ORIGINAL ARTICLE

Detection of Asymptomatic Carotid Artery Stenosis in High-Risk Individuals of Stroke Using a Machine-Learning Algorithm

Junxiong Yin¹, Cheng Yu², Lixia Wei², Chuanyong Yu¹,
Hongxing Liu¹, Mingyang Du¹, Feng Sun¹,
Chongjun Wang², Xiaoshan Wang^{1*}

¹Department of Neurology, Brain Hospital Affiliated to Nanjing
Medical University, Nanjing 210029, China

²Department of Computer Science and Technology,
Nanjing University, Nanjing 210093, China

Key words: high-risk population; stroke; asymptomatic carotid stenosis; risk factors; machine learning

Objective Asymptomatic carotid stenosis (ACS) is closely associated to the incidence of severe cerebrovascular diseases. Early identifying the individuals with ACS and its associated risk factors could be beneficial for primary prevention of stroke. This study aimed to investigate a machine-learning algorithm for the detection of ACS among high-risk population of stroke based on the associated risk factors.

Methods A novel model of machine learning was utilized to screen the associated predictors of ACS based on 30 potential risk factors. The algorithm of this model adopted a random forest pattern based on the training data and then was verified using the testing data. All of the original data were retrieved from the China National Stroke Screening and Prevention Project (CNSSPP), including demographic, clinical and laboratory characteristics. The individuals with high risk of stroke were enrolled and randomly divided into a training group and a testing group at a ratio of 4:1. The identification of carotid stenosis by carotid artery duplex scans was set as the golden standard. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) was used to evaluate the efficacy of the model in detecting ACS.

Results Of 2841 high risk individual of stroke enrolled, 326 (11.6%) were diagnosed as ACS by ultrasonography. The top five risk factors contributing to ACS in this model were identified as family history of dyslipidemia, high level of low-density lipoprotein cholesterol (LDL-c), low level of high-density lipoprotein cholesterol (HDL-c), aging, and low body

Received December 16, 2019; accepted March 24, 2020; published online September 28, 2020.

*Corresponding author Tel: 86-13852297339; Fax: 86-25-82296208; E-mail: professor_wxs@163.com.

Fund supported by the Medical Science and Technology Development Foundation (YKK18114), and the General Social Development Medical and Health Project of Nanjing Science and Technology Commission (201803029).

mass index (BMI). Their weights were 11.8%, 7.6%, 7.1%, 6.1%, and 6.1%, respectively. The total weight of the top 15 risk factors was 85.5%. The AUC values of the model for detecting ACS with training dataset and testing dataset were 0.927 and 0.888, respectively.

Conclusions This study demonstrated that the machine-learning algorithm could be used to identify the risk factors for ACS among high risk population of stroke. Family history of dyslipidemia may be the most important risk factor for ACS. This model could be a suitable tool to optimize the clinical approach for the primary prevention of stroke.

ASYMPTOMATIC carotid stenosis (ACS) is defined as a significant ($\geq 50\%$) narrowing of carotid artery in patients without ipsilateral neurological deficit symptoms during the past 6 months.^[1] ACS is often neglected clinically. Under asymptomatic conditions, ACS patients could develop cerebral infarction rapidly with serious consequences. Several studies have confirmed that ACS is an independent risk factor for cerebral infarction.^[2, 3] In order to identify ACS as early as possible, extensive screening is necessary among "normal" population. However, given the cost-effectiveness for the screening, the U.S. Preventive Services Task Force (USPSTF) and the American Society of Neuroimaging both do not recommend screening ACS in general adult population.^[4-6] Meanwhile, other researchers suggested the targeted screening in people with multiple risk factors for atherosclerosis.^[7] Therefore, most previous studies chose individuals with increased risk for stroke as screening subjects, but how to identify high-risk subjects effectively and conveniently is still uncertain.

Machine-learning model was initially designed to analyze largely complex datasets. Compared with traditional statistical methods, machine learning has advantages in dealing with complex data. Some researches^[8-10] showed that machine learning in prediction model worked better than those of general statistical methods. Nowadays, it has been widely used to learn complex relationships from data to predict cardio-cerebrovascular disease and generate medical imaging reports.^[8, 9] Recently, machine-learning techniques for predicting carotid plaque progression have been successfully developed, which mainly involve analyzing vulnerable plaque progression in apoE^{-/-} mice.^[10] However, the study was based on animal models and overlooked some traditional risk factors, such as hypertension and diabetes. Several clinical studies^[11, 12] demonstrated that elder age, smoking, high blood pressure and diabetes were independent risk factors for carotid stenosis ($>50\%$ stenosis), and have generate some predictive models for

ACS. Nevertheless, the analytic approaches above were mainly traditional frequentist biostatistical methods.

Machine learning consists a variety of algorithms, such as support vector machine, decision tree, deep learning, ensemble learning, neural network learning, and so on. Random forest is a kind of machine-learning algorithm combining ensemble learning and decision tree.^[13] By combining several weak classifiers, the final result can be voted or averaged, which makes the results of the whole model with higher accuracy and better generalization performance.

The aim of this study was to identify the effectiveness of random forest model in differentiating ACS patients from high risk individuals of stroke using data from the China National Stroke Screening and Prevention Project (CNSSPP).

MATERIALS AND METHODS

Study design and participants

This cross-sectional study was designed and approved by the institutional review board of Nanjing Brain Hospital (2017-kyy119-01). All participants received information on the study and provided written informed consents. The data of this study was based on data from Nanjing Brain Hospital, one of study centers participating the CNSSPP. During 2012-2016, more than 34000 residents participated in the CNSSPP. According to the risk factors screening, 5250 participants were identified as high risk individuals for stroke. Among them, 2309 cases were excluded due to a history of stroke or transient ischemic attack (TIA), and 100 cases were excluded for the incomplete data. Thus, totally 2481 high risk individual for strokes were enrolled in the current study.

The screening work complied with the protocol of CNSSPP.^[14] A high-risk individual for stroke was defined as: ≥ 40 years old with at least three of the following risk factors: hypertension, atrial fibrillation, current smoking, dyslipidemia, diabetes mellitus (DM),

physical inactivity, overweight ($\text{BMI} \geq 26 \text{ kg/m}^2$),^[8] and family history of stroke.

Carotid vascular ultrasound scanning

The high risk individuals were subjected to take carotid artery duplex scans following Chinese stroke vascular ultrasound examination guidelines,^[15] which consisted of ultrasound imaging of the distal common carotid artery, bulb, proximal internal, and external carotid arteries, with Doppler signal evaluation for 3 to 5 beats in each location on both sides. Plaque was defined as intima-media thickness (IMT) greater than 1.5 mm based on Doppler-derived.^[16] According to previous studies,^[12, 17] ACS was identified if diameter of carotid lumen decreased by $\geq 50\%$, along with at least one of the following indicators: a peak systolic velocity (PSV) $\geq 125 \text{ cm/s}$, end diastolic velocity $\geq 40 \text{ cm/s}$, or the ratio of internal carotid artery (ICA) PSV/common carotid artery (CCA) PSV > 2.0 . The carotid duplex scans were performed by four experienced registered vascular technicians who were independent of clinical information.

Data collection and management

Each candidate was asked to complete a standardized questionnaire designed by CNSSPP *via* face-to-face interviews by well-trained staff before the carotid artery duplex scan. Information we collected for 53 risk factors included demographic characteristics, lifestyle factors, physical examination results, medical history, and family history of stroke. Subjects also took fasting blood tests for laboratory indexes, such as plasma glucose (FPG), homocysteine (Hcy), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c), hemoglobin A1c (HbA1c), triglyceride (TG), etc.

