**Methods**

Hyperparameters for the Naïve Bayes model were tuned using a grid search approach with cross-validation. The following hyperparameters were optimized:

* **Smoothing parameter (α):** Values ranged from 0.1 to 1.0 in increments of 0.1.
* **Feature selection:** Recursive Feature Elimination (RFE) was used to identify the top 10 most relevant features.

The best-performing model was selected based on its performance on the validation set.

**Evaluation Metrics**

Model performance was evaluated using precision, recall, F1-score, and accuracy for each skill level. A confusion matrix and ROC-AUC score were also generated to assess classification performance and overall discriminatory ability.

**Results**

**Classification Performance**

The Naïve Bayes model achieved an overall accuracy of 55%, with the following performance metrics for each skill level:

* **Expert:** Precision = 0.58, Recall = 0.26, F1-score = 0.35
* **Intermediate:** Precision = 0.53, Recall = 0.67, F1-score = 0.59
* **Novice:** Precision = 0.56, Recall = 0.54, F1-score = 0.55

The weighted average precision, recall, and F1-score were 0.55, 0.55, and 0.54, respectively. The confusion matrix (Table 1) illustrates the distribution of true and predicted classifications:

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted \ True** | **Expert** | **Intermediate** | **Novice** |
| **Expert** | 11 | 23 | 9 |
| **Intermediate** | 5 | 72 | 31 |
| **Novice** | 3 | 40 | 51 |

The model achieved a ROC-AUC score of 0.65, indicating moderate discriminatory power.

#### Explainability

LIME explanations highlighted the localized contributions of features for individual predictions. For instance, higher values of **Mean\_Short\_LyE** and **Var\_MSE** were strongly associated with intermediate skill levels, while higher **Var\_DFA\_alpha** and **Var\_ApEn** values differentiated novice from expert surgeons. These insights underscore the potential of XAI in identifying critical physiological and biomechanical factors influencing surgical performance.

The integration of XAI techniques with wearable sensor data provides a transparent framework for assessing surgical skill levels. The identified biomarkers, such as Lyapunov exponents and entropy measures, offer actionable insights that can inform personalized training interventions. Despite moderate accuracy, the interpretability of the model makes it a valuable tool for enhancing surgical education and rehabilitation strategies.