http://dx.doi.org/10.2106/JBJS.21.00718

The following content was supplied by the authors as supporting material and has not been copy-edited or verified by JBJS.

Appendix I. Descriptions of basic performance evaluation metrics for machine learning algorithms.

1. Discrimination Analysis: Evaluated through the use of receiver operating characteristic (ROC) analysis and generates C-statistic. The c-statistic is described as the probability that the machine learning model will assign a greater predicted probability to a randomly selected positive case (patient who experienced hyponatremia) relative to a randomly selected negative case (false positive case, i.e., a patient who did not develop hyponatremia).

2. Calibration Analysis: Describes concordance between machine learning predictions and true observed outcomes in the data set. A calibration slope = 1 indicates perfect prediction and represents the precision of predictions. A calibration intercept of zero also indicates perfect prediction and represents the tendency of the predictions to overestimate or underestimate the observed outcome.

3. Brier Score Analysis: The Bier score is equal to the mean squared difference between the model prediction probabilities and the true observed outcomes. This metric is a benchmark measure of performance. Model brier scores lower than the null Brier score (when the probability of a prediction is equal to the prevalence in the population) indicates that predictions are well calibrated (with zero being perfect calibration).

4. Decision-Curve Analysis: Decision-curve analysis is a visual and quantitative method of describing clinical utility and real-world application of use of the machine learning model in practice. This analysis compares changes in management based off of the model, the most predictive variable alone, changes for all patients, and changes for no patients. It then assigns a net decision benefit for varying patient risk thresholds. As the risk threshold increases, the cost-to-benefit ratio (and consequently the weight attributed to false positive classifications made by the model) increases.

Appendix II. Machine learning performance metrics.

A. Algorithm Performance

Performance of the machine learning algorithms (**Table 2**) demonstrated that the c-statistic ranged between 0.65 – 0.75. Calibration intercepts ranged between -0.01 – -0.03, while calibration slope ranged between 0.73 – 1.23. The Brier score ranged between 0.12 – 0.13, which in all cases was below the Null model Brier score. Analysis of relative performance indicated that the SGB algorithm was the best performing algorithm, with a c-statistic = 0.75 (**Figure 3**), calibration intercept = -0.02 and calibration slope = 1.02, and Brier score = 0.12 (**Figure 4**). Decision-curve analysis indicated that this model conferred a greater standardized net benefit to patients in terms of predicting postoperative hyponatremia compared with treatment of no patients, treatment of all patients, and treatment based off of preoperative serum sodium concentration alone (**Figure 5**). The global weighted importance of each variable in the SGB model for making the prediction of hyponatremia is depicted in **Figure 6**.

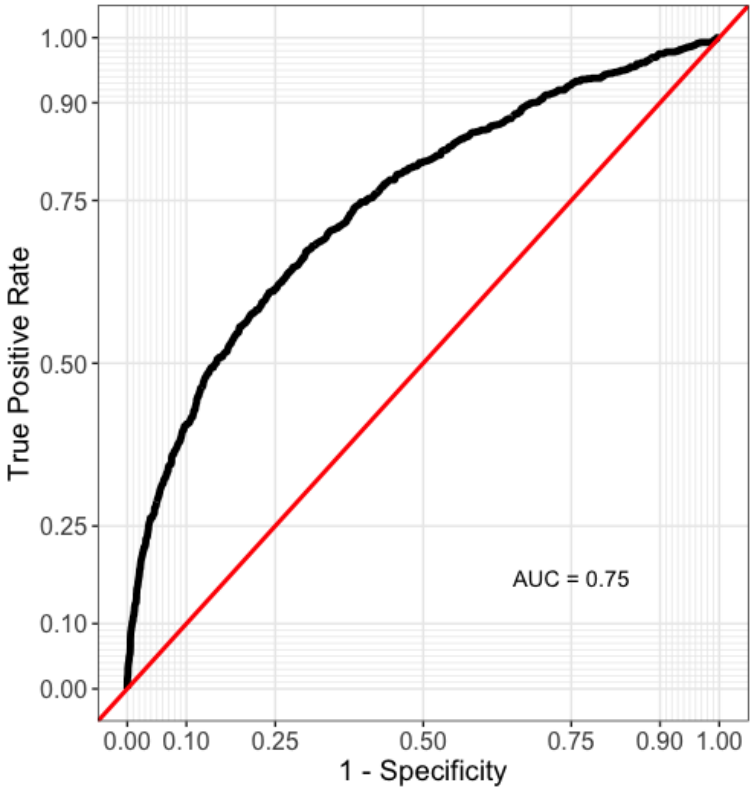
Table 2. Machine learning algorithm performance in independent testing set of patients (95% confidence interval),

n = 6,137.

