Stroke

CLINICAL AND POPULATION SCIENCES

Deep Learning Algorithm Enables Cerebral Venous Thrombosis Detection With Routine Brain Magnetic Resonance Imaging

Xiaoxu Yang, MD*; Pengxin Yu, MSc*; Haoyue Zhang, MSc; Rongguo Zhang, PhD; Yuehong Liu, MD; Haoyuan Li, MD; Penghui Sun, MSc; Xin Liu, MD; Yu Wu, MSc; Xiuqin Jia, PhD; Jiangang Duan, MD, PhD; Xunming Ji, MD, PhD; Qi Yang, MD, PhD

BACKGROUND: Cerebral venous thrombosis (CVT) is a rare cerebrovascular disease. Routine brain magnetic resonance imaging is commonly used to diagnose CVT. This study aimed to develop and evaluate a novel deep learning (DL) algorithm for detecting CVT using routine brain magnetic resonance imaging.

METHODS: Routine brain magnetic resonance imaging, including T1-weighted, T2-weighted, and fluid-attenuated inversion recovery images of patients suspected of CVT from April 2014 through December 2019 who were enrolled from a CVT registry, were collected. The images were divided into 2 data sets: a development set and a test set. Different DL algorithms were constructed in the development set using 5-fold cross-validation. Four radiologists with various levels of expertise independently read the images and performed diagnosis within the test set. The diagnostic performance on per-patient and per-segment diagnosis levels of the DL algorithms and radiologist's assessment were evaluated and compared.

RESULTS: A total of 392 patients, including 294 patients with CVT (37±14 years, 151 women) and 98 patients without CVT (42±15 years, 65 women), were enrolled. Of these, 100 patients (50 CVT and 50 non-CVT) were randomly assigned to the test set, and the other 292 patients comprised the development set. In the test set, the optimal DL algorithm (multisequence multitask deep learning algorithm) achieved an area under the curve of 0.96, with a sensitivity of 96% (48/50) and a specificity of 88% (44/50) on per-patient diagnosis level, as well as a sensitivity of 88% (129/146) and a specificity of 80% (521/654) on per-segment diagnosis level. Compared with 4 radiologists, multisequence multitask deep learning algorithm showed higher sensitivity both on per-patient (all $P<0.05$) and per-segment diagnosis levels (all $P<0.001$).

CONCLUSIONS: The CVT-detected DL algorithm herein improved diagnostic performance of routine brain magnetic resonance imaging, with high sensitivity and specificity, which provides a promising approach for detecting CVT.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: algorithm ■ area under the curve ■ brain ■ cerebral venous thrombosis ■ magnetic resonance imaging

Cerebral venous thrombosis (CVT) has shown an increasing incidence, estimated as 11.6/1 000 000 per year in recent years.[1,2] Rapid identification and treatment of CVT are of essential importance. However, missed diagnosis often occur, especially in the emergency department, due to various risk factors and atypical clinical presentation.[3]

Magnetic resonance imaging (MRI) is widely used as an initial imaging evaluation for admitted patients with cerebrovascular disorders and plays an important role in

Correspondence to: Qi Yang, MD, PhD, Department of Radiology, Beijing Chaoyang Hospital, Capital Medical University, No. 8 Gongti South Rd, Chaoyang District, Beijing 100020, China. Email yangyangqiqi@gmail.com
*X. Yang and P. Yu authors contributed equally.
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.122.041520.
For Sources of Funding and Disclosures, see page 1365.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

BBOX	bounding box
CVT	cerebral venous thrombosis
DL	deep learning
FLAIR	fluid-attenuated inversion recovery
MRI	magnetic resonance imaging
MSMT-DL	multisequence multitask DL algorithm
T1w	T1-weighted
T2w	T2-weighted

detecting CVT.[4] According to the guideline for the diagnosis of CVT,[5] routine brain MRI is used to diagnose CVT with altered signal intensity of thrombus and the absence of a flow-void signal. Nevertheless, diagnosis based on routine MRI is time-consuming, laborious, and subjective, leading to inconsistent diagnosis (with sensitivity ranging from 55% to 76%) among radiologists with different training and levels of experience.[6,7]

Deep learning (DL) is a branch of machine learning that has made great advances in diagnosing stroke by capturing morphological structures and texture features from images.[8,9] Thus, we hypothesize that DL can provide information on imaging features which were overlooked and improve the diagnostic ability of routine brain MRI. To the best our knowledge, no studies have been conducted to apply DL algorithm of routine brain MRI in the diagnosis of CVT. The aim of our study was to develop a DL algorithm based on routine brain MRI sequences (T1-weighted [T1w], T2-weighted [T2w], and fluid-attenuated inversion recovery [FLAIR]) in order to identify thrombi in the cerebral venous system and to compare the diagnostic performance of this algorithm with that of radiologists' assessment.

Data Availability

The study code is available at https://github.com/zhanghaoyue/MSMT-DL. The authors do not have permission to share data. All clinical and imaging data used in our study are not publicly available.

METHODS

This was a retrospective study approved by our institutional review board, and the written informed consent was obtained. The article was prepared in accordance with STARD (Standards for Reporting Diagnostic Accuracy Studies).

Study Participants and Data Acquisition

Data of patients suspected of CVT between April 2014 and December 2019, whose main clinical manifestations were headache, dizziness, and blurred vision, were consecutively collected from a Registry Study of Cerebral Venous Sinus Thrombosis in China (https://www.clinicaltrials.gov; Unique identifier: NCT05448248). Based on comprehensive imaging and clinical information, patients were diagnosed with CVT and non-CVT (including white matter degeneration, chronic ischemic foci, lacunar infarct, and idiopathic cranial hypertension). Patients with poor image quality, insufficient imaging examination, or uncertain diagnosis were excluded. Fifty patients with and without CVT were randomly selected for the test set. The remaining patients were included in the development set.

Imaging examination was performed on 3T MR scanners (Magnetom Verio and Magnetom Trio; Siemens, Germany) with 12-channel head coils. In routine MRI brain imaging protocol, the sequences of T1w, T2w, and FLAIR were acquired using the following parameters: T1w: repetition time (TR)/echo time (TE) 10 000/119 ms, field of view (FOV) 23.0×23.0 cm^2, matrix 308×308, slice thickness 5 mm; T2w: TR/TE 2750/85 ms, FOV 23.0×23.0 cm^2, matrix 308×308, slice thickness 5 mm; FLAIR: TR/TE 10 000/119 ms, FOV 24.0×24.0 cm^2, matrix 308×308, slice thickness 5 mm.

