Deep learning model for automated diagnosis of moyamoya disease based on magnetic resonance angiography

Mingming Lu,[a,b,f] Yijia Zheng,[c,d,f] Shitong Liu,[b,f] Xiaolan Zhang,[d] Jiahui Lv,[d] Yuan Liu,[b] Baobao Li,[b] Fei Yuan,[a] Peng Peng,[a] Cong Han,[e] Chune Ma,[d] Chao Zheng,[d,g,***] Hongtao Zhang,[b,g,**] and Jianming Cai[b,g,*]

[a]Department of Radiology, Pingjin Hospital, Characteristic Medical Center of Chinese People's Armed Police Force, Tianjin, China
[b]Department of Radiology, The Fifth Medical Center, Chinese PLA General Hospital, Beijing, China
[c]Center for Biomedical Imaging Research, School of Biomedical Engineering, Tsinghua University, Beijing, China
[d]Shukun Technology Co., Ltd, Beijing, China
[e]Department of Neurosurgery, Chinese PLA General Hospital, Beijing, China

eClinicalMedicine
2024;77: 102888
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102888

Summary

Background This study explores the potential of the deep learning-based convolutional neural network (CNN) to automatically recognize MMD using MRA images from atherosclerotic disease (ASD) and normal control (NC).

Methods In this retrospective study in China, 600 participants (200 MMD, 200 ASD and 200 NC) were collected from one institution as an internal dataset for training and 60 from another institution were collected as external testing set for validation. All participants were divided into training (N = 450) and validation sets (N = 90), internal testing set (N = 60), and external testing set (N = 60). The input to the CNN models comprised preprocessed MRA images, while the output was a tripartite classification label that identified the patient's diagnostic group. The performances of 3D CNN models were evaluated using a comprehensive set of metrics such as area under the curve (AUC) and accuracy. Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the CNN's decision-making process in MMD diagnosis by highlighting key areas. Finally, the diagnostic performances of the CNN models were compared with those of two experienced radiologists.

Findings DenseNet-121 exhibited superior discrimination capabilities, achieving a macro-average AUC of 0.977 (95% CI, 0.928–0.995) in the internal test sets and 0.880 (95% CI, 0.786–0.937) in the external validation sets, thus exhibiting comparable diagnostic capabilities to those of human radiologists. In the binary classification where ASD and NC were group together, with MMD as the separate group for targeted detection, DenseNet-121 achieved an accuracy of 0.967 (95% CI, 0.886–0.991). Additionally, the Grad-CAM results for the MMD, with areas of intense redness indicating critical areas identified by the model, reflected decision-making similar to human experts.

Interpretation This study highlights the efficacy of CNN model in the automated diagnosis of MMD on MRA images, easing the workload on radiologists and promising integration into clinical workflows.

Funding National Natural Science Foundation of China, Tianjin Science and Technology Project and Beijing Natural Science Foundation.

Keywords: Artificial intelligence; Convolutional neural network; Automated diagnosis; Moyamoya disease

Introduction

Moyamoya disease (MMD) is a relatively rare cerebrovascular disease.[1] Compared with other cerebrovascular diseases such as atherosclerosis, MMD occurs more frequently in children and young people, with a faster disease course and a higher incidence of cerebral hemorrhage.[2–4] Thus, early diagnosis and intervention of MMD are necessary. Compared with digital subtraction

*Corresponding author. Department of Radiology, The Fifth Medical Center, Chinese PLA General Hospital, Beijing, 100853, China.
**Corresponding author. Department of Radiology, The Fifth Medical Center, Chinese PLA General Hospital, Beijing, 100853, China.
***Corresponding author. Shukun Technology Co., Ltd, Beijing, 102200, China.
E-mail addresses: beili12345@sina.cn (J. Cai), zhanghongtao192@sina.com (H. Zhang), zchao@shukun.net (C. Zheng).
[f]These authors have contributed equally to this work and share first authorship.
[g]These authors are co-corresponding authors and contributed equally to this manuscript.

Research in context

Evidence before this study

We searched PubMed and Google Scholar database up to July 30, 2024, for papers published in English, using the terms "moyamoya disease (MMD)", "MRA", "automatic diagnosis" and "deep learning". Our search showed that previous studies primarily reported deep learning models for the diagnosis of moyamoya disease using DSA or X-ray. However, magnetic resonance angiography (MRA) is less invasive and doesn't require contrast agents like in DSA. These may dissatisfy the requirement of a deep learning-based convolutional neural network (CNN) to automatically recognize MMD on MRA images.

Added value of this study

The DenseNet-121 exhibited superior discrimination capabilities among MMD, atherosclerotic disease (ASDs), and normal controls (NCs) in both the internal and external datasets, achieving almost the same levels as those of human radiologists.

Implications of all the available evidence

This study highlights the efficacy of CNN model in the automated diagnosis of MMD on MRA images, easing the workload on radiologists and promising integration into clinical workflows.

angiography (DSA) and computed tomography angiography (CTA), magnetic resonance angiography (MRA) is more commonly used for the screening diagnosis of MMD in clinical practice because it is less invasive and does not require contrast agents.[5,6] However, because of the relative rarity of MMD and the large number of cerebrovascular examination images, relying solely on the radiologist's visual evaluation is time-consuming and labor-intensive and is prone to misdiagnosis and missed diagnosis. Thus, providing rapid and accurate diagnosis of MMD is increasingly warranted as the utilization of cerebrovascular MR angiography expands.

In recent years, deep learning algorithms have developed rapidly and shown significant promise in medical imaging, including chest CT, coronary CTA, and musculoskeletal magnetic resonance imaging (MRI).[7–9] Deep learning algorithms have been widely recognized as powerful tools for automatically extracting information and making rapid and accurate diagnoses.[10] Thus, this study explored the potential of a deep learning-based convolutional neural network (CNN) to automatically recognize MMD on MRA images and distinguish it from atherosclerotic disease (ASD) and normal control (NC).

