

Deep Learning–Based Automatic Classification of Ischemic Stroke Subtype Using Diffusion–Weighted Images

Wi-Sun Ryu,^{1,2} Dawid Schellingerhout,³ Hoyoun Lee,² Keon-Joo Lee,⁴ Chi Kyung Kim,⁴ Beom Joon Kim,⁵ Jong-Won Chung,⁶ Jae-Sung Lim,⁷ Joon-Tae Kim,⁸ Dae-Hyun Kim,⁹ Jae-Kwan Cha,⁹ Leonard Sunwoo,¹⁰ Dongmin Kim,² Sang-Il Suh,¹¹ Oh Young Bang,⁶ Hee-Joon Bae,⁵ Dong-Eog Kim^{1,12}

¹Department of Neurology, Dongguk University Ilsan Hospital, Goyang, Korea

²Artificial Intelligence Research Center, JLK Inc., Seoul, Korea

³Department of Neuroradiology and Imaging Physics, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

⁴Department of Neurology, Korea University Guro Hospital, Seoul, Korea

⁵Department of Neurology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Korea

⁶Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

⁷Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

⁸Department of Neurology, Chonnam National University Hospital, Chonnam National University Medical School, Gwangju, Korea

⁹Department of Neurology, Dong-A University Hospital, Busan, Korea

¹⁰Department of Radiology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Korea

¹¹Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, Korea

¹²National Priority Research Center for Stroke, Goyang, Korea

Background and Purpose Accurate classification of ischemic stroke subtype is important for effective secondary prevention of stroke. We used diffusion-weighted image (DWI) and atrial fibrillation (AF) data to train a deep learning algorithm to classify stroke subtype.

Methods Model development was done in 2,988 patients with ischemic stroke from three centers by using U-net for infarct segmentation and EfficientNetV2 for subtype classification. Experienced neurologists (n=5) determined subtypes for external test datasets, while establishing a consensus for clinical trial datasets. Automatically segmented infarcts were fed into the model (DWI-only algorithm). Subsequently, another model was trained, with AF included as a categorical variable (DWI+AF algorithm). These models were tested: (1) internally against the opinion of the labeling experts, (2) against fresh external DWI data, and (3) against clinical trial dataset.

Results In the training-and-validation datasets, the mean (\pm standard deviation) age was 68.0 \pm 12.5 (61.1% male). In internal testing, compared with the experts, the DWI-only and the DWI+AF algorithms respectively achieved moderate (65.3%) and near-strong (79.1%) agreement. In external testing, both algorithms again showed good agreements (59.3%–60.7% and 73.7%–74.0%, respectively). In the clinical trial dataset, compared with the expert consensus, percentage agreements and Cohen's kappa were respectively 58.1% and 0.34 for the DWI-only vs. 72.9% and 0.57 for the DWI+AF algorithms. The corresponding values between experts were comparable (76.0% and 0.61) to the DWI+AF algorithm.

Conclusion Our model trained on a large dataset of DWI (both with or without AF information) was able to classify ischemic stroke subtypes comparable to a consensus of stroke experts.

Keywords Deep learning; Artificial intelligence; Diffusion magnetic resonance imaging; Atrial fibrillation; Ischemic stroke

Correspondence: Dong-Eog Kim
Department of Neurology, Dongguk University College of Medicine,
27 Dongguk-ro, Ilsandong-gu, Goyang 10326, Korea
Tel: +82-31-961-7207
E-mail: kdongeog@duih.org
<https://orcid.org/0000-0002-9339-6539>

Received: February 6, 2024

Revised: April 4, 2024

Accepted: May 7, 2024

Introduction

Studies have shown that the volume¹ and pattern² of ischemic lesions on diffusion-weighted image (DWI) are associated with stroke subtype and predictive of post-stroke functional outcomes and future cerebrovascular events. Approximately a quarter of patients with ischemic stroke experience recurrence.^{3,4} In a previous study of 7,101 patients with acute ischemic stroke, we observed that large artery atherosclerosis (LAA) and cardioembolic strokes were associated with an approximately 5-times higher risk of recurrence at 1-year, compared with small vessel occlusion (SVO) stroke.⁵ The etiology of stroke is critical to the correct implementation of future preventative strategies.

The Trial of Org10172 in Acute Stroke (TOAST) classification has been the most frequently method employed for etiologic stroke subtyping in clinical practice and research.⁶ The original TOAST classification required clinical features and data from tests including brain imaging (computed tomography/magnetic resonance imaging [CT/MRI]), cardiac evaluation (electrocardiography [ECG], echocardiography, and etc.), duplex imaging of extracranial arteries, arteriography, and laboratory assessments for a pro-thrombotic state.⁶ Additional tests, such as Holter monitoring, implantable loop recorder, and high-resolution vessel wall MRI have enabled more precise stroke subtyping.⁷ However, these tests increase the cost and the length of hospital stay. Moreover, many countries lack enough access to these advanced techniques. A diagnosis support system using initial or simple exams, such as DWI and ECG, to detect acute infarcts and atrial fibrillation (AF) could reduce costs^{8,9} and assist clinicians who do not have access to other resources to determine stroke etiology.

To date, a few previous studies have developed automated systems for classifying stroke subtypes using deep learning algorithms and DWI.^{10,11} However, no study has externally validated these algorithms, which is critically important given the low inter-rater reliability in the classification of stroke subtypes.¹² In the present multi-center study, we enrolled about 6,500 patients with acute ischemic stroke. Using 2,489 patients' DWI data with and without information on the presence of AF, we developed a deep learning algorithm to classify stroke subtypes. We then externally validated the deep learning algorithm on a new set of 3,384 patients, using three temporally and regionally different datasets. In addition, we compared stroke subtype classifications by the deep learning algorithm versus neurovascular experts. Finally, we outlined practical applications of the deep learning-based stroke subtype classification for cardioembolism (CE) risk stratification based solely on initial DWI assessments, for use when AF information is not available or becomes available after continuous ECG monitoring (for days–years).^{13,14}

Methods

Participants

Dataset for training, validation, and internal testing

From May 2011 to March 2014, we consecutively enrolled 4,514 patients with acute ischemic stroke, who were admitted to three university hospitals (Dongguk University Hospital, Seoul National University Bundang Hospital, and Dong-A University Hospital). We included a consecutive series of patients who were admitted within 7 days of onset, while excluding the following patients with: (1) unavailable or poor-quality of DWI (n=342), (2) other causes of stroke (n=241), and (3) undetermined stroke subtype (n=933) (Supplementary Figure 1). The remaining 2,998 patients' data were used for training, validation, and internal test, using random sub-setting in ratio of 7:2:1. The institutional review board of Dongguk University Hospital (IRB No. 2017-09-017) and each participating center approved the study protocol, and patients or their legal proxies provided a written informed consent.

Datasets for external testing

A total of 3,384 fresh stroke imaging datasets were used for external testing, comprised of the following components.

External test dataset 1

From May 2011 to March 2014, 2,787 patients with acute ischemic stroke who were admitted within 7 days of symptom onset were consecutively enrolled from Chonnam National University Hospital. After excluding 868 patients, 1,919 were finally included (Supplementary Figure 1).

External test dataset 2

From October 2021 to August 2022, 1,315 patients with acute ischemic stroke who were admitted within 7 days of symptom onset were consecutively enrolled from the Chonnam National University Hospital. After excluding 491 patients, 824 were finally included (Supplementary Figure 1).

External test dataset 3

From March 2021 to April 2022, 931 patients with acute ischemic stroke who were admitted within 7 days of symptom onset were consecutively enrolled from Korea University Guro Hospital. After excluding 290 patients, 641 were finally included (Supplementary Figure 1).

Clinical trial dataset

A pivotal clinical trial was conducted to assess the efficacy of deep learning algorithms in comparison to a standard reference

established through expert consensus, and to measure the level of agreement between the deep learning algorithm and the consensus as well as among the experts themselves. From March 2016 to May 2017, 1,701 patients who met the following inclusion criteria were enrolled from the two stroke centers (Dongguk University Hospital and Seoul National University Bundang Hospital): (1) age between 20 and 95 years, (2) patients with acute ischemic stroke who visited the hospitals within 7 days after symptom onset, and (3) patients who underwent DWI. According to the pre-planned exclusion criteria, we excluded 612 patients for the following reasons: inadequate or poor-quality DWI ($n=148$), other causes of stroke ($n=114$), undetermined causes of stroke ($n=315$), and CE stroke attributable to causes other than AF ($n=35$). Thus, data from 900 patients remained available for clinical testing.

Clinical data collection

Using a standardized protocol,¹⁵ we prospectively collected demographic data, prior medication history, and the presence of vascular risk factors including hypertension, diabetes mellitus, hyperlipidemia, coronary artery disease, AF, and smoking history.

Imaging acquisition and infarct segmentation

For the training data, brain MRIs were performed on 1.5 tesla ($n=2,471$) or 3.0 tesla ($n=527$) MRI systems. The DWI protocol was as follows: b-values of 0 and 1,000 s/mm^2 , echo time 50–99 ms, repetition time 2,400–9,000 ms, voxel size $1 \times 1 \times 3\text{--}5\text{ mm}^3$, interslice gap of 0–2 mm, and slice thickness of 3–7 mm. Using a validated 3D U-net algorithm that we recently developed,^{16,17} we automatically segmented infarct lesions on DWIs.

Ischemic stroke subtype classification

For the datasets for training and validation, internal testing, and external test datasets 1–3, stroke subtypes were determined by experienced vascular neurologists at each hospital, using a validated MRI-based classification system built on the TOAST criteria (details provided in Supplementary Methods and Supplementary Figure 2).⁷ Briefly, the modified TOAST classification is composed of the following five steps: (1) consideration of other determined etiologies of stroke, (2) screening for SVO on DWI, (3) consideration of relevant artery stenosis or occlusion, (4) consideration of recanalization status after recanalization therapy, and (5) consideration of follow-up recanalization status without recanalization therapy. For the clinical trial dataset, stroke subtypes were determined through consensus among three experienced vascular neurologists (J-W Chung, J-S Lim, and D-E Kim).

Development of a deep learning algorithm for ischemic stroke subtype classification

Brain DWIs were preprocessed by (1) skull stripping using the Gaussian blur and Otsu's threshold,¹⁸ (2) applying N4 bias field correction using the SimpleITK library, and (3) performing image signal normalization. After the preprocessing, infarct areas on DWI were automatically segmented using the aforementioned validated 3D U-net algorithm (JLK-DWI, JLK Inc., Seoul, Korea).^{16,17} The segmented infarct masks from raw DWIs were stacked and condensed into three 2D X, Y, Z-axis images to ensure consistent data input regardless of the number of slices (Supplementary Figure 3). These condensed 2D X, Y, Z-axis images were resized to 256×256 pixels using bilinear interpolation. Thus, the training data for the algorithm was comprised of three 2D images representing X, Y, Z-axis projections of segmented infarct area and a label (LAA, SVO, and CE).