Conventional Imaging Evaluation and Thrombus Labeling

Two neuroradiologists (X.Y. with 6 years' experience and Q.Y. with 15 years' experience) performed a consensus reading of all available imaging studies (magnetic resonance venography, contrast-enhancement MRI, contrast-enhancement magnetic resonance venography, and black-blood thrombus imaging) and clinical information for each patient to obtain the reference standard, including the presence of thrombus as well as the thrombosed segments (superior sagittal sinus, straight sinus, left transverse sinus, right transverse sinus, left sigmoid sinus, right sigmoid sinus, confluence of sinus, and cortical veins). Based on the reference standard, one neuroradiologist (X.Y.) annotated the thrombus on the routine MRI images by drawing a bounding box (BBOX) that fully contained the region of the thrombus on each slice. To ensure the accuracy of annotations, another neuroradiologist (Q.Y.) performed a careful review and the refined annotations were served as ground truth.

Development of the DL Algorithm

The DL algorithm was developed based on 2-dimensional slices. Inspired by the Faster-RCNN (Region Based Convolutional Neural Network),[10] our DL algorithm adopted a 2-stage architecture: feature extraction and CVT detection. Two different feature extraction schemes corresponding to different input sequences were used (Figure 1).

The single-sequence scheme used single sequence (T1w, T2w, or FLAIR) to develop the corresponding DL algorithm (T1-DL, T2-DL, or FLAIR-DL). The single-sequence feature extraction module was Resnet-50,[11] where the output of each block would be fed into the Feature Pyramid Network[12] in the CVT detection module.

The multisequence scheme used all 3 sequences of routine MRI simultaneously and combined the multitask learning strategy to develop a multisequence multitask DL algorithm (MSMT-DL). For multisequence input, each sequence had an independent feature extraction branch, except for the first block, which shared model parameters among the 3 branches. In subsequent blocks, the outputs of the 3 branches underwent feature fusion in feature fusion block and were then fed into 2 different parts. The output of each block in the multisequence feature

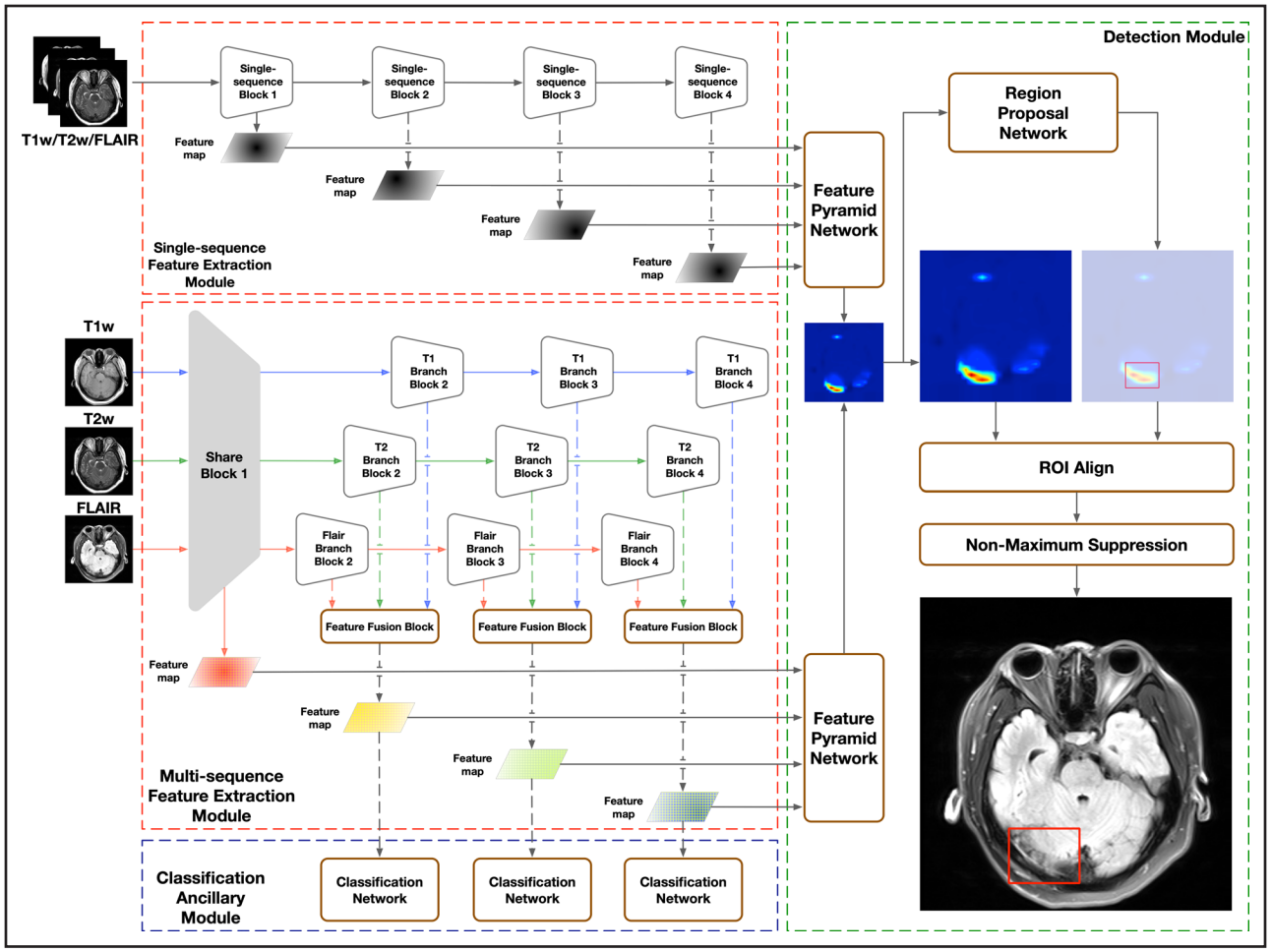

Figure 1. Deep learning model architecture.
FLAIR indicates fluid-attenuated inversion recovery; T1w, T1-weighted; and T2w, T2-weighted.

extraction module was fed into the Feature Pyramid Network. In addition, the output of each feature fusion block would also be fed into the Classification Ancillary Module, which was used to guide the algorithm to pay attention to the presence of thrombus in the slice, thus forming a multitask learning paradigm.

For each algorithm, 5-fold cross-validation was used to train models and select optimal hyper-parameters for the development set. In the test set, the final prediction of each algorithm was obtained by combining the outputs of all 5 models using the nonmaximum weighting suppression method.[13] The module design and training process of the DL algorithms was detailed in the Supplemental Material, and the code is available.

Evaluation of Diagnosis Performance of the DL Algorithm

Diagnostic performance was evaluated on per-patient and per-segment diagnosis levels in the test set.

On per-patient diagnosis level, patients were diagnosed with CVT if one or more thrombosed segments were present. The BBOX for predicting a thrombus using DL algorithm was termed the detection BBOX. The average of the top 3 largest predictive probabilities of the detection BBOXs for each patient was used as the probability of CVT positivity in evaluating per-patient diagnostic performance.

On per-segment diagnosis level, if one or more thrombi were present, the corresponding cerebral venous segment was defined as a thrombosed segment. Specifically, the ground truth BBOXs belonging to the same segment formed the ground truth volume, and the adjacent detection BBOXs were stacked to form a detection volume that was termed the D-volume. As shown in Figure 2, true-positive, false-positive, and false-negative readings on per-segment level were calculated based on the intersection-over-union between D-volumes and ground truth volumes. True-positive, false-positive, and false-negative readings were further used for the evaluation of per-segment diagnostic performance.

Reader Study of Radiologists

The reader study was conducted on the test set, and the diagnostic performance of radiologists' assessment was used for comparison with the best DL algorithm. In the first round, 4 radiologists (reader 1 with 10 years, reader 2 with 5 years, reader 3 with 5 years, and reader 4 with 3 years of experience, respectively) who were blinded to the clinical information, radiological reports, and other imaging examinations reviewed the imaging data independently. The radiologists labeled the existence of thrombus and the corresponding thrombosed venous segments. The results of the 4 readers were recorded

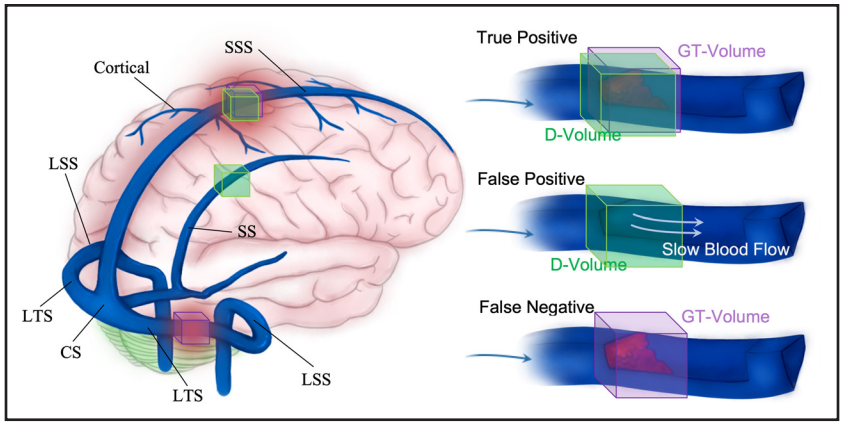

Figure 2. Examples of 3-dimensional detection performance on per-segment diagnosis level.
Adjacent detection bounding boxes (D-BBOXs) were stacked to form a detection volume, termed the D-volume. The ground truth bounding boxes (GT-BBOXs) belong to the same segment formed the GT-volume. For each patient, true-positive (TP), false-positive (FP), and false-negative (FN) readings were calculated according to the intersection-over-union (IoU) between D-volumes and GT-volumes. A D-volume was defined as a TP reading if at least one D-BBOX belonged to this detection volume, with any GT-BBOX whose IoU was not less than the judgment threshold; otherwise, the reading was defined as a FP. For a GT-Volume, readings were defined as FNs if the IoU of all the GT-BBOXs belonging to the ground truth volume with any D-BBOX was less than the judgment threshold. The judgment threshold was set to 0.5. CS indicates confluence of sinus; LSS, left sigmoid sinus; LTS, left transverse sinus; SS, straight sinus; and SSS, superior sagittal sinus.

as R1-F (reader 1 for first time), R2-F (reader 2 for first time), R3-F (reader 3 for first time), and R4-F (reader 4 for first time), respectively. In addition to the individual diagnostic results of each radiologist, a pooled determination was reported for each patient by a majority vote among 4 radiologists. If the votes were the same, the radiologist with the highest seniority was selected as the decisive voice. Majority vote results were recorded as RV-F (reader vote for first time).

After a washout period, reader 1 and reader 3 were invited to conduct a second-round reader study with access to clinical information. Similar to the first round reader study, 2 radiologists independently labeled the existence of a thrombus and the corresponding thrombosed venous segments. The results of the 2 radiologist assessments were recorded as R1-S (reader 1 for second time) and R3-S (reader 3 for second time), respectively.

Statistical Analysis

On per-patient diagnosis level, receiver operating characteristic curves and corresponding area under the curve of different algorithms were calculated and compared using the Delong test.[14] For each algorithm, Youden J statistics were used on the development set to determine a classification threshold and to then compute the accuracy, sensitivity, and specificity of the test set based on this threshold.[15] On per-segment diagnosis level, sensitivity and specificity were calculated based on the per-segment TP, false positive, and false negative.[16]

The 95% CIs were estimated with bootstrapping.[17] The McNemar test was used to compare accuracy, sensitivity, and specificity. Cohen and Fleiss κ were used to measure the inter-reader agreement between 2 radiologists and among all the radiologists, respectively. The agreement value was defined as 0 to 0.20 as slight, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1.00 as almost perfect agreement.[18] Two-tailed $P<0.05$ was considered to indicate a statistically significant difference. Statistical analyses were performed on R software (version 3.6.1; https://www.r-project.org/).

RESULTS

Study Participants

A total of 489 patients were initially recruited (358 with CVT and 131 non-CVT) in our study. Ninety-seven patients were excluded due to poor image quality (n=8), insufficient imaging examination (n=83), and uncertain diagnosis (n=6). A total of 392 participants, including 294 patients with CVT (37±14 years, 151 women) and 98 non-CVT patients (42±15 years, 65 women), were enrolled in the study. The medical and demographic characteristics of the patients are summarized in Table S1. There were no differences in the frequencies of different thrombus locations and sizes between men and women.

The development set included 244 patients with CVT with 698 thrombosed venous segments and 48 non-CVT patients. In the test set, there were 50 patients with CVT with 146 thrombosed venous segments and 50 non-CVT patients. Parenchymal lesions were visualized in 44% (22/50) of the patients with CVT in the test set.

