medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Title

Potential of machine learning to predict early ischemic events after carotid endarterectomy or stenting: A comparison with surgeon predictions

Authors

Kazuya Matsuo [1], Atsushi Fujita [1], Kohkichi Hosoda [2], Jun Tanaka [3], Taichiro Imahori [4], Taiji Ishii [5], Masaaki Kohta [1], Kazuhiro Tanaka [1], Yoichi Uozumi [1], Hidehito Kimura [1], Takashi Sasayama [1], Eiji Kohmura [1]

Affiliations

[1] Department of Neurosurgery, Kobe University Graduate School of Medicine, Kobe, Japan

[2] Department of Neurosurgery, Kobe City Nishi-Kobe Medical Center, Kobe, Japan

[3] Department of Neurosurgery, Konan Hospital, Kobe, Japan

[4] Department of Neurosurgery, Hyogo Brain and Heart Center at Himeji, Himeji, Japan

[5] Department of Neurosurgery, Toyooka Hospital, Toyooka, Japan

*Correspondence: Kazuya Matsuo, MD, PhD, Department of Neurosurgery, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Tel: +81-78-382-5966

Fax: +81-78-382-5979

E-mail: kkmatsuo@outlook.jp

Abstract

Background

Carotid endarterectomy (CEA) and carotid artery stenting (CAS) are recommended for high stroke-risk patients with carotid artery stenosis to reduce ischemic events. However, we often face difficulty in determining the best treatment method. In this study, we aimed to develop an accurate post-CEA/CAS outcome prediction model using machine learning (ML) algorithms that will serve as a basis for a new decision support tool for patient-specific treatment planning.

Methods

Retrospectively collected data from 165 consecutive patients with carotid stenosis underwent CEA or CAS were divided into training and test samples. The following six ML algorithms were tuned and their predictive performance evaluated by comparison with surgeon predictions: an artificial neural network, logistic regression, support vector machine, Gaussian naïve Bayes, random forest, and extreme gradient boosting (XGBoost). Seventeen clinical factors were introduced into the models, and outcome was defined as any ischemic stroke within 30 days after

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

treatment.

Results

The XGBoost model performed the best in the evaluation; its sensitivity, specificity, positive predictive value, and accuracy were 66.7%, 89.5%, 50.0%, and 86.4%, respectively. All statistical measures are comparable with those of surgeons. Internal carotid artery peak systolic velocity, low density lipoprotein cholesterol, and procedure (CEA or CAS) were the most contributing factors according to the XGBoost algorithm.

Conclusion

We were able to develop a post-procedural outcome prediction model comparable to surgeons in performance. The accurate outcome prediction model will make it possible to make a more appropriate patient-specific selection of CEA or CAS for the treatment of carotid stenosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Carotid artery stenosis is a main cause of ischemic stroke, which remains a major public health problem worldwide.[1] To reduce the risk of ischemic stroke, carotid endarterectomy (CEA) and carotid artery stenting (CAS) are recommended for patients at high stroke risk with carotid artery stenosis. Based on the evidence that there are no significant differences in long-term outcomes after CEA and CAS,[2-4] there are general guidelines for patient selection for CEA and CAS.[4-6] However, we often face difficulty in determining the best treatment method. Therefore, it is necessary to develop a useful decision support tool to identify an appropriate patient-specific treatment for carotid artery stenosis. Recently, artificial intelligence (AI) has been used widely in medical fields because of tremendous advances in computer technology including robust machine learning (ML) algorithms for successful prediction or diagnosis.[7] However, no studies have investigated the effects of modern ML models on any prediction or assessment for carotid artery stenosis.

In this study, we aimed to develop an accurate ML model for outcome prediction within 30 days after CEA or CAS with 17 clinical factors. The usefulness of the ML models was evaluated by comparing their predictions with those of surgeons. Because early periprocedural major and minor stroke has been known to affect outcomes at 1 year,[8] our post-procedural outcome prediction model can serve as the basis for an effective decision support tool for patient-specific

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

adaptation of CEA and CAS. Additionally, the relative importance of the clinical features was measured using an ML method.

Methods

The patient data that support the findings of this study are available from the corresponding author upon reasonable request. A subset of the program code generated for this study is available at GitHub and can be accessed at https://gist.github.com/kkmatsuo/54baea963568d184f347400a486eba45. We conducted and reported this study in compliance with the TRIPOD statement for multivariate prediction models.[9]

Study population

We enrolled 170 consecutive cases of carotid stenosis treated with CEA or CAS at a single institution in Japan between January 2013 and December 2018. Patient information was retrospectively collected from the hospital carotid stenosis database and complemented by individual search of medical records. The names of the patients and their hospital admission numbers were anonymized before analysis. This retrospective study was approved by the Ethical Committee of Kobe University Graduate School of Medicine (approval no. B200444). Written

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

informed consent was obtained from all patients before treatment. We excluded patients with arterial dissection and those who could not undergo an MRI because of pacemaker implantation. The missing values were imputed by the k-nearest-neighbours (kNN) method.[10] We used the clinical data of patients until March 2018 as training data to optimize the hyperparameters and train the ML models, and used the data of more recent patients, from April to December 2018, as test data to evaluate the predictive performance of each model. The sample size was determined by previous research investigating the appropriate sample size for efficient ML model development.[11]

Treatments

Therapeutic approach was in accordance with the guidelines [5, 6] based on stenosis degree assessed by NASCET criteria.[12] Basically, CAS was performed for CEA high-risk patients according to the inclusion criteria in the SAPPHIRE study.[4] The final treatment strategy was made by a multidisciplinary team. CEA was performed under general anaesthesia by three surgeons (K.H., H.K., K.T.). Continuous neurophysiological monitoring was performed by neurophysiologists during surgery with a multimodality protocol involving electroencephalogram (EEG), median nerve somatosensory evoked potentials (SSEP), and bilateral regional cerebral oximetry (rSO2). Shunt placement was determined by the onset of

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

alarm criteria, which were reported previously,[13] for either EEG, SSEP, or rSO2. CAS was performed under local anaesthesia by four surgeons (A.F., T.I., T.I., J.T.). The rSO2 was monitored during the procedure. Each patient treated with CAS was administered 100 mg aspirin and 75 mg clopidogrel daily for at least 7 days before and 90 days after the CAS. The protection methods (proximal protection, or filter- or balloon-based distal protection), the size and type of stents, and the size of the pre- and post-balloon were determined by the interventional team.