In a pilot study, deep learning models using EfficientNetV2¹⁹ outperformed those using ResNet,²⁰ MobileNetV3,²¹ and EfficientNet²² in stroke subtyping (data not shown). The EfficientNetV2¹⁹ is a new family of convolutional networks that have faster training speed and better parameter efficiency, while adding a `global_average_pooling2d` layer to minimize overfitting by reducing the total number of parameters. In addition, we incorporated a sequence of one inner dense layer with dropout layers. In total, a 30% dropout rate was randomly chosen to avoid overfitting. Finally, one output dense layer contained 3 output units for multi-class (LAA, SVO, and CE) classification, which were designated as the DWI-only based subtype classification. The details of the layers, their order in the proposed model, and the output shape of each layer are presented in Supplementary Figure 3. The total number of parameters was 52,862,199.

To develop a deep learning algorithm that takes account for AF, we concatenated a binary value (0 vs. 1: the absence vs. presence of AF) to previous outputs, and then applied a fully connected layer. The output was then designated as the DWI+AF based subtype classification.

For all the procedures, including preprocessing and model development, we used Python 3.7.9 and 3.8.13, PyTorch 1.12.0, Torchvision 0.13.0, pandas 1.2.4, NumPy 1.19.5/1.22.3, SciPy 1.4.1/1.6.3, scikit-image 0.15.0/0.18.1, SimpleITK 2.1.1, and Pydicom 2.1.2. Each model was trained for approximately 9 hours using a hardware system comprising Intel Xeon Silver 4314 @2.40 GHz, 640 GB RAM, and NVIDIA Quadro RTX A6000 with 48GB GDDR6.

Expert consensus for the classification of stroke subtype in the clinical trial dataset

For the clinical trial dataset, we first assessed the inter-observer

agreement of stroke subtype classification between two experts (J-W Chung and J-S Lim, board-certified neurologists with more than 5-year experience in both stroke practice and research), who had served as stroke neurologists at least 5 years and independently reviewed the brain MRI and patients' data. Information provided to the reviewers included age, sex, the presence of AF, DWIs, and magnetic resonance or CT angiography. Based on the aforementioned ischemic stroke subtype-classification system,⁷ they independently determined etiologies (i.e., LAA, SVO, or CE). In cases of disagreement between the two reviewers, a third reviewer (D-E Kim) served as the tie-breaker. When the final consensus on stroke subtype was undetermined or other determined stroke, the case was excluded from the analysis. The experts' consensus classifications were compared with the deep learning algorithm's classifications.

Statistical analysis

The baseline characteristics between datasets were compared using the analysis of variance or Kruskal–Wallis test for continuous variables and chi-square test for categorical variables, as appropriate. To compare the subtype classifications made by experts and those made by deep learning algorithms, we used percentage agreement and Cohen's kappa. To assess performance metrics of deep learning algorithms, we used the one-vs-rest method²³ and calculated the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value, and negative predictive value for each subtype (LAA, SVO, and CE). In the clinical trial dataset, a paired t-test was used to compare CE probabilities estimated by the DWI-only and DWI+AF algorithms for CE cases that artificial intelligence (AI) algorithms misclassified as LAA. Additionally, the rates of inter-expert disagreements for LAA, SVO, and CE were compared with the rates of disagreements between the expert consensus and the DWI+AF algorithm-based classifications for these three subtypes. Further, disagreements for the following three subtypes of LAA were also assessed (Supplementary Figure 2): LAA-negative (LAA-NG), LAA-branch atheromatous disease (LAA-BR), and LAA-lacune (LAA-LC). To evaluate the performance of deep learning algorithms depending on the onset-to-imaging time (<24 hours vs. 24 hours–7 days), we calculated the percentage agreements of stroke subtyping between both the early and late imaging groups using the external test dataset 3 as well as the clinical trial dataset, both of which had information on the time of DWI acquisition. To examine the clinical implications of AI prediction of CE using DWI, participants in each dataset were stratified into ten groups based on the probability of having CE estimated by deep learning algorithm. The trend of the observed frequency of CE stroke, as determined by experts,

was examined using a Wilcoxon-type test for trend.²⁴ All the statistical analyses described above were performed using Stata 16.0 (Stata Corp., College Station, TX, USA), and a *P*-value <0.05 was considered statistically significant.

Results

Baseline characteristics

In the training and validation datasets, the mean (standard deviation [SD]) age was 68.0 (12.5) and 61.1% were men (Table 1). Mean ages were similar in all datasets. Other demographic characteristics, such as sex, admission National Institute of Health Stroke Scale scores, and risk factors for stroke varied significantly among the datasets. The distribution of stroke subtypes also differed among the datasets, indicating their heterogeneity.

Deep learning prediction of stroke subtype using DWI data only versus DWI plus AF data

In the internal test dataset (Table 2), the percentage agreement between the DWI-only algorithm and stroke experts was 65.3% (95% confidence interval [CI]: 60.0%–70.6%); the AUC values for LAA, SVO, and CE were 0.75, 0.93, and 0.81, respectively (Supplementary Figure 4). After incorporating the information regarding the presence of AF (DWI+AF algorithm), the percentage agreement increased to 79.1% (95% CI: 74.6%–83.6%), and the AUC values for LAA, SVO, and CE increased to 0.90, 0.93, and 0.95, respectively (Figure 1).

In the external test datasets (Table 2), both algorithms again showed good agreements. The DWI-only algorithm achieved 59.3%–60.7% levels of agreements (Table 2 and Supplementary Table 1); the AUC values for LAA, SVO, and CE were 0.69–0.72, 0.83–0.90, and 0.79–0.82, respectively (Supplementary Figure 4). The DWI+AF algorithm again showed higher agreements, ranging from 73.7% to 74.0%, with Cohen's kappa ranging from 0.57 to 0.59. In addition, the accuracy of stroke subtype classification reached 0.83 (Table 3), and the AUC values for LAA, SVO, and CE increased to 0.84–0.88, 0.85–0.91, and 0.95–0.97, respectively (Figure 1).

In the clinical trial dataset (Table 2), the percentage agreements and Cohen's kappa were respectively 58.1% (95% CI: 54.9%–61.3%) and 0.34 (0.29–0.39) for the DWI-only algorithm, and the values were 72.9% (95% CI: 69.1%–76.7%) and 0.57 (0.51–0.62) for the DWI+AF algorithm, respectively. In addition, the AUC values for LAA, SVO, and CE improved from 0.68 to 0.90, 0.86 to 0.87, and 0.77 to 0.996, respectively (Supplementary Figure 4 and Figure 1).

Alluvial plots for the five datasets (Figure 2) showed that additional information regarding the presence of AF on ECG changed

Table 1. Baseline characteristics

	Training and validation (n=2,687)	Internal test (n=311)	External test dataset 1 (n=1,919)	External test dataset 2 (n=824)	External test dataset 3 (n=641)	Clinical trial dataset (n=900)	P
Age (yr)	68.0±12.5	68.2±12.9	68.7±12.0	70.2±12.4	67.5±12.0	68.6±12.4	0.229
Male sex	1,642 (61.1)	174 (56.0)	1,090 (56.8)	504 (61.2)	426 (66.5)	551 (61.3)	<0.001
Admission NIHSS score*	4 (2–9)	4 (2–9)	4 (2–10)	2 (1–5)	3 (1–5)	4 (2–7)	<0.001 [†]
Prestroke mRS score 0–2*	2,346 (87.3)	274 (88.1)	1,667 (86.9)	791 (96.0)	126 (74.6)	832 (92.5)	<0.001
Stroke subtype							<0.001
LAA	1,224 (45.6)	142 (45.7)	1,044 (54.4)	434 (52.7)	300 (46.8)	574 (63.8)	
SVO	667 (24.8)	75 (24.1)	221 (11.5)	154 (18.7)	222 (34.6)	155 (17.2)	
CE	796 (29.6)	94 (30.2)	654 (34.1)	236 (28.6)	119 (18.6)	171 (19.0)	
Prior stroke history	570 (21.2)	68 (21.9)	301 (15.7)	161 (19.5)	101 (15.8)	168 (18.7)	<0.001
Coronary artery disease	257 (9.6)	29 (9.3)	60 (3.1)	67 (8.1)	56 (8.7)	82 (9.1)	<0.001
Hypertension	1,884 (70.1)	222 (71.4)	1,203 (62.7)	484 (58.7)	426 (66.5)	660 (73.4)	<0.001
Diabetes	966 (36.0)	126 (40.5)	577 (30.1)	247 (30.0)	205 (32.0)	341 (37.9)	<0.001
Hyperlipidemia	1,246 (46.4)	148 (47.6)	299 (15.6)	154 (18.7)	68 (10.6)	385 (42.8)	<0.001
Smoking	1,122 (41.8)	121 (38.9)	717 (37.4)	219 (26.6)	215 (33.5)	394 (43.8)	<0.001
Atrial fibrillation	645 (24.0)	74 (24.0)	539 (28.1)	230 (27.9)	98 (15.3)	172 (19.1)	<0.001
Recanalization therapy	458 (17.1)	53 (17.0)	448 (23.4)	119 (14.4)	72 (11.2)	135 (15.0)	<0.001

Values are expressed as mean±standard deviation, n (%), or medians (interquartile ranges).

NIHSS, National Institute of Health Stroke Scale; mRS, modified Rankin Scale; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism.

*Data were missing in 472 and 1 patients of the external test dataset-3 and clinical trial dataset, respectively; [†]Kruskal–Wallis test was used.

Table 2. Agreements of stroke subtype classification between deep learning algorithm and stroke neurologists (experts)

	Internal test (n=311)	External test dataset 1 (n=1,919)	External test dataset 2 (n=824)	External test dataset 3 (n=641)	Clinical trial dataset (n=900)
	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts' consensus
DWI-only algorithm					
Percentage agreement (95% CI)	65.3 (60.0–70.6)	60.7 (58.5–62.9)	59.8 (56.5–63.2)	59.3 (55.5–63.1)	58.1 (54.9–61.3)
Cohen's kappa (95% CI)	0.47 (0.39–0.56)	0.38 (0.34–0.41)	0.37 (0.32–0.42)	0.37 (0.31–0.43)	0.34 (0.29–0.39)
DWI+AF algorithm					
Percentage agreement (95% CI)	79.1 (74.6–83.6)	73.7 (71.8–75.7)	74.0 (71.0–77.0)	73.9 (70.5–77.4)	72.9 (69.1–76.7)
Cohen's kappa (95% CI)	0.68 (0.61–0.75)	0.57 (0.54–0.60)	0.59 (0.54–0.64)	0.59 (0.54–0.64)	0.57 (0.51–0.62)
Between experts					
Percentage agreement (95% CI)					76.0 (74.0–79.0)
Cohen's kappa (95% CI)					0.61 (0.57–0.65)

DWI, diffusion-weighted image; CI, confidence interval; AF, atrial fibrillation.

