

1 **Title**

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

4

5 **Author information**

6 Kazuya Matsuo MD, PhD ¹, ORCID ID 0000-0001-6721-1353, Atsushi Fujita MD, PhD ¹,
7 ORCID ID 0000-0003-1294-6584, Kohkichi Hosoda MD, PhD ², ORCID-ID
8 0000-0002-0419-9497, Jun Tanaka MD, PhD ³, Taichiro Imahori MD, PhD ⁴, Taiji Ishii MD,
9 PhD ⁵, Masaaki Kohta MD, PhD ¹, Kazuhiro Tanaka MD, PhD ¹, Yoichi Uozumi MD, PhD ¹,
10 Hidehito Kimura MD, PhD ¹, Takashi Sasayama MD, PhD ¹, Eiji Kohmura MD, PhD ¹

11

12 **Affiliations**

13 ¹ Department of Neurosurgery, Kobe University Graduate School of Medicine, Kobe, Hyogo,
14 Japan

15 ² Department of Neurosurgery, Kobe City Nishi-Kobe Medical Center, Kobe, Hyogo, Japan

16 ³ Department of Neurosurgery, Konan Hospital, Kobe, Hyogo, Japan

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

18 ⁵ Department of Neurosurgery, Toyooka Hospital, Toyooka, Hyogo, Japan

19

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

22 Tel: +81-78-382-5966

23 Fax: +81-78-382-5979

24 E-mail: kkmatsuo@outlook.jp

1 **Abstract**

2 **Background**

3 Carotid endarterectomy (CEA) and carotid artery stenting (CAS) are recommended for
4 high stroke-risk patients with carotid artery stenosis to reduce ischemic events. However, we
5 often face difficulty in determining the best treatment strategy.

6 **Objective**

7 We aimed to develop an accurate post-CEA/CAS outcome prediction model using machine
8 learning that will serve as a basis for a new decision support tool for patient-specific treatment
9 planning.

10 **Methods**

11 Retrospectively collected data from 165 consecutive patients with carotid stenosis
12 underwent CEA or CAS were divided into training and test samples. The following five
13 machine learning algorithms were tuned, and their predictive performance evaluated by
14 comparison with surgeon predictions: an artificial neural network, logistic regression, support
15 vector machine, random forest, and extreme gradient boosting (XGBoost). Seventeen clinical
16 factors were introduced into the models. Outcome was defined as any ischemic stroke within 30
17 days after treatment including asymptomatic diffusion-weighted imaging abnormalities.

18 **Results**

19 The XGBoost model performed the best in the evaluation; its sensitivity, specificity,
20 positive predictive value, and accuracy were 31.9%, 94.6%, 47.2%, and 86.2%, respectively.
21 These statistical measures were comparable to those of surgeons. Internal carotid artery peak
22 systolic velocity, low density lipoprotein cholesterol, and procedure (CEA or CAS) were the
23 most contributing factors according to the XGBoost algorithm.

24 **Conclusion**

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

5

6 **Keywords**

7 Carotid artery stenting, carotid endarterectomy, carotid stenosis, decision support tool, machine
8 learning.

9

10 **Declarations**

11 **Funding**

12 Matsuo K is supported in part by a research fund from the Alumni Association of the
13 Department of Neurosurgery, Kobe University School of Medicine.

14

15 **Conflicts of interest/Competing interests**

16 Not applicable.

17

18 **Availability of data and material (data transparency)**

19 The patient data that support the findings of this study are available from the corresponding
20 author upon reasonable request.

21

22 **Code availability (software application or custom code)**

23 A subset of the program code generated for this study is available at GitHub and can be
24 accessed at <https://gist.github.com/kkmatsuo/e77e78a0346280d5570829164760132f>.

1

2 **Ethics approval (include appropriate approvals or waivers)**

3 This retrospective study was approved by the Ethical Committee of Kobe University Graduate
4 School of Medicine (approval no. B200444).

5

6 **Consent to participate (include appropriate statements)**

7 Written informed consent was obtained from all patients before treatment.

8

9 **Consent for publication (include appropriate statements)**

10 Not applicable.

11

12 **Authors' contributions**

13 Kazuya Matsuo conceptualized the experimental design, analyzed the data, developed the
14 machine learning model, and wrote the manuscript. Atsushi Fujita contributed to study design,
15 interpreted the results, and critically revised the manuscript. Kohkichi Hosoda provided the
16 background data for the study, interpreted the results, and reviewed the manuscript. Jun Tanaka
17 carried out the data acquisition, conducted the analyses, and reviewed the manuscript. Taichiro
18 Imahori carried out the data acquisition and reviewed the manuscript. Taiji Ishii carried out the
19 data acquisition and reviewed the manuscript. Masaaki Kohta participated in the study design,
20 interpreted the results, and reviewed the manuscript. Kazuhiro Tanaka carried out the data
21 acquisition, conducted the analyses, and interpreted the results. Yoichi Uozumi conducted the
22 analyses and interpreted the results. Hidehito Kimura carried out the data acquisition and
23 interpreted the results. Takashi Sasayama supervised the whole study and critically revised the
24 manuscript. Eiji Kohmura supervised the whole study, contributed to study design, and critically

1 revised the manuscript. All authors read and approved the final manuscript.

