Original Article

Yonsei Med J 2021 Nov;62(11):1052-1061
https://doi.org/10.3349/ymj.2021.62.11.1052

Yonsei Medical Journal
YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

A Deep Learning Model with High Standalone Performance for Diagnosis of Unruptured Intracranial Aneurysm

Bio Joo[1], Hyun Seok Choi[2,3], Sung Soo Ahn[3], Jihoon Cha[3], So Yeon Won[3], Beomseok Sohn[3], Hwiyoung Kim[3], Kyunghwa Han[3], Hwa Pyung Kim[4], Jong Mun Choi[4], Sang Min Lee[4], Tae Gyu Kim[4], and Seung-Koo Lee[3]

[1]Department of Radiology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul;
[2]Department of Radiology, Seoul Medical Center, Seoul;
[3]Department of Radiology, Research Institute of Radiological Science and Center for Clinical Image Data Science, Yonsei University College of Medicine, Seoul;
[4]DEEPNOID, Seoul, Korea.

Purpose: This study aimed to investigate whether a deep learning model for automated detection of unruptured intracranial aneurysms on time-of-flight (TOF) magnetic resonance angiography (MRA) can achieve a target diagnostic performance comparable to that of human radiologists for approval from the Korean Ministry of Food and Drug Safety as an artificial intelligence-applied software.

Materials and Methods: In this single-center, retrospective, confirmatory clinical trial, the diagnostic performance of the model was evaluated in a predetermined test set. After sample size estimation, the test set consisted of 135 aneurysm-containing examinations with 168 intracranial aneurysms and 197 aneurysm-free examinations. The target sensitivity and specificity were set as 87% and 92%, respectively. The patient-wise sensitivity and specificity of the model were analyzed. Moreover, the lesion-wise sensitivity and false-positive detection rate per case were also investigated.

Results: The sensitivity and specificity of the model were 91.11% [95% confidence interval (CI): 84.99, 95.32] and 93.91% (95% CI: 89.60, 96.81), respectively, which met the target performance values. The lesion-wise sensitivity was 92.26%. The overall false-positive detection rate per case was 0.123. Of the 168 aneurysms, 13 aneurysms from 12 examinations were missed by the model.

Conclusion: The present deep learning model for automated detection of unruptured intracranial aneurysms on TOF MRA achieved the target diagnostic performance comparable to that of human radiologists. With high standalone performance, this model may be useful for accurate and efficient diagnosis of intracranial aneurysm.

Key Words: Artificial intelligence, deep learning, intracranial aneurysm, magnetic resonance angiography

INTRODUCTION

Rupture of an intracranial aneurysm is a major cause of nontraumatic subarachnoid hemorrhage, with an annual rupture risk of approximately 1%, and is associated with a high risk of morbidity and mortality.[1,2] Therefore, identification and risk stratification of intracranial aneurysms before rupture are clinically crucial and represent an active field of investigation.[3,4]

Time-of-flight (TOF) magnetic resonance angiography (MRA) is the preferred modality for identifying unruptured intracranial aneurysms in asymptomatic patients; this is not only because there is no risk of exposure to ionizing radiation or contrast agents,

Received: April 22, 2021 **Revised:** July 29, 2021 **Accepted:** August 30, 2021
Corresponding author: Hyun Seok Choi, MD, PhD, Department of Radiology, Seoul Medical Center, 156 Sinnae-ro, Jungnang-gu, Seoul 02053, Korea.
Tel: 82-2-2276-7232, Fax: 82-2-2276-7241, E-mail: chlgustjr1@gmail.com

•The authors have no potential conflicts of interest to disclose.

but also because it shows high diagnostic performance for intracranial aneurysm detection.[5,6] However, a dramatic increase in the number of radiologic examinations, coupled with a relative shortage of experienced neuroradiologists and the resultant excess workload, is a valid concern as it might lead to the missing of unruptured intracranial aneurysms.[7,8]

In this context, computer-aided diagnostic (CAD) systems, which help improve the interpretation accuracy of radiologists by providing a second opinion, are considered to be one of the possible options.[9] Moreover, along with recent advances in the field of deep learning, several studies have evaluated deep learning algorithms in terms of their applicability as a CAD algorithm for intracranial aneurysm detection on TOF MRA.[10-14] These studies achieved high detection sensitivities and proved that deep learning algorithms were useful as a supportive and complementary tool for reducing the number of missed aneurysms. However, to achieve such a high sensitivity, these algorithms showed substantially high rates of false-positive detection. The high rate of false-positive detection may potentially hinder radiologists from using CAD systems; they may not use these systems if distinguishing true-positive aneurysms from false-positive aneurysms requires additional meticulous care.[9]

Previously, we investigated a deep learning model for automated detection and localization of intracranial aneurysms on TOF MRA and validated its high sensitivity and specificity in internal and external test sets.[15] However, similar to other previous studies, the study was also suboptimal for the systematic evaluation of the performance of the algorithm, as it used an arbitrarily chosen number of test data.

Therefore, the primary endpoint of this study was to investigate whether a deep learning model for automated detection of unruptured intracranial aneurysms on TOF MRA can achieve a target diagnostic performance comparable to that of human radiologists for approval from the Korean Ministry of Food and Drug Safety (MFDS) as an artificial intelligence-applied software. For this purpose, we performed a clinical trial designed according to the guidance of the Korean MFDS.

MATERIALS AND METHODS

This single-center, retrospective, confirmatory clinical trial was approved by the Institutional Review Board of Severance Hospital, Yonsei University Health System (1-2019-0053). The requirement for informed consent was waived.

Target performance and sample size estimation

The evaluation of target performance and sample size estimation were performed by a statistical expert, who is one of the authors of this study. The MFDS suggested that the target performance for approval should be set higher than that in recently published articles. The inclusion criteria for the reference articles were as follows: 1) reports regarding sensitivity and specificity of clinicians for intracranial aneurysm detection, 2) results of TOF MRA, and 3) results of patient-wise analysis.[6,16-19] Five published articles met the criteria for the references for determination of target performance. Based on the list of references, the mean sensitivity was 88.1%, and the mean specificity was 93.7%. In addition, half of the 95% confidence intervals (CIs) of the sensitivities and specificities were 1.4%–18.0% and 3.6%–26.5%, respectively (Supplementary Table 1, only online). Therefore, the target sensitivity and target specificity were set as 87% and 92%, respectively. Also, the marginal error for sensitivity and specificity were set as 6% and 4%, respectively.