the categorization of stroke subtype by the DWI-only algorithm from CE to LAA more often (22.1%–38.2%) than from LAA to CE (13.7%–16.2%) or from SVO to CE (4.2%–7.2%). There was no reclassification from CE to SVO. In the clinical trial dataset, we found (1) one CE stroke (with a small subcortical infarct and a tiny cortical infarct in the presence of AF; Supplementary Figure 5A) that the DWI+AF algorithm classified as SVO and (2) 65 LAA strokes cases (without AF) that the DWI+AF algorithm classified as CE. For the former CE case, one of the experts misclas-

sified it as undetermined (with two or more causes) prior to the consensus meeting, because he failed to detect the cortical infarct lesion and had to rule out both SVO and CE as possible causes of the single subcortical infarct lesion. Among the 65 LAA stroke cases, 14 cases (21.5%) had been initially classified as undetermined (two or more; AF+relevant artery stenosis) by one of the two experts. Subsequently in the consensus meeting, the degree of stenosis was determined not to be significant. As shown by the scatter plots in Supplementary Figure 5B, the DWI+AF al-

gorithm yielded significantly lower CE probabilities for the 65 cases than the DWI-only algorithm (mean±SD 0.66±0.10 vs. 0.86±0.06, respectively; $P<0.001$ by paired t test). However, add-

ing the AF-absence information did not sufficiently lower the estimated CE probability to reclassify to LAA, probably due in part to the absence of the arterial stenosis information. Review-

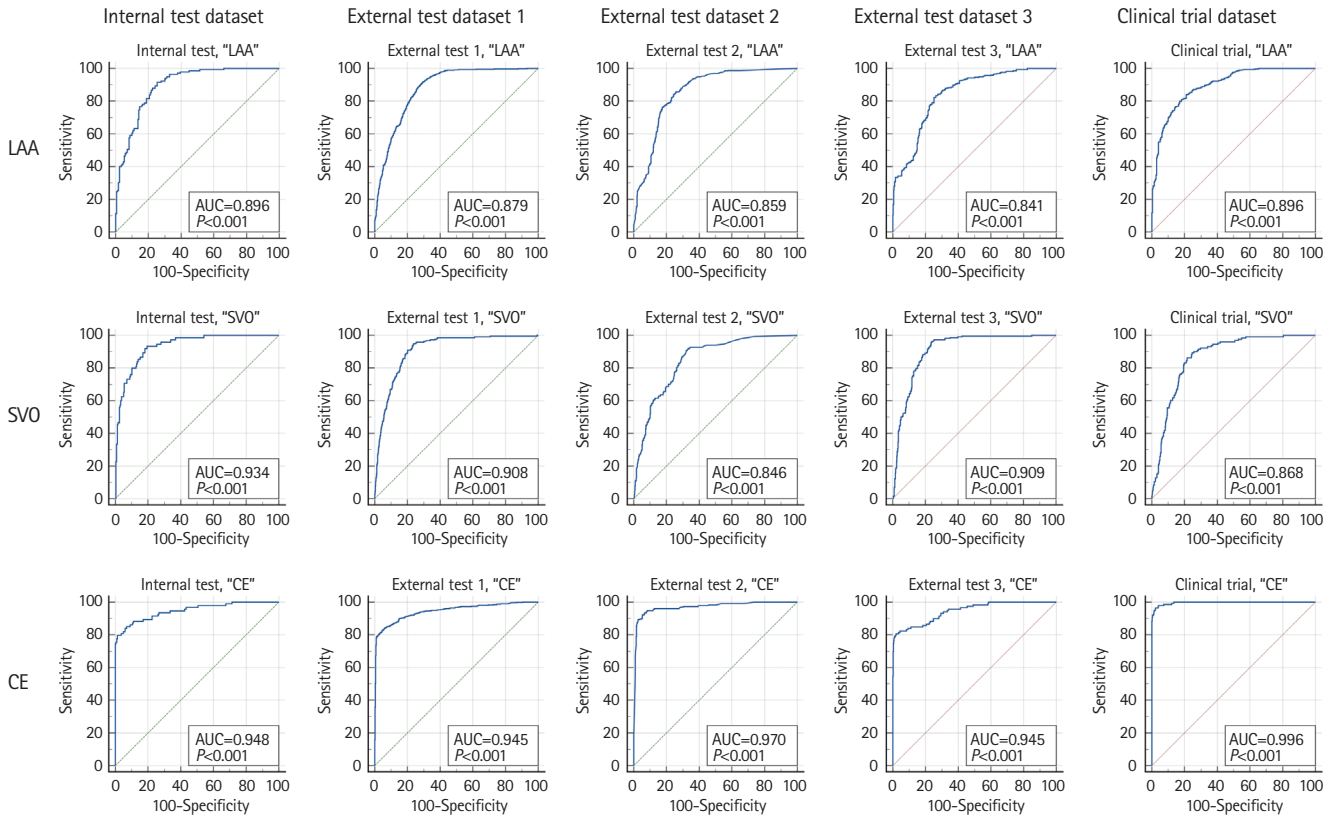


Figure 1. Receiver operating characteristic curves for subtype classification of ischemic stroke for by deep learning algorithm using DWIs and AF information. LAA, large artery atherosclerosis; AUC, area under the curve; SVO, small vessel occlusion; CE, cardioembolism; DWI, diffusion-weighted image; AF, atrial fibrillation.

Table 3. Confusion matrix for deep learning algorithm versus expert classification of stroke subtype using DWIs and AF information

	Internal test dataset				External test dataset 1				External test dataset 2				External test dataset 3				Clinical trial dataset				
	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	
	LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		LAA
Experts																	Experts' consensus				
LAA	106*	22	14		678*	179	187		291*	87	56		210*	49	41		LAA	300*	123	57	
SVO	14	60*	1		66	151*	4		50	95*	9		54	165*	3		SVO	19	110*	2	
CE	11	3	80*		61	7	586*		9	3	224*		17	3	99*		CE	0	2	136*	
Percentage agreement	79.1				73.7				74.0				73.9				Percentage agreement 72.9				
Sensitivity	0.75	0.80	0.85	0.80	0.65	0.68	0.90	0.74	0.67	0.62	0.95	0.75	0.70	0.74	0.83	0.76	Sensitivity	0.63	0.84	0.99	0.82
Specificity	0.85	0.89	0.93	0.89	0.85	0.89	0.85	0.86	0.85	0.87	0.89	0.87	0.79	0.88	0.92	0.86	Specificity	0.93	0.80	0.90	0.88
PPV	0.81	0.71	0.84	0.79	0.84	0.45	0.75	0.68	0.83	0.51	0.78	0.71	0.75	0.76	0.69	0.73	PPV	0.94	0.47	0.70	0.70
NPV	0.80	0.93	0.94	0.89	0.67	0.96	0.94	0.86	0.70	0.91	0.98	0.86	0.75	0.87	0.96	0.86	NPV	0.58	0.96	1.00	0.85
Accuracy	0.80	0.87	0.91	0.86	0.74	0.87	0.87	0.82	0.75	0.82	0.91	0.83	0.75	0.83	0.90	0.83	Accuracy	0.73	0.81	0.92	0.82

For each stroke subtype, sensitivity, specificity, PPV, NPV, and accuracy were evaluated. The average value of each statistic was shown in the last column. DWI, diffusion-weighted image; AF, atrial fibrillation; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; Avg, average; PPV, positive predictive value; NPV, negative predictive value; AI, artificial intelligence.

*The values indicate that the results of the AI algorithm align with those of the experts or the experts' consensus.

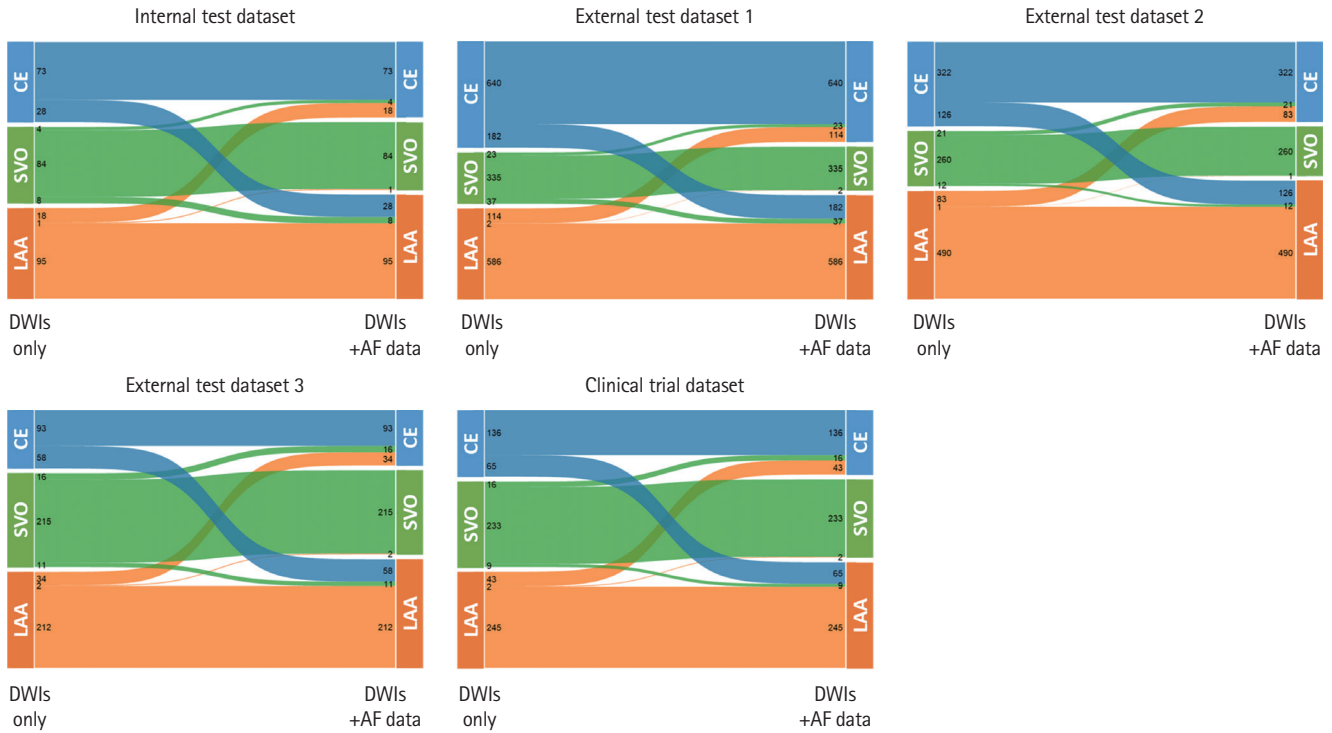


Figure 2. Alluvial plot depicting changes of stroke subtype classification after using AF data in addition to DWIs. Numbers indicates the number of patients in each stroke subtype. LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; DWI, diffusion-weighted image; AF, atrial fibrillation.

ing five cases with the highest CE probabilities (blue dotted square in the left Supplementary Figure 5B) showed that four had large wedge-shaped territorial infarcts (Supplementary Figure 5C-F) while the other one had multiple infarcts in the posterior circulation (Supplementary Figure 5G).

While stroke experts disagreed more on SVO strokes, the experts and the DWI+AF algorithm disagreed more on LAA strokes (Supplementary Table 2). Compared with LAA and LAA-NG strokes, LAA-BR and LAA-LC showed higher disagreements between the experts or between the experts' consensus and the DWI+AF algorithm-based subtyping (Supplementary Table 3).

In the external test dataset 3 and the clinical trial dataset, where information regarding the time of image acquisition within the 7-day period was available, there was no significant difference in the classification performance of either the DWI-only algorithm or the DWI+AF algorithm (vs. expert consensus) between the early (within 24 hours from last known well) and late (24 hours–7 days) imaging groups (Supplementary Table 4).

