EBioMedicine 69 (2021) 103466

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Classification of cardioembolic stroke based on a deep neural network using chest radiographs

Han-Gil Jeong[a,b,#], Beom Joon Kim[a,#], Tackeun Kim[b,*], Jihoon Kang[a], Jun Yup Kim[a], Joonghee Kim[c], Joon-Tae Kim[d], Jong-Moo Park[e], Jae Guk Kim[f], Jeong-Ho Hong[g], Kyung Bok Lee[h], Tai Hwan Park[i], Dae-Hyun Kim[j], Chang Wan Oh[b], Moon-Ku Han[a], Hee-Joon Bae[a]

[a] Department of Neurology, Seoul National University Bundang Hospital, Seongnam, Korea
[b] Department of Neurosurgery, Seoul National University Bundang Hospital, Seongnam, Korea
[c] Department of Emergency Medicine, Seoul National University Bundang Hospital, Seongnam, Korea
[d] Department of Neurology, Chonnam National University Hospital, Gwangju, Korea
[e] Department of Neurology, Nowon Eulji Medical Center, Eulji University, Seoul, Korea
[f] Department of Neurology, Eulji University Hospital, Daejeon, Korea
[g] Department of Neurology, Keimyung University Dongsan Medical Center, Daegu, Korea
[h] Department of Neurology, Soonchunhyang University Hospital, Seoul, Korea
[i] Department of Neurology, Seoul Medical Center, Korea
[j] Department of Neurology, Dong-A University Hospital, Busan, Korea

ARTICLE INFO

Article History:
Received 10 March 2021
Revised 11 June 2021
Accepted 11 June 2021
Available online xxx

Keywords:
Stroke
Chest radiograph
Deep learning
Cardioembolism
Classification

ABSTRACT

Background: Although chest radiographs have not been utilised well for classifying stroke subtypes, they could provide a plethora of information on cardioembolic stroke.
This study aimed to develop a deep convolutional neural network that could diagnose cardioembolic stroke based on chest radiographs.
Methods: Overall, 4,064 chest radiographs of consecutive patients with acute ischaemic stroke were collected from a prospectively maintained stroke registry. Chest radiographs were randomly partitioned into training/validation (n = 3,255) and internal test (n = 809) datasets in an 8:2 ratio. A densely connected convolutional network (ASTRO-X) was trained to diagnose cardioembolic stroke based on chest radiographs. The performance of ASTRO-X was evaluated using the area under the receiver operating characteristic curve. Gradient-weighted class activation mapping was used to evaluate the region of focus of ASTRO-X. External testing was performed with 750 chest radiographs of patients with acute ischaemic stroke from 7 hospitals.
Findings: The areas under the receiver operating characteristic curve of ASTRO-X were 0.86 (95% confidence interval [CI], 0.83–0.89) and 0.82 (95% CI, 0.79–0.85) during the internal and multicentre external testing, respectively. The gradient-weighted class activation map demonstrated that ASTRO-X was focused on the area where the left atrium was located. Compared with cases predicted as non-cardioembolism by ASTRO-X, cases predicted as cardioembolism by ASTRO-X had higher left atrial volume index and lower left ventricular ejection fraction in echocardiography.
Interpretation: ASTRO-X, a deep neural network developed to diagnose cardioembolic stroke based on chest radiographs, demonstrated good classification performance and biological plausibility.

1. Introduction

Stroke classification is important because many therapeutic decisions are dependent on its pathophysiology [1]. Exploration for a cardioembolic source particularly matters as most cases require anticoagulation therapy for secondary prevention [2]. The incidence of cardioembolic stroke has been increasing, and it causes more severe stroke than other subtypes [3]. Cardioembolic stroke is characterized by focal cerebral infarction due to cardiac problems and not due to the pathology of the cerebral vasculature [4]. To diagnose cardioembolic stroke, it is imperative to consider not only the clinical and radiological patterns of cerebral infarction, but also the underlying

* Corresponding author.
E-mail address: tackeun.kim@snu.ac.kr (T. Kim).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2021.103466

Research in context

Evidence before this study

Classification of cardioembolic stroke is important because its prevention requires anticoagulation therapy or intervention for cardiac disease. Diverse cardiac workups such as electrocardiogram, echocardiography, cardiac computed tomography and cardiac magnetic resonance imaging have been used to diagnose cardioembolic stroke. Although chest radiographs can provide a plethora of information about the cardiac structure, they have not been well utilized to classify stroke subtypes due ambiguity and limited inter-rater reliability.

Added value of this study

This is the first study to evaluate whether deep learning can distinguish cardioembolic stroke using chest radiographs in patients with acute ischaemic stroke. Our deep learning model demonstrated significant performance in classifying cardioembolic stroke from non-cardioembolic stroke on chest radiograph. Gradient-weighted class activation mapping analysis showed that classification of cardioembolic stroke could be made by attention to the heart, especially the area in which the left atrium is located. This analysis provides interpretability and biological plausibility for the results of the deep learning algorithm.

Implications of all the available evidence

These results suggest the possibility of using chest radiographs to diagnose stroke subtypes when deep learning approach is used. Because chest radiograph is relatively inexpensive and non-invasive, it can be easily used in resource-poor countries. Moreover, our deep learning model may complement human decision-making processes for stroke work-up and diagnosis of cardioembolic stroke in current clinical practice. However, to validate clinical usefulness of this model, it is necessary to evaluate what additional roles chest radiograph can play in established stroke work-up in future studies.

cardiac substrates. Diverse workups, including 12-lead electrocardiogram, extended electrocardiographic monitoring, echocardiography, cardiac computed tomography, and even cardiac magnetic resonance imaging, have been used to reveal underlying cardiac pathologies [5–7]. However, some of these tests are expensive, often take a long time, and sometimes performed irrespectively of the requirement of such tests in patients. As such, diagnosing cardioembolic stroke requires significant efforts and clinical skills, yet sometimes such diagnoses fail [8, 9].