Clinical parameters

A total of 17 clinical parameters were used for the ML model development based on their known or expected influence on the outcome. These parameters consisted of age,[14] pretreatment modified Rankin scale (mRS),[15] hypertension,[15] diabetes mellitus,[16-18] medical history of arteriosclerotic disease,[19] serum low-density lipoprotein (LDL) cholesterol value (mg/dL),[16] internal carotid artery peak systolic velocity (ICA-PSV, cm/sec), symptomatic,[14, 16] crescendo transient ischemic attack (TIA) or stroke in evolution,[17] previous neck irradiation,[20] type III aorta,[16] contralateral carotid occlusion,[18] stenosis at a high position, mobile plaque, plaque ulceration,[16,17] vulnerable plaque, and procedure (CEA or CAS).[8, 21, 22] History of arteriosclerotic disease was defined as a history of acute coronary syndrome or peripheral artery disease requiring treatment. The most recent LDL-cholesterol

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

value was used. Crescendo TIA was defined as at least two similar TIAs in one week. Contralateral carotid occlusion, stenosis at a high position, and type III aorta were assessed using MRA, CTA, or angiography. Stenosis at a high position was defined as carotid stenosis that extends distally to the height of the vertebral body of C2. The ICA-PSV, mobile plaque, and plaque ulceration were assessed using echocardiogram. A vulnerable plaque was defined as a plaque that presents a hyperintense signal on an MRI time-of-flight image.

Outcomes

Outcome was defined as minor or major ischemic stroke including asymptomatic diffusion weighted imaging (DWI) hyperintense lesions within 30 days after CEA or CAS. Postprocedural MRI was performed the day after the procedure for CAS and less than one week afterwards for CEA. Additional MRI was performed if any neurological deficit was observed.

Development of machine learning models

The following six ML models were applied: artificial neural network (ANN), logistic regression, random forest, support vector machine (SVM), Gaussian naïve Bayes, and extreme gradient boosting (XGBoost).[23] The logistic regression, random forest, and SVM were implemented using scikit-learn, which is a free ML library for Python.[24] The ANN model was

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

implemented using the Keras library with a TensorFlow backend.[25] All the ML models were developed using Python version 3.7.7, Scikit-learn version 0.22.1, and TensorFlow version 2.2.0. First, we performed hyperparameter tuning of all models except for ANN using a grid-search algorithm with a log loss as the objective function on the training data. All numerical variables were standardized using centring and scaling before splitting the data into training and test sets. When applying grid-search, the value of the objective function was evaluated by stratified 5-fold cross validation. The hyperparameters of the ANN model were hand-tuned using the holdout method on the training data with cross-entropy as the objective function. The base ANN model consisted of three dense layers with two dropout layers. After identifying the optimal hyperparameters for every model that minimize the log loss value, we evaluated the predictive performance of each model using the stratified 5-fold cross validation method on the training data. We also created and evaluated an ensemble model using the four models with the highest accuracy on the training data. The usefulness of a multi-model ensemble was reported in a previous report.[26] Each model was ranked according to the sensitivity, specificity, positive prediction value (PPV), prediction accuracy, and area under the receiver operating characteristic curve (ROC-AUC). All models were then further analysed on the test data as external validation. Additional re-tuning of the hyperparameters was not permitted when the model was evaluated on the test data.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Test of machine learning models

The data of 22 consecutive patients treated by CEA or CAS at the same institution from April to December 2018 were used as a test dataset. Their post-procedural outcomes were predicted using the trained ML models with the 17 factors. Using the same 17 factors, four surgeons (board-certified neurosurgeons who had at least 10 years of experience) also predicted the post-procedural outcomes for each patient based on the test data within 10 minutes. When surgeons performed the outcome prediction test, to ensure that other information was never leaked, a paper test with information on only the 17 clinical factors for each patient was used (Supplemental Figure 1). The average sensitivity, specificity, PPV, and prediction accuracy of the predictions of the four surgeons were compared with those of the ML models.

Feature importance measurement

The relative importance of the clinical features was measured by the total gain of the XGBoost algorithm. The gain is the relative contribution of a feature to the model, calculated by taking each feature's contribution to each tree in the gradient boosting decision tree model. Thus, the features with higher gain are more important for generating the prediction of the XGBoost model.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Statistics

We used EZR version 1.38 (Saitama Medical Center, Jichi Medical University, Saitama, Japan) for statistical measurement. We performed a statistical comparison between training and test groups using Welch's t-test for numerical values, Fisher's exact test for categorical variables, and the Mann–Whitney U test for pretreatment mRS. Statistical significance was assumed if probability was measured at less than 0.05.

One author (K.M.) had sufficient access to the data to verify the manuscript's scientific integrity.