2

1 **Introduction**

2 Carotid artery stenosis is an important cause of ischemic stroke, which remains a major
3 public health problem worldwide [26]. To reduce the risk of ischemic stroke, carotid
4 endarterectomy (CEA) and carotid artery stenting (CAS) are recommended for patients at high
5 stroke risk with carotid artery stenosis. Based on the evidence that there are no significant
6 differences in long-term outcomes after CEA and CAS [1,10,23], there are general guidelines
7 for patient selection for CEA and CAS [10,13,27]. However, we often face difficulty in
8 determining the best treatment method. Therefore, it is necessary to develop a useful decision
9 support tool for patient-specific treatment planning for carotid artery stenosis. Recently, the use
10 of artificial intelligence (AI) has been increasing in medical field because of advances in
11 technology including robust machine learning (ML) algorithms for successful prediction and
12 diagnosis [14]. However, no studies have applied modern ML models in the carotid stenosis
13 cohort.

14 In this study, we aimed to develop an accurate ML model for the prediction of ischemic
15 events within 30 days after CEA or CAS with 17 clinical factors. The usefulness of the ML
16 models was evaluated by comparing their predictions with those of surgeons. Because early
17 periprocedural major and minor stroke is associated with long-term outcomes [18], our model to
18 predict post-procedural ischemic events can serve as the basis for an effective decision support
19 tool for patient-specific adaptation of CEA and CAS. Additionally, the relative importance of
20 the clinical features was measured using an ML method.

21

22 **Materials and Methods**

23 **Study population**

24 We enrolled 170 consecutive cases of carotid stenosis treated with CEA or CAS at a single

1 center in Japan between January 2013 and December 2018. Patient information was
2 retrospectively collected from the hospital carotid stenosis database and anonymized before
3 analysis. This retrospective study was approved by the institutional review board. Written
4 informed consent was obtained from all patients before treatment. We excluded patients with
5 arterial dissection and those who could not undergo MRI because of pacemaker implantation.
6 The missing values were imputed by the k-nearest-neighbours (kNN) method [30]. We used the
7 clinical data of patients until March 2018 as training data to optimize the hyperparameters and
8 train the ML models, and used the data of more recent patients, from April to December 2018,
9 as test data to evaluate the predictive performance of each model. The sample size was
10 determined based on a previous study [8]. We conducted and reported this study in compliance
11 with the transparent reporting of a multivariable prediction model for individual prognosis or
12 diagnosis (TRIPOD) statement for multivariate prediction models [19].

13

14 **Treatments**

15 Therapeutic approach was in accordance with the guidelines [13,27] based on stenosis
16 degree assessed by NASCET criteria [7]. Basically, CAS was performed for CEA high-risk
17 patients according to the inclusion criteria in the SAPHIRE study [10]. The final treatment
18 strategy was made by a multidisciplinary team. CEA was performed under general anaesthesia
19 by three surgeons. Continuous neurophysiological monitoring was performed by
20 neurophysiologists during surgery with a multimodality protocol involving
21 electroencephalogram (EEG), median nerve somatosensory evoked potentials (SSEP), and
22 bilateral regional cerebral oximetry (rSO₂). Shunt placement was determined by the onset of
23 alarm criteria for either EEG or SSEP, which was defined as a >50% decrease in amplitude [9].
24 CAS was performed under local anaesthesia by four surgeons. The rSO₂ was monitored during

1 the procedure. Each patient treated with CAS was administered 100 mg aspirin and 75 mg
2 clopidogrel daily for at least 7 days before and 90 days after the CAS. The choice of stent and
3 interventional strategy were determined by the interventional team.

4

5 **Clinical parameters**

6 A total of 17 clinical parameters were used for the ML model development based on their
7 known or expected influence on the outcome. These parameters consisted of age [28],
8 pretreatment modified Rankin scale (mRS) [21], hypertension [21], diabetes mellitus [11,24,31],
9 medical history of arteriosclerotic disease [5], serum low-density lipoprotein (LDL) cholesterol
10 value (mg/dL) [24], internal carotid artery peak systolic velocity (ICA-PSV, cm/sec),
11 symptomatic [24,28], crescendo transient ischemic attack (TIA) or stroke in evolution [11],
12 previous neck irradiation [6], type III aorta [24], contralateral carotid occlusion [31], stenosis at
13 a high position, mobile plaque, plaque ulceration [11,24], plaque with hyperintense signal on
14 time-of-flight (TOF), and procedure (CEA or CAS) [4,18,20]. History of arteriosclerotic disease
15 was defined as a history of acute coronary syndrome or peripheral artery disease requiring
16 treatment. Crescendo TIA was defined as at least two similar TIAs in one week. Contralateral
17 carotid occlusion, stenosis at a high position, and type III aorta were assessed using MRA, CTA,
18 or angiography. Stenosis at a high position was defined as carotid stenosis that extends distally
19 to the height of the vertebral body of C2. The ICA-PSV, mobile plaque, and plaque ulceration
20 were assessed using echocardiogram. A plaque with hyperintense signal on TOF was defined as
21 a plaque that appeared hyperintense on TOF-MRI compared with the signal of the adjacent
22 sternocleidomastoid muscle.

23 In our preliminary study, we first created a prediction model with 22 clinical parameters
24 which included gender, arterial fibrillation, estimated glomerular filtration rate, stenosis degree

1 assessed by NASCET criteria, and the side of the lesion. We then find the effective parameters
2 and narrowed it down to the 17 parameters with the highest predictive performance (data not
3 shown).