Chest radiograph is the most common medical image taken in medicine, which can inexpensively and easily visualize diverse structures in the thoracic cavity [10]. Since changes in the configuration of the heart and great vessels are reflected in chest radiographs, they can provide useful information related to stroke etiology. However, stroke guidelines have changed and currently, such guidelines recommend that most patients do not necessarily need chest radiographs as a part of the initial evaluation [11]. A previous study demonstrated that the routine acquisition of chest radiographs impacted clinical management in only 3.8% of stroke patients [12]. The limitations of routine chest radiographs have been attributed to the difficulty in defining or quantifying ambiguous features related to stroke etiology based on radiographic analysis. In addition, many acute stroke patients undergo chest radiography in the anteroposterior projection in the emergency room, which makes it difficult for the human eye to accurately measure the dimension of the heart.

Using deep learning, however, it may be feasible to create an algorithm for diagnosing cardioembolic stroke based on chest radiographs. The classification of cardioembolic stroke from unclassified chest radiographs is challenging and has not been previously investigated. In the field of artificial intelligence, chest radiograph research has so far focused on how to mimic the reading performance of radiology experts [13–16]. However, a deep neural network is powerful in finding features that are difficult for the human eye to extract or classify [17]. Therefore, training strategies using labels of stroke classification could produce a much more elaborate deep learning algorithm than conventional labels of radiological readings. If cardioembolic stroke can be identified based on the initial chest radiographs on presentation, it will not only save time and reduce diagnostic costs, but also help in patient care. This study aimed to develop a classification algorithm for cardioembolic strokes using more than 4000 chest radiographs from an ischemic stroke registry with prospectively coded etiological information [18].

2. Methods

2.1. Study design

This study was a retrospective analysis of prospectively collected data from a clinical registry. It is an exploratory study to develop a deep learning model which classifies cardioembolic stroke from non-cardioembolic stroke using chest radiograph.

2.2. Data

Among 5430 consecutive patients with ischemic stroke who were admitted to the Seoul National University Bundang Hospital between January 2014 and March 2019, we identified 5358 patients who had chest radiographs as part of the initial evaluation. The first chest radiographs taken after admission were used regardless of projection (anteroposterior or posteroanterior) and without specific selection criteria. We excluded cases with undetermined (n = 866) or other determined (n = 428) aetiologies in the modified TOAST (Trial of ORG 10,172 in Acute Stroke Treatment) classification system [19]. Finally, 4064 patients were included in this analysis. External testing was performed in 750 chest radiographs of consecutive patients with acute ischemic stroke containing etiological information from 7 hospitals (Chonnam National University Hospital, Gwangju [University hospital]; Nowon Eulji Medical Center, Seoul [University hospital]; Eulji University Hospital, Daejeon [University hospital]; Keimyung University Dongsan Medical Center, Daegu [University hospital]; Soonchunhyang University Hospital, Seoul [University hospital]; Seoul Medical Center, Seoul [Public hospital]; Dong-A University Hospital, Busan [University hospital]) that participated in the Clinical Research Center for Stroke-Fifth Division Registry in South Korea [18]. Previously known atrial fibrillation was defined when the patient had a past medical history or was diagnosed with an initial electrocardiogram in the emergency department. Newly diagnosed atrial fibrillation was defined if it was diagnosed during hospitalization for acute ischemic stroke. This study was approved by the institutional review boards of Seoul National University Bundang Hospital (B-2004–604–118) and 7 hospitals participating in external testing. Informed consent was waived by the institutional review boards. The data are not available for public access because of patient privacy concerns but are available from the corresponding author on reasonable request approved by the institutional review boards of Seoul National University Bundang Hospital.

2.3. Ground truth

Using modified TOAST classification system which has gained wide acceptance in clinical practice and research, cardioembolic

stroke and non-cardioembolic stroke (large artery atherosclerosis or small vessel occlusion) were defined. In our stroke registry, modified TOAST classification has been prospectively performed using MRI-based algorithm consisting of multiple steps dealing with clinical information, comprehensive work-up and vessel status (Supplemental Fig. 1). The details and the intra-class coefficients of modified TOAST classification in our stroke registry were previously reported [19].

2.4. Data partition and pre-processing

The sample size was determined a priori without power calculation. As the first chest radiograph is used, the number of images and patients was same. The eligible patients were partitioned into training/validation (n = 3255) and test (n = 809) datasets at an 8:2 ratio using permutation. The training/validation datasets was further partitioned into training (n = 2605) and validation (n = 650) dataset (Fig. 1). All chest radiographs, including external test dataset, were cropped to square dimension and resized to 224 × 224 pixels. We processed images in all datasets for histogram equalization to adjust sample-wise contrast, remove confounding histogram differences between image classes, and locally improve the contrast of the images [20]. Only images in the training dataset underwent a data augmentation procedure that permitted a rotation within 10°, horizontal and vertical shifts within 10%, and zoom range between 0.95 and 1.05. To adjust class imbalance, chest radiographs of cardioembolic and non-cardioembolic stroke patients were augmented 6 and 3 folds, respectively. Collectively, a total of 10,116 augmented, 650 non-augmented, 809 non-augmented, and 750 non-augmented images were used for training, validation, internal and external testing, respectively.

2.5. Modeling and training

ASTRO-X (Acute STROke classification by chest X-ray) is designed as a classifier based on previously reported 121-layer Densely Connected Convolutional Network (DenseNet-121) [21] and trained on etiologically classified chest radiographs of patients with acute ischemic stroke. We used the TensorFlow platform (ver.2.1, https://www.tensorflow.org) as the deep-learning library to implement the software to train, validate, and test the convolutional neural network (CNN). The weights of the network were initialized with weights from a model pretrained on the chest X-ray 14 datasets, which contained 112,120 chest radiographs labelled with 14 different diagnoses including pneumonia [13,14,22]. We replaced the final fully connected layer with another layer that has a single output with a sigmoid function. The binary cross-entropy function was adopted as a loss function, and Adam was used as an optimizer function (β1=0.9, and β2=0.999) following a method proposed in pre-print [23]. Hyperparameters were determined by grid search method with combination of a learning rate of (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7}) and a batch size of (8, 16, 32, 64 and 128). The model and codes used in model training are available online (https://github.com/han-gil/astro-x/)

2.6. Evaluation

The performance of the model was evaluated using the institutional and external test datasets. A confusion matrix was created using 0.5 as a cut-off of sigmoid output, where positive and negative instances represented cardioembolic and non-cardioembolic strokes, respectively. The accuracy, area under the receiver operating characteristic (AUROC) curve, sensitivity, specificity, and positive and negative predictive values were calculated. Confidence intervals (CIs) for each value were calculated using the exact binomial confidence limits [24]. Then, we compared the performance of ASTRO-X with that of multivariable logistic regression models using clinical variables including previously known atrial fibrillation selected by backward elimination. The performance of models trained with the ImageNet initial weight, without histogram equalization or without image augmentation were compared, respectively. As a sensitivity analysis, models were developed and tested through 5-fold cross validation.

