Improving the Accuracy of Scores to Predict Gastrostomy after Intracerebral Hemorrhage with Machine Learning

Ravi Garg, BS, MSc[1], **Shyam Prabhakaran, MD, MS**[1], **Jane L Holl, MD, MPH**[1], **Yuan Luo, PhD**[2], **Roland Faigle, Md**[3], **Konrad Kording, PhD**[4], and **Andrew M Naidech, mD, MSPH**[1]

[1]Center for Healthcare Studies, University of Pennsylvania

[2]Department of Preventive Medicine, Northwestern University, Chicago, IL, University of Pennsylvania

[3]Department of Neurology, Johns Hopkins University, Baltimore, MD, University of Pennsylvania

[4]Department of Neuroscience, University of Pennsylvania

Abstract

Background.—Gastrostomy placement after intracerebral hemorrhage (ICH) indicates the need for continued medical care and predicts patient dependence. Our objective was to determine the optimal machine learning technique to predict gastrostomy.

Methods.—We included 531 patients in a derivation cohort and 189 patients from another institution for testing. We derived and tested predictions of the likelihood of gastrostomy placement with logistic regression using the GRAVo score (composed of Glasgow Coma Scale <=12, age>50 years, black race, and hematoma volume >30 mL), compared to other machine learning techniques (kth nearest neighbor, support vector machines, random forests, extreme gradient boosting, gradient boosting machine, stacking). Receiver Operating Curves (Area Under the Curve, AUC) between logistic regression (the technique used in GRAVo score development) and other machine learning techniques were compared. Another institution provided an external test data set.

Results.—In the external test data set, logistic regression using the GRAVo score components predicted gastrostomy (P<0.001), however, with a lower AUC (0.66) than kth nearest neighbors (AUC 0.73), random forests (AUC 0.74), Gradient boosting machine (AUC 0.77), extreme gradient boosting (AUC 0.77), (P<0.01 for all compared to logistic regression). Results from the internal test set were similar.

Corresponding Author: Dr. Naidech, 710 N Lake Shore Dr. 11th floor, Chicago, IL 60611, Tel. 312-503-5781, a-naidech@northwestern.edu.

Conclusions.—Machine learning techniques other than logistic regression (e.g., random forests, extreme gradient boost, and kth nearest neighbors) were significantly more accurate for predicting gastrostomy using the same independent variables. Machine learning techniques may assist clinicians in identifying patients likely to need interventions.

Keywords

Intracerebral hemorrhage; gastrostomy; outcomes; machine learning

Introduction

Survivors of intracerebral hemorrhage (ICH), the most morbid form of stroke, often require gastrostomy, a percutaneous feeding tube in the abdomen to provide nutrition [1]. Reliably predicting grastrostomy after ICH is important because gastrostomy placement predicts the need for future healthcare services and patient dependence at follow-up, and unexplained racial disparities in gastrostomy have been noted.[2]

Like outcomes for ICH generally, [3,4] ordinal predictive scores for gastrostomy have been validated,[5,6] including variables that measure the severity of neurologic injury and other established risk factors. The GRAVo score is composed of categorical variables (age over 50 years, black race, Glasgow Coma Scale 12 or less, and hematoma volume more than 30 mL), with higher GRAVo scores predicting increased odds of the patient undergoing gastrostomy in a logistic regression model.[6] Other prediction models of gastrostomy after ICH have identified similar predictors.[7] Predicting outcomes, including gastrostomy, with ordinal scores, however, has suboptimal accuracy. The GRAVo score may not distinguish between components that sum to the same score, but are different (e.g., a black patient over 50 years may have the same score as a patient with reduced Glasgow Coma Scale and a large hematoma, but their outcomes may be different). Therefore, prediction methods other than regression are needed that may more accurately predict the likelihood of a patient undergoing gastrostomy after ICH. Techniques that improve prediction of gastrostomy may be broadly applicable to other procedures, and other diseases.

