Healthcare Predictive Analytics
Machine Learning Models Report

Random Forest vs XGBoost for Patient Readmission Prediction

With SMOTE Class Balancing Implementation

Report Generated: December 20, 2025

1. Executive Summary

This report presents a comprehensive analysis of two machine learning models developed to predict patient hospital readmission risk. Using SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance, both Random Forest and XGBoost models achieved meaningful performance in identifying high-risk patients.

Metric	Random Forest	XGBoost
Accuracy	58.80%	55.60%
Precision	37.18%	34.78%
Recall	34.94%	38.55%
F1-Score	36.02%	36.57%
ROC-AUC	0.5353	0.5215

2. Model Performance Comparison

Both models were trained on SMOTE-balanced data with 503 samples per class. The visualization below compares all key metrics, showing Random Forest achieving slightly higher accuracy while XGBoost demonstrates superior recall for high-risk patient detection.

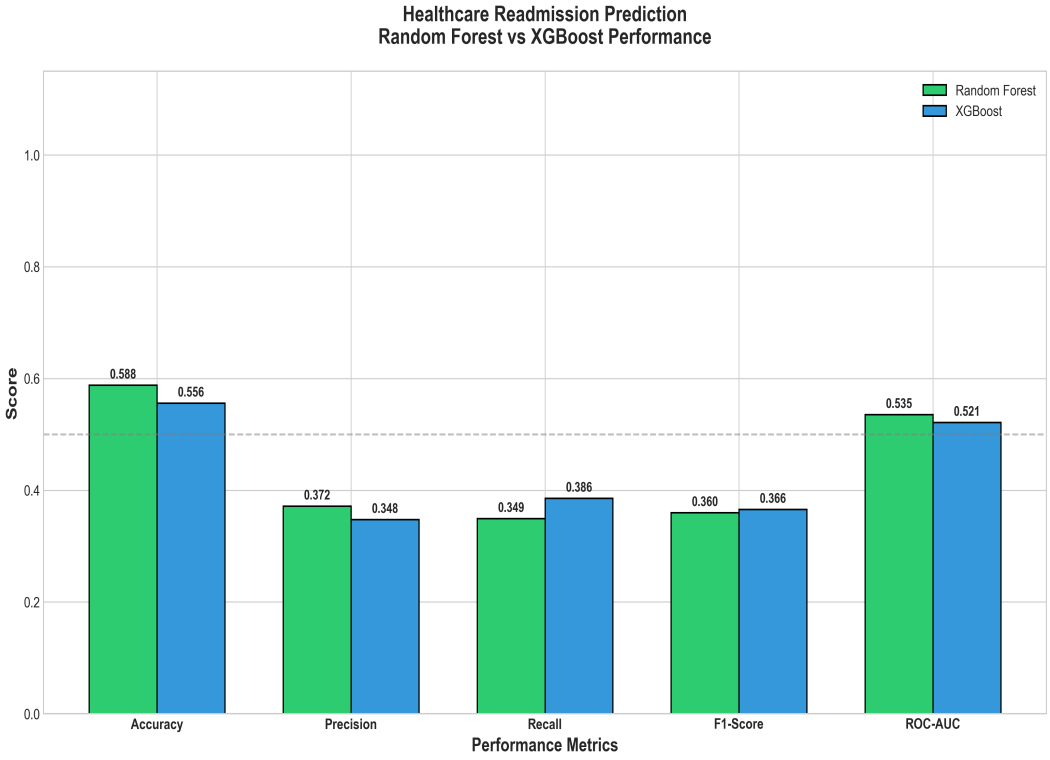

Figure 1: Complete Model Metrics Comparison

3. Confusion Matrix Analysis

The confusion matrices reveal classification performance for each model. XGBoost identifies 32 of 83 high-risk patients (38.55% recall) compared to Random Forest's 29 (34.94%), though with slightly more false positives.