We applied the gradient-weighted class activation mapping (Grad-CAM) to produce visual representations of our model [25]. Using the gradient of weights for cardioembolic stroke flowing into the final convolutional layer, a localization map highlighting important regions for predicting cardioembolic stroke can be created in the original image. We also made Grad-CAM of cardiomegaly using

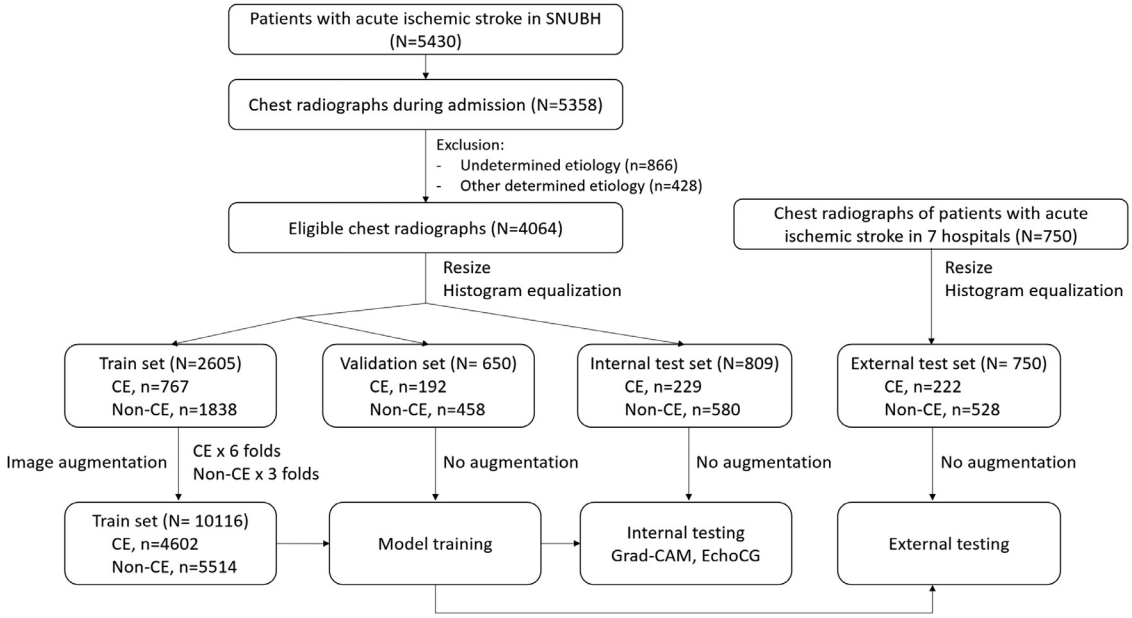


Fig. 1. Study flowchart. SNUBH, Seoul National University Bundang Hospital; CE, Cardioembolism; Grad-CAM, Gradient-weighted Class Activation Mapping; EchoCG, Echocardiography.

weights from CheXNet which was the initial weights of our model. Then, we compared the attention map of ASTRO-X and CheXNet using overlays of activation maps (Cardioembolism by ASTRO-X vs. Cardiomegaly by CheXNet.)

To enhance the interpretability of ASTRO-X classifier, we further analysed the findings of transthoracic echocardiography from the internal test dataset. A total of 650 patients in the internal test dataset who underwent transthoracic echocardiography between 1 week before and 1 month after stroke were analysed. Specific parameters such as the left atrial and ventricular size, systolic and diastolic functions, and valvular dysfunction were compared based on the classes predicted by ASTRO-X.

2.7. Statistics

Continuous variables are presented as mean ± standard deviation or median [interquartile range] were analysed by Student's *t*-test or analysis of variance as appropriate. Categorical variables are presented as number (percent) and were analysed by Pearson's chi-squared test. AUROC was calculated with true labels and sigmoid output values, and its 95% confidence interval was computed with bootstrapping. The 95% confidence intervals of sensitivity, specificity and accuracy were calculated by the exact binomial test. For comparison, a multivariable logistic regression model using clinical information obtained prior to hospitalization, including previously known atrial fibrillation, was constructed with backward elimination. A simple ensemble model was created using the mean of the sigmoid outputs of the logistic regression model and ASTRO-X. All models were tested on both internal and external test sets and the AUROC of each model was compared using the DeLong's test [26]. All statistical tests were performed using R statistical software (version 4.0.4).

2.8. Role of funding source

The funders had no role in study design, data collection, data analyses, interpretation, and writing of report.