Machine learning [8] refers to a collection of techniques intended to predict a result from data; regression is the most commonly utilized technique in clinical medicine, typically using ordinal scales.[9,10] Several machine learning techniques inherently account for non-linear predictions, such as proximity-based methods (e.g., kth nearest neighbors, which predicts a classification based on similar patients) and decision-tree based methods (e.g., random forests, which "grows" decision trees and identifies the most significant class). Machine learning techniques have been utilized to predict cardiovascular events in asymptomatic patients [11] and arteriovenous malformations,[12,13] but have not been utilized after ICH. We sought to derive and validate machine learning techniques other than logistic regression to predict gastrostomy (the technique used for GRAVo score development and validation).[6] In this study, we tested the hypothesis that machine learning techniques improve the accuracy of gastrostomy prediction derived from traditional logistic regression in patients with ICH using the components of the GRAVo score as independent variables.

Methods

Patient Identification.

We used prospectively collected data from Johns Hopkins Medicine and Northwestern Medicine. The methods of patient identification have been previously described.[6,14] A board-certified neurologist confirmed the diagnosis of spontaneous ICH, using each patient's head computed tomography results and the appropriate clinical history. Patients with trauma, hemorrhagic conversion of ischemic stroke, or structural lesions (e.g., tumor) were excluded. Gastrostomy was specifically coded.

Standard Protocol Approvals, Registrations, and Patient Consents.

The Northwestern University Institutional Review Board (IRB), and the Johns Hopkins IRB, separately, as previously reported.[6,15]

Statistical Analysis.

We predicted gastrostomy placement with standard machine learning techniques, using previously identified predictors from the GRAVo score [6], a logistic regression model that accounts for age over 50 years, Glasgow Coma Scale (GCS) of 12 or less, black race, and hematoma volume > 30 mL. In addition to logistic regression, the following machine learning techniques were employed: K nearest neighbors algorithm [16] predicts a data point will belong to the same class as its nearest neighbors (k, the number of nearest neighbors, which is typically set between 3 and 7). Support Vector Machines builds a predictive model by mapping example data points in a space where points are divided into separate categories with as wide a gap as possible. Random forest [17] algorithms construct multiple decision trees and identify the mode of the classes predicted across the trees. Gradient boosting [18] adds trees to maximally reduce variability at every step to improve on previously imperfect models. Extreme Gradient Boosting (Xgboost) [19] is built upon gradient boosting and mitigates over-fitting by adding regularization in order to specify the complexity of the model (accurately predicting the derivation data set while performing poorly in test and other data sets). We also performed stacking [20] wherein predictions from these models are used by a meta-classifier (Logistic Regression and Xgboost) to perform further classification. We used the same variables, classifying rules, and weights as specified by the GRAVo model. The Northwestern data set was divided into training (80%) and testing (20%) subsets. We performed 10-fold crossvalidation to find best hyper-parameters for machine learning algorithms and optimize base classifiers. We then tested the trained classifier on the test data set. To generate point estimates and test the significance of the results, we performed bootstrapping with 100 sets of random samples (with replacement) of training and test dataset. Further, the trained machine learning classifiers were applied to the Hopkins data as an additional external test data set. Area Under the Curve (AUC) values from Receiver Operating Characteristic curves for each of the respective techniques were compared. Analyses and graphics were produced using Scikitlearn package (v0.19) available in Python language (v2.7). For illustration purposes, heat maps of the likelihood of undergoing gastrostomy were produced in Stata (v.14, College Station, TX).

Results

Of 544 patients in the Northwestern data, 13 had incomplete data for calculating the GRAVo score, leaving 531 patients, of whom 424 patients (351 without and 73 with gastrostomy) were randomly selected for training, while the remaining 107 patients (89 without and 18 with gastrostomy) were selected for testing. Demographics of the Northwestern data are shown in Table 1. The demographics of the Hopkins data set have been previously published.[6] Thus, the data sets were large enough to apply machine learning techniques.

Logistic regression with the GRAVo score was associated with gastrostomy in the Northwestern data set with moderate accuracy (Figure 1). For example, while both lower Glasgow Coma Scale and older age increase the likelihood of gastrostomy, some black and non-black patients who were the oldest and had the lowest Glasgow Coma Scale did not undergo gastrostomy, even though the logistic regression model predicted the highest likelihood of gastrostomy for such patients.

