Evaluation of Machine Learning Models for Predicting Postoperative Adverse Events in Conversion and Revision Bariatric Surgery Using MBSAQIP Data

Abdallah Attia, MD; Eman Toraih, MD, PhD; Danielle Tatum, PhD, Shauna Levy, MD, FACS, MS, Dietric Hennings, MD

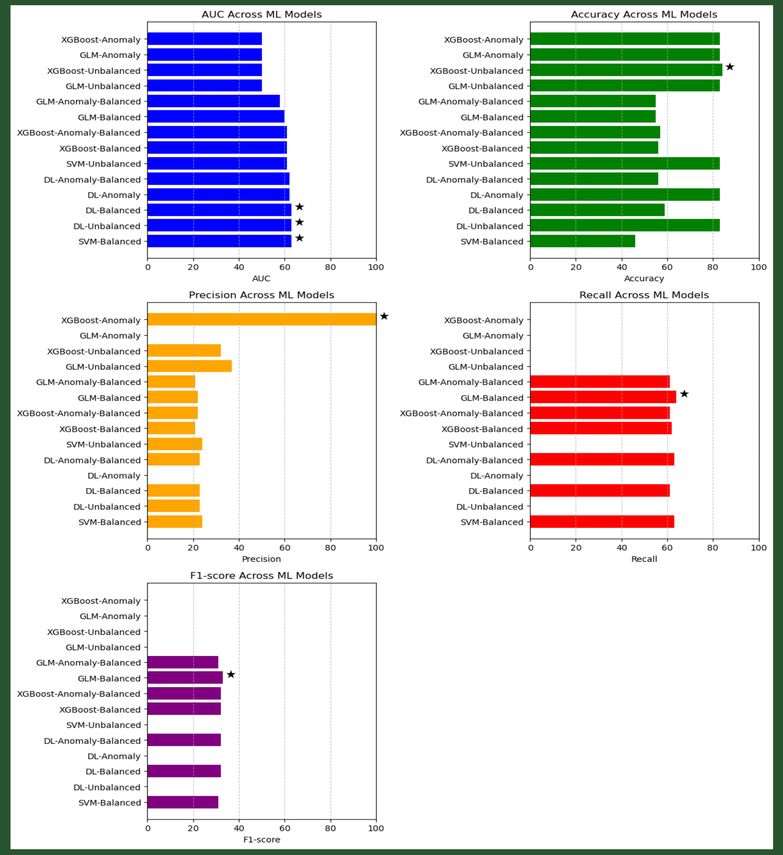
Department of Surgery, Tulane University School of Medicine, New Orleans

Introduction

- Accurate prediction of postoperative complications is essential for patient safety.
- Machine learning (ML) offers advantages over traditional statistical methods.
- Study evaluates ML models for predicting complications after conversion/revision MBS.
- **Hypothesis:** Addressing class imbalance and collinearity will affect model sensitivity and performance.

Methods

- **Dataset:** MBSAQIP (2020-2022)
- **Models Tested:**
  - Generalized Linear Models (GLM)
  - Support Vector Machines (SVM)
  - Extreme Gradient Boosting (XGBoost)
  - Deep Learning (DL) Neural Networks
- **Training Conditions:**
  - **Unbalanced data**
  - **Balanced data** (sampling techniques)
  - **Anomaly detection applied**
- **Performance Metrics:** AUC, accuracy, precision, recall, F1-score



The figure illustrates the **AUC, Accuracy, Precision, Recall, and F1-score** across various machine learning models used to predict postoperative complications in conversion/revision bariatric surgery. Models are **sorted by AUC** for clearer comparison. A **star (★) highlights the highest value** for each performance metric.

Results

- **Sample Size:**
- 68,498 patients; 11,411 (16.7%) had complications.

Impact of Data Balancing:

- Improved AUC across all models.

Unbalanced data:

- Higher accuracy (DL model = 83%) but very low recall (0.2%).

Balanced data:

- Recall improved (DL model = 61%) but accuracy dropped (59%).

- **Key Findings:**
  - Anomaly detection had minimal impact.
  - Pulmonary embolism history = **major risk factor** for complications.
  - History of adjustable gastric banding (AGB) = **lower risk** than Roux-en-Y gastric bypass (RYGB).

Conclusion:

- ML models struggled with prediction due to **class imbalance & collinearity**.
- MBSAQIP dataset **not structured for ML applications**, limiting model effectiveness.
- **High accuracy in unbalanced models was misleading** due to poor sensitivity.

Future Direction:

- Improve data quality by adding key variables.
- Address collinearity to enhance ML model performance.