












# Prediction of Clinical Outcome in Patients with Large-Vessel Acute Ischemic Stroke: Performance of Machine Learning versus SPAN-100

 B. Jiang,  G. Zhu,  Y. Xie,  J.J. Heit,  H. Chen,  Y. Li,  V. Ding,  A. Eskandari,  P. Michel,  G. Zaharchuk, and  M. Wintermark



## ABSTRACT

**BACKGROUND AND PURPOSE:** Traditional statistical models and pretreatment scoring systems have been used to predict the outcome for acute ischemic stroke patients (AIS). Our aim was to select the most relevant features in terms of outcome prediction on the basis of machine learning algorithms for patients with acute ischemic stroke and to compare the performance between multiple models and the Stroke Prognostication Using Age and National Institutes of Health Stroke Scale (SPAN-100) index model.

**MATERIALS AND METHODS:** A retrospective multicenter cohort of 1431 patients with acute ischemic stroke was subdivided into recanalized and nonrecanalized patients. Extreme Gradient Boosting machine learning models were built to predict the mRS score at 90 days using clinical, imaging, combined, and best-performing features. Feature selection was performed using the relative weight and frequency of occurrence in the models. The model with the best performance was compared with the SPAN-100 index model using area under the receiver operating curve analysis.

**RESULTS:** In 3 groups of patients, the baseline NIHSS was the most significant predictor of outcome among all the parameters, with relative weights of 0.36~0.69; ischemic core volume on CTP ranked as the most important imaging biomarker with relative weights of 0.29~0.47. The model with the best-performing features had a better performance than the other machine learning models. The area under the curve of the model with the best-performing features was higher than SPAN-100 model and reached statistical significance for the total ( $P < .05$ ) and the nonrecanalized patients ( $P < .001$ ).

**CONCLUSIONS:** Machine learning–based feature selection can identify parameters with higher performance in outcome prediction. Machine learning models with the best-performing features, especially advanced CTP data, had superior performance of the recovery outcome prediction for patients with stroke at admission in comparison with SPAN-100.

**ABBREVIATIONS:** AIS = acute ischemic stroke; CBS = clot burden score; GBM = Gradient Boosting Machine; IQR = interquartile range; NECT = non-contrast-enhanced CT; SPAN = Stroke Prognostication Using Age and National Institutes of Health Stroke Scale; TIMI = Thrombolysis in Myocardial Infarction; XGB = Extreme Gradient Boosting; AUC = area under the receiver operating curve


Ischemic stroke still ranks as the fifth leading cause of death and the second leading cause of disability in the United States.<sup>1</sup> Although recent reports show a trend toward a decreasing incidence of ischemic stroke for individuals 65 years of age or older, the incidence remains stable for individuals


18~65 years of age.<sup>1</sup> Revascularization therapies such as endovascular thrombectomy have extended the treatment window up to 16–24 hours after symptom onset as demonstrated in selected patients in Endovascular Therapy Following Imaging Evaluation for Ischemic Stroke (DEFUSE 3)<sup>2</sup> and Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention with Trevo (DAWN) trials.<sup>3</sup> However, up to 55% of patients in the endovascular therapy group and 83% in the medical therapy group remained functionally dependent, with 90-day mRS scores of  $>2$ .<sup>2</sup> Therefore, physicians taking care of patients with acute ischemic stroke (AIS) not only need to predict the individual benefit of endovascular treatment but should also be able to estimate prognosis in both treated and untreated patients and to select patients for acute treatment, inform all involved persons about the prognosis, and plan for rehabilitation and long-term care.<sup>4</sup>

Received April 24, 2020; accepted after revision September 12.

From the Department of Radiology, Neuroradiology Section (B.J., G.Z., Y.X., J.J.H., H.C., Y.L., G.Z., M.W.), Stanford University School of Medicine, Palo Alto, California; Department of Medicine (V.D.), Quantitative Sciences Unit, Stanford University, Stanford, California; and Neurology Service (A.E., P.M.), Centre Hospitalier Universitaire Vaudois and Lausanne University, Lausanne, Switzerland.

Please address correspondence to Max Wintermark, MD, Stanford University, Department of Radiology, Neuroradiology Division, 300 Pasteur Dr, Room 5047, Stanford, CA 94305-5105; e-mail: max.wintermark@gmail.com

 Indicates article with supplemental online tables.

 Indicates article with supplemental online photos.