3. Results

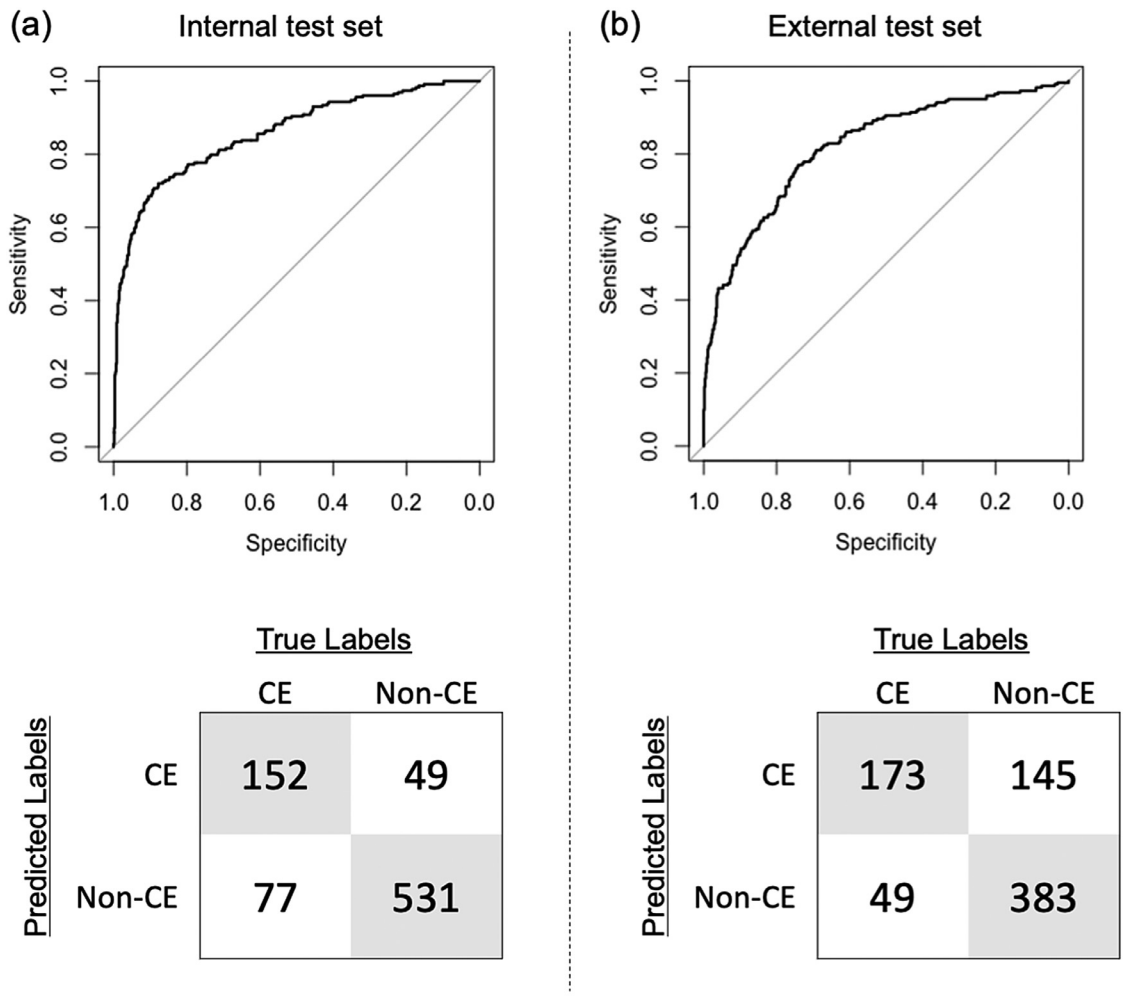
Of the 4064 cases with acute ischemic stroke, 61% (n = 2478) of the sample were men, while the mean age was 68.7 ± 12.6 years. In these patients, risk factors included hypertension [2868 (70.6%)], diabetes mellitus [1367 (33.6%)], and atrial fibrillation [919 (22.6%)]. The median National Institutes of Health Stroke Scale score at admission was 3 [interquartile range (IQR) 1–7] points, and the median time delay from symptom onset to arrival was 16.5 (IQR 4.5–57.4) hours. Endovascular treatment and intravenous thrombolysis were performed in 473 (11.6%) and 386 (9.5%) patients, respectively. The modified Rankin Scale of 0–2 at 3 months was achieved in 2731 (67.2%) patients, and the mortality rate at 3 months was 4.1% (Supplemental Table 1). While the baseline characteristics were comparable between the training/validation and internal test datasets, those were different between the internal and external test datasets regarding demographics, hyperacute treatment and clinical outcomes (Table 1).

Validation loss achieved the lowest value at the 23rd epoch of training process with a learning rate of 10^{-5} and a batch size of 32 with an accuracy of 87.4% and 81.5% in the training and validation sets, respectively. In the internal test set, the accuracy was 84.4% (95% confidence interval [CI], 81.7%–86.9%), sensitivity was 0.66 (95% CI, 0.60–0.72), specificity was 0.92 (95% CI, 0.89–0.94), and AUROC was 0.86 (95% CI, 0.83–0.89). The positive and negative predictive values were 0.76 (95% CI, 0.69–0.81) and 0.87 (95% CI, 0.84–0.90), respectively (Fig. 2a and c).

Table 1
Baseline characteristics of the training/validation, internal and external test datasets.

	Training /validation (n = 3255)	Internal test (n = 809)	External test (n = 750)	P-value*	P-value†
Demographic information					
Male sex	1988 (61.1%)	490 (60.6%)	381 (50.8%)	0.82	<0.01
Age, years	68.7 ± 12.6	68.6 ± 12.6	70.4 ± 12.6	0.75	0.01
Premorbid mRS score, 0–1	2914 (89.5%)	727 (89.9%)	644 (85.9%)	0.83	0.02
Stroke information					
Onset to arrival, hours	16.4 [4.2–57.6]	16.5 [4.5–57.4]	13.3 [3.4–33.6]	0.36	<0.01
NIHSS score at arrival	3 [1–7]	3 [1–7]	3 [1–7]	0.71	0.10
Systolic BP, mmHg	155.5 ± 37.3	154.1 ± 26.3	149.6 ± 27.5	0.21	<0.01
Diastolic BP, mmHg	84.4 ± 32.6	83.0 ± 16.2	84.3 ± 16.2	0.08	0.13
Hyperacute treatment					
IV thrombolysis	319 (9.8%)	67 (8.3%)	92 (12.3%)	0.21	0.01
Endovascular therapy	379 (11.6%)	94 (11.6%)	55 (7.3%)	1.00	0.01
Risk factors					
Hypertension	2304 (70.8%)	564 (69.7%)	496 (66.1%)	0.58	0.14
Diabetes	1103 (33.9%)	264 (32.6%)	269 (35.9%)	0.53	0.20
Dyslipidaemia	1199 (36.8%)	277 (34.2%)	201 (26.8%)	0.18	<0.01
Current smoker	738 (22.7%)	176 (21.8%)	145 (19.3%)	0.61	0.26
Atrial fibrillation	737 (22.6%)	182 (22.5%)	195 (26.0%)	0.97	0.12
Previously known	420 (12.9%)	99 (12.2%)	105 (14.0%)	0.65	0.34
Newly diagnosed	317 (9.7%)	83 (10.3%)	90 (12.0%)	0.71	0.31
Laboratory information					
Haemoglobin, g/dL	13.7 ± 2.0	13.7 ± 2.1	13.4 ± 2.1	0.92	0.01
Leukocyte count, 10^9	8104 ± 3051	8070 ± 2911	8084 ± 3135	0.78	0.93
Glucose, mg/dL	139.1 ± 58.8	139.4 ± 58.3	146.3 ± 65.1	0.88	0.03
HbA1c,%	6.3 ± 1.3	6.3 ± 1.2	6.3 ± 1.4	0.66	0.19
Total cholesterol, mg/dL	168.1 ± 41.6	166.9 ± 38.9	169.8 ± 44.0	0.43	0.17
LDL cholesterol, mg/dL	99.2 ± 32.2	98.7 ± 31.2	103.9 ± 56.3	0.70	0.03
Outcomes					
mRS 0–2 at 3 months	2177 (66.9%)	554 (68.5%)	447 (59.6%)	0.41	<0.01
Mortality at 3 months	126 (3.9%)	41 (5.1%)	43 (5.7%)	0.15	0.64