Results

Study participants

The flow diagram of model development and validation is presented in Figure 1. Among the 170 patients with carotid stenosis, five (2.9%) were excluded. Thus, the data of a total of 165 patients with carotid stenosis were included for analysis and separated into training and test data. The baseline characteristics before missing value imputation are shown in Table 1. There were 36 (22%) patients over 80 years of age, and 127 (77%) patients in good condition (mRS 0-1). Severe carotid stenosis, which was defined as an ICA-PSV of more than 200 cm/s, was observed

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

in 115 (70%) patients. CEA and CAS were performed on 95 (58%) and 70 (42%) patients, respectively. All missing values were imputed using the kNN method. No significant difference was found in the comparison before and after imputation of the missing values (data not shown). Although age and follow-up duration were significantly older and shorter, respectively, in the test data, there was no significant difference between the training and test data for the other factors that were used for analysis after imputation (Table 1). The test data tended to have more CAS and fewer outcomes, but there were no significant differences.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Patient characteristics

Variable	Before imputation	After imputation		Comparison between train and test data		
	Total (n=165)	Training data (n=143)	Test data (n=22)	p value	95% confidence interval	
Age, year (mean ±SD)	74.2 ± 7.61	73.8 ± 7.82	76.7 ± 5.62	0.04	0.15	5.69
Sex, male, n (%)	141 (85%)	122 (85%)	19 (86%)	1.00	0.16	3.55
pre-treatment mRS (median, IQR)	0 (0-1)	0 (0-1)	0.5 (0-1.75)	0.55	na	na
LDL-cholesterol, mg/dL (mean ±SD)	90.0 ± 29.5	90.4 ± 30.1	87.1 ± 23.3	0.56	-14.88	8.27
Prior medical histories						
HT, n (%)	136 (82%)	118 (83%)	18 (82%)	1.00	0.24	3.58
DM, n (%)	58 (35%)	50 (35%)	8 (36%)	1.00	0.34	2.78
ACS, n (%)	53 (32%)	42 (29%)	11 (50%)	0.08	0.15	1.15
PAD, n (%)	28 (17%)	24 (17%)	4 (18%)	0.77	0.27	4.02
Anatomical and pathophysiological features						
Contralateral occlusion, n (%)	14 (8.5%)	11 (7.7%)	3 (14%)	0.40	0.12	3.22
Stenosis at a high position, n (%)	13 (7.9%)	12 (8.4%)	1 (4.5%)	1.00	0.26	86.07
Type III Aorta, n (%)	64 (39%)	56 (39%)	8 (36%)	1.00	0.41	3.31
ICA-PSV, cm/sec (mean ±SD)	276 ± 130	272 ± 126	291 ± 147	0.56	-48.58	87.34
Mobile plaque, n (%)	19 (12%)	15 (10%)	4 (18%)	0.29	0.15	2.43
Plaque ulceration, n (%)	39 (24%)	36 (25%)	3 (14%)	0.29	0.57	11.85
Vulnerable plaque, n (%)	61 (37%)	54 (38%)	7 (32%)	0.64	0.46	4.01
Previous neck irradiation, n (%)	15 (9.1%)	13 (9.1%)	2 (9.1%)	1.00	0.20	9.79
Symptomatic, n (%)	64 (39%)	56 (39%)	8 (36%)	1.00	0.41	3.31
Crescendo TIA or stroke in evolution, n (%)	10 (6.1%)	8 (5.6%)	2 (9.1%)	0.62	0.11	6.14
Treatment, CEA, n (%)	95 (58%)	85 (59%)	10 (45%)	0.25	0.65	4.86
Ischemic stroke within 30 days, n (%)	46 (28%)	43 (30%)	3 (14%)	0.13	0.74	15.04
Follow-up duration, days (mean ±SD)	833 ± 565	921 ± 555	259 ± 259	<0.0001	-762	-561

ACS, acute coronary syndrome; CEA, carotid endarterectomy; CAS, carotid artery stenting; IQR, interquartile range; LDL, low density lipoprotein; mRS, modified Rankin scale; na, not available; PAD, peripheral artery disease; ICA-PSV, internal carotid artery-peak systolic velocity; SD, standard deviation; TIA, transient ischemic attack.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Prediction of post-procedural outcomes on the training dataset

To evaluate the predictive performance of each ML model for post CEA/CAS outcomes, stratified 5-fold cross-validation was first performed on the 143-patient training data. The prediction results showed that the ROC-AUC and sensitivity of naïve Bayes were highest at 0.730 and 55.6%, respectively, and the specificity, PPV, and accuracy of random forest were highest at 99.0%, 80.0%, and 74.1%, respectively (Table 2). Then, an ensemble model of the logistic regression, random forest, XGBoost, and ANN models, which were the four most accurate models, was also created and evaluated. It yielded a sensitivity of 23.3%, specificity of 97.0%, PPV of 65.0%, accuracy of 74.9%, and ROC-AUC of 0.704. Thus, the ensemble model obtained the highest accuracy scores on the training data.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Prediction results on the training dataset evaluated by 5-fold cross validation and sorted by accuracy