While testing on Northwestern data without any random sampling, modern machine learning algorithms achieved a higher area under the curve with *k* nearest neighbors (0.73), Random Forests (0.74), Gradient Boosting Machine (0.74), Xgboost (0.76) and stacking with logistic regression (0.74) and stacking with Xgboost (0.74), compared to the GRAVo logistic regression model (0.70) ($P < 0.01$ for all comparisons to logistic regression.

To test the generalizability of the results, we performed bootstrapping with 100 iterations of random sampling of training and test data sets with replacements. We found that decision-tree based methods Random Forests 0.73 (0.718 – 0.742, 95% CI), Gradient Boosting Machine 0.72 (0.706 – 0.733), Xgboost 0.74 (0.729 – 0.752), Stacking using logistic regression 0.73 (0.718 – 0.742) and Stacking using Xgboost 0.72 (0.704 – 0.734) outperformed linear classifiers logistic regression 0.71 (0.700–0.727) and Support Vector Machines 0.54 (0.522–0.565) ($P < 0.01$ for each modern machine learning technique compared to logistic regression).

We observed similar results in the Hopkins (external test) data set, in which all machine learning techniques (other than support vector machines) predicted gastrostomy more accurately than the logistic regression model, using the GRAVo score components as independent variables ($P < 0.01$) (Figure 2).

Discussion

We found that machine learning techniques predicted gastrostomy after ICH with higher accuracy than logistic regression, particularly proximity-based (*k* nearest neighbors) and decision-tree based techniques (e.g., random forests). Machine learning techniques may be an important tool, not only for predicting gastrostomy, but also for other categories of treatments and outcomes, expanding the usefulness of existing data sets to provide new insights on patient management and outcomes.

Logistic regression may not accurately differentiate patients with similar ordinal scores. For example, the GRAVo score implies that patients with the lowest GCS and highest age would

be most likely to undergo gastrostomy placement; however, we found that these patients were unlikely to do so. While one potential explanation is possible (e.g., a patient over 80 years with a GCS of 3 is likely to have limitations placed on medical care),[21,22] this underscores the value of machine learning techniques to account for unanticipated confounders compared to ordinal scales, where specific independent variables are defined and complex prediction rules are not possible. One difference between the Northwestern and Hopkins data sets is that patients with early limitations in medical care were excluded from the derivation and validation of the GRAVo score (Hopkins data set), however, advanced machine learning techniques outperformed logistic regression in both test data sets, so this difference in patient inclusion is unlikely to be significant. There are other occasions where other machine learning techniques are likely to be insightful.

We chose to study gastrostomy placement after ICH because it is reasonably common, predictable, well defined, and has well-described risk factors, as do many other severity of injury scores.[23] An ideal outcome for the derivation and validation of a prediction score would be one that is universally assessed with complete accuracy, not subject to potential bias on the part of clinicians or unintended consequences (e.g. limitations in medical care), and can be easily retrieved from the electronic health record; we are unaware of such an outcome measure. Machine learning techniques may also be particularly helpful for large multi-center or anonymous data sets (e.g., the Nationwide Inpatient Sample) where it is not possible to review the medical record of individual patients for context.

Logistic regression using the GRAVo score was associated with gastrostomy placement overall in the Northwestern data set, like the Hopkins data set.[6] Like many ordinal scores, the major advantage of the GRAVo is that it is practicable for humans to calculate and interpret. No ordinal predictive score can account for all the predictive characteristics and still be practicable. For example, dysphagia is not accounted for in the GRAVo score, but would also be expected to be predictive of gastrostomy placement. As machine learning becomes practicable in electronic health record generally, the advantages of ordinal scores are likely to wane in favor of machine learning techniques. Machine learning is able to accommodate a higher number of independent predictors and may achieve greater accuracy, i.e. by including embedded algorithms for predicting complications. Future research might leverage machine learning techniques that perform well with large data sets, such as random forests. These techniques may allow for new insights to be gleaned from existing data where regression models have been previously utilized and no association was found. Alternatively, Xgboost, random forests, and gradient boosting machine techniques are relatively resistant to "over-fitting" when many potential independent variables are present,[24,25] and run relatively quickly on large data sets.