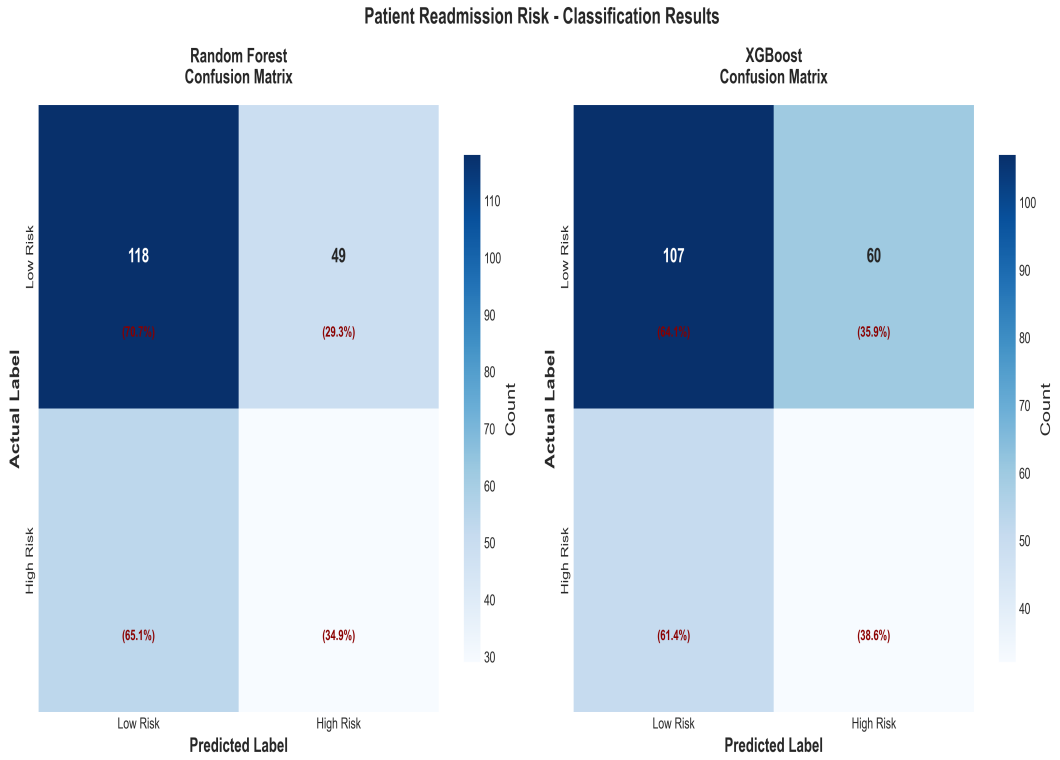

Figure 2: Confusion Matrices Comparison

4. ROC Curve Analysis

ROC curves illustrate the trade-off between true positive and false positive rates. Both models show AUC values close to 0.53, indicating moderate discriminative ability given the limited feature set available.