Diagnostic Performance of DL Algorithms

The diagnostic performance of the DL algorithms on per-patient and per-segment diagnosis levels in the test set was summarized in Table 1. On per-patient diagnosis level, all DL algorithms, except T1-DL, achieved area under the curve above 0.90. The area under the curve of 3 single-sequence algorithms were between 0.82 and 0.94: T1-DL was 0.82 (95% CI, 0.73–0.90), T2-DL was 0.93 (95% CI, 0.87–0.97), and FLAIR-DL was 0.94 (95% CI, 0.90–0.98). The MSMT-DL achieved the highest area under the curve of 0.96 (95% CI, 0.92–0.99). The Delong test showed that the diagnostic performance of MSMT-DL was better

Table 1. Performances of Deep Learning Algorithms and Radiologists

Algorithm	Per-patient				Per-segment	
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]
T1-DL	0.82 [0.73–0.90]	77 (77/100) [69–85]	78 (39/50) [66–90]	76 (38/50) [64–86]	68 (99/146) [56–78]	64 (416/654) [60–67]
T2-DL	0.93 [0.87–0.97]	81 (81/100) [73–88]	78 (39/50) [66–89]	84 (42/50) [73–94]	79 (115/146) [71–86]	78 (513/654) [74–82]
FLAIR-DL	0.94 [0.90–0.98]	84 (84/100) [76–91]	84 (42/50) [73–94]	84 (42/50) [73–94]	72 (105/146) [61–82]	80 (522/654) [75–84]
MSMT-DL	0.96 [0.92–0.99]	92 (92/100) [86–97]	96 (48/50) [90–100]	88 (44/50) [78–96]	88 (129/146) [82–94]	80 (521/654) [75–84]
P value (versus MSML-DL)						
T1-DL	<0.001	0.001	0.01	0.11	<0.001	<0.001
T2-DL	0.08	0.007	0.004	0.69	<0.001	0.54
FLAIR-DL	0.23	0.04	0.03	0.69	<0.001	0.99
R1-F	…	85 (85/100) [77–92]	78 (39/50) [67–88]	92 (46/50) [84–98]	45 (65/146) [35–55]	98 (638/654) [96–99]
R2-F	…	86 (86/100) [79–93]	72 (36/50) [58–84]	100 (50/50) [100–100]	47 (68/146) [35–58]	96 (630/654) [94–98]
R3-F	…	85 (85/100) [78–92]	78 (39/50) [66–89]	92 (46/50) [84–98]	50 (73/146) [41–59]	98 (639/654) [96–99]
R4-F	…	78 (78/100) [70–85]	76 (38/50) [64–88]	80 (40/50) [68–90]	58 (85/146) [47–69]	96 (629/654) [94–98]
RV-F	…	86 (86/100) [79–92]	78 (39/50) [66–89]	94 (47/50) [87–100]	71 (104/146) [61–81]	92 (604/654) [89–95]
R1-S	…	92 (92/100) [86–97]	92 (46/50) [84–98]	92 (46/50) [83–98]	57 (83/146) [46–67]	95 (622/654) [93–97]
R3-S	…	86 (86/100) [79–92]	84 (42/50) [73–94]	88 (44/50) [78–96]	56 (82/146) [45–66]	93 (606/654) [90–96]
P value (versus MSML-DL)						
R1-F	…	0.17	0.01	0.73	<0.001	<0.001
R2-F	…	0.24	<0.001	0.03	<0.001	<0.001
R3-F	…	0.17	0.01	0.73	<0.001	<0.001
R4-F	…	0.004	0.002	0.39	<0.001	<0.001
RV-F	…	0.26	0.01	0.51	<0.001	<0.001
R1-S	…	0.99	0.69	0.69	<0.001	<0.001
R3-S	…	0.24	0.07	0.99	<0.001	<0.001

AUC indicates area under the curve; FLAIR-DL, deep learning algorithm using FLAIR-weighted image; MSMT-DL, multisequence multitask deep learning algorithm; R1-F, reader 1 for first time; R2-F, reader 2 for first time; R3-F, reader 3 for first time; R4-F, reader 4 for first time; RV-F, reader vote for first time; R1-S, reader 1 for second time; R3-S, reader 3 for second time; T1-DL, deep learning algorithm using T1-weighted image; and T2-DL, deep learning algorithm using T2-weighted image.

than that of T1-DL ($P<0.001$) but was not significantly better than that of the other algorithms. The per-patient' receiver operating characteristic curves of the different DL algorithms were shown in Figure 3.

The accuracy and sensitivity of MSMT-DL were higher than those of any single-sequence algorithm (accuracy: T1-DL, $P=0.001$; T2-DL, $P=0.007$; FLAIR-DL, $P=0.04$; sensitivity: T1-DL, $P=0.01$; T2-DL, $P=0.004$; FLAIR-DL, $P=0.03$). The specificity of MSMT-DL was 88%, which was not significantly different from that of the other algorithms (T1-DL, $P=0.11$; T2-DL, $P=0.69$; FLAIR-DL, $P=0.69$). The per-patient sensitivity of the MSMT-DL was 95% (21/22) in patients with CVT with parenchymal lesions and 96% (27/28) in patients with CVT without parenchymal lesions.

On per-segment diagnosis level, the sensitivity of the single-sequence DL algorithms was not higher than 80% (T1-DL, 68%; T2-DL, 79%; and FLAIR-DL, 72%). MSMT-DL achieved the highest sensitivity of 88% (129/146 [95% CI, 82%–94%) compared to any single-sequence algorithm (all $P<0.001$). The specificity of MSMT-DL was higher than that of T1-DL (80% versus 64%, $P<0.001$) but was similar to that of the other 2 algorithms (T2-DL, $P=0.54$; FLAIR-DL, $P=0.99$). The per-segment sensitivities of MSMT-DL were 85% (33/39), 91% (41/45), and 89% (49/55) in the acute, subacute, and chronic stage groups, respectively.

The top 3 sensitivities of MSMT-DL for detecting specific thrombosed segments were found for the left sigmoid sinus, left transverse sinus, and straight sinus (Table 2). Examples of true-positive, false-positive, and false-negative diagnosis using MSMT-DL in different thrombosed segments were illustrated in Figure S3.