4

5 **Outcomes**

6 Outcome was defined as minor or major ischemic stroke including asymptomatic diffusion
7 weighted imaging (DWI) hyperintense lesions within 30 days after CEA or CAS. Minor stroke
8 was defined as a National Institutes of Health Stroke Scale score of less than 3, and major stroke
9 was defined as a score of 3 or higher. Postprocedural MRI was performed the day after the
10 procedure for CAS and less than one week afterwards for CEA. Additional MRI was performed
11 if any neurological deficit was observed.

12

13 **Development of machine learning models**

14 The following five ML models were applied: artificial neural network (ANN), logistic
15 regression, random forest, support vector machine (SVM), and extreme gradient boosting
16 (XGBoost) [3]. The logistic regression, random forest, and SVM were implemented using
17 scikit-learn, which is a free ML library for Python. The ANN model was implemented using the
18 Keras library with a TensorFlow backend. All the ML models were developed using Python
19 version 3.7.7, Scikit-learn version 0.22.1, and TensorFlow version 2.2.0. First, we performed
20 hyperparameter tuning of all models except for ANN using a grid-search algorithm with
21 log-loss as the objective function on the training data. All numerical variables were standardized
22 using centring and scaling. When applying grid-search, the value of the objective function was
23 evaluated by stratified 5-fold cross-validation. The hyperparameters of the ANN model were
24 hand-tuned using the holdout method on the training data with cross-entropy as the objective

1 function. The base ANN model consisted of three dense layers with two dropout layers and two
2 batch normalization layers (Supplemental table 1). After identifying the optimal
3 hyperparameters for every model that minimize the log-loss value, we evaluated the predictive
4 performance of each model using 10 times repeated stratified 5-fold cross-validation on the
5 training data. The averages of the area under the receiver operating characteristic curve (ROC
6 AUC), sensitivity, specificity, positive prediction value (PPV), and prediction accuracy were
7 calculated through the cross-validation, and 95% confidence intervals were estimated. We also
8 created and evaluated an ensemble model which comprises of the three models with the highest
9 ROC AUC on the training data.

10

11 **Test of machine learning models**

12 The data of 22 consecutive patients treated by CEA or CAS at the same institution from
13 April to December 2018 were used as a test data. Their post-procedural outcomes were
14 predicted using the trained ML models with the 17 factors. In this procedure, we used the
15 standard bootstrap method as an additional internal validation technique. In this study, the
16 number of resampling repetitions, which should be as large as possible to ensure the stability of
17 the estimates, was set to 1000 repetitions. The bootstrap averages of sensitivity, specificity, PPV,
18 and prediction accuracy were calculated, and 95% confidence intervals were estimated. Using
19 the same 17 factors, four surgeons (board-certified neurosurgeons who had at least 10 years of
20 experience) also predicted the post-procedural outcomes for each patient based on the test data
21 within 10 minutes. When surgeons performed the outcome prediction test, to ensure that other
22 information was never leaked, a paper test with information on only the 17 clinical factors for
23 each patient was used. These surgeons were independent of the CEA and CAS performed during
24 the designated study period. The average sensitivity, specificity, PPV, and prediction accuracy of

1 the predictions of the four surgeons were compared with those of the ML models.

2

3 **Feature importance measurement**

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

9

10 **Statistical Analysis**

11 Descriptive and comparative statistics were used to describe clinical characteristics of the
12 patients. We performed a statistical comparison between training and test groups using Welch's
13 t-test for numerical values, Fisher's exact test for categorical variables, and the Mann–Whitney
14 U test for pretreatment mRS. A 2-tailed probability value of 0.05 or lower was considered
15 statistically significant. Statistical analysis was performed using EZR version 1.38.

16

17 **Results**

18 **Study participants**

19 The flow diagram of model development and validation is presented in Figure 1. Among
20 the 170 patients with carotid stenosis, five (2.9%) were excluded. Thus, the data of a total of
21 165 patients with carotid stenosis were included for analysis and separated into training and test
22 data. The baseline characteristics before missing value imputation are shown in Table 1. There
23 were 36 (22%) patients over 80 years of age, and 127 (77%) patients in good condition (mRS
24 0-1). Severe carotid stenosis, which was defined as an ICA-PSV of more than 200 cm/s, was

1 observed in 115 (70%) patients. CEA and CAS were performed on 95 (58%) and 70 (42%)
2 patients, respectively. The outcome was observed in 45 (27%) patients. Major stroke, minor
3 stroke, and asymptomatic DWI hyperintense lesions were diagnosed in 3 (1.8%), 3 (1.8%), and
4 39 (24%) patients, respectively. All missing values were imputed using the kNN method. No
5 significant difference was found in the comparison before and after imputation of the missing
6 values (data not shown). Although age and follow-up duration were significantly older and
7 shorter, respectively, in the test data, there was no significant difference between the training
8 and test data for the other factors that were used for analysis after imputation (Table 1). The test
9 data tended to have more CAS and fewer outcomes, but there were no significant differences.

10 The differences in patient characteristics between the CEA and CAS groups were shown in
11 Supplemental table 1. The patients with hypertension, plaque ulceration, plaque with
12 hyperintense signal on TOF, and symptomatic stenosis were significantly more common in the
13 CEA group. Asymptomatic DWI hyperintense lesion was significantly increased in CAS group.
14 Other variables did not differ between the CEA and CAS groups.