Methods

Data collection

A total of 600 consecutive participants, including 200 MMDs, 200 ASDs, and 200 NCs, who underwent brain MRA examinations at the Fifth Center of Chinese PLA General Hospital between July 2020 and September 2023, were retrospectively collected for training, validation, and testing purposes (Table 1). Sample weights were not specifically applied to the cohort as the data collection was retrospective, and all participants were consecutively collected without survey design considerations. Stratified random sampling was applied to the internal dataset to ensure a balanced distribution across the training, validation, and test sets, following a common split ratio of 7.5:1.5:1. Specifically, the training set included 150 participants per group (MMD, ASD, and NC), the validation set included 30 participants per group, and the internal test set included 20 participants per group. To better assess the model's performance across both internal and external data sets, the external validation sets was deliberately constructed to mirror the internal test set's composition, including 20 MMDs, 20 ASDs, and 20 normal controls from the Characteristic Medical Center of Chinese People's Arme Police Force between May 2016 and December 2023. This consistency in sample distribution allows for a more direct comparison of the model's performance across different datasets. Table 1 presents the datasets and patient demographics for each data sets.

All patients in our database were diagnosed independently by two neuroradiologists and when the results between the two readers were inconsistent, the diagnosis was decided by consensus. MMD was finally diagnosed using digital subtraction angiography (DSA) and/or MR angiography (MRA) based on the diagnostic criteria of the 2012 Research Committee on Moyamoya Disease of the Ministry of Health and Welfare of Japan.[11] ASD was diagnosed when luminal stenosis of intracranial vessels with eccentric wall thickening or an increase or no change in the outer diameter was detected on cine MRA and risk factors of atherosclerosis were identified.[12] Participants whose MRA images showed completely normal were identified as normal controls.[13] When the grouping results between the two readers were inconsistent, grouping was decided by consensus. A total of 15 cases (2.5%) were found to be inconsistent between the two neuroradiologists when being grouped into MMD, NC and ASD based on clinical and radiological information of included patients. Of them, 8 cases that were inconsistent were finally grouped into MMD and 7 were determined as ASD by consensus. The following conditions were excluded: (1) secondary moyamoya phenomenon caused by other clear disease entities; (2) other cerebrovascular diseases; (3) brain

	Internal datasets			External datasets		
	NC	MMD	ASD	NC	MMD	ASD
Age (years)	53.2 ± 16.3	40.9 ± 10.7	69.2 ± 11.1	49.8 ± 13.9	43.1 ± 12.5	69.0 ± 13.2
Sex (male)	105 (52.5%)	83 (41.5%)	136 (68%)	11 (55%)	8 (40%)	14 (70%)
Height (cm)	167.2 ± 10.2	166.5 ± 9.1	164.6 ± 10.3	166.7 ± 9.8	164.6 ± 10.2	165.3 ± 10.4
Weight (kg)	65.4 ± 13.2	64.3 ± 12.5	63.2 ± 11.6	64.6 ± 10.7	63.2 ± 12.0	63.9 ± 11.6

Data are mean ± standard deviation for Age, Height and Weight or number (percentage) for Sex. NC: normal control, MMD: moyamoya disease, ASD: atherosclerotic disease.

Table 1: **Basic characteristics of enrolled participants.**

tumors or trauma; (4) any medical history of neurosurgery. The included cohorts in our study can be considered to be representative. First, consecutive patients in the daily clinical practice were included in the present study. Second, standardized including criteria for MMD and ASD was adopted. Third, strict screening was performed according to the including criteria. Finally, no irrational exclusion criteria were specially set.

MR imaging parameters

The internal datasets were acquired on a 3-T MR system (Magnetom SKYRA, Siemens Healthcare) with an 8-channel head coil. The scan sequences of MRA were as follows: time of flight (TOF), repetition time (TR)/echo time (TE) of 20/3.5 ms, slice gap of 0.5 mm, field of view (FOV) of 200 × 200 × 60 mm^3, matrix size of 360 × 360, slice thickness of 0.5 mm. The external validation datasets were obtained using a 3-T MR system (Magnetom Verio, Siemens Healthcare) with a 12-channel head coil. MRA images were obtained with the following parameters: TOF, TR/TE of 20/3.6 ms, slice gap of 0.5 mm, FOV of 200 × 200 × 60 mm^3, matrix size of 360 × 360, and slice thickness of 0.5 mm.

Data partition and preprocessing

The model employed a TOF-MRA image dataset, along with the corresponding subject groups (NC, MMD, and ASD). Each group was randomly divided into training, validation, and test datasets in a ratio of 75:15:10, respectively.

To enhance the adaptability of our model to TOF-MRA images, we implemented various data-augmentation techniques. These included mirroring, random translations, rotations, scaling (magnification and reduction), random elastic transformations, pixel enhancements based on proportionality or pixel values, random contrast enhancements, random Gaussian smoothing, and random Gaussian noise.[14] Such enhancements were crucial in ensuring the diversity and robustness of our training dataset. This approach to data augmentation is particularly prevalent in deep learning applications, particularly when dealing with smaller datasets, such as those associated with medical conditions.

Deep learning network

In this study, we used the Shukun Technology research platform (https://medresearch.shukun.net/) to implement five commonly used deep learning classification networks: DenseNet121, ResNet50, SENet154, SEResNet50, and SEResNext50. Each network architecture was configured using default parameters corresponding to those in MONAI (https://monai.io/).[15–17]

All the classification models maintained consistent input and output formats. Specifically, the input to the Convolutional Neural Network (CNN) models consisted of preprocessed MRA images resampled to dimensions of 256 × 256 × 128 with a single channel. The output was structured as a tripartite classification label by employing a SoftMax activation function to estimate the probability of belonging to one of the three diagnostic categories. The architectural details of the neural networks, along with their inputs and outputs, are illustrated in Fig. 1. The number of epochs for each study was 200. As shown in Fig. 2, our model's performance begins to stabilize after 100 epochs, with the loss curve flattening and accuracy showing minimal improvement. However, we extended training to 200 epochs to ensure full convergence, capture complex patterns, and validate stability across mini-batch variations and data fluctuations. While some redundancy exists beyond 100 epochs, this additional buffer helped avoid underfitting and provided robustness without signs of overfitting.

During the training phase, we meticulously optimized hyperparameters based on the performance on the validation set, employing back-propagation and gradient descent algorithms to minimize the Cross-Entropy Loss function. Specifically, the momentum stochastic gradient descent method (Adam optimization) was utilized for this purpose. Adjustments to hyperparameters during training, such as the learning rate, weight decay, and batch size, were informed by ongoing performance evaluations, aiming to enhance the model's overall effectiveness.