Improved prediction of gastrostomy is likely to improve the clinical care of patients with ICH. Physicians could be alerted to a high likelihood of a patient requiring gastrostomy by the electronic health record, a widely supported functionality, which could lead to earlier consideration. More timely gastrostomy placement in patients highly likely to require one could lead to shorter length of stay and a reduced risk of complications.[26] Conversely, a low predicted probability of gastrostomy could encourage physicians to delay and reconsider a potentially unneeded procedure. Our algorithms were highly accurate not only derivation

and validation sets at one institution, but also at an independent institution, and were robust to differences in limitations of medical care, underscoring their generalizability.

Machine learning algorithms are likely to have wide applicability. Several machine learning algorithms are already in use, and user-generated algorithms can be uploaded to some electronic health records for use in real time. A potential downside of machine learning algorithms is that they are likely to be less intuitive to humans, e.g., a GRAVo score of 4 is likely to be more intuitive than a prediction from a random forest model seen by clinicians as a "black box."

There are limitations to this investigation. We used the components of the GRAVo score, derived and validated from another institution,[6] for advanced machine learning techniques, but not another published ordinal score for gastrostomy placement[5] because we did not have information on all of the components of the score, such as midline shift on computed tomography scans. The machine learning techniques we employed here are standard and well-described, and might be applicable to other complications and outcomes. Machine learning as a field is rapidly evolving, however, and while we utilized well-described techniques, new techniques may emerge quickly that may require additional software and knowledge of how to apply them.

Summary

We found that machine learning techniques had higher accuracy in predicting gastrostomy after ICH compared to a validated ordinal prediction score using regression, even when using the same score components as independent variables. Machine learning techniques are likely to be useful to predict outcomes that have heretofore been predicted by logistic regression using ordinal scores. Machine learning may be broadly applicable for predicting complications and outcomes, given the ubiquitous need for accurate predictions and risk assessments in clinical medicine.

Acknowledgements.

References

1. Morgenstern LB, Hemphill JC, 3rd, Anderson C, et al. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage. A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2010;41:2108–29. [PubMed: 20651276]

2. Faigle R, Bahouth MN, Urrutia VC, Gottesman RF. Racial and Socioeconomic Disparities in Gastrostomy Tube Placement After Intracerebral Hemorrhage in the United States. Stroke 2016;47:964–70. [PubMed: 26892281]

3. Hemphill JC, Farrant M, Neill TA. Prospective validation of the ICH Score for 12-month functional outcome. Neurology 2009;73:1088–94. [PubMed: 19726752]

4. Hemphill J, Bonovich D, Besmertis L, Manley G, Johnston SC, Tuhrim S. The ICH Score: a simple, reliable grading scale for intracerebral hemorrhage. Stroke 2001;32:891–7. [PubMed: 11283388]

5. Dubin PH, Boehme AK, Siegler JE, et al. New model for predicting surgical feeding tube placement in patients with an acute stroke event. Stroke 2013;44:3232–4. [PubMed: 23963332]

6. Faigle R, Marsh EB, Llinas RH, Urrutia VC, Gottesman RF. Novel score predicting gastrostomy tube placement in intracerebral hemorrhage. Stroke 2015;46:31–6. [PubMed: 25468881]

7. Crisan D, Shaban A, Boehme A, et al. Predictors of recovery of functional swallow after gastrostomy tube placement for Dysphagia in stroke patients after inpatient rehabilitation: a pilot study. Ann Rehabil Med 2014;38:467–75. [PubMed: 25229025]

8. Bishop CM. Pattern recognition and machine learning: springer; 2006.

9. Bath PM, Lees KR, Schellinger PD, et al. Statistical analysis of the primary outcome in acute stroke trials. Stroke 2012;43:1171–8. [PubMed: 22426314]

10. Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke 2007;38:1911–5. [PubMed: 17463316]

11. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circ Res 2017;121:1092–101. [PubMed: 28794054]