<http://dx.doi.org/10.3174/ajnr.A6918>

Many publications have addressed the issues of predicting outcome in patients with acute large-vessel ischemic stroke. These include (but are not limited to) traditional logistic regression statistical models and pretreatment scoring systems such as the DRAGON score (Dense cerebral artery sign/early infarct signs on admission CT scan, prestroke modified Rankin Scale, Age, Glucose level at baseline, Onset-to-treatment time, and baseline National Institutes of Health Stroke Scale score),<sup>5-7</sup> the Stroke Prognostication Using Age and National Institutes of Health Stroke Scale (SPAN-100) index,<sup>8,9</sup> the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score,<sup>7</sup> the Pittsburgh Response to Endovascular Therapy (PRE) score,<sup>10</sup> the Total Health Risks in Vascular Events (THRIVE) score,<sup>11</sup> the Houston Intra-Arterial Therapy (HIAT) score, and the HIAT2 score.<sup>12</sup> The components considered in these predicting scoring systems were either clinical parameters only such as age and the NIHSS or non-contrast-enhanced CT (NECT) parameters such as ASPECTS. None of these models take into account advanced imaging parameters. In addition, these models were built on the basis of the hypothesis of a linear relationship between the parameters and the outcome, but some studies have highlighted a nonlinear correlation.<sup>13,14</sup>

In comparison with traditional modeling methods, machine learning algorithms have much higher scalability, allowing large numbers of features and parameters to be incorporated into the models. Machine learning models have been trained not only for outcome prediction following intravenous thrombolysis<sup>15</sup> and intra-arterial therapy<sup>16,17</sup> after AIS but also for subtype classification,<sup>18</sup> hemorrhagic transformation,<sup>19</sup> and clot-characteristic identification.<sup>20</sup> All the above-mentioned models use clinical features as input; 2 studies also used baseline NECT<sup>14,16</sup> or MR imaging gradient recalled-echo sequence features,<sup>20</sup> and 1 study used MR perfusion.<sup>19</sup>

The hypothesis of our study was that machine learning algorithms can help select the most powerful features in outcome prediction, and the model with features from advanced perfusion CTP data would have more robust prognostic ability in comparison with the other machine learning models and SPAN-100 model.<sup>9</sup>

## **MATERIALS AND METHODS**

### **Study Population**

This retrospective study was conducted using a registry of 1782 patients with AIS from January 2008 to December 2018 at the Lausanne University Hospital (1310 patients) and Stanford University (472 patients). Institutional review board approval was obtained from both institutional review boards, with a waiver of informed consent due to the retrospective nature of the study. Inclusion criteria were the following: 18 years of age or older; clinical examination and baseline CT imaging confirming acute ischemic infarction with the infarct area within the ICA/MCA territory; availability of complete clinical (onset-to-baseline time; baseline NIHSS; glucose, lipid, and blood pressure levels at admission; history of cardiac disease, statin use, smoking status; stroke mechanism according to the Trial of Org 10172 in Acute Stroke Treatment [TOAST] trial;<sup>21</sup> and treatment and 90-day mRS) and imaging parameters (baseline NECT, CTP, and CTA);

early [ $<72$  hours from baseline] recanalization CTA). Patients with subacute, chronic, remote, and/or hemorrhagic infarctions were excluded from this study. The type of revascularization treatment (intravenous thrombolysis and endovascular treatment) was recorded if performed on the basis of the treating physician's decision.

### **Initial Clinical and Imaging Data**

All the clinical and imaging parameters assessed in our study are summarized in Online Table 1. The 90-day mRS was dichotomized into favorable (mRS 0–2) and unfavorable outcome (mRS 3–6).

NECT, CTP, and CTA data were collected at admission as baseline studies. A blinded neuroradiologist evaluated the imaging features for all of the imaging studies. Features including the ASPECTS and hyperdense middle cerebral artery sign were extracted from the NECT. CTP datasets were processed on a workstation (Brain Perfusion, Version 6.0.0; Philips Healthcare). Automatic segmentation of ischemic core and penumbra volumes was performed on the basis of previously published thresholds.<sup>22</sup> The sidedness of cerebral ischemia was evaluated as well. The site of occlusion, Thrombolysis in Myocardial Infarction (TIMI) score, and collateral status were interpreted on the MIP CTA images. The TIMI<sup>23</sup> score was assessed as follows: 0, complete occlusion; 1, subocclusion with no distal branch filling; 2, subocclusion with incomplete or slow distal branch filling; and 3, completely open artery. A previously reported scoring system<sup>24</sup> was used for grading the collaterals into 4 levels in comparison with the normal side on baseline CTA. In addition, the clot burden score<sup>25</sup> (CBS), reflecting the extent of intracranial clot, and degree of stenosis of the carotid bifurcation according to the NASCET criteria were assessed on baseline CTA images. The total cohort was divided into 2 subgroups depending on the recanalization status. A TIMI score of  $\geq 2$  on recanalization studies was considered recanalization, while  $<2$  was considered persistent arterial occlusion.