Some risk factors of the record were removed from analyses due to missing value, duplicated or invalid values. Consequently, 30 risk factors of 2841 subjects retained for further random forest analyses

(Supplementary **Table S1**), which consisted of atrial fibrillation, diabetes mellitus (DM), hypertension, dyslipidemia, overweight, physical activity, family history of stroke/coronary heart disease (CHD)/DM/dyslipidemia, smoking, alcohol drinking, diet habits (positive/negative/blend), occupation, PayStyle categories, etc.

Random forest

We trained a machine-learning algorithm using random forest to distinguish the patients with ACS based on the data from CNSSPP. Some of these individuals were tested using this algorithm to evaluate the accuracy rate. The random forest was implemented by the researchers from the department of Computer Science and Technology, Nanjing University.

The good results of random forest mainly depended on "random" and "forest". "Random" made it resist over-fitting, and "forest" made it more accurate, as shown in **Figure 1**. The weak classifier used in random forest was CART tree, also called classification regression tree. A decision tree was a predictive model, representing a mapping relationship between object attributes and object values. Each node in the tree represented an object. Each branching path represented a possible attribute value. Each leaf node corresponded to the value of the object represented by the path that the root node went through the leaf node. Decision tree had only a single output. A risk classification for unilateral carotid stenosis was presented in **Figure 2** for illustration.

Random forest was applied by using Tensorflow code style.^[18] The running environment of random forest graph algorithm program was as follows: system version, Ubuntu 16.4; python version, 3.6; CPU, Intel (R) Xeon (R) CPU e5-2630 V4 @ 2.20GHz; memory, 320G. We inputted all the data into the edited random forest algorithm, setting the diagnosis of CAS on carotid artery color Doppler ultrasound as the task marker. The processes of modeling and verification are presented as in **Figure 3**.

Figure 1. The architecture of ensemble learning.

Figure 2. Decision tree: risk classification for unilateral carotid stenosis.

Algorithm 1 Random Forest

Input:

The set of training data $S = \{(X_i, Y_i), i = 1, 2, \dots, n\}$,
 $(X, Y) \in R^d \times R$;

Output:

The set of trees $\{h_i, i = 1, 2, \dots, N_{\text{tree}}\}$;

- 1: for $i = 1, 2, \dots, N_{\text{tree}}$ do:
- 2: Sampling the original training set S by Bootstrap Method to generate a new training set S_i ;
- 3: Randomly select k features from d features;
- 4: Each node of the decision tree selects the best features for splitting based on the Gini index of k features;
- 5: According to the previous step, until all the training examples of the node belong to the same class;
- 6: Save this decision tree based on data set S_i ;
- 7: end for
- 8: The Classification results $f(x_i) = \text{majorityvote}\{h(x_i)\}_{i=1}^{N_{\text{tree}}}$,
 x_i is the test sample.

Figure 3. The algorithm using CNSSPP data to detect ACS.

Training and validation

First, we randomly sampled the original data set to generate a new data set, and then selected the feature that causes the largest change in the Gini index when the node was split to grow decision tree. Stop generating until the samples in each node were of the same category to get a trained decision tree. Repeated the above steps until a specified number of decision trees were generated to form a random forest.

After each sample in the test dataset was judged by each decision tree in the trained random forest model, the prediction result was obtained. They are subsequently voted by the majority to obtain the final

prediction label of the model for the sample. The final predictions were compared to the true label (diagnosis by ultrasound) of the sample for receiver operating characteristic (ROC) evaluation. The area under the curve (AUC) was used as the model measurement standard.

Data preprocessing

The data preprocessing procedures of modeling were explained briefly as follows:

Firstly, we removed similar, duplicated, or not relevant variables, then we normalized the variable from text to number in the form of one to one, and used -1 to fill in the missing value. Then the dataset was split randomly into a section of 80% as training set and a section of 20% as testing set.

In the second step, we gave the values of several model hyper parameters. Then sample extraction was used for the training set to build a new sample set, which generated a new decision tree. We repeated the process to generate a specified number of decision trees to form a forest.^[19]

The third step was to adjust the hyper parameters of model, and then to repeat the second step to generate numerous numbers of different random forests. The testing set was used to evaluate the performance of each detective model based on accuracy, true positive rate, false positive rate. The effectiveness of the model was evaluated by area under receiver operating characteristic curve.

