

Original Article

The impact of presurgical comorbidities on discharge disposition and length of hospitalization following craniotomy for brain tumor

Whitney E. Muhlestein, Dallin S. Akagi¹, Silky Chotai, Lola B. ChamblessDepartment of Neurological Surgery, Vanderbilt University, Vanderbilt University Medical Center, Nashville, Tennessee, ¹DataRobot, Inc., Boston, Massachusetts, USAE-mail: *Whitney E. Muhlestein - Whitney.e.muhlestein@vanderbilt.edu; Dallin S. Akagi - dallin@datarobot.com; Silky Chotai - silky.chotai@vanderbilt.edu; Lola B. Chambless - lola.chambless@vanderbilt.edu

*Corresponding author

Received: 03 February 17 Accepted: 22 June 17 Published: 07 September 17

Abstract

Background: Identifying risk factors for negative postoperative outcomes is an important part of providing quality care. Here, we build machine learning (ML) ensembles to model the independent impact of presurgical comorbidities on discharge disposition and length of stay (LOS) following brain tumor resection from the HCUP National Inpatient Sample (NIS).

Methods: We performed a retrospective cohort study of 41,222 patients who underwent craniotomy for brain tumors during 2002–2011 and were registered in the NIS. Twenty-six ML algorithms were trained on prehospitalization variables to predict nonhome discharge and extended LOS (>7 days), and the most predictive algorithms combined to create ensemble models. Models were validated to demonstrate generalizability. Analysis was done to identify which and how specific comorbidities influence ensemble predictions.

Results: Receiver operating curve analysis showed area under the curve of 0.796 and 0.824 for the disposition and LOS ensembles, respectively. The disposition ensemble was most strongly influenced by preoperative paralysis and fluid/electrolyte abnormalities, which independently increased the risk of nonhome discharge in craniotomy patients by 35.4% and 13.9%, respectively. The LOS ensemble was most strongly influenced by the presence of preoperative paralysis, fluid/electrolyte abnormalities, and other nonparalysis neurological deficits, which independently increased the risk of extended LOS in craniotomy patients by 20.4%, 22.5%, and 38.3%, respectively.

Conclusions: In this study, we used ML ensembles to identify preoperative comorbidities that increased the risk of nonhome discharge and extended LOS following craniotomy for brain tumor. Recognizing these risk factors for poor

Access this article online

Website:www.surgicalneurologyint.com**DOI:**

10.4103/sni.sni_54_17

Quick Response Code:

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of presurgical comorbidities on discharge disposition and length of hospitalization following craniotomy for brain tumor. *Surg Neurol Int* 2017;8:220.

<http://surgicalneurologyint.com/The-impact-of-presurgical-comorbidities-on-discharge-disposition-and-length-of-hospitalization-following-craniotomy-for-brain-tumor/>

postsurgical outcomes can improve patient counseling and offer opportunities for quality improvement.

Key Words: Comorbidities, intracranial tumor, machine learning, postoperative outcomes

INTRODUCTION

It is important for neurosurgeons to identify risk factors for poor outcomes following craniotomy for brain tumor (CFBT) as it impacts care on both the individual patient and healthcare systems level. An improved understanding of risk factors allows better patient counseling by providers, helps facilitate coordination of postoperative care between patients, providers, patient caregivers, and insurance companies, and affords patients and providers the opportunity to address potentially reversible conditions that could adversely impact postoperative outcomes. On the healthcare systems level, accurate identification of patients at high risk for poor postoperative outcomes helps surgeons and hospital administrators more efficiently apportion scarce hospital resources, while also helping hospitals and payers more easily determine appropriate rates of reimbursement for high-resource patients. Prediction of high-risk patients also allows more accurate calculation of observed-to-expected morbidity and mortality ratios, which can benefit quality improvement initiatives and avoid undue penalties for providers and hospitals that take on challenging cases.

Several operative risk assessments have already been developed and are in regular use. Some, such as the American Society of Anesthesiologists Physical Status (ASA-PS) classification system, are very simple risk indices. While intuitive and user-friendly, this risk index is not specific to neurosurgical patients and does not look at specific comorbidities, limiting its usefulness in terms of targeting potentially reversible preoperative patient characteristics.^[23] This is particularly important in light of the fact that baseline patient characteristics have been identified as critical drivers of poor outcomes following CFBT. Grossman *et al.* showed that preoperative Charlson Comorbidity Score (a weighted patient comorbidity index) is predictive of inpatient postoperative mortality, length of stay (LOS), total hospital charges, and postoperative complications for elderly patients undergoing craniotomy for meningioma^[10] and for metastatic brain tumors.^[11] Dasenbrock *et al.* identified specific comorbidities recorded in the National Surgical Quality Improvement Program (NSQIP) database that increased the risk of extended LOS in patients undergoing CFBT.^[8] It is clear, therefore, that risk prediction tools that incorporate patient comorbidities are needed to provide the most useful information to providers and their patients.

Here, we use a novel machine learning (ML) technique to predict two postoperative outcomes for patients registered in the National Inpatient Sample (NIS) undergoing CFBT – discharge disposition and length of hospital stay (LOS). We directly compare the predictive ability of 26 unique ML algorithms and combine the top performing algorithms to create final predictive ensembles. We then interrogate the ML ensembles to investigate the independent impact of 29 different patient comorbidities on discharge disposition and LOS. We also build ML ensembles for specific tumor subtypes to determine whether comorbidity risk factors vary based on pathologic diagnosis.

MATERIALS AND METHODS

Database

We used the NIS in-hospital discharge database for the years 2002–2011. NIS is the largest all-payer inpatient database publicly available in the United States, containing approximately 8 million hospital stays from ~1000 hospitals, sampled to approximate a 20% stratified sample of US hospitals.^[13] NIS is compiled and maintained by the Agency for Healthcare Research and Quality (AHRQ, Rockville, Maryland, USA). This publicly available, de-identified database was considered exempt from IRB review.

Patient selection

All 79,742,743 admissions registered in the NIS between 2002 and 2011 were screened for inclusion in the study. Eligible admissions were first identified by ICD9 diagnosis codes for brain tumor (225.0–225.4, 225.8, 225.9, 199.1, 191.0–191.9). Admissions in this subset were then screened for ICD9 procedure codes matching craniotomy (01.20–01.29, 01.31, 01.32, 01.39, 01.59). We further restricted our cohort to patients 18 years or older. A total of 41,222 admissions met our criteria.

To determine whether trends in the entire CFBT population are mimicked within specific brain tumor diagnoses, we derived the following tumor subsets using the appropriate ICD9 diagnosis codes: meningioma (225.2), nonmeningioma benign tumor (225.0, 225.1, 225.8, 225.9), and malignant tumor (199.1, 191.0–191.9).

Variable selection and primary outcomes

Data was collected on a variety of preoperative patient characteristics, including age, race, sex, specific neurosurgical diagnosis, comorbidities, admission

Table 1: AHRQ comorbidities included in ensemble training

AIDS	Lymphoma
Alcohol abuse	Metastatic cancer
Chronic blood loss anemia	Obesity
Chronic pulmonary disease	Other neurologic disorders
Coagulopathies	Paralysis
Congestive heart failure	Peptic ulcer disease excluding bleeding
Deficiency anemias	Peripheral vascular disorders
Depression	Psychoses
Diabetes, uncomplicated	Pulmonary circulation disorders
Diabetes, with chronic complications	Renal failure
Drug abuse	Rheumatoid arthritis/collagen vascular diseases
Fluid/electrolyte abnormalities	Solid tumor without metastasis
Hypertension	Valvular disease
Hypothyroidism	Weight loss
Liver disease	

AHRQ: Agency for Healthcare Research and Quality

type, emergent vs nonemergent surgery, expected payer, and hospital characteristics. The 29 included comorbidity variables were identified using the Elixhauser Comorbidity Software administered by AHRQ [Table 1]. In total, 61 different variables were considered [Supplementary Table 1]. The primary outcome for the discharge disposition model was home vs other (including death, short-term hospital, AMA, home health, and other). Based on previous studies of LOS in the CFBT population, the primary outcome for the LOS model was defined as greater than 7 inpatient days versus less than or equal to 7 inpatient days.^[8] According to this definition, patients with extended LOS made up approximately one-third of the CFBT population.

Data preprocessing

Numerous data preprocessing approaches are represented in the collection of algorithms evaluated in the leaderboard. This section describes the approaches used in the algorithms included in the final ensembles. Missing numerical data was dealt with by imputing the median of the column and creating a new binary column to indicate the imputation. Numerical data was standardized in each column by subtracting the mean and dividing by the standard deviation. For linear algorithms (Support Vector Machine, Elastic Net Classifier, Regularized Logistic Regression, Stochastic Gradient Descent Classifier, and Vowpal Wabbit Classifier), categorical data was turned into many binary columns by one-hot encoding. Missing categorical values were treated as their own categorical level and get their own column. For tree-based algorithms, categorical data was encoded with integers.

The assignment of category values to integers was done randomly.

Leaderboard construction and model validation

Before training, 20% of the dataset was randomly selected as the holdout, which was never used in training or cross-validation. The remaining data was divided into five mutually exclusive folds of data, four of which were used together as training, with the final fold used for validation.^[16] Training was performed five times per algorithm, with each fold used once for validation. Cross-validation scores were calculated by taking the average area under the curve (AUC) of the receiver-operating characteristic (ROC) of the five possible validation folds. Twenty-six ML algorithms were trained and scored, and the top performing algorithms were selected for use in each ensemble. The algorithms were combined with an Elastic Net Classifier and an Average Blender for the disposition and LOS ensembles, respectively. Following ensemble model creation, validation was performed on the holdout to demonstrate the ability of the model to generalize never-before-seen data. The holdout was taken as a single sample, and hence no confidence intervals were calculated. Model construction was performed using DataRobot (version 3.0; DataRobot, Inc, Boston, MA).