Taking the target performance and a 10% dropout rate into account, the required number of examinations was calculated as 135 for aneurysm-containing examinations and 197 for aneurysm-free examinations by using a sample size calculation formula.[20]

Training set

Six hundred TOF MRA examinations of unruptured intracranial aneurysms in patients without a history of surgical clipping, coil embolization, or stent insertion were randomly extracted from our hospital database based on radiologic reports from January 2014 to December 2016. Twelve examinations were excluded due to pronounced artifacts or poor image quality (n=5), fusiform aneurysm (n=3), and junctional dilatation (n=4). Of the final 588 examinations, 468 examinations were randomly assigned as the training set, which were also randomly assigned to training data and validation data for hyperparameter tuning with a ratio of 9:1. These TOF MRA examinations had been performed using both 1.5-T (Achieva, Philips Medical Systems, Best, The Netherlands) and various 3-T scanners (Achieva and Ingenia, Philips Medical Systems; Trio Tim, Siemens, Erlangen, Germany; Discovery MR750, GE Healthcare, Waukesha, WI, USA). The details of the training set have been reported in our previous research (Fig. 1).[15]

Test set

For the test set of this study, intracranial TOF MRA examinations that were performed after January 2018 were extracted from our institution in consecutive order, which were temporally separate from the training data set. The inclusion criteria for aneurysm-containing examinations were as follows: 1) >18 years of age; 2) TOF MRA performed using 1.5-T or 3-T scanners; and 3) untreated unruptured intracranial aneurysms on radiologic report. The exclusion criteria for aneurysm-containing examinations were as follows: 1) non-saccular aneurysm, such as mycotic aneurysm, dissecting aneurysm, or pseudoaneurysm (n=8); 2) giant aneurysm, which is larger than 25 mm in diameter (n=0); 3) ruptured aneurysm (n=0); 4) significant displacement of the intracranial vascular structure due to intracranial hemorrhage or tumor (n=1); 5) pronounced artifacts (n=4); 6) equivocal aneurysm cases in which consensus among three neuroradiologists (JBO, CJH, and CHS with 2, 11, and 15

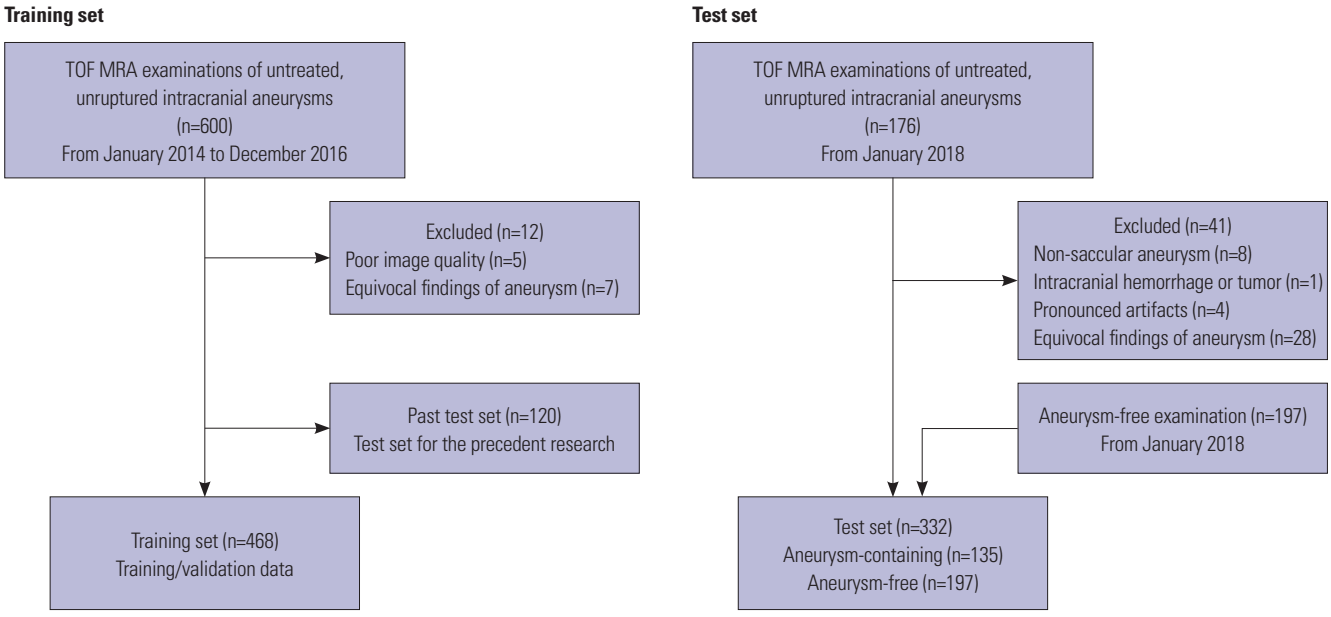

Fig. 1. Flowchart of datasets. TOF, time-of-flight; MRA, magnetic resonance angiography.

years of experience in neuroradiology, respectively) on the number and location of aneurysms was not reached (n=28). Finally, 135 aneurysm-containing examinations were eligible for inclusion in the test set.

For the aneurysm-free examinations in the test set, 197 aneurysm-free TOF MRA examinations, which were taken from January 2018 to June 2019, were randomly extracted based on radiologic reports from our institution. The three neuroradiologists reviewed the recruited examinations and reached a consensus that these examinations did not display any discernible aneurysm, significant vascular steno-occlusion, any significant structural abnormalities, such as intracranial hemorrhages and tumors, or pronounced artifacts. Finally, the test set consisted of 135 aneurysm-containing examinations and 197 aneurysm-free examinations. All data were made anonymous (Fig. 1).

TOF MRA data acquisition

All examinations were performed using both 1.5-T (Achieva) and 3-T scanners (Achieva and Ingenia; Trio Tim; Discovery MR750). The number of examinations according to the different scanners in both the training set and test set is shown in Supplementary Table 2 (only online).

The following parameters were used for the 3D TOF MRA. For 3-T scanners: repetition time (TR), 20–24 ms; echo time (TE), 3.4 ms; flip angle, 18–20°; field of view (FOV), 200–250 mm; section thickness, 0.5–0.7 mm; and matrix, 384×310–880×637. For 1.5-T scanners: TR, 24 ms; TE, 6.9 ms; flip angle, 20°; FOV, 220 mm; section thickness, 0.6 mm; matrix, 576×440.

Model development

After our previous research was reported,[15] we developed a new model that hierarchically reduces the search space by leveraging prior knowledge of the cerebral vessel morphology using the same training set. For the training set, manual segmentation for each aneurysm was performed by two radiologists (JBO, SBS) in a voxel-wise manner using in-house software (DEEP:PHI, DEEPNOID) and validated by another radiologist (SSA, 15 years of experience in neuroradiology). Model training was performed using the aneurysm segmentation as the ground truth mask.

