

Performance Comparison: Decision Tree vs Random Forest

We evaluated the best **Decision Tree** and **Random Forest** models on the Breast Cancer Wisconsin dataset. The performance metrics are summarized below:

Metric	Decision Tree (Test)	Random Forest (Test)
Accuracy	0.9767	0.9535
Precision	0.9833	0.9516
Recall	0.9833	0.9833
F1-Score	0.9833	0.9672
Confusion Matrix	[[25, 1], [1, 59]]	[[23, 3], [1, 59]]

Observations

1. Decision Tree Performance

- The single Decision Tree achieved slightly higher test accuracy (97.7%) than the Random Forest (95.3%).
- Precision and F1-score are also higher for the Decision Tree.
- Confusion matrix shows only **2 misclassifications** on the test set.

2. Random Forest Performance

- Random Forest shows slightly lower accuracy and precision but maintains **perfect recall** for the malignant class (all positives correctly detected).
- Confusion matrix indicates **4 misclassifications** (slightly more than the single tree).
- The lower test accuracy is likely due to **variance reduction** from averaging many trees:
 - Individual trees in the forest are trained on **bootstrap samples** with **random feature subsets**, making them slightly less overfitted than the single tree.
 - This can slightly reduce test performance on a small dataset or particular test split.

3. Bias–Variance Interpretation

- The Decision Tree has **low bias but higher variance**, fitting the training data almost perfectly (train accuracy 98.96% vs test accuracy 97.67%).
- The Random Forest reduces variance via bagging and feature randomness:
 - Training may be slightly underfit compared to the single tree.
 - Generalization is more stable across different splits, even if a single test set shows slightly lower accuracy.

Conclusion

- While the single Decision Tree shows marginally higher test accuracy on this dataset, the **Random Forest is more robust** and less sensitive to fluctuations in training data.
 - With more trees or averaging over multiple random seeds, the Random Forest would likely **match or surpass the single tree**, especially on larger datasets.
 - These results highlight the classical **bias–variance trade-off**:
 - Decision Tree → low bias, high variance
 - Random Forest → slightly higher bias, much lower variance → more stable predictions
-