Permutation importance

The relative importance of a feature to the final ensemble model was assessed using permutation importance, as described by Breiman.^[3] Using training data only, for each variable, the ensemble was retrained on data with the values for the variable randomly permuted. The difference in performance in AUC between the ensemble built on the reference data and that of the data with the permuted variable was used to rank and compare the relative importance of the features to the ensemble.

Partial dependence

To understand the independent impact of race on the disposition and LOS ensembles, we constructed Partial Dependence plots, as described by Friedman.^[9] A subset of the training data was selected. For any variable, we made predictions from the ensemble after replacing all values for the variable with a constant test value and computing the mean of those predictions. We tested many values to observe how the ensemble reacts to changes in the variable of interest.

Other statistical methods

Traditional statistical analysis was performed on selected patient and hospital characteristics. Continuous variables were compared using the Mann-Whitney *U* test. Categorical variables were compared using Pearson's χ^2 test. Statistical analysis was performed using the open source statistical tools in SciPy (SciPy ver 0.17).

RESULTS

Patient characteristics

A total of 41,222 admissions for CFBT were reviewed for analysis; 25,406 resulted in discharge to home and 15,705 admissions did not. One hundred and eleven admissions had no or unknown discharge disposition recorded and were excluded from the study. 69.3% of patients had at least one comorbidity. With the exception of drug abuse and peptic ulcer disease, all comorbidities were associated with a significantly increased risk of nonhome discharge [Supplementary Table 2]. Patients at higher risk for nonhome discharge tended to be females ($P < .001$), black ($P < .001$), older (average age 60.7 years versus 50.5 years $P < .001$), have Medicare as their primary payer ($P < .001$), and have nonelective surgeries ($P < .001$). They were seen in medium-sized ($P = 0.02$), private ($P < .001$) urban nonteaching ($P < .001$), northeastern hospitals ($P < .001$) [Supplementary Table 3].

A total of 27,314 admissions lasted less than or equal to 7 days, and 13,907 admissions lasted more than 7 days. One admission had no recorded LOS and was excluded from analysis. 69.4% of the patients had at least one comorbidity. Except lymphoma and rheumatoid arthritis or other collagen vascular diseases, all comorbidities were associated with a significant risk of extended LOS [Supplementary Table 4]. Patients at higher risk for extended hospitalization also tended to be older (average age 57.8 years vs 52.7 years, $P < .001$), black or other nonblack minority race ($P < .001$), and have nonelective surgeries ($P < .001$). They were seen in medium-sized ($P < .001$), private ($P < .001$), urban nonteaching ($P < .001$), southern hospitals ($P < .001$) [Supplementary Table 5].

Receiver-operator characteristics curve and other classifier statistics

An ensemble model including a Nystroem Kernel Support Vector Machine (SVM) Classifier, Elastic-Net Classifier, and Extreme Gradient Boosted Trees Classifier was best able to predict discharge disposition, and an ensemble comprising an Elastic-Net Classifier, a Vowpal Wabbit Classifier, a Stochastic Gradient Descent Classifier, two Extreme Gradient Boosted Trees Classifiers, a Gradient Boosted Tree Classifier, a Nystroem Kernel SVM, and a Regularized Logistic Regression was best able to predict extended LOS. The disposition ensemble model had an AUC on the validation set of 0.796 (95% CI, 0.790–0.801), and the LOS ensemble had an AUC of 0.824 (95% CI, 0.823–0.826). Validating on the holdout set yielded an AUC of 0.807 for the disposition ensemble and 0.818 for the LOS ensemble. When optimizing the F1 score, the discharge ensemble had a positive predictive value (PPV) of 60.0% (95% CI, 59.1–60.9%), a negative predictive value (NPV) of 82.0% (95% CI, 81.5–82.4%), sensitivity of 75.5% (95% CI, 75.0–76.0%), and specificity of 68.9% (95% CI, 67.9–69.8%). The LOS ensemble had a PPV of 58.8% (95% CI, 58.6–59.0%), an NPV of 86.2% (95% CI, 86.0–86.4%), sensitivity of 77.2% (95% CI, 76.9–77.5%), and specificity 72.5% (95% CI, 72.2–72.7%). All additional ensembles predicted their respective outcomes with similar discrimination [Table 2].

We also built logistic regressions to compare the predictive abilities of our ensembles to those of traditional regression-based models. Both ensembles performed better than the corresponding logistic regression on the data hold out (AUC = 0.807 vs 0.802 for the disposition models; AUC = 0.818 vs 0.816 for the LOS models).