The fourth step involved calculation of variable importance. After identifying the best performing model among the large number of models generated, we calculated the Gini index of each node in each decision tree, which was the probability that two samples randomly selected from the node m had inconsistent class labels. Then the change of Gini index caused by branching of the node was calculated. Finally, the importance of the feature, also called "weight", was calculated by the branching in which the feature participated.

Parameter optimization

To characterize the random forest algorithm, the following important parameters were used in this study:

- 1) N_estimators: the maximum number of iterations of a weak learner, or the largest number of the weak learners. This referred to the number of decision trees in a random forest.
- 2) Max_features: the maximum number of

features considered when partitioning nodes. The value was used to flexibly control the generation time of decision tree.

3) Max_depth: the maximum depth of decision tree. It could limit the times of node splitting.

4) Min_samples_split: the minimum number of samples required for internal node splitting, which limited the conditions for the continuation of subtree splitting.

5) Min_samples_leaf: minimum sample number for leaf node. If the number of samples for a leaf node was less than the specified number, it would be pruned together with sibling nodes.

Statistical analysis

The demographic, clinical and laboratory results between ACS group and non-ACS group were compared by using *t*-test for continuous variables and Chi-square test for categorical variables. Continuous variables were described as mean and standard deviation (SD), and categorical variables were described as percentages. All statistical analyses were performed using SPSS (Version 20.0, SPSS Inc., Chicago, IL, USA). *P* value <0.05 was considered statistically significant.

RESULTS

Of 2841 individuals enrolled in the study, according to ultrasound findings, the carotid plaque were found in 1000 (35.2%), and the ACS was found in 326 (11.5%). Comparison of the characteristics of ACS and non-ACS participants (**Table 1**) revealed the significant differences existed between the two groups in sex, age, marriage, BMI, waistline, IsDrink, overweight, whether concurrent with atrial fibrillation, DM, hypertension, whether has family history of stroke, CHD, DM, dyslipidemia, and the levels of FPG, HDL-c, LDL-c.

Through calculation and optimization, the top five risk factors for ACS in high-risk population of stroke and their corresponding weights were as follow: positive family history of dyslipidemia, 11.8%; high level of LDL-c, 7.6%; low level of HDL-c, 7.1%; older age, 6.1%; and low BMI, 6.1%. The accumulated weight of the top 15 risk factors was 85.5%, which were presented in **Table 2**.

The parameters of this model for detecting ACS based on machine learning were as follow:

1) N_estimators:10

2) Min_samples_split:100

3) Max_features: 4

4) Min_samples_leaf: 21

5) Max_depth: 19

Based on the above parameters, we input the 30 risk factors into the random forest model. The AUC value of the model predicting ACS in training dataset was 0.927, and in testing dataset was 0.888. The ROC curves were plotted as shown in **Figure 4**.

DISCUSSION

The novel finding of our study was that the machine learning method could be applied to detect potential ACS patients among the high risk individuals of stroke. Our results further demonstrated the potential utility of random forest approach in detecting ACS with our datasets.

The area under the ROC curve of our random forest algorithm was 0.888 in testing set, which was superior to that in previous studies.^[11, 12, 20] Greco *et al.* created a model for predicting the risk of carotid artery disease based on data of carotid Duplex scans from 2,885,257 individuals in the Life Line Screening between 2003 and 2008.^[12] The independent risk factors for ACS were advanced age, smoking, peripheral arterial disease, high blood pressure, coronary artery disease, diabetes, cholesterol, etc. According to these independent risk factors, a predictive scoring system was created, and the AUC of the ROC was 0.753. This was a remarkable study which predicting ACS with a national cohort across the population. The AUC value of our present study was better than that of the Life Line Screening cohort, which was probably because: 1) the population in our study was mainly the selected high-risk individuals of stroke, whereas the latter was the general population;^[12] 2) the predictive model of the two studies were different.