Reader Study of Radiologists

The diagnostic performances of the 4 radiologists as well as a comparison with MSMT-DL were summarized in Table 1. In the first round, the per-patient sensitivities of the 4 radiologists ranged from 72% to 78%, and all were lower than that of MSMT-DL (R1-F, $P=0.01$; R2-F, $P<0.001$; R3-F, $P=0.01$; R4-F, $P=0.002$). The specificities were above 90% across the first 3 readers, whereas R4-F was slightly lower with 80% (95% CI, 68%–90%).

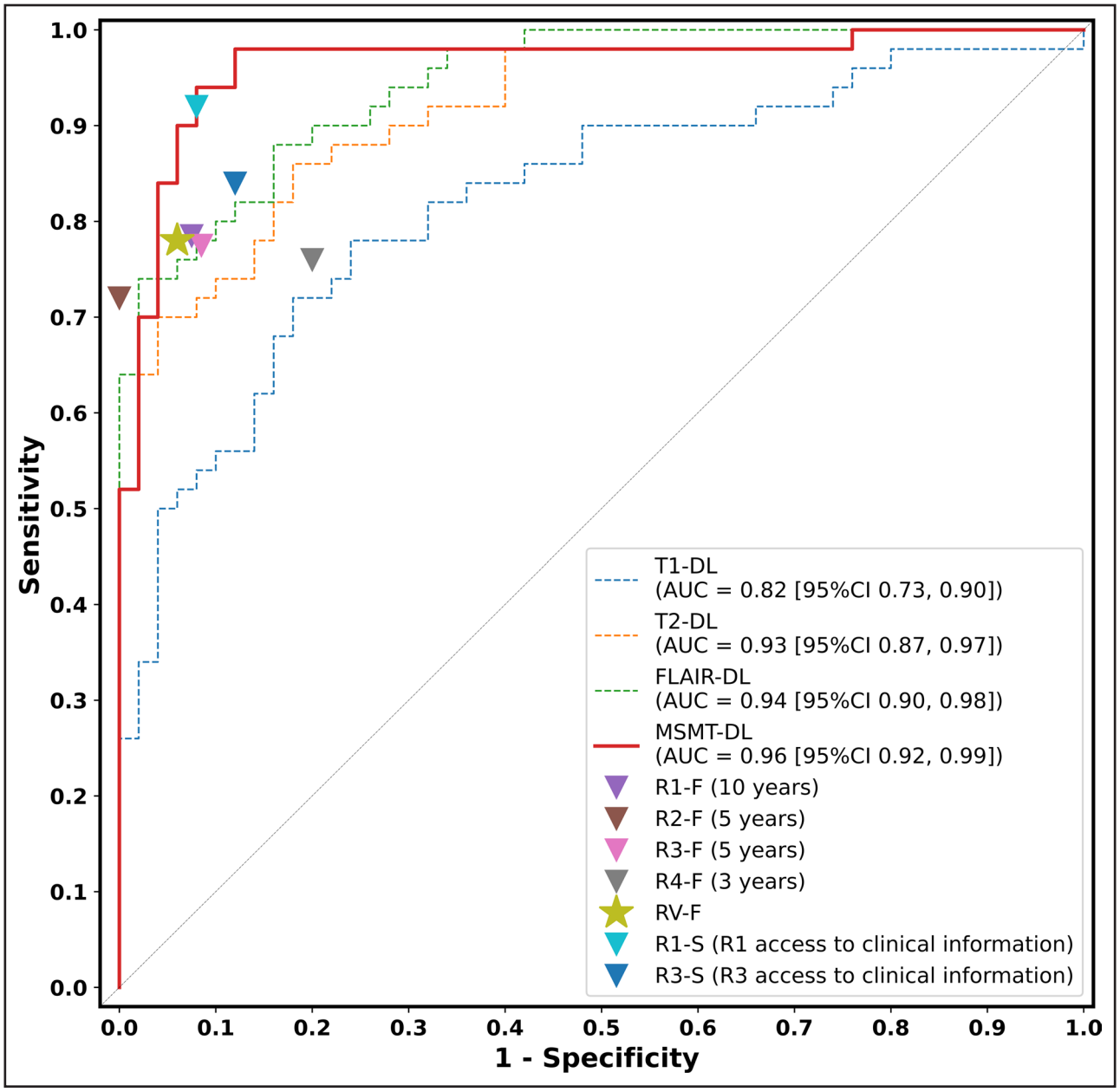

Figure 3. Comparison of deep learning (DL) algorithms and radiologists' assessment on per-patient diagnosis level. The diagnostic performances of DL algorithms were described by receiver operating characteristic (ROC) curves. According to the derived sensitivity and specificity, the diagnostic performance of the reader study was plotted on the ROC curve and was represented using inverted triangles and 5-pointed stars. AUC indicates area under the curve; FLAIR, fluid-attenuated inversion recovery; and MSMT-DL, multisequence multitask deep learning algorithm.

The specificity of MSMT-DL was not significantly different from that of all radiologists' assessments, except for R2-F (R1-F, P=0.73; R2-F, P=0.03; R3-F, P=0.73; R4-F, P=0.39). As shown in Figure 3, the corresponding points of the readers on the receiver operating characteristic curve were all inside the MSMT-DL curve, except for R2-F. The specificity of R2-F was 100% (95% CI, 100%–100%), but the corresponding sensitivity was only 72% (95% CI, 58%–84%).

The per-segment sensitivities of all 4 readers were below 60%, which were also lower than that of the MSMT-DL (all P<0.001). An additional 30% (44/146) to 44% (64/146) of thrombosed segments were revealed by MSMT-DL than 4 readers. The per-segment specificities of R1-F to R4-F were 98%, 96%, 98%, and 96%, respectively.