### **Model Construction**

Our dataset had 2 distinctive characteristics: low dimensionality with  $<100$  features and high nonlinearity for both qualitative and quantitative clinical/imaging features. We, therefore, decided to use Extreme Gradient Boosting (XGB), which is a specialized Gradient Boosting Machine (GBM), for our dataset. There are 2 core elements of the GBM. The first is a decision tree, which is the approach to generate and approximate non-linear-relationship mapping between input features and final outcome. The second is boosting. Initially raised by the authors of Adaptive Boosting (AdaBoost),<sup>26</sup> the concept of boosting consists of first creating many weaker, simpler machine learning classifiers during training. Then, the final model is constructed by pooling the results from all weaker models and creating a fine-tuned, stronger classifier. XGB was developed on the basis of the GBM with superiority of performance in multiple data science contests, and its multicore algorithms allow multiple computations to run simultaneously in parallel, thus enabling the algorithm to scale to large datasets.<sup>27</sup>

A previous study<sup>28</sup> using GBM demonstrated that machine learning methods with decision tree and boosting algorithms were capable of predicting patient outcomes after AIS. In that study, both XGB and GBM were used, and XGB was found to have a relatively better performance when the cohort was divided into subgroups. XGB was also shown to perform very well in another study when segmenting stroke infarct regions using both clinical and imaging features.<sup>29</sup>

Sixteen clinical and 11 imaging parameters were introduced in our models (Online Table 1). The dataset was broken down into 5 groups with a relatively equal number of patients in each group for 5-fold cross-validations. Data of each patient were randomly enrolled into 1 of the 5 folds as a testing set. In the remaining 4 folds, the patient data were used as a training set. For each model's training and testing phase, 5 identical models were trained, each using 1 group as the test set, with the remaining 4 groups as a training set. Then the overall model performance was evaluated on the basis of results from all 5 models on 5 test sets. At first, 3 types of feature group combinations, clinical features, imaging features, and clinical plus imaging features, were used in the XGB models to predict the 90-day mRS of the entire cohort and recanalized and nonrecanalized subgroups, respectively, creating 9 total combinations. To improve the performance of the machine learning models, we selected a subset of clinical and imaging features from all the predictors according to their contributions to the models. Features were selected on the basis of the following criteria: They had a relative weight of  $\geq 0.2$  or a relative weight of  $\geq 0.1$  and were in the top 5 highest weights in the 9 above-mentioned models. The SPAN-100 XGB model was built by introducing age and the NIHSS at admission based on the definition.

### Statistical Analysis

Overall and by recanalization status, continuous characters were summarized as medians and interquartile ranges (IQRs) and as counts and percentages for categoric characters. For each of the 3 cohorts, measures of prediction sensitivity, specificity, accuracy, and area under the receiver operating curve (AUC) were estimated for the machine learning models, as well as for the reference SPAN-100 index model, with SPAN-100 defined as the sum of patient age and the NIHSS score.<sup>9</sup> The machine learning model with the highest AUC was then compared with the SPAN-100 index model, with the Delong test of pair-wise AUCs assessed using the pROC R package (<https://www.rdocumentation.org/packages/pROC/versions/1.16.2>).<sup>30,31</sup>

Finally, confusion matrices for 90-day mRS prediction were constructed, by cohort, for all models on the basis of 7-fold cross-validation and visualized as heatmaps. All analyses were conducted in the R statistical computing framework,<sup>32</sup> Version 3.6 (<http://www.r-project.org/>), and statistical significance was assessed at the .05  $\alpha$  level.

## RESULTS

There were 1431 patients included in this study, including 899 patients with recanalization and 532 patients with no recanalization (Online Fig 1). Online Table 1 illustrates the clinical and

imaging characteristics for the total cohort and for the 2 subgroups.

### Feature Selection with Machine Learning

Among the clinical and imaging parameters, the baseline NIHSS was the most important predictor of outcome for the whole cohort, as well as in the recanalized and nonrecanalized groups, with relative weights ranging from 0.36 to 0.69. Age and glucose levels at admission ranked as the next most important parameters in both the model using only clinical parameters and the model using all the clinical and imaging parameters (Online Table 2). The NIHSS and age are both components of the SPAN-100 scoring system.

Among the imaging parameters, ischemic core volume on CTP came in first place for all 3 groups of patients, with relative weights of 0.29~0.47 (Online Table 2). The CTA-CBS score, penumbra volume on CTP, and infarct side were the second strongest imaging predictors in the full cohort, the recanalized patients, and the nonrecanalized patients, respectively.