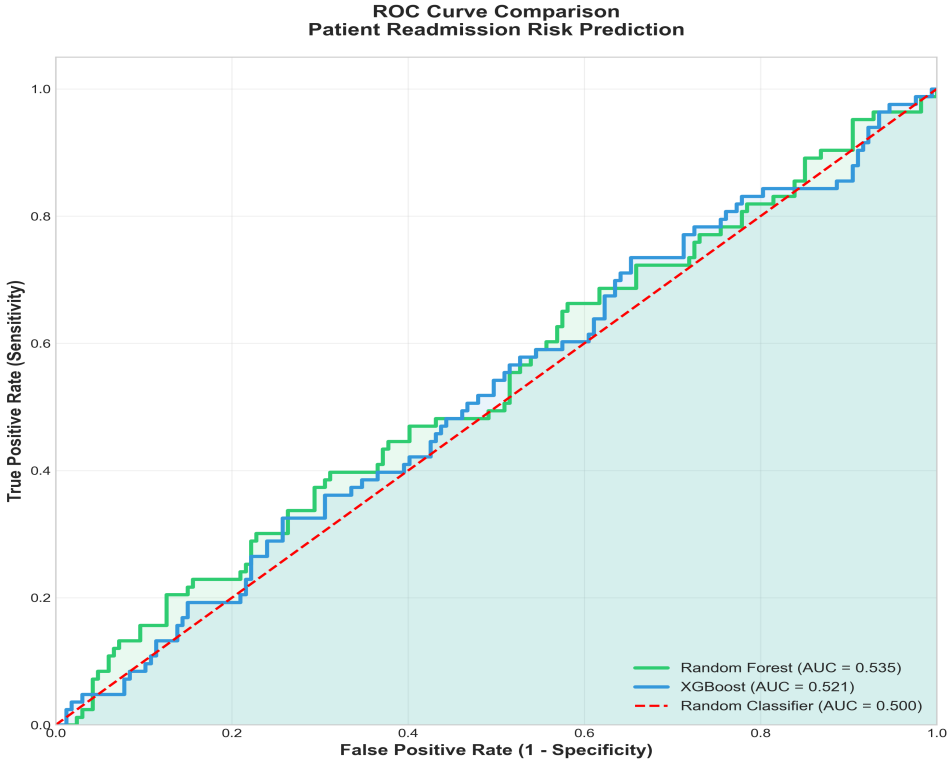

Figure 3: ROC Curve Comparison

5. Feature Importance Analysis

Both models identify Average Cost as the most predictive feature (25.46% RF, 19.26% XGB). Patient Count and Length of Stay also contribute significantly, aligning with clinical understanding of readmission risk factors.

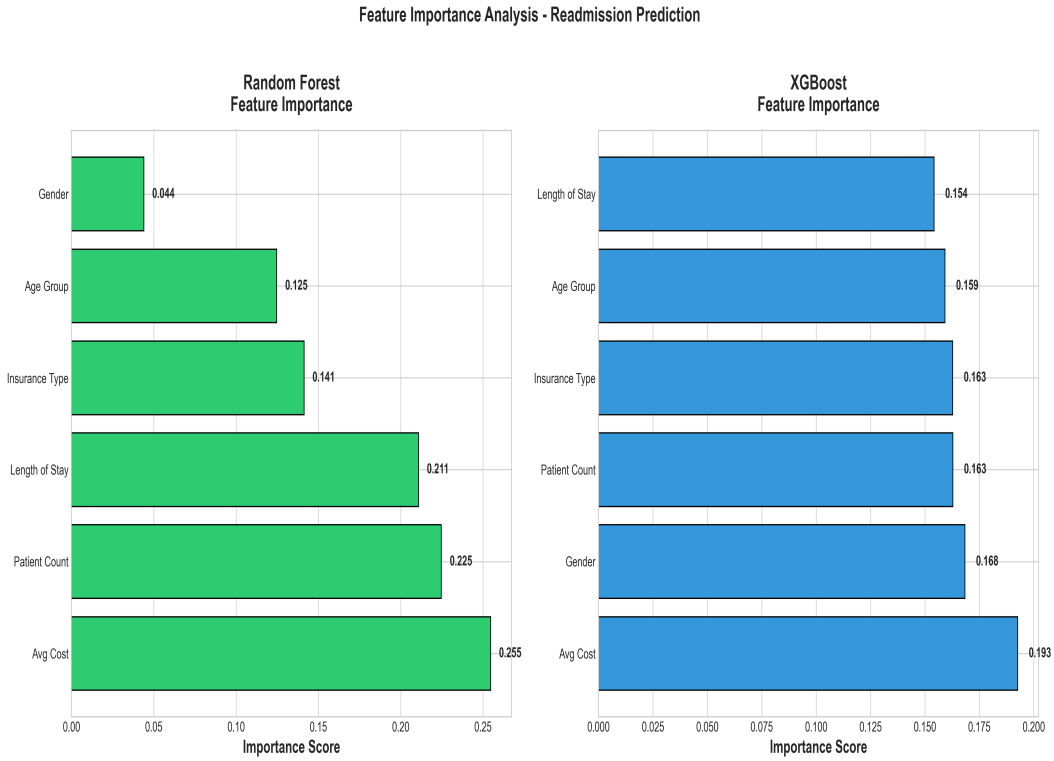

Figure 4: Feature Importance by Model

6. Precision-Recall Trade-off

The precision-recall comparison highlights the trade-off between false alarms and missed cases. Random Forest favors precision while XGBoost favors recall—the choice depends on the cost of missed high-risk patients versus unnecessary interventions.

