AJNR

Automated detection of steno-occlusive lesion on time-of-flight magnetic resonance angiography: an observer performance study

Hunjong Lim, Dongjun Choi, Leonard Sunwoo, Jae Hyeop Jung, Sung Hyun Baik, Se Jin Cho, Jinhee Jang, Tackeun Kim and Kyong Joon Lee

AJNR Am J Neuroradiol published online 7 May 2024
http://www.ajnr.org/content/early/2024/05/07/ajnr.A8334

Published May 7, 2024 as 10.3174/ajnr.A8334

ORIGINAL RESEARCH

NEUROVASCULAR/STROKE IMAGING

Automated detection of steno-occlusive lesion on time-of-flight magnetic resonance angiography: an observer performance study

Hunjong Lim, Dongjun Choi, Leonard Sunwoo, Jae Hyeop Jung, Sung Hyun Baik, Se Jin Cho, Jinhee Jang, Tackeun Kim, Kyong Joon Lee

ABSTRACT

BACKGROUND AND PURPOSE: Intracranial steno-occlusive lesions are responsible for acute ischemic stroke. However, the clinical benefits of artificial intelligence-based methods for detecting pathologic lesions in intracranial arteries have not been evaluated. We aimed to validate the clinical utility of an artificial intelligence model for detecting steno-occlusive lesions in the intracranial arteries.

MATERIALS AND METHODS: Overall, 138 TOF-MRA images were collected from two institutions, which served as internal (n = 62) and external (n = 76) test sets, respectively. Each study was reviewed by five radiologists (two neuroradiologists and three radiology residents) to compare the usage and non-usage of our proposed artificial intelligence model for TOF-MRA interpretation. They identified the steno-occlusive lesions and recorded their reading time. Observer performance was assessed using the area under the Jackknife free-response receiver operating characteristic curve and reading time for comparison.

RESULTS: The average area under the Jackknife free-response receiver operating characteristic curve for the five radiologists demonstrated an improvement from 0.70 without artificial intelligence to 0.76 with artificial intelligence (*P* = .027). Notably, this improvement was most pronounced among the three radiology residents, whose performance metrics increased from 0.68 to 0.76 (P = .002). Despite an increased reading time upon using artificial intelligence, there was no significant change among the readings by radiology residents. Moreover, the use of artificial intelligence resulted in improved inter-observer agreement among the reviewers (the intraclass correlation coefficient increased from 0.734 to 0.752).

CONCLUSIONS: Our proposed artificial intelligence model offers a supportive tool for radiologists, potentially enhancing the accuracy of detecting intracranial steno-occlusion lesions on TOF-MRA. Less-experienced readers may benefit the most from this model.

ABBREVIATIONS: AI = Artificial intelligence; AUC = Area under the receiver operating characteristic curve; AUFROC = Area under the Jackknife free-response receiver operating characteristic curve; DL = Deep learning; ICC = Intraclass correlation coefficient; IRB = Institutional Review Boards; JAFROC = Jackknife free-response receiver operating characteristic.

Received month day, year; accepted after revision month day, year.
From the Department of Radiology (H.L., L.S., J.H.J., S.H.B., S.J.C., K.J.L.), Center for Artificial Intelligence in Healthcare (L.S.), Seoul National University Bundang Hospital, Seongnam, Korea; lululab Inc. (D.C.), Seoul, Korea; Remote Reading Team (J.H.J.), Korea Armed Forces Capital Hospital, Seongnam, Korea; Department of Radiology (J.J.), Seoul St. Mary's Hospital, Seoul, Korea; TALOS Corp. (T.K.), Seoul, Korea; Monitor Corp. (K.J.L.), Seoul, Korea.

This research was funded by the SNUBH Research Fund (No. 14-2023-0014).

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dongjun Choi is currently an employee of lululab Inc., Seoul, Korea. Leonard Sunwoo is currently employed part-time at JLK Inc., Seoul, Korea. Tackeun Kim reports a relationship with TALOS Corp., Seoul, Korea, that includes: equity or stocks. Kyong Joon Lee reports a relationship with Monitor Corporation, Seoul, Korea, that includes: equity or stocks. Dongjun Choi, Leonard Sunwoo, Tackeun Kim, Kyong Joon Lee are listed as inventors on a patent related to the work. The patentee is Seoul National University Hospital, Seoul, Korea.

Please address correspondence to Leonard Sunwoo, MD, PhD, Department of Radiology, Seoul National University Bundang Hospital, 82 Gumi-ro 173 Beon-gil, Bundang-gu, Seongnam, 13620, Korea; e-mail: leonard.sunwoo@gmail.com.

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies have utilized deep learning algorithms to detect intracranial steno-occlusive lesions, leveraging semi- or fully automated techniques and image reconstruction methods. Despite advancements, accurate detection and localization remain challenging due to the complex nature of intracranial arteries and limitations in existing methods, such as the time-consuming extraction of multiple arteries and the inability to accurately measure the width of occluded lesions. Our recently

proposed approach integrates classification and localization within established medical imaging networks, aiming to overcome these challenges by simultaneously segmenting blood vessels and detecting lesions without extensive image reconstruction or patch-based analysis.

KEY FINDINGS: The use of our artificial intelligence model improved the detection accuracy of intracranial steno-occlusive lesions on TOF-MRA, with an improvement from 0.70 to 0.76 in the AUFROC for radiologists. Radiology residents, in particular, benefited significantly from AI assistance, highlighting its potential to enhance diagnostic accuracy.

