

SHAP Explanation Report - Breast Cancer Classification

Objective

This report aims to explain the predictions made by a Random Forest model trained on the Breast Cancer Wisconsin dataset using SHAP (SHapley Additive exPlanations), a game-theoretic approach to interpreting machine learning models.

Global Feature Importance (SHAP Summary)

The global feature importance is visualized using SHAP's bar plot, which displays the average absolute SHAP value for each feature, indicating its overall impact on the model.

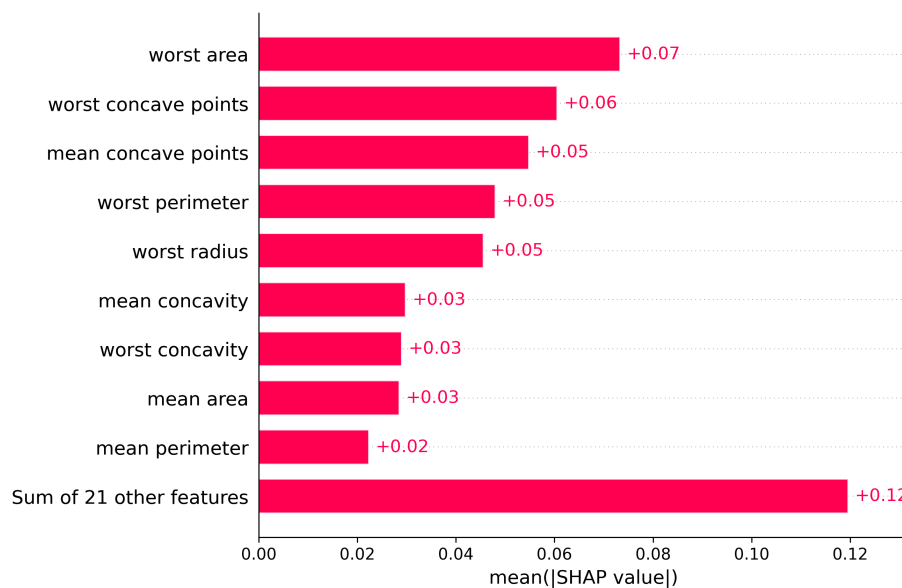


Figure 1: SHAP Bar Plot

Top important features:

Rank	Feature	Mean SHAP Value
1	worst area	+0.07
2	worst concave points	+0.06
3	mean concave points	+0.05
4	worst perimeter	+0.05
5	worst radius	+0.05

These features have the highest contribution to the model’s prediction of whether a tumor is malignant or benign.

Local Explanation (SHAP Force Plot)

We selected one example (index 0) and visualized the individual prediction explanation using SHAP’s `force_plot`.

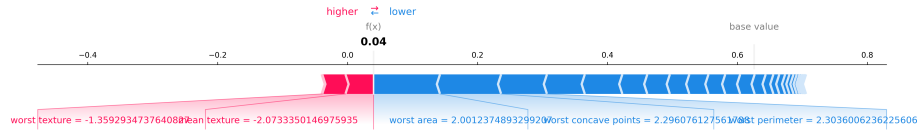


Figure 2: SHAP Force Plot

Interpretation:

- Features like `worst area` and `worst concave points` strongly pushed the prediction toward **malignant (higher output)**.
- Features such as `worst texture` and `mean texture` pushed the prediction toward **benign (lower output)**.
- The predicted output ($f(x) = 0.04$) is close to the **base value** (expected output), meaning the example is likely classified as benign.

Conclusion

- The Random Forest model relies heavily on shape-related features (area, concavity, perimeter) to make decisions.

- SHAP visualizations help interpret both **global importance** and **individual decisions**.
 - This enhances the **transparency** and **trust** in the model's behavior, especially in sensitive domains like healthcare.
-

Appendix

- Model: RandomForestClassifier (n_estimators=100)
- Preprocessing: StandardScaler
- Explanation Tool: SHAP 0.41.0