12. Asadi H, Kok HK, Looby S, Brennan P, O'Hare A, Thornton J. Outcomes and Complications After Endovascular Treatment of Brain Arteriovenous Malformations: A Prognostication Attempt Using Artificial Intelligence. World Neurosurg 2016;96:562–9.e1. [PubMed: 27693769]

13. Oermann EK, Rubinsteyn A, Ding D, et al. Using a Machine Learning Approach to Predict Outcomes after Radiosurgery for Cerebral Arteriovenous Malformations. Sci Rep 2016;6:21161. [PubMed: 26856372]

14. Naidech AM, Garg RK, Liebling S, et al. Anticonvulsant Use and Outcomes After Intracerebral Hemorrhage. Stroke 2009;40:3810–5. [PubMed: 19797183]

15. Naidech AM, Beaumont JL, Berman M, et al. Dichotomous "Good Outcome" Indicates Mobility More Than Cognitive or Social Quality of Life. Crit Care Med 2015;43:1654–9. [PubMed: 25978337]

16. Peterson LE. K-nearest neighbor. Scholarpedia 2009;4:1883.

17. Liaw A, Wiener M. Classification and regression by random Forest. R news 2002;2:18–22.

18. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics 2001:1189–232.

19. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016: ACM p. 785–94.

20. Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one? Machine learning 2004;54:255–73.

21. Creutzfeldt CJ, Becker KJ, Weinstein JR, et al. Do-not-attempt-resuscitation orders and prognostic models for intraparenchymal hemorrhage. Crit Care Med 2011;39:158–62. [PubMed: 21037471]

22. Becker KJ, Baxter AB, Cohen WA, et al. Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies. Neurology 2001;56:766–72. [PubMed: 11274312]

23. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 1997;277:488–94. [PubMed: 9020274]

24. Bolandzadeh N, Kording K, Salowitz N, et al. Predicting cognitive function from clinical measures of physical function and health status in older adults. PLoS One 2015;10:e0119075. [PubMed: 25734446]

25. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Statis Soc B 2005;67:301–20.