KNOWLEDGE ADVANCEMENT: Our study advances knowledge by demonstrating the clinical utility of an AI model in improving radiologists' accuracy in detecting intracranial steno-occlusive lesions. This suggests that AI can be a valuable support tool, especially for less-experienced readers, potentially increasing diagnostic performance and contributing to better patient outcomes.

INTRODUCTION

Acute ischemic stroke is the second leading cause of death and a major cause of disability worldwide.[1, 2] One of the primary underlying factors responsible for ischemic stroke is intracranial steno-occlusive lesions.[3-5] Thus, the prompt and precise identification of steno-occlusive lesions is of paramount importance in diagnosing patients with ischemic stroke and in selecting appropriate therapeutic strategies.[6, 7]

TOF-MRA is one commonly used non-invasive imaging technique for evaluating intracranial arteries.[6] However, accurate detection and the precise localization of steno-occlusive lesions present a challenge because of the intricate shapes of intracranial arteries. Meticulous evaluation requires considerable time and effort, which leads to an increased workload and the subsequent risk of detection failure.

Methods using deep learning (DL) algorithms have emerged to detect steno-occlusive lesions automatically in intracranial arteries.[6, 8-10] Previous approaches have utilized semi- or fully automated labeling and techniques such as straightened multi-planar reformatted (MPR) images, along with extracting blood vessels in advance of measuring the width, aiming to detect stenoses. Despite advancements, the detection of steno-occlusive lesions in intracranial arteries remains challenging due to the time-consuming nature of extracting multiple arteries with MPR, loss of vascular bifurcation, and inability to measure the width of occluded lesions even with pre-extracted blood vessels. Our novel approach leverages multi-task learning to segment blood vessels and detect lesions simultaneously, without the need for extensive image reconstruction or patch-based analysis.[10] This method aids in enhancing lesion detection efficiency by integrating classification and localization modules, thus offering a comprehensive solution to the limitations identified in prior studies. However, the clinical benefits of such artificial intelligence (AI) methods have not yet been evaluated sufficiently.

Therefore, we aimed to investigate the potential benefits and limitations of an AI-based model to aid radiologists in detecting steno-occlusive lesions. Specifically, we assessed the model's lesion detection accuracy and its impact on interpretation time compared to those of conventional methods.

MATERIALS AND METHODS

This multicenter retrospective study was approved by the Institutional Review Boards (IRBs) of the Seoul National University Bundang Hospital (SNUBH) and Seoul St. Mary's Hospital (SSMH), both of which waived the requirement for informed consent (IRB Nos.: B-2204-753-106 and KC20RIDI0197, respectively).

Study Cohort

Sixty-two individuals were collected from the SNUBH database between October 2014 and August 2019 as an internal test set, including 30 with intracranial stenosis or occlusion (stenosis group) and 32 without steno-occlusive lesions (healthy group). The inclusion criteria for the stenosis group were as follows: 1) age >18 years, 2) TOF-MRA and DSA performance within 1 month's interval, and 3) moderate or severe degree stenosis according to the DSA report (>50% stenosis using the Warfarin vs. Aspirin for Symptomatic Intracranial Disease method as follows: % stenosis = $(1-[D_{stenosis}/D_{normal}]) \times 100$). The inclusion criteria for the healthy group were as follows: 1) age >18 years, 2) TOF-MRA performance, and 3) normal MRA findings according to the radiologic report.

In addition, 76 individuals were collected as an external test set from the SSMH database from January 2016 to December 2019, comprising 30 with intracranial stenosis or occlusion (stenosis group) and 46 without steno-occlusive lesions (healthy group). The inclusion criteria for the SSMH cohort were as follows: 1) age >18 years and 2) TOF-MRA results.

Patients in the stenosis group were randomly selected. For the healthy group, the patients were randomly chosen among individuals who underwent brain MRI as part of health screenings and received normal MRA reports. In total, we collected 138 individuals from two institutions, with 60 in the stenosis group and 78 in the healthy group.

TOF-MRA Acquisition

For SNUBH, three-dimensional TOF-MRA examinations were performed using a 1.5-T (Amira, Siemens Healthcare, Erlangen, Germany; or Intera, Philips Healthcare, Best, Netherlands) or 3.0-T scanner (Achieva or Ingenia, Philips Healthcare). The scan parameters were as follows: repetition time (TR), 20 ms to 27 ms; echo time (TE), 3.45 ms to 7.15 ms; flip angle (FA), 18° to 25°; field of view (FOV), 132 mm to 230 mm; section thickness, 0.5 mm to 1.6 mm; matrix, 256 to 704×163 to 360.

For SSMH, 3D TOF-MRA was performed using a 1.5-T (Avanto, Siemens Healthcare; Achieva, Philips Healthcare) or 3.0-T (Verio or Vida, Siemens Healthcare; Ingenia, Philips Healthcare) scanner. The scan parameters were as follows: TR, 17.9 ms to 25 ms; TE, 3.5 ms to 7 ms; FA, 18° to 23°; FOV, 170 mm to 240 mm; section thickness, 0.4 mm to 1.2 mm; matrix, 384 to 512×214 to 331.

AI Model

We used a DL algorithm to detect steno-occlusive lesions using traces of intracranial arteries (Figure 1). Our model utilizes an image segmentation model such as U-Net[11] as a backbone, augmented by additional modules to detect steno-occlusive lesions. Specifically, we designed a backbone model termed Spider U-Net, which is a modified version of U-Net that adds a long short-term memory network. The vessel segmentation performance of Spider U-Net outperformed that of U-Net.[12] A multitask learning method based on Spider U-Net demonstrated that the detection performance of steno-occlusive lesions while extracting blood vessels was higher than that while detecting steno-occlusive lesions without extracting blood vessels.[10] The model used a training set from SNUBH similar to that used previously and demonstrated superior overall detection performance compared to other models.[10] Details of the model have been published elsewhere.[10, 12] Code details can be accessed through the following link: https://github.com/djchoi1742/MRA_ICAD.