DWI-based prediction of CE

When we divided subjects into deciles of the expected CE probability (estimated by the DWI-only algorithm; Supplementary Table 5), the observed frequency of the CE subtype determined by experts increased with a nearly linear fashion ($P < 0.001$; Figure 3), showing good agreement. A similar trend was observed

in all external test datasets. In the 8th, 9th, and 10th decile groups, approximately 40%–70% of subjects were shown to have CE strokes. Furthermore, in the clinical trial dataset, there was a strong correlation between the expected probability and observed frequency ($P < 0.001$).

Discussion

In the present study, we developed a fully automated deep learning algorithm to classify ischemic stroke subtype using DWI and AF data from 2,998 ischemic stroke patients from three stroke centers. The deep learning algorithm was externally validated with three external datasets. The algorithm demonstrated good agreement with stroke experts, achieving Cohen's kappa coefficients of 0.57–0.59 for three external datasets, which were lower than the value (0.68) for the internal dataset. Furthermore, the clinical trial also demonstrated that the AI classification of stroke subtypes was comparable to the expert consensus.

To date, few studies have developed deep learning algorithms to classify stroke subtypes. According to a study that exclusively utilized electronic medical records, deep learning algorithms demonstrated moderate agreement (kappa=0.57) when compared with expert decisions.²⁵ Another study reported that a deep learning algorithm to classify stroke subtypes using DWI showed an average accuracy of 81.9%.²⁶ However, these investigations

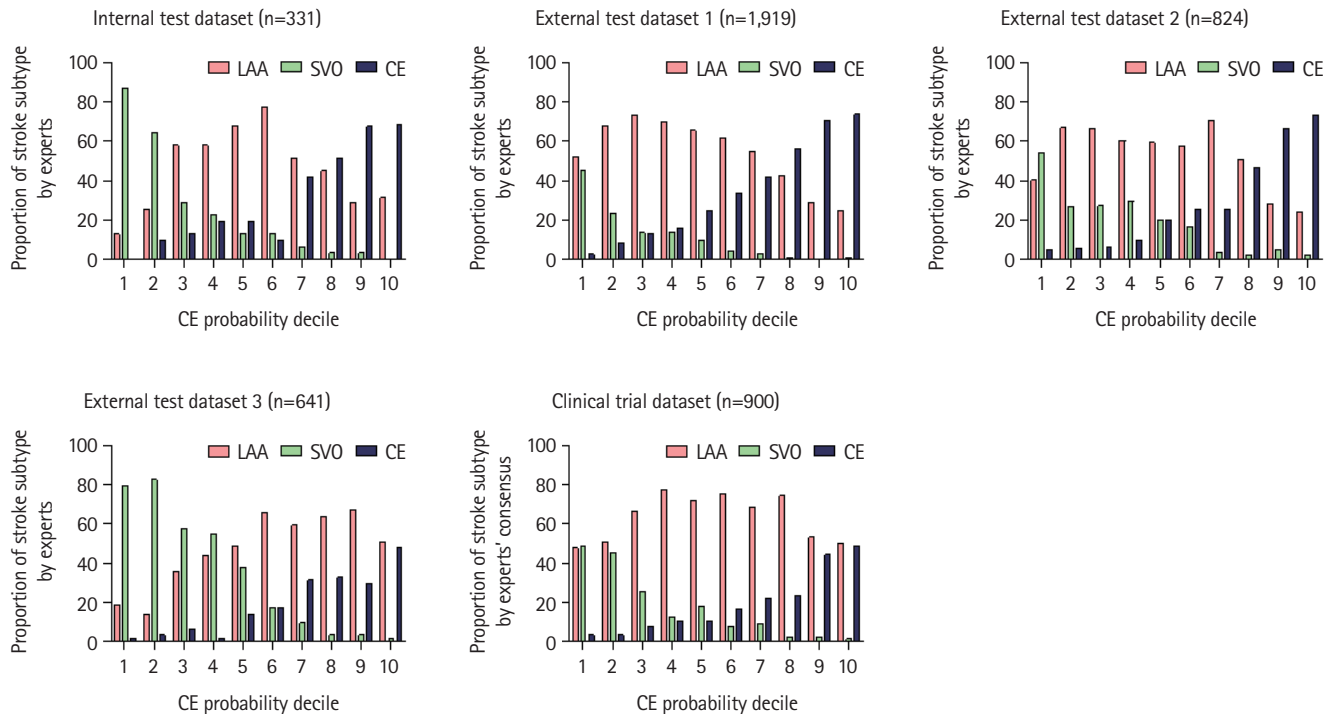


Figure 3. Proportions of stroke subtypes determined by experts in each decile of increasing CE probability that was estimated by the DWI-only based deep learning algorithm. Using DWIs only, a deep learning algorithm estimated probabilities of CE stroke. Then, the probabilities of every case were categorized into deciles in each dataset. Bars indicate observed frequency of each stroke subtype determined by expert or experts' consensus. Note that the proportion of CE stroke diagnosed rises proportionally with the estimated CE probability, suggesting that both human experts and the AI are examining the same underlying entity. LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; DWI, diffusion-weighted image; AF, atrial fibrillation; AI, artificial intelligence.

did not conduct an external validation. As described, the present study validated our deep learning algorithm in three different external datasets and in a clinical trial involving two hospitals. This represents the largest dataset and best external validation currently available in the literature, to our knowledge. In all datasets, the deep learning algorithm achieved a similarly high mean accuracy (between 0.82 and 0.83), supporting its robustness. It is notable that there was a comparable level of agreement between the consensus of experts and deep learning predictions as there was between the experts themselves.

Studies have demonstrated that stroke subtypes are closely associated with the pattern and extent of ischemic lesions.^{2,27,28} Cardioembolic strokes were associated with corticosubcortical single lesions, multiple lesions in anterior and posterior circulations, and multiple lesions in multiple cerebral circulations ($P=0.008$).² LAA stroke lesions were located more frequently in the same vascular territory than CE strokes.^{23,28} SVO stroke could be distinguished from other stroke subtypes based on distinctive morphological properties.²⁷ Thus, our deep learning algorithm trained on extensive DWI data may infer morphological and geometrical patterns associated with stroke etiologies. This could be one of the explanations for why, in the clinical trial dataset, the DWI+AF algorithm classified a CE stroke (with AF) as SVO and

65 LAA strokes (without AF) as CE.

The training dataset and the three external datasets included CE cases with AF or other potential cardiac embolic sources, while the clinical trial dataset did not include CE cases with potential cardiac embolic sources other than AF. Undetermined strokes, such as large infarcts with both relevant large artery stenosis and AF or single small subcortical infarcts with AF, were excluded from all the datasets including the training dataset. Intriguingly, as depicted by the alluvial plot, our AI algorithms trained on this training dataset classified some AF-positive cases as non-CE strokes in internal and external validations. Although further investigation is required, we speculate that the AI algorithms may have indeed identified AF-positive LAA strokes due to a covert source of artery-to-artery embolism such as non-significant (remnant) carotid stenosis or aortic atheroma, with AF acting as a bystander. Further studies are required to investigate how the presence or absence of vessel information, as well as AF information, affects our AI-based stroke subtyping, particularly in distinguishing between artery-to-artery embolism-mediated LAA and AF-mediated CE.

Guidelines for secondary prevention of stroke underscore a tailored therapeutic approach based on stroke subtypes,^{29,30} recommending strict blood pressure management for SVO strokes,³¹

intensive antiplatelet and lipid-lowering therapy for LAA strokes,³²⁻³⁵ and anticoagulant therapy for CE strokes.³⁶ However, a quarter of strokes are classified as embolic stroke with undetermined source (ESUS).³⁷ Repeated failures of randomized clinical trials to compare the effectiveness of antiplatelets and direct oral anticoagulants in preventing stroke in patients with ESUS³⁸⁻⁴⁰ have highlighted the need for new biomarkers or tools to identify people at high risk of CE stroke. A few machine learning algorithms using clinical and echocardiography data have demonstrated promising results in identifying individuals with an increased risk of AF within ESUS subjects.^{37,41} However, these algorithms relied on extensive data input such as patients' demographics, vascular risk factors, comorbidities, vital signs, laboratory results, and echocardiographic findings. The comprehensive data requirement poses a challenge in real-world scenarios, where data acquisition varies and resources are often limited. Our deep learning algorithm identified CE strokes based solely on DWI, suggesting its potential clinical utility in predicting an occult cardioembolic source in ESUS without additional clinical and laboratory data.

In the CRYSTAL-AF (Cryptogenic Stroke [CS] and Underlying AF) trial, stroke was classified as cryptogenic when the cause remained uncertain after extensive diagnostic evaluation, including 12-lead ECG, 24 hours or more of ECG monitoring, transesophageal echocardiography, angiographic or ultrasonographic evaluation of intracranial and extracranial vessels, and screening for thrombophilic states (in patients <55 years of age).¹⁴ In this study, ECG monitoring with an insertable cardiac monitor detected AF in 12.4% of patients by 1 year.¹⁴ We hypothesize that AI algorithms can increase the yield of testing, by helping to select patients who are more likely to test positive for AF during long-term ECG monitoring. To test the hypothesis, further research should investigate prospectively whether an occult cardioembolic source is more often found during post-ESUS or post-CS follow-up in patients with higher CE probabilities predicted by our DWI-only algorithm.

Including AF information changed the DWI-only algorithm-based original categorization of stroke subtype in about 20% of cases, which highlights the importance of detecting AF. In the NAVIGATE ESUS (New Approach Rivaroxaban Inhibition of Factor Xa in a Global Trial Versus ASA to Prevent Embolism in Embolic Stroke of Undetermined Source) trial, rivaroxaban failed to show superiority over aspirin in preventing recurrent ischemic stroke (4.7% per year in both groups).³⁹ It was suggested that the eligibility assessment may not have effectively identified strokes due to embolism and that AF was not a major cause of recurrent stroke.^{39,42} Indeed, AF was identified in only 3% of the patients at a median follow-up of 5 months, although systematic screening for arrhythmia was not performed during the trial.³⁹ How-

ever, the role of AF in patients with ESUS, whether it is the underlying cause of the index stroke or not, and its effect on stroke recurrence remain unclear,⁴³ requiring further investigations. In the NAVIGATE ESUS trial, about two-thirds of carotid plaques were present in the carotid artery ipsilateral to the index stroke, showing a strong trend of a higher risk of recurrent ischemic stroke.³⁹ Thus, future ESUS trials for direct oral anticoagulants may have to exclude strokes due to carotid atherosclerosis.⁴⁴ Our deep learning algorithms, which effectively classify stroke subtypes using DWIs with or without AF data, would facilitate these research, such as by improving eligibility assessments.

Many disparity studies have shown that primary and comprehensive stroke centers provide different levels of care for treating ischemic strokes. In a recent study involving 750,594 stroke patients from 1,474 stroke centers,⁴⁵ Chinese investigators found lower levels of care for quality measures such as thrombolysis, rehabilitation access, and medication at discharge, suggesting the need to increase the awareness on guideline-recommended treatments. In addition, a Korean study involving 10,399 patients from 201 healthcare facilities showed that approximately 40% of general hospitals provided relatively low quality stroke care (grade 3-5), while only one third of stroke patients received treatment at grade 1 hospitals.⁴⁶ Thus, we believe that our AI algorithms with accuracy comparable to stroke experts (working in grade 1 hospitals) may assist physicians in appropriately triaging stroke patients, particularly in hospitals with limited resources. The algorithms allow more expert guidance to be available to caregivers in resource-poor circumstances and should help to provide more optimal care to stroke patients.