Model	**ROC-AUC** mean (95%CI)	**Sensitivity (%)** mean (95%CI)	**Specificity (%)** mean (95%CI)	**PPV (%)** mean (95%CI)	**Accuracy (%)** mean (95%CI)
Ensemble model *	0.704 (0.593 - 0.816)	23.3 (0.8 - 47.4)	97.0 (93.6 - 100)	65.0 (13.1 - 100)	74.9 (66.0 - 83.7)
Random forest	0.677 (0.523 - 0.831)	15.8 (0.2 - 31.4)	99.0 (96.2 - 100)	80.0 (24.5 - 100)	74.1 (68.8 - 79.4)
Logistic regression	0.678 (0.551 - 0.805)	27.8 (1.1 - 54.5)	92.0 (82.6 - 100)	54.6 (7.1 - 100)	72.8 (61.2 - 84.4)
XGBoost	0.706 (0.634 - 0.778)	20.8 (4.2 - 37.5)	94.0 (83.8 - 100)	60.0 (6.0 - 100)	72.1 (62.6 - 81.6)
Neural network	0.637 (0.458 - 0.817)	30.0 (5.3 - 54.7)	90.0 (81.2 - 98.8)	53.9 (12.0 - 95.8)	72.0 (60.8 - 83.2)
SVM	0.675 (0.555 - 0.796)	36.9 (5.1 - 68.8)	85.0 (80.6 - 89.4)	44.8 (10.3 - 79.2)	70.6 (60.1 - 81.2)
Naïve Bayes	0.730 (0.604 - 0.856)	55.6 (22.0 - 89.1)	68.0 (30.0 - 100)	48.7 (26.0 - 71.4)	64.5 (39.2 - 89.7)

* Ensemble model is created by using random forest, logistic regression, XGBoost, and neural network models.

ROC-AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; SVM, support vector machine; 95%CI, 95% confidence interval.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Prediction of post-procedural outcomes on the test dataset and comparison with surgeons

Next, the post-CEA/CAS outcome prediction performances of the seven trained ML models including the ensemble model were further evaluated on the 22-patient test data. The results are shown in Table 3. XGBoost achieved the highest sensitivity, specificity, PPV, and accuracy scores, which were 66.7%, 89.5%, 50.0%, and 86.4%, respectively. The ensemble model, which obtained the highest accuracy on the training data, did not perform well on the test data.

The average of the outcome predictions made by four surgeons had a sensitivity of 41.7%, specificity of 75.0%, PPV of 20.1%, and accuracy of 70.5%. Therefore, all of the current ML models trained with 143 cases and 17 factors outperformed the surgeons' predictions in terms of specificity and accuracy. Only XGBoost outperformed the surgeons' predictions in terms of sensitivity (Table 4). A statistical analysis was not performed on these results because of the small sample size. The tuned hyperparameters of these models are listed in Supplemental Table 1.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 3. Prediction results on the test dataset sorted by accuracy

Model	Sensitivity (%)	Specificity (%)	PPV (%)	Accuracy (%)
XGBoost	66.7	89.5	50.0	86.4
Random forest	33.3	84.2	25.0	77.3
Naïve Bayes	33.3	84.2	25.0	77.3
Ensemble model *	33.3	84.2	25.0	77.3
Neural network	33.3	78.9	20.0	72.7
SVM	33.3	78.9	20.0	72.7
Logistic regression	33.3	78.9	20.0	72.7
Surgeons **	41.7 (-9.1 – 92.4)	75.0 (60.7 – 89.3)	20.1 (-2.5 – 42.8)	70.5 (57.9 – 83.0)

* Ensemble model is created by using random forest, logistic regression, XGBoost, and neural network models.

** The average of 4 surgeon's prediction results with 95% confidence interval.

PPV, positive predictive value; SVM, support vector machine.

Table 4. Differences in surgeon and XGBoost model outcome predictions for the test dataset

	XGBoost	Surgeons *	Difference	95%CI	
Sensitivity (%)	66.7	41.7	25.0	-52.2	102
Specificity (%)	89.5	75.0	14.5	-9.4	38.3
PPV (%)	50.0	20.1	29.9	-33.0	92.7
Accuracy (%)	86.4	70.5	15.9	-7.9	39.8

* The average of 4 surgeon's prediction results.

PPV, positive predictive value; 95%CI, 95% confidence interval.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Importance values of the clinical factors

The feature importances were measured using a function of the XGBoost algorithm, which obtained the best predictive performance on the test data. The results reveal that ICA-PSV, serum LDL-cholesterol value, and procedure (CEA or CAS) are the most important in this order (Figure 2).

Discussion

In this study, we identified two notable findings. First, using an appropriate model construction process with effective clinical factors, we were able to develop a post-CEA/CAS outcome prediction model that is comparable to surgeons in terms of sensitivity, specificity, PPV, and accuracy, although our models were developed with a relatively small number of patients. Second, in our validation process, the XGBoost model had the highest predictive performance, and the factors that contributed most to the accurate model were ICA-PSV, serum LDL-cholesterol value, and procedure (CEA or CAS).

ML and ICAS

Using an appropriate model construction process with effective clinical factors, we were able to develop a post-CEA/CAS outcome prediction model that is comparable to surgeons in terms

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of sensitivity, specificity, PPV, and accuracy, although the ML models were developed using relatively few patients. This is the first study to predict the outcome after treatment of carotid artery stenosis using ML models. Our model makes it possible to preoperatively calculate the post-CEA/CAS stroke risk as a concrete numeric value; for example, 10% with CEA and 15% with CAS for postprocedural stroke risk, which could contribute to better treatment and prevention designs. However, a problem of the current models is that sensitivity is relatively low. Because every surgeon chooses the best treatment strategy to avoid treatment-associated ischemic complications, it would be somewhat difficult for surgeons to predict the likelihood of complications before treatment. Therefore, a higher sensitivity will be required for the ML prediction model to be a good decision support tool. To increase the prediction sensitivity, we tuned the models with sensitivity as the objective function during model development. However, such sensitivity-oriented models showed a considerable decrease in accuracy, although sensitivity increased slightly (data not shown). Therefore, these sensitivity-oriented models were not adopted this time. To further improve the predictive performance of an ML model, it would be necessary to use big data and more dominant influential factors.