15

16 **Prediction of post-procedural ischemic events on the training dataset**

17 To evaluate the predictive performance of five ML models for post CEA/CAS outcomes,
18 10 times repeated stratified 5-fold cross-validation was first performed on the 143-patient
19 training data. The prediction results showed that the ROC AUC of XGBoost was highest at
20 0.719, sensitivity of SVM was highest at 36.2%, and the specificity, PPV, and accuracy of
21 random forest were highest at 98.3%, 78.9%, and 75.4%, respectively (Table 2). Then, an
22 ensemble model of the logistic regression, XGBoost, and ANN models, which were the three
23 most highest ROC AUC models, was also created and evaluated. It yielded a ROC AUC of
24 0.739, sensitivity of 15.1%, specificity of 97.4%, PPV of 75.1%, and accuracy of 72.7%. Thus,

1 the ensemble model obtained the highest ROC AUC on the training data.

2

3 **Prediction of post-procedural ischemic events on the test dataset and comparison with** 4 **surgeons**

5 Next, to confirm the post-CEA/CAS outcome prediction performances of the six ML
6 models including the ensemble model, these models were further evaluated on the 22-patient
7 test data with the bootstrap method. The results are shown in Table 3. XGBoost achieved the
8 highest PPV, and accuracy scores, which were 47.2% and 86.2%, respectively. The highest
9 sensitivity was 34.0%, which was achieved by random forest, SVM, and logistic regression. The
10 highest specificity was 95.4%, which was achieved by ANN. The ensemble model, which
11 obtained the highest ROC AUC on the training data, did not perform so well on the test data.

12 The average of the outcome predictions made by four surgeons had a sensitivity of 41.7%,
13 specificity of 75.0%, PPV of 20.1%, and accuracy of 70.5%. Therefore, all of the current ML
14 models trained with 143 cases and 17 factors outperformed the surgeons' predictions in terms of
15 specificity and accuracy. However, surgeons showed a higher predictive sensitivity than the
16 current ML models. A statistical analysis was not performed on these results because of the
17 small sample size. The optimized hyperparameters of these models are listed in Supplemental
18 table 2.

19

20 **Importance values of the clinical factors**

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

1

2 **Discussion**

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

10

11 **ML and ICAS**

12 We were able to develop a post-CEA/CAS outcome prediction model that is comparable to
13 surgeons, even with a relatively small training sample size. This is the first study to predict the
14 outcome after treatment of carotid artery stenosis using ML models. Our model makes it
15 possible to preoperatively calculate the post-CEA/CAS stroke risk as a concrete numeric value;
16 for example, 10% with CEA and 15% with CAS for postprocedural stroke risk. This function
17 can help surgeons to determine whether the patient suitable for CEA or CAS. However, a
18 problem of the current models is that sensitivity is relatively low. To increase the prediction
19 sensitivity, we tuned the models with sensitivity as the objective function during model
20 development. However, such sensitivity-oriented models showed a considerable decrease in
21 accuracy, although sensitivity increased slightly (data not shown). Therefore, these
22 sensitivity-oriented models were not adopted this time. In this study, we did not perform
23 statistical analyses on differences of predictive performance between ML models and surgeons
24 because of the small test sample size. Therefore, we did not determine the statistical significance

1 of ML models. However, our prediction models could be a milestone in the development of new
2 decision support tool for treatment choice in carotid stenosis, because as the training data
3 continues to grow in size, we can expect the prediction performance of the models to be higher
4 than the present study. To further improve the predictive performance of ML models, it would
5 be necessary to use big data and more important contributing factors.

6 In this study, the incidence rate of postprocedural ischemic events was 27%, which was
7 similar to the results of previous studies [29]. Most of these postprocedural ischemic events
8 were asymptomatic DWI small lesions. Although such asymptomatic DWI lesions are
9 sometimes underestimated in clinical practice, it was reported that the volume of postoperative
10 DWI lesions could correlate with cognitive deterioration at 6 months after CEA or CAS [36].
11 Accordingly, even if the lesion is small, we should try to avoid the postoperative DWI lesions.
12 In general, MRI DWI can detect ischemic changes for at least 2 weeks after a stroke onset, and
13 the postoperative DWI lesions have been considered to be derived from cerebral microembolism
14 during CAS or CEA [25]. Thus, the difference in the timing of postoperative MRI between the
15 CEA and CAS groups should not have a significant impact on our results.

16 The number of studies using ML to explore stroke and neurosurgical diseases has increased
17 rapidly over the past decade. The most frequently applied algorithms are ANN, logistic
18 regression, random forest, and SVM [2,12,32], which suggests that our model selection is
19 appropriate. In our study, the predictive performance of the XGBoost was better than those of
20 the ANN, logistic regression, SVM, and random forest. Although the superior performance of
21 the GBDT model has been shown in many data-science contests in recent years [33], there is
22 still very little research on neurosurgery or stroke using the GBDT model [22,33]. The ensemble
23 model, which had the highest ROC AUC on our training data, did not perform so well on the
24 test data, probably because of overfitting to the training data.

