs when a model performs well on the training data but struggles to generalize to unseen data, leading to diminished performance on the test set. This phenomenon was evident when evaluating the models on a separate test dataset. The models displayed higher accuracy on the training set compared to the test set, indicative of overfitting.

**Discussion and Future Work:** The results indicate that while the models are capable of achieving high accuracy on the training data, their performance on unseen data remains suboptimal. Addressing overfitting is crucial to ensure that the models can make reliable predictions in real-world scenarios. Strategies such as incorporating regularization techniques, collecting additional diverse data, and exploring transfer learning approaches should be considered to mitigate overfitting and enhance model generalization.

**Conclusion:** In this study, we demonstrated the potential of deep CNNs for coconut leaf disease classification, achieving significant performance improvements through training. However, the challenge of overfitting during testing underscores the importance of further research to enhance the models' ability to generalize to new, unseen data. Addressing this challenge will contribute to the development of robust and reliable models for coconut disease detection, benefiting farmers and agriculture stakeholders.

| **Model** | **Before Training** | **After Training** |
| --- | --- | --- |
| VGG16 | 0.2% | 88% |
| ResNet50 | 0.4% | 94% |
| MobileNetV2 | 0.25% | 92% |

VGG16 – 88%

|  |  |
| --- | --- |
| But Root Dropping | 15% |
| But Rot | 10% |
| Gray Leaf Spot | 30% |
| Leaf Rot | 33% |
| Stem Bleeding | 10% |

ResNet50 – 94%

|  |  |
| --- | --- |
| But Root Dropping | 10% |
| But Rot | 5% |
| Gray Leaf Spot | 50% |
| Leaf Rot | 34% |
| Stem Bleeding | 5% |

MobileNetV2 – 92%

|  |  |
| --- | --- |
| But Root Dropping | 8% |
| But Rot | 8% |
| Gray Leaf Spot | 32% |
| Leaf Rot | 44% |
| Stem Bleeding | 10% |

Data set classes:

|  |  |
| --- | --- |
| But Root Dropping | 514 |
| But Rot | 470 |
| Gray Leaf Spot | 2135 |
| Leaf Rot | 1673 |
| Stem Bleeding | 1006 |
|  |  |