The number of studies applying ML to neurosurgical decision support has increased rapidly over the past decade. The top three most frequently applied algorithms are ANN, logistic regression, and SVM.[27] Similarly, the number of studies on ML and stroke has increased.[28,

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

29] Most of these reports also use ANN, logistic regression, SVM, and random forest, which suggests that our model selection is appropriate. In our study, the predictive performance of the XGBoost was better than those of the ANN, logistic regression, SVM, and random forest. XGBoost is one of the major gradient boosting decision tree (GBDT) models. In a conventional GBDT algorithm, simple but weak predictive trees are repetitively generated and added to the predicting machine until the prediction is close to the ground truth. Although the superior performance of the GBDT model has been shown in many data-science contests in recent years,[30] there is still very little research on neurosurgery or stroke using the GBDT model.[30, 31] The ensemble model, which had the highest accuracy on our training data, did not perform well on the test data, probably because of overfitting on the training data, which likely also happened with random forest and naïve Bayes.

Feature Importance

In our validation process, the XGBoost model showed the highest predictive performance, and the factors that contributed most to its accuracy were ICA-PSV, serum LDL-cholesterol value, and procedure (CEA or CAS). The procedure type, CEA or CAS, has previously been reported as a potential predictor of periprocedural stroke.[8, 15, 21, 22] Thus, it seems reasonable that the procedure type is the third most important factor in this study. However, no studies have

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

suggested that ICA-PSV or serum LDL-cholesterol value are effective predictors of postprocedural outcome. Although ICA-PSV should reflect the degree of stenosis, many studies have reported that the degree of stenosis is not associated with postprocedural outcome.[14, 16] ICA-PSV might be used in our model as a valid predictor in combination with other clinical factors. For LDL-cholesterol, a previous study showed that preprocedural statin therapy significantly reduced major adverse cardiac and cerebrovascular events after carotid revascularization.[32] However, its primary outcome was a 1-year composite risk including any stroke, and serum LDL-cholesterol value was not measured. In a study investigating the components of emboli captured by the filter-protection device during CAS, more postoperative DWI high intensity lesions were observed in patients whose main component of the emboli was cholesterol.[33] It may be inferred from such results that serum LDL-cholesterol might be related to post-procedural outcome because it should be possible for a higher serum LDL-cholesterol value to leads to a higher cholesterol content of the plaque, which produces emboli.

There should be other important predictive factors that have not been adopted in our model. For example, the stenotic lesion length,[16] surgeon's experience,[16, 34] centre's experience,[14, 19], or detailed surgical methods are potential predictive factors. Moreover, the accuracy of image classification and recognition using deep neural networks has recently remarkably increased. Therefore, it might be possible to improve our model performances by replacing some

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

clinical factors from radiological or echocardiogram findings with feature extraction using deep neural networks. However, because our XGBoost model gives importance to LDL-cholesterol, there must be important factors other than imaging findings that should be further explored.

Comparison of clinician and AI performance

Many studies have compared clinicians and AI with image interpretation or diagnostic performance on, for example, CT or gastroscopic images, and have shown that ML models are equivalent to or superior to specialists.[35, 36] However, it is difficult to find a study that compares the predictive performance of clinicians and AI as in this study. Although it is an unfamiliar task for surgeons to predict the post-procedural outcome by looking only at tabulated letters and numerals, the surgeons' predictions were, surprisingly, not inadequate. The sensitivity of the surgeons' prediction was higher than those of all ML models except XGBoost, presumably because the surgeons have experience of more than 143 patients, which comprise the dataset that our ML models learned. Therefore, instead of a difference in predictive ability, the number of experienced and learned cases might have influenced the results. The advantage of an ML model is that if only the question and answer are provided correctly, it can learn an enormous number of cases that one surgeon would not be able to experience. Therefore, in future, as the number of training cases increases, the predictive performance of the ML model should definitely improve.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Studies comparing the performance of AI with expert clinicians have been criticized because in most of them, adherence to reporting standards is suboptimal and code availability is lacking.[37] This study was conducted in accordance with the TRIPOD statement, and the main source code is provided.

Limitations

One of the main limitations of our study is the small sample size. Because big data generally makes ML models more accurate, a larger sample size would be needed for more accurate predictions. However, a previous report suggested that 80–560 samples are required for ML algorithms excluding deep neural networks, and the required sample size depends on the dataset and sampling method.[11] Furthermore, a systematic review of AI in neurosurgery has shown that the median number of patients in each study was 120.[27] Thus, the sample size might not be insufficient for our ML models (excepting the ANN model). Next, optimizing hyperparameters for neural network models is generally difficult because they have many hyperparameters that need to be adjusted. In addition to hyperparameters, the neural network architecture should be optimized for better performance.[38] In this study, because the ANN was hand-tuned by multiple trial-and-error sessions, a more effective hyperparameter set might be found by other, more sophisticated optimization methods. Finally, because the dataset was

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

collected retrospectively from a single institution, it is prone to selection bias, and the proposed ML models may not be applicable to other institutions where different treatment strategies or patient demographics might exist. Although internal validation was performed with cross-validation and additional validation was performed on test data from new patients, further external validation and prospective studies are required in other settings that differ in time or place to validate the performance of our models.

Conclusion

We were able to develop a post-CEA/CAS outcome prediction model with a performance comparable to that of surgeons. The XGBoost model showed the best predictive performance, and the factors that contributed most to model building were ICA-PSV, serum LDL-cholesterol value, and procedure (CEA or CAS). Larger datasets and analysis of potential prognostic factors would be necessary to further improve the predictive performance of the ML models, which will enable a more appropriate patient-specific selection of CEA or CAS for the treatment of carotid stenosis.

Competing Interests

None.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Sources of Funding

This study was partially supported by research funds from the Alumni Association of the Department of Neurosurgery, Kobe University School of Medicine.