Permutation and partial dependence analysis

We performed permutation importance and partial dependence analyses to determine which variables are

Table 2: Cross-validation AUC and other metrics for constructed ensembles

Ensemble and patient population	AUC (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Holdout
Discharge						
CFBT	0.796 (0.790-0.801)	60.0 (59.1-60.9)	82.0 (81.5-82.4)	75.5 (75.0-76.0)	68.9 (67.9-69.8)	0.807
Meningioma	0.789 (0.777-0.799)	57.8 (56.7-58.7)	81.3 (79.7-82.6)	78.7 (76.2-80.8)	61.6 (59.6-63.4)	0.804
Nonmeningioma benign tumor	0.761 (0.744-0.776)	50.0 (47.3-51.9)	83.2 (81.9-84.4)	71.4 (69.1-73.5)	66.3 (63.4-68.8)	0.782
Malignant tumor	0.793 (0.788-0.797)	60.2 (59.9-60.4)	82.2 (81.9-82.5)	77.2 (76.5-77.7)	67.3 (66.8-67.8)	0.793
Length of Stay						
CFBT	0.824 (0.823-0.826)	58.8 (58.6-59.0)	86.2 (86.0-86.4)	77.2 (76.9-77.5)	72.5 (72.2-72.7)	0.818
Meningioma	0.815 (0.809-0.820)	58.7 (60.0-60.3)	86.4 (85.5-87.2)	78.7 (77.0-80.2)	70.9 (68.7-72.8)	0.811
Non-meningioma benign tumor	0.788 (0.761-0.812)	56.0 (52.8-58.8)	84.1 (81.5-86.4)	73.3 (68.1-77.7)	71.0 (68.1-74.5)	0.779
Malignant tumor	0.819 (0.815-0.823)	55.8 (55.1-56.3)	88.2 (87.8-88.5)	82.6 (81.9-83.2)	66.4 (65.5-67.2)	0.817

Cross-validation receiver operating characteristic and other metrics for each ensemble. Metrics are calculated at the optimized F1 score for each ensemble. The holdout for each ensemble is taken as a single sample, and so no confidence intervals are calculated. AUC: Area under the curve, CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value, CFBT: Craniotomy for brain tumor

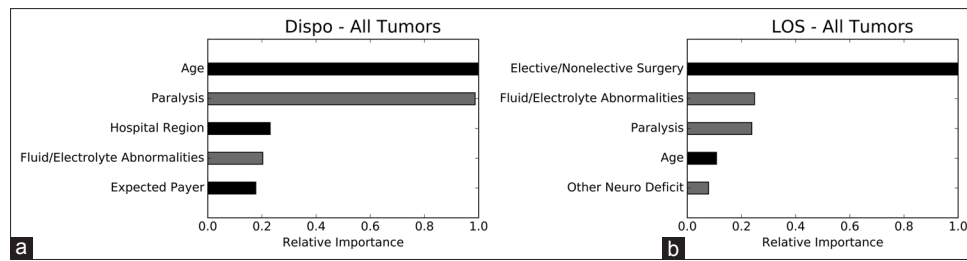


Figure 1: Permutation importance. Permutation importance demonstrating the relative importance of individual variables to the disposition and LOS ensembles. The most important variable is given an importance value of 1.0 and the importance of other variables is shown relative to 1.0. Gray bars represent comorbidities. Dispo, disposition; LOS, length of stay. (a) disposition ensemble; (b) LOS ensemble

most important to, and how they independently impact, the ensembles. The strongest risk factors for nonhome discharge are, in order, increasing age, preoperative paralysis, hospitalization in the northeast, nonelective surgery, and preoperative fluid/electrolyte abnormalities. The strongest risk factors for extended LOS in the LOS ensemble are nonelective surgery, preoperative paralysis, preoperative fluid/electrolyte abnormalities, increasing age, and other preoperative neurological deficits [Figure 1]. Preoperative paralysis and preoperative fluid/electrolyte abnormalities increase the risk of nonhome discharge by 35.4% and 13.9%, respectively. Preoperative paralysis increases the risk of extended LOS by 20.4%, whereas preoperative fluid/electrolyte abnormalities increase the risk by 22.5% and other preoperative neurological deficits increase the risk by 38.3%. Subgroup analysis showed very little variance from the trends identified in the overall CFBT population, suggesting that specific tumor diagnosis does not strongly influence the effect of specific patient comorbidities on either discharge disposition or LOS [Figure 2].