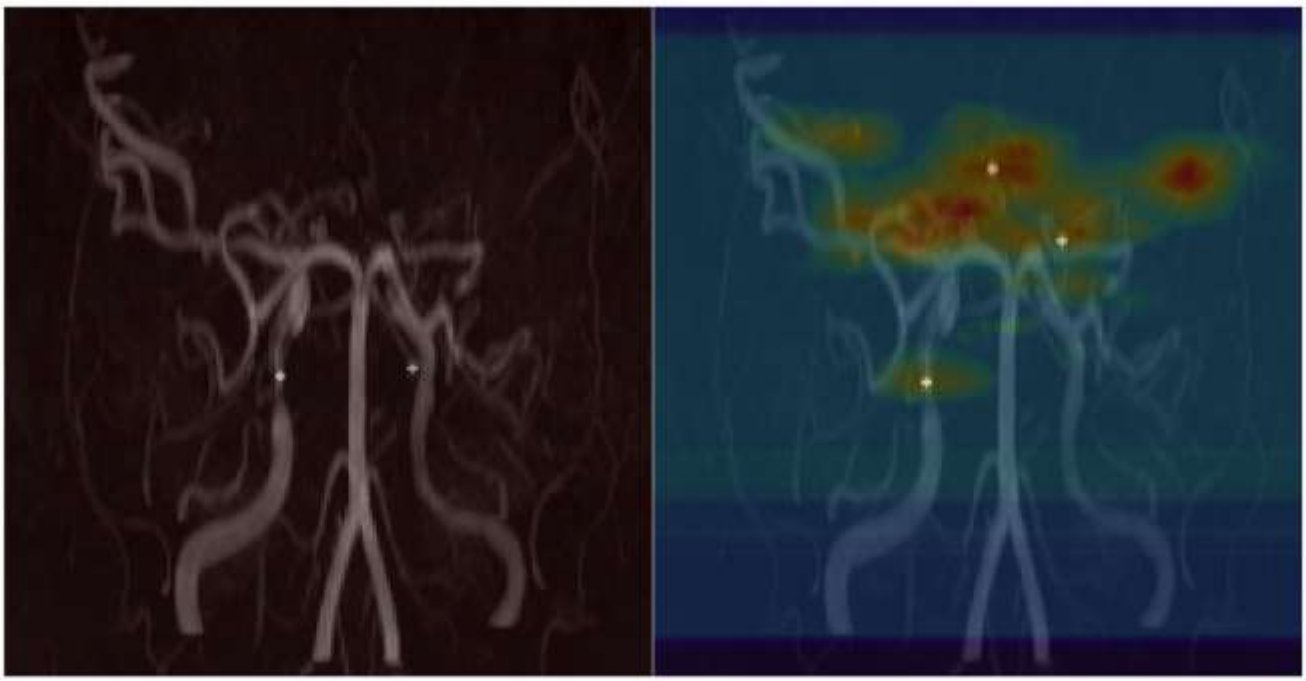

FIG 1. Representative image of our proposed model. (a) Initial TOF-MRA shows multiple steno-occlusive lesions, including left middle cerebral artery occlusion. (b) Our proposed artificial intelligence (AI) model identified steno-occlusive lesions, shown as red and yellow markings on the heat map.

Image Interpretation and Observer Performance Study

A neuroradiologist (L.S., with 13 years of clinical experience), a board-certified radiologist (H.L., with 5 years of clinical experience), and a neurosurgeon (T.K., with 13 years of clinical experience) thoroughly reviewed all TOF-MRA examinations and accessible DSA studies, establishing a reference standard by consensus for the number and location of steno-occlusive lesions. The reference standards for steno-occlusive lesion detection were confined to the distal ICA, A1-2 segments of ACA, M1-2 segments of MCA, P1-2 segments of PCA, V4 segments of the vertebral artery, and the basilar artery.

Five radiologists, including two neuroradiologists (S.H.B. and S.J.C., with 10 and 9 years of experience in neuroradiology, respectively) and three radiology residents (J.H.J., H.C., and H.U.C., with 4, 3, and 3 years of clinical experience, respectively), participated as observers. The radiologists who determined the reference standard did not participate. Each reviewer conducted two separate assessments of all TOF-MRA studies (n = 138) across two sessions. The studies were randomly divided into two blocks (Block A and Block B, each containing 69 studies). During the first session, studies in Block A were reviewed with AI assistance, and those in Block B were reviewed without AI. To mitigate bias and memory recall effects, a compulsory 4-week washout period was placed between the sessions. After this interval, the review conditions were swapped: during the second session, studies initially reviewed with AI (Block A) were assessed without AI, and vice versa for Block B (Figure 2). The reviewers were blinded to the patient information and reference standards. They were instructed

to assess whether there is stenosis or occlusion in the intracranial artery using a commercial picture archiving and communication system (PACS) (RadiAnt viewer, Medixant, Poland). Upon conducting an AI-assisted review, the heat map generated by the AI model was superimposed onto the maximum intensity projection (MIP) TOF-MRA images. The heatmap images were reviewed using an additional viewer (Windows Photo Viewer, Microsoft, Redmond, WA, USA). The participants were instructed to mark a stenotic or occlusive lesion by recording the coordinates of the central portion of the lesion. The confidence rating for each lesion ranged from 1 (very uncertain) to 5 (absolutely certain). The overall reading time was recorded for each case. During the review, they were encouraged to evaluate both the source and MIP TOF-MRA images.