Clinical features such as baseline NIHSS score and age outweighed all the imaging features in importance in all 3 groups. Glucose level at admission appeared to be the third most important clinical biomarker in the total cohort and in recanalized patients, but not in nonrecanalized patients. In the nonrecanalized group, infarct and penumbra volume on CTP and time from onset to the baseline study came before the glucose level. Accordingly, the model with the best-performing features (total of 6 features) was built by including 3 clinical features (baseline NIHSS, age, glucose at admission) and 3 imaging features (ischemic core volume on CTP, penumbra volume on CTP, and CTA-CBS) (Online Table 3).

### Model Performance in the Full Cohort and Recanalized and Nonrecanalized Cohorts

The sensitivity, specificity, accuracy, AUC, and heatmap of each model in the full cohort, as well as in the recanalized and the nonrecanalized subgroups are demonstrated in the Table, Figure, and Online Fig 2. The models with both imaging and clinical features performed better than those with only clinical or imaging input. The model with 6 features performed better than models with clinical features only, models with imaging features only, and models with both clinical and imaging features. This finding was true in all 3 groups of participants, with the highest AUC value of 0.83 for the nonrecanalized patients.

### Comparison between Machine Learning Models and the SPAN Scoring Model

Our best model, the model with the best-performing features, was compared with the SPAN-100 index (Figure and Online Fig 2). The AUCs for the machine learning models with the 6 best-performing features in the total cohort and recanalized and nonrecanalized groups were 0.80, 0.79, and 0.82, respectively. The AUCs for SPAN-100 were 0.78, 0.76, and 0.78, respectively. The optimal cutoff values of SPAN-100 were 85, 94, and 64 for the total, recanalized, and nonrecanalized cohorts, respectively. The AUCs of the XGB models with the 6 best-performing features were higher than those of SPAN-100 and reached the

## Performance of machine learning models and the SPAN-100 index in 3 cohorts

Cohorts/Models	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Full cohort				
Clinical features only (16 features)	78.1	65.5	73.5	0.77
Imaging features only (11 features)	53.5	79.9	63.2	0.69
Both clinical and imaging features (27 features)	74.4	69.8	72.8	0.79
Best-performing clinical and imaging features (6 features)	72.2	74.0	72.8	0.80 <sup>a</sup>
SPAN-100	80.6	64.3	73.5	0.78
Recanalized				
Clinical features only (16 features)	73.1	70.4	71.9	0.76
Imaging features only (11 features)	53.7	69.4	60.9	0.61
Both clinical and imaging features (27 features)	74.5	68.9	72.0	0.77
Best-performing clinical and imaging features (6 features)	76.9	69.9	73.8	0.79 <sup>a</sup>
SPAN-100	78.8	63.8	71.9	0.76
Nonrecanalized				
Clinical features only (16 features)	80.0	65.8	74.1	0.78
Imaging features only (11 features)	63.3	67.8	64.3	0.70
Both clinical and imaging features (27 features)	71.3	80.5	73.3	0.81
Best-performing clinical and imaging features (6 features)	81.9	75.4	80.5	0.82 <sup>a</sup>
SPAN-100	65.5	77.1	68.1	0.78

<sup>a</sup> Model with the highest AUC value.

statistical significance for the total cohort ( $P < .05$ ) and the nonrecanalized patients ( $P < .001$ ). In the recanalized group, the difference was not significant ( $P = .05$ ).

## DISCUSSION

Our study shows that machine learning models trained with best-performing clinical and imaging features, including advanced CTP parameters, can predict the outcome of patients with stroke more accurately than a conventional scoring system.

Bacchi et al<sup>33</sup> used deep learning models to predict the outcome in patients with AIS who underwent intravenous thrombolysis. The combined convolutional-plus-artificial neural network model based on both clinical and imaging data performed best in predicting patient outcomes. Heo et al<sup>34</sup> attempted to predict favorable outcome in a large group of 2043 patients with stroke using 3 machine learning models. By incorporating 38 demographic/clinical variables into their models, they found that the deep neural network model performed better than the other 2 models (random forest and logistic regression) and the ASTRAL score, while the performance of the deep neural network did not differ significantly from the ASTRAL score when trained on only the same 6 variables used for calculating the ASTRAL score. Nishi et al<sup>17</sup> built 9 models, including 5 previously reported scoring models, 1 logistic regression statistical model, and 3 machine learning models to predict the clinical outcome in a cohort of 387 patients with stroke who underwent endovascular treatment. Machine learning models were superior to the other models. These above-mentioned models used ASPECTS as the only imaging variable to make the outcome prediction, and the overwhelming clinical variables in these models seemed not quite practical in an emergency scenario because a physician has to input many variables to get valuable prognostic information. Our models with the best-performing features were trained on more advanced imaging data such as CTP and CTA parameters, which provide improved accuracy compared with models using only parameters from the NECT. Furthermore, clinical features are