* Comparison between training/validation and internal test set. † Comparison between internal test and external test set. mRS, modified Rankin Scale; NIHSS, National Institute of Health Stroke Scale; BP, blood pressure; IV, intravenous; LDL, low-density lipoprotein.



(c)

Metric	Internal test set	External test set
AUROC	0.86 (0.83 – 0.89)	0.82 (0.79 – 0.85)
Sensitivity	0.66 (0.60 – 0.72)	0.78 (0.72 – 0.83)
Specificity	0.92 (0.89 – 0.94)	0.73 (0.69 – 0.76)
Accuracy	0.84 (0.82 – 0.87)	0.74 (0.71 – 0.77)
Weighted accuracy	0.79 (0.76 – 0.82)	0.75 (0.72 – 0.78)
F1 score	0.707	0.641

Fig. 2. Performance of ASTRO-X in the classification of cardioembolic strokes based on chest radiographs. The probability cut-off to predict cardioembolic stroke was 0.5. AUC, area under the curve; CE, cardioembolism.

The external testing from 7 hospitals demonstrated that the accuracy was 74.1% (95% CI, 70.8–77.2%), sensitivity was 0.78 (95% CI, 0.72–0.83), specificity was 0.73 (95% CI, 0.69–0.76), and AUROC was 0.82 (95% CI, 0.79–0.85) (Fig. 2a and c). ASTRO-X, at high sensitivity or high specificity, demonstrated similar performances as a screening tool for both internal and external test sets (Supplemental Tables 2 and 3).

Grad-CAM showed that the prediction of cardioembolic stroke was primarily done with a focus on the upper middle part of the heart in chest radiographs, where the left atrium was usually located posteriorly (Fig. 3). Echocardiography was performed in 80.3% (n = 650) of the test dataset (Supplemental Table 4). Echocardiographic findings demonstrated that cases predicted as cardioembolic stroke by ASTRO-X had a lower ejection fraction, higher E/e', left atrial and

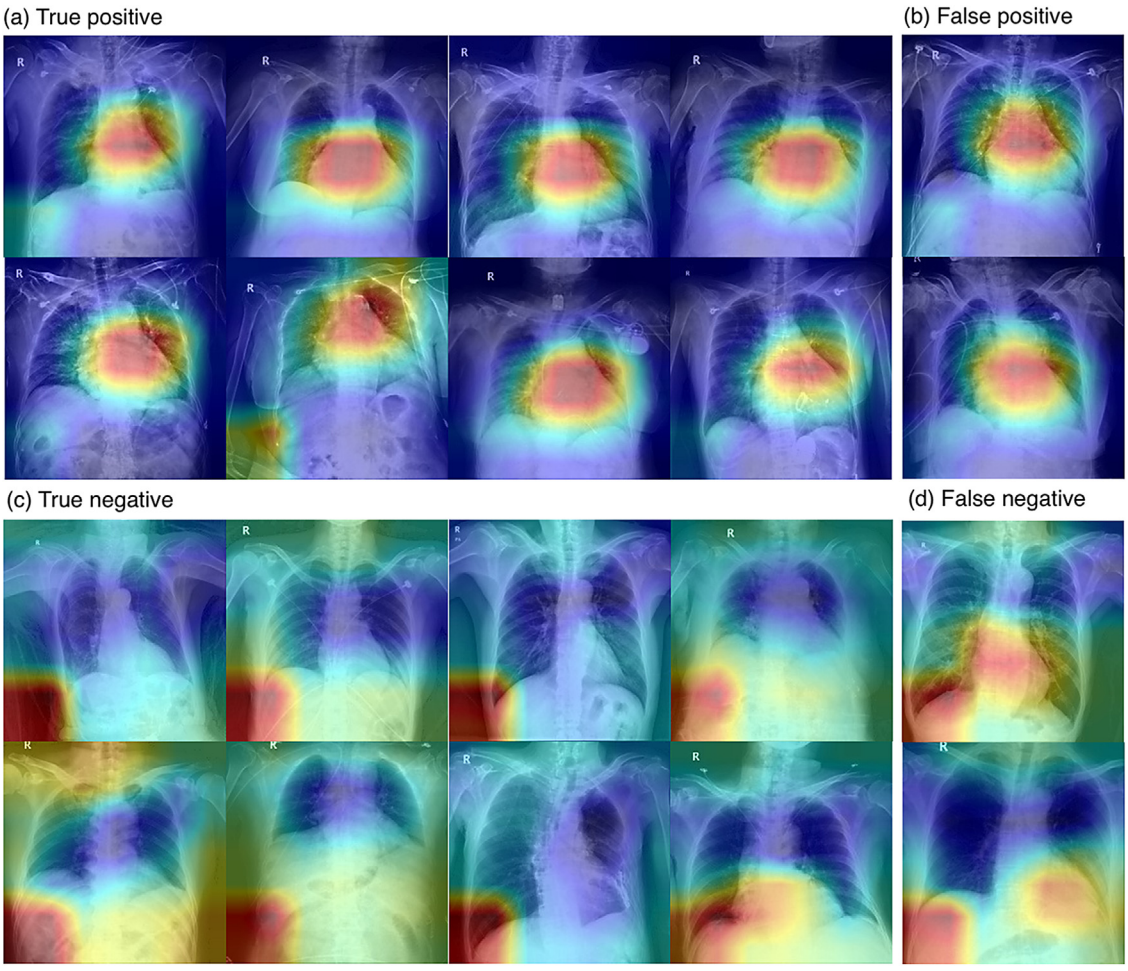


Fig. 3. Representative gradient-weighted class activation mapping results according to predictions by ASTRO-X.

ventricular size indices, and were more likely to have moderate to severe mitral stenosis or mitral/tricuspid regurgitation compared to those cases predicted as non-cardioembolic stroke by ASTRO-X (Table 2 and Supplemental Table 5).

