# Model Optimization and Tuning Phase Report

During the model development process, the performance of machine learning models was further enhanced through careful optimization and hyperparameter tuning. This phase aimed to select the most suitable model parameters that improve accuracy, reduce overfitting, and ensure generalizability of the model.

## 1. Hyperparameter Tuning Techniques Used

**- Grid Search CV**

Exhaustively searches through a specified subset of hyperparameters using cross-validation.

**- Randomized Search CV**

Randomly samples parameter combinations to find the best settings more quickly than Grid Search.

**- Cross-Validation (CV)**

Divides the dataset into k folds, trains on k-1 and tests on the remaining to ensure robust results.

| Hyperparameter | Description | Final Value |
|---|---|---|
| n_estimators | Number of trees | 100 |
| learning_rate | How quickly the model adapts | 0.1 |
| max_depth | Maximum tree depth to control complexity | 4 |
| subsample | Fraction of samples used per tree | 0.8 |
| colsample_bytree | Fraction of features used per tree | 0.8 |
| gamma | Minimum loss reduction to split a node | 0 |
| reg_lambda | L2 regularization term | 1 |

## 2. Parameters Tuned for Selected Models

### - Random Forest Classifier

Parameters Tuned: n_estimators, max_depth, min_samples_split, criterion

Best Values: 100, 10, 2, 'gini'

### - XGBoost

Parameters Tuned: n_estimators, learning_rate, max_depth, subsample, colsample_bytree

Best Values: 100, 0.1, 5, 0.8, 0.8

### - Gradient Boost Classifier

Parameters Tuned: learning_rate, n_estimators, max_depth

Best Values: 0.05, 150, 4

### - Logistic Regression

Parameters Tuned: C, solver, penalty

Best Values: 1.0, 'liblinear', 'l2'

## 3. Evaluation Metrics

To assess model performance post-tuning, the following metrics were analyzed:

Accuracy, Precision & Recall, F1 Score, ROC-AUC Score, Confusion Matrix

These metrics confirmed that the tuned models performed better with balanced precision and recall, especially important for healthcare predictions like anemia detection.

## 4. Final Selection

Based on the performance metrics after tuning:

- XGBoost, Gradient Boost, and Random Forest emerged as top models with near-perfect scores.

- **Grid Search** was used to explore combinations of hyperparameters.

**5-Fold Cross Validation** ensured performance was robust across different data splits.

## 5. Results After Tuning

- **Accuracy:** 100%
- **Precision / Recall / F1-Score:** All very high (perfect classification)
- **Confusion Matrix:** No false positives or false negatives in test data