This study leverages data from the National Surgical Quality Improvement Program (NSQIP) spanning the years 2014-2020. A rigorous data cleaning process was employed, involving the removal of empty columns and identification of columns with a substantial proportion of missing values, using a threshold of 60%. Additionally, steps were taken to address potential bias in the dataset.

Feature selection was performed, resulting in the identification of 33 pertinent features, encompassing demographic, procedural, and health-related variables. These features include sex, principal diagnosis, admission type, transfer status, age, discharge destination, anesthesia type, surgical specialty, elective status, height, weight, comorbidities (diabetes, smoking, dyspnea), functional status, ventilation status, and various medical conditions.

The target variable for prediction is ASA classification. In the preprocessing phase, categorical variables were converted into numerical format, and unique values were consolidated into binary categories (e.g., yes/no, in/out). For multi-class variables, the top three most frequent categories were retained, while all others were grouped under an 'others' category.

Encoding techniques, specifically label encoding, were applied to the data to facilitate model training. Six distinct algorithms (Neural Network, Logistic Regression, Stochastic Gradient Descent, Random Forest, Decision Tree, and Ridge Classifier) were deployed to identify the optimal predictive model. Model performance was assessed using key metrics including precision, recall, F1-score, ROC-AUC, and the confusion matrix, ultimately highlighting the superiority of the Ridge Classifier in this predictive task.