ASTRO-X was better in the diagnosis of cardioembolic strokes in high-risk sources than in medium-risk sources (P-value < 0.01 [chi-squared test]). The majority of high-risk sources was atrial fibrillation, 75.3% of which was correctly classified as cardioembolic stroke. The majority of medium-risk sources was patent foramen ovale, approximately 91% of which was classified as non-cardioembolic stroke by ASTRO-X (Table 3).

Table 2
Comparison of echocardiographic findings according to predictions by ASTRO-X.

Variables	Total (n = 650)	Predicted CE (n = 163)	Predicted Non-CE (n = 487)	Ratio*	P-value
Ejection fraction,%	61.1 ± 8.7	56.8 ± 11.9	62.5 ± 6.9	0.91	<0.01
E/e'	12.3 ± 6.2	15.9 ± 9.0	11.1 ± 4.2	1.44	<0.01
LA AP diameter, mm	38.2 ± 7.1	43.6 ± 8.0	36.4 ± 5.8	1.20	<0.01
LA volume, ml	70.7 ± 33.0	102.9 ± 41.6	59.8 ± 20.1	1.72	<0.01
LA volume index, ml/m^2	42.2 ± 20.3	62.4 ± 25.7	35.3 ± 12.0	1.77	<0.01
LV end-diastolic diameter, mm	45.5 ± 5.6	47.2 ± 6.5	44.9 ± 5.1	1.05	<0.01
LV end-diastolic volume, ml	74.9 ± 25.0	78.0 ± 33.9	73.9 ± 21.2	1.06	0.14
LV end-systolic diameter, mm	29.3 ± 6.1	32.0 ± 7.6	28.4 ± 5.2	1.13	<0.01
LV end-systolic volume, ml	30.1 ± 17.5	35.6 ± 26.1	28.2 ± 13.0	1.26	<0.01
LV mass, g	166.4 ± 49.6	181.1 ± 58.9	161.5 ± 45.1	1.12	<0.01
LV mass index, g/m^2	98.5 ± 27.2	109.1 ± 32.8	95.0 ± 24.1	1.15	<0.01
Any wall motion abnormality	64 (9.9%)	31 (19.1%)	33 (6.8%)	2.81	<0.01
Aortic regurgitation**	13 (2.0%)	6 (3.7%)	7 (1.4%)	2.56	0.15
Aortic stenosis**	7 (1.1%)	4 (2.5%)	3 (0.6%)	3.98	0.13
Mitral regurgitation**	5 (0.8%)	4 (2.5%)	1 (0.2%)	12.00	0.02
Mitral stenosis**	3 (0.5%)	3 (1.8%)	0 (0%)	N/A	0.02
Tricuspid regurgitation**	21 (3.2%)	15 (9.2%)	6 (1.2%)	7.47	<0.01

E/e', ratio of early mitral inflow velocity and mitral annular early diastolic velocity; LV, left ventricle. *CE to Non-CE ratio; calculated as the mean and percentage for continuous and categorical variables, respectively. ** Moderate to severe degree. The missing values are less than 1% except LA volume and LA volume index (n = 37, 5.7%).

Table 3
Comparison of the presumptive cause of cardioembolic stroke according to predictions by ASTRO-X.

	Cardioembolic stroke (n = 209)		
	Predicted CE (n = 152)	Predicted non-CE (n = 77)	P-value
Risk of cardioembolism			<0.01
High risk sources	147 (73.5%)	53 (26.5%)	
Medium risk sources	5 (17.2%)	24 (82.8%)	
High risk sources			
Atrial fibrillation/flutter	137 (75.3%)	45 (24.7%)	<0.01
Left ventricular thrombus	1 (33.3%)	2 (66.7%)	0.22
Mechanical prosthetic valve	1 (25%)	3 (75%)	0.08
Atrial myxoma	1 (50%)	1 (50%)	0.62
Dilated cardiomyopathy	1 (50%)	1 (50%)	0.62
Infective endocarditis	2 (100%)	0 (0%)	0.31
Recent myocardial infarct	1 (50%)	1 (50%)	0.62
Akinetic left ventricular segments	1 (100%)	0 (0%)	0.48
Sick sinus syndrome	1 (100%)	0 (0%)	0.48
Other cause*	1 (100%)	0 (0%)	0.48
Medium risk sources			
Patent foramen ovale	2 (9.1%)	20 (90.9%)	<0.01
Left atrial turbulence (smoke)	1 (25%)	3 (75%)	0.08
Hypokinetic left ventricular segment	1 (50%)	1 (50%)	0.62
Congestive heart failure	1 (100%)	0 (0%)	0.48

All percentages were calculated row-wise. CE, cardioembolic stroke. *A case with multiple embolic infarcts and severe aortic stenosis with left atrial enlargement was categorized as a high-risk source based on the attending physician's clinical reasoning.

The AUROC of ASTRO-X for classifying cardioembolic stroke was comparable to the AUROC of the multivariable logistic regression model using various clinical information, including previously known atrial fibrillation, in both internal and external test sets (Supplemental Figure 2 and 3). The AUROCs of the ensemble of the logistic regression model and ASTRO-X were 0.89 and 0.90 in the internal and external test, respectively, which were significantly higher than the AUROC of both models (Supplemental Figure 4 and Table 6). The performance of models trained with ImageNet initial weight, without histogram equalization or without image augmentation were lower than ASTRO-X (Supplemental Table 7). The mean AUROC from 5-fold cross validation was 0.81 (Supplemental Table 8).

4. Discussion

We developed a deep neural network that can successfully classify cardioembolic strokes based on chest radiographs. Our network showed good predictive performance with high accuracy and AUROC. Our results could be generalized to patients with acute ischemic stroke from 7 academic hospitals in Korea.