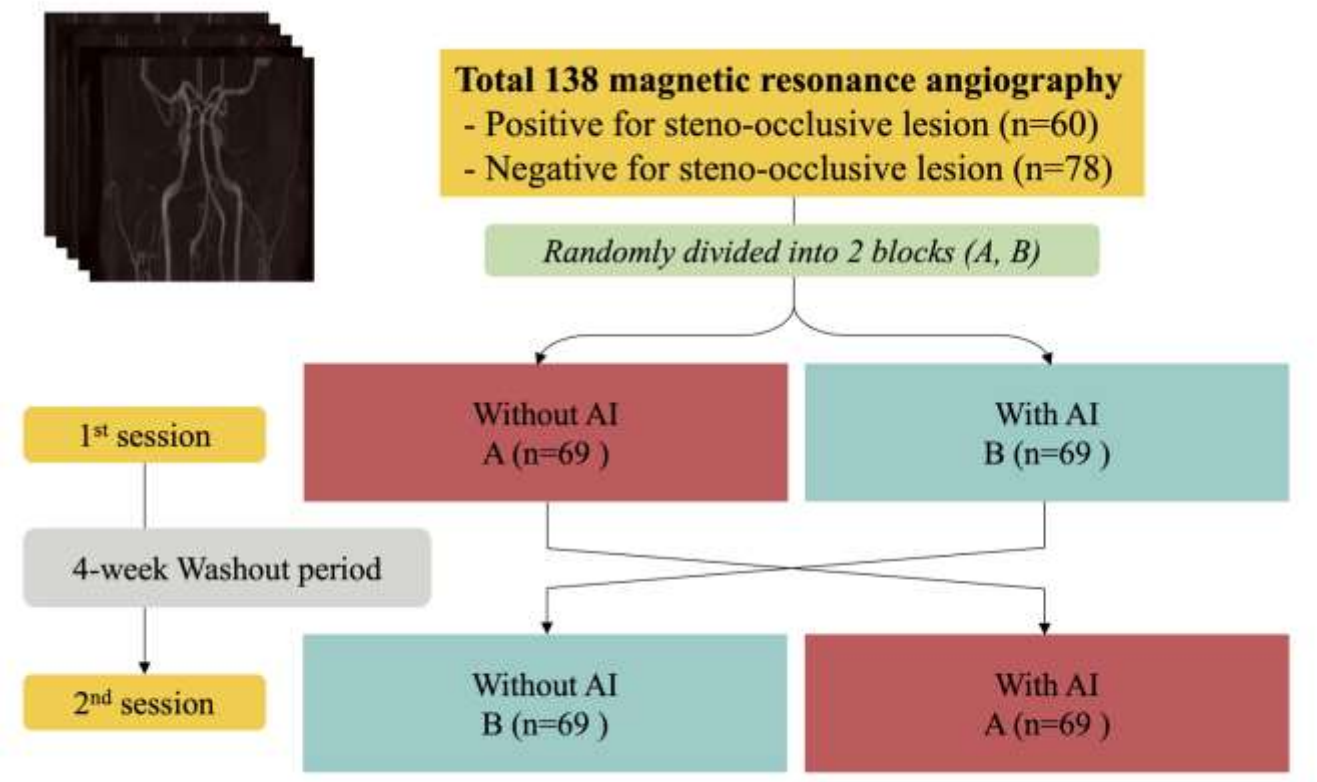

FIG 2. A diagram illustrating the observer performance study. Each observer conducted two separate TOF-MRA reviews across two reading sessions; one without AI and another with AI. There was a washout period of 4 weeks or longer between these sessions.

The observers' performances were stratified into neuroradiologist and radiology resident groups and were combined eventually. An overlap between the lesions marked by the observer (indicated by the coordinates) and reference standard (annotated across the entire length) was classified as a true-positive finding; else, it was classified as a false-positive finding.

Statistical Analysis

We performed a Jackknife free-response receiver operating characteristic (JAFROC) analysis to evaluate the reviewers' localization performance on a per-lesion basis.[13] The area under the JAFROC (AUFROC) indicates the probability that the lesion rating marked in the diseased case is greater than the highest rating in the healthy case. We calculated the AUFROC according to AI use and performed a comparison test using the Dorfman-Berbaum-Metz method.[14] Comparisons based on AI use were calculated as fixed-reader random cases; for pooled reviewers, they were calculated as random-reader random cases. We computed the area under the receiver operating characteristic curve (AUC) to measure the performance on a per-patient basis. We calculated the average of the reviewers' ratings per patient. Similar to JAFROC, we performed a comparison test of the AUC using identical method for each reviewer and pooled reviewers.

We calculated the sensitivity and specificity to measure the diagnostic accuracy of the reviewers. The sensitivity was calculated for each lesion and per patient. Sensitivity per lesion was calculated under a cut-off of the sum of the 1 – false-positive fraction, and the lesion localization fraction was maximized. The sensitivity per patient was calculated using an optimal cut-off to maximize Youden's J statistic.[15] Specificity was calculated for each patient.

We used the intraclass correlation coefficient (ICC)[16] to measure interobserver agreement among the reviewers. The ICC values < 0.5, 0.5 to 0.75, 0.75 to 0.9, or ≥ 0.9 indicated poor, moderate, good, or excellent agreement, respectively.[17] We calculated the ICC based on AI use. All statistical analyses were performed using the R statistical software version 3.6.3 (The R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $P < 0.05$.

RESULTS

Patients

Table 1 summarizes the demographic and lesion characteristics of the study cohort. The median age was 58 years (range, 28–84 years), and the male-to-female ratio was 1:1. Sixty patients in the stenosis group had 115 steno-occlusive lesions (1.92 lesions per patient). In the internal test set, steno-occlusive lesions were caused by atherosclerosis (n = 26), Moyamoya disease (n = 3), and dissection (n = 1), and in the external test set, by atherosclerosis (n = 27) and Moyamoya disease (n = 3). The patients predominantly had a single lesion, accounting for 55% of the cases (33/60). The lesion distribution did not differ between the right and left sides (51 vs. 61, P = .345). We recorded 91 (79.1%) and 24 (20.9%) lesions in the anterior and posterior circulations, respectively. In the anterior circulation, they were predominantly located in the middle cerebral artery, accounting for 53.8% (49/91) of the total lesions. No significant difference was observed in the patient ratios of 1.5T (n = 12) compared to 3T (n = 48) between the two test sets (P = 0.748).

Observer Performance Assessments

Table 2 summarizes results of the per-lesion and per-patient analyses. In the per-lesion analysis, AI-assisted review exhibited a higher pooled AUFROC (0.76; 95% CI, 0.67, 0.85) than non-AI-assisted review (0.70; 95% CI: 0.56, 0.81) (P = .027). The pooled AUFROC for all residents increased from 0.68 (95% CI: 0.56, 0.77) to 0.76 (95% CI: 0.67, 0.85) (P = .002), whereas that for all neuroradiologists did not demonstrate a statistically significant increase. In terms of per-patient analysis, the AUC for all reviewers improved marginally, but did not reach statistical significance. Figure 3 illustrates the pooled JAFROC curves for all reviewers, neuroradiologists, and radiology residents.

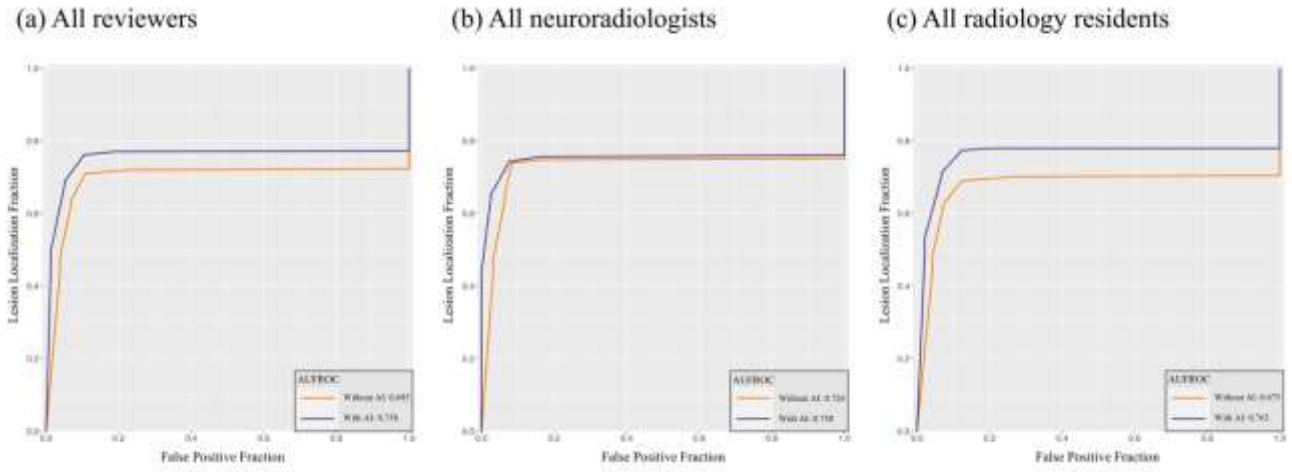

FIG 3. Pooled Jackknife free-response receiver operating characteristic (JAFROC) curves of all reviewers (a), neuroradiologists (b), and radiology residents (c). With AI, the area under the pooled JAFROC (AUFROC) for all reviewers improved significantly from 0.70 to 0.76 (P = .027). Similarly, the use of AI improved AUFROC for radiology residents from 0.68 to 0.76 (P = .002). The AUFROC for neuroradiologists did not show statistical difference between the results without and with AI.

Table 3 presents the results in the internal (SNUBH) and external (SSMH) test sets. Notably, all resident groups in both test sets exhibit significant differences between AUFROCs with and without AI. From a lesion-based perspective, the sensitivity of the four reviewers (Reviewers 2, 3, 4, and 5) improved upon using AI (Table 4). However, we observed a marginal reduction in the sensitivity reported by Reviewer 1 (with AI: 80.9% vs. without AI: 80.0%). Supplementary Table 1 presents a comparison of the lesion detection sensitivity with and without AI, based on the lesion number and circulation type. Overall, AI use led to a marginal improvement in sensitivity across the subgroups. Notably, with AI-assistance, an increased number of lesions correlated with a greater increase in sensitivity. For example, in cases with more than two lesions, the sensitivity improved from 62.8% to 73.7% with AI. Similar trends were observed in both the anterior and posterior circulation groups. The interobserver agreement among the reviewers was increased from 0.734 (95% CI: 0.670, 0.792) without AI to 0.752 (95% CI: 0.693, 0.805) with AI (Supplementary Table 2).

The average reading time for all five reviewers was longer in AI readings than in non-AI readings (71.8 ± 37.0 s with AI vs. 63.5 ± 31.7 s without AI, P = .044). Specifically, AI use increased the reading time for neuroradiologists (45.4 ± 19.5 s with AI vs. 37.6 ± 16.5 s without AI, P <.001). However, the reading time for radiology residents did not demonstrate a statistically significant difference (89.5 ± 49.9 s with AI vs. 80.7 ± 43.0 s without AI, P = .118) (Table 5).