There were also other studies on predictive model. Jacobowitz *et al.*^[11] created a model to predict carotid artery stenosis in population older than 60 years and had at least one of the following risk factors: history of CHD, previously diagnosed hypertension, current smoking, and family history of stroke in the first-degree relatives. The study was performed on 394 individuals, of which 9.6% had either unilateral or bilateral carotid artery stenosis. Risk factors entered into the logistic regression analysis were hypertension, cardiac disease, current smoking, and hypercholesterolemia.

Table 1. Comparison of demographic, clinical and laboratory characteristics between ACS and Non-ACS subjects (n=2841)

Characteristics	ACS (n=326)	Non ACS (n=2515)	Sub-total	χ^2 or t	P
Sex-male [n(%)]	175 (53.7)	1150 (45.7)	1325 (46.6)	7.339	<0.01
Age (years, mean $\pm$ SD)	64.0 $\pm$ 7.7	60.1 $\pm$ 8.7	60.6 $\pm$ 8.6	7.637	<0.001
Education-primary school [n(%)]	132 (40.5)	1000 (39.8)	1132 (39.8)	0.064	0.800
Marriage-married [n(%)]	292 (89.6)	1737 (69.1)	2029 (71.4)	59.444	<0.001
Physical exam index (mean $\pm$ SD)					
BMI	25.0 $\pm$ 3.3	26.0 $\pm$ 3.2	25.7 $\pm$ 3.2	-4.169	<0.001
Height (cm)	164.3 $\pm$ 7.5	163.2 $\pm$ 28.0	163.3 $\pm$ 26.5	0.712	0.476
Waistline (cm)	82.6 $\pm$ 17.2	67.1 $\pm$ 36.0	68.9 $\pm$ 34.6	7.664	<0.001
Weight (kg)	67.4 $\pm$ 9.9	68.1 $\pm$ 9.3	68.0 $\pm$ 9.4	-1.320	0.187
Concurrent diseases [n(%)]					
Atrial fibrillation	38 (11.7)	149 (5.9)	187 (6.6)	15.420	<0.001
DM	107 (32.8)	601 (23.9)	708 (24.9)	12.288	0.001
Hypertension	275 (84.4)	1971 (78.4)	2246 (79.1)	6.246	0.05
Hypercholesterolemia	129 (39.6)	961 (38.2)	1090 (38.4)	0.226	0.635
Life style [n(%)]					
Overweight	84 (25.8)	1171 (46.6)	1255 (44.2)	50.599	<0.001
Current smoking	118 (36.2)	855 (34.0)	973 (34.2)	0.620	0.431
Current drinking	77 (23.6)	312 (12.4)	389 (13.7)	30.710	<0.001
Past smoking	62 (19.0)	520 (20.7)	582 (20.5)	0.487	0.485
Diet-positive	47 (14.4)	449 (17.9)	496 (17.5)	2.364	0.124
Lack Physical activity	235 (72.1)	1935 (76.9)	2170 (76.4)	3.767	0.053
Family history [n(%)]					
Stroke	86 (26.4)	798 (31.7)	884 (31.1)	3.853	<0.05
CHD	31 (9.5)	119 (4.7)	150 (5.3)	13.172	<0.001
DM	48 (14.7)	241 (9.6)	289 (10.2)	8.349	<0.01
Dyslipidemia	2 (0.6)	71 (2.8)	73 (2.6)	4.780	<0.05
Laboratory results (mmol/L, mean $\pm$ SD)					
FPG	6.6 $\pm$ 2.1	6.1 $\pm$ 1.8	6.1 $\pm$ 1.9	4.503	<0.001
Hcy	11.3 $\pm$ 7.7	11.1 $\pm$ 7.7	11.1 $\pm$ 7.7	0.506	0.613
HDL-c	1.4 $\pm$ 0.4	1.6 $\pm$ 0.8	1.5 $\pm$ 0.8	-4.039	<0.001
LDL-c	3.2 $\pm$ 0.8	3.0 $\pm$ 0.8	3.0 $\pm$ 0.8	4.878	<0.001
TC	5.1 $\pm$ 0.9	5.1 $\pm$ 1.0	5.1 $\pm$ 1.0	1.006	0.315
TG	1.8 $\pm$ 1.3	1.7 $\pm$ 1.3	1.7 $\pm$ 1.3	0.327	0.744