Metric	Stochastic Gradient Boosting	Random Forest	Support Vector Machine	eXtreme Gradient Boosting	Neural Network	Elastic-Net Penalized Logistic Regression
C-statistic	0.75 (0.73, 0.76)	0.71 (0.69, 0.73)	0.65 (0.63, 0.67)	0.74 (0.73, 0.76)	0.74 (0.72, 0.76)	0.74 (0.72, 0.75)
Calibration intercept	-0.02 (-0.09, 0.06)	-0.01 (-0.08, 0.06)	-0.03 (-0.10, 0.04)	-0.01 (-0.08, 0.06)	-0.01 (-0.08, 0.06)	-0.01 (-0.08, 0.06)
Calibration slope	1.02 (0.94, 1.10)	0.73 (0.67, 0.80)	0.84 (0.74, 0.95)	1.01 (0.93, 1.09)	1.01 (0.93, 1.09)	1.23 (1.13, 1.33)
Brier score	0.12 (0.11, 0.13)	0.13 (0.12, 0.14)	0.13 (0.12, 0.14)	0.12 (0.11, 0.13)	0.13 (0.12, 0.14)	0.13 (0.12, 0.14)

Null model Brier score = 0.18

Figure 3. Receiver operative curve demonstrating discrimination of stochastic gradient boosting algorithm on independent testing set of patients, n = 6,140.



AUC, area under the curve.

Figure 4. Calibration plot for stochastic gradient boosting algorithm on independent testing set of patients, n = 6,140. The calibration slope represents the precision of predictions, while the calibration intercept represents the tendency for the model to overestimate or underestimate the observed outcome.