26. Naidech AM, Bendok BR, Tamul P, et al. Medical complications drive lengtlh of stay after brain hemorrhage: a cohort study. Neurocrit Care 2009;10:11–9. [PubMed: 18821035]
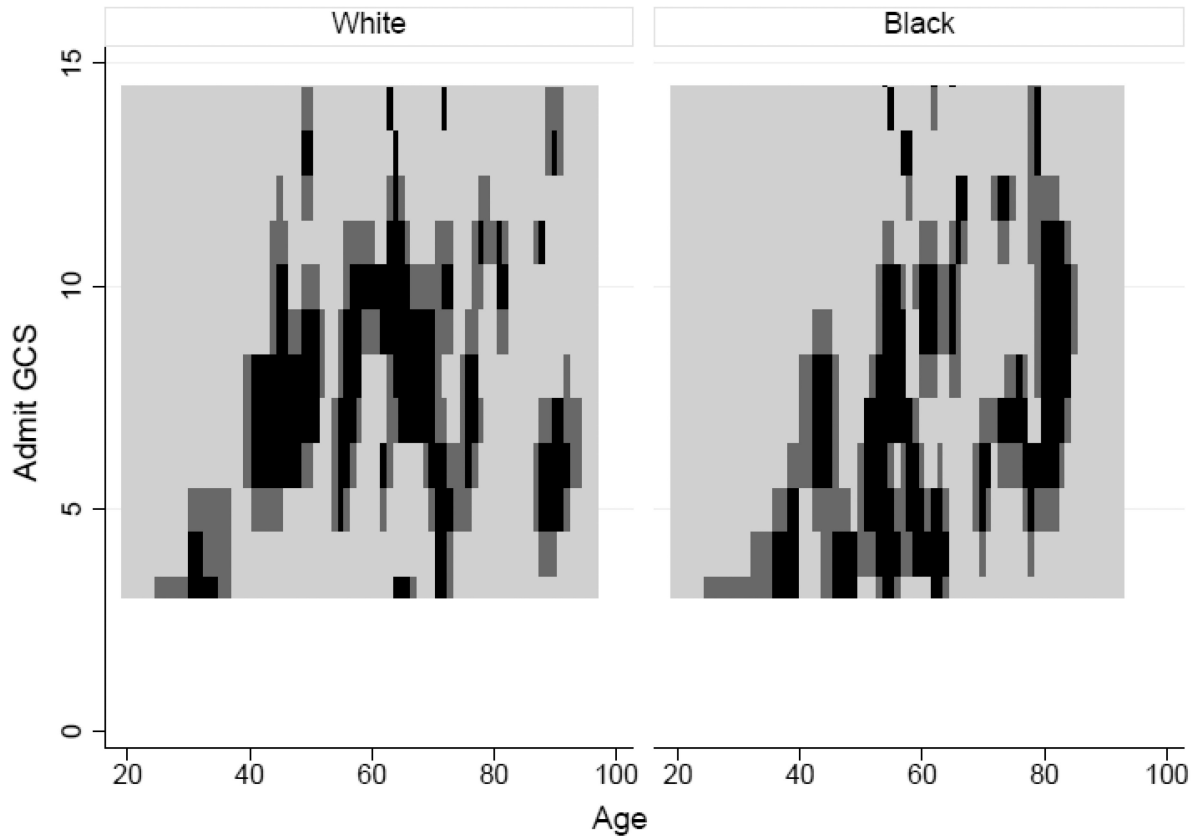
Figure 1.
Heat map of likelihood of undergoing gastrostomy shown in black. Both black and nonblack patients at the extremes of Glasgow Coma Scale and age were unlikely to undergo gastrostomy, although patients at lowest Glasgow Coma Scale and highest age would be predicted to be most likely to undergo gastrostomy.
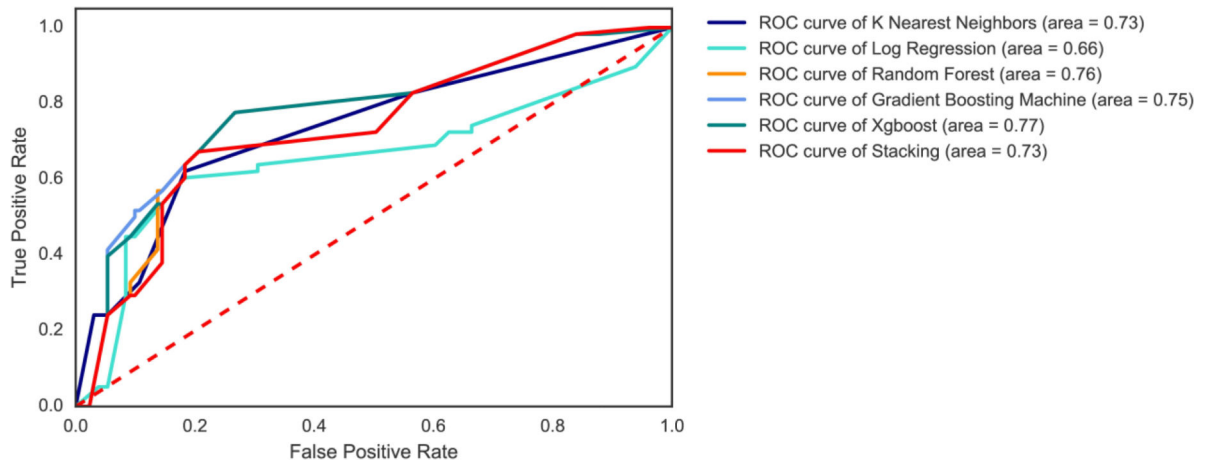
Figure 2.
Comparison of receiver operating characteristic curves for gastrostomy in the Hopkins (external) test data set. Logistic regression had worse performance than other machine learning techniques (P<0.01 for all).

Table 1.

Demographics of 531 patients from the derivation data set.

Variable	N(%), Median [Q1-Q3], or Mean ± SD
Age, years	65.5 ± 14.6
Hematoma volume, mL	12 [5–32]
ICH Score	1 [1 – 3]
Black race	221 (41.6)
Glasgow Coma Scale on admission	13 [8–15]