DISCUSSION

In this study, we built ML ensemble models to identify preoperative comorbidities that most strongly predict nonhome discharge and extended LOS following CFBT. Previous work to identify patient characteristics that predict poor neurosurgical outcomes has tended to rely on traditional statistical techniques, in particular logistic regression (LR). While a powerful technique, LR works best on a limited range of datasets, particularly those that contain few, independent variables.^[1] ML is a type of analysis that uses diverse computer algorithms to identify patterns in large, complicated data sets.^[12] Newer ML algorithms are particularly adept at handling vast numbers of variables in complex and nonlinear ways to generate sophisticated predictions.^[20] Given the complexity of human disease and the large amounts of patient data now collected in electronic medical records, some researchers have suggested that algorithms other than LR may be better suited to predicting patient outcomes.^[21,24] It is impossible to know, however, which ML algorithm will best predict an outcome of interest for a given dataset without

directly comparing the predictive abilities of the different algorithms. The advantage of our guided ML ensemble technique lies in the fact that we directly compare 26 different ML algorithms using the same validation metrics, allowing us to objectively identify the algorithms that best predict outcomes of interest from our data.

Our ensemble models drew from a library of open source ML algorithms representing major classes of ML algorithms currently available. The algorithms selected for inclusion in each ensemble were selected based solely on their predictive abilities after training on the dataset – we did not bias our study by selecting algorithms *a priori*. The disposition ensemble comprised a Nystroem Kernel SVM classifier, Elastic-Net classifier, and Extreme Gradient Boosted Trees classifier, whereas the LOS ensemble comprised an Elastic-Net classifier, Stochastic Gradient Descent classifier, Vowpal Wabbit classifier, two Extreme Gradient Boosted Trees classifiers, and a Gradient Boosted Tree classifier, Nystroem Kernel SVM, and Regularized Logistic Regression. We briefly describe several of the machine learning algorithms that may be unfamiliar to the practicing neurosurgeon.

- Nystroem Kernel SVMs plot inputs as vectors in higher dimensional space, with each axis corresponding to a different variable. The algorithm then calculates a plane that separates the inputs into two different classes. New inputs are then assigned to a class based on which side of the plane they fall on.^[26]
- Elastic Nets are logistic regressions that employ lasso and ridge regularization to improve predictive accuracy. Lasso and ridge regularizations work to minimize model overfitting (when a model describes random error or noise inherent to the database rather than the true, underlying associations) by shrinking or eliminating large regression coefficients.^[27]
- Tree-based models use decision rules to classify data. Large trees may have many decision rules, the outcome of each contributing to the overall likelihood that a data input will fall into one class or another. Extreme Gradient Boosted Trees continually add weak decision rules to the overall tree to produce models that are optimized to classify edge cases.^[4,7,9]
- Stochastic Gradient Descent classifiers are linear models that initially assign each variable a random

