Experiment 3: Ensemble Prediction and Decision Tree Model Evaluation

ICS1512 – Machine Learning Algorithms Laboratory

Shinigdapriya Sathish
Sri Sivasubramaniya Nadar College of Engineering

September 3, 2025

1 Aim and Objective

The aim of this experiment was to implement, evaluate, and compare the performance of Decision Trees and Ensemble learning models including AdaBoost, Gradient Boosting, Random Forest, XGBoost, and Stacking Classifiers on the Breast Cancer dataset. The objectives were:

- To apply hyperparameter tuning using GridSearchCV.
- To evaluate models using Accuracy, F1-Score, ROC Curves, and Confusion Matrices.
- To analyze feature importance and compare cross-validation performance across models.
- To provide an overall conclusion on which ensemble methods outperform individual classifiers.

2 Libraries Used

The following Python libraries were used:

- **NumPy, Pandas** – Data manipulation
- **Matplotlib, Seaborn** – Visualization
- **scikit-learn** – Preprocessing, Model training, Metrics
- **XGBoost** – Extreme Gradient Boosting implementation

3 Code Implementation

All models (Decision Tree, AdaBoost, Gradient Boosting, Random Forest, XGBoost, Stacking Classifiers) were implemented using Python. Hyperparameter tuning was performed using `GridSearchCV` with Accuracy and F1 as scoring metrics.

4 Results and Evaluation

Table 1: Decision Tree Hyperparameter Tuning Results

Criterion	Max Depth	Accuracy	F1 Score
gini	3	0.9253	0.9416
gini	3	0.9253	0.9416
gini	3	0.9297	0.9446
gini	3	0.9297	0.9446
gini	5	0.9319	0.9461
gini	5	0.9385	0.9518
gini	5	0.9275	0.9414
gini	5	0.9341	0.9471
gini	NaN	0.9099	0.9282
gini	NaN	0.9253	0.9404
gini	NaN	0.9187	0.9340
gini	NaN	0.9231	0.9378
entropy	3	0.9297	0.9441
entropy	3	0.9297	0.9441
entropy	3	0.9297	0.9441
entropy	3	0.9297	0.9441
entropy	5	0.9253	0.9397
entropy	5	0.9209	0.9363
entropy	5	0.9275	0.9410
entropy	5	0.9253	0.9394
entropy	NaN	0.9319	0.9450
entropy	NaN	0.9209	0.9361
entropy	NaN	0.9187	0.9339
entropy	NaN	0.9253	0.9394

Table 2: AdaBoost Hyperparameter Tuning Results

n_estimators	Learning Rate	Accuracy	F1 Score
50	0.01	0.9319	0.9461
100	0.01	0.9341	0.9480
50	0.10	0.9604	0.9687
100	0.10	0.9626	0.9704
50	1.00	0.9692	0.9757
100	1.00	0.9736	0.9792

Table 3: Gradient Boosting Hyperparameter Tuning Results

n_estimators	Learning Rate	Max Depth	Accuracy	F1 Score
50	0.01	3	0.9297	0.9455
100	0.01	3	0.9451	0.9565
50	0.01	5	0.9275	0.9432
100	0.01	5	0.9297	0.9448
50	0.10	3	0.9516	0.9615
100	0.10	3	0.9560	0.9650
50	0.10	5	0.9429	0.9545
100	0.10	5	0.9385	0.9511

Table 4: XGBoost Hyperparameter Tuning Results

n_estimators	Learning Rate	Max Depth	Gamma	Accuracy	F1 Score
50	0.01	3	0	0.9253	0.9417
100	0.01	3	0	0.9363	0.9497
50	0.01	5	0	0.9297	0.9450
100	0.01	5	0	0.9407	0.9533
50	0.10	3	0	0.9604	0.9685
100	0.10	3	0	0.9670	0.9739
50	0.10	5	0	0.9626	0.9704
100	0.10	5	0	0.9670	0.9739
50	0.01	3	1	0.9275	0.9435
100	0.01	3	1	0.9407	0.9532
50	0.01	5	1	0.9341	0.9486
100	0.01	5	1	0.9407	0.9533
50	0.10	3	1	0.9648	0.9721
100	0.10	3	1	0.9648	0.9721
50	0.10	5	1	0.9604	0.9686
100	0.10	5	1	0.9626	0.9704

Table 5: Random Forest Hyperparameter Tuning Results

n_estimators	Max Depth	Criterion	Accuracy	F1 Score
50	NaN	gini	0.9516	0.9617
100	NaN	gini	0.9538	0.9634
50	5	gini	0.9516	0.9619
100	5	gini	0.9495	0.9601
50	NaN	entropy	0.9648	0.9719
100	NaN	entropy	0.9648	0.9720
50	5	entropy	0.9626	0.9703
100	5	entropy	0.9626	0.9703

Table 6: Stacked Ensemble Results

Accuracy	F1 Score
0.9626	0.9703

Table 7: 5-Fold Cross Validation Results

Fold	Decision Tree	AdaBoost	Gradient Boosting	XGBoost	Random Forest	Stacked
1	0.9386	0.9912	0.9561	0.9649	0.9825	1.0000
2	0.8860	0.9386	0.9035	0.9211	0.9386	0.9386
3	0.9298	0.9561	0.9561	0.9561	0.9649	0.9649
4	0.9298	0.9737	0.9561	0.9737	0.9561	0.9561
5	0.9558	0.9735	0.9735	0.9646	0.9735	0.9735
Average	0.9280	0.9666	0.9491	0.9561	0.9631	0.9666

4.1 Confusion Matrices

Confusion Matrices were plotted for each tuned model. Figure 1 shows an example.