It is important to diagnose cardioembolic stroke since the secondary prevention strategy is different for other stroke subtypes such as large artery atherosclerosis and small vessel occlusion [3]. Clinical evaluation, neuroimaging findings, and cardiac evaluation such as electrocardiogram, echocardiography, or prolonged cardiac rhythm monitoring have been comprehensively used to diagnose cardioembolic stroke (5). Chest radiographs are frequently taken during hospitalization of stroke patients (98.7% in our cohort), but have not been actively used to classify stroke aetiologies, because it is not only challenging to define features related to cardioembolic stroke but also difficult for the human eye to consistently evaluate such features. However, ASTRO-X can distinguish cardioembolic strokes based on chest radiographs with high discriminative power, because deep neural networks, especially convolutional neural networks, can automatically extract various features related to the classification process [27].

From the Grad-CAM results, it is evident that the main focus of ASTRO-X, which learned through the clinical diagnosis of cardioembolic strokes, was the left atrium of the heart. In contrast, CheXNet, which learned using radiologists' cardiomegaly readings, focused on the whole contour of the heart (Supplemental Figure 5 and 6). This comparison suggests that the CheXNet was converted into ASTRO-X through transfer learning, which can classify cardioembolic stroke by evaluating specific characteristics of the heart associated with stroke. The left atrium, especially its appendage, is the most common site of cardiac thrombosis and closely related to atrial fibrillation and atrial flutter [28]. Many studies recently suggested that atrial cardiopathy could be a cause of cardioembolic stroke even in the absence of atrial fibrillation [29]. Therefore, ASTRO-X's primary focus on the left atrium on chest radiographs would be the most effective strategy for a convolutional neural network to distinguish cardioembolic strokes.

Since chest radiographs are a 2-dimensional representation of 3-dimensional objects, echocardiographic results were further analysed to assess the biological plausibility of ASTRO-X's predictions. Among the echocardiographic measures that were significantly higher in cases predicted as cardioembolism by ASTRO-X, the left atrial volume index, known to have a high correlation with cardioembolic stroke and atrial fibrillation, is more likely to be detected in patients with a higher left atrial volume index [30]. Left atrial enlargement is also associated with spontaneous echo contrast and embolic events regardless of atrial fibrillation [31,32]. Interestingly, moderate to severe tricuspid regurgitation was found to be 7.5 times higher in cases predicted as cardioembolic stroke than in those cases predicted as non-cardioembolic stroke. This may be secondary to the left-sided heart disease, but recent studies have shown that atrial fibrillation could cause isolated tricuspid regurgitation through tricuspid annular dilatation without right ventricular remodeling in elderly patients [33,34].

Another interesting finding is that the activation maps were similar between true and false positives, as well as true and false negatives. The TOAST classification is the most widely used etiological classification of ischemic stroke, where cardioembolism encompasses both high-risk (i.e., mechanical prosthetic valve) and medium-risk (i.e., patent foramen ovale [PFO]) sources [35]. However, the weak ground truth (TOAST classification) with modest inter-rater reliability made it fundamentally impossible to train ASTRO-X to work perfectly [9]. Some false positive cases would have been true positives, if more extensive stroke work-ups such as long-term continuous ambulatory electrocardiographic monitoring had been performed [6,36]. In the interpretation of activations maps, the heterogeneity of cardioembolic stroke regarding cardiac morphology should also be considered [3]. PFO is distinct among these since the paradoxical emboli, passing through a PFO, is not literally of cardiac origin [37]. Thus, the observation that ASTRO-X classified 91% of PFO-related strokes as non-cardioembolism (false negatives) is considered reasonable and may support the proposal to move PFO-related strokes into the other determined category [37].

Despite the good performance of ASTRO-X, it could not replace definite measures to diagnose atrial fibrillation or evaluate other cardiac pathology associated with stroke. However, ASTRO-X may help in classifying cardioembolic stroke through reducing human errors or guiding more thorough work-up for cardiac problems based on the probability. The performance of ASTRO-X was therefore evaluated by benchmarking the model using information available prior to hospitalization including previously known atrial fibrillation (past medical history or diagnosis at emergency room). The results showed that the AUROC of ASTRO-X was comparable to the complex multivariable model including previously known atrial fibrillation, and the ensemble models showed better performance with AUROC up to 0.90,

suggesting that its potential utility for etiologic evaluation during hospitalization.

This study has a few limitations. First, although the generalizability of our data was confirmed through multicentre external testing in Korea, additional testing, particularly in multi-ethnic populations, are warranted. As patients with other-determined and undetermined aetiologies were excluded in our analysis, ASTRO-X's performance would be difficult to measure prospectively in a population of all comers with ischemic stroke. Second, differences in the performance metric between the internal and external test sets could be attributable to differences in the clinical profiles, mode of radiograph acquisition or slight overfitting of ASTRO-X to the internal test set. Thus, further recalibration should be performed for the model to be used in clinical practice due to the sensitivity-specificity trade-off shifts between internal and external tests. The TOAST classification, used as the ground truth in training the network, is a clinical diagnosis with inter-rater reliability issues and without a gold standard [8,9]. Consequently, ASTRO-X could only complement and/or correct human decision-making processes. Third, ASTRO-X is required to be compared to human performance to detect left atrial enlargement on chest radiographs, though it is generally not evaluated in routine reading process. Finally, although a lot of effort has been made to explain the classification algorithm, there is still an inherent interpretability issue associated with a deep convolutional neural network. [38]

Conclusion

Among patients with acute ischemic stroke, ASTRO-X could differentiate between cardioembolic versus non-cardioembolic stroke. ASTRO-X demonstrated good classification feasibility and biological plausibility. Chest radiographs have not been utilized well in the classification of stroke subtypes; however, in the future, ASTRO-X can help in the detection of cardioembolic stroke based on chest radiographs.

Contributors

H-G.J, B.J.K., and T.K. conceived the study and wrote the manuscript. H-G.J and T.K. performed training of convolutional neural network. J.K., J.Y.K., J-T.K., J-M.P., J.G.K., J-H.H., K.B.L., T.H.P., D.H.K, C.W.O., M-K.H., H-J.B. prepared patient samples and radiologic images, and critically revised drafted manuscript. H-G.J, T.K and H-J.B. had verified and full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final version of the manuscript.