DISCUSSION

In this study, we assessed the impact of AI on observer performance in detecting steno-occlusive lesions in the intracranial arteries using data from two separate institutions. We found that the pooled AUFROC for the five radiologists demonstrated improvement, increasing from 0.70 without AI to 0.76 with AI (P = .027). This improvement was particularly pronounced among radiology residents (AUFROC improved from 0.675 to 0.763, P=.002). For the neuroradiologists, there was a trend toward improvement with the AUFROC increasing from 0.726 to 0.750; however, this change was not statistically significant (P > .05). The average reading time of the five reviewers was slightly longer when using AI assistance than that without AI. However, this difference was not statistically significant when the analysis was limited to radiology residents.

The Stroke Outcomes and Neuroimaging of Intracranial Atherosclerosis trial suggested that TOF-MRA for evaluating intracranial artery stenosis has a relatively lower positive predictive value than DSA.[18] In line with a previous study, our findings demonstrated low sensitivity of observers without AI-assistance in detecting intracranial stenosis. DSA is the gold standard for evaluating intracranial artery stenosis; however, it poses risks of radiation, nephrotoxicity caused by iodinated contrast agents, and thromboembolic complications. Therefore, the accurate and reliable assessment of intracranial steno-occlusive lesions using noninvasive angiography is essential. Our results suggest that AI use can improve the accuracy of detecting intracranial steno-occlusive lesion by radiologists upon evaluating TOF-MRA.

The performance of the radiology residents significantly improved with the use of AI. The AUFROC calculated with AI assistance by radiology residents was comparable to that of neuroradiologists. Moreover, unlike that of the neuroradiologists, the reading time for radiology residents did not extend between the reading sessions with and without the use of AI. Specifically, their average reading time for the healthy group remained unchanged, while for the stenosis group, they allocated more time when using AI. In contrast, neuroradiologists experienced increased reading times with AI for both stenosis and healthy groups. Thus, the use of AI assistance has greater utility for relatively less experienced observers, such as radiology residents or specialists from other fields, highlighting its utility in supporting diagnostic accuracy without compromising efficiency.

Radiologists encounter diagnostic errors that include visual perception and cognitive errors.[17, 19] The “satisfaction of search” is a prevalent cognitive error, which signifies halting visual exploration upon identifying an initial abnormality during image interpretation.[19] Without AI, the sensitivity for detecting steno-occlusive lesions decreased with increased number of lesions. However, AI use helped maintain the sensitivity, despite statistical insignificance (Supplementary Table 1). Integrating AI can enhance the radiologists’ performance by sustaining vigilance.

Furthermore, we observed an improvement in the interobserver agreement among the reviewers, increasing from a moderate to a good level (ICC increased from 0.734 to 0.752). This improvement could potentially elevate the level of consensus between radiologists, thereby alleviating the interobserver variability associated with MRA in determining intracranial stenosis. Similarly, Lin et al. demonstrated that AI use not only enhances accuracy but also reduces the interobserver variability in delineating nasopharyngeal carcinoma for radiation therapy.[20]

In our previous study, we validated the stand-alone performance of our algorithm exclusively in patients with steno-occlusive lesions, achieving an AUC of up to 0.874 and an AUFROC of up to 0.855.[10] To assess the clinical utility of our model, we conducted this observer performance study using a study cohort including both healthy individuals and patients with steno-occlusive lesions from two different institutions. Despite a slightly lower performance on the external test set compared to the internal test set, the significant improvement in residents' performance on both test sets underscores the generalizability of our AI model. In accordance with the recently published guidelines, the evidence of this study can be classified as Level 5A, signifying a retrospective study that integrates internal and external data for the purpose of concluding performance assessment.[21, 22]

Our study has several limitations. First, only patients in the SNUBH stenosis group underwent confirmatory DSA. The reference standard assessment for participants in the healthy group from SNUBH and all participants from SSMH was based on expert consensus. The small caliber of the intracranial arteries and the limited spatial resolution of TOF-MRA may have under- or overestimated the stenosis degree. However, the observers were instructed to grade each stenotic lesion using a 5-point Likert scale consistently. Thus, JAFROC analyses could mitigate potential calibration issues, where the optimal threshold was applied individually. Second, we measured the reading time to simulate a clinical reading session; however, the need for an additional viewer to detect the AI suggestions may have introduced a bias in accurate time measurement. The observed increase in reading time could be a natural consequence of the additional steps required. These aspects warrant future studies using software implemented into the in-hospital PACS software to eliminate the need for a separate viewer and more accurately reflect the impact of reading time. Third, the scope of our model was confined to the distal ICA, A1-2 segments of ACA, M1-2 segments of MCA, P1-2 segments of PCA, V4 segments of the vertebral artery, and the basilar artery. This focus was necessitated by the inherent spatial resolution limitations of TOF-MRA. In addition, the relatively small sample size limits the depth of subgroup analyses, such as diagnostic performance according to vascular subsegments. Consequently, this constraint may limit the utility of our model in detecting steno-occlusive lesions in the distal branches of intracranial arteries. A future study with a larger patient cohort may be needed to enable detailed subgroup analyses.

CONCLUSIONS

Our study suggests that the proposed AI model offers a supportive tool for radiologists, potentially enhancing the accuracy of detecting intracranial steno-occlusive lesions on TOF-MRA. Although the value for neuroradiologists may be limited, less-experienced readers may benefit from this model.