References

1. Taussky P, Hanel RA, Meyer FB. Clinical considerations in the management of asymptomatic carotid artery stenosis. *Neurosurg Focus* 2011;31(6):E7. doi: 10.3171/2011.9.FOCUS11222

2. Rosenfield K, Matsumura JS, Chaturvedi S, et al. Randomized Trial of Stent versus Surgery for Asymptomatic Carotid Stenosis. *N Engl J Med* 2016;374(11):1011-20. doi: 10.1056/NEJMoa1515706

3. Brott TG, Howard G, Roubin GS, et al. Long-Term Results of Stenting versus Endarterectomy for Carotid-Artery Stenosis. *N Engl J Med* 2016;374(11):1021-31. doi: 10.1056/NEJMoa1505215

4. Hitinder SG, Jay SY, Pierre F, et al. Long-term results of carotid stenting versus endarterectomy in high-risk patients. *N Engl J Med* 2008;358:1572-9.

5. Kernan WN, Ovbiagele B, Black HR, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014;45(7):2160-236. doi: 10.1161/STR.0000000000000024

6. The Japan Stroke Society SGC. Japanese guidelines for the management of stroke 2015. Tokyo: Kyowa Kikaku 2015.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

7. Koch M. Artificial Intelligence Is Becoming Natural. *Cell* 2018;173(3):531-3. doi: 10.1016/j.cell.2018.04.007

8. Mantese VA, Timaran CH, Chiu D, et al. The Carotid Revascularization Endarterectomy versus Stenting Trial (CREST): stenting versus carotid endarterectomy for carotid disease. *Stroke* 2010;41(10 Suppl):S31-4. doi: 10.1161/STROKEAHA.110.595330

9. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1-73. doi: 10.7326/M14-0698

10. Olga T, Michael C, Gavin S, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17(6):520-5.

11. Figueroa RL, Zeng-Treitler Q, Kandula S, et al. Predicting sample size required for classification performance. *BMC Med Inform Decis Mak* 2012;12(8):1-10.

12. Gary GF, Michael E, Hugh WK, et al. The North American Symptomatic Carotid Endarterectomy Trial Surgical Results in 1415 Patients. *Stroke* 1999;30:1751-8.

13. Friedell ML, Clark JM, Graham DA, et al. Cerebral oximetry does not correlate with electroencephalography and somatosensory evoked potentials in determining the need for shunting during carotid endarterectomy. *J Vasc Surg* 2008;48(3):601-6. doi: 10.1016/j.jvs.2008.04.065

14. Theiss W, Hermanek P, Mathias K, et al. Predictors of death and stroke after carotid angioplasty and stenting: a subgroup analysis of the Pro-CAS data. *Stroke* 2008;39(8):2325-30. doi: 10.1161/STROKEAHA.108.514356

15. Muller MD, von Felten S, Algra A, et al. Immediate and Delayed Procedural Stroke or Death in Stenting Versus Endarterectomy for Symptomatic Carotid Stenosis. *Stroke*

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2018;49(11):2715-22. doi: 10.1161/STROKEAHA.118.020684

16. Setacci C, Chisci E, Setacci F, et al. Siena carotid artery stenting score: a risk modelling study for individual patients. *Stroke* 2010;41(6):1259-65. doi: 10.1161/STROKEAHA.110.578583

17. Halm EA, Tuhrim S, Wang JJ, et al. Risk factors for perioperative death and stroke after carotid endarterectomy: results of the new york carotid artery surgery study. *Stroke* 2009;40(1):221-9. doi: 10.1161/STROKEAHA.108.524785

18. Tu JV, Wang H, Bowyer B, et al. Risk factors for death or stroke after carotid endarterectomy: observations from the Ontario Carotid Endarterectomy Registry. *Stroke* 2003;34(11):2568-73. doi: 10.1161/01.STR.0000092491.45227.0F

19. Dua A, Romanelli M, Upchurch GR, Jr., et al. Predictors of poor outcome after carotid intervention. *J Vasc Surg* 2016;64(3):663-70. doi: 10.1016/j.jvs.2016.03.428

20. Favre JP, Nourissat A, Duprey A, et al. Endovascular treatment for carotid artery stenosis after neck irradiation. *J Vasc Surg* 2008;48(4):852-8. doi: 10.1016/j.jvs.2008.05.069

21. Choi JC, Johnston SC, Kim AS. Early outcomes after carotid artery stenting compared with endarterectomy for asymptomatic carotid stenosis. *Stroke* 2015;46(1):120-5. doi: 10.1161/STROKEAHA.114.006209

22. Moresoli P, Habib B, Reynier P, et al. Carotid Stenting Versus Endarterectomy for Asymptomatic Carotid Artery Stenosis: A Systematic Review and Meta-Analysis. *Stroke* 2017;48(8):2150-7. doi: 10.1161/STROKEAHA.117.016824

23. 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;785-94.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

24. Fabian P, Ga¨el V, Alexandre G, et al. Scikit-learn_ Machine Learning in Python. *Journal ofMachine Learning Research* 2011;12:2825-30.