1

2 **Feature Importance**

3 XGBoost, which showed the best predictive performance, identified ICA-PSV, serum
4 LDL-cholesterol value, and procedure (CEA or CAS), as the most effective factors. The
5 procedure type, CEA or CAS, has previously been reported as a potential predictor of
6 periprocedural stroke [4,18,20,21]. Thus, it seems reasonable that the procedure type is the third
7 most important factor in this study. However, no studies have suggested that ICA-PSV or serum
8 LDL-cholesterol value are associated to postprocedural outcome. Although ICA-PSV should
9 reflect the degree of stenosis, many studies reported that the degree of stenosis is not associated
10 with postprocedural outcome [24,28]. ICA-PSV might be used in our model as a valid predictor
11 in combination with other clinical factors. In a study investigating the components of emboli
12 captured by the filter-protection device during CAS, more postoperative DWI high intensity
13 lesions were observed in patients whose main component of the emboli was cholesterol [17]. It
14 may be inferred from such results that serum LDL-cholesterol might be related to
15 post-procedural outcome because it should be possible for a higher serum LDL-cholesterol
16 value to leads to a higher cholesterol content of the plaque, which produces emboli. In this study,
17 the presence and intensity of lipid-lowering therapies, such as statins, were not evaluated. How
18 LDL-cholesterol levels and lipid-lowering therapies are associated with clinical outcome need
19 to be further investigated in future studies.

20

21 **Comparison of clinician and AI performance**

22 Many studies have compared clinicians and AI with image interpretation or diagnostic
23 performance and have shown that ML models are equivalent to or superior to specialists [16,34].
24 However, there were very few studies that compared the predictive performance of clinicians

1 and AI as in this study [15]. The sensitivity of the surgeons' prediction was higher than our ML
2 models, presumably because the surgeons have experience of more than 143 patients, which
3 comprise the dataset that our ML models learned. Therefore, instead of a difference in predictive
4 ability, the number of experienced and learned cases might have influenced the results. The
5 advantage of an ML model is that if only the question and answer are provided correctly, it can
6 learn an enormous number of cases that one surgeon would not be able to experience. Therefore,
7 in future, as the number of training cases increases, the predictive performance of the ML model
8 would further improve.

9

10 **Limitations**

11 One of the main limitations of our study is the small sample size. Although a larger sample
12 size would be needed for more accurate ML models, a previous report suggested that 80–560
13 samples are required for ML algorithms excluding deep neural networks, and the required
14 sample size depends on the dataset and sampling method [19]. Furthermore, a systematic review
15 of AI in neurosurgery has shown that the median number of patients in each study was 120 [2].
16 Thus, the sample size might not be insufficient for our ML models (excepting the ANN model).
17 Second, several potentially important clinical parameters, such as tandem stenotic lesions or
18 other inappropriate anatomical features for the procedure, were not considered in this study.
19 Third, optimizing hyperparameters for neural network models is generally difficult because they
20 have many hyperparameters that need to be adjusted. In addition to hyperparameters, the neural
21 network architecture should be optimized for better performance [35]. In this study, because the
22 ANN was hand-tuned by multiple trial-and-error sessions, a more effective hyperparameter set
23 might be found by other, more sophisticated optimization methods. Finally, because the dataset
24 was collected retrospectively from a single institution, it is prone to selection bias, and our ML

1 models may not be applicable to other institutions where different treatment strategies or patient
2 demographics might exist. Although internal validation was applied with repeated
3 cross-validation and bootstrap methods, further external validation is necessary in another
4 setting that differs in time or place to validate the performance of our prognostic models.

5

6 **Conclusion**

7 We developed a post-CEA/CAS outcome prediction model with a performance comparable
8 to that of surgeons. The XGBoost model showed the best predictive performance which
9 achieved more than 85% accuracy, and the most contributing factors were ICA-PSV, serum
10 LDL-cholesterol value, and procedure (CEA or CAS). Our model can help surgeons to
11 determine whether the patient suitable for CEA or CAS based on the calculated probability
12 estimates for postprocedural ischemic event. Larger datasets and analysis of potential prognostic
13 factors would be necessary to further improve the predictive performance of the ML models.

14

15 **Conflict of Interest**

16 None.

17

18 **References**

- 19 1. Brott TG, Howard G, Roubin GS, Meschia JF, Mackey A, Brooks W, et al (2016) Long-Term
20 Results of Stenting versus Endarterectomy for Carotid-Artery Stenosis. *N Engl J Med*
21 374:1021-1031. doi:10.1056/NEJMoa1505215
- 22 2. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, et al (2019)
23 Machine learning applications to clinical decision support in neurosurgery: an artificial
24 intelligence augmented systematic review. *Neurosurg Rev*. doi:10.1007/s10143-019-01163-8

- 1 3. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. Paper presented at
2 the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge
3 Discovery and Data Mining,
- 4 4. Choi JC, Johnston SC, Kim AS (2015) Early outcomes after carotid artery stenting compared
5 with endarterectomy for asymptomatic carotid stenosis. *Stroke* 46:120-125.
6 doi:10.1161/STROKEAHA.114.006209
- 7 5. Dua A, Romanelli M, Upchurch GR, Jr., Pan J, Hood D, Hodgson KJ, et al (2016) Predictors
8 of poor outcome after carotid intervention. *J Vasc Surg* 64:663-670.
9 doi:10.1016/j.jvs.2016.03.428
- 10 6. Favre JP, Nourissat A, Duprey A, Nourissat G, Albertini JN, Becquemin JP (2008)
11 Endovascular treatment for carotid artery stenosis after neck irradiation. *J Vasc Surg*
12 48:852-858. doi:10.1016/j.jvs.2008.05.069
- 13 7. Ferguson GG, Eliasziw M, Barr HW, Clagett GP, Barnes RW, Wallace MC, et al (1999) The
14 North American Symptomatic Carotid Endarterectomy Trial Surgical Results in 1415
15 Patients. *Stroke* 30:1751-1758. doi: 10.1161/01.str.30.9.1751
- 16 8. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH (2012) Predicting sample size required for
17 classification performance. *BMC Med Inform Decis Mak* 12:8. doi: 10.1186/1472-6947-12-8
- 18 9. Friedell ML, Clark JM, Graham DA, Isley MR, Zhang XF (2008) Cerebral oximetry does not
19 correlate with electroencephalography and somatosensory evoked potentials in determining
20 the need for shunting during carotid endarterectomy. *J Vasc Surg* 48:601-606.
21 doi:10.1016/j.jvs.2008.04.065
- 22 10. Gurm HS, Yadav JS, Fayad P, Katzen BT, Mishkel GJ, Bajwa TK, et al (2008) Long-term
23 results of carotid stenting versus endarterectomy in high-risk patients. *N Engl J Med*
24 358:1572-1579.