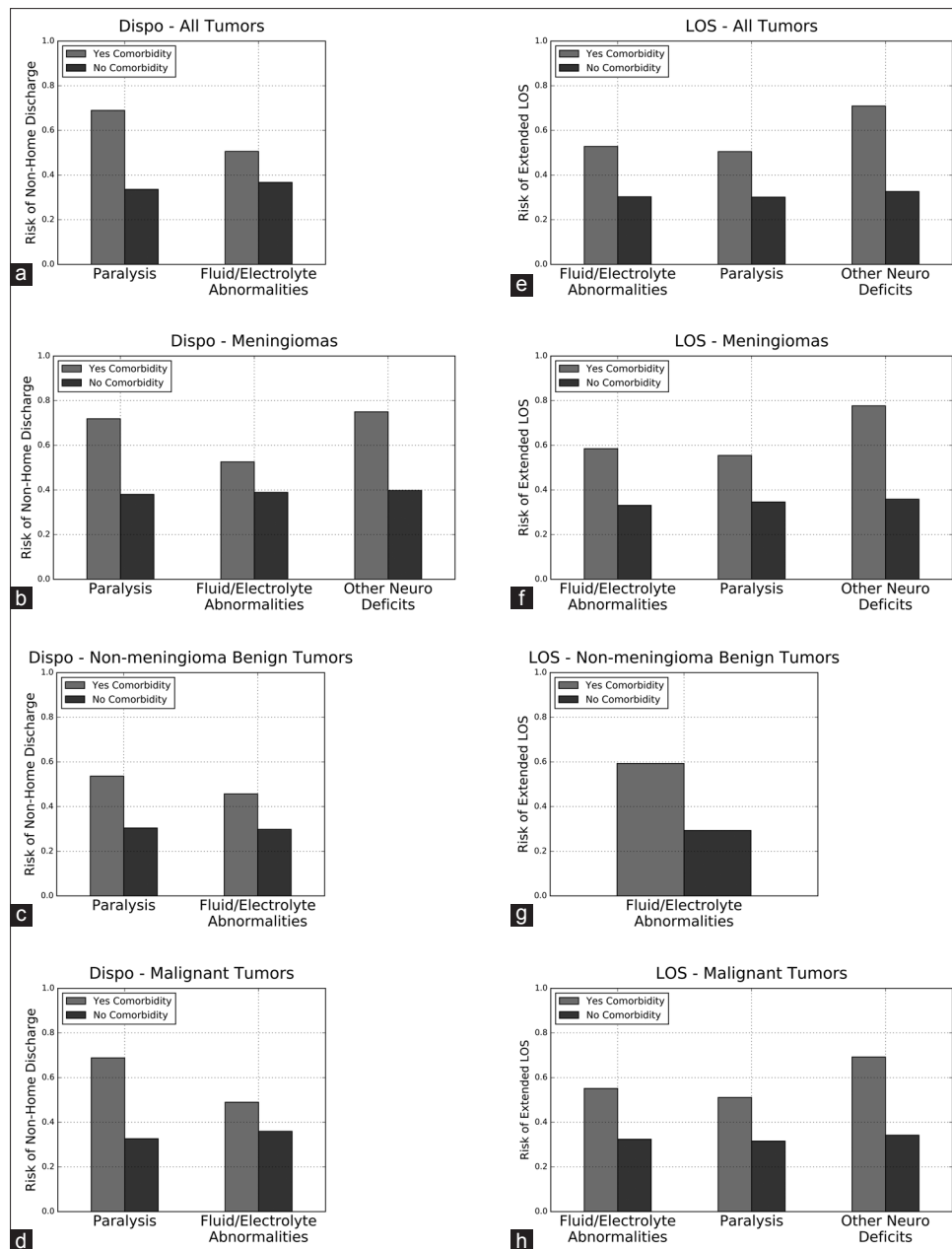


Figure 2: Partial dependence plots for comorbidities in the disposition and LOS ensembles. Partial dependence plots demonstrating the independent impact of comorbidities included in the top five most important variables for predicting discharge disposition or LOS. X-axis represents probability of non-home discharge or extended LOS, with 1 equivalent to 100% likelihood of non-home discharge or extended LOS and 0 equivalent to 0% likelihood. Dispo, disposition; LOS, length of stay. (a) disposition ensemble for all tumors, (b) Disposition ensemble for meningiomas, (c) Disposition ensemble for non-meningioma benign tumors, (d) Disposition ensemble for malignant tumors, (e) LOS ensemble for all tumors, (f) LOS ensemble for meningiomas, (g) LOS for non-meningioma benign tumors, (h) LOS for malignant tumors

coefficient. The error function of the model is then calculated, and coefficient values updated in the direction that minimizes the error function. This process continues in a stepwise manner until minimization is achieved^[3]

- Vowpal Wabbit classifiers represent a specific implementation of a Stochastic Gradient Descent classifier built specifically to handle large volumes of streaming (continually refreshing and updating) data.^[25]

We point interested readers toward the references for more details.

We take further advantage of ML by combining the top performing models into ensembles, allowing us to utilize the unique advantages of different classes of algorithms within the same predictive ensemble. Our ML ensembles have good discrimination for both nonhome discharge (AUC = 0.796) and extended

LOS (AUC = 0.824) on internal validation. Importantly, the ensemble models also have good discrimination on the data hold out (AUC = 0.807 for the disposition model; AUC = 0.818 for the LOS model), demonstrating the ability of the ensembles to correctly categorize never-before-seen patients into home and nonhome discharge and extended and nonextended LOS. Ultimately, by building the most predictive ensembles for discharge disposition and LOS from our data, we can draw more accurate conclusions about the impact of comorbidities on the outcomes of interest.

Univariate analysis of the NIS database demonstrated that the presence of at least one comorbidity is associated with nonhome discharge and extended LOS. Analysis of each of the 29 comorbidities individually also revealed a statistically significant association with nonhome discharge and extended LOS for nearly all recorded comorbidities. Using permutation analysis and partial dependence plots of our ensembles, we identified preoperative paralysis and fluid/electrolyte abnormalities as the strongest independent comorbidity risk factors for nonhome discharge, and preoperative paralysis, fluid/electrolyte abnormalities, and other neurological deficits as the strongest independent comorbidity risk factors for extended LOS. Knowing that these, among the myriad comorbidities represented in the ensembles, most strongly influence the risk of nonideal postoperative outcomes helps neurosurgeons provide more accurate risk assessments for patients. Importantly, the independent impact of these preoperative comorbidities remains even in the presence of many other covariates, including age, gender, elective vs nonelective surgery, race, specific tumor diagnosis, and hospital characteristics.

While we controlled for a large number of variables before identifying the impact of preoperative paralysis, fluid/electrolyte abnormalities, and nonparalysis neurological deficits on extended LOS and nonhome discharge, these comorbidities may act as partial surrogates for variables not included in the original ensembles. Disease severity, for example, is a particularly important variable that was not directly addressed in the ensembles. It is likely that preoperative paralysis and nonparalysis neurological deficits correlate with more advanced preoperative disease, reflecting large tumor size or proximity of tumors to sensitive brain structures, which in turn may be associated with a host of negative prognostic factors, including extent of tumor resection and length of operation.