Our model consisted of three stages. First, TOF MRA images were preprocessed by following steps: resampled to a fixed voxel size of 0.5×0.5×0.5 mm^3, cropped around the center of the brain to a size of (320, 320, 140), and performed the Z-score normalization. Automated vessel segmentation was conducted to the preprocessed MRA images using both the region growing method and Otsu thresholding method.[21] By applying the 3D skeletonization algorithm to the segmented vessel, volume of interests (VOIs) were extracted around each skeleton point with a patch size of (32, 32, 32) and a voxel size of 0.5×0.5×0.5 mm^3. Approximately 2000 VOIs were extracted for each TOF MRA examination. Next, we selectively suppressed aneurysm-free VOIs by determining whether each VOI contains intracranial aneurysms. We used a deep learning algorithm of 3D ResNet that maps from the VOIs to binary values.[22] If the intersection over union (IOU) between an aneurysm mask and a VOI was 80% or greater, the VOI was annotated as positive for intracranial aneurysms. In contrast, if the IOU was 20% or less, the VOI was annotated as negative. The other cases were excluded. To train the 3D ResNet, we used the following parameters: 200 epochs, batch size of 32, without dropout, RMSProp optimizer, learning rate of 0.001, and loss of binary cross entropy. Finally, another deep learning algorithm of 3D Unet was trained to locate intracranial aneurysms in the aneurysm-containing VOIs.[23] For the 3D Unet, we used the following parameters: 150 epochs, batch size of 64, without dropout, RMSProp optimizer, learning rate of 0.001, and loss of binary cross entropy. Model training was performed us-

https://doi.org/10.3349/ymj.2021.62.11.1052

ing Tensorflow r1.15 on a single GPU server with 512 GB RAM memory, Xeon E5-2640 v4 CPU (Intel), and TITAN Xp (NVIDIA). Based on this patch-wise segmentation algorithm, we obtained the 3D locations of intracranial aneurysms on TOF MRA images. We reduced false-positive detection by discarding aneurysms that had a probability value lower than the cutoff value of 0.5. The model finally presented a 1.6×1.6×1.6 cm^3-sized bounding box that was expected to contain an aneurysm on TOF MRA images. If the algorithm predicted that there was no aneurysm on the examination, no bounding box was presented. Flow charts of the model pipeline are shown in Fig. 2. The structure of the modified 3D ResNet and 3D Unet algorithms used in this study is shown in Supplementary Table 3 (only online).

Model evaluation

For the test set, manual segmentation for each aneurysm was performed by two neuroradiologists (JBO, CJH) and validated by another neuroradiologist (CHS). These segmentation masks were used as the ground truth for the evaluation of model performance.

Two authors (WSY, KHY) who had not been involved in the ground truth annotation evaluated model performance in the test set that was randomly shuffled. The sensitivity and specificity were evaluated using a patient-wise method. For a case to be classified as a true-positive, the model had to present bounding boxes including all aneurysms in an aneurysm-containing examination. If the model missed any aneurysm in an aneurysm-containing examination, the case was classified as a false-negative. Conversely, if the model presented any bounding box in an aneurysm-free examination, the case was classified as a false-positive. If the model identified all presenting aneurysms but showed additional bounding boxes where there was no aneurysm, the case was still classified as a true-positive, but the false-positive detection was separately recorded. A true-negative was defined as a case in which the model presented no bounding box in an aneurysm-free examination.

Statistical analysis

The patient-wise sensitivity and specificity of the model were analyzed to determine if the model achieved the target sensitivity and specificity. The lesion-wise sensitivity and false-positive detection rate per case were also analyzed. A subgroup