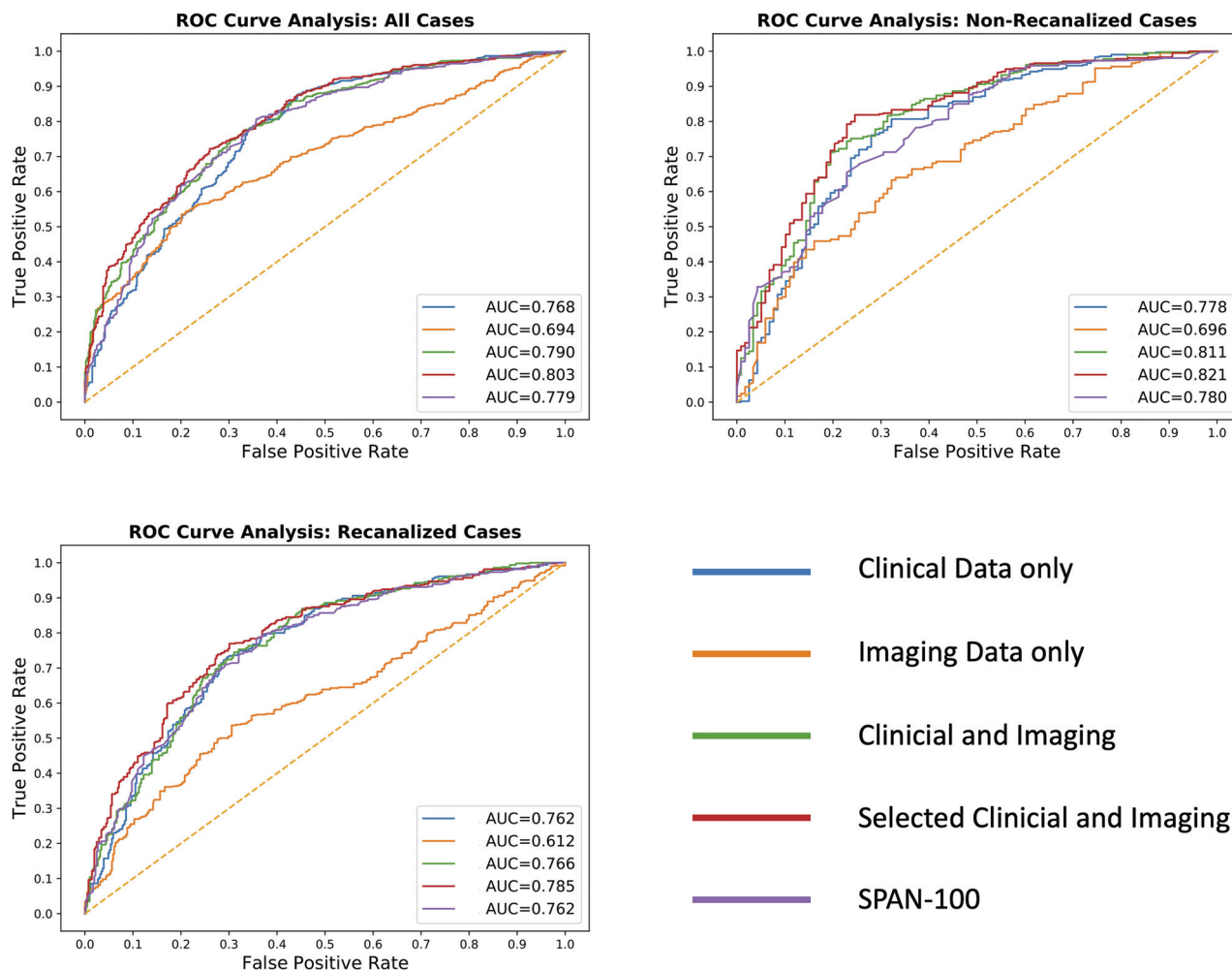
important predictors, but when they are broken down into recanalized and nonrecanalized groups, CTP imaging data were a more potent contributor, especially for those nonrecanalized patients.

The commonly used machine learning models in cerebrovascular diseases include random forest, support-vector machines, the neural network, decision trees, and logistic regression. In this study, we used a supervised XGB model, which is a decision tree-based machine learning method. Previous publications<sup>28,29,35</sup> highlighted the adaptability of XGB in dealing with redundant and nonlinear datasets. Compared with other machine learning models, XGB makes more powerful predictions with less chance of overfitting, especially in predictions of binary outcomes.