Our study has limitations. First, stroke experts typically determine ischemic stroke etiology by using clinical, angiographic, and laboratory data in a comprehensive manner. The validity of relying solely on DWI and AF information could be questioned. However, an earlier study demonstrated that TOAST diagnoses without DWI matched final diagnoses in 48%, improving to 83% after DWI alone and to 94% after DWI plus magnetic resonance angiography,⁴⁷ indicating that DWI features has a major impact on classification accuracy enhancement. Second, although we validated the algorithm using multiple external datasets of Korean stroke populations, further investigation is required for multi-ethnic populations. Third, due to the rarity of other determined stroke subtype, deep learning algorithms to identify this subtype are challenging to develop. Thus, further investigation with a larger sample size is required.

Conclusions

In conclusion, our deep learning algorithm trained on a large da-

taset of DWI and AF information was able to classify ischemic stroke subtypes with a level of accuracy comparable to that of stroke experts. The AI algorithm, which performed well with the minimal data input in three different external test datasets and a multi-center clinical trial dataset, could be useful for stroke management by less experienced physicians or general practitioners.

Supplementary materials

Supplementary materials related to this article can be found online at <https://doi.org/10.5853/jos.2024.00535>.

Funding statement

This study was supported by the Multiminsty Grant for Medical Device Development (KMDF_PR_20200901_0098), the National Priority Research Center Program Grant (NRF-2021R1A6A1A03038865), and the Basic Science Research Program Grant (NRF-2020R1A2C3008295) of National Research Foundation, funded by the Korean government.

Conflicts of interest

Wi-Sun Ryu, Hoyoun Lee, and Dongmin Kim are employees of JLK Inc., Seoul, Korea. Dong-Eog Kim has stocks of JLK Inc.

Author contribution

Conceptualization: WSR, DEK. Study design: WSR, DK, DEK. Methodology: WSR, HL, DEK. Data collection: WSR, KJL, CKK, BJK, JTK, DHK, JKC, SIS, OYB, HJB, DEK. Investigation: WSR, JWC, JSL, DK, DEK. Statistical analysis: WSR, DEK. Writing—original draft: WSR, DS, DEK. Writing—review & editing: all authors. Funding acquisition: DK, HJB, DEK. Approval of final manuscript: all authors.

Acknowledgments

The authors appreciate the contributions of all members of the Clinical Research Collaboration for Stroke in Korea to this study.

References

1. Albers GW. Diffusion-weighted MRI for evaluation of acute stroke. *Neurology* 1998;51(3 Suppl 3):S47-S49.
2. Kang DW, Chalela JA, Ezzeddine MA, Warach S. Association of ischemic lesion patterns on early diffusion-weighted imaging with TOAST stroke subtypes. *Arch Neurol* 2003;60:1730-1734.
3. Chin YY, Sakinah H, Aryati A, Hassan BM. Prevalence, risk factors and secondary prevention of stroke recurrence in eight countries from South, East and Southeast Asia: a scoping review. *Med J Malaysia* 2018;73:90-99.
4. Flach C, Muruet W, Wolfe CDA, Bhalla A, Douiri A. Risk and secondary prevention of stroke recurrence: a population-based cohort study. *Stroke* 2020;51:2435-2444.
5. Ryu WS, Schellingerhout D, Hong KS, Jeong SW, Jang MU, Park MS, et al. White matter hyperintensity load on stroke recurrence and mortality at 1 year after ischemic stroke. *Neurology* 2019;93:e578-e589.
6. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;24:35-41.
7. Ko Y, Lee S, Chung JW, Han MK, Park JM, Kang K, et al. MRI-based algorithm for acute ischemic stroke subtype classification. *J Stroke* 2014;16:161-172.
8. Mamoli A, Censori B, Casto L, Sileo C, Cesana B, Camerlingo M. An analysis of the costs of ischemic stroke in an Italian stroke unit. *Neurology* 1999;53:112-116.
9. Jung WS, Seo KD, Suh SH. National trends in medical costs and prognosis of acute ischemic stroke patients in endovascular thrombectomy era: analysis using medical claim data in Korea. *Neurointervention* 2022;17:152-160.
10. Fang G, Xu P, Liu W. Automated ischemic stroke subtyping based on machine learning approach. *IEEE Access* 2020;8:18426-118432.
11. Zhang S, Wang J, Pei L, Liu K, Gao Y, Fang H, et al. Interpretable CNN for ischemic stroke subtype classification with active model adaptation. *BMC Med Inform Decis Mak* 2022;22:3.
12. Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chaturvedi V, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. *Stroke* 2001;32:1091-1098.
13. Buck BH, Hill MD, Quinn FR, Butcher KS, Menon BK, Gulamhusein S, et al. Effect of implantable vs prolonged external electrocardiographic monitoring on atrial fibrillation detection in patients with ischemic stroke: the PER DIEM randomized clinical trial. *JAMA* 2021;325:2160-2168.
14. Sanna T, Diener HC, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, et al. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med* 2014;370:2478-2486.
15. Kim BJ, Park JM, Kang K, Lee SJ, Ko Y, Kim JG, et al. Case characteristics, hyperacute treatment, and outcome information with TOAST stroke subtypes. *Arch Neurol* 2003;60:1730-1734.

- tion from the clinical research center for stroke-fifth division registry in South Korea. *J Stroke* 2015;17:38-53.
16. Noh YG, Ryu WS, Schellingerhout D, Park J, Chung J, Jeong SW, et al. Deep learning algorithms for automatic segmentation of acute cerebral infarcts on diffusion-weighted images: effects of training data sample size, transfer learning, and data features. medRxiv [Preprint]. 2023 [accessed July 2, 2023]. Available from: <https://doi.org/10.1101/2023.07.02.23292150>.
 17. Ryu WS, Kang YR, Noh YG, Park JH, Kim D, Kim BC, et al. Acute infarct segmentation on diffusion-weighted imaging using deep learning algorithm and RAPID MRI. *J Stroke* 2023; 25:425-429.
 18. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1979;9:62-66.
 19. Tan M, Le Q. EfficientNetV2: smaller models and faster training. In Lawrence N editor. Proceedings of Machine Learning Research. ICML 2021: Proceedings of the 38th International Conference on Machine Learning, 2021 Jul 18-24; Online. Norfolk, MA: Journal of Machine Learning Research; 2021. p. 10096-10106.
 20. 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 (CVPR); 2016 Jun 27-30; Las Vegas, NV, USA. New York: IEEE; 2016. p.770-778.
 21. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv [Preprint]. 2017 [accessed 2023 July 2]. Available from: <https://arxiv.org/abs/1704.04861>.
 22. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In Lawrence N editor. Proceedings of Machine Learning Research. ICML 2019: Proceedings of the 36th International Conference on Machine Learning, 2019 Jun 9-15; Long Beach, CA. Norfolk, MA: Journal of Machine Learning Research; 2019. p. 6105-6114.
 23. De Bruijn B. Revisiting the area under the ROC. In: Moen A, Andersen SK, Aarts J, Hurlen P. User centred networked health care (volume 169). Amsterdam: IOS Press, 2011;532-536.
 24. Cuzick J. A Wilcoxon-type test for trend. *Stat Med* 1985;4:87-90.
 25. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating ischemic stroke subtype classification using machine learning and natural language processing. *J Stroke Cerebrovasc Dis* 2019;28:2045-2051.
 26. Kim BK, Park S, Han MK, Hong JH, Lee DI, Yum KS. Deep learning for prediction of mechanism in acute ischemic stroke using brain diffusion magnetic resonance image. *J Neurocrit Care* 2023;16:85-93.
 27. Cheng B, Knaack C, Forkert ND, Schnabel R, Gerloff C, Thomalla G. Stroke subtype classification by geometrical descriptors of lesion shape. *PLoS One* 2017;12:e0185063.
 28. Kang DW, Kwon SU, Yoo SH, Kwon KY, Choi CG, Kim SJ, et al. Early recurrent ischemic lesions on diffusion-weighted imaging in symptomatic intracranial atherosclerosis. *Arch Neurol* 2007;64:50-54.
 29. Kleindorfer DO, Towfighi A, Chaturvedi S, Cockcroft KM, Gutierrez J, Lombardi-Hill D, et al. 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the American Heart Association/American Stroke Association. *Stroke* 2021;52:e364-e467.
 30. Dawson J, Béjot Y, Christensen LM, De Marchis GM, Dichgans M, Hagberg G, et al. European Stroke Organisation (ESO) guideline on pharmacological interventions for long-term secondary prevention after ischaemic stroke or transient ischaemic attack. *Eur Stroke J* 2022;7:I-II.
 31. Benavente OR, Coffey CS, Conwit R, Hart RG, McClure LA, Pearce LA, et al. Blood-pressure targets in patients with recent lacunar stroke: the SPS3 randomised trial. *Lancet* 2013;382: 507-515.
 32. Jing J, Meng X, Zhao X, Liu L, Wang A, Pan Y, et al. Dual antiplatelet therapy in transient ischemic attack and minor stroke with different infarction patterns: subgroup analysis of the CHANCE randomized clinical trial. *JAMA Neurol* 2018;75:711-719.
 33. Gao Y, Chen W, Pan Y, Jing J, Wang C, Johnston SC, et al. Dual antiplatelet treatment up to 72 hours after ischemic stroke. *N Engl J Med* 2023;389:2413-2424.
 34. Lee M, Ovbiagele B, Saver JL. Intensive medical management to prevent large and small artery atherothrombotic stroke: time to expand the horizon. *JAMA* 2021;326:217-218.
 35. Kim JS. Role of blood lipid levels and lipid-lowering therapy in stroke patients with different levels of cerebral artery diseases: reconsidering recent stroke guidelines. *J Stroke* 2021; 23:149-161.
 36. Sharma M, Cornelius VR, Patel JP, Davies JG, Molokhia M. Efficacy and harms of direct oral anticoagulants in the elderly for stroke prevention in atrial fibrillation and secondary prevention of venous thromboembolism: systematic review and meta-analysis. *Circulation* 2015;132:194-204.
 37. Hart RG, Catanese L, Perera KS, Ntaios G, Connolly SJ. Embolic stroke of undetermined source: a systematic review and clinical update. *Stroke* 2017;48:867-872.
 38. Poli S, Keller T, Martus P, Poli K, Ziemann U, Geisler T. The ATTICUS randomized controlled trial-subgroup analyses. *Stroke* 2023;54(Suppl 1):A31.
 39. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI,

- Berkowitz SD, et al. Rivaroxaban for stroke prevention after embolic stroke of undetermined source. *N Engl J Med* 2018; 378:2191-2201.
40. Diener HC, Sacco RL, Easton JD, Granger CB, Bernstein RA, Uchiyama S, et al. Dabigatran for prevention of stroke after embolic stroke of undetermined source. *N Engl J Med* 2019; 380:1906-1917.
41. Kamel H, Navi BB, Parikh NS, Merkle AE, Okin PM, Devereux RB, et al. Machine learning prediction of stroke mechanism in embolic strokes of undetermined source. *Stroke* 2020;51: e203-e210.
42. Veltkamp R, Pearce LA, Korompoki E, Sharma M, Kasner SE, Toni D, et al. Characteristics of recurrent ischemic stroke after embolic stroke of undetermined source: secondary analysis of a randomized clinical trial. *JAMA Neurol* 2020;77:1233-1240.
43. Kim AS, Kamel H, Bernstein RA, Manchanda M, Caprio FZ. Controversies in stroke: should patients with embolic stroke of undetermined source undergo intensive heart rhythm monitoring with an implantable loop recorder? *Stroke* 2022;53: 3243-3247.
44. Hong KS. Non-vitamin K antagonist oral anticoagulants in medical conditions at high risk of thromboembolism beyond atrial fibrillation. *J Stroke* 2019;21:259-275.
45. Liu Z, Gu H, Wei M, Feng X, Yu F, Feng J, et al. Comparison between healthcare quality in primary stroke centers and comprehensive stroke centers for acute stroke patients: evidence from the Chinese Stroke Center Alliance. *Lancet Reg Health West Pac* 2023;38:100863.
46. Kim J, Cho S, Lee H, Lee JY. [Nationwide acute stroke care quality and disparity in Korea: focusing on the type of health-care facilities and the socioeconomic status of patients]. *Public Health Aff* 2021;5:e6. Korean
47. Lee LJ, Kidwell CS, Alger J, Starkman S, Saver JL. Impact on stroke subtype diagnosis of early diffusion-weighted magnetic resonance imaging and magnetic resonance angiography. *Stroke* 2000;31:1081-1089.