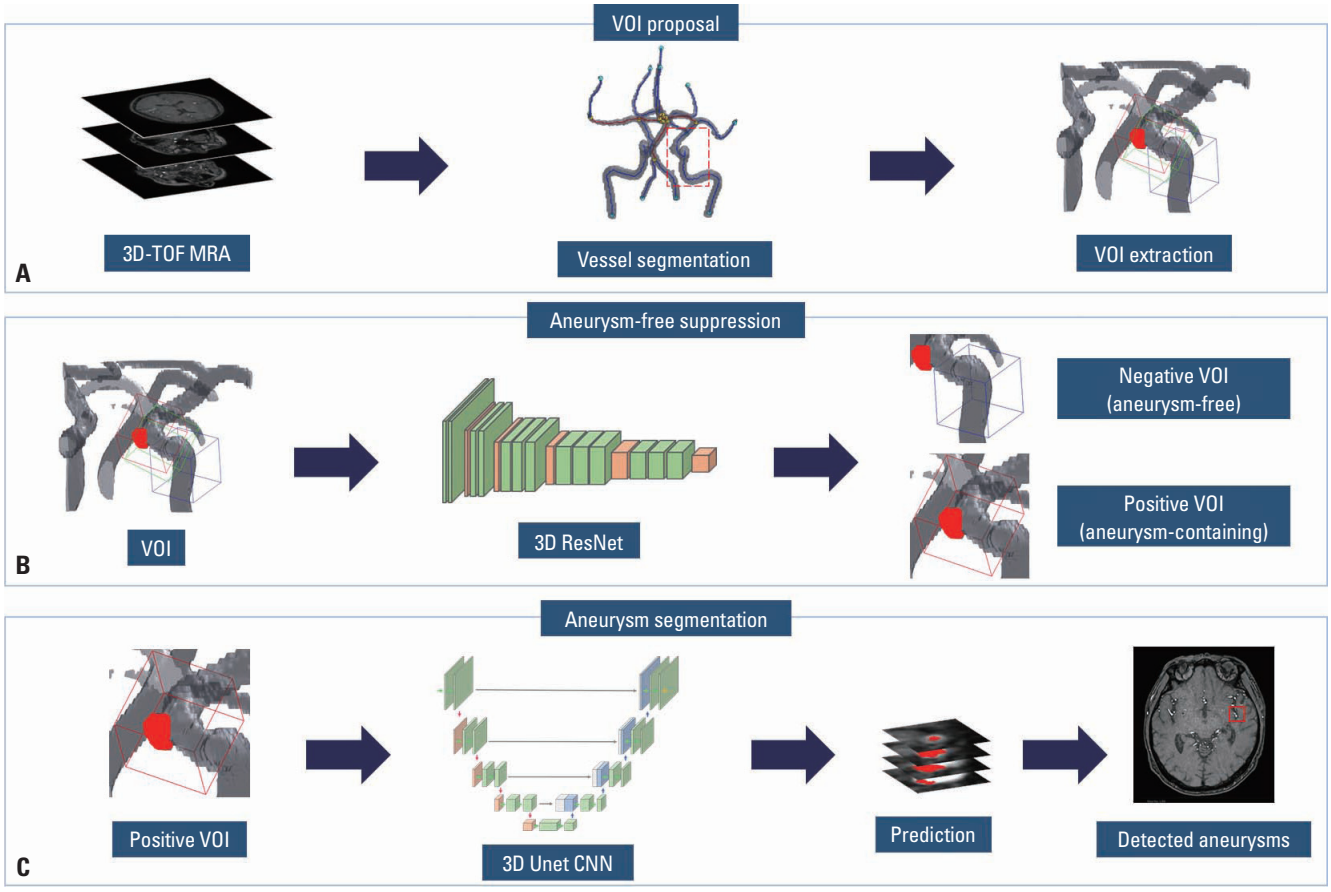

Fig. 2. Pipeline of the model. (A) After automated vessel segmentation and 3D skeletonization, approximately 2000 VOIs were extracted around each skeleton point for each TOF MRA examination. (B) For each VOI, a deep learning algorithm of 3D ResNet determined whether each VOI contains intracranial aneurysms or not. (C) Another deep learning algorithm of 3D Unet was applied to the resultant VOIs which were expected to contain intracranial aneurysms to predict the location of intracranial aneurysms. The model finally presents a 1.6×1.6×1.6 cm^3 sized bounding box expected to contain an aneurysm on TOF MRA images. VOI, volume of interest; TOF, time-of-flight; MRA, magnetic resonance angiography.

analysis was additionally performed in relation to the location and size of aneurysms. The median size difference between the detected and missed aneurysms was compared using the Mann–Whitney U test. Statistical analyses were performed using MedCalc (version 9.3.6.0; MedCalc Software, Ostend, Belgium).

RESULTS

Dataset

The 135 aneurysm-containing examinations in the test set included 168 aneurysms. The mean size of the aneurysms was 3.24±1.12 mm (range: 1.40–7.50 mm). The sizes of 168 aneurysms in the test set were as follows: 3 mm or less, 82; larger than 3 mm and up to 5 mm, 74; and larger than 5 mm, 12. Among the aneurysm-containing examinations, 107 examinations with one aneurysm, 24 examinations with two aneurysms, three examinations with three aneurysms, and one examination with four aneurysms were identified. The locations of the ground truth aneurysms were as follows: 12 aneurysms in the cavernous segment of the internal carotid artery, 106 in the paraclinoid internal carotid artery, 19 in the supraclinoid internal carotid artery, three in the bifurcation of the internal carotid artery, four in the M1 segment of the middle cerebral artery, 14 in the bifurcation of the middle cerebral artery, one in the M2 segment of the middle cerebral artery, four in the anterior communicating artery, one in the A2 segment of the anterior cerebral artery, and four in the basilar artery (Table 1).

Table 1. Characteristics and Diagnostic Performance of the Model and Target Performance

Number of examinations	135 aneurysm-containing and 197 aneurysm-free	
Number of aneurysms	168	
Size of aneurysm, mm	3.24±1.12	
Patient-wise sensitivity and specificity		
	Value	Target value (marginal error)
Sensitivity	91.11 (123/135) (95% CI: 84.99, 95.32)	87% (6%)
Specificity	93.91 (185/197) (95% CI: 89.60, 96.81)	92% (4%)
Lesion-wise sensitivity		
Overall	92.26 (155/168)	
According to the aneurysm size		
≤3 mm	84.15 (69/82)	
3–5 mm	100 (74/74)	
>5 mm	100 (12/12)	
According to the aneurysm location		
Internal carotid artery	93.57 (131/140)	
Middle cerebral artery	94.74 (18/19)	
Anterior cerebral artery	100 (1/1)	
Anterior communicating artery	50 (2/4)	
Basilar artery	75 (3/4)	
False-positive detection rate per case	0.123 (41/332)	

CI, confidence interval.
Data are presented as mean±standard deviation or %.

Diagnostic performance of the model

Of the 135 aneurysm-containing examinations, the numbers of true-positive and false-negative cases were 123 and 12, respectively. In contrast, of the 197 aneurysm-free examinations, the numbers of true-negative and false-positive cases were 185 and 12, respectively. Therefore, the sensitivity and specificity of the model were 91.11% (95% CI: 84.99, 95.32) and 93.91% (95% CI: 89.60, 96.81), respectively. These values exceeded the target sensitivity of 87% and target specificity of 92%, respectively. Also, the half width of the 95% CI of sensitivity and specificity were 5.17 and 3.61, respectively, which were lower than the predetermined marginal error of 6% and 4% for sensitivity and specificity, respectively (Table 1).

On lesion-wise analysis, the model correctly detected 155 out of 168 aneurysms, resulting in a lesion-wise sensitivity of 92.26%. The detection sensitivity was low for aneurysms 3 mm or smaller (84.15%, 69/82). On the other hand, the model showed a sensitivity of 100% for aneurysms larger than 3 mm. Of the 168 aneurysms, 13 aneurysms from 12 examinations were missed by the model: seven aneurysms in the paraclinoid internal carotid artery, two in the cavernous segment of the internal carotid artery, two in the anterior communicating artery, one in the middle cerebral artery, and one in the basilar artery. The median size of the missed aneurysms was 1.92 mm (range: 1.40–2.56 mm), which was significantly lower than that of the detected aneurysms (1.92 mm vs. 3.11 mm, $p<0.0001$). The model made 41 false-positive detections: 14 detections from 12 aneurysm-free examinations, and 27 detections from 22 aneurysm-containing examinations. Hence, the overall false-positive detection rate per case was 0.123. The aneurysm locations in these false-positive detections were as follows: 11 aneurysms in the paraclinoid internal carotid artery, eight in the cavernous segment of the internal carotid artery, five in the bifurcation of the internal carotid artery, four in the anterior communicating artery, four in the internal carotid artery-posterior communicating artery, three in the basilar artery, 3 in the bifurcation of the middle cerebral artery, 2 in the M1 segment of the middle cerebral artery, and one in the posterior cerebral artery. One of the false-positive detections in the basilar artery turned out to be a persistent primitive trigeminal artery orifice, and another false-positive detection in the anterior communicating artery was a fenestration rather than a true aneurysm. Examples of true-positive, false-negative, and false-positive cases are shown in Figs. 3–5.

https://doi.org/10.3349/ymj.2021.62.11.1052

DISCUSSION

In this confirmatory clinical trial, the results demonstrated that the primary endpoint was met: the deep learning model for detecting saccular unruptured intracranial aneurysms on TOF MRA exceeded the target sensitivity of 87% and the target speci-

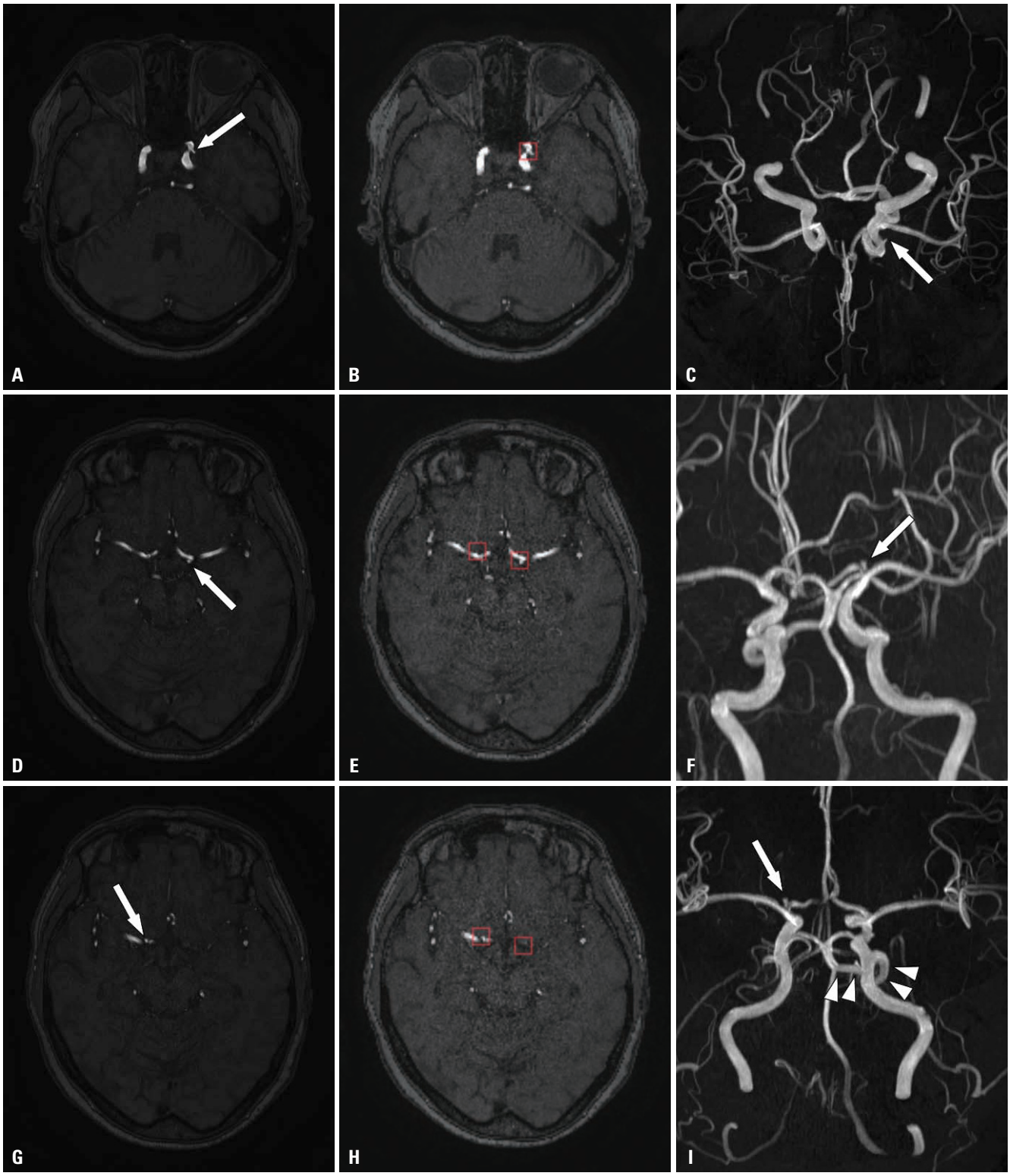

Fig. 3. True-positive case. Three aneurysms (arrow) at the left paraclinoid internal carotid artery, the bifurcation of the left internal carotid artery, and the bifurcation of the right internal carotid artery were noted on TOF MRA (A, D, G) and MIP image (C, F, I), respectively. The model correctly presented all the aneurysms (red box in B, E, H). A left persistent trigeminal artery was accidentally noted (arrowheads in I). TOF, time-of-flight; MRA, magnetic resonance angiography.

ficity of 92%, by achieving a sensitivity of 91.11% and a specificity of 93.91%. Also, the half width of the 95% CI of sensitivity and specificity were within the predetermined marginal error.

In this study, we proposed a hierarchical deep learning model that integrated 3D Unet and 3D ResNet, by leveraging prior knowledge of the intracranial vessel morphology. We reduced the search space by extracting only VOIs from near the vessels and applied 3D Unet to effectively detect intracranial aneu-

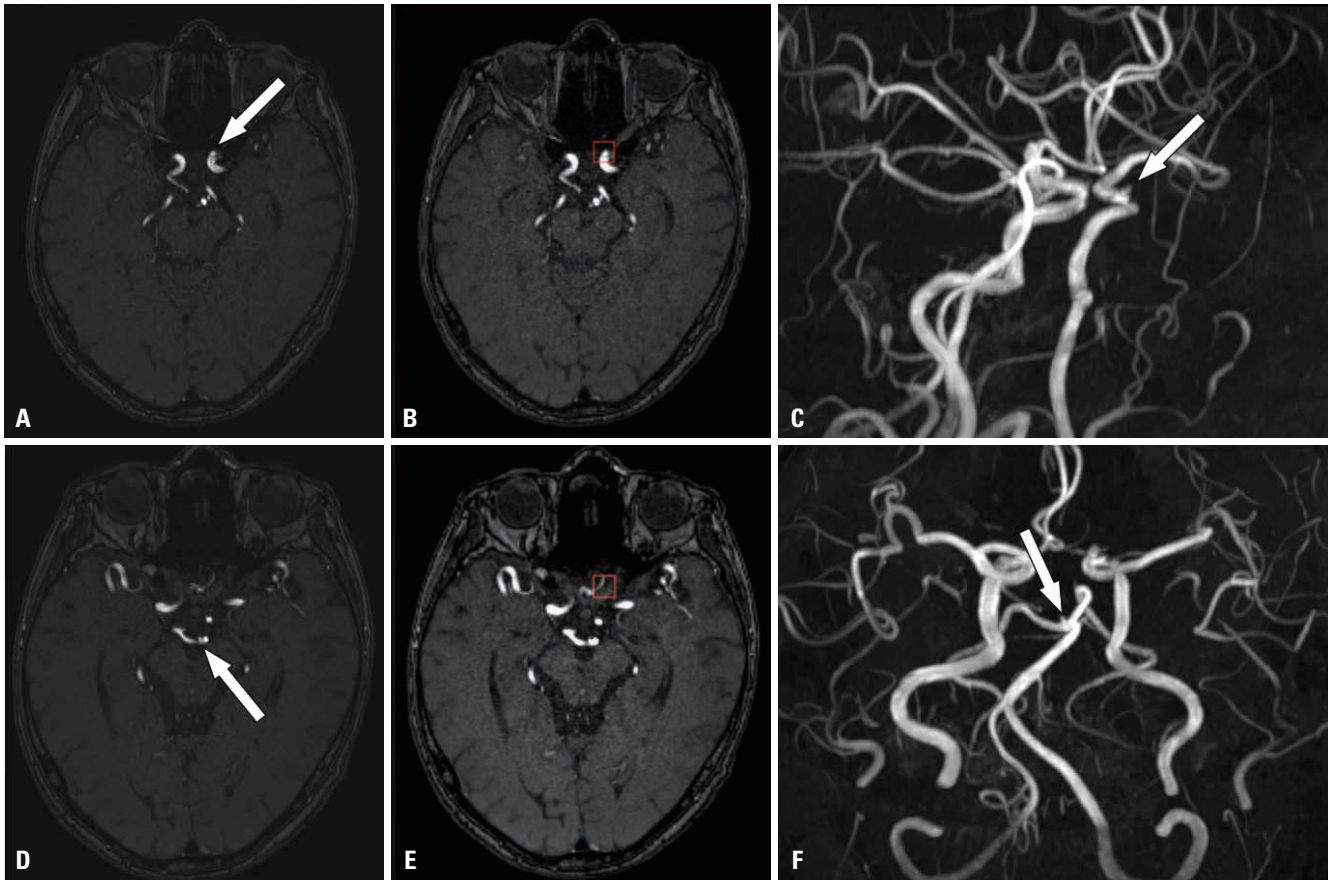

Fig. 4. False-negative case. A small aneurysm (arrow) at left paraclinoid internal carotid artery was noted on TOF MRA (A) and MIP image (C). Another small aneurysm (arrow) at the tip of the basilar artery was also noted on TOF MRA (D) and MIP image (F). The model correctly detected the left paraclinoid aneurysm (red box in B). However, the model failed to detect the small aneurysm at the basilar artery (E). The bounding box shown on (E) was for the left paraclinoid aneurysm, which was clearly seen on other slices (B). Since the model failed to detect all of the aneurysms in this examination, this case was considered as false-negative. TOF, time-of-flight; MRA, magnetic resonance angiography; MIP, maximum intensity projection.

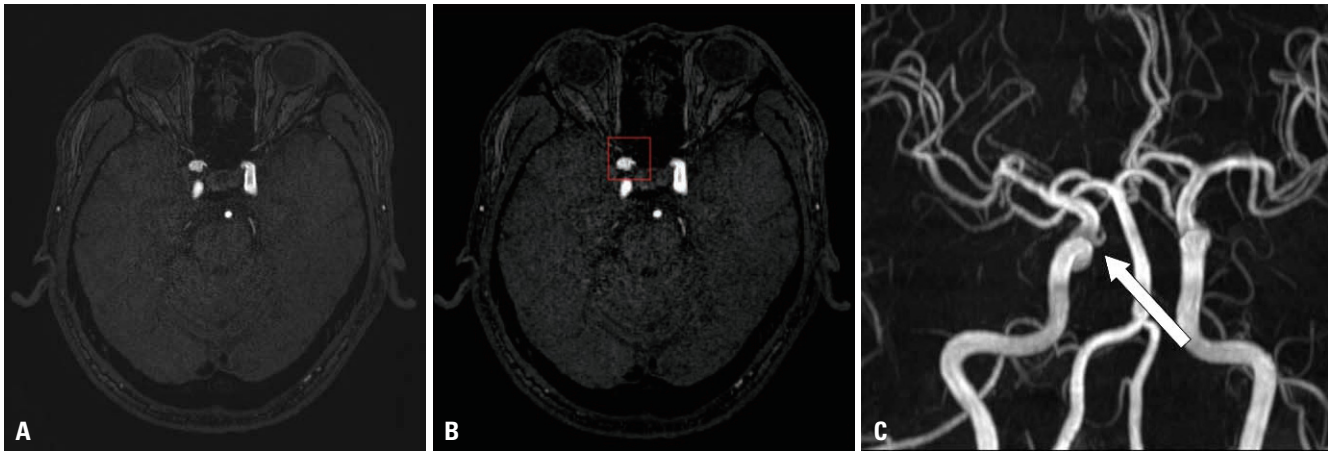

Fig. 5. False-positive case. A bulging contour at the right paraclinoid internal carotid artery was noted on TOF MRA (A). The lesion was not an aneurysm but the orifice of the ophthalmic artery (arrow), and it was clearly seen on MIP image (C). The model falsely presented the bulging contour at the orifice of the right ophthalmic artery as an aneurysm (red box in B). TOF, time-of-flight; MRA, magnetic resonance angiography; MIP, maximum intensity projection.

https://doi.org/10.3349/ymj.2021.62.11.1052

rysms. Since this approach allowed us to drastically reduce the search space, it was possible to reduce the processing time to less than 1 minute. In addition, false-positive detection rate could be significantly reduced by suppressing aneurysm-free VOIs by using 3D ResNet. Furthermore, unexplainable factors, known as the "black box problems" of deep learning models, were minimized by making decisions only on the extracted VOIs instead of entire images and by representing bounding boxes where the model predicted intracranial aneurysms are located. This efficient and accurate model with 3D approach was in contrast with previous studies for automated detection of intracranial aneurysm on TOF MRA or CTA. Nakao, et al.[10] used a 2.5D approach in which 9-direction of maximum intensity projection (MIP) images were applied to CNNs, resulting in the sensitivity of 94.2% with 2.9 false-positive detections per case. Ueda, et al.[11] reported at least four false-positive detections per case to obtain the sensitivity of 90% and reported the sensitivity of 61.1%–66.2% when the number of false-positive detections per case was reduced to one. Stember, et al.[13] applied Unet to 2D MIP image to automatically determine the size of basilar tip aneurysms. Recently, several studies have been reported to overcome the limitations of the 2D approach.[12,24] For example, Sichtermann, et al.[12] established deep learning systems that applied 3D patches on TOF MRA at multiple scales. Although they demonstrated a potential of 3D approach, however, the model took over 5 minutes for preprocessing and showed more than two false-positive detections per case to achieve the sensitivity of approximately 90%.