Our modeling filtered 6 parameters that best predicted the 90-day mRS score. Baseline NIHSS, age, and glucose on admission are clinical components of most of the conventional pretreatment prognostic systems developed for patients with stroke.<sup>5-12</sup> Previous studies have shown that baseline NIHSS and age are strongly associated with prognosis.<sup>13,36,37</sup> Hyperglycemia on admission is known to be an independent predictor of worse outcome because of its association with lactic acidosis and accelerated conversion of penumbra to infarct.<sup>38,39</sup> The relevant imaging features (CTP ischemic core volume, penumbra volume, and CTA-CBS) are also well-established stroke imaging biomarkers.<sup>13</sup> Collateral scores and the CBS have been reported to be equally important in outcome prediction.<sup>40</sup> In our study, collaterals played an important role in the recanalized group, but not in the nonrecanalized group.

It is beneficial to have a simple model because it makes clinical deployment faster and easier. A model requiring few features to yield a useful prediction is also less prone to overfitting. In addition, the 3 imaging features used in our model can be automatically extracted within a machine learning pipeline embedded in the daily workflow. It is practical for our best-performing model to provide a prompt outcome prediction.

The SPAN-100 index has been shown to have the ability to predict patient outcome and the risk of complications after endovascular therapy in several stroke cohorts.<sup>9,13</sup> Möbius et al<sup>8</sup> found



**FIGURE.** Receiver operating characteristics (ROCs) of XGB prediction models with clinical features, imaging features, both clinical and imaging features, best-performing features, and SPAN-100 for predicting a 90-day mRS score of >2. For all patients and recanalized and nonrecanalized patients, the AUCs of models with the best-performing features were higher than those in SPAN-100, and statistical significance was reached in the total and nonrecanalized groups. The AUCs for machine learning models with the 6 best-performing features in the total cohort and recanalized and nonrecanalized groups were 0.80, 0.79, and 0.82, respectively. The AUCs for SPAN-100 were 0.78, 0.76, and 0.78, respectively. The AUCs of XGB models with the best-performing features were higher than those in SPAN-100 and reached statistical significance for the total cohort ( $P < .05$ ) and the nonrecanalized patients ( $P < .001$ ). In the recanalized group, the difference was not significant ( $P = .05$ ).

that the patients positive on the basis of SPAN-100 demonstrated a 9-fold increase in the odds ratio of poor outcome compared with those negative on the basis of SPAN-100, with an AUC of 0.74. The NIHSS ranked as the most highly relevant parameter among all of the clinical and imaging biomarkers in our study, while age was the second-best predictor in nonrecanalized patients and the third-best predictor in all and recanalized patients. When combined with imaging features, the ability of outcome prediction improved from 0.78, 0.76, and 0.78 to 0.80, 0.79, and 0.82 for all and recanalized and nonrecanalized patients. The major limitation of SPAN-100 is its inapplicability to younger patients, for it cannot reach a positive status because of the age component. However, our model overcomes this limitation and is applicable to any patient with AIS older than 18 years of age.

There are several limitations to this study. First, this was a retrospective study, and our model will need to be validated

prospectively. Second, we used only XGB models in this machine learning study, and other machine learning algorithms need be considered in future study designs. Third, prognostic models other than the SPAN-100 may have superior long-term predictive values for handicap and mortality, which will be incorporated into our future study design.<sup>41</sup>

## CONCLUSIONS

Machine learning-based feature selection can identify parameters with higher performance in long-term recovery-outcome prediction for patients with stroke at admission, while removing redundant and less predictive parameters. Moreover, the models with input from the best-performing features had better predictive value than the other models using clinical features only, imaging features only, both clinical and imaging features, and the SPAN-100 index. Finally, the prognostic

ability of machine learning models with advanced imaging features such as CTP data can be improved, especially for nonrecanalized patients.

Disclosures: Yuan Xie—UNRELATED: Employment: Subtle Medical Inc; Stock/Stock Options: Subtle Medical Inc. Jeremy J. Heit—UNRELATED: Board Membership: iSchemaView, Comments: Medical and Scientific Advisory Board member; Consultancy: Medtronic and MicroVention. Patrik Michel—RELATED: Grant: Swiss National Science Foundation, Swiss Heart Foundation\*; Consulting Fee or Honorarium: Medtronic\*; Other Relationships: Steering Committees/Data Safety and Monitoring Board of BASICS, ELAN, International PFO Consortium PROMISE, CLOSE (no payments). Greg Zaharchuk—UNRELATED: Board Membership: Subtle Medical Inc; Consultancy: Subtle Medical Inc; Grants/Grants Pending: various National Institutes of Health projects, GE Healthcare, Bayer AG\*; Royalties: Cambridge University Press; Stock/Stock Options: Equity, Subtle Medical Inc. Max Wintermark—UNRELATED: Consultancy: MORE Health, Magnetic Insight, Subtle Medical Inc, icometrix, Nous. \*Money paid to the institution.

