We evaluate how accurate serology tests are by checking whether each test result matches a confirmed gold standard reference. We turn this into a yes/no, where 1 means the test result was correct and 0 means it was incorrect. We build and compare two models: Logistic Regression, which gives us a baseline, and a Random Forest model, which captures more complex patterns. To measure how well the models work, we use ROC and Precision-Recall curves, and we also check calibration to see if the predicted probabilities match the gold standard reference.

To understand what influences the predictions, we look at the most important features for each model. For Logistic Regression, we examine the biggest positive and negative coefficients, which tell us which factors make a correct result more or less likely. For the Random Forest, we review feature importance scores to see which inputs the model relied on most.

ROC Curve

Shows the trade-off between true positive rate and false positive rate.
A higher curve and area under the curve indicate better overall ability to rank correct vs. incorrect test outcomes.
The diagonal reference line represents random performance.

Precision-Recall Curve

Shows the relationship between precision and recall across decision thresholds.
Average Precision shows overall performance across thresholds.

Calibration Curve

Compares predicted probabilities to observed accuracy across probability bins.
 A perfectly calibrated model follows the diagonal line:

- Above diagonal - model is under-confident
- Below diagonal - model is over-confident

Top Logistic Regression Coefficients

- Positive coefficients increase the likelihood of a correct test result
- Negative coefficients decrease the likelihood

<u>Random Forest Feature Importances</u>

Ranks features based on their contribution to decision splits across the forest. Higher values indicate greater overall importance.