Supplementary Methods

Magnetic resonance–based stroke subtype classification

The magnetic resonance (MR)–based stroke subtype classification that we previously developed¹⁸ is composed of the five following steps: (1) consideration of other determined etiology of stroke, (2) screening for small vessel occlusion (SVO) on diffusion-weighted image (DWI), (3) consideration of relevant artery stenosis or occlusion, (4) consideration of recanalization status after thrombolytic therapy, and (5) consideration of follow-up recanalization status without thrombolytic therapy.

Step 1. Consideration of other determined etiology of stroke

The other causes category includes patients with a diverse array of stroke mechanisms. Disorders included in this category are difficult to categorize into more homogenous groups. A patient who has a rare cause of ischemic stroke would be classified as "other determined cause" or "two or more undetermined causes (UD ≥ 2)," according to coexistence of other stroke etiology such as large artery atherosclerosis (LAA), SVO, and cardioembolism.

Step 2. Screening for SVO using DWI

A single lesion with the largest diameter of ≤ 20 mm in an axial DWI for penetrating artery infarction of the basal ganglia, corona radiata, thalamus, or pons would be classified as a SVO. If high-risk cardioembolic sources coexist, the subtype is classified as UD ≥ 2 . If accompanied by relevant stenosis of a corresponding cerebral artery on angiographic evaluation, including CT angiography, MR angiography, or conventional angiography, then it is classified as a "large artery atherosclerosis with lacunae." Infarctions in the midline extending from the base of the pons into the tegmentum without significant relevant artery stenosis would be classified as "branch atheromatous disease (LAA-BR)."

Step 3. Consideration of relevant artery stenosis or occlusion

Relevant arterial pathology was defined as stenosis or occlusion of arteries supplying the vascular territory of acute ischemic lesions detected on DWI. Stenosis less than 50% was also regarded as being relevant when clinical syndromes, lesions patterns on DWI, and new imaging techniques such as high-resolution wall imaging supported its relevance. In cases of a single lesion with the largest diameter >20 mm or multiple lesions with no steno-occlusion of relevant artery on angiographic evaluation, a possibility of cardioembolic stroke should be considered. "Exten-

sive embolic source evaluation," including 24-h Holter monitoring (24-h Holter), transthoracic echocardiography, and transesophageal echocardiography are recommended. Infarctions in which a definite cardioembolic source is not revealed despite a comprehensive work-up would be classified as "undetermined negative (UD-negative)." However, when relevant lesions are located at the anterior choroidal artery territory, single territory of cerebellum, or medulla oblongata, where SVOs do not seem to be causing infarctions, traditional MRI techniques cannot detect vascular pathologies of a relevant artery, and atherosclerosis may be a dominant vascular pathology, the infarctions are classified as "large artery atherosclerosis with normal angiography" instead of "UD-negative."

When a relevant pathology of a corresponding artery is observed, it is divided into stenosis and occlusion. If medical history or electrocardiography identifies high-risk cardioembolic sources with coexistence of relevant stenosis, then that infarction is classified as "UD ≥ 2 ." When there is evidence of chronic occlusion, or no or low risk cardioembolic source with relevant stenosis, it is classified as "LAA." Occlusion on pre-stroke angiographic evaluation, border zone infarction with clinical settings suggestive of hemodynamic failure, or recent (within 1 month of stroke onset) transient ischemic attack corresponding to occlusion site is considered as evidence for chronic occlusion.

Step 4. Consideration of recanalization status of occluded artery after recanalization therapy

If there is occlusion but no evidence of chronic occlusion and recanalization therapy, including mechanical thrombectomy, is performed, then the recanalization status after recanalization therapy should be considered. When residual stenosis exists, or angioplasty or stenting is performed for atherosclerotic stenosis, the underlying vascular pathology is considered primarily as atherosclerotic. When occlusion is resolved completely, comprehensive cardioembolic work-up is recommended. In this situation, low-risk cardioembolic sources are regarded as explaining the etiology of stroke.

Step 5. Consideration of follow-up recanalization status of occluded artery without recanalization therapy

When there is occlusion and recanalization therapy is not performed, follow-up angiographic evaluation is recommended and recanalization status on that evaluation guides further investigation and determination of stroke subtypes.

Supplementary Table 1. Confusion matrix of ischemic stroke subtype classification by deep learning algorithms versus experts using DWI only (DWI-only algorithm)

	Internal test dataset				External test dataset 1				External test dataset 2				External test dataset 3				Clinical trial dataset				
	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	Prediction			Avg.	
	LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		LAA	SVO	CE		
Experts																	Experts' consensus				
LAA	79*	26	37		518*	201	325		236*	95	103		154*	55	91		LAA	290*	151	133	
SVO	11	62*	2		41	165*	15		47	97*	10		48	170*	4		SVO	17	133*	5	
CE	24	8	62*		143	29	482*		60	16	160*		46	17	56*		CE	53	18	100*	
Percentage agreement	65.3				60.7				59.8				59.3				Percentage agreement 58.1				
Sensitivity	0.56	0.83	0.66	0.68	0.50	0.75	0.74	0.66	0.54	0.63	0.68	0.62	0.51	0.77	0.47	0.58	Sensitivity	0.51	0.86	0.58	0.65
Specificity	0.79	0.86	0.82	0.82	0.79	0.86	0.73	0.80	0.73	0.83	0.81	0.79	0.72	0.83	0.82	0.79	Specificity	0.79	0.77	0.81	0.79
PPV	0.69	0.65	0.61	0.65	0.74	0.42	0.59	0.58	0.69	0.47	0.59	0.58	0.62	0.70	0.37	0.56	PPV	0.81	0.44	0.42	0.56
NPV	0.68	0.94	0.85	0.82	0.57	0.96	0.84	0.79	0.59	0.91	0.86	0.79	0.63	0.87	0.87	0.79	NPV	0.47	0.96	0.89	0.78
Accuracy	0.68	0.85	0.77	0.77	0.63	0.85	0.73	0.74	0.63	0.80	0.77	0.73	0.63	0.81	0.75	0.73	Accuracy	0.61	0.79	0.77	0.72

For each stroke subtype, sensitivity, specificity, PPV, NPV, and accuracy were evaluated. The average value of each statistic was shown in the last column. DWI, diffusion-weighted image; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; Avg, average; PPV, positive predictive value; NPV, negative predictive value; AI, artificial intelligence.

*The values indicate that the results of the AI algorithm align with those of the experts or the experts' consensus.

Supplementary Table 2. Disagreement rates between experts and between experts' consensus and the DWI+AF algorithm after stratification by stroke subtypes in the clinical trial dataset

Consensus subtype	LAA	SVO	CE
Disagreement between experts	95/574 (16.6)	72/155 (46.5)	49/171 (28.7)
Disagreement between experts' consensus and the DWI+AF algorithm	204/574 (35.5)	19/155 (12.3)	1/171 (0.6)

DWI, diffusion weighted image; AF, atrial fibrillation; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism.

Supplementary Table 3. Sensitivity of the DWI+AF algorithm after stratification by subcategories of LAA

	LAA	LAA-BR	LAA-LC	LAA-NG	P
Internal test dataset*	60/78 (76.9)	4/8 (50.0)	4/7 (57.1)	9/14 (64.3)	0.001 [‡]
External test dataset 1 [†]	357/519 (68.8)	87/157 (55.4)	36/75 (48.0)	17/22 (77.3)	<0.001
External test dataset 2 [†]	209/287 (72.8)	59/106 (55.7)	21/37 (56.8)	2/3 (66.7)	<0.001 [‡]
Clinical trial dataset [§]	250/336 (74.4)	30/41 (73.2)	11/21 (52.3)	31/47 (66.0)	<0.001
Agreement between experts in the clinical trial dataset [§]	298/336 (88.7)	29/41 (70.7)	13/21 (61.9)	35/47 (74.5)	<0.001

Data are presented as numbers classified as LAA by the DWI+AF algorithm/numbers of LAA subgroup by an expert or experts' consensus (percentage). LAA, large artery atherosclerosis; LAA-BR, LAA-branch atheromatous disease; LAA-LC, LAA-lacune; LAA-NG, LAA-negative; DWI, diffusion-weighted image; AF, atrial fibrillation.

*Data were missing in 35 patients; [†]Data were missing in 271 patients; [‡]Data were missing in 1 patient; [§]Data were missing in 129 patients; [‡]Fisher's exact test was used.