## REFERENCES

1. Benjamin EJ, Muntner P, Alonso A, et al; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. **Heart disease and stroke statistics-2019 update: a report from the American Heart Association.** *Circulation* 2019;139:e56–528 [CrossRef Medline](#)
2. Albers GW, Marks MP, Kemp S, et al; DEFUSE 3 Investigators. **Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging.** *N Engl J Med* 2018;378:708–18 [CrossRef Medline](#)
3. Nogueira RG, Jadhav AP, Haussen DC, et al; DAWN Trial Investigators. **Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct.** *N Engl J Med* 2018;378:11–21 [CrossRef Medline](#)
4. Winzeck S, Hakim A, McKinley R, et al. **ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI.** *Front Neurol* 2018;9:679 [CrossRef Medline](#)
5. Turc G, Apoil M, Nagara O, et al. **Magnetic resonance imaging-dragon score: 3-month outcome prediction after intravenous thrombolysis for anterior circulation stroke.** *Stroke* 2013;44:1323–28 [CrossRef Medline](#)
6. Strbian D, Seiffge DJ, Breuer L, et al. **Validation of the dragon score in 12 stroke centers in anterior and posterior circulation.** *Stroke* 2013;44:2718–21 [CrossRef Medline](#)
7. Cooray C, Mazya M, Bottai M, et al. **External validation of the astral and dragon scores for prediction of functional outcome in stroke.** *Stroke* 2016;47:1493–99 [CrossRef Medline](#)
8. Möbius C, Blinzler C, Schwab S, et al. **Re-evaluation of the Stroke Prognostication Using Age and NIH Stroke Scale index (SPAN-100 index) in IVT patients: the-SPAN 100(65) index.** *BMC Neurol* 2018;18:129 [CrossRef Medline](#)
9. Saposnik G, Guzik AK, Reeves M, et al. **Stroke prognostication using age and NIH Stroke Scale: SPAN-100.** *Neurology* 2013;80:21–28 [CrossRef Medline](#)
10. Rangaraju S, Aghaebrahim A, Streib C, et al. **Pittsburgh Response to Endovascular Therapy (PRE) score: optimizing patient selection for endovascular therapy for large vessel occlusion strokes.** *J NeuroIntervent Surg* 2015;7:783–88 [CrossRef Medline](#)
11. Flint AC, Faigeles BS, Cullen SP, et al; VISTA Collaboration. **Thrive score predicts ischemic stroke outcomes and thrombolytic hemorrhage risk in VISTA.** *Stroke* 2013;44:3365–69 [CrossRef Medline](#)
12. Sarraj A, Albright K, Barreto AD, et al. **Optimizing prediction scores for poor outcome after intra-arterial therapy in anterior circulation acute ischemic stroke.** *Stroke* 2013;44:3324–30 [CrossRef Medline](#)
13. Jiang B, Ball RL, Michel P, et al. **Factors influencing infarct growth including collateral status assessed using computed tomography in acute patients with stroke with large artery occlusion.** *Int J Stroke* 2019;14:603–12 [CrossRef Medline](#)
14. Hakimelahi R, Vachha BA, Copen WA, et al. **Time and diffusion lesion size in major anterior circulation ischemic strokes.** *Stroke* 2014;45:2936–41 [CrossRef Medline](#)
15. Bentley P, Ganesalingam J, Carlton Jones AL, et al. **Prediction of stroke thrombolysis outcome using CT brain machine learning.** *Neuroimage Clin* 2014;4:635–40 [CrossRef Medline](#)
16. Asadi H, Dowling R, Yan B, et al. **Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy.** *PLoS One* 2014;9:e88225 [CrossRef Medline](#)
17. Nishi H, Oishi N, Ishii A, et al. **Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning.** *Stroke* 2019;50:2379–88 [CrossRef Medline](#)
18. Garg R, Oh E, Naidech A, et al. **Automating ischemic stroke subtype classification using machine learning and natural language processing.** *J Stroke Cerebrovasc Dis* 2019;28:2045–51 [CrossRef Medline](#)
19. Yu Y, Guo D, Lou M, et al. **Prediction of hemorrhagic transformation severity in acute stroke from source perfusion MRI.** *IEEE Trans Biomed Eng* 2018;65:2058–65 [CrossRef Medline](#)
20. Chung JW, Kim YC, Cha J, et al. **Characterization of clot composition in acute cerebral infarct using machine learning techniques.** *Ann Clin Transl Neurol* 2019;6:739–47 [CrossRef Medline](#)
21. Adams HP Jr, Bendixen BH, Kappelle LJ, et al. **Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial—TOAST. Trial of Org 10172 in Acute Stroke Treatment.** *Stroke* 1993;24:35–41 [CrossRef Medline](#)
22. Wintermark M, Flanders AE, Velthuis B, et al. **Perfusion-CT assessment of infarct core and penumbra: receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke.** *Stroke* 2006;37:979–85 [CrossRef Medline](#)
23. TIMI Study Group. **The Thrombolysis in Myocardial Infarction (TIMI) trial: Phase I findings.** *N Engl J Med* 1985;312:932–36 [CrossRef Medline](#)
24. Tan JC, Dillon WP, Liu S, et al. **Systematic comparison of perfusion-CT and CT-angiography in acute patients with stroke.** *Ann Neurol* 2007;61:533–43 [CrossRef Medline](#)
25. Puetz V, Dzialowski I, Hill MD, et al; Calgary CTA Study Group. **Intracranial thrombus extent predicts clinical outcome, final infarct size and hemorrhagic transformation in ischemic stroke: the clot burden score.** *Int J Stroke* 2008;3:230–36 [CrossRef Medline](#)
26. Freund Y. **An adaptive version of the boost by majority algorithm.** *Machine Learning* 2001;43:293–18. file:///C:/Users/mrudi/Downloads/Freund2001\_Article\_AnAdaptiveVersionOfTheBoostByM.pdf. Accessed January 18, 2019
27. Chen T, Guestrin C. **Xgboost: a scalable tree boosting system.** In: *KDD '16: Proceedings of the 22nd Association for Computing Machinery Special Interest Group for Knowledge Discovery from Data International Conference on Knowledge Discovery and Data Mining.* San Francisco, California; August 2016:785–94
28. Xie Y, Jiang B, Gong E, et al. **Journal club: use of Gradient Boosting Machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information.** *AJR Am J Roentgenol* 2019;212:44–51 [CrossRef Medline](#)
29. Livne M, Boldsen JK, Mikkelsen IK, et al. **Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke.** *Stroke* 2018;49:912–18 [CrossRef Medline](#)
30. DeLong ER, DeLong DM, Clarke-Pearson DL. **Comparing the areas under two or more correlated receiver operation characteristic curves: a nonparametric approach.** *Biometrics* 1988;44:837 [CrossRef Medline](#)
31. Xavier Robin NT, Hainard A, Tiberti N, et al. **Proc: An open-source package for R and S+ to analyze and compare roc curves.** *BMC Bioinformatics* 2011;12:77 [CrossRef](#)
32. The R Core Team. *A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing; 1999–2012
33. Bacchi S, Zerner T, Oakden-Rayner L, et al. **Deep learning in the prediction of ischaemic stroke thrombolysis functional**