It is similarly possible that fluid/electrolyte abnormalities reflect a patient's baseline health status, rather than acting as true independent predictors of poor postsurgical outcomes. Cecconi *et al.* demonstrated, for example, that only severe

hyponatremia is independently associated with increased postoperative mortality, and that poor outcomes associated with other types of sodium abnormalities likely reflect the underlying cause of the abnormality rather than the abnormality itself.^[6] In our study, we attempt to control for baseline health characteristics by including a fairly comprehensive array of 29 preoperative comorbidities in our ensembles. It is possible, however, that they do not fully describe a patient's preoperative state of health.

Understanding the impact of these comorbidities on postsurgical outcomes is important from the perspective of patient counseling and treatment planning. We were surprised to find that some of the comorbidities identified here as strongly informing risk of poor postoperative outcomes are not accounted for in well-established surgical risk calculators, including the American College of Surgeons (ACS) Universal Surgical Risk Calculator.^[2] This suggests that these tools may underestimate risk for patients undergoing CFBT. Notable among the omissions are preoperative fluid and abnormalities, which represent one of the strongest risk factors for both nonhome discharge and extended LOS for patients following CFBT. Critically, electrolyte and fluid abnormalities are potentially reversible conditions, giving providers an opportunity to intervene preoperatively to improve patient outcomes postoperatively.

Identifying important risk factors for nonhome discharge and extended LOS is also important on a health-systems level. In recognition of the importance of efficient coordination of discharge to post-acute facilities or programs, some Medicare reimbursement models are now bundling acute hospitalizations and post-acute care into single episodes of care.^[5] Predicting which patients may require a higher level of care following CFBT will help acute and post-acute care providers better coordinate transitions of care, potentially improving resource utilization and decreasing costs. Identifying patients at high risk for complications can also improve reimbursement for healthcare services. Missios *et al.* recently identified LOS as a primary driver of increasing hospitalization cost after CFBT – preemptively identifying patients who will require extended LOS following craniotomy can therefore help providers and hospitals justify appropriately higher rates of reimbursement for these patients to payers.^[17]

Finally, recognizing risk factors for poor postoperative outcomes can aid hospitals and providers accurately gauge quality of care at their institutions. High quality care, which is defined by the Institute of Medicine as effective, efficient, equitable, safe, and timely provision of services, is being increasingly emphasized and incentivized in health care reimbursement strategies.^[14]

Given the abstract nature of quality, surrogate measures for quality, including postoperative outcomes such as extended LOS, are seeing more extensive use. In agreement with other researchers, the findings of this study suggest that baseline patient characteristics play an important role in determining whether or not patients undergoing CFBT will have “poor quality” postoperative outcomes.^[8,10,11] This has enormous implications for hospitals and providers who take care of patients with preoperative comorbidities, particularly those identified in this study as being strongly associated with nonhome discharge and extended LOS. Recognizing patients who will likely have nonhome discharge or extended LOS gives providers and hospitals a better understanding of the true quality of care that they provide, helping them not only avoid undue penalty for “poor quality” care for patients with these comorbidities but also helping them to target quality improvement projects to areas over which they have direct control.

Our study has several advantages; we use a well-established, validated, multi-institution database to train our predictive algorithms. We also use a standardized, validated method of identifying patient comorbidities within our database. In addition, our guided-ML ensemble technique allows us to objectively select algorithms that best predict outcomes of interest from our database.

Despite these advantages, there are several limitations to the study; the guided-ML ensemble is trained and validated on the same retrospective database. Though we demonstrate generalizability of the model to data never used in algorithm training, a stronger validation of the model would involve demonstrating generalizability to a prospective or to an entirely separate database. It is possible, for example, that nuances in the way data is collected in the NIS influence the model in a way that weakens generalizability to non-NIS data. Furthermore, though a powerful modeling tool, ML and the associated algorithm analysis is unfamiliar to most surgeons and requires training and expertise to wield appropriately. Our strategy in particular is novel, and will require further study and validation. Finally, we were unable, due to limitations of the database, to correlate decreased LOS or home discharge with functional outcome measures. However, other work has shown that decreased LOS^[15,18,22] and home discharge^[19] may be associated with decreased readmission and all-cause mortality rates for a variety of pathologies. It is possible that these improvements reflect milder preoperative disease severity. At least one study, however, demonstrated a decreased risk of 30-day all-cause mortality for a cohort of patients with hip fracture discharged to home as opposed to permanent nursing homes after controlling for preoperative functional status and comorbidities.^[19]

CONCLUSION

In this study, we identify preoperative paralysis and fluid/electrolyte abnormalities as the strongest risk factors for nonhome discharge in craniotomy patients, whereas preoperative fluid/electrolyte abnormalities, paralysis, and other neurological deficits are the strongest risk factors for extended LOS. Recognizing these risk factors for poor postoperative outcomes provides opportunities to improve patient counseling, address potentially reversible causes of negative postoperative outcomes, improve reimbursement and resource allocation, and institute more effective quality improvement initiatives.