REFERENCES

1. Group GBDNDC. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Neurol* 2017;16:877-897. DOI: https://doi.org/10.1016/s1474-4422(17)30299-5

2. Katan M, Luft A. Global Burden of Stroke. *Semin Neurol* 2018;38:208-211. DOI: https://doi.org/10.1055/s-0038-1649503

3. Arenillas JF. Intracranial atherosclerosis: current concepts. *Stroke* 2011;42:S20-23. DOI: https://doi.org/10.1161/strokeaha.110.597278

4. Sacco RL, Kargman DE, Gu Q, Zamanillo MC. Race-ethnicity and determinants of intracranial atherosclerotic cerebral infarction. The Northern Manhattan Stroke Study. *Stroke* 1995;26:14-20. DOI: https://doi.org/10.1161/01.str.26.1.14

5. Wityk RJ, Lehman D, Klag M, et al. Race and sex differences in the distribution of cerebral atherosclerosis. *Stroke* 1996;27:1974-1980. DOI: https://doi.org/10.1161/01.str.27.11.1974

6. Choi CG, Lee DH, Lee JH, et al. Detection of intracranial atherosclerotic steno-occlusive disease with 3D time-of-flight magnetic resonance angiography with sensitivity encoding at 3T. *AJNR Am J Neuroradiol* 2007;28:439-446

7. Li Q, Tian CL, Yang YW, et al. Conventional T2-Weighted Imaging to Detect High-Grade Stenosis and Occlusion of Internal Carotid Artery, Vertebral Artery, and Basilar Artery. *J Stroke Cerebrovasc Dis* 2015;24:1591-1596. DOI: https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.03.028

8. Chen L, Mossa-Basha M, Balu N, et al. Development of a quantitative intracranial vascular features extraction tool on 3D MRA using semiautomated open-curve active contour vessel tracing. *Magn Reson Med* 2018;79:3229-3238. DOI: https://doi.org/10.1002/mrm.26961

9. Qiu J, Tan G, Lin Y, et al. Automated detection of intracranial artery stenosis and occlusion in magnetic resonance angiography: A preliminary study based on deep learning. *Magn Reson Imaging* 2022;94:105-111. DOI: https://doi.org/10.1016/j.mri.2022.09.006

10. Choi D, Kim T, Jang J, et al. Intracranial steno-occlusive lesion detection on time-of-flight MR angiography using multi-task learning. *Comput Med Imaging Graph* 2023;107:102220. DOI: https://doi.org/10.1016/j.compmedimag.2023.102220

11. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Cham: Springer International Publishing; 2015:234-241

12. Lee K, Sunwoo L, Kim T, Lee KJ. Spider U-Net: Incorporating Inter-Slice Connectivity Using LSTM for 3D Blood Vessel Segmentation. *Applied Sciences* 2021;11:2014. DOI: https://doi.org/10.3390/app11052014

13. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys* 2004;31:2313-2330. DOI: https://doi.org/10.1118/1.1769352

14. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 1992;27:723-731

15. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-35. DOI: https://doi.org/10.1002/1097-0142(1950)3:1%3C32::aid-cncr2820030106%3E3.0.co;2-3

16. Koch GG. Intraclass correlation coefficient. *Encyclopedia of statistical sciences* 2004

17. Benchoufi M, Matzner-Lober E, Molinari N, et al. Interobserver agreement issues in radiology. *Diagnostic and Interventional Imaging* 2020;101:639-641. DOI: https://doi.org/10.1016/j.diii.2020.09.001

18. Feldmann E, Wilterdink JL, Kosinski A, et al. The Stroke Outcomes and Neuroimaging of Intracranial Atherosclerosis (SONIA) trial. *Neurology* 2007;68:2099-2106. DOI: https://doi.org/10.1212/01.wnl.0000261488.05906.c1

19. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and System Factors Contributing to Diagnostic Errors in Radiology. *American Journal of Roentgenology* 2013;201:611-617. DOI: https://doi.org/10.2214/ajr.12.10375

20. Lin L, Dou Q, Jin Y-M, et al. Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. *Radiology* 2019;291:677-686. DOI: https://doi.org/10.1148/radiol.2019182012

21. Klontzas ME, Gatti AA, Tejani AS, Charles E. Kahn J. AI Reporting Guidelines: How to Select the Best One for Your Research. *Radiology: Artificial Intelligence* 2023;5:e230055. DOI: https://doi.org/10.1148/ryai.230055

22. Pham N, Hill V, Rauschecker A, et al. Critical Appraisal of Artificial Intelligence–Enabled Imaging Tools Using the Levels of Evidence System. *American Journal of Neuroradiology* 2023;44:E21-E28. DOI: https://doi.org/10.3174/ajnr.a7850

264
265 **Table 1:** Patient demographics and lesion characteristics.

	Positive for steno-occlusive lesion			Negative for steno-occlusive lesion	All
	Internal test set	External test set	Subtotal		
No. of patients*	30	30	60	78	138
Age, yr (range)	58 (28-78)	58 (38-80)	58 (28-80)	58 (29-84)	58 (28-84)
Sex (female ratio)	18 (0.60)	13 (0.43)	31 (0.52)	38 (0.49)	69 (0.50)
Proportion of 3T	0.83	0.77	0.80		
Underlying causes					
Atherosclerosis	26	27	53		
Moyamoya disease	3	3	6		
Dissection	1	0	1		
No. of lesions	67	48	115		
1	12	21	33		
2	9	3	12		
3	2	3	5		
4	5	3	8		
5	1	0	1		
6	1	0	1		
Lesion laterality					
Right	30	21	51		
Left	35	26	61		
BA	2	1	3		
Lesion segment					
Anterior	55	36	91		
ACA	15	2	17		
MCA	27	22	49		
Distal ICA	13	12	25		
Posterior	12	12	24		

VA	4	6	10
BA	2	1	3
PCA	6	5	11

Note. ICA, internal carotid artery; VA, vertebral artery; BA, basilar artery; ACA, anterior cerebral artery; MCA, middle cerebral artery; PCA, posterior cerebral artery

*****Numbers in parentheses are interquartile ranges or percentages.

Table 2: Diagnostic performance of reviewers.