Data sharing statement

The data are not available for public access because of patient privacy concerns but are available from the corresponding author on reasonable request approved by the institutional review boards of Seoul National University Bundang Hospital. The model and codes used in model training are available online (https://github.com/han-gil/astro-x/)

Declaration of Competing Interest

Seoul National University Hospital (inventors: H-G. J and T. K) has a pending patent (KR10-2020-0125100).

Funding

Grant No. 14−2020−046 and 08−2016−051 from the Seoul National University Bundang Research Fund and NRF-2020M3E5D9079768 from the National Research Foundation of Korea

Acknowledgements

This work was supported by grant No. 14-2020-046 and No. 08-2016-051 from the Seoul National University Bundang Research Fund and by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (Grant No. NRF-2020M3E5D9079768).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103466.

References

[1] Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. Classification of stroke subtypes. Cerebrovasc Dis 2009;27(5):493–501.
[2] Ferro JM. Cardioembolic stroke: an update. Lancet Neurol 2003;2(3):177–88.
[3] Kamel H, Healey JS. Cardioembolic stroke. Circ Res 2017;120(3):514–26.
[4] Arboix A, Alió J. Cardioembolic stroke: clinical features, specific cardiac disorders and prognosis. Curr Cardiol Rev 2010;6(3):150–61.
[5] Ustrell X, Pellisé A. Cardiac workup of ischemic stroke. Curr Cardiol Rev 2010;6(3):175–83.
[6] Sanna T, Diener H-C, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, et al. Cryptogenic stroke and underlying atrial fibrillation. N Engl J Med 2014;370(26):2478–86.
[7] Baher A, Mowla A, Kodali S, Polsani VR, Nabi F, Nagueh SF, et al. Cardiac MRI improves identification of etiology of acute ischemic stroke. Cerebrovasc Dis 2014;37(4):277–84.
[8] Marnane M, Duggan CA, Sheehan OC, Merwick A, Hannon N, Curtin D, et al. Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, ASCO, and causative classification system: direct comparison in the North Dublin population stroke study. Stroke 2010;41(8):1579–86.
[9] Gordon D, Bendixen B, Adams H, Clarke W, Kappelle L, Woolson R, et al. Interphysician agreement in the diagnosis of subtypes of acute ischemic stroke: implications for clinical trials. Neurology 1993;43(5):1021.
[10] Radiation UNSCotEoA. Sources and effects of ionizing radiation: sources. United Nations Publications; 2000.
[11] Adams Jr HP, Del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: the American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2007;38(5):1655–711.
[12] Sagar G, Riley P, Vohrah A. Is admission chest radiography of any clinical value in acute stroke patients? Clin Radiol 1996;51(7):499–502.
[13] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:171105225. 2017.
[14] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit; 2017. p. 3462–71.
[15] Bar Y, Diamant I, Wolf L, Greenspan H. Deep learning with non-medical training used for chest pathology identification. In: Proc SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis.; 2015. p. 94140V.
[16] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172(5):1122–31 e9.
[17] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018;2018:1–13.
[18] Kim BJ, Park J-M, Kang K, Lee SJ, Ko Y, Kim JG, et al. Case characteristics, hyperacute treatment, and outcome information from the clinical research center for stroke-fifth division registry in South Korea. J Stroke 2015;17(1):38–53.
[19] Ko Y, Lee S, Chung J-W, Han M-K, Park J-M, Kang K, et al. MRI-based algorithm for acute ischemic stroke subtype classification. J Stroke 2014;16(3):161.
[20] Behzadi-khormouji H, Rostami H, Salehi S, Derakhshande-Rishehri T, Masoumi M, Salemi S, et al. Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. Comput Methods Programs Biomed 2020;185:105162.
[21] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.; 2017. p. 2261–9.
[22] Chou B. CheXNet-Keras 2018. Available from: https://github.com/bruce-chou1983/CheXNet-Keras.
[23] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

[24] Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26(4):404–13.
[25] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis. 2017:618–26.
[26] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988:837–45.
[27] Shaheen F, Verma B, Asafuddoula M. Impact of automatic feature extraction in deep learning architecture. DICTA 2016:1–8.
[28] Di Biase L, Natale A, Romero J. Thrombogenic and Arrhythmogenic Roles of the Left Atrial Appendage in Atrial Fibrillation: clinical Implications. Circulation 2018;138(18):2036–50.
[29] Kamel H, Okin PM, Elkind MS, Iadecola C. Atrial fibrillation and mechanisms of stroke: time for a new model. Stroke 2016;47(3):895–900.
[30] Psaty BM, Manolio TA, Kuller LH, Kronmal RA, Cushman M, Fried LP, et al. Incidence of and risk factors for atrial fibrillation in older adults. Circulation 1997;96(7):2455–61.
[31] Benjamin EJ, D'Agostino RB, Belanger AJ, Wolf PA, Levy D. Left atrial size and the risk of stroke and death: the Framingham Heart Study. Circulation 1995;92(4):835–41.
[32] Yaghi S, Kamel H, Elkind MS. Atrial cardiopathy: a mechanism of cryptogenic stroke. Expert Rev Cardiovasc Ther 2017;15(8):591–9.
[33] Topilsky Y, Khanna A, Le Tourneau T, Park S, Michelena H, Suri R, et al. Clinical context and mechanism of functional tricuspid regurgitation in patients with and without pulmonary hypertension. Circ Cardiovasc Imaging 2012;5(3):314–23.
[34] Mutlak D, Lessick J, Reisner SA, Aronson D, Dabbah S, Agmon Y. Echocardiography-based spectrum of severe tricuspid regurgitation: the frequency of apparently idiopathic tricuspid regurgitation. J Am Soc Echocardiogr 2007;20(4):405–8.
[35] Adams Jr HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24(1):35–41.
[36] Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, et al. Atrial fibrillation in patients with cryptogenic stroke. N Engl J Med 2014;370:2467–77.
[37] Elgendy AY, Saver JL, Amin Z, Boudoulas KD, Carroll JD, Elgendy IY, et al. Proposal for Updated Nomenclature and Classification of Potential Causative Mechanism in Patent Foramen Ovale–Associated Stroke. JAMA Neurol 2020.
[38] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206–15.