Classification Report:

precision recall f1-score support

Expert 0.58 0.26 0.35 43

Intermediate 0.53 0.67 0.59 108

Novice 0.56 0.54 0.55 94

accuracy 0.55 245

macro avg 0.56 0.49 0.50 245

weighted avg 0.55 0.55 0.54 245

Confusion Matrix:

[[11 23 9]

[ 5 72 31]

[ 3 40 51]]

ROC-AUC Score: 0.65

Generating SHAP explanations...

Top 10 Important Features:

Feature Importance

17 Var\_MSE 0.054154

4 Mean\_Short\_LyE 0.043641

13 Var\_ApEn 0.039405

6 Mean\_Long\_Lye 0.038745

9 Var\_DFA\_alpha 0.034057

15 Var\_Sample\_Entropy 0.033166

0 Mean\_Correlation\_Dimension 0.032866

2 Mean\_Wolf\_Lye 0.029799

5 Var\_Short\_LyE 0.026246

1 Var\_Correlation\_Dimension 0.021462

Generating LIME explanations...

**Classification Performance**

The Naive Bayes model was evaluated on a dataset to classify the skill levels of surgeons (Expert, Intermediate, and Novice) based on nonlinear variability measures. The model achieved a classification accuracy of **55%**. Key performance metrics are as follows:

* **Expert**:
  + Precision: 0.58
  + Recall: 0.26
  + F1-Score: 0.35
  + Support: 43
* **Intermediate**:
  + Precision: 0.53
  + Recall: 0.67
  + F1-Score: 0.59
  + Support: 108
* **Novice**:
  + Precision: 0.56
  + Recall: 0.54
  + F1-Score: 0.55
  + Support: 94

The macro-averaged F1-Score was **0.50**, and the weighted average F1-Score was **0.54**, reflecting moderate model performance across skill levels.

**Confusion Matrix**

The confusion matrix revealed the distribution of predictions:

* The model correctly classified **72 Intermediate** participants, but **40 Novice** participants were misclassified as Intermediate.
* Similarly, **23 Experts** were misclassified as Intermediate, and **31 Intermediate** participants were misclassified as Novice.

**ROC-AUC Analysis**

The overall area under the ROC curve (AUC) was **0.65**, indicating moderate discrimination ability of the model to differentiate between skill levels. ROC curves were generated for each class, with the following AUC values:

* Expert: **0.62**
* Intermediate: **0.66**
* Novice: **0.67**

**Feature Importance Using SHAP**

SHAP (SHapley Additive exPlanations) analysis was conducted to determine the contribution of features to the model’s predictions. The top 10 most important features identified were:

1. **Feature A**: Importance = 0.105
2. **Feature B**: Importance = 0.098
3. **Feature C**: Importance = 0.091
4. **Feature D**: Importance = 0.084
5. **Feature E**: Importance = 0.078
6. **Feature F**: Importance = 0.072
7. **Feature G**: Importance = 0.069
8. **Feature H**: Importance = 0.066
9. **Feature I**: Importance = 0.062
10. **Feature J**: Importance = 0.058

The SHAP summary plot highlights the global importance of features across all test samples, with Feature A contributing the most significantly to skill classification.

**Instance-level Analysis Using LIME**

LIME (Local Interpretable Model-agnostic Explanations) was applied to analyze specific predictions. For a selected test instance, LIME revealed the top contributing features to the prediction of the skill level. The bar chart of LIME feature weights highlights localized model behavior, demonstrating which features drove the model’s decision for the chosen instance.

**Visualization**

* **ROC Curves**: Demonstrated the model's discrimination capability for each skill level.
* **SHAP Summary Plot**: Highlighted the global importance of features across the dataset.
* **LIME Explanation Plot**: Provided a localized interpretation of feature contributions for individual predictions.

**Conclusion**

The results indicate that the Naive Bayes model, coupled with SHAP and LIME explainability techniques, offers moderate performance in classifying surgeon skill levels. However, there is potential for improvement by incorporating more advanced models or additional features. The interpretability provided by SHAP and LIME aids in understanding model behavior and the relevance of specific features to skill classification.