A Comparative Analysis of Logistic Regression and Machine Learning Models for Predicting Surgical Site Infections in Bariatric Surgery Patients

Juan S. Barajas-Gamboa[1], Gustavo Romero-Velez[2,3], Terrence J. Lee- St. John[1], Andrew T. Strong[2,3], Salvador Navarrete[2,3], Ricard Corcelles[2,3], Carlos Abril[1,3], Juan Pablo Pantoja[1], Alfredo D. Guerron[1], John Rodriguez[1,3], Matthew Kroh[2,3], Jerry T. Dang[2]

[1] Digestive Disease Institute, Cleveland Clinic Abu Dhabi, UAE: [2] Digestive Disease & Surgery Institute, Cleveland Clinic, Ohio, USA: [3] Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio, USA

Background

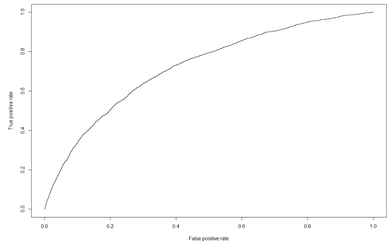
- Surgical site infections (SSI) remain a significant concern in bariatric surgery
- Identifying risk factors for SSI can improve significant morbidity and mortality
- Previous efforts to determine risk factors mainly used logistic regression (LR), limiting the integration of innovative approaches like machine learning (ML)
- The objective was to evaluate and compare SSI prediction models in bariatric surgery using LR methods and ML

Methods

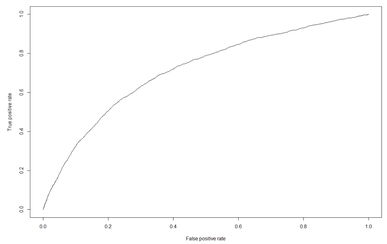
- A cross-sectional study of the MBSAQIP database was conducted from 2015 to 2019
- Patients undergoing sleeve gastrectomy (SG) and Roux-en-Y gastric bypass (RYGB) were included
- A prediction model for SSI was developed using: (1) LR, (2) neural network (NN), and (3) XGBoost model
- A randomized 50/50 split of the complete dataset were used for both training and validation of each model
- Overall model performance was assessed using the area under the receiver operating curve (AUC)

Results

- A total of 750,417 patients were analyzed in this study, of which 0.4% experienced SSI
- Of the total cohort, 79.6% were female with a mean age of 44.4 ± 11.9 years
- The initial mean body mass index (BMI) was 45.2 ± 7.8 kg/m^2
- The most common comorbidities included hypertension 47.7%, diabetes 26.1%, hyperlipidemia 23.6% and tobacco used within 1 year 8.4%
- The AUC for the three different SSI prediction models was: 0.719 for logistic regression, 0.725 for neural network and 0.718 for XGBoost
- Although the AUC suggests a comparable discriminant function, the risk prediction histogram for the neural network displayed a more gradual shift

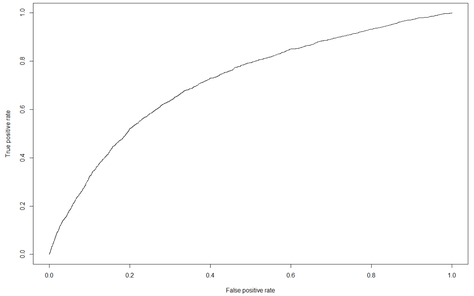


- **Figure 1.** Receiver operating characteristic curve for logistic regression model. AUC of 0.719



- **Figure 2.** Receiver operating characteristic curve for neural networks model. AUC of 0.725

Results



- **Figure 3.** Receiver operating characteristic curve for extreme gradient boosting model. AUC of 0.718

Conclusions

- The ML models demonstrated significant discriminant function in predicting SSI compared with LR
- ML may be used to identify patients at increased risk and to target preventive interventions for SSI after bariatric surgery
- Future research should aim to integrate these predictive models within the clinical practice, so that their real-time efficacy in clinical environments can be assessed

Disclosures

- Authors have no conflicts of interest relevant to this publication