- 1 11. Halm EA, Tuhim S, Wang JJ, Rockman C, Riles TS, Chassin MR (2009) Risk factors for
2 perioperative death and stroke after carotid endarterectomy: results of the new york carotid
3 artery surgery study. *Stroke* 40:221-229. doi:10.1161/STROKEAHA.108.524785
- 4 12. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH (2019) Machine Learning-Based Model
5 for Prediction of Outcomes in Acute Stroke. *Stroke* 50:1263-1265.
6 doi:10.1161/STROKEAHA.118.024293
- 7 13. Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, et al
8 (2014) Guidelines for the prevention of stroke in patients with stroke and transient ischemic
9 attack: a guideline for healthcare professionals from the American Heart
10 Association/American Stroke Association. *Stroke* 45:2160-2236.
11 doi:10.1161/STR.0000000000000024
- 12 14. Koch M (2018) Artificial Intelligence Is Becoming Natural. *Cell* 173:531-533.
13 doi:10.1016/j.cell.2018.04.007
- 14 15. Lammers RL, Hudson DL, Seaman ME (2003) Prediction of traumatic wound infection with
15 a neural network-derived decision model. *Am J Emerg Med* 21:1-7.
16 doi:10.1053/ajem.2003.50026
- 17 16. Li L, Chen Y, Shen Z, Zhang X, Sang J, Ding Y, et al (2020) Convolutional neural network
18 for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric*
19 *Cancer* 23:126-132. doi:10.1007/s10120-019-00992-2
- 20 17. Maekawa K, Shibata M, Nakajima H, Kitano Y, Seguchi M, Kobayashi K, et al (2018)
21 Cholesterol Crystals in Embolic Debris are Associated with Postoperative Cerebral
22 Embolism after Carotid Artery Stenting. *Cerebrovasc Dis* 46:242-248.
23 doi:10.1159/000495795
- 24 18. Mantese VA, Timaran CH, Chiu D, Begg RJ, Brott TG, Investigators C (2010) The Carotid

- 1 Revascularization Endarterectomy versus Stenting Trial (CREST): stenting versus carotid
2 endarterectomy for carotid disease. *Stroke* 41:S31-34.
3 doi:10.1161/STROKEAHA.110.595330
- 4 19. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al (2015)
5 Transparent Reporting of a multivariable prediction model for Individual Prognosis or
6 Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 162:W1-73.
7 doi:10.7326/M14-0698
- 8 20. Moresoli P, Habib B, Reynier P, Secrest MH, Eisenberg MJ, Filion KB (2017) Carotid
9 Stenting Versus Endarterectomy for Asymptomatic Carotid Artery Stenosis: A Systematic
10 Review and Meta-Analysis. *Stroke* 48:2150-2157. doi:10.1161/STROKEAHA.117.016824
- 11 21. Muller MD, von Felten S, Algra A, Becquemin JP, Brown M, Bulbulia R, et al (2018)
12 Immediate and Delayed Procedural Stroke or Death in Stenting Versus Endarterectomy for
13 Symptomatic Carotid Stenosis. *Stroke* 49:2715-2722.
14 doi:10.1161/STROKEAHA.118.020684
- 15 22. Oermann EK, Rubinsteyn A, Ding D, Mascitelli J, Starke RM, Bederson JB, et al (2016)
16 Using a Machine Learning Approach to Predict Outcomes after Radiosurgery for Cerebral
17 Arteriovenous Malformations. *Sci Rep* 6:21161. doi:10.1038/srep21161
- 18 23. Rosenfield K, Matsumura JS, Chaturvedi S, Riles T, Ansel GM, Metzger DC, et al (2016)
19 Randomized Trial of Stent versus Surgery for Asymptomatic Carotid Stenosis. *N Engl J Med*
20 374:1011-1020. doi:10.1056/NEJMoa1515706
- 21 24. Setacci C, Chisci E, Setacci F, Iacoponi F, de Donato G, Rossi A (2010) Siena carotid artery
22 stenting score: a risk modelling study for individual patients. *Stroke* 41:1259-1265.
23 doi:10.1161/STROKEAHA.110.578583
- 24 25. Skjelland M, Krohg-Sørensen K, Tennøe B, Bakke SJ, Brucher R, Russell D (2009)

- 1 Cerebral microemboli and brain injury during carotid artery endarterectomy and stenting.
2 Stroke 40:230-234. doi:10.1161/STROKEAHA.107.513341
- 3 26. Taussky P, Hanel RA, Meyer FB (2011) Clinical considerations in the management of
4 asymptomatic carotid artery stenosis. Neurosurg Focus 31:E7.
5 doi:10.3171/2011.9.FOCUS11222
- 6 27. The Japan Stroke Society SGC (2015) Japanese guidelines for the management of stroke
7 2015. Kyowa Kikaku, Tokyo
- 8 28. Theiss W, Hermanek P, Mathias K, Bruckmann H, Dembski J, Hoffmann FJ, et al (2008)
9 Predictors of death and stroke after carotid angioplasty and stenting: a subgroup analysis of
10 the Pro-CAS data. Stroke 39:2325-2330. doi:10.1161/STROKEAHA.108.514356
- 11 29. Traenka C, Engelter ST, Brown MM, Dobson J, Frost C, Bonati LH (2019) Silent brain
12 infarcts on diffusion-weighted imaging after carotid revascularisation: A surrogate outcome
13 measure for procedural stroke? A systematic review and meta-analysis. Eur Stroke J
14 4:127-143. doi:10.1177/2396987318824491
- 15 30. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al (2001)
16 Missing value estimation methods for DNA microarrays. Bioinformatics 17:520-525
- 17 31. Tu JV, Wang H, Bowyer B, Green L, Fang J, Kucey D (2003) Risk factors for death or
18 stroke after carotid endarterectomy: observations from the Ontario Carotid Endarterectomy
19 Registry. Stroke 34:2568-2573. doi:10.1161/01.STR.0000092491.45227.0F
- 20 32. van Os HJA, Ramos LA, Hilbert A, van Leeuwen M, van Walderveen MAA, Kruijff ND, et
21 al (2018) Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke:
22 Potential Value of Machine Learning Algorithms. Front Neurol 9:784.
23 doi:10.3389/fneur.2018.00784
- 24 33. Xie Y, Jiang B, Gong E, Li Y, Zhu G, Michel P, et al (2019) Use of Gradient Boosting

- 1 Machine Learning to Predict Patient Outcome in Acute Ischemic Stroke on the Basis of
2 Imaging, Demographic, and Clinical Information. *AJR Am J Roentgenol* 212:44-51.
3 doi:10.2214/AJR.18.20260
- 4 34. Ye H, Gao F, Yin Y, Guo D, Zhao P, Lu Y, et al (2019) Precise diagnosis of intracranial
5 hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural
6 network. *Eur Radiol* 29:6191-6201. doi:10.1007/s00330-019-06163-2
- 7 35. Yoo Y (2019) Hyperparameter optimization of deep neural network using univariate
8 dynamic encoding algorithm for searches. *Knowledge-Based Systems* 178:74-83.
9 doi:10.1016/j.knosys.2019.04.019
- 10 36. Zhou W, Baughman BD, Soman S, Wintermark M, Lazzeroni LC, Hitchner E, et al (2017)
11 Volume of subclinical embolic infarct correlates to long-term cognitive changes after carotid
12 revascularization. *J Vasc Surg* 65:686-694. doi:10.1016/j.jvs.2016.09.057

13

14 **Figure Captions**

15 **Fig. 1** Flow diagram describing the general framework of the study. Models were built using the
16 training dataset. The test dataset was used for measuring the predictive performance and
17 comparison with the surgeons. CAS, carotid artery stenting; CEA, carotid endarterectomy;
18 ICA-PSV, internal carotid artery peak systolic velocity; kNN, k-nearest-neighbours; LDL, low
19 density lipoprotein; PPV, positive predictive value

20

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

1

2 **Supplementary Information**

3 **Supplemental Table 1** Patient characteristics of CEA group and CAS group.

4 **Supplemental Table 2** Optimized hyperparameters of five machine learning models.

5

1 Table 1. Patient characteristics

Variable	Before imputation	After imputation		<i>p</i> value
	Total (n=165)	Training data (n=143)	Test data (n=22)	
Age, year, mean (SD)	74.2 (7.61)	73.8 (7.82)	76.7 (5.62)	0.04
Male sex, n (%)	141 (85)	122 (85)	19 (86)	1
pre-treatment mRS, median [IQR]	0 [0-1]	0 [0-1]	0.5 [0-1.75]	0.55
LDL-cholesterol, mg/dL, mean (SD)	90.0 (29.5)	90.4 (30.1)	87.1 (23.3)	0.56
Prior medical histories, n (%)				
Hypertension	136 (82)	118 (83)	18 (82)	1
Diabetes mellitus	58 (35)	50 (35)	8 (36)	1
Acute coronary syndrome	53 (32)	42 (29)	11 (50)	0.08
Peripheral artery disease	28 (17)	24 (17)	4 (18)	0.77
Anatomical and pathophysiological features				
Contralateral occlusion, n (%)	14 (8.5)	11 (7.7)	3 (14)	0.40
Stenosis at a high position, n (%)	13 (7.9)	12 (8.4)	1 (4.5)	1
Type III Aorta, n (%)	64 (39)	56 (39)	8 (36)	1
ICA-PSV, cm/sec, mean (SD)	276 (130)	272 (126)	291 (147)	0.56
Mobile plaque, n (%)	19 (12)	15 (10)	4 (18)	0.29
Plaque ulceration, n (%)	39 (24)	36 (25)	3 (14)	0.29
plaque with hyperintense signal on TOF, n (%)	61 (37)	54 (38)	7 (32)	0.64
Previous neck irradiation, n (%)	15 (9.1)	13 (9.1)	2 (9.1)	1
Symptomatic, n (%)	64 (39)	56 (39)	8 (36)	1
Crescendo TIA or stroke in evolution, n (%)	10 (6.1)	8 (5.6)	2 (9.1)	0.62
Treatment, CEA, n (%)	95 (58)	85 (59)	10 (45)	0.25
Outcome, n (%)				
Ischemic stroke within 30 days	45 (27)	42 (29)	3 (14)	0.13
Major ischemic stroke	3 (1.8)	2 (1.4)	1 (4.5)	0.36
Minor ischemic stroke	3 (1.8)	3 (2.1)	0 (0)	1
Asymptomatic DWI hyperintense lesions	39 (24)	37 (26)	2 (9.1)	0.11

Follow-up duration, days, mean (SD)	833 (565)	921 (555)	259 (259)	<.0001
-------------------------------------	-----------	-----------	-----------	--------

- 1 CEA = carotid endarterectomy; DWI = diffusion weighted imaging; IQR = interquartile range;
- 2 LDL = low density lipoprotein; mRS = modified Rankin scale; ICA-PSV = internal carotid
- 3 artery-peak systolic velocity; TIA = transient ischemic attack; TOF = time-of-flight.
- 4

1 Table 2. Prediction results on the training dataset evaluated by repeated 5-fold cross validation
 2 and sorted by ROC AUC

Model	ROC AUC	Sensitivity (%)	Specificity (%)	PPV (%)	Accuracy (%)
	mean (95%CI)	mean (95%CI)	mean (95%CI)	mean (95%CI)	mean (95%CI)
Ensemble model *	0.739 (0.714 - 0.764)	15.1 (11.3 – 18.8)	97.4 (96.2 – 98.6)	75.1 (65.6 – 84.7)	72.7 (71.3 – 74.0)
XGBoost	0.719 (0.692 - 0.746)	14.6 (10.3 – 19.0)	94.9 (93.1 – 96.7)	54.5 (44.6 – 64.3)	70.8 (69.1 – 72.4)
Logistic regression	0.702 (0.671 - 0.732)	26.5 (22.4 – 30.5)	95.5 (94.2 – 96.8)	71.2 (62.7 – 79.7)	74.8 (73.1 – 76.4)
Neural network	0.692 (0.659 - 0.726)	29.0 (23.8 – 34.1)	91.2 (89.1 – 93.3)	61.3 (54.3 – 68.3)	72.5 (70.8 – 74.3)
Random forest	0.683 (0.653 - 0.712)	22.0 (18.0 – 26.0)	98.3 (97.4 – 99.2)	78.9 (69.2 – 88.7)	75.4 (74.0 – 76.8)
SVM	0.680 (0.650 - 0.711)	36.2 (32.2 – 40.3)	85.3 (83.5 – 87.1)	51.7 (46.6 – 56.8)	70.6 (68.8 – 72.4)

3

4 * Ensemble model is created by using XGBoost, neural network, and logistic regression, which
 5 are the three most highest ROC AUC models.

6 ROC AUC = area under the receiver operating characteristic curve; PPV = positive predictive
 7 value; SVM = support vector machine.

8

1 Table 3. Prediction results on the test dataset evaluated using the bootstrap technique and sorted
 2 by accuracy.

Model	Sensitivity (%) mean (95%CI)	Specificity (%) mean (95%CI)	PPV (%) mean (95%CI)	Accuracy (%) mean (95%CI)
XGBoost	31.9 (30.0 – 33.8)	94.6 (94.2 – 94.9)	47.2 (44.9 – 49.6)	86.2 (85.7 – 86.6)
Neural network	4.1 (3.3 – 4.8)	95.4 (95.3 – 95.5)	12.4 (10.4 – 14.4)	83.3 (82.9 – 83.8)
Ensemble model *	31.8 (29.9 – 33.7)	89.6 (89.2 – 90.0)	32.0 (30.0 – 33.9)	82.0 (81.5 – 82.5)
Random forest	34.0 (32.1 – 35.9)	84.3 (83.8 – 84.8)	25.1 (23.6 – 26.6)	77.7 (77.1 – 78.2)
SVM	34.0 (32.1 – 35.9)	79.1 (78.6 – 79.7)	19.9 (18.7 – 21.1)	73.2 (72.6 – 73.7)
Logistic regression	34.0 (32.1 – 35.9]	79.0 (78.4 – 79.6)	20.2 (18.9 – 21.5)	73.0 (72.4 – 73.6)
Surgeons **	41.7 (0 – 92.4)	75.0 (60.7 – 89.3)	20.1 (0 – 42.8)	70.5 (57.9 – 83.0)

3
 4 * Ensemble model is created by using XGBoost, neural network, and logistic regression, which
 5 are the three most highest ROC AUC models on the training dataset.
 6 ** The average of 4 surgeon’s prediction results with 95% confidence interval.
 7 PPV = positive predictive value; SVM = support vector machine.

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

- Hyperparameter optimization with grid-search
- Performance evaluation with 10 times repeated stratified 5-fold cross validation
- Evaluation metrics: ROC AUC, sensitivity, specificity, PPV, and accuracy

Dataset for model performance test

- 22 patients underwent CEA or CAS from April to December 2018
- Outcome was observed in 3 patients

- Model validation with bootstrap method (n=22: 1000 replicates)
- Performance comparison with 4 surgeons
- Evaluation metrics: sensitivity, specificity, PPV, and accuracy

Total gain of the XGBoost model