The main purpose of using a CAD system is to avoid missing pathologic lesions and eventually reduce the workload of the radiologist.[25] However, if the CAD system presents too many false-positive detections to achieve a certain level of sensitivity, checking those false-positive detections would increase the workload. Therefore, in order for a CAD system to be clinically useful, the standalone performance of the CAD system should be comparable to that of expert radiologists.[26] In this context, we defined a true-positive detection as one in which the model correctly detected all presenting aneurysms on an aneurysm-containing examination without missing any aneurysms. In addition, true-negative detection was defined as no presentation of bounding boxes on an aneurysm-free examination. Even though the definitions were strict, this model achieved high levels of both sensitivity and specificity, which were comparable to those of human radiologists, while achieving a false-positive detection rate per case of 0.123. The high diagnostic performance with a low level of false-positive detection and the short processing time of this model may facilitate an accurate and prompt diagnosis of intracranial aneurysms.

What makes this study more distinguishable is that a priori sample size calculation was conducted to obtain valid results. To date, the present study is the first to apply sample size estimation in developing deep learning models for the detection of intracranial aneurysms on TOF MRA, ensuring the diagnostic performance in clinical settings.[10-15] Based on the high diagnostic performance in this confirmatory clinical trial, the present model obtained MFDS approval in South Korea.

The median size of the missed aneurysms by the model was significantly smaller than that of the detected aneurysms, which is in line with previous studies where deep learning models were applied for intracranial aneurysm detection.[12,15] This tendency might be attributed to insufficient training data on these small aneurysms or a lack of consistency in differentiating aneurysm versus non-aneurysmal bulging contours in equivocal sizes. Of note, regarding the detection sensitivity according to the aneurysm location, the present model showed lower detection sensitivities for aneurysms in the anterior communicating artery and basilar artery, which were shown to be 50% (2/4) and 75% (3/4), respectively. These low sensitivities in these areas might have proceeded from the small size of the missed aneurysms, which were measured 1.40 mm and 1.92 mm in the anterior communicating artery and 1.91 mm in the basilar artery, as well as the small number of cases included in both training and test sets. Further investigation with more cases of small aneurysms and various aneurysm locations may be needed to improve the accuracy of the model.

During the determination of target sensitivity and specificity and the preparation of the test set, the aneurysm size was not primarily considered, except for giant aneurysms, which can hardly be missed by radiologists. On the other hand, TOF MRA data were included in consecutive order, without being selected using size criteria, in order to reflect the real clinical settings. The resultant mean size of the aneurysms in the test set was not significantly different from the mean size of unruptured and untreated intracranial aneurysms from a previous study that analyzed the natural course of unruptured intracranial aneurysms in South Korea.[27] Nevertheless, given that the natural course and features might vary among different populations, further studies with various populations may be required.[2,28]

In addition to the retrospective study design, this study had several limitations. First, this study was conducted based on single-center data. Altough we included images from various scanners in this study, further study with external validation may be warranted. Despite recent advances in deep learning algorithms, however, a fully generalized multi-center data-based algorithm does not exist. As one of the realistic approaches, this institution-specific model can be used under the supervision of radiologists. This institution-specific model has been integrated into the PACS system of our institution and is being partially used for further investigation. Second, giant aneurysms that were 25 mm or larger in size were excluded from the test set, even though there were no cases of giant aneurysms among candidates for the test set in this study. We thought that the inclusion of giant aneurysms would result in bias in model performance due to their substantial rarity.[29] Therefore, the performance of this model for intracranial giant aneurysms is not guaranteed. Third, the model was validated only for saccular,

unruptured, and untreated intracranial aneurysms. Also, there were no cases of significant steno-occlusive lesions in both aneurysm-containing and aneurysm-free examinations of the test set. The performance of the model might be different in those conditions. Therefore, further studies are warranted to investigate the model performance with inclusion of not only the aforementioned conditions but also other vascular conditions such as ruptured aneurysm, vasculitis, vascular malformation, or dissection. Since these conditions are usually symptomatic, radiologists can be informed in advance. However, this study was designed primarily for the surveillance of non-symptomatic aneurysms. Fourth, the ground truth annotation of aneurysms was conducted by a consensus of three neuroradiologists based not only on TOF MRA data but also on data from other examinations (as additional references), such as CT angiography or DSA, only when they were available. DSA is currently considered to be the reference standard for the imaging of intracranial aneurysms. However, in real clinical practice, DSA is not always performed, except when a confirmatory diagnosis is needed or an endovascular treatment is considered, due to its invasiveness and procedure-related risks.[30] Therefore, the inclusion of only intracranial aneurysms that were confirmed on DSA would have caused higher bias in this study. Fifth, this model used a patch-wise approach rather than original images, which may require a large amount of computation. Several approaches, such as parallel computing, sparse coding, or image compression, would be beneficial to deal with the computation efficiently. Finally, a comparison of diagnostic performance between the model and human radiologists was not conducted in this study. We are planning to conduct a further study regarding this issue.

In conclusion, the present deep learning model for automated detection of unruptured intracranial aneurysms on TOF MRA met the primary endpoint by achieving the target diagnostic performance comparable to that of human radiologists. With the high standalone performance and the short processing time, this model may be useful for making a correct diagnosis of intracranial aneurysms and for the eventual reduction of clinician workload.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Information and Communication Technology, Korea, under the Information Technology Research Center support program (IITP-2020-2020-0-01461), supervised by the Institute for Information and Communications Technology Planning and Evaluation.

AUTHOR CONTRIBUTIONS

Conceptualization: Hyun Seok Choi. Data curation: Bio Joo, Jihoon Cha, and So Yeon Won. Formal analysis: Bio Joo. Funding acquisition: Hyun Seok Choi. Investigation: Bio Joo. Methodology: Hyun Seok Choi. Project administration: Bio Joo. Resources: Hyun Seok Choi. Software: Hwa Pyung Kim, Jong Mun Choi, Sang Min Lee, and Tae Gyu Kim. Supervision: Hyun Seok Choi. Validation: Hwa Pyung Kim, Jong Mun Choi, Sang Min Lee, and Tae Gyu Kim. Visualization: Bio Joo. Writing—original draft: Bio Joo. Writing—review & editing: Sung Soo Ahn, Jihoon Cha, So Yeon Won, Beomseok Sohn, Hwiyoung Kim, Kyunghwa Han, Seung-Koo Lee, and Hyun Seok Choi. Approval of final manuscript: all authors.