Financial support and sponsorship

We Muhlestein received financial support from the Vanderbilt Medical Scholars Program and the UL1 RR 024975 NIH CTSA grant.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. *Cancer Lett* 2016;382:110-7.
2. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY, et al. Development and evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013;217:833-42.e3.
3. Bottou L. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. Heidelberg, Physica-Verlag; 2010. p. 177-86.
4. Breiman L. Random forests. *Machine Learning*, 2001;45:5-32.
5. Bundled Payments for Care Improvement (BPCI) Initiative: General Information. Centers for Medicare & Medicaid Services; [cited 2016 Nov 11]. Available from: <https://innovation.cms.gov/initiatives/bundled-payments/>.
6. Cecconi M, Horchrieser H, Chew M, Grocott M, Hoefl A, Hoste A, et al. Preoperative abnormalities in serum sodium concentrations are associated with higher in-hospital mortality in patients undergoing major surgery. *Br J Anaesth* 2016;116:63-9.
7. Chen T, Guestrin C. XGBoost: A scalable tree boosting system, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017 Aug 13-17; San Francisco, CA, USA.
8. Dasenbrock HH, Liu KX, Devine CA, Chavakula V, Smith TR, Gormley WB, et al. Length of hospital stay after craniotomy for tumor: A National Surgical Quality Improvement analysis. *Neurosurg Focus* 2015;39:E12.
9. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Statistics* 2001;29:1189-232.
10. Grossman R, Mukherjee D, Chang DC, Bennett R, Brem H, Olivi A, et al. Preoperative Charlson Comorbidity Score predicts postoperative outcomes among older intracranial meningioma patients. *World Neurosurg* 2011;75:279-85.
11. Grossman R, Mukherjee D, Chang DC, Purtil M, Lim M, Brem H, et al. Predictors of inpatient death and complications among postoperative elderly patients with metastatic brain tumors. *Ann Surg Oncol* 2011;18:521-8.
12. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer Science & Business Media; 2009.
13. Healthcare Cost and Utilization Project Databases: Nationwide Inpatient Sample. Agency for Healthcare Research and Quality; Available from: <http://www.hcup-us.ahrq.gov/nisoverview.jsp>. [Last accessed on 2016 Nov 16].

14. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, D.C.: National Academies Press; 2001.
15. Kaboli PJ, GO JT, Hockenberry J, Glasgow JMR, Johnson SR, Rosenthal GE, et al. Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 Veterans Affairs hospitals. *Ann Intern Med* 2012;157:837-45.
16. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for prediction radiation therapy outcomes: A clinician's perspective. *Int J Radiat Oncol Biol Phys* 2015;93:1127-35.
17. Missios S, Bekelis K. Drivers of hospitalization cost after craniotomy for tumor resection: Creation and validation of a predictive model. *BMC Health Serv Res* 2015;15:85.
18. Nikkel LE, Kates SL, Scherck M, Maceroli M, Mahmood B, Elfar JC. Length of hospital stay after hip fracture and risk of early mortality after discharge in New York state: A retrospective cohort study. *Br Med J* 2015;351:h6246.
19. Nordstrom P, Michaelsson K, Hommel A, Norrman PO, Thorngren KG, Nordstrom A. Geriatric rehabilitation and discharge location after hip fracture in relation to the risks of death and readmission. *J Am Med Dir Assoc* 2016;17:91.e1-91.e7.
20. Obermeyer Z, Emmanuel EJ. Predicting the future – big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216-9.
21. Oermann EK, Kress MS, 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* 2013;72:944-51.
22. Reynolds K, Butler MG, Kimes TM, Rosales AG, Chan W, Nichols GA. Relation of acute heart failure hospital length of stay to subsequent readmission and all-cause mortality. *Am J Cardiol* 2015;116:400-5.
23. Sankar A, Beattie WS, Wijesundera DN. How can we identify the high-risk patient? *Curr Opin Crit Care* 2015;21:328-35.
24. Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: A nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J Neurosurg* 2013;118:746-52.
25. Vowpal Wabbit (Fast Learning); Available from: <http://hunch.net/~vw/>. [Last accessed on 2017 Apr 12].
26. Williams CKI, Seeger M. Using the Nystroem method to speed up kernel machines In: Proceedings of the 13th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press; 2000. p. 661-7.
27. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc Series B* 2005;67:301-20.