25. Martín A, Paul B, Jianmin C, et al. TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)* 2016

26. Xiao Y, Wu J, Lin Z, et al. A deep learning-based multi-model ensemble method for cancer prediction. *Comput Methods Programs Biomed* 2018;153:1-9. doi: 10.1016/j.cmpb.2017.09.005

27. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev* 2019 doi: 10.1007/s10143-019-01163-8

28. Heo J, Yoon JG, Park H, et al. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke* 2019;50(5):1263-5. doi: 10.1161/STROKEAHA.118.024293

29. van Os HJA, Ramos LA, Hilbert A, et al. Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Front Neurol* 2018;9:784. doi: 10.3389/fneur.2018.00784

30. Xie Y, Jiang B, Gong E, et al. 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(1):44-51. doi: 10.2214/AJR.18.20260

31. 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. doi: 10.1038/srep21161

32. Hussain MA, Saposnik G, Raju S, et al. Association Between Statin Use and Cardiovascular Events After Carotid Artery Revascularization. *J Am Heart Assoc* 2018;7(16):e009745. doi:

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10.1161/JAHA.118.009745

33. Maekawa K, Shibata M, Nakajima H, et al. Cholesterol Crystals in Embolic Debris are Associated with Postoperative Cerebral Embolism after Carotid Artery Stenting. *Cerebrovasc Dis* 2018;46(5-6):242-8. doi: 10.1159/000495795

34. Hussain MA, Mamdani M, Tu JV, et al. Association between operator specialty and outcomes after carotid artery revascularization. *J Vasc Surg* 2018;67(2):478-89 e6. doi: 10.1016/j.jvs.2017.05.123

35. Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol* 2019;29(11):6191-201. doi: 10.1007/s00330-019-06163-2

36. Li L, Chen Y, Shen Z, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* 2020;23(1):126-32. doi: 10.1007/s10120-019-00992-2

37. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020;368:m689. doi: 10.1136/bmj.m689

38. Yoo Y. Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches. *Knowledge-Based Systems* 2019;178:74-83. doi: 10.1016/j.knosys.2019.04.019