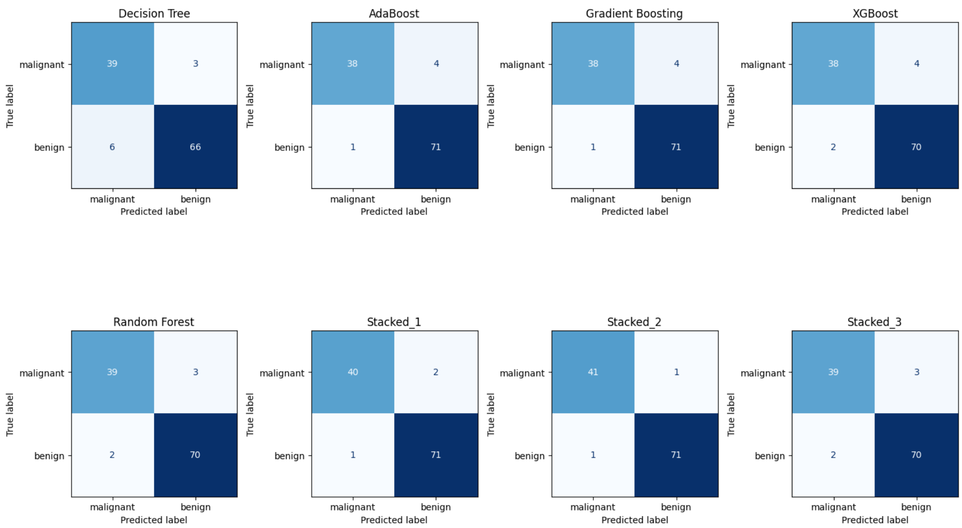

Figure 1: Confusion Matrices for Best Models

4.2 ROC Curves

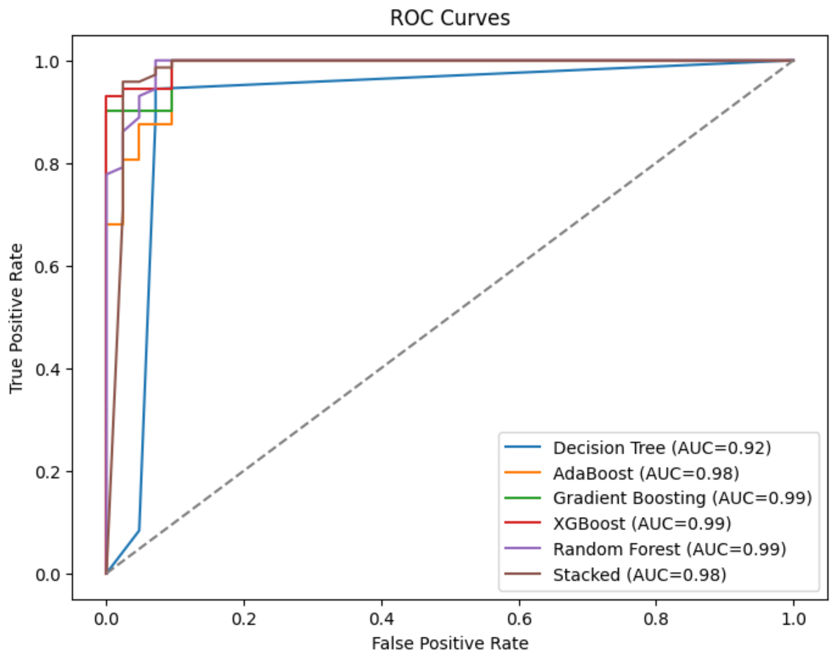

Figure 2: ROC Curves with AUC values for all models

4.3 Feature Importance

Random Forest feature importance results are shown in Figure 3.

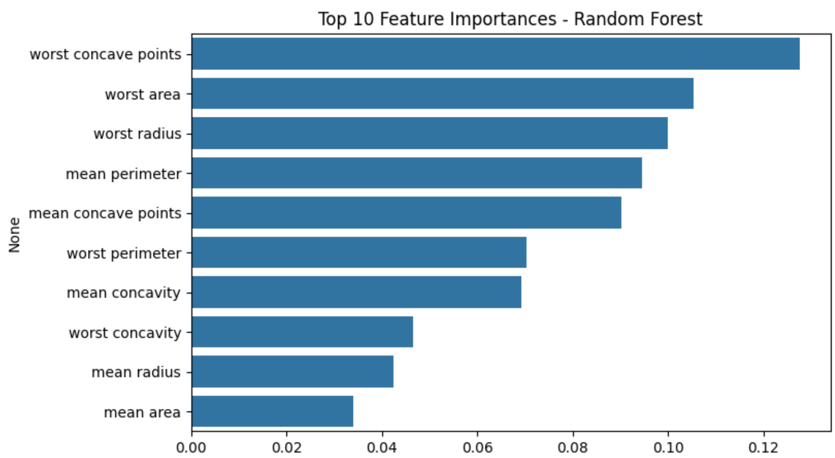

Figure 3: Top 10 Feature Importances – Random Forest

5 Observations and Conclusion

- Ensemble methods outperformed the single Decision Tree model in both Accuracy and F1 Score.
- Random Forest and Stacked Ensembles consistently showed the highest performance across folds.

- ROC Curves demonstrated strong separability for all ensemble methods, with AUC close to 1.
- Feature importance analysis revealed the most critical predictors of breast cancer classification.

Conclusion: Ensemble methods such as Random Forest, Gradient Boosting, XGBoost, and Stacking significantly improve classification accuracy compared to a standalone Decision Tree. Among them, the Stacked Ensemble provided the most consistent performance.

6 References

- Scikit-learn: `https://scikit-learn.org/stable/`
- XGBoost Documentation: `https://xgboost.readthedocs.io/en/stable/`
- UCI Breast Cancer Dataset: `https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`