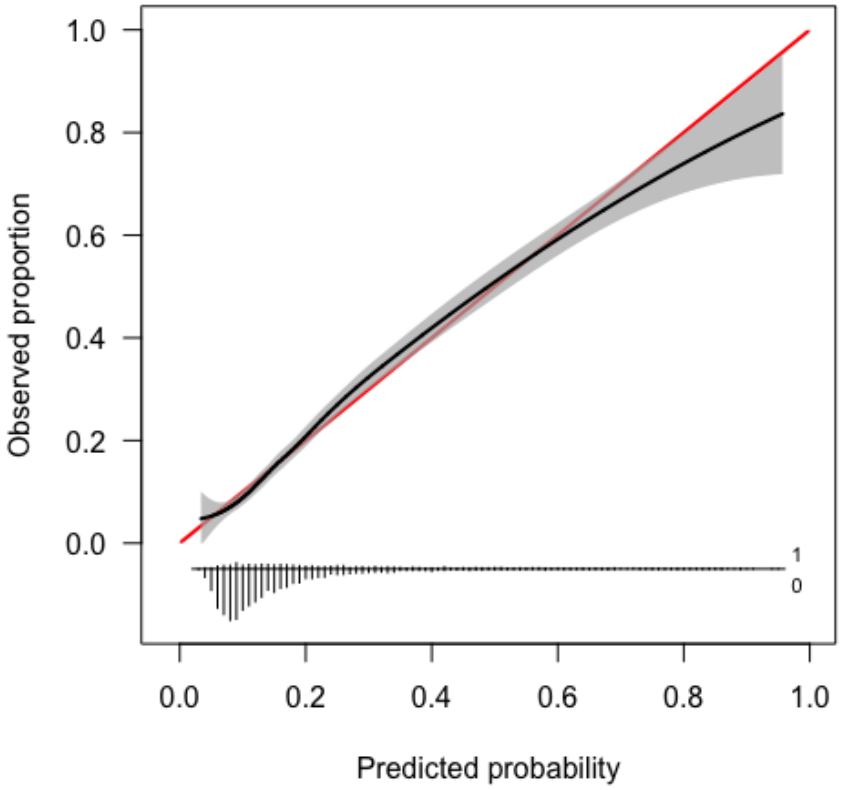
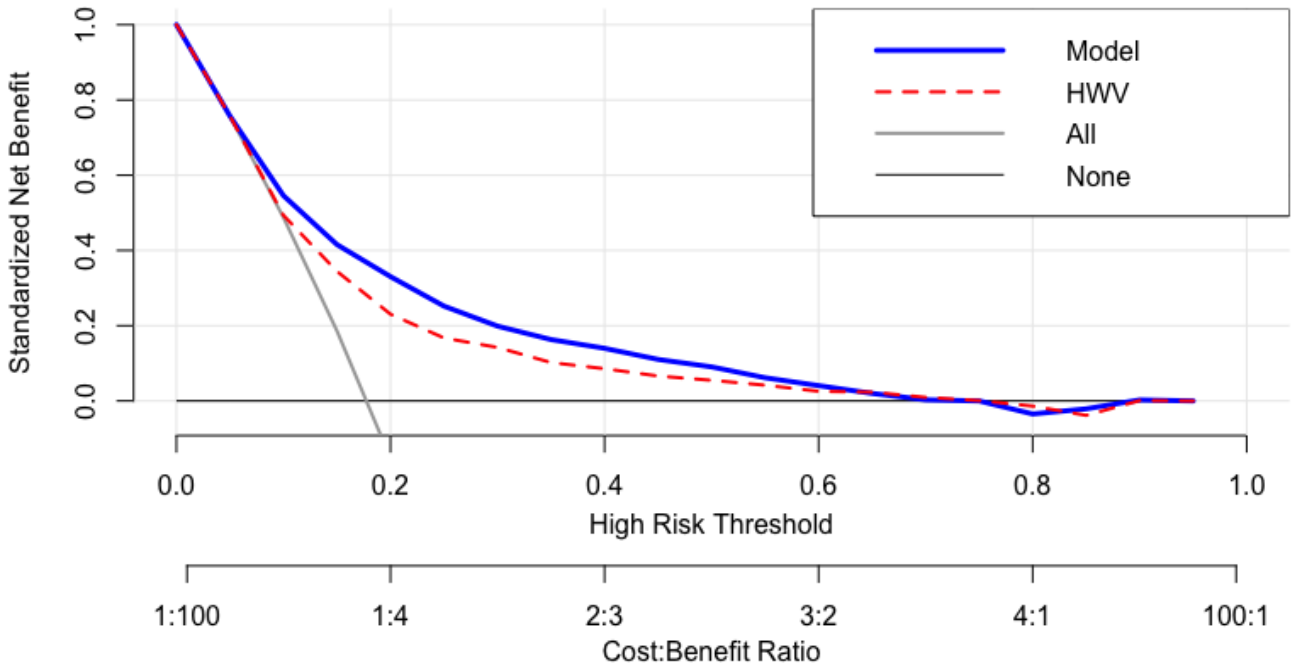


Figure 5. Decision curve analysis of stochastic gradient boosting algorithm. In the decision curve analysis, the net benefit of the model (blue line) relative to default strategies of changing management for patients ("all" for all patients, "none" for no patients, or the highest weighted variable (HWV), preoperative serum sodium level. The ("all") line represents the net benefit from changing management for all patients. The line slopes down because at a threshold of zero, false positives are given no weight relative to true positives; as the threshold increases, false positives gain increased weight relative to true positives and changing management for all patients results in decreasing net benefit. The horizontal line ("none") represents the default strategy of changing management for no patients (net benefit is zero at all thresholds). The relative net benefit of using the current model to predict hyponatremia is demonstrated to be superior to all other options (attempting to predict hyponatremia based off of preoperative serum sodium levels, treating all patients regardless of sodium level, or treating no patients).



http://dx.doi.org/10.2106/JBJS.21.00718

Figure 6. Global variable importance plot, with variables ranked in decreasing order of importance based on weighted contribution to overall prediction of hyponatremia.