Visualization of CNN decision-making using Grad-CAM

Finally, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the decision-making

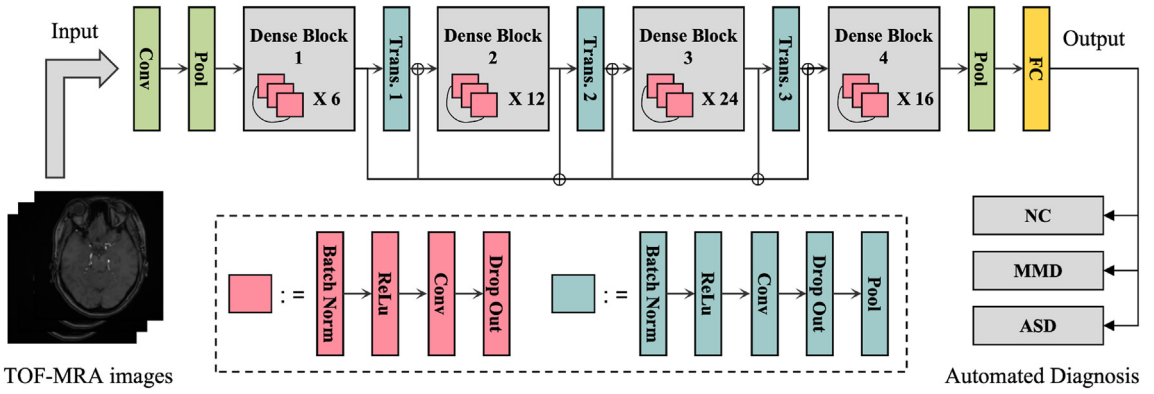

Fig. 1: DenseNet121 Architecture with Inputs and Outputs. **Abbreviations**: MMD: moyamoya disease, ASD: atherosclerotic disease, NC: normal control.

process of the CNN in MMD diagnosis by highlighting key areas. Grad-CAM leverages the gradients of the target concepts from the final convolutional layer to generate a localization map.[18] This map emphasizes critical areas within images that significantly influence the diagnostic decisions of the network, thereby clarifying the pathways through which decisions are made.

Specifically applied to the diagnosis of NC, MMD, and ASD, this method highlights regions crucial to the AI's diagnostic processes. By visually demystifying the AI's operational basis, Grad-CAM aids in fostering a deeper understanding of CNN operations and promotes the development of more explainable AI systems in the field of medical diagnostics.

Human performance evaluation

For the internal and external datasets, two experienced human radiologists (Mingming Lu and Shitong Liu), who were blinded to the patients' information, independently classified the MRA images and evaluated their performance. The human performances were compared with those of the CNN models.

Statistical analysis and model evaluation

All continuous variables in this study were presented as mean ± standard deviation and were analyzed using either Student's t-test or paired t-test, depending on their paired or unpaired nature, respectively. Categorical variables were reported as numbers (percentages) and analyzed using Pearson's chi-squared test. The Shapiro–Wilk test was performed and a normality distribution was verified when $p > 0.05$.

Upon completion of the model training, the performances of the 3D CNN models for three-class classification were first evaluated using a comprehensive set of metrics. These included macro-average area under the curve (AUC), macro-average sensitivity (SEN), macro-average positive predictive value (PPV), macro-average F1-score (F1), macro-average accuracy (mACC), and accuracy (ACC).

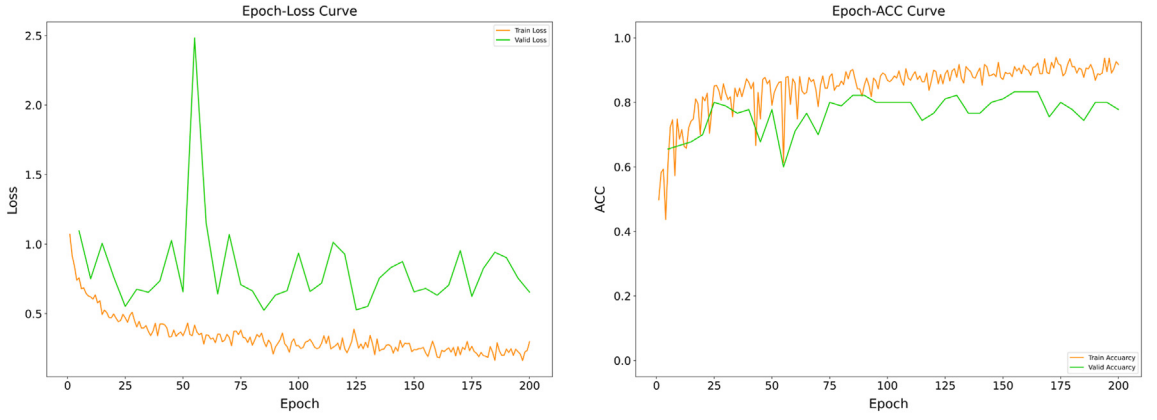

Fig. 2: The accuracy and loss curves for the DenseNet121 network during the training.

Subsequently, each category was treated as a positive value in isolation, with the remaining two categories serving as negative values to measure the detection performance for each category. Specific metrics included AUC, ACC, SEN, PPV, and F1-score. For ACC, SEN, and PPV, the Wilson score interval was used to calculate 95% confidence intervals (CIs), while the bootstrap method was employed for AUC.

The performance metrics for both the three-class and binary classifications were evaluated across all datasets, including the training, validation, testing, and external validation datasets. In addition, to gain a deeper understanding of the algorithm's performance, confusion matrices were calculated and the area under the receiver operating characteristic (ROC) curves (AUC) were plotted for both the testing and external validation sets.

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki after approval by the Institutional Review Boards of the fifth medical center of PLA general hospital (KY-2020-4-29-1) and the Characteristic Medical Center of Chinese People's Arme Police Force (2016-0029). Informed consent was obtained from all participants or their legal guardians.

Role of the funding source

The funder of the study had no role in study design, in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the manuscript for publication.

Results

Basic characteristics of enrolled participants

Among all eligible participants, the mean age was 55.2 ± 12.1 years, and the proportion of female participants was 54.1% (357 of 660). The mean ages of the internal participants in the MMD, ASD and control group were 40.9 ± 10.7 years, 69.2 ± 11.1 years and 53.2 ± 16.3 years, respectively. The mean ages of the external participants in the MMD, ASD and control group were 43.1 ± 12.5 years, 69.0 ± 13.2 years and 49.8 ± 13.9 years, respectively (Table 1).

Three-class classification performance of deep learning models

For the internal dataset, 450 images were utilized for training, and 90 images were set aside as the validation set. A total of 60 TOF-MRA images were separated into a test set using random selection, with equal distribution among the three diagnostic groups: NC, MMD, and ASD, each contributing 20 images.

All the models underwent 200 epochs of computation, selecting the parameters from the epoch with the highest accuracy on the validation set as the optimal model configuration. The performance metrics for the optimal parameters of all models are listed in Table 2. Specifically, the accuracy and loss curves for the DenseNet121 network during the training process are illustrated in Fig. 2.

For the internal validation sets, which comprised 20 images per category (NC, MMD, and ASD), the mACCs for the three-class classification in the test dataset were observed as follows: 0.911 for DenseNet121, 0.887 for Resnet50, 0.867 for SENet154, 0.922 for SEResNet50, and 0.833 for SEResNext50.

Three-class classification using best-performing deep learning models and human radiologists

Notably, DenseNet-121 exhibited superior discrimination capabilities, achieving a high Area Under the Curve (AUC) of 0.977 (95% CI, 0.928–0.995) in the test sets and 0.870 in the external validation sets. The SEN, PPV, F1-score, mACC, and ACC for DenseNet-121 in the internal test sets were recorded as 0.867, 0.864, 0.864, 0.911, and 0.867 respectively. In the external validation, the SEN, PPV, F1, mACC, and ACC were 0.733, 0.753, 0.733, 0.822, and 0.733, respectively, using DenseNet-121. Furthermore, a subgroup exploratory analysis on model performance stratified by age and sex was also performed.

The per-class ROC curves for the best-performing models in both the test dataset and external validation sets, along with heatmaps for the confusion matrix, are presented in Fig. 3. These visualizations provide a clear representation of the model's performances across different classes and datasets.

In the subgroup analysis stratified by age, we divided the patients in the test set into three age groups, ensuring that each group had a balanced number of participants. The results demonstrate that the model's classification performance (MMD, ASD, and NC) in terms of accuracy (0.857, 0.895, 0.850) and recall (0.833, 0.889, 0.857) is consistent across different age groups. This indicates that age did not significantly interfere with the model's performance, thereby confirming the model's stability and robustness across various age ranges. It is noteworthy that the 49–62 age group exhibited the best overall performance, particularly in terms of precision (0.911) and F1 score (0.885). We hypothesize that this could be related to the sample distribution and the model's adaptability to this specific age range (Supplemental Table S1).

In the subgroup analysis stratified by sex, we conducted chi-square tests on the internal test set to examine whether there are significant differences in sex distribution across different groups. Additionally, we explored the potential impact of sex on model prediction results through chi-square tests. The chi-square test indicated no significant difference in sex distribution across different groups ($p = 0.233$) and the relationship between sex and model prediction results showed no significant association ($p = 0.156$). The analysis results

CNN model	AUC	SEN	PPV	F1	mACC	ACC
DenseNet121						
Train (n = 450)	1.000 (1.000, 1.000)	1.000	1.000	1.000	1.000	1.000
Val (n = 90)	0.936 (0.882, 0.967)	0.833	0.838	0.835	0.889	0.833
Test (n = 60)	**0.977 (0.928, 0.995)**	**0.867**	**0.864**	**0.864**	**0.911**	**0.867**
Ext (n = 60)	0.880 (0.786, 0.937)	0.733	0.733	0.733	0.822	0.733
ResNet50						
Train (n = 450)	1.000 (1.000, 1.000)	0.991	0.991	0.991	0.994	0.991
Val (n = 90)	0.946 (0.889, 0.979)	0.856	0.863	0.857	0.904	0.856
Test (n = 60)	0.965 (0.918, 0.989)	0.817	0.821	0.817	0.878	0.817
Ext (n = 60)	0.831 (0.723, 0.894)	0.633	0.728	0.630	0.755	0.633
SENet154						
Train (n = 450)	1.000 (1.000, 1.000)	1.000	1.000	1.000	1.000	1.000
Val (n = 90)	0.932 (0.877, 0.963)	0.833	0.844	0.833	0.889	0.833
Test (n = 60)	0.968 (0.916, 0.989)	0.800	0.797	0.795	0.867	0.800
Ext (n = 60)	0.765 (0.649, 0.844)	0.567	0.556	0.560	0.711	0.567
SEResNet50						
Train (n = 450)	1.000 (1.000, 1.000)	0.993	0.993	0.993	0.996	0.993
Val (n = 90)	0.929 (0.867, 0.968)	0.844	0.846	0.843	0.896	0.844
Test (n = 60)	0.970 (0.917, 0.993)	0.883	0.884	0.883	0.922	0.883
Ext (n = 60)	0.755 (0.633, 0.834)	0.550	0.589	0.558	0.700	0.550
SEResNext50						
Train (n = 450)	1.000 (1.000, 1.000)	1.000	1.000	1.000	1.000	1.000
Val (n = 90)	0.906 (0.833, 0.954)	0.822	0.846	0.826	0.881	0.822
Test (n = 60)	0.909 (0.817, 0.956)	0.750	0.788	0.755	0.833	0.750
Ext (n = 60)	0.740 (0.629, 0.822)	0.517	0.483	0.486	0.678	0.517

Date are presented as value or value (lower bound, upper bound, 95% confidence interval). NC: normal control, MMD: moyamoya disease, ASD: atherosclerotic disease, CNN: convolutional neural work, AUC: macro-average area under curve, SEN: macro-average sensitivity, PPV: macro-average positive predictive value, F1: macro-average F1-score, mACC: macro-average accuracy, ACC: accuracy, Val: validation, Ext: external validation. The bold font represents the optimal diagnostic performance.

***Table 2*: The performance metrics for all models in the tripartite classification of MMD, ASD and NC.**

indicated that there were no significant differences in sex distribution across groups, nor was there a significant impact of sex on model prediction outcomes in both datasets. The detailed results have been saved in the supplementary materials as Supplemental Table S2.

Compared to the diagnostic performances of radiologists for internal test datasets (AUC and ACC were 0.981 (95% CI, 0.933–0.996) and 0.891) and external validation datasets (AUC and ACC were 0.921 (95% CI, 0.857–0.961) and 0.853), the DenseNet-121 model exhibited a comparable ability to human radiologists in Three-Class Classification for internal test datasets (AUC and ACC were 0.977 (95% CI, 0.928–0.995) and 0.867) and external validation datasets (AUC and ACC were 0.880 (95% CI, 0.786–0.937) and 0.836). The kappa values of inter-reader agreement for classification of MMD, ASD and NC were 0.87 (95% CI: 0.74–0.95).

Binary classification using best-performing deep learning models

Overall, the DenseNet121 model demonstrated a high performance in differentiating each type within the internal test set, as detailed in Table 3. Specifically, the DenseNet121 model excelled in detecting MMD within the test set. The performance metrics for MMD detection using DenseNet121 in the test set were as follows: AUC, 0.992 (95% CI, 0.974–1.000); ACC, 0.967 (95% CI, 0.886–0.991); SEN, 1.000 (95% CI, 0.839–1.000); PPV, 0.909 (95% CI, 0.722–0.975); and F1-score, 0.952.

In the external validation sets, the DenseNet121 model performed better in detecting NC than the internal test set: AUC, 0.974 vs. 0.965; ACC, 0.933 vs. 0.883; SEN, 0.900 vs. 0.800; PPV, 0.900 vs. 0.842; and F1-score, 0.900 vs. 0.821. However, the performance of DenseNet121 in detecting MMD and ASD was lower in the external validation sets than that in the internal test set.

Analysis of Grad-CAM images by CNN model

Fig. 4 shows representative Grad-CAM images of the MMD, ASD, and NC groups. Grad-CAM images were reviewed by a radiologist, and the region of interest of the deep learning model was consistent with that of humans. In Grad-CAM images of MMD, the CNN model (DenseNet-121) successfully detected bilateral stenotic/occlusive ICAs and moyamoya vessels from the

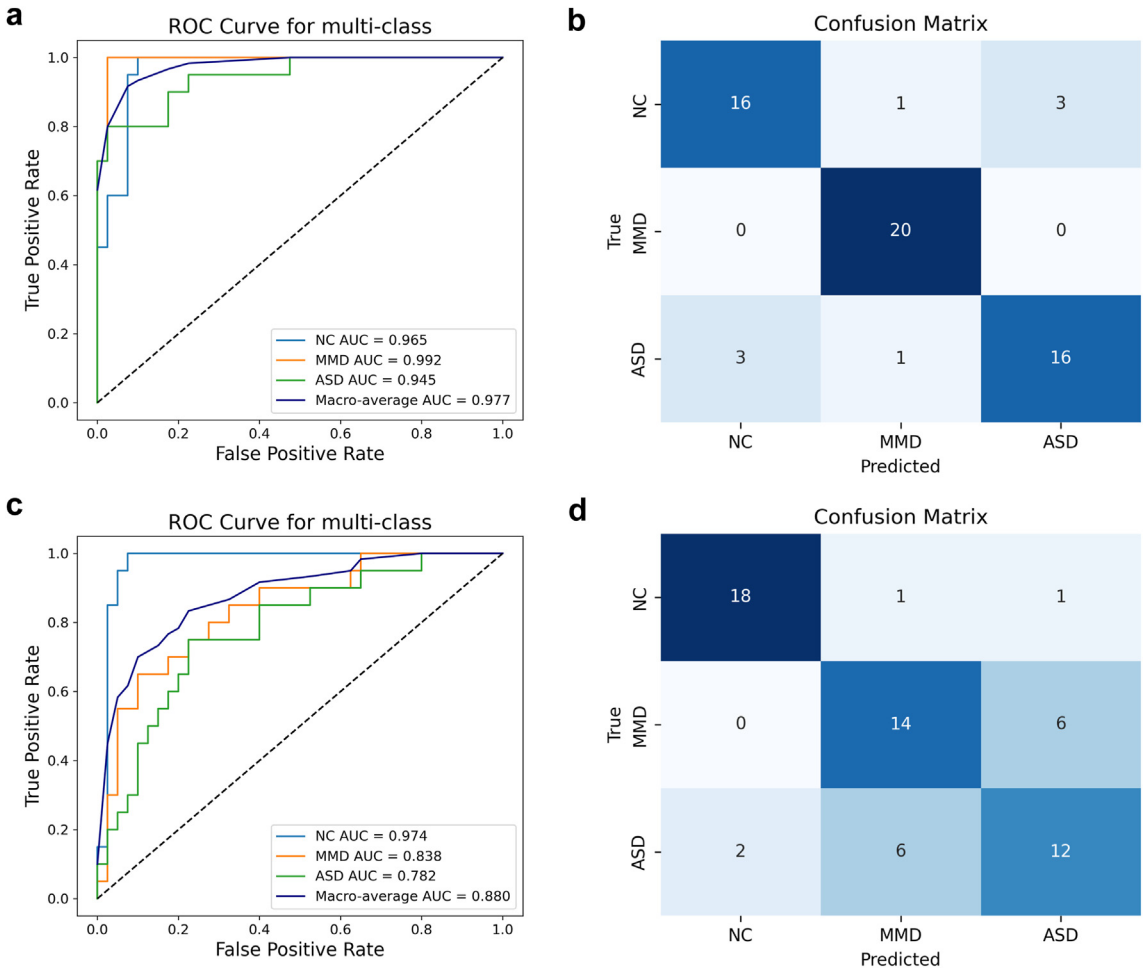

Fig. 3: The per-class Receiver Operating Characteristic (ROC) curves for the best-performing models in both the test dataset and external validation sets (a, c), along with heatmaps for the confusion matrix (b, d). **Abbreviations**: MMD: moyamoya disease, ASD: atherosclerotic disease, NC: normal control.

image slices of basal brain which exhibited "hot areas" in these regions. For ASD, the Grad-CAM images showed that the highlighted areas crowded together around the image areas of the involved ICAs and their branches.

Discussion

The purpose of this study was to develop and evaluate a learning model capable of automatically detecting MMD in real-world clinical practice using the original MRA imaging examination with model validation on an external dataset. We found that DenseNet-121 exhibited superior discrimination capabilities among MMD, ASDs, and NCs in both the internal and external datasets, achieving almost the same levels as those of human radiologists. CNN-based deep learning models enhance the accuracy and efficiency of MMD detection, easing the workload of radiologists and promising integration into clinical workflows.

Previous studies have investigated the diagnosis of MMD using AI technology. Kim innovatively developed deep learning (DL) algorithms to distinguish MMD in plain skull radiograph images and they thought that MMD patients might have distinct features in the viscerocranium.[19] Akiyama et al. selected two axial continuous slices of conventional T2-weighed imaging at the level of the basal cistern and basal ganglia to detect MMD using deep learning techniques, and their results showed excellent accuracy in the differential diagnosis of MMD.[20] Additionally, Hao et al. evaluated the diagnostic performance of CNN in MMD detection on preprocessed DSA and they found that the CNN model performed well.[21] However, several problems still need to be solved, despite the deep learning algorithms performing well in MMD detection in previous studies.

DenseNet121	AUC	ACC	SEN	PPV	F1
NC vs. MMD、ASD					
Train (n = 450)	1.000 (1.000, 1.000)	1.000 (0.992, 1.000)	1.000 (0.975, 1.000)	1.000 (0.975, 1.000)	1.000
Val (n = 90)	0.920 (0.860, 0.971)	0.878 (0.794, 0.930)	0.800 (0.627, 0.905)	0.828 (0.655, 0.924)	0.814
Test (n = 60)	0.965 (0.921, 1.000)	0.883 (0.778, 0.942)	0.800 (0.584, 0.919)	0.842 (0.624, 0.945)	0.821
Ext (n = 60)	0.974 (0.922, 1.000)	0.933 (0.841, 0.974)	0.900 (0.699, 0.972)	0.900 (0.699, 0.972)	0.900
MMD vs. NC、ASD					
Train (n = 450)	1.000 (1.000, 1.000)	1.000 (0.992, 1.000)	1.000 (0.975, 1.000)	1.000 (0.975, 1.000)	1.000
Val (n = 90)	0.974 (0.942, 0.996)	0.933 (0.862, 0.969)	0.867 (0.703, 0.947)	0.929 (0.774, 0.980)	0.897
Test (n = 60)	**0.992 (0.974, 1.000)**	**0.967 (0.886, 0.991)**	**1.000 (0.839, 1.000)**	**0.909 (0.722, 0.975)**	**0.952**
Ext (n = 60)	0.838 (0.71, 0.933)	0.783 (0.664, 0.879)	0.700 (0.481, 0.855)	0.677 (0.454, 0.828)	0.683
ASD vs. NC、MMD					
Train (n = 450)	1.000 (1.000, 1.000)	1.000 (0.992, 1.000)	1.000 (0.975, 1.000)	1.000 (0.975, 1.000)	1.000
Val (n = 90)	0.891 (0.819, 0.952)	0.856 (0.768, 0.914)	0.833 (0.664, 0.927)	0.758 (0.590, 0.872)	0.794
Test (n = 60)	0.945 (0.873, 0.992)	0.883 (0.778, 0.942)	0.800 (0.584, 0.919)	0.842 (0.624, 0.945)	0.821
Ext (n = 60)	0.782 (0.648, 0.897)	0.750 (0.628, 0.842)	0.600 (0.387, 0.781)	0.632 (0.410, 0.809)	0.615

NC: normal control, MMD: moyamoya disease, ASD: atherosclerotic disease, AUC: area under curve, SEN: sensitivity, PPV: positive predictive value, F1: F1-score, ACC: accuracy, Val: validation, Ext: external validation. Data are presented as value or value (lower bound, upper bound, 95% confidence interval). The bold font represents the optimal diagnostic performance.

Table 3: **The performance metrics for DenseNet121 in the binary classification.**

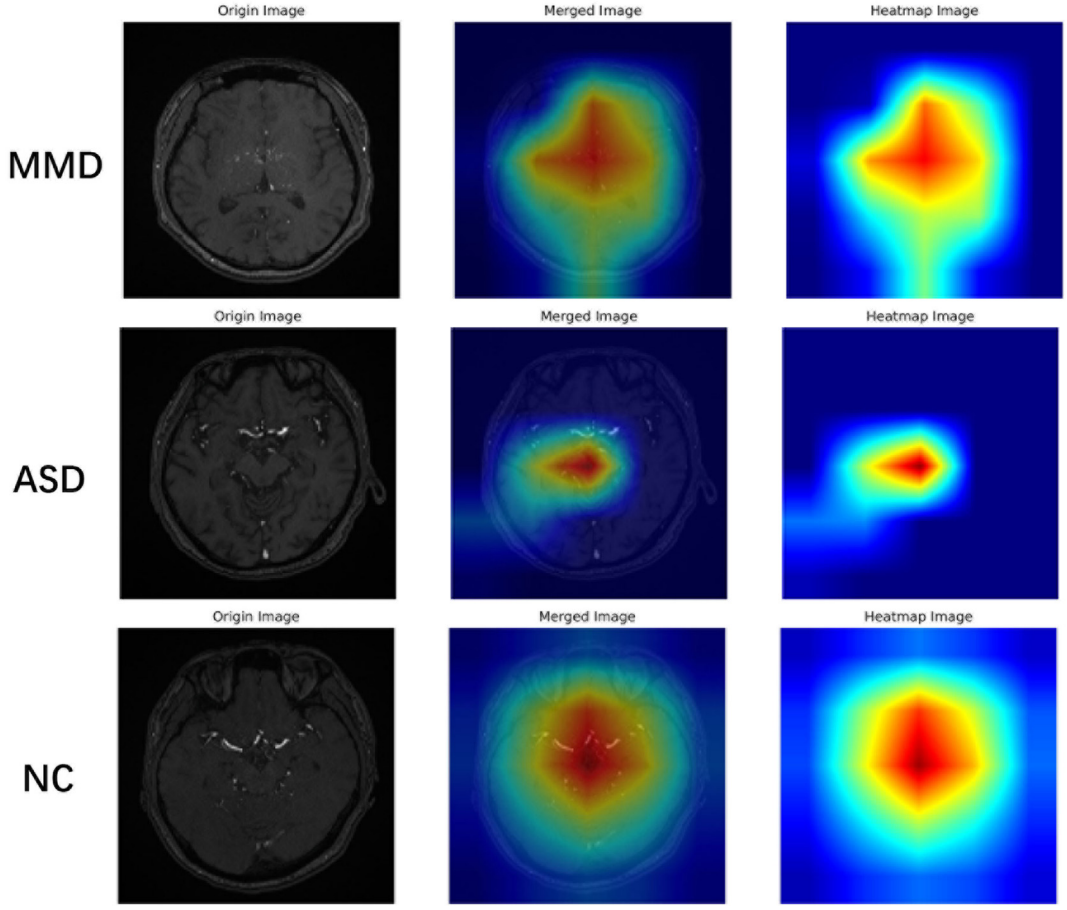

Fig. 4: Grad-CAM results (including origin image, merged image and heatmap image) for the moyamoya disease (MMD), atherosclerotic disease (ASD) and normal control (NC). The Grad-CAM images of MMD showed that "hot areas" were present around the regions of bilateral stenotic ICAs and moyamoya vessels at the base of brain. For ASD, the Grad-CAM images showed that the highlighted areas crowded together around the image areas of the right ICA and their branches which exhibited luminal stenosis. The last row showed the characteristics of Grad-CAM mages of NC.

First, MRA, as one of the most utilized imaging methods for detecting MMD, was not evaluated using AI technology in previous studies. Second, MMD patients were differentially diagnosed only with normal controls, and atherosclerotic diseases were not considered previously. Furthermore, special selection of image slices or regions of interest is usually required before further AI analysis in previous studies.[22,23] Accordingly, we developed convolutional neural network algorithms to automatically recognize MMD and distinguish MMD from atherosclerotic disease (ASD) and normal control (NC) on MRA images without special pre-processing or feature engineering.

In the binary classification, we observed that the CNN models had a higher accuracy and sensitivity in identifying MMD than ASDs in both internal and external datasets. This might be because MMD has more recognizable features than ASD. In MMD, color mapping on MRA images showed that the "hot areas" crowded together around the image areas of terminal bilateral intracranial internal carotid arteries (ICAs) and image slices of the basal ganglia, which is in accordance with the fact that MMD is characterized by luminal stenosis or occlusion of the bilateral intracranial internal carotid arteries (ICAs) and collateral network at the base of the brain.[24,25] In contrast, the hot spots only focused on the image areas where stenotic vessels and their branches were located in ASDs, and the decreased number of their distal branches on image slices was one of the main features of ASDs to be identified by CNN. In addition, our model performed best in detecting MMD for binary classification in internal test datasets. However, the model identified NC best in the external datasets. The differences in MR scanners and imaging protocol settings between the internal and external datasets might have resulted in inconsistent results. In future studies, analyzing MRA images from more institutes would be helpful in increasing the diagnostic accuracy of the CNN models.

Our model proved to be almost the same as that of human radiologists in detecting MMD on MRA. In recent years, radiologists worldwide have been reading increasing numbers of medical images with ever-increasing complexity, which brings a large burden to modern healthcare systems.[26,27] Furthermore, even experienced radiologists are subject to human limitations, including fatigue, perceptual biases, and cognitive biases, all of which can lead to errors.[28,29] These forces strongly motivate leveraging deep learning to perform tasks in medical imaging and, as demonstrated here, could be used to detect MMD automatically via worklist prioritization, allowing the sickest patients to receive quicker diagnosis treatment. Fortunately, our 3D convolution model for the automatic diagnosis of MMD on MRA images achieved almost the same levels as those of human radiologists. Furthermore, the model was tested on patient data from two hospital systems that used different MR scanners and imaging protocol settings. In addition, the instability of image interpretation is one of the disadvantages of manual interpretation. Although good agreement was observed, there was still inconsistency between the two human radiologists. We found that inconsistent interpretation mainly existed in identifying MMD and ASD, especially for MMD patients in early and late stages of Suzuki, which may be misdiagnosed as ASD patients due to the absence of smoky blood vessels. Anyway, compared to manual interpretation methods, artificial intelligence models have higher stability in MMD diagnosis.

We aimed to develop an artificial technique for detecting MMD from the original MRA images of MRA, some practical difficulties were observed. First, the moyamoya vessels in MMD are easier to identify. However, because of the absence or decrease in moyamoya vessels,[30] it may be difficult to distinguish MMD from ASD in the early and late Suzuki stages. Next, we will further evaluate the role of CNN artificial intelligence technology in the different Suzuki stages of MMD with a larger sample size. Second, a deep learning model was developed and trained on data from one institution, and its performance was sustained in a new dataset from another institution. However, a universally generalizable performance is not known. Multicenter studies were warranted to validate the value of CNN artificial intelligence technology in the automatic diagnosis of MMD using different imaging scanners among different institutions. Thirdly, the lack of interpretability and transparency of neural networks in decision -making processes and learned features is another limitation of the study.

In conclusion, this study highlights the efficacy of the CNN model in the automated diagnosis of MMD using MRA images. The CNN-based deep learning models enhances the accuracy and efficiency of MMD detection, easing the workload on radiologists and promising integration into clinical workflows.

Contributors

Mingming Lu, Jiayi Zheng and Shitong Liu accessed and verified the data. Concept and design: Mingming Lu and Jianming Cai. Acquisition, analysis, or interpretation of data: Yijia Zheng, Xiaolan Zhang, Jiahui Lv, Cong Han, Chune Ma, and Chao Zheng. Drafting of the manuscript: Mingming Lu and Yijia Zheng. Critical review of the manuscript for important intellectual content: Yuan Liu, Baobao Li, Fei Yuan, and Peng Peng. Statistical analysis: Hongtao Zhang. Supervision: Fei Yuan, and Jianming Cai.

Data sharing statement

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

Declaration of interests

All coauthors report no conflict of interest.

Acknowledgements

This study was supported by the grants of National Natural Science Foundation of China (82371948 and 82001774), Tianjin Science and Technology Project (TJWJ2021MS043), Beijing Natural Science Foundation (7212100).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102888.

References

1 Ihara M, Yamamoto Y, Hattori Y, et al. Moyamoya disease: diagnosis and interventions. *Lancet Neurol*. 2022;21(8):747–758. https://doi.org/10.1016/S1474-4422(22)00165-X.

2 Boulouis G, Blauwblomme T, Hak JF, et al. Nontraumatic pediatric intracerebral hemorrhage. *Stroke*. 2019;50(12):3654–3661. https://doi.org/10.1161/STROKEAHA.119.025783.

3 Kaseka ML, Slim M, Muthusami P, et al. Distinct clinical and radiographic phenotypes in pediatric patients with moyamoya. *Pediatr Neurol*. 2021;120:18–26. https://doi.org/10.1016/j.pediatrneurol.2021.03.002.

4 Rafay MF, Smith SE, Dirks P, Armstrong D, deVeber GA. Hemorrhage predisposing to cerebral infarction in children with moyamoya disease. *Pediatr Neurol*. 2006;34(5):400–404. https://doi.org/10.1016/j.pediatrneurol.2005.10.003.

5 Guo Q, Pei S, Wang QN, et al. Risk factors for preoperative cerebral infarction in infants with moyamoya disease. *Transl Stroke Res*. 2023. https://doi.org/10.1007/s12975-023-01167-z.

6 Hayashi T, Kimiwada T, Karibe H, et al. Preoperative risks of cerebral infarction in pediatric moyamoya disease. *Stroke*. 2021;52(7):2302–2310. https://doi.org/10.1161/STROKEAHA.120.032699.

7 Wang J, Sui X, Zhao R, et al. Value of deep learning reconstruction of chest low-dose CT for image quality improvement and lung parenchyma assessment on lung window. *Eur Radiol*. 2024;34(2):1053–1064. https://doi.org/10.1007/s00330-023-10087-3.

8 Mu D, Bai J, Chen W, et al. Calcium scoring at coronary CT angiography using deep learning. *Radiology*. 2022;302(2):309–316. https://doi.org/10.1148/radiol.2021211483.

9 Xie Y, Tao H, Li X, et al. Prospective comparison of standard and deep learning-reconstructed turbo spin-echo MRI of the shoulder. *Radiology*. 2024;310(1):e231405. https://doi.org/10.1148/radiol.231405.

10 Claudio Quiros A, Coudray N, Yeaton A, et al. Mapping the landscape of histomorphological cancer phenotypes using self-supervised learning on unannotated pathology slides. *Nat Commun*. 2024;15(1):4596. https://doi.org/10.1038/s41467-024-48666-7.

11 Hoshino H, Izawa Y, Suzuki N, Research Committee on Moyamoya Disease. Epidemiological features of moyamoya disease in Japan. *Neurol Med Chir*. 2012;52(5):295–298. https://doi.org/10.2176/nmc.52.295.

12 Kim SJ, Schneider DJ, Feldmann E, Liebeskind DS. Intracranial atherosclerosis: review of imaging features and advances in diagnostics. *Int J Stroke*. 2022;17(6):599–607. https://doi.org/10.1177/17474930211066427.

13 Chen G, Yifang B, Jiajun Z, et al. Automated unruptured cerebral aneurysms detection in TOF MR angiography images using dual-channel SE-3D UNet: a multi-center research. *Eur Radiol*. 2023;33(5):3532–3543. https://doi.org/10.1007/s00330-022-09385-z.

14 Abdullah-Al-Wadud M, Kabir MH, Akber Dewan MA, Chae O. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans Consum Electron*. 2007;53(2):593–600. https://doi.org/10.1109/TCE.2007.381734.

15 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. https://arxiv.org/pdf/1512.03385.

16 Huang G, Liu Z, Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. https://arxiv.org/pdf/1608.06993.pdf.

17 Zhu Y, Newsam S. Densenet for dense flow. In: *2017 IEEE international conference on image processing (ICIP)*. IEEE; 2017. https://arxiv.org/pdf/1707.06316.pdf.

18 Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C]. In: *Proceedings of the IEEE international conference on computer vision*. 2017:618–626.

19 Kim T, Heo J, Jang DK, et al. Machine learning for detecting moyamoya disease in plain skull radiography using a convolutional neural network. *eBioMedicine*. 2019;40:636–642. https://doi.org/10.1016/j.ebiom.2018.12.043.

20 Akiyama Y, Mikami T, Mikuni N. Deep learning-based approach for the diagnosis of moyamoya disease. *J Stroke Cerebrovasc Dis*. 2020;29(12):105322. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105322.

21 Hao X, Xu L, Liu Y, et al. Construction of diagnosis model of moyamoya disease based on convolution neural network algorithm. *Comput Math Methods Med*. 2022;2022:4007925. https://doi.org/10.1155/2022/4007925.

22 Song Z, Zhang W, Jiang Q, et al. Artificial intelligence-aided detection for prostate cancer with multimodal routine health check-up data: an Asian multi-center study. *Int J Surg*. 2023;109(12):3848–3860. https://doi.org/10.1097/JS9.0000000000000862.

23 Huang SC, Kothari T, Banerjee I, et al. PENet-a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. *NPJ Digit Med*. 2020;3:61. https://doi.org/10.1038/s41746-020-0266-y.

24 Tigchelaar SS, Wang AR, Vaca SD, Li Y, Steinberg GK. Incidence and outcomes of posterior circulation involvement in moyamoya disease. *Stroke*. 2024;55(5):1254–1260. https://doi.org/10.1161/STROKEAHA.123.044693.

25 Kim DY, Son JP, Yeon JY, et al. Infarct pattern and collateral status in adult moyamoya disease: a multimodal magnetic resonance imaging study. *Stroke*. 2017;48(1):111–116. https://doi.org/10.1161/STROKEAHA.116.014529.

26 Zheng D, He X, Jing J. Overview of artificial intelligence in breast cancer medical imaging. *J Clin Med*. 2023;12(2):419. https://doi.org/10.3390/jcm12020419.

27 Zhou L, Yu L, Wang L. RECIST-induced reliable learning: geometry-driven label propagation for universal lesion segmentation. *IEEE Trans Med Imaging*. 2024;43(1):149–161. https://doi.org/10.1109/TMI.2023.3294824.

28 Yoon SY, Lee KS, Bezuidenhout AF, Kruskal JB. Spectrum of cognitive biases in diagnostic radiology. *Radiographics*. 2024;44(7):e230059. https://doi.org/10.1148/rg.230059.

29 Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am J Roentgenol*. 2013;201(3):611–617. https://doi.org/10.2214/AJR.12.10375.

30 Ge P, Zhang Q, Ye X, et al. Long-term outcome after conservative treatment and direct bypass surgery of moyamoya disease at late Suzuki stage. *World Neurosurg*. 2017;103:283–290. https://doi.org/10.1016/j.wneu.2017.03.101.