### Methods: Model Training and Hyperparameter Selection

#### Random Forest

The Random Forest classifier was implemented using the RandomForestClassifier from the sklearn.ensemble module. The following hyperparameters were used:

* **Number of Estimators (n\_estimators)**: 100
* **Random State (random\_state)**: 42 (to ensure reproducibility)
* **Criterion**: Default ("gini")
* **Max Depth**: None (allowing the trees to expand until all leaves are pure or contain fewer than the minimum number of samples)
* **Min Samples Split (min\_samples\_split)**: 2
* **Min Samples Leaf (min\_samples\_leaf)**: 1
* **Bootstrap**: True

These parameters were selected based on the standard defaults provided by the sklearn library, as initial exploratory experiments indicated reasonable performance without extensive hyperparameter tuning.

#### XGBoost

The XGBoost classifier was implemented using the XGBClassifier from the xgboost library. The hyperparameters were:

* **Number of Estimators (n\_estimators)**: Default (100)
* **Learning Rate (learning\_rate)**: Default (0.3)
* **Max Depth (max\_depth)**: Default (6)
* **Subsample**: Default (1.0, using all samples for each tree)
* **Random State (random\_state)**: 42
* **Evaluation Metric (eval\_metric)**: "mlogloss" (multi-class logarithmic loss)
* **Use Label Encoder**: False (to avoid warnings with newer versions of the library)

Similar to the Random Forest classifier, default parameters were used for the XGBoost model to establish a baseline. Hyperparameter tuning was not performed in this study to maintain consistency across models and focus on explainability rather than optimization.

**Model Performance**

**Random Forest**

The Random Forest classifier achieved an overall accuracy of **60%** in classifying skill levels into "Expert," "Intermediate," and "Novice." The detailed performance metrics for each class are:

* **Expert**: Precision = 0.62, Recall = 0.12, F1-Score = 0.20
* **Intermediate**: Precision = 0.60, Recall = 0.75, F1-Score = 0.67
* **Novice**: Precision = 0.60, Recall = 0.66, F1-Score = 0.63

The weighted average F1-score was 0.57. The ROC-AUC score for the Random Forest model was **0.97**, indicating strong discriminative performance. The confusion matrix showed that most misclassifications occurred for the "Expert" class, which had the lowest recall (12%).

**XGBoost**

The XGBoost model achieved an accuracy of **59%**. Class-wise performance metrics are as follows:

* **Expert**: Precision = 0.33, Recall = 0.16, F1-Score = 0.22
* **Intermediate**: Precision = 0.63, Recall = 0.71, F1-Score = 0.67
* **Novice**: Precision = 0.59, Recall = 0.64, F1-Score = 0.61

The weighted average F1-score was 0.57. The ROC-AUC score for XGBoost was **0.96**, slightly lower than Random Forest. Similar to Random Forest, the "Expert" class had the lowest recall (16%), indicating challenges in predicting this class accurately. The confusion matrix highlighted consistent misclassifications across all classes, with a notable overlap between "Novice" and "Intermediate."

**Explainability Analyses**

**Random Forest**

Using SHAP values, the top 10 most important features for the Random Forest model were identified. The most influential features included:

1. **Mean\_Long\_Lye** (SHAP Importance: 0.0278)
2. **Mean\_Short\_LyE** (SHAP Importance: 0.0246)
3. **Mean\_Correlation\_Dimension** (SHAP Importance: 0.0243)
4. **Mean\_Generalized\_Hurst\_Exp** (SHAP Importance: 0.0240)

Additional key features such as **Mean\_Wolf\_Lye**, **Mean\_ApEn**, and **Mean\_Sample\_Entropy** contributed significantly to the model’s predictions. LIME-based explanations corroborated these findings, demonstrating that features related to Lyapunov exponents and entropy measures were pivotal in distinguishing skill levels.

**XGBoost**

SHAP analysis for XGBoost revealed similar trends, with top features including:

1. **Mean\_Generalized\_Hurst\_Exp** (SHAP Importance: 0.3154)
2. **Mean\_Long\_Lye** (SHAP Importance: 0.3128)
3. **Mean\_Correlation\_Dimension** (SHAP Importance: 0.3082)
4. **Mean\_ApEn** (SHAP Importance: 0.2819)

Interestingly, XGBoost placed higher relative importance on features like **Mean\_DFA\_alpha** and **Var\_Wolf\_Lye**, which were less prominent in the Random Forest model. The LIME explanations supported these results, indicating that similar entropy-based features drove predictions, but XGBoost relied more heavily on Hurst exponent metrics.

**Feature Importance Insights**

Both models emphasized the significance of features capturing **long-term dynamics (e.g., Mean\_Long\_Lye)** and **nonlinear system properties (e.g., Mean\_Correlation\_Dimension, Mean\_ApEn)**. These results suggest that skill levels are influenced by complex and dynamic patterns in the data.

**Conclusion**

Random Forest and XGBoost exhibited comparable accuracy and F1-scores, with strong ROC-AUC values (>0.95). Explainability tools (SHAP and LIME) highlighted consistent feature importance trends, indicating the robustness of the selected features in capturing skill-level dynamics. Future efforts could focus on improving the predictive performance for the "Expert" class, potentially through tailored feature engineering or advanced ensemble techniques.