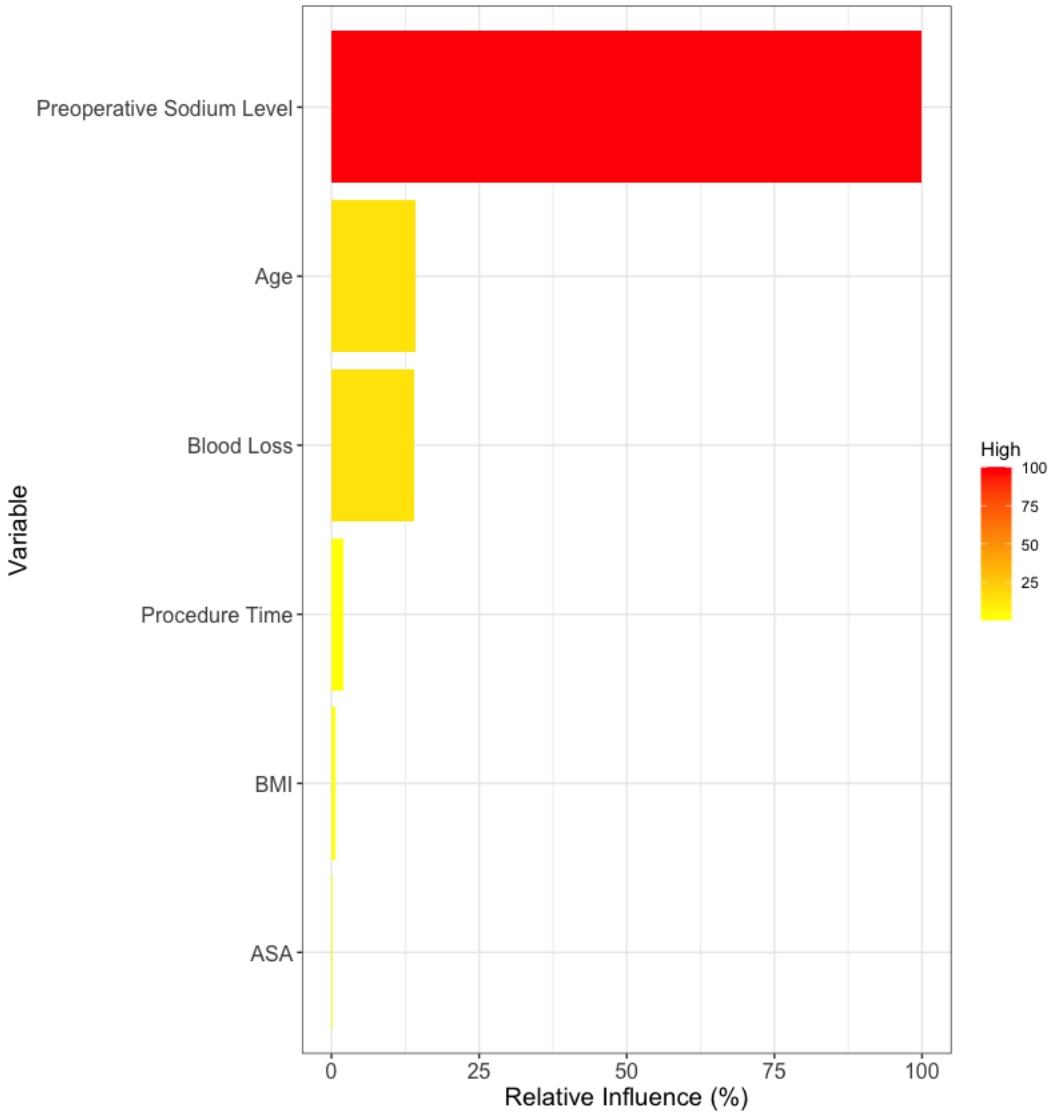
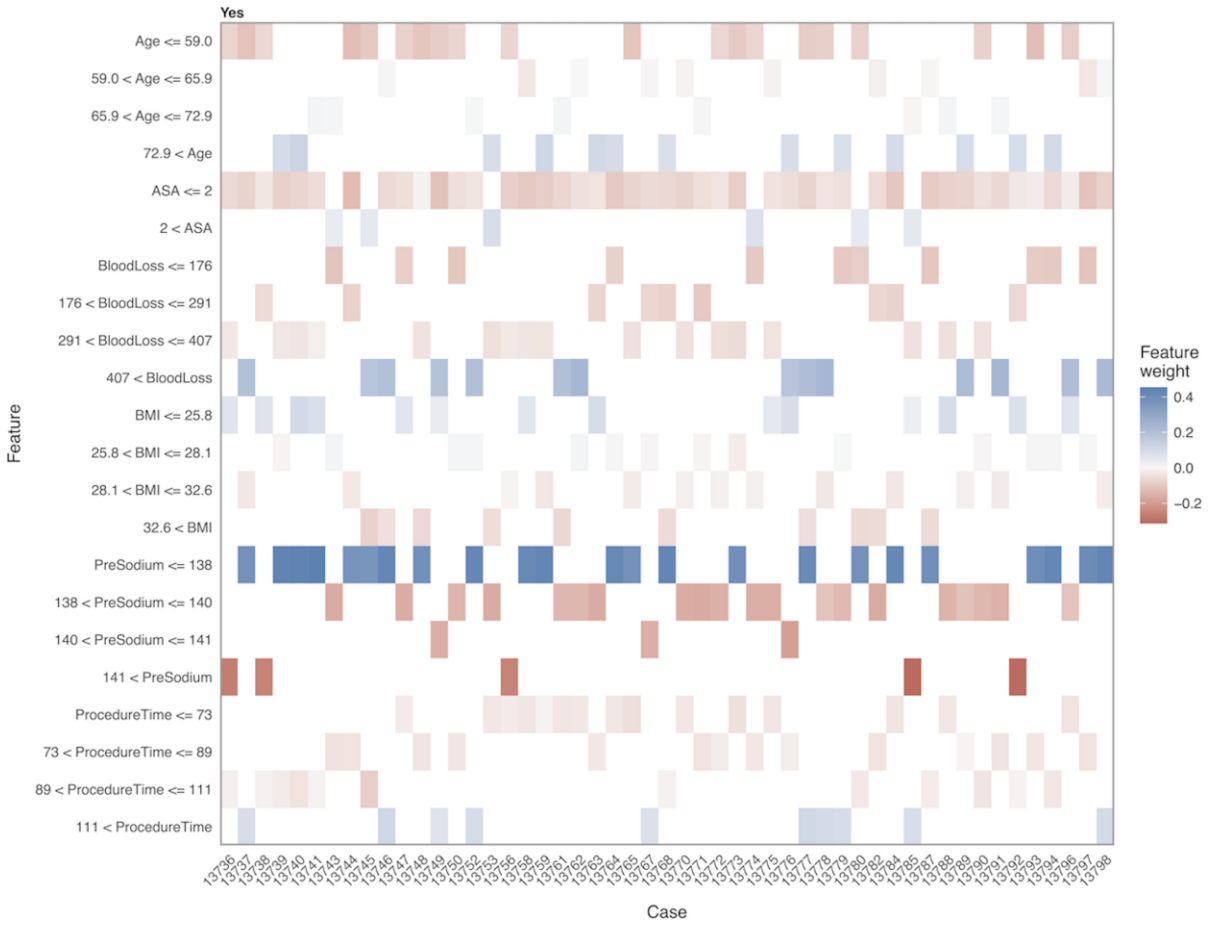


Figure 7. Heat map depicting threshold feature values associated with relative weights (and risk) of developing postoperative hyponatremia. Darker blue boxes indicate feature thresholds that confer greater risk of hyponatremia, while darker red boxes indicate feature thresholds that are protective. Therefore, the darker the shade of blue or red, the stronger the effect a particular variable had in the individual prediction. **Each column indicates an individual patient case, while each row demonstrates how the risk factors identified by the novel algorithms tend to influence patient risk.** This example demonstrates how the SGB model came to a risk prediction in 50 unique cases among the testing population and helps leverage insight into how the model makes individual predictions. Furthermore, this heat map demonstrates how the model can account for individual medical profiles by making specific predictions based on each patient's different numeric combination of these six factors. BMI, body mass index; PreSodium, preoperative serum sodium concentration (mmol/L); ASA, American Society of Anesthesiologists score.



http://dx.doi.org/10.2106/JBJS.21.00718

Appendix III. Patient case example of anticipating hyponatremia risk after total joint arthroplasty. A: This case represents a 62-year-old patient with a body mass index (BMI) of 28 kg/m^2 and American Society of Anesthesiologists (ASA) score of three who undergoes a primary TJA. During preoperative medical clearance, their basic metabolic panel (BMP) revealed a serum sodium concentration of 132 mEq/L. Their TJA lasts 97 minutes where the patient experiences 450mL of blood loss. With this information, the anesthesiologist uses the predictive tool developed here while the patient is still in the operating room to find that the calculated risk of developing hyponatremia is found to be 86.0%. Based on this information, the anesthesiologist orders more frequent postoperative BMPs, switches the patient's fluids from lactated ringers to normal saline, and orders salt tablets prophylactically. They also notify the receiving unit that this patient is at risk of hyponatremia and to be more vigilant for signs and symptoms of hyponatremia in this patient.