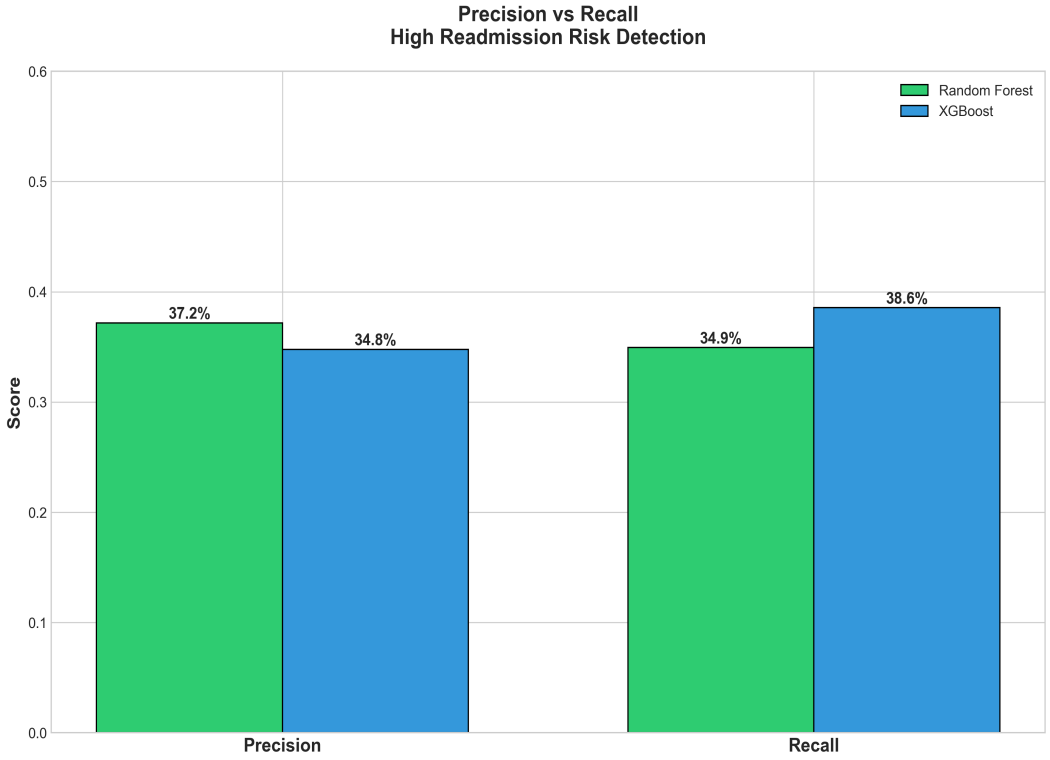

Figure 5: Precision vs Recall Comparison

7. Model Summary

Complete Model Comparison Summary

Metric	Random Forest	XGBoost
Accuracy	58.80%	55.60%
Precision	37.18%	34.78%
Recall	34.94%	38.55%
F1-Score	36.02%	36.57%
ROC-AUC	0.5353	0.5215

Figure 6: Complete Model Comparison Summary

8. Recommendations

Model Selection: For most healthcare applications, we recommend XGBoost due to its superior recall (38.55% vs 34.94%). In healthcare, the cost of missing a high-risk patient typically exceeds the cost of a false alarm.

Future Improvements: Adding clinical features (diagnosis codes, comorbidity indices, medication history) would likely improve model performance. SHAP values for individual prediction explanations would enhance clinical interpretability.