ORCID iDs

Bio Joo https://orcid.org/0000-0001-7460-1421
Hyun Seok Choi https://orcid.org/0000-0003-4999-8513
Sung Soo Ahn https://orcid.org/0000-0002-0503-5558
Jihoon Cha https://orcid.org/0000-0002-1662-8041
So Yeon Won Https://orcid.org/0000-0003-0570-3365
Beomseok Sohn https://orcid.org/0000-0002-6765-8056
Hwiyoung Kim https://orcid.org/0000-0001-7778-8973
Kyunghwa Han https://orcid.org/0000-0002-5687-7237
Hwa Pyung Kim https://orcid.org/0000-0003-0103-6200
Jong Mun Choi https://orcid.org/0000-0002-3267-8989
Sang Min Lee https://orcid.org/0000-0001-5954-9442
Tae Gyu Kim https://orcid.org/0000-0003-1260-9369
Seung-Koo Lee https://orcid.org/0000-0001-5646-4072

REFERENCES

1. van Gijn J, Kerr RS, Rinkel GJ. Subarachnoid haemorrhage. Lancet 2007;369:306-18.
2. UCAS Japan Investigators; Morita A, Kirino T, Hashi K, Aoki N, Fukuhara S, Hashimoto N, et al. The natural course of unruptured cerebral aneurysms in a Japanese cohort. N Engl J Med 2012;366:2474-82.
3. Backes D, Vergouwen MD, Tiel Groenestege AT, Bor AS, Velthuis BK, Greving JP, et al. PHASES score for prediction of intracranial aneurysm growth. Stroke 2015;46:1221-6.
4. Thompson BG, Brown RD Jr, Amin-Hanjani S, Broderick JP, Cockroft KM, Connolly ES Jr, et al. Guidelines for the management of patients with unruptured intracranial aneurysms: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2015;46:2368-400.
5. Sailer AM, Wagemans BA, Nelemans PJ, de Graaf R, van Zwam WH. Diagnosing intracranial aneurysms with MR angiography: systematic review and meta-analysis. Stroke 2014;45:119-26.
6. Li MH, Li YD, Gu BX, Cheng YS, Wang W, Tan HQ, et al. Accurate diagnosis of small cerebral aneurysms ≤5 mm in diameter with 3.0-T MR angiography. Radiology 2014;271:553-60.
7. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad Radiol 2015;22:1191-8.
8. OECD iLibrary. Magnetic resonance imaging (MRI) exams [accessed on 2021 July 28]. Available at: https://doi.org/10.1787/1d89353f-en.
9. Shi Z, Hu B, Schoepf UJ, Savage RH, Dargis DM, Pan CW, et al. Artificial intelligence in the management of intracranial aneurysms: current status and future perspectives. AJNR Am J Neuroradiol 2020;41:373-9.
10. Nakao T, Hanaoka S, Nomura Y, Sato I, Nemoto M, Miki S, et al. Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. J Magn Reson Imaging 2018;47:948-53.

https://doi.org/10.3349/ymj.2021.62.11.1052

11. Ueda D, Yamamoto A, Nishimori M, Shimono T, Doishita S, Shimazaki A, et al. Deep learning for MR angiography: automated detection of cerebral aneurysms. Radiology 2019;290:187-94.
12. Sichtermann T, Faron A, Sijben R, Teichert N, Freiherr J, Wiesmann M. Deep learning-based detection of intracranial aneurysms in 3D TOF-MRA. AJNR Am J Neuroradiol 2019;40:25-32.
13. Stember JN, Chang P, Stember DM, Liu M, Grinband J, Filippi CG, et al. Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. J Digit Imaging 2019;32:808-15.
14. Faron A, Sichtermann T, Teichert N, Luetkens JA, Keulers A, Nikoubashman O, et al. Performance of a deep-learning neural network to detect intracranial aneurysms from 3D TOF-MRA compared to human readers. Clin Neuroradiol 2020;30:591-8.
15. Joo B, Ahn SS, Yoon PH, Bae S, Sohn B, Lee YE, et al. A deep learning algorithm may automate intracranial aneurysm detection on MR angiography with high diagnostic performance. Eur Radiol 2020;30:5785-93.
16. Hiratsuka Y, Miki H, Kiriyama I, Kikuchi K, Takahashi S, Matsubara I, et al. Diagnosis of unruptured intracranial aneurysms: 3T MR angiography versus 64-channel multi-detector row CT angiography. Magn Reson Med Sci 2008;7:169-78.
17. Li MH, Li YD, Tan HQ, Gu BX, Chen YC, Wang W, et al. Contrast-free MRA at 3.0 T for the detection of intracranial aneurysms. Neurology 2011;77:667-76.
18. White PM, Teasdale EM, Wardlaw JM, Easton V. Intracranial aneurysms: CT angiography and MR angiography for detection prospective blinded comparison in a large patient cohort. Radiology 2001;219:739-49.
19. White PM, Wardlaw JM, Lindsay KW, Sloss S, Patel DK, Teasdale EM. The non-invasive detection of intracranial aneurysms: are neuroradiologists any better than other observers? Eur Radiol 2003;13:389-96.
20. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform 2014;48:193-204.
21. Wang R, Li C, Wang J, Wei X, Li Y, Zhu Y, et al. Threshold segmentation algorithm for automatic extraction of cerebral vessels from brain magnetic resonance angiography images. J Neurosci Methods 2015;241:30-6.
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27-30; Las Vegas, NV. New Jersey: Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 770-8.
23. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention 2015 (MICCAI 2015): 2015 Oct 5-9; Munich, Germany. Basel: Springer, Cham; 2015. p. 234-41.
24. Park A, Chute C, Rajpurkar P, Lou J, Ball RL, Shpanskaya K, et al. Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model. JAMA Netw Open 2019;2:e195600.
25. Miki S, Hayashi N, Masutani Y, Nomura Y, Yoshikawa T, Hanaoka S, et al. Computer-assisted detection of cerebral aneurysms in MR angiography in a routine image-reading environment: effects on diagnosis by radiologists. AJNR Am J Neuroradiol 2016;37:1038-43.
26. van Ginneken B, Schaefer-Prokop CM, Prokop M. Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology 2011;261:719-32.
27. Byoun HS, Huh W, Oh CW, Bang JS, Hwang G, Kwon OK. Natural history of unruptured intracranial aneurysms: a retrospective single center analysis. J Korean Neurosurg Soc 2016;59:11-6.
28. Wiebers DO, Whisnant JP, Huston J 3rd, Meissner I, Brown RD Jr, Piepgras DG, et al. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. Lancet 2003;362:103-10.
29. Sekhar LN, Tariq F, Mai JC, Kim LJ, Ghodke B, Hallam DK, et al. Unyielding progress: treatment paradigms for giant aneurysms. Clin Neurosurg 2012;59:6-21.
30. Howard BM, Hu R, Barrow JW, Barrow DL. Comprehensive review of imaging of intracranial aneurysms and angiographically negative subarachnoid hemorrhage. Neurosurg Focus 2019;47:E20.