

Deep learning-based automatic classification of ischemic stroke subtype using diffusion-weighted images

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

Affiliations

1. Department of Neurology, Dongguk University Ilsan Hospital, Goyang, Republic of Korea
2. Artificial Intelligence Research Center, JLK Inc., Seoul, Republic of Korea
3. Department of Neuroradiology and Imaging Physics, The University of Texas M.D. Anderson Cancer Center, Houston, USA
4. Department of Neurology, Korea University Guro Hospital, Seoul, Republic of Korea
5. Department of Neurology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
6. Department of Neurology, Samsung Medical Center, Sungkyunkwan University, Seoul, Republic of Korea
7. Department of Neurology, Asan Medical Center, Seoul, Republic of Korea
8. Department of Neurology, Chonnam National University Hospital, Gwangju, Republic of Korea
9. Department of Neurology, