

A machine learning approach to predict early outcomes after pituitary adenoma surgery

Todd C. Hollon, MD,¹ Adish Parikh, BS,² Balaji Pandian, BA,² Jamaal Tarpeh, BS,²
Daniel A. Orringer, MD,¹ Ariel L. Barkan, MD,^{1,3} Erin L. McKean, MD,^{1,4} and Stephen E. Sullivan, MD¹

Departments of ¹Neurosurgery, ³Internal Medicine, and ⁴Otolaryngology, and ²School of Medicine, University of Michigan, Ann Arbor, Michigan

OBJECTIVE Pituitary adenomas occur in a heterogeneous patient population with diverse perioperative risk factors, endocrinopathies, and other tumor-related comorbidities. This heterogeneity makes predicting postoperative outcomes challenging when using traditional scoring systems. Modern machine learning algorithms can automatically identify the most predictive risk factors and learn complex risk-factor interactions using training data to build a robust predictive model that can generalize to new patient cohorts. The authors sought to build a predictive model using supervised machine learning to accurately predict early outcomes of pituitary adenoma surgery.

METHODS A retrospective cohort of 400 consecutive pituitary adenoma patients was used. Patient variables/predictive features were limited to common patient characteristics to improve model implementation. Univariate and multivariate odds ratio analysis was performed to identify individual risk factors for common postoperative complications and to compare risk factors with model predictors. The study population was split into 300 training/validation patients and 100 testing patients to train and evaluate four machine learning models using binary classification accuracy for predicting early outcomes.

RESULTS The study included a total of 400 patients. The mean \pm SD patient age was 53.9 ± 16.3 years, 59.8% of patients had nonfunctioning adenomas and 84.7% had macroadenomas, and the mean body mass index (BMI) was 32.6 ± 7.8 (58.0% obesity rate). Multivariate odds ratio analysis demonstrated that age < 40 years was associated with a 2.86 greater odds of postoperative diabetes insipidus and that nonobese patients (BMI < 30) were 2.2 times more likely to develop postoperative hyponatremia. Using broad criteria for a poor early postoperative outcome—major medical and early surgical complications, extended length of stay, emergency department admission, inpatient readmission, and death—31.0% of patients met criteria for a poor early outcome. After model training, a logistic regression model with elastic net (LR-EN) regularization best predicted early postoperative outcomes of pituitary adenoma surgery on the 100-patient testing set—sensitivity 68.0%, specificity 93.3%, overall accuracy 87.0%. The receiver operating characteristic and precision-recall curves for the LR-EN model had areas under the curve of 82.7 and 69.5, respectively. The most important predictive variables were lowest perioperative sodium, age, BMI, highest perioperative sodium, and Cushing's disease.

CONCLUSIONS Early postoperative outcomes of pituitary adenoma surgery can be predicted with 87% accuracy using a machine learning approach. These results provide insight into how predictive modeling using machine learning can be used to improve the perioperative management of pituitary adenoma patients.

<https://thejns.org/doi/abs/10.3171/2018.8.FOCUS18268>

KEYWORDS pituitary adenoma; machine learning; risk stratification; outcome prediction; predictive modeling; obesity

THE ability to predict patient outcomes after a specific treatment is fundamental to providing optimal surgical care. Pituitary adenomas present a unique predictive challenge due to significant heterogeneity among the patient population. This heterogeneity stems from both

the diverse at-risk patient population and the underlying tumor pathophysiology. Pituitary adenomas can occur at any age, with age-adjusted incidence for patients 15–75+ years old ranging from 1.5 to 7.5 tumors per 100,000 people.¹⁷ Endocrinopathies that result from functioning ade-

ABBREVIATIONS AUC = area under the curve; BMI = body mass index; DVT = deep vein thrombosis; LR-EN = logistic regression with elastic net; PE = pulmonary embolism; PR = precision recall; ROC = receiver operating characteristic.

SUBMITTED June 7, 2018. **ACCEPTED** August 27, 2018.

INCLUDE WHEN CITING DOI: 10.3171/2018.8.FOCUS18268.

nomas can produce severe preoperative comorbidity, such as obesity, diabetes mellitus, and cardiomyopathies. Complication rates after transsphenoidal surgery for Cushing's disease range up to 42%.¹⁹ However, nonfunctioning adenomas are more likely to present in older patients, who may have multiple chronic medical conditions that can increase perioperative surgical risk.⁸ The clinical diversity of pituitary adenoma patients makes it challenging to use traditional biostatistical techniques or scoring systems to stratify surgical risk or predict postoperative outcomes given that specific patient characteristics (e.g., tumor type, age, and body mass index [BMI]) are likely to vary in predictive importance across the entire patient population.

Advances in applied predictive modeling using machine learning have provided a novel method for predicting outcomes in healthcare.⁵ Machine learning models have an advantage over other predictive methods because machine learning enables a predictive computer model to automatically learn the best predictive features present in training data. As opposed to the use of a human operator to manually identify these features, which is time and labor intensive, machine learning models can automatically identify the most robust predictive features and can potentially generalize this information to new patient cohorts. Previous studies have used these methods to predict outcomes of stereotactic radiosurgery for brain metastases¹⁵ and arteriovenous malformations,¹⁶ stratify cardiovascular risk,¹⁴ predict mortality/readmission/length of stay,¹ and make cancer prognoses.²⁶

To improve the perioperative management and risk stratification of pituitary adenoma patients, we aimed to predict early outcomes of pituitary adenoma surgery using a machine learning approach. By analyzing a large cohort of pituitary adenoma patients treated at a tertiary care center, we sought to develop an accurate predictive model built via modern machine learning methods that will identify patients at high risk for poor early outcomes after pituitary adenoma surgery.

Methods

Study Design

We designed a retrospective analysis of 400 consecutive pituitary adenoma patients treated with surgical resection via an endoscopic endonasal approach by the senior authors (E.L.M., S.E.S.). After IRB approval, two independent reviewers completed a systematic chart review using a standardized database template. In addition to formal chart review, the University of Michigan Electronic Medical Record Search Engine (UM-EMERSE)⁹ was used to confirm patient details and/or any discrepancy between reviewers.

The study aims were to 1) perform exploratory data analysis of a large series of pituitary adenoma patients treated at a high-volume medical center with an integrated neuroendocrine center and 2) develop and validate a supervised machine learning model that can predict early postoperative outcomes. We defined a poor early postoperative outcome using broad and inclusive criteria, which included the following: 1) major adverse medical event within 30 days of surgery (including deep vein thrombosis [DVT]/

pulmonary embolism [PE], myocardial infarction, severe arrhythmia, or stroke), 2) early surgical complication (CSF leak with or without symptomatic pneumocephalus or postoperative meningitis), 3) expected length of stay (2 days for non-Cushing's disease, 4 days for Cushing's disease) exceeded by 2 days, or any of the following within 30 days of surgery: 4) emergency department admission, 5) inpatient admission, or 6) death. Extended length of stay for nonmedical reasons (transportation, rehabilitation bed availability, social reasons, etc.) was not included as a poor outcome. Because these outcomes can be overlapping (e.g., extended hospital stay due to PE), patients who experienced any or all of these outcomes were assigned to the poor outcome group using a binary classification. Sodium dysregulation (diabetes insipidus or hyponatremia) itself was not considered a poor early postoperative outcome, as it can often be managed effectively in outpatients without complication (i.e., unrestricted free water intake for diabetes insipidus or fluid restriction for hyponatremia). Patients with sodium dysregulation that resulted in unanticipated postoperative care, such as extended length of stay or readmission, were included as poor postoperative outcomes. Patient characteristics/model predictors were established prior to initiating chart review. To improve future model implementation, model predictors were chosen to include only standard clinical information common to all pituitary adenoma patients. Disease-specific characteristics (e.g., preoperative adrenocorticotropic hormone [ACTH] levels) and advanced radiographic features (e.g., Knosp score) were avoided to eliminate missing/not applicable data values and data sparsity.

Descriptive Statistics and Data Exploration

All patient characteristics and outcomes were divided into continuous or nonordered categorical variables for statistical analysis. In addition to the poor early postoperative outcomes defined above, risk factors for common postoperative complications after pituitary adenoma surgery were also explored. Using the full 400-patient dataset, pairwise odds ratio analysis was performed to explore risk factors for diabetes insipidus, hyponatremia, transient cranial nerve palsy, cerebrospinal fluid leaks, symptomatic pneumocephalus, DVT/PE, and postoperative meningitis. For univariate analysis, continuous variables were converted to an indicator variable using binary encoding (e.g., age > 40 years, BMI > 30) to allow for odds ratio calculation. Univariate statistical significance was calculated using Fisher exact testing and defined as $p < 0.05$. Multivariate logistic regression was done for postoperative complications with multiple statistically significant predictors to account for covariance among variables. The R Environment for Statistical Computing (version 3.3.1; <http://www.r-project.org>) and Python-based SciPy library (version 0.19.1, <https://www.scipy.org>) were used for statistical analysis.

Supervised Machine Learning

Four supervised machine learning algorithms were trained and tested as binary classifiers to predict early postoperative outcomes in pituitary adenoma patients:

naïve Bayes, logistic regression with elastic net (LR-EN) regularization (linearly combined L_1 and L_2 regularization penalties), support vector machines with linear kernel, and random forest. These methods were selected for algorithm diversity (i.e., Bayesian model, generalized linear model, margin classifier, and decision trees). Twenty-six patient characteristics were used as predictive variables. Model hyperparameters were selected using a grid search, and 10-fold cross-validation was performed for each model. The training/cross-validation set and testing set were selected by random sampling without replacement from the full 400-patient study population using a 75%/25% (300/100 patient) split. To improve clinical relevance and allow patient risk to be recalculated in the perioperative setting (“rolling” risk assessment), perioperative lowest and highest sodium levels were used as predictors. Data preprocessing included rescaling continuous variables to between 0 and 1. Model training and performance was evaluated using prediction accuracy: model accuracy = (true positives + true negatives)/(true positives + false positives + true negatives + false negatives).

To further evaluate the models, both receiver operating characteristic (ROC) and precision-recall (PR) curves were generated, and area under the curve (AUC) was calculated. To determine the best-performing model, McNemar’s test was used to evaluate marginal homogeneity and determine statistically significant differences between model predictions. Variable importance for the best-performing model is reported to improve model interpretability and assessment of clinical relevance. The R “caret” package (<http://caret.r-forge.r-project.org>) was used for model training, hyperparameter search, validation, and testing. The R and Python code can be downloaded at https://github.com/toddhollon/pituitary_ml.

Results

Patient Population and Early Postoperative Outcomes

The mean age of the study population was 53.9 ± 16.3 years, ranging from 13 to 91 years, and 54% were male. Caucasians made up 84% of patients and blacks 10%. Nonfunctioning pituitary tumors were the most common (59.8%) followed by growth hormone–secreting adenomas (22.8%) and ACTH-secreting adenomas (13.0%). Previous treatment with transsphenoidal surgery or radiation therapy had been performed in 16.5% and 4.0% of patients, respectively. A listing of patient characteristic can be found in Table 1. Differences in sex, tumor size, age, and BMI with respect to tumor type are shown in Fig. 1.

Sodium dysregulation was the most common complication after pituitary adenoma surgery (Table 2). Diabetes insipidus and hyponatremia occurred in 14.8% and 14.3% of patients, respectively. Prevalence of cerebrospinal fluid leak was 7%, and 2% of patients developed symptomatic pneumocephalus. Acute DVT/PE was found in 1.5% of patients, and 1.3% developed postoperative meningitis. Extended length of stay occurred in 20.7% of non-Cushing’s disease patients and 30.8% of Cushing’s disease patients. Thirty-day emergency department admission and subsequent inpatient readmission occurred in 17.0% and 11.8% of patients, respectively. Thirty-day mortality rate

TABLE 1. Preoperative patient characteristics

Characteristic	Value
Age in yrs	53.9 ± 16.3 (13–91)
Male	219 (54%)
Race	
White	336 (84%)
Black	40 (10%)
Other	24 (6%)
Tumor type	
Nonfunctioning	239 (59.8%)
Acromegaly	91 (22.8%)
Cushing’s disease	52 (13.0%)
Prolactinoma	16 (4.0%)
TSHoma	2 (0.5%)
Tumor size	
Macroadenoma	339 (84.7%)
BMI in kg/m ²	32.6 ± 7.8 (19.4–69.7)
Previous TS	66 (16.5%)
Previous skull base radiation	17 (4%)
Preoperative visual deficit	179 (44.8%)
Diabetes mellitus, type II	91 (22.8%)
Heart disease*	34 (8.5%)
Pulmonary disease	26 (6.5%)
Liver disease	17 (4.3%)
Renal disease	7 (1.8%)
Preop antiplatelet/anticoagulant	115 (28.8%)

TS = transsphenoidal surgery; TSHoma = thyroid-stimulating hormone–secreting tumor.

Values are presented as mean ± SD (range) or number of patients (%).

* Includes congestive heart failure, ischemic cardiomyopathy, history of myocardial infarction, arrhythmias.

was 1% (4/400). Based on the study defined criteria, 31% (124/400) of patients had a poor early postoperative outcome, with the top four inclusion criteria being emergency department admission, extended length of stay, inpatient readmission, and CSF leak (Fig. 2, left). A single inclusion outcome occurred in 13% (52/400) of patients, while 18% (72/400) had two or more (Fig. 2, right).

Data Exploration and Odds Ratio Analysis

To explore risk factors for specific complications after pituitary adenoma surgery, we performed a pairwise univariate odds ratio analysis of patient characteristics and comorbidities (Fig. 3). Diabetes insipidus was associated with age < 40 years, Cushing’s disease, microadenomas, and no history of anticoagulation/antiplatelet use. On multivariate logistic regression, age was the only predictor that remained statistically significant, with patients younger than 40 years having 2.86 greater odds of postoperative diabetes insipidus (95% CI 1.52–5.27, $p = 0.001$). Patients with microadenomas were 1.9 times more likely to develop diabetes insipidus; however, this trend did not reach statistical significance on multivariate regression (OR 1.93, 95% CI 0.91–3.98, $p = 0.076$). A relationship

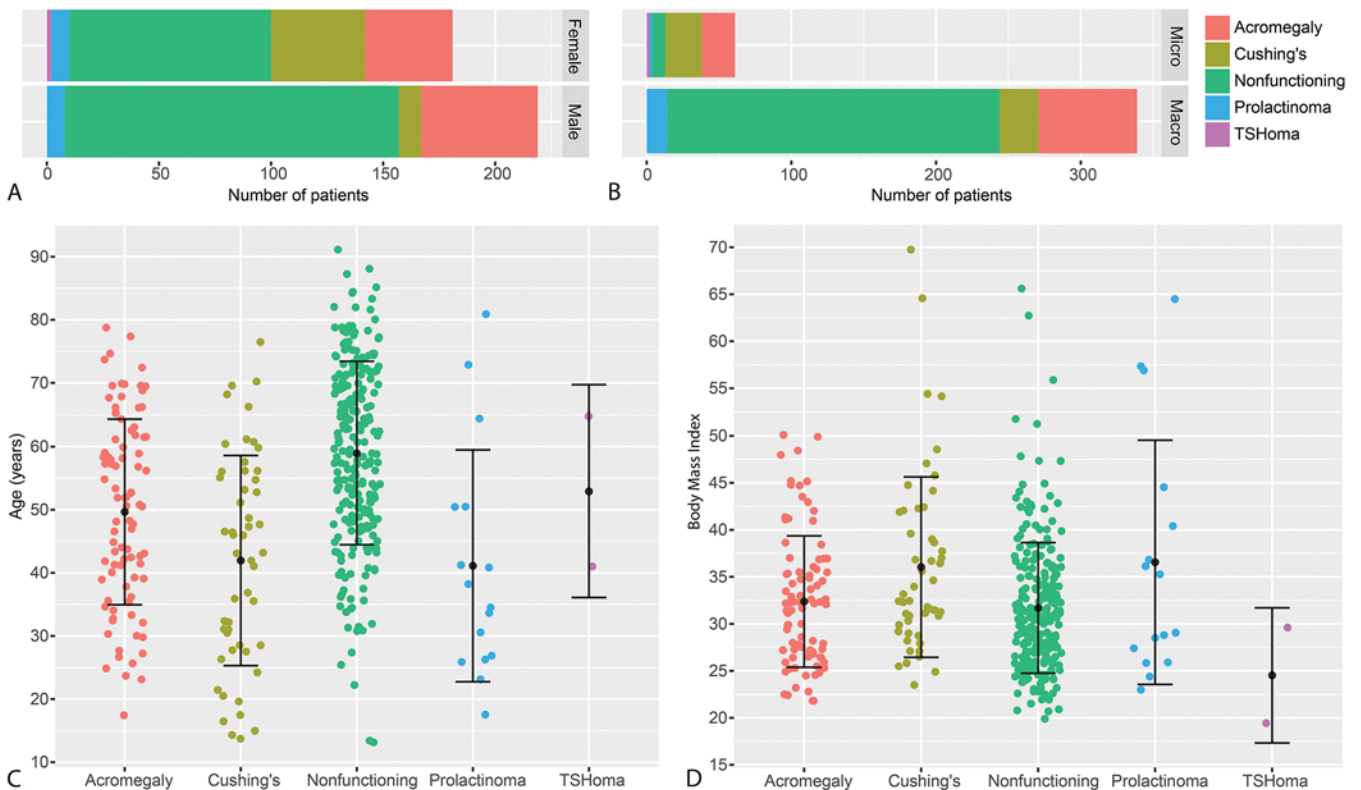


FIG. 1. Patient characteristics by pituitary adenoma diagnosis. **A:** Nonfunctioning adenomas (62.3%) and acromegaly (63.4%) were more common in male patients. Cushing's disease was more common in female patients (80.7%). **B:** Macroadenomas were more common in our study population (84.8%) and the majority were nonfunctioning adenomas (67.8%). Cushing's disease had almost equal distribution between microadenomas (51.9%) and macroadenomas (48.1%). **C:** Mean age of patients with nonfunctioning adenomas was 58.9 ± 14.4 years and was significantly greater than the age of patients with functioning adenomas (mean 46.3 ± 16.1 years, $p < 0.000$). **D:** Prolactinoma patients had the greatest BMI (36.5 ± 13.0), followed by Cushing's disease patients (36.0 ± 9.6) and acromegaly patients (32.4 ± 7.0). TSHoma = thyroid-stimulating hormone–secreting tumor.

was identified between age and tumor size, with patients older than 40 years having a 2.02 (95% CI 1.08–3.69) greater odds of being diagnosed with a macroadenoma ($p = 0.021$).

Obesity was inversely correlated with postoperative hyponatremia on multivariate analysis (OR 0.46, 95% CI 0.25–0.82, $p = 0.009$), and a clinically significant trend toward older patients being more likely to develop hyponatremia was observed (OR 2.48, 95% CI 1.03–7.00, $p = 0.058$). History of skull base radiation was associated with postoperative symptomatic pneumocephalus (OR 8.6, 95% CI 1.1–42.9, $p = 0.040$), and recurrent pituitary adenomas/previous resection was associated with postoperative meningitis (OR 7.7, 95% CI 1.1–67.0, $p = 0.03$). Cushing's disease (OR 12.2, 95% CI 2.2–92.3, $p = 0.006$) and a history of congestive heart failure (OR 7.8, 95% CI 0.91–49.8, $p = 0.04$) significantly increased the odds of DVT/PE on both univariate and multivariate logistic regression. Multivariate analysis included preoperative antiplatelet/anticoagulant use to account for perioperative cessation of medications. Of the 4 patients who died within 30 days of surgery, 3 had Cushing's disease ($p = 0.008$). To further explore the relationships among age, BMI, and sodium dysregulation, the distribution of post-

operative sodium values was plotted with respect to age and BMI (Fig. 4).

Predicting Early Postoperative Outcomes Using Machine Learning

After training and cross-validation of the four machine learning models, they were tested on an independent testing set of 100 patients. Performance data of each model can be found in Table 3. The LR-EN model achieved the highest accuracy at 87.0% (95% CI 78.8–92.9; optimized hyperparameters: $\alpha = 0.05$, $\lambda = 0.005$), followed by the random forest model (85.0%, 95% CI 76.5–91.4; optimized hyperparameter: $mtry = 7$). A significant improvement in model sensitivity was noted for LR-EN and random forest over naïve Bayes classifier and support vector machines. A statistically significant difference in model prediction accuracy was found between LR-EN versus support vector machines and naïve Bayes, but not random forest. Areas under the ROC and PR curves are presented in Table 3. The LR-EN model had the largest AUC-PR (69.5%) and second largest AUC-ROC (82.7%).

ROC and PR curves for each model are presented in Fig. 5A and B. To better understand the output prediction

TABLE 2. Summary of early postoperative complications and outcomes

Complication/Outcome	Value
Lowest postop sodium in mEq/L	138.1 ± 4.9
Highest postop sodium in mEq/L	141.9 ± 3.7
Diabetes insipidus*	59 (14.8%)
Diabetes insipidus requiring desmopressin	40 (10.0%)
Hyponatremia (Na <135 mEq/L)	54 (14.3%)
CSF leak	28 (7.0%)
Symptomatic pneumocephalus	8 (2.0%)
DVT/PE	6 (1.5%)
Transient diplopia/cranial nerve palsy	5 (1.3%)
Meningitis	5 (1.3%)
Extended length of stay	
Non-Cushing's disease	72/348 (20.7%)
Cushing's disease	16/52 (30.8%)
Emergency department admission	68 (17.0%)
Inpatient readmission	47 (11.8%)
Death	4 (1.0%)
Poor early postop outcome (per study criteria)	124 (31.0%)

Values are presented as mean ± SD or number of patients (%).

* Diagnosis of diabetes insipidus was made on the clinical basis of urine output and urine-specific gravity. No absolute serum sodium value was used as a threshold.

probabilities from the LR-EN classifier, the probability of a poor early postoperative outcome for each test set patient is shown in Fig. 5C. The majority of patients who did not have a poor outcome had a low prediction probability (mean 0.201 ± 0.189). The LR-EN classifier correctly identified 17/25 (68%) of patients who did have a poor early postoperative outcome and reflects the improvement in LR-EN sensitivity compared with that of the other trained models. The top six most important predictive variables are shown in Fig. 5D. Lowest perioperative sodium level was the most important predictor, followed by patient age and BMI. These findings are concordant with the calculated odds ratios and the relationships identified above among age, BMI, and sodium dysregulation.

Discussion

Our findings demonstrate that early outcomes of pituitary adenoma surgery can be accurately predicted using a machine learning approach. Using the full patient cohort, we were first able to identify risk factors for common postoperative complications, including diabetes insipidus, hyponatremia, and DVT/PE, using univariate and multivariate odds ratio analysis. By using a large cohort of pituitary adenoma patients to train a machine learning classifier, we were then able to identify patients at high risk for poor postoperative outcomes with an accuracy of 87% and AUC of 83% on ROC analysis on a 100-patient testing set. We identified sodium dysregulation, age, obesity, Cushing's disease, and sex as the most predictive features for stratifying a patient's risk of a poor postoperative outcome. These results provide insight into how predictive

modeling using a machine learning approach can improve the surgical management of pituitary tumors.

A major motivation for the study resulted from the high prevalence of pituitary adenomas among central nervous system tumors, coupled with the lack of any system to meaningfully predict postoperative outcomes. Pituitary adenomas represent approximately 16% of all newly diagnosed brain tumors and are among the top three most common primary central nervous system tumors in the United States.¹⁷ Moreover, they are the second most common nonmalignant brain tumor with surgical resection as a potential curative treatment. While scoring systems have been developed that use radiographic features to classify invasion into adjacent structures^{4,10,11} and hormone levels to predict treatment response,^{2,25} no scoring system has been developed to comprehensively include patient characteristics and stratify surgical risk. Such scoring systems have been developed for meningiomas,²² gliomas (both low-grade^{3,20} and malignant^{13,18}), brain metastases,^{6,7} and arteriovenous malformations^{12,23,24} to predict both early and long-term outcomes. These scoring systems help to determine indications for surgery and improve patient counseling, intraoperative decision-making, and postoperative management.

While scoring systems can apply well to homogenous patient populations, such as those seen in glioblastoma, they are not well suited for the clinical heterogeneity found in pituitary adenoma patient populations. Unlike gliomas and meningiomas, pituitary adenomas are unique among brain tumors in that the presence of the tumor can result in severe systemic illness due to the stimulation or suppression of a neuroendocrine axis. As a result, perioperative risk can stem both from tumor morphology and from secondary systemic comorbidities, rather than lesion morphology alone (e.g., eloquent tumor location in gliomas and deep venous drainage in arteriovenous malformations). The complex interplay between tumor morphology, patient characteristics, and secondary comorbidities associated with endocrinopathies necessitates a more robust method for applied predictive modeling. Machine learning methods offer the opportunity to improve predictive accuracy by learning the complex interactions among risk factors.

The application of machine learning techniques to healthcare has increased over the last 5 years, mainly due to larger datasets, electronic medical records, and better application programming interfaces.^{5,21} Leveraging these aforementioned tools, we were able to build a machine learning classifier that captured the complex risk factor interactions of pituitary adenoma patients and provide accurate predictions of early postoperative outcomes. Via the odds ratio analysis and model feature importance, one complex interaction that we identified was that among age, BMI, tumor size, and postoperative sodium dysregulation. For example, we found that younger age (< 40 years), microadenomas, and Cushing's disease were associated with postoperative diabetes insipidus. The underlying mechanism for this is unclear but may be related to microadenomas and Cushing's disease presenting in younger patients, and resection of these microadenomas can require more pituitary gland manipulation, and subsequent diabetes

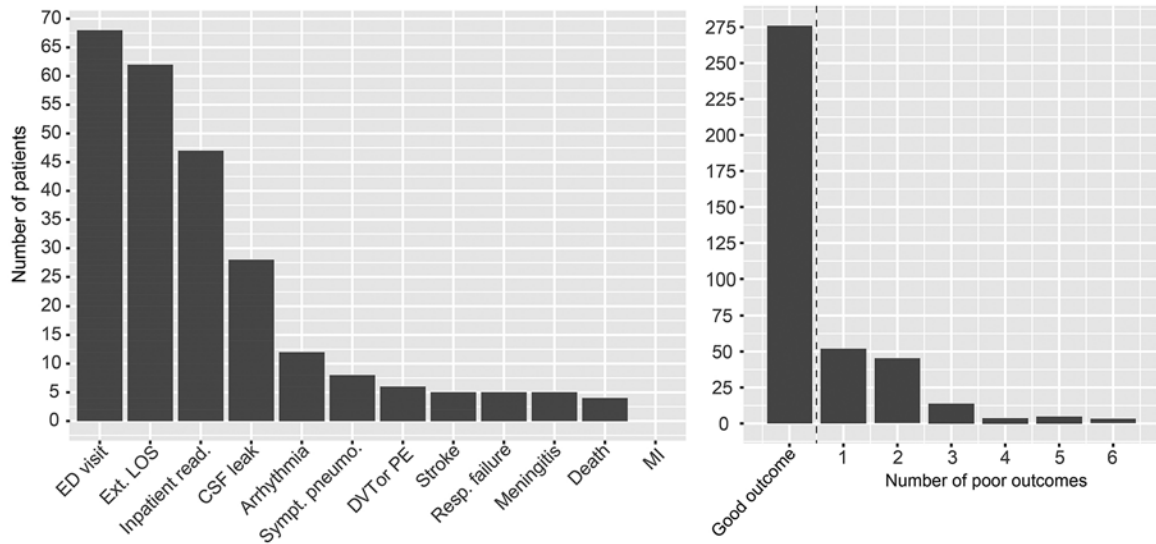


FIG. 2. Study-defined early postoperative outcomes. **Left:** Distribution of inclusion criteria met for the study-defined early postoperative outcomes across the study population (n = 400). The most common criteria met for poor early postoperative outcome were emergency department (ED) admission, extended length of stay (Ext. LOS), inpatient readmission (read.), and CSF leak. No patient suffered myocardial infarction (MI), and 4 patients died within 30 days of surgery. Resp. = respiratory; Sympt. pneumon. = symptomatic pneumonia. **Right:** Good early postoperative outcome occurred in the majority of patients (69%, 276/400). A single inclusion criterion was met in 13% (52/400) of patients and 2 or more criteria were met in 18% (72/400) of patients.

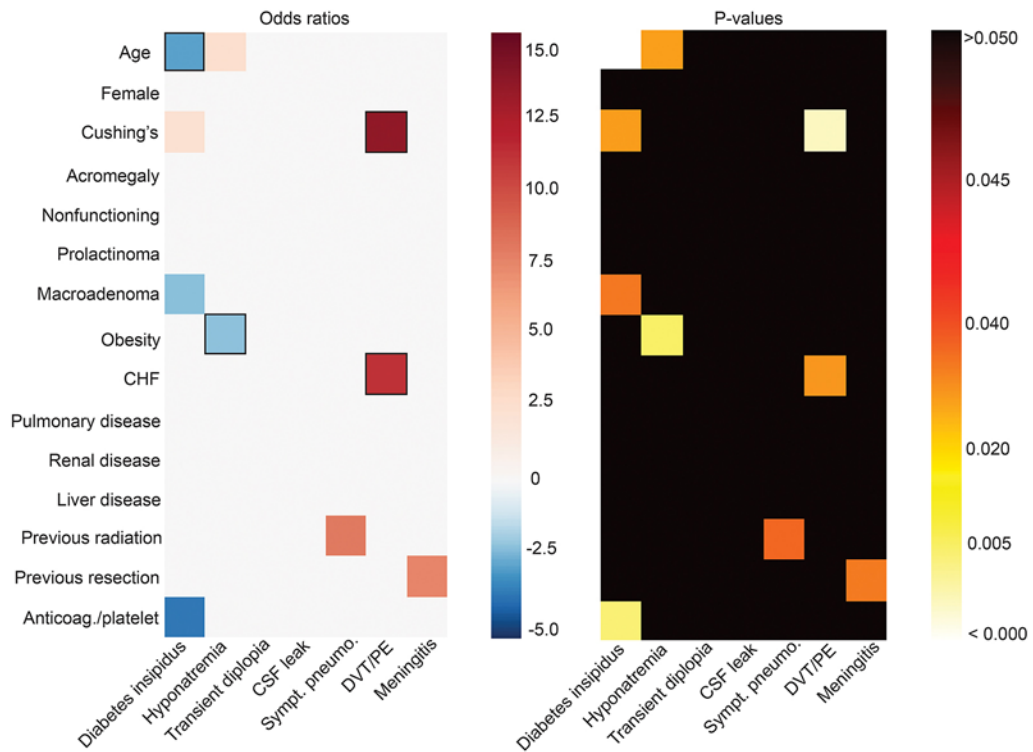


FIG. 3. Univariate and multivariate odds ratio analysis. Odds ratios (**left**) and p values (**right**) are presented as tiled heat maps comparing patient characteristics with early complications. Odds ratio values are color coded such that red indicates a patient characteristic as a risk factor and blue indicates a protective factor for a given outcome. Black boxes identify patient characteristic–outcome pairs that remained statistically significant on multivariate analysis. p values are presented as a single continuous variable on a logarithmic scale. Solid black squares are non–statistically significant comparisons. Anticoag. = anticoagulant use; CHF = congestive heart failure.

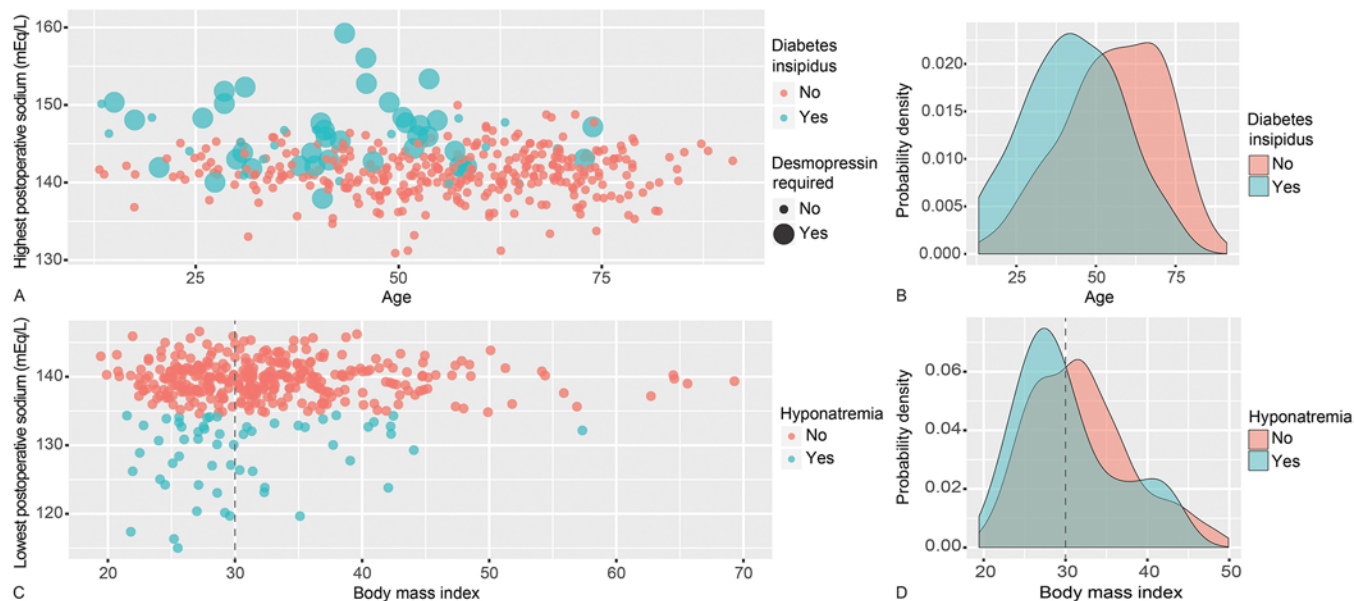


FIG. 4. Early postoperative sodium dysregulation. **A:** Scatter plot showing the distribution of highest postoperative sodium levels with respect to patient age. Younger patients had a higher probability of being diagnosed with diabetes insipidus and requiring desmopressin for treatment. **B:** Probability density function of patient age shows two distinct distributions separable by a diagnosis of diabetes insipidus. **C:** Scatter plot showing the distribution of postoperative lowest sodium with respect to BMI (*dashed black line*, BMI = 30 or clinical obesity). **D:** Probability density function shows unique distributions for patients with BMI less than versus greater than 30 (i.e., obesity diagnosis) when diagnosis of hyponatremia is indicated.

insipidus, compared to nonfunctioning macroadenomas presenting in older patients. Additionally, it is unclear how younger age and obesity, as independent risk factors, could protect against hyponatremia. This observation may be explained as the inverse of the previous; nonobese older patients with macroadenomas undergo less pituitary gland manipulation, and thus these patients are less susceptible to diabetes insipidus but more vulnerable to hyponatremia. While any attempt to interpret these results must be tentative, high-quality training data allow the machine learning model to identify these complex interactions and latent variables, which can then be used to make accurate predictions on new patients.

Our study is limited by being completed at a single institution. Patients treated at other institutions and by other surgeons will be needed to further test the generalizability of the predictive model. The current model is designed as

a binary classifier. With a larger dataset, a multiclass classifier can be trained that may allow for prediction and risk stratification of multiple outcomes (e.g., medical complications, surgical complications, and readmissions). With longer follow-up data, the model can be further tailored to include long-term treatment response and predict tumor recurrence. Our study population will be followed longitudinally in preparation for expanding our predictive model and will provide additional data for model training using machine learning methods similar to those described here.

Conclusions

Pituitary adenomas occur in a heterogeneous patient population, which makes predicting postoperative outcomes a challenge. To address this challenge, we analyzed a large cohort of 400 consecutive pituitary adenoma pa-

TABLE 3. Machine learning model performance

Factor	Naïve Bayes	Support Vector Machines	Random Forest	LR-EN Regularization
Accuracy	79.0 (69.7–86.5)	83.0 (74.2–89.8)	85.0 (76.5–91.4)	87.0 (78.8–92.9)
Sensitivity	24.0	48.0	56.0	68.0
Specificity	97.3	94.7	94.7	93.3
PPV	75.0	75.0	77.8	77.3
NPV	79.4	84.5	86.6	89.7
AUC-ROC	79.5	82.6	84.8	82.7
AUC-PR	64.6	67.2	67.2	69.5

NPV = negative predictive value; PPV = positive predictive value.

Boldface value indicates the highest value for the corresponding metric.

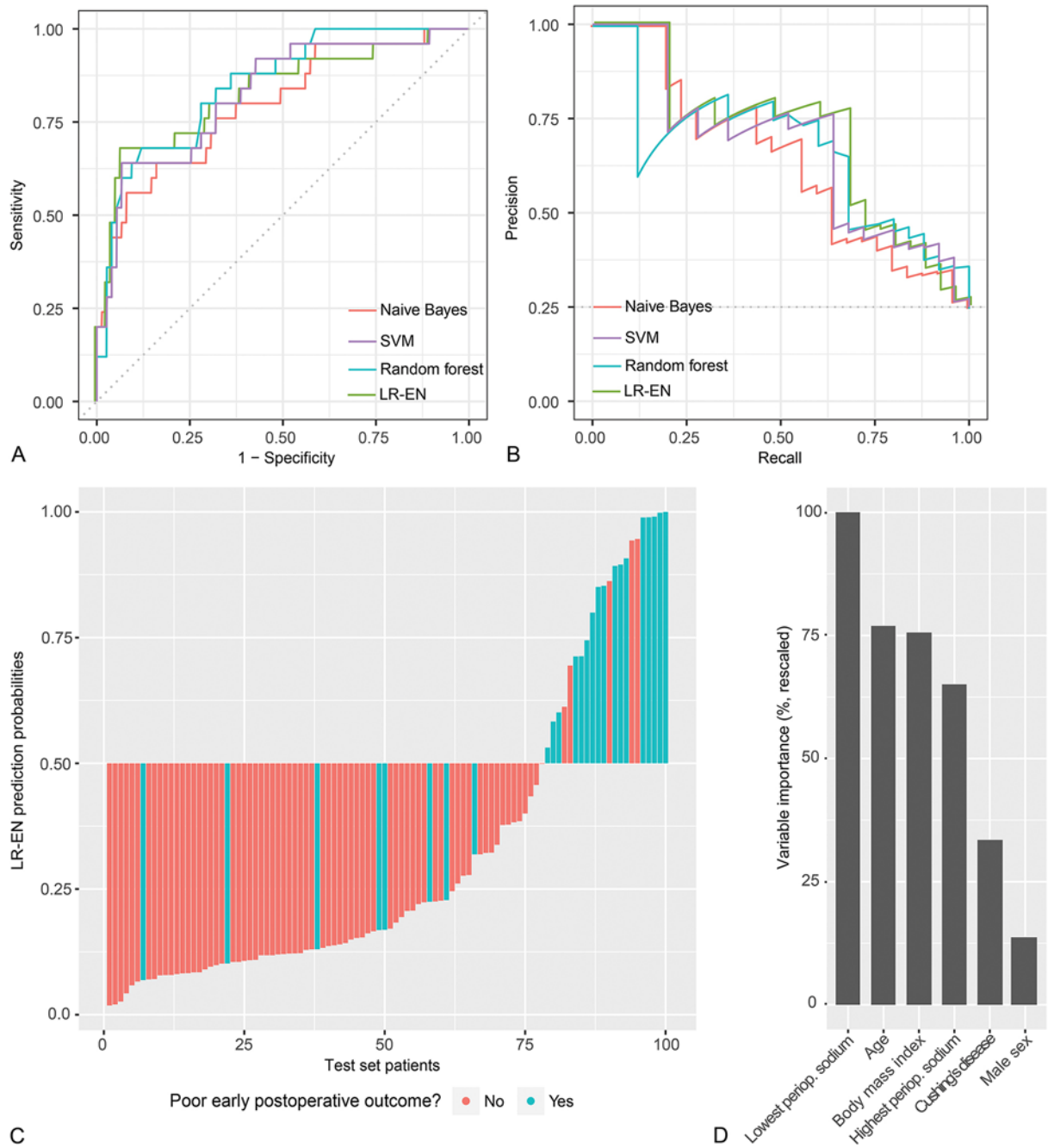


FIG. 5. Machine learning model evaluation, prediction probabilities, and variable importance. ROC curves (**A**) and PR curves (**B**) are shown. Random forest and LR-EN regularization had the top two AUCs for both curves. **C:** Because the LR-EN model had the highest accuracy on the testing set, prediction probabilities for each patient were plotted in ascending order. Ground truth outcome labels are color coded, and LR-EN output probabilities greater than 50% were predicted as having a poor early postoperative outcome. Of the 100 patients, the LR-EN classifier made 5 false-positive and 8 false-negative errors. Rescaled variable importance for the LR-EN model is shown in (**D**). Lowest perioperative (periop.) sodium levels, age, and BMI were the top three predictive variables in the trained model. SVM = support vector machine.

tients and found that with the use of a machine learning approach, early outcomes of pituitary adenoma surgery can be predicted with an accuracy of 87%. These results provide insight into how machine learning can be used to improve the perioperative management of pituitary adenoma patients.

References

1. Cai X, Perez-Concha O, Coiera E, Martin-Sanchez F, Day R, Roffe D, et al: Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *J Am Med Inform Assoc* 23:553-561, 2016
2. Chandler WF, Barkan AL, Hollon T, Sakharova A, Sack J,

- Brahma B, et al: Outcome of transsphenoidal surgery for Cushing disease: a single-center experience over 32 years. **Neurosurgery** **78**:216–223, 2016
3. Chang EF, Smith JS, Chang SM, Lamborn KR, Prados MD, Butowski N, et al: Preoperative prognostic classification system for hemispheric low-grade gliomas in adults. **J Neurosurg** **109**:817–824, 2008
 4. Di Ieva A, Rotondo F, Syro LV, Cusimano MD, Kovacs K: Aggressive pituitary adenomas—diagnosis and emerging treatments. **Nat Rev Endocrinol** **10**:423–435, 2014
 5. Dua S, Acharya UR, Dua P (eds): **Machine Learning in Healthcare Informatics**. Berlin: Springer, 2014
 6. Gaspar L, Scott C, Rotman M, Asbell S, Phillips T, Wasserman T, et al: Recursive partitioning analysis (RPA) of prognostic factors in three Radiation Therapy Oncology Group (RTOG) brain metastases trials. **Int J Radiat Oncol Biol Phys** **37**:745–751, 1997
 7. Gaspar LE, Scott C, Murray K, Curran W: Validation of the RTOG recursive partitioning analysis (RPA) classification for brain metastases. **Int J Radiat Oncol Biol Phys** **47**:1001–1006, 2000
 8. Gondim JA, Almeida JP, de Albuquerque LA, Gomes E, Schops M, Mota JJ: Endoscopic endonasal transsphenoidal surgery in elderly patients with pituitary adenomas. **J Neurosurg** **123**:31–38, 2015
 9. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K: Supporting information retrieval from electronic health records: a report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). **J Biomed Inform** **55**:290–300, 2015
 10. Hardy J: Transphenoidal microsurgery of the normal and pathological pituitary. **Clin Neurosurg** **16**:185–217, 1969
 11. Knosp E, Steiner E, Kitz K, Matula C: Pituitary adenomas with invasion of the cavernous sinus space: a magnetic resonance imaging classification compared with surgical findings. **Neurosurgery** **33**:610–618, 1993
 12. Lawton MT, Kim H, McCulloch CE, Mikhak B, Young WL: A supplementary grading scale for selecting patients with brain arteriovenous malformations for surgery. **Neurosurgery** **66**:702–713, 2010
 13. Li J, Wang M, Won M, Shaw EG, Coughlin C, Curran WJ Jr, et al: Validation and simplification of the Radiation Therapy Oncology Group recursive partitioning analysis classification for glioblastoma. **Int J Radiat Oncol Biol Phys** **81**:623–630, 2011
 14. Myers PD, Scirica BM, Stultz CM: Machine learning improves risk stratification after acute coronary syndrome. **Sci Rep** **7**:12692, 2017
 15. Oermann EK, Kress MA, Collins BT, Collins SP, Morris D, Ahalt SC, et al: Predicting survival in patients with brain metastases treated with radiosurgery using artificial neural networks. **Neurosurgery** **72**:944–52, 2013
 16. Oermann EK, Rubinsteyn A, Ding D, Mascitelli J, Starke RM, Bederson JB, et al: Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. **Sci Rep** **6**:21161, 2016
 17. Ostrom QT, Gittleman H, Xu J, Kromer C, Wolinsky Y, Kruchko C, et al: CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2009–2013. **Neuro Oncol** **18** (Suppl 5):v1–v75, 2016
 18. Park CK, Kim JH, Nam DH, Kim CY, Chung SB, Kim YH, et al: A practical scoring system to determine whether to proceed with surgical resection in recurrent glioblastoma. **Neuro Oncol** **15**:1096–1101, 2013
 19. Patil CG, Lad SP, Harsh GR, Laws ER Jr, Boakye M: National trends, complications, and outcomes following transsphenoidal surgery for Cushing's disease from 1993 to 2002. **Neurosurg Focus** **23**(3):E7, 2007
 20. Pignatti F, van den Bent M, Curran D, Debruyne C, Sylvester R, Therasse P, et al: Prognostic factors for survival in adult patients with cerebral low-grade glioma. **J Clin Oncol** **20**:2076–2084, 2002
 21. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj M, Liu PJ, et al: Scalable and accurate deep learning for electronic health records. **NPJ Digit Med** **1**:18, 2018
 22. Simpson D: The recurrence of intracranial meningiomas after surgical treatment. **J Neurol Neurosurg Psychiatry** **20**:22–39, 1957
 23. Spetzler RF, Martin NA: A proposed grading system for arteriovenous malformations. **J Neurosurg** **65**:476–483, 1986
 24. Spetzler RF, Ponce FA: A 3-tier classification of cerebral arteriovenous malformations. Clinical article. **J Neurosurg** **114**:842–849, 2011
 25. Yano S, Shinojima N, Kawashima J, Kondo T, Hide T: Intraoperative scoring system to predict postoperative remission in endoscopic endonasal transsphenoidal surgery for growth hormone-secreting pituitary adenomas. **World Neurosurg** **105**:375–385, 2017
 26. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. **Nat Commun** **7**:12474, 2016

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Sullivan, Hollon, Parikh, Barkan, McKean. Acquisition of data: Hollon, Parikh, Pandian, Tarpeh. Analysis and interpretation of data: Sullivan, Hollon, Parikh, Orringer, Barkan, McKean. Drafting the article: Hollon. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Sullivan. Statistical analysis: Hollon, Pandian. Administrative/technical/material support: Hollon, Pandian, Orringer. Study supervision: Sullivan, Orringer, Barkan, McKean.

Correspondence

Stephen E. Sullivan: University of Michigan, Ann Arbor, MI. ssulliva@med.umich.edu.