- outcomes: a pilot study.** *Acad Radiology* 2019;27:e19–23 [CrossRef Medline](#)
34. Heo J, Yoon JG, Park H, et al. **Machine learning-based model for prediction of outcomes in acute stroke.** *Stroke* 2019;50:1263–65 [CrossRef Medline](#)
35. Torlay L, Perrone-Bertolotti M, Thomas E, et al. **Machine learning-XGBoost analysis of language networks to classify patients with epilepsy.** *Brain Inform* 2017;4:159–69 [CrossRef Medline](#)
36. Adams HP Jr, Davis PH, Leira EC, et al. **Baseline NIH stroke scale score strongly predicts outcome after stroke: a report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST).** *Neurology* 1999;53:126–31 [CrossRef](#)
37. Hankey GJ, Spiesser J, Hakimi Z, et al. **Rate, degree, and predictors of recovery from disability following ischemic stroke.** *Neurology* 2007;68:1583–87 [CrossRef Medline](#)
38. Weir CJ, Murray GD, Dyker AG, et al. **Is hyperglycaemia an independent predictor of poor outcome after acute stroke? Results of a long-term follow-up study.** *BMJ* 1997;314:1303–06 [CrossRef Medline](#)
39. Parsons MW, Barber PA, Desmond PM, et al. **Acute hyperglycemia adversely affects stroke outcome: a magnetic resonance imaging and spectroscopy study.** *Ann Neurol* 2002;52:20–28 [CrossRef Medline](#)
40. Tan IY, Demchuk AM, Hopyan J, et al. **CT angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct.** *AJNR Am J Neuroradiol* 2009;30:525–31 [CrossRef Medline](#)
41. Quinn TJ, Singh S, Lees KR, et al; VISTA Collaborators. **Validating and comparing stroke prognosis scales.** *Neurology* 2017;89:997–1002 [CrossRef Medline](#)