Supplementary Table 4. Agreements of stroke subtype classification between deep learning algorithm and stroke neurologists (experts) after stratification by onset-to-imaging time

	External test dataset 3		Clinical trial dataset	
	Onset-to-imaging within 24 hours (n=412)	Onset-to-imaging between 24 hours–7 days (n=229)	Onset-to-imaging within 24 hours (n=595)	Onset-to-imaging between 24 hours–7 days (n=305)
	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts	Deep learning algorithm vs. experts' consensus	Deep learning algorithm vs. experts' consensus
DWI-only algorithm				
Percentage agreement (95% CI)	57.0 (52.2–61.8)	63.3 (57.0–69.6)	61.2 (57.3–65.1)	58.7 (53.1–64.3)
Cohen's kappa (95% CI)	0.35 (0.27–0.42)	0.40 (0.31–0.50)	0.40 (0.35–0.46)	0.38 (0.30–0.46)
DWI+AF algorithm				
Percentage agreement (95% CI)	73.5 (69.3–77.8)	74.7 (69.0–80.4)	75.6 (72.2–79.1)	74.1 (69.2–79.0)
Cohen's kappa (95% CI)	0.59 (0.53–0.66)	0.57 (0.48–0.67)	0.62 (0.57–0.67)	0.55 (0.47–0.63)
Between experts				
Percentage agreement (95% CI)			76.7 (73.2–80.1)	74.8 (69.9–79.7)
Cohen's kappa (95% CI)			0.60 (0.55–0.66)	0.52 (0.43–0.60)

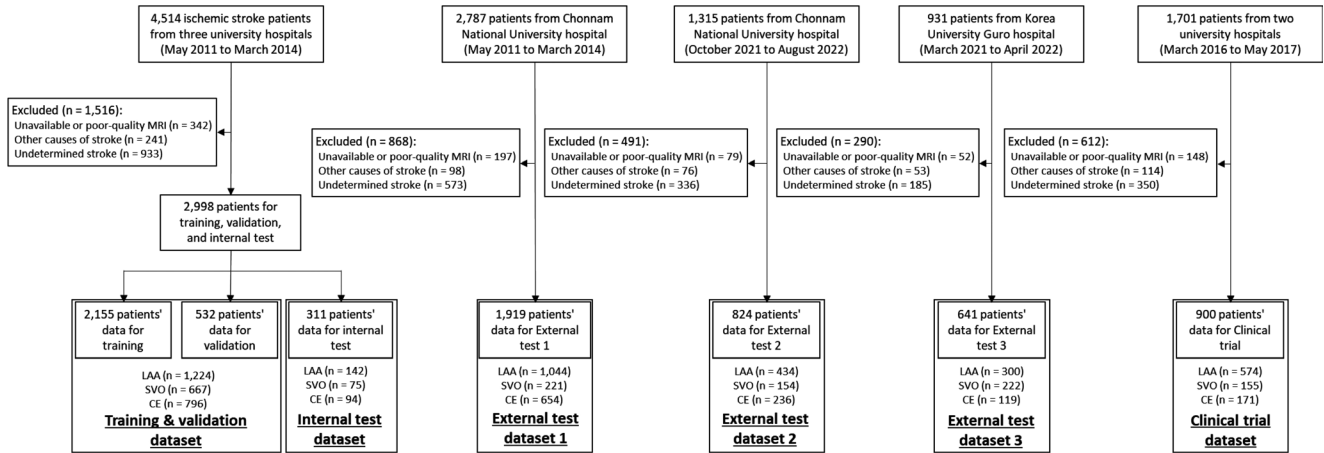
DWI, diffusion-weighted image; CI, confidence interval; AF, atrial fibrillation.

Supplementary Table 5. Mean probability of CE after decile stratification in the datasets for internal testing, external testing, and clinical trial

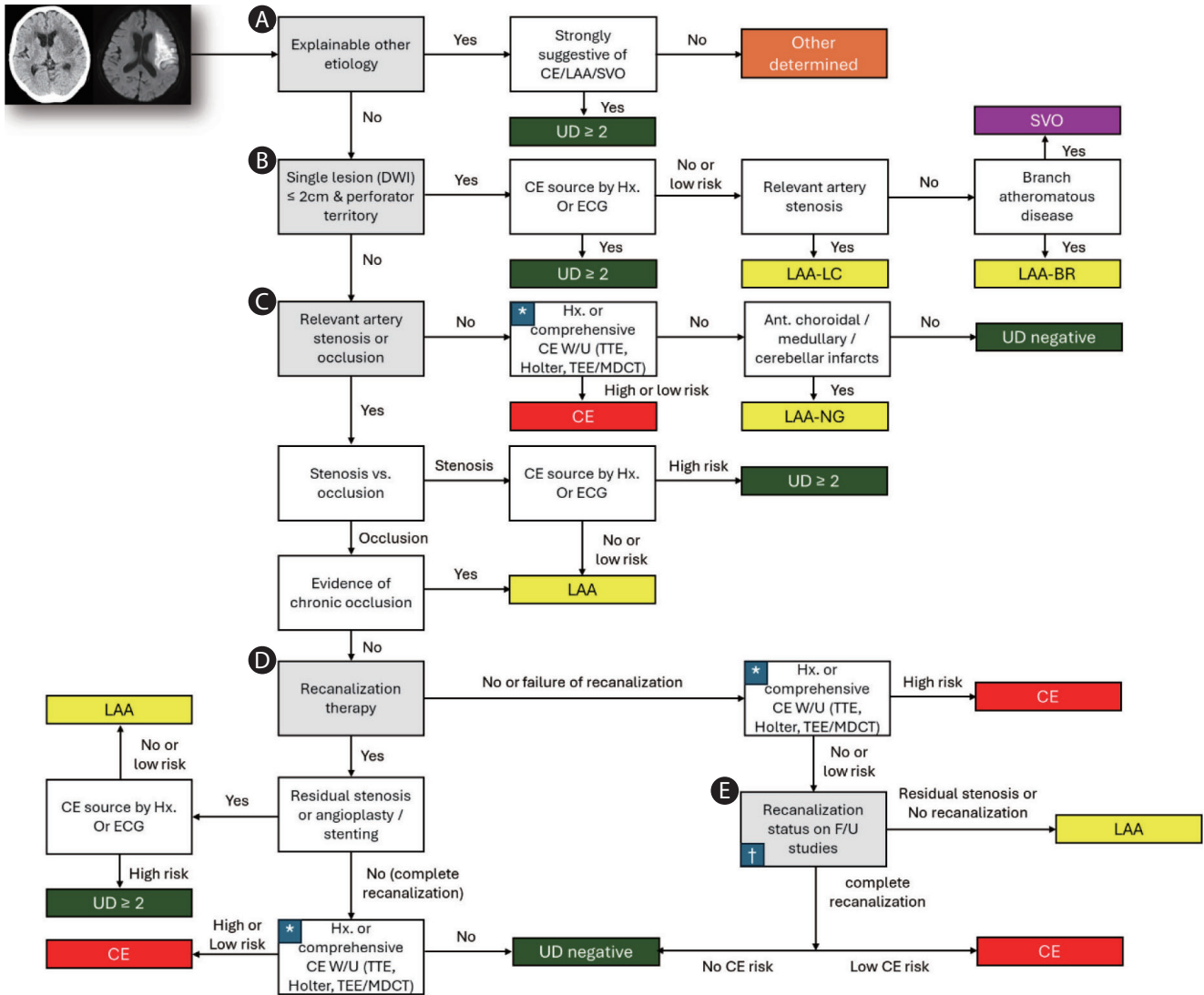
Decile	Probability of CE (%)				
	Internal test dataset	External test dataset 1	External test dataset 2	External test dataset 3	Clinical trial dataset
1	0.5 (0.2–0.8)	1.2 (0.3–2.1)	0.8 (0.3–1.2)	0.5 (0.2–0.8)	0.6 (0.2–0.9)
2	1.2 (0.8–1.6)	3.0 (2.1–4.0)	1.6 (1.3–2.1)	1.0 (0.8–1.2)	1.2 (0.9–1.7)
3	2.1 (1.6–2.5)	5.5 (4.0–7.4)	2.9 (2.2–3.8)	1.6 (1.2–2.1)	2.3 (1.7–2.9)
4	4.3 (2.6–6.5)	10.4 (7.5–14.4)	5.2 (3.9–7.1)	2.7 (2.1–3.5)	3.8 (3.0–5.0)
5	10.3 (6.5–15.4)	21.4 (14.4–30.3)	9.8 (7.2–13.2)	4.6 (3.5–6.0)	6.9 (5.1–9.7)
6	23.6 (15.8–31.1)	41.8 (30.3–54.3)	20.5 (13.4–30.0)	9.2 (6.1–14.0)	14.6 (9.7–20.2)
7	40.3 (31.3–51.7)	66.1 (54.4–75.5)	42.7 (30.2–56.2)	20.9 (14.3–28.1)	30.5 (20.2–40.3)
8	62.8 (51.8–72.3)	81.3 (75.5–85.8)	68.4 (57.2–78.7)	42.5 (28.3–58.2)	52.2 (40.7–64.0)
9	78.3 (72.6–86.5)	88.6 (85.8–90.7)	84.7 (79.1–89.3)	70.2 (58.4–79.5)	75.0 (64.4–83.0)
10	91.0 (86.6–96.1)	93.0 (90.7–96.9)	92.9 (89.5–97.8)	87.0 (79.5–96.0)	89.9 (83.3–96.0)

Data are presented as mean (minimum–maximum).

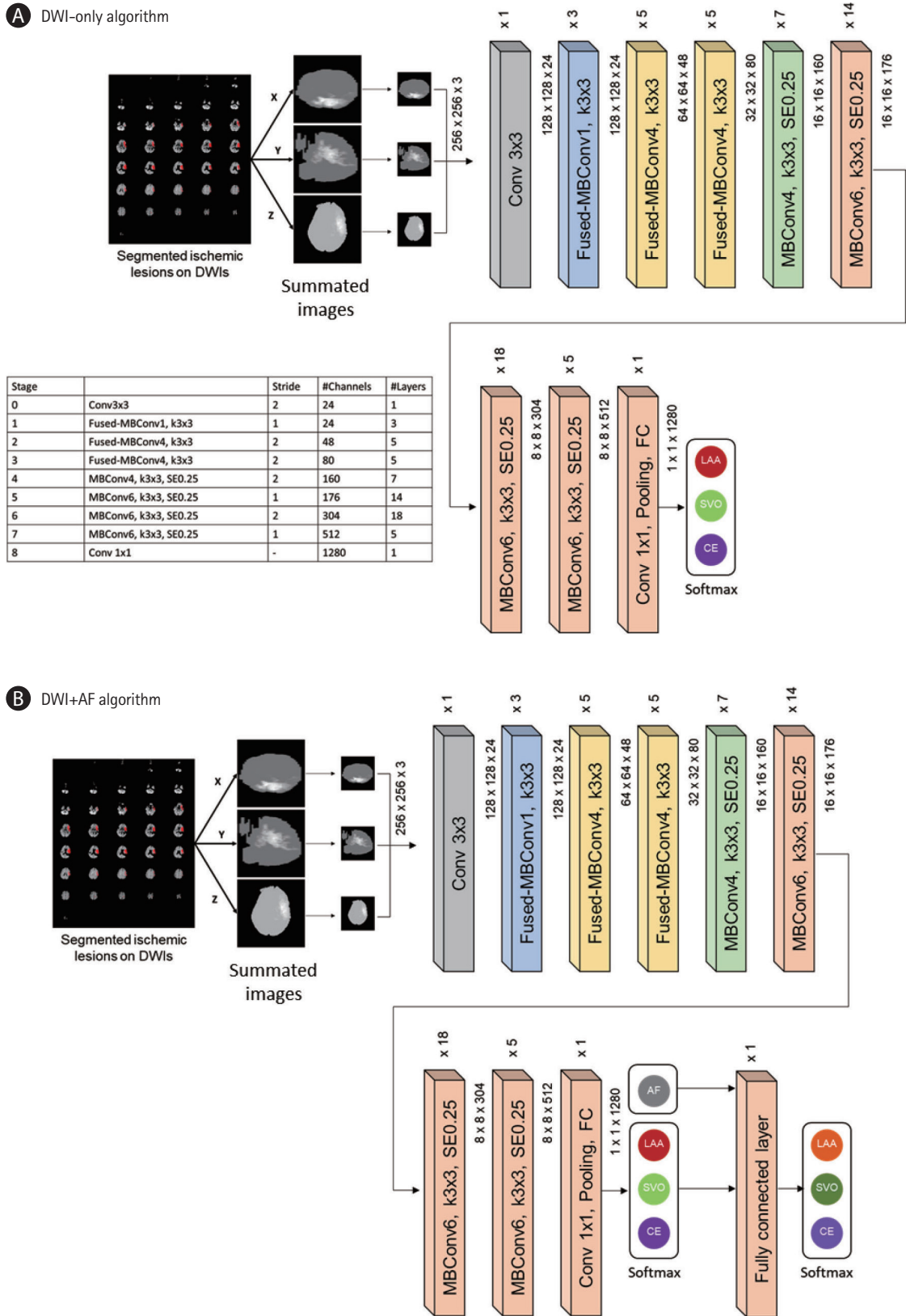
CE, cardioembolism.