By aggregating all 4 readers through majority voting, the sensitivity and specificity of the per-patient diagnosis level were found to be 78% (95% CI, 66%–89%) and 94% (95% CI, 87%–100%), respectively. On per-segment diagnosis level, the sensitivity and specificity were 71% (95% CI, 61%–81%) and 92% (95% CI,

Table 2. Subgroup Analysis of Performances according to Sensitivities of MSMT-DL and Radiologists for Detecting CVT in Different Segments on Test Set

Segment	MSMT-DL	Reader 1	Reader 2	Reader 3	Reader 4	Aggregate reader
Cor (n=8)	63 (5/8)	0 (0/8)	63 (5/8)	13 (1/8)	50 (4/8)	75 (6/8)
LSS (n=17)	100 (17/17)	35 (6/17)	12 (2/17)	71 (12/17)	53 (9/17)	65 (11/17)
LTS (n=19)	100 (19/19)	21 (4/19)	16 (3/19)	37 (7/19)	47 (9/19)	47 (9/19)
RSS (n=28)	93 (26/28)	64 (18/28)	54 (15/28)	68 (19/28)	64 (18/28)	79 (22/28)
RTS (n=30)	90 (27/30)	50 (15/30)	53 (16/30)	30 (9/30)	50 (15/30)	77 (23/30)
CS (n=5)	80 (4/5)	0 (0/5)	60 (3/5)	0 (0/5)	0 (0/5)	60 (3/5)
SS (n=6)	100 (6/6)	50 (3/6)	33 (2/6)	0 (0/6)	50 (3/6)	50 (3/6)
SSS (n=33)	94 (31/33)	58 (19/33)	67 (22/33)	76 (25/33)	82 (27/33)	82 (27/33)

Data are percentages, with numerator and denominator in parentheses. Cor indicates cortical veins; CS, confluence of sinus; CVT, cerebral venous thrombosis; LSS, left sigmoid sinus; LTS, left transverse sinus; MSMT-DL, multisequence multitask deep learning algorithm; RSS, right sigmoid sinus; RTS, right transverse sinus; SS, straight sinus; and SSS, superior sagittal sinus.

89%–95%), respectively. Compared with RV-F, MSMT-DL had higher sensitivity at the per-patient diagnosis level (96% versus 78%; P=0.01) and per-segment diagnosis level (88% versus 71%; P<0.001).

On per-patient diagnosis level, the Fleiss κ coefficient was 0.71 across all 4 readers, and the Cohen κ coefficients between any 2 readers ranged from 0.62 to 0.77. On the per-segment diagnosis level, the Fleiss κ coefficient was 0.56, and the Cohen κ coefficients between any 2 readers were <0.60 (ranging from 0.47 to 0.59).

The sensitivities of MSMT-DL and the readers for detecting thrombus in different segments are shown in Table 2. In 6 of 8 segments, including left sigmoid sinus, left transverse sinus, right sigmoid sinus, right transverse sinus, straight sinus, and superior sagittal sinus, MSMT-DL had a sensitivity of over 90%. Representative CVT-positive cases (admitted with headache as the first symptom) with correct diagnosis by MSMT-DL but that were overlooked by the radiologist for case of thrombus in different venous sinus were shown in Figure 4.

Second-Round Reader Study of Radiologists

As shown in Table 1 and Figure 3, the diagnostic performance on the per-patient diagnosis level improved for both radiologists when clinical information was available. The accuracy, sensitivity, and specificity of the R1-S were 92% (95% CI, 86%–97%), 92% (95% CI, 84%–98%), and 92% (95% CI, 83%–98%), respectively. The accuracy, sensitivity, and specificity of the R3-S were 86% (95% CI, 79%–92%), 84% (95% CI, 73%–94%), and 88% (95% CI, 78%–96%), respectively. Neither radiologist performed significantly differently from MSMT-DL at the per-patient diagnosis level (all P>0.05).

On per-segment diagnosis level, there was a slight improvement in the sensitivity of radiologist assessment for 2 radiologists (R1-F to R1-S, 45% to 57%; R3-F to R3-S, 50% to 56%), but this was still significantly lower than that of MSMT-DL (all P<0.001). The specificities of R1-S and R3-S were 95% (95% CI, 93%–97%) and 93% (95% CI, 90%–96%), respectively.

On per-patient and per-segment diagnosis levels, the Cohen κ coefficients between R1-S and R3-S were 0.64 and 0.50, respectively.

DISCUSSION

In our study, we developed a MSMT-DL algorithm tailored for CVT detecting via routine brain MRI. The MSMT-DL takes T1w, T2w, and FLAIR images jointly as inputs to exploit complementary multisequence information and combined the global and local information of images through a multitask learning strategy. The MSMT-DL achieved high sensitivity on per-patient and per-segment diagnosis levels without introducing excessive false-positive detection. An additional 30% (44/146) to 44% (64/146) of thrombosed segments were further revealed by MSMT-DL as compared to radiologist.

Routine brain MRI is an essential and common imaging examination in detecting CVT at admission, which can guide the subsequent strategy of imaging examination and treatment.[19,20] However, the MRI findings of CVT vary with the composition of the thrombus and identifying the signal of the thrombus in a special stage using naked eye is laborious. Deoxyhemoglobin in acute thrombi with isointensity on T1w and hypointensity on T2w is similar to the signal of slow blood flow. Hemosiderin in chronic thrombus causes isointensity or hypointensity on T1w and hypointensity on T2w, similar to the acute thrombus signal. These situations lead to varied diagnostic performances of different routine MRI sequences in CVT. Sadigh et al[6] showed that the sensitivities of T1w and T2w images were 55% and 58%, respectively. Patel et al[21] reported that the combination of routine MRI sequences showed an overall sensitivity of 79.2% and a specificity of 89.9%, with moderate interrater agreement.[22–24]

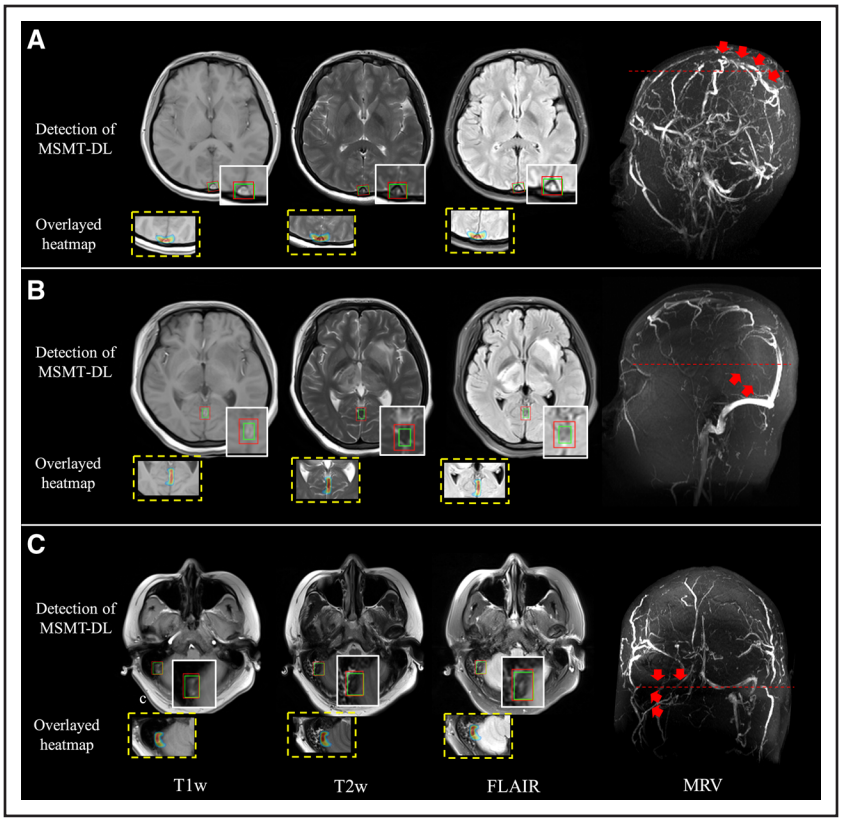

Figure 4. Three cases with correct diagnosis by the multisequence multitask deep learning algorithm (MSMT-DL) and misdiagnosis by radiologist.
A, A 19-year-old woman with thrombus in the superior sagittal sinus; (**B**) a 14-year-old girl with thrombus in the sagittal sinus; and (**C**) a 27-year-old woman with thrombus in the right transverse sinus. The green bounding box (BBOX) indicates the ground truth, and the red BBOX indicates the detection of MSMT-DL in magnetic resonance imaging images. The red line of the corresponding level is shown in the magnetic resonance venography (MRV) image. FLAIR indicates fluid-attenuated inversion recovery; T1w, T1-weighted; and T2w, T2-weighted.

In our study, DL algorithms (T1-DL, T2-DL, and FLAIR-DL) constructed based on different single-sequence images improved the diagnostic performance of the 3 MRI sequences compared with those found in the literature, especially T2w and FLAIR. This could be explained by the flow-void effect of cerebral venous sinus on T2w and FLAIR images being more obvious than within T1w images. Moreover, we developed MSMT-DL with multisequence image input (T1w, T2w, and FLAIR) through multitask learning. The multisequence feature extraction module of MSMT-DL effectively integrated information from multiple sequences into enhanced feature representations. By adding auxiliary classification branches to perform the multitask learning strategy, the MSMT-DL could observe the image characteristics from a global perspective and realize the complementarity of global information and local information, which lead to a more accurate diagnosis. Therefore, compared with the single-sequence DL algorithm, the sensitivity of MSMT-DL was increased by 12% to 18% on per-patient diagnosis level and 9% to 20% on the per-segment diagnosis level.

The results of the reader study illustrated that the diagnostic performance of MSMT-DL in diagnosing CVT was as good as or better than that of radiologist assessment. MSMT-DL may perform better on the following venous abnormalities: (1) acute/chronic thrombus (signal capture); (2) the tiny arachnoid granulations and septa (morphology capture), and (3) thrombus in the deep venous sinus (position capture). These results show that the DL algorithm has unique advantages in identifying subtle features of venous thrombus, such as weak signals, venous size, and abnormal shape. We found that the per-segment sensitivity of MSMT-DL was lower than that of assessment by one radiologist and an aggregate reader only in the diagnosis of cortical vein thrombosis, and higher than that of all radiologists and an aggregate reader in all other venous segments when diagnosing CVT based on imaging alone. We note that some patients with cortical vein thrombosis in the test set were accompanied by subarachnoid hemorrhage. Both cortical vein thrombosis and subarachnoid hemorrhage on T1w, T2w, and FLAIR images showed hyperintense signals, which may interfere with the judgment of MSMT-DL. In addition to the advantages in diagnostic performance, as a computer-aided diagnosis system, MSMT-DL provided objective results, which led to better agreement. Through the results of the 2-round reader study, we found that clinical information may help radiologists improve the diagnostic accuracy of CVT. This has inspired our research group to develop DL algorithms incorporating clinical information in future research efforts.

There are several limitations in our study. First, owing to the relatively rare incidence of the evaluated condition, sample size was not calculated and was determined based on prior study dealing with DL in acute stroke imaging.[25] Thus, the sample size of the CVT data set

might not be enough to achieve the best-performing algorithm. The generalizability of the algorithm is also limited by the fact that the study data is derived from a single medical center. External validation and multicenter studies are necessary in future studies. Second, the time of diagnosis and the presence of parenchymal abnormalities might have an effect on the algorithm. Therefore, we will enlarge the patient sample and input more information to analyze the diagnostic performance of the DL algorithm in different situations in future work. Third, there was insufficient disease diversity among patients included in this study. More patients with ischemic strokes or other brain lesions will be included in our future work to enhance the DL algorithm's discrimination and to enable it to be more adaptable to clinical complexity. Fourth, we did not evaluate radiologist's performance in detecting CVT assisted with our DL algorithm. It will be an interesting study to explore how readers and the DL algorithm interact in clinical practice and the changes in the diagnostic performance of radiologists with the aid of DL algorithm. Finally, due to SWI and T2* sequences are not standard MRI protocol in our institution for diagnosing CVT; thus, these 2 sequences were not included in this study. In the future, we will add these sequences into another algorithm construction with a larger data set.

CONCLUSIONS

In the present study, we developed a MSMT-DL algorithm based on routine brain MRI. MSMT-DL offers numerous advantages, including high sensitivity, consistency of diagnosis, and reduce time. The DL model may assist in the initial screening of patients suspected of CVT. It is critical in helping radiologists accurately and rapidly diagnose patients with CVT and locate thrombus by routine MRI.

ARTICLE INFORMATION

Received October 4, 2022; final revision received January 30, 2023; accepted February 20, 2023.

Affiliations

Department of Radiology, Beijing Chaoyang Hospital, Capital Medical University, China (X.Y., Y.L., H.L., P.S., X.L., X.J., Q.Y.). Institute of Advanced Research, Infervision Medical Technology Co, Ltd, Beijing, China (P.Y., H.Z., R.Z.). Department of Bioengineering, University of California Los Angeles (H.Z.). Department of Radiology, Beijing Ditan Hospital, Capital Medical University, China (Y.W.). Department of Emergency (J.D.) and Department of Neurosurgery (X.J.), Xuanwu Hospital, Capital Medical University, Beijing, China.

Sources of Funding

This study was funded by the National Natural Science Foundation of China (No. 92249301, No. 82025018), Capital's Funds for Health Improvement and Research (No. 2022-1-2031), Beijing Hospitals Authority's Ascent Plan (No. DFL20220303), and Beijing Key Specialists in Major Epidemic Prevention and Control.

Disclosures

None.

Supplemental Material

Model architecture
Training settings
Lesion-level detection performance
Ablation study
Tables S1–S4
Figure S1–S4
References 26–34

REFERENCES

1. Bousser MG, Ferro JM. Cerebral venous thrombosis: an update. *Lancet Neurol*. 2007;6:162–170. doi: 10.1016/S1474-4422(07)70029-7
2. Rezoagli E, Bonaventura A, Coutinho JM, Vecchié A, Gessi V, Re R, Squizzato A, Pomero F, Bonzini M, Ageno W, et al. Incidence rates and case-fatality rates of cerebral vein thrombosis: a population-based study. *Stroke*. 2021;52:3578–3585. doi: 10.1161/STROKEAHA.121.034202
3. Alluri RK, Kashefi MJ. Initial misdiagnosis of cerebral venous thrombosis: a case report. *J Med Cases*. 2013;4:734–737. doi: 10.4021/jmc1612e
4. Chalela JA, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L, Lancet SWJ. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet*. 2007;369:293–298 10.1016/S0140-6736(07)60151-2
5. Ferro JM, Bousser M-G, Canhão P, Coutinho JM, Crassard I, Dentali F, Minno M, Maino A, Martinelli I, Masuhr F, et al. European stroke organization guideline for the diagnosis and treatment of cerebral venous thrombosis - endorsed by the european academy of neurology. *Eur J Neurol*. 2017;24:1203–1213. doi: 10.1111/ene.13381
6. Sadigh G, Mullins ME, Saindane AM. Diagnostic performance of MRI sequences for evaluation of dural venous sinus thrombosis. *AJR Am J Roentgenol*. 2016;206:1298–130610.2214/AJR.15.15719
7. Pai V, Khan I, Sitoh YY, Purohit B. Pearls and pitfalls in the magnetic resonance diagnosis of dural sinus thrombosis: a comprehensive guide for the trainee radiologist. *J Clin Imaging Sci*. 2020;10:77. doi: 10.25259/JCIS_187_2020
8. Mouridsen K, Thurner P, Zaharchuk G. Artificial intelligence applications in stroke. *Stroke*. 2020;51:2573–2579. doi: 10.1161/STROKEAHA.119.027479
9. Cao C, Liu Z, Liu G, Jin S, Xia S. Ability of weakly supervised learning to detect acute ischemic stroke and hemorrhagic infarction lesions with diffusion-weighted imaging. *Quant Imaging Med Surg*. 2022;12:321–332. doi: 10.21037/qims-21-324
10. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2015;28:1137. doi: 10.1109/TPAMI.2016.2577031
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;770–778.
12. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;2117–2125.
13. Solovyev R, Wang W, Gabruseva T. Weighted boxes fusion: ensembling boxes from different object detection models. *Image Vision Comput*. 2021;107:104117. doi: 10.1016/j.imavis.2021.104117
14. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845. doi: 10.2307/2531595
15. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J*. 2005;47:458–47210.1002/bimj.200410135
16. Sung J, Park S, Lee SM, Bae W, Park B, Jung E, Seo JB, Jung KH. Added value of deep learning-based detection system for multiple major findings on chest radiographs: *A* randomized crossover study. *Radiology*. 2021;299:450–459. doi: 10.1148/radiol.2021202818
17. Wang M, Xia C, Huang L, Xu S, Qin C, Liu J, Cao Y, Yu P, Zhu T, Zhu H, et al. Deep learning-based triage and analysis of lesion burden for covid-19: a retrospective study with external validation. *Lancet Digit Health*. 2020;2:e506–e515. doi: 10.1016/S2589-7500(20)30199-0
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174. doi: 10.2307/2529310
19. Bianchi D, Maeder P, Bogousslavsky J, Schnyder P, Meuli RA. Diagnosis of cerebral venous thrombosis with routine magnetic resonance: an update. *Eur Neurol*. 1998;40:179–190. doi: 10.1159/000007978
20. McArdle CB, Mirfakhraee M, Amparo EG, Kulkarni MV. MR imaging of transverse/sigmoid dural sinus and jugular vein thrombosis. *J Comput Assist Tomogr*. 1987;11:831–838. doi: 10.1097/00004728-198709000-00017

21. Patel D, Machnowska M, Symons S, Yeung R, Fox AJ, Aviv RI, Maralani PJ. Diagnostic performance of routine brain MRI sequences for dural venous sinus thrombosis. *AJNR Am J Neuroradiol.* 2016;37:2026–203210.3174/ajnr.A4843
22. Liberman AL, Gialdini G, Bakradze E, Chatterjee A, Kamel H, Merkler AE. Misdiagnosis of cerebral vein thrombosis in the emergency department. *Stroke.* 2018;49:1504–1506. doi: 10.1161/STROKEAHA.118.021058
23. Long B, Koyfman A, Runyon MS. Cerebral venous thrombosis: a challenging neurologic diagnosis. *Emerg Med Clin North Am.* 2017;35:869–878. doi: 10.1016/j.emc.2017.07.004
24. Chiewvit P, Piyapittayanan S, Poungvarin N. Cerebral venous thrombosis: diagnosis dilemma. *Neurol Int.* 2011;3:e13. doi: 10.4081/ni.2011.e13
25. Lee H, Lee E-J, Ham S, Lee H-B, Lee JS, Kwon SU, Kim JS, Kim N, Kang D-W, . Machine learning approach to identify stroke within 4.5 hours. *Stroke.* 2020;51:860–866. doi: 10.1161/STROKEAHA.119.027611
26. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018;7794–7803.
27. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision. 2017;2961–2969.
28. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018;7132–7141.
29. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. Proceedings of the European conference on computer vision. 2018;3–19.
30. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard CVD, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal.* 2017;42:1–13. doi: 10.1016/j.media.2017.06.015
31. Yang J, Xie M, Hu C, Alwalid O, Xu Y, Liu J, Jin T, Li C, Tu D, Liu X, et al. Deep learning for detecting cerebral aneurysms with CT angiography. *Radiology.* 2021;298:155–163. doi: 10.1148/radiol.2020192154
32. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J.* 2009;2:1–35. https://github.com/ANTsX/ANTs
33. Gonzales RC, Fittes BA. Gray-level transformations for interactive image enhancement. *Mech Mach Theory.* 1977;12:111–122. doi: 10.1016/0094-114X(77)90062-3
34. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint. 2017, arXiv:1711.05101.