	AUFROC (per-lesion)			AUC (per-patient)		
Reviewers	Without AI	With AI	*P*-value	Without AI	With AI	*P*-value
Neuroradiologists						
Reviewer 1	0.807	0.812	0.892	0.902	0.898	0.874
Reviewer 2	0.644	0.687	0.389	0.899	0.910	0.710
All neuroradiologists	0.726	0.750	0.423	0.900	0.904	0.779
Radiology residents						
Reviewer 3	0.555	0.675	0.023*	0.894	0.906	0.758
Reviewer 4	0.770	0.854	0.023*	0.894	0.942	0.081
Reviewer 5	0.700	0.759	0.218	0.928	0.919	0.716
All residents	0.675	0.763	0.002*	0.905	0.922	0.420
All reviewers	0.695	0.758	0.027*	0.903	0.915	0.364

Note: AUFROC, Area under the Jackknife free-response receiver operating characteristic curve; AUC, area under the receiver operating characteristic curve

An asterisk is added to the *P*-values at a significance level of 0.05.

Table 3: Diagnostic performance in the internal (SNUBH) and external (SSMH) test sets.

	AUFROC (per-lesion)					
	Internal test set			External test set		
Reviewers	Without AI	With AI	*P*-value	Without AI	With AI	*P*-value
All neuroradiologists	0.774	0.774	0.985	0.671	0.717	0.410
All residents	0.705	0.787	0.034*	0.637	0.732	0.033*
All reviewers	0.733	0.782	0.110	0.650	0.726	0.086

Note: AUFROC, Area under the Jackknife free-response receiver operating characteristic curve

An asterisk is added to the *P*-values at a significance level of 0.05.

Table 4: Results of diagnostic accuracy of reviewers.

	Sensitivity (per-lesion)			Sensitivity (per-patient)			Specificity		
Reviewers	Without AI	With AI	*P*-value	Without AI	With AI	*P*-value	Without AI	With AI	*P*-value
Neuroradiologists									
Reviewer 1	80.9% (93/115)	80.0% (92/115)	0.500	88.3% (53/60)	81.7% (49/60)	0.778	92.3% (72/78)	94.9% (74/78)	0.372
Reviewer 2	67.0% (77/115)	68.7% (79/115)	0.444	90.0% (54/60)	85.0% (51/60)	0.710	91.0% (71/78)	89.7% (70/78)	0.500
All neuroradiologists	73.9% (170/230)	74.8% (171/230)	0.500	90.0% (107/120)	83.3% (100/120)	0.746	91.7% (143/156)	92.3% (144/156)	0.500
Radiology residents									
Reviewer 3	55.7% (64/115)	68.7% (78/115)	0.028	81.7% (49/60)	86.7% (52/60)	0.308	96.2% (75/78)	93.6% (73/78)	0.642
Reviewer 4	75.7% (87/115)	86.1% (99/115)	0.033	90.0% (54/60)	95.0% (57/60)	0.244	87.2% (68/78)	88.5% (69/78)	0.500
Reviewer 5	67.0% (77/115)	73.0% (84/115)	0.194	88.3% (53/60)	81.7% (49/60)	0.778	92.3% (72/78)	93.6% (73/78)	0.500
All residents	66.1% (228/345)	75.9% (261/345)	0.067	86.7% (156/180)	87.8% (158/180)	0.500	91.9% (215/234)	91.9% (215/234)	0.500

All reviewers	69.2% (398/575)	75.3% (432/575)	0.188	87.7% (263/180)	86.0% (258/180)	0.500	91.8% (358/390)	92.1% (359/390)	0.500

Note. The count and number of target groups are shown in parentheses.

Table 5: Results of reviewer performance test (reading time).

	Reading time [seconds]		
	Without AI	With AI	*P*-value
Total	63.5±31.7	71.8±37.0	0.0442
Neuroradiologists	37.6±16.5	45.4±19.5	0.0004
Radiology residents	80.7±43.0	89.5±49.9	0.1179
Normal Group	45.3±17.9	50.3±20.1	0.1033
Neuroradiologists	28.8±9.8	34.7±9.6	0.0002
Radiology residents	56.3±24.4	60.7±27.9	0.2959
Disease Group	87.0±30.4	99.8±35.3	0.0353
Neuroradiologists	49.0±16.5	59.3±20.4	0.0031
Radiology residents	112.3±41.3	126.8±47.5	0.0770

SUPPLEMENTAL FILES

Supplementary Table 1. Subgroup analysis of diagnostic accuracy

	Sensitivity (per-lesion)	
Group	Without AI	With AI
Number of lesions		
1	68.8% (22/32)	70.0% (22/32)
≧2	69.4% (58/83)	77.3% (64/83)
≦2	75.5% (44/58)	76.9% (45/58)
≧3	62.8% (36/57)	73.7% (42/57)
Circulation		
Anterior	73.2% (67/91)	78.9% (72/91)
Posterior	54.2% (13/24)	61.7% (15/24)

Note. The count and number of target groups are shown in parentheses.

Supplementary Table 2. Interobserver variability in reviewers

	Without AI	With AI
ICC	0.734 (0.670–0.792)	0.752 (0.693–0.805)

Note. ICC: intraclass correlation coefficient. 95% confidence intervals are given in parentheses.