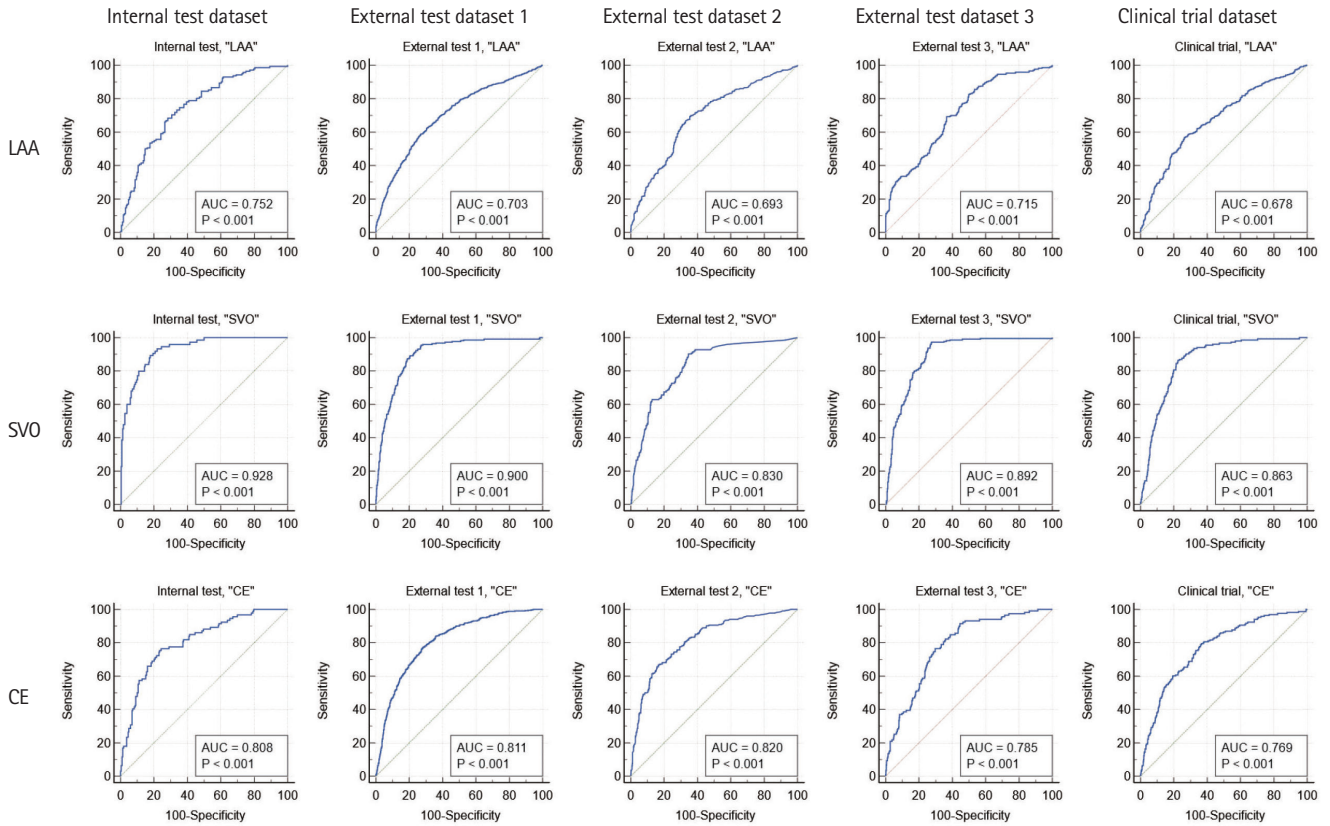
Supplementary Figure 1. Study flowchart. MRI, magnetic resonance imaging; LAA, large artery atherosclerosis; SVO, small vessel occlusion.



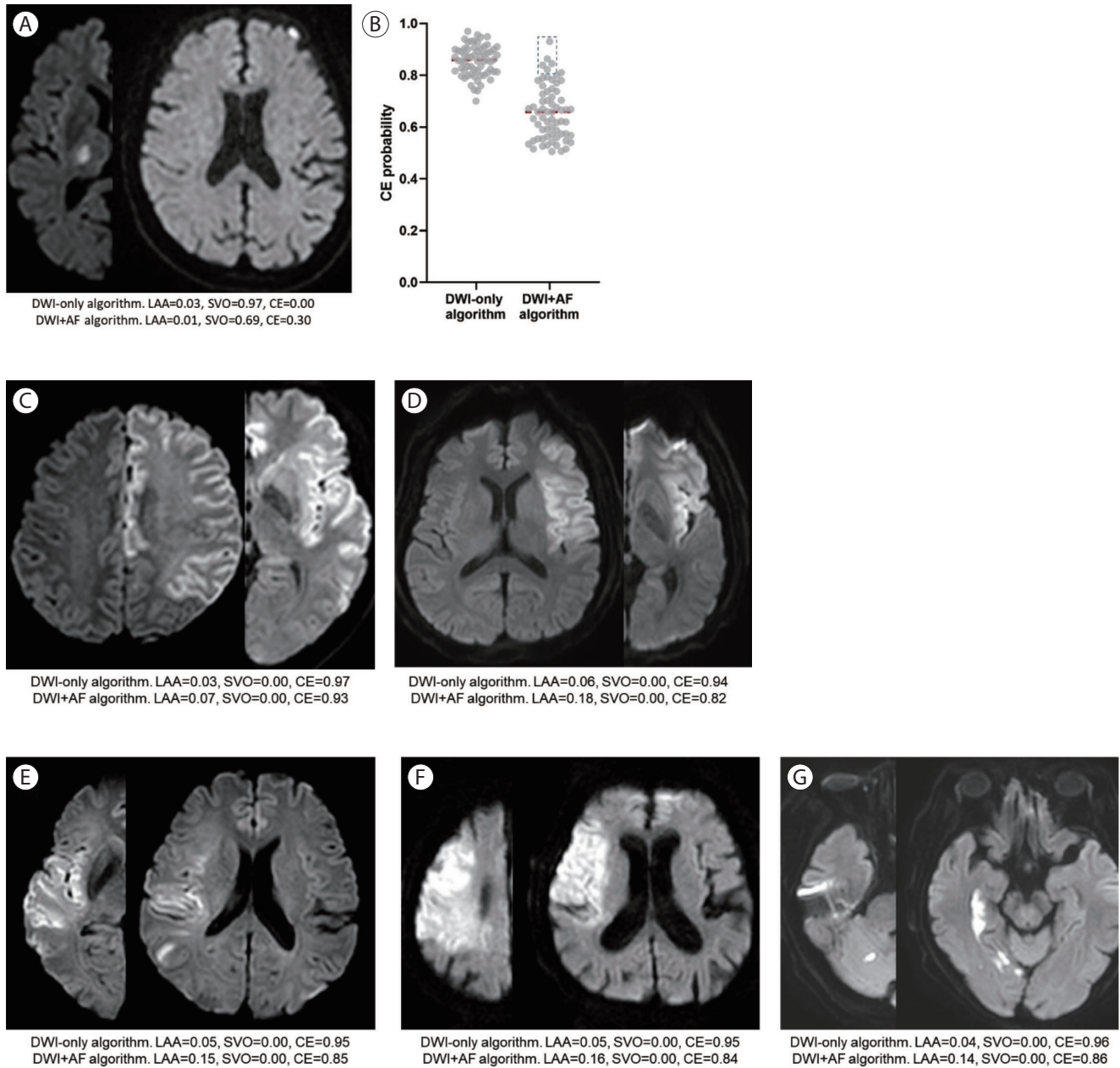
Supplementary Figure 2. MRI-based algorithm for subtype classification of ischemic stroke. (A) Step 1: consideration of other determined etiology of stroke. (B) Step 2: screening for SVO using MRI. (C) Step 3: consideration of relevant artery stenosis or occlusion. (D) Step 4: consideration of recanalization status of occluded artery after recanalization therapy. (E) Step 5: consideration of follow-up recanalization status of occluded artery without recanalization therapy. *If one of three examinations (TTE, 24-hr Holter monitoring, and TEE [or MDCT]) was not performed, then the patient was classified as 'undetermined incomplete'; †The follow-up vascular status would be evaluated by MR/CT angiography or transcranial Doppler. If no examinations are performed, then the patient should be classified as 'undetermined incomplete.' LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; UD, undetermined cause; UD ≥2, undetermined with two or more causes; DWI, diffusion weighted imaging; Hx, history; ECG, electrocardiography; LAA-LC, large artery atherosclerosis with lacunae; LAA-BR, branch atheromatous disease; W/U, work-up; TTE, transthoracic echocardiography; TEE, transesophageal echocardiography; MDCT, multi-detector row computed tomography; Ant, Anterior; LAA-NG, large artery atherosclerosis with normal angiography; F/U, follow-up; MRI magnetic resonance imaging. Modified with permission from Ko et al. *J Stroke* 2014;16:161-172.⁷



Supplementary Figure 3. Deep learning model to classify stroke subtype. We condensed the segmented infarct mask on DWI into three images in each of the X, Y, and Z-axis planes in order to ensure consistent input regardless of the number of DWI slices. Images were resized and then fed into EfficientNetV2 (A). The table in the lower left corner provided a summary of the details of each layer. The final layer generated probability for each subtype of stroke. To incorporate atrial fibrillation, we concatenated a binary value (0 or 1) representing its presence to previous outputs and then applied additional fully connected layers (B). DWI, diffusion-weighted image; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism; AF, atrial fibrillation.



Supplementary Figure 4. Receiver operating characteristic curves for subtype classification of ischemic stroke by the deep learning algorithm using DWIs. LAA, large artery atherosclerosis; AUC, area under the curve; SVO, small vessel occlusion; CE, cardioembolism; DWI, diffusion-weighted image.



Supplementary Figure 5. Representative cases misclassified by the DWI+AF algorithm in the clinical trial dataset. (A) A CE case misclassified as SVO by the DWI+AF algorithm, showing a single small infarct in the right striatocapsular area and a tiny cortical infarct. (B-G) Scatter plots showing probabilities of CE estimated by the DWI-only and the DWI+AF algorithms for cases misclassified as CE (B; gray dots and red bars represent individual cases and group means, respectively) and review of five cases (C-G) in the blue dotted squares in B (cases with highest CE probabilities on estimation by the DWI+AF algorithm). DWI, diffusion-weighted image; AF, atrial fibrillation; LAA, large artery atherosclerosis; SVO, small vessel occlusion; CE, cardioembolism.