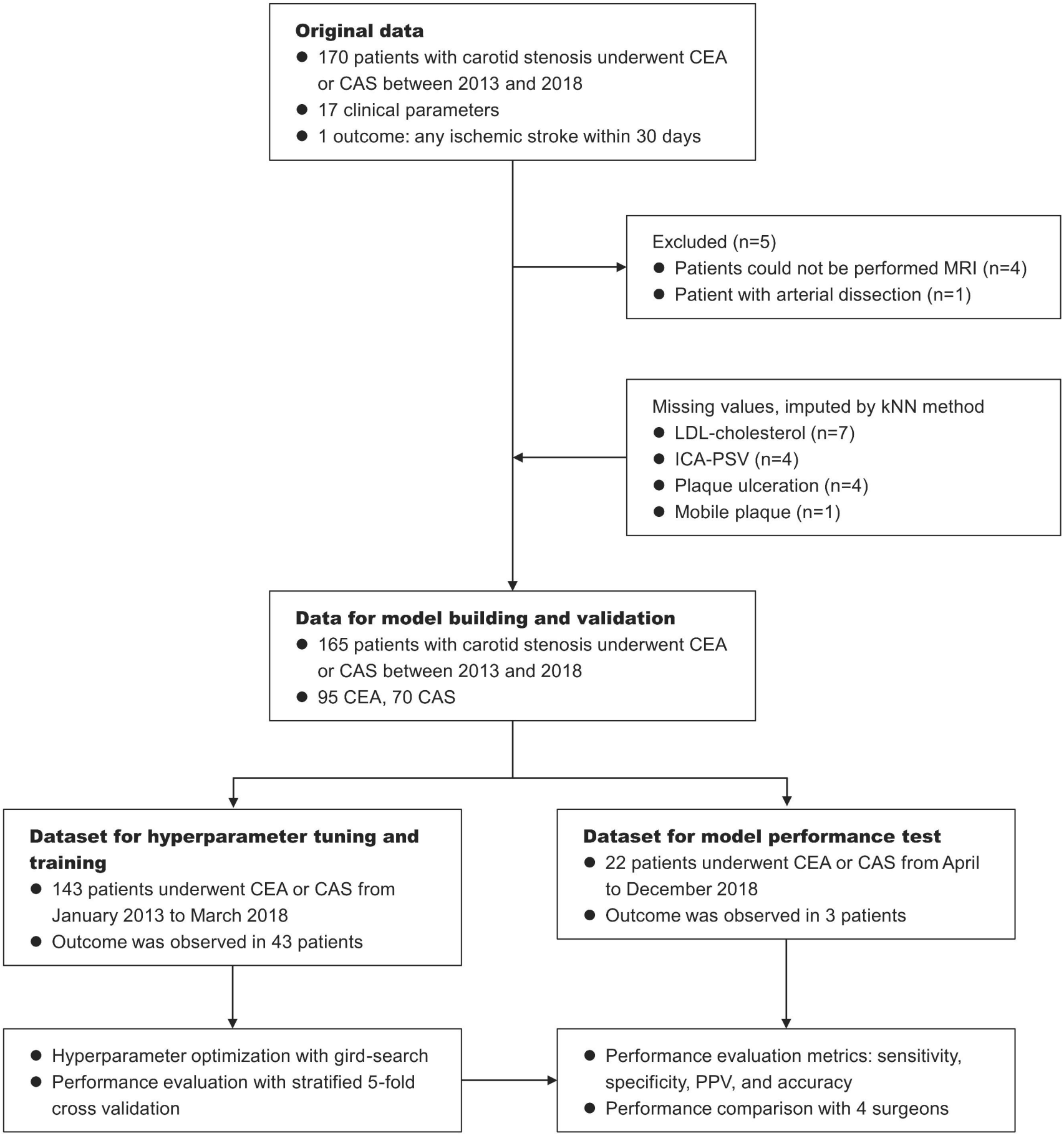
Figure legends

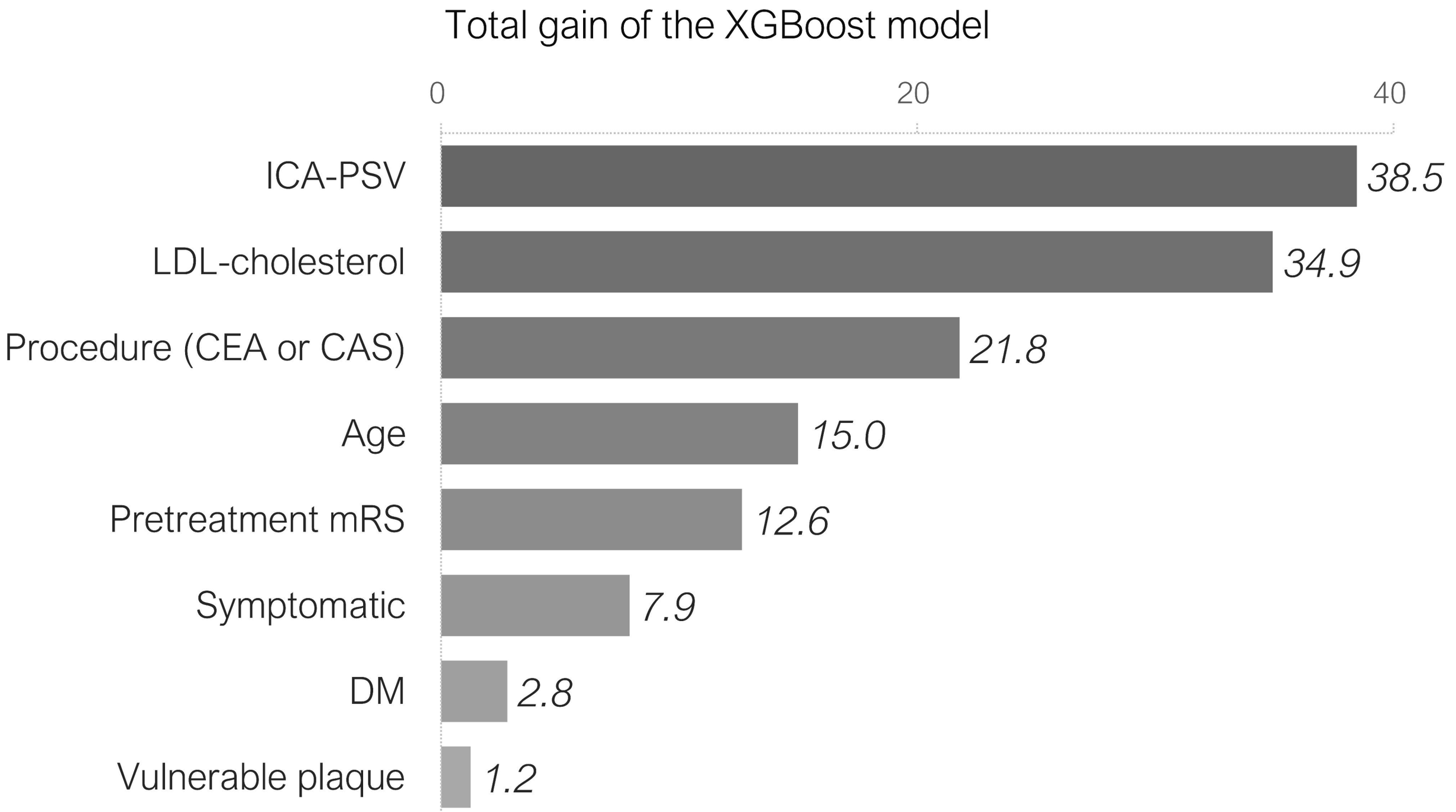
Figure 1. Flow diagram describing the general framework of the study. Models were built using

medRxiv preprint doi: https://doi.org/10.1101/2020.11.16.20231639; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the training dataset. The test dataset was used for measuring the predictive performance and comparison with the surgeons. CAS, carotid artery stenting; CEA, carotid endarterectomy; ICA-PSV, internal carotid artery peak systolic velocity; kNN, k-nearest-neighbours; LDL, low density lipoprotein; PPV, positive predictive value.

Figure 2. Importance values of the clinical factors measured using the total gain of the XGBoost algorithm. CAS, carotid artery stenting; CEA carotid endarterectomy; DM, diabetes mellitus; ICA-PSV, internal carotid artery peak systolic velocity; LDL, low density lipoprotein; mRS, modified Rankin scale.


Original data
170 patients with carotid stenosis underwent CEA or CAS between 2013 and 2018
17 clinical parameters
1 outcome: any ischemic stroke within 30 days
Excluded (n=5)
Patients could not be performed MRI (n=4)
Patient with arterial dissection (n=1)
Missing values, imputed by kNN method
LDL-cholesterol (n=7)
ICA-PSV (n=4)
Plaque ulceration (n=4)
Mobile plaque (n=1)
Data for model building and validation
165 patients with carotid stenosis underwent CEA or CAS between 2013 and 2018
95 CEA, 70 CAS
Dataset for hyperparameter tuning and training
143 patients underwent CEA or CAS from January 2013 to March 2018
Outcome was observed in 43 patients
Dataset for model performance test
22 patients underwent CEA or CAS from April to December 2018
Outcome was observed in 3 patients
Hyperparameter optimization with gird-search
Performance evaluation with stratified 5-fold cross validation
Performance evaluation metrics: sensitivity, specificity, PPV, and accuracy
Performance comparison with 4 surgeons


Total gain of the XGBoost model
0
20
40
ICA-PSV
38.5
LDL-cholesterol
34.9
Procedure (CEA or CAS)
21.8
Age
15.0
Pretreatment mRS
12.6
Symptomatic
7.9
DM
2.8
Vulnerable plaque
1.2