ACS, asymptomatic carotid stenosis; DM, diabetes mellitus; CHD, coronary heart disease; FPG, fasting blood glucose; Hcy, homocysteine; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglyceride; BMI, body mass index.

The model showed that, the prevalence of carotid stenosis was 1.8% with no risk factor, 5.8% with one risk factor, 13.5% with two risk factors, 16.7% with three risk factors, and 66.7% with four risk factors. Qureshi *et al.* [20] created a simple scoring system to identify high risk individuals for ACS based on routine data collected in a community health screening. They evalu-

ated 1331 unselected volunteers who had no previous history of stroke, TIA, or carotid artery surgery, and found four risk factors that were significantly associated with ACS: age older than 65, current smoking, coronary artery disease, and hypercholesterolemia. The AUC of this system was 0.706.

In recent years, many studies using various

Table 2. Risk factors and the weights for predicting moderate to severe carotid stenosis by algorithm model based on random forest machine learning

Rank	Risk factor	Weight	Rank	Risk factor	Weight
1	Fh-Dyslipidemia	0.118606411	16	Dyslipidemia	0.014717
2	LDL-c	0.076347664	17	Education level	0.013718
3	HDL-c	0.071114458	18	Marriage	0.013375
4	Age	0.061408055	19	Fh-DM	0.01327
5	BMI	0.061151368	20	Current drinking	0.013036
6	Waistline	0.059055337	21	Fh-CHD	0.011855
7	Glucose	0.05701681	22	DM	0.01076
8	Hcy	0.054882752	23	Fh-Stroke	0.007797
9	TC	0.054851202	24	Current smoking	0.007762
10	Height	0.049662527	25	Lack physical activity	0.007582
11	TG	0.047633703	26	Hypertension	0.007054
12	Weight	0.047228036	27	Past smoking	0.007031
13	Occupation	0.039596269	28	Overweight	0.006679
14	PayStyle	0.035715073	29	Sex	0.005956
15	Diet	0.020405623	30	Atrial fibrillation	0.004731

Fh, family history.

Figure 4. The ROC curves of the model for predicting ACS in high risk population of stroke. AUC, area under curve.

machine learning models to predict medical events in domain of healthcare have been published. These studies involve workflow of clinical care, cancer survival, and cardiovascular diseases.^[21-23] Some studies applied machine learning to predict carotid atherosclerosis. Hu *et al.*^[24] used machine learning to predict rapid progression of carotid atherosclerosis in 382 patients with impaired glucose tolerance, which suggested that the best machine learning method was Naïve Bayes (AUC 0.797), and machine learning could be applicable in a relatively small number of subjects. Li *et al.*^[10] predicted carotid plaque progression in ApoE^{-/-} mice using support vector machine and decision tree, and concluded

that the method could be suitable for identifying vulnerable plaque progression in mice. Yet, there has been no clinical study on detecting ACS in high-risk population of cerebral infarction using machine learning.

The results of our study showed that family history of dyslipidemia was the most important risk factor, which had not been reported yet. Risk factors for ACS reported^[20,25] include advanced age, current smoking, peripheral arterial disease, hypercholesterolemia, hypertension, diabetes mellitus, and coronary artery disease. Many studies had paid close attentions to family history of stroke/hypertension/diabetes, whereas few had looked at family history of dyslipidemia.^[26] Familial

hypercholesterolaemia is a common genetic cause of premature coronary heart disease,^[27] and can lead to cardiovascular disease (CVD). Most cases are caused by autosomal dominant mutations in low density lipoprotein receptor (LDLR) gene.^[28] The family history of dyslipidemia, as a risk factor for ACS, did not gain attention in the early-stage study with logistic regression.^[29] The result of current machine learning study suggests that more attention should be given to the family history of hyperlipidemia in studies on carotid atherosclerosis in future, especially in population with LDLR genes mutation.

The current study has several limitations. The data we retrieved from the CNSSPP were a single centric dataset, the number of subjects in this study was limited. Additionally, the model is only applicable in high risk individuals of stroke to detect ACS, and did not consider the influence of drug using, e.g., statins which is an important aspect for ACS. Furthermore, the detection model provides weight and ranking of the risk factor, but how to apply them conveniently in a practical setting needs further study.

Conflict of interests disclosed

None.

Supplementary materials

Table S1: 30 risk factors for random forest analyses.

Available online at <http://cmsj.cams.cn/EN/10.24920/003703>.

REFERENCES

1. DE Waard DD, Morris D, DE Borst GJ, et al. Asymptomatic carotid artery stenosis: who should be screened, who should be treated and how should we treat them? *J Cardiovasc Surg* 2017; 58(1):3-12. doi: 10.23736/S0021-9509.16.09770-6.
2. Faxon DP, Creager MA, Smith SC, et al. American Heart Association. Atherosclerotic Vascular Disease Conference: Executive summary: Atherosclerotic Vascular Disease Conference proceeding for healthcare professionals from a special writing group of the American Heart Association. *Circulation* 2004;109(21):2595-604. doi: 10.1161/01.CIR.0000128517.52533.DB.
3. Fayad ZA, Fuster V. Clinical imaging of the high-risk or vulnerable atherosclerotic plaque. *Circ Res* 2001; 89(4):305-16. doi: 10.1161/hh1601.095596.
4. U.S. Preventive Services Task Force. Screening for carotid artery stenosis: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2007; 147(12):854-9. doi: 10.7326/0003-4819-147-12-200712180-00005.
5. Summaries for patients. Screening for carotid artery stenosis: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014; 161(5). doi: 10.7326/P14-9030.
6. Qureshi AI, Alexandrov AV, Tegeler CH, et al; American Society of Neuroimaging; Society of Vascular and Interventional Neurology. Guidelines for screening of extracranial carotid artery disease: a statement for healthcare professionals from the multidisciplinary practice guidelines committee of the American Society of Neuroimaging; cosponsored by the Society of Vascular and Interventional Neurology. *J Neuroimaging* 2007; 17(1):19-47. doi: 10.1111/j.1552-6569.2006.00085.x.
7. Brott TG, Halperin JL, Abbara S, et al. 2011 ASA/ACCF/AHA/AANN/AANS/ACR/ASNR/CNS/SAIP/SCAI/SIR/SNIS/SVM/SVS Guideline on the management of patients with extracranial carotid and vertebral artery disease. *Stroke* 2011; 42(8):e464-540. doi: 10.1161/STR.0b013e3182112cc2. Corrected and republished: *Stroke* 2011; 42(8):e542. *Stroke* 2012; 43(8):e80.
8. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res* 2017; 121(9):1092-101. doi: 10.1161/CIRCRESA-HA.117.311312.
9. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging. *Radiographics* 2017; 37(2):505-15. doi: 10.1148/rg.2017160130.
10. Li B, Jiao Y, Fu C, et al. Contralateral artery enlargement predicts carotid plaque progression based on machine learning algorithm models in apoE^{-/-} mice. *Biomed Eng Online* 2016; 15(Suppl 2):146. doi: 10.1186/s12938-016-0265-z.
11. Jacobowitz GR, Rockman CB, Gagne PJ, et al. A model for predicting occult carotid artery stenosis: Screening is justified in a selected population. *J Vasc Surg* 2003; 38(4):705-9. doi: 10.1016/s0741-5214(03)00730-4.
12. Greco G, Egorova NN, Moskowitz AJ, et al. A model for predicting the risk of carotid artery disease. *Ann Surg* 2013; 257(6):1168-73. doi: 10.1097/SLA.0b013e31827b9761.
13. Breiman, Leo. Random forests. *Machine learning* 2001; 45(1):5-32. Available at <https://www.stat>.

- berkeley.edu/~breiman/randomforest2001.pdf. Accessed: December 1, 2019.
14. Wang X, Li W, Song F, et al. Carotid atherosclerosis detected by ultrasonography: A national cross-sectional study. *J Am Heart Assoc* 2018; 7(8):e008701. doi: 10.1161/JAHA.118.008701.
 15. Hua Y, Hui PJ, Xing YQ. Chinese stroke vascular ultrasound examination guidelines. *Chin J Med Ultrasound (Electronic Edition)* 2015; 12(8):599-610. Chinese. doi: 10.3877/cma.j.issn.1672-6448.2015.08.004.
 16. Touboul PJ, Hennerici MG, Meairs S, et al. Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the Advisory Board of the 3rd and 4th Watching the Risk Symposium, 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006. *Cerebrovasc Dis* 2007; 23(1):75-80. doi: 10.1159/000097034.
 17. Abbott AL. Medical (nonsurgical) intervention alone is now best for prevention of stroke associated with asymptomatic severe carotid stenosis: results of a systematic review and analysis. *Stroke* 2009; 40(10):e573-83. doi: 10.1161/STROKEAHA.109.556068.
 18. Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning. In: OSDI' 16 proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. USA: Usenix Association; 2016. p.265-83.
 19. Jiao Y, Chen R, Ke X, et al. Predictive models of autism spectrum disorder based on brain regional cortical thickness. *Neuroimage* 2010; 50(2):589-99. doi: 10.1016/j.neuroimage.2009.12.047.
 20. Qureshi AI, Janardhan V, Bennett SE, et al. Who should be screened for asymptomatic carotid artery stenosis? Experience from the Western New York Stroke Screening Program. *J Neuroimaging* 2001; 11(2):105-11. doi: 10.1111/j.1552-6569.2001.tb00019.x.
 21. Jauk S, Kramer D, Quehenberger F, et al. Information adapted machine learning models for prediction in clinical workflow. *Stud Health Technol Inform* 2019; 260:65-72. doi: 10.3233/978-1-61499-971-3-65.
 22. Weng SF, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017; 12(4):e0174944. doi: 10.1371/journal.pone.0174944.
 23. Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2014; 13:8-17. doi: 10.1016/j.csbj.2014.11.005.
 24. Hu X, Reaven PD, Saremi A, et al. Machine learning to predict rapid progression of carotid atherosclerosis in patients with impaired glucose tolerance. *EURASIP J Bioinform Syst Biol* 2016; 2016(1):14. doi: 10.1186/s13637-016-0049-6.
 25. Forsblad J, Gottsater A, Matzsch T, et al. Predictors of carotid endarterectomy in middle-aged individuals. *Vasc Med* 2001; 6(2):81-5. doi: 10.1177/1358836X0100600203.
 26. Zmysłowski A, Szterk A. Current knowledge on the mechanism of atherosclerosis and pro-atherosclerotic properties of oxysterols. *Lipids Health Dis* 2017; 16(1):188. doi: 10.1186/s12944-017-0579-2.
 27. Wiegman A, Gidding SS, Watts GF, et al. Familial hypercholesterolaemia in children and adolescents: gaining decades of life by optimizing detection and treatment. *Eur Heart J* 2015; 36(36):2425-37. doi: 10.1093/eurheartj/ehv157.
 28. Defesche JC, Gidding SS, Harada-Shiba M, et al. Familial hypercholesterolaemia. *Nat Rev Dis Primers* 2017; 3:17093. doi: 10.1038/nrdp.2017.93.
 29. Yin J, Yu C, Liu H, et al. A model to predict unstable carotid plaques in population with high risk of stroke. *BMC Cardiovasc Disord* 2020; 20(1):164. doi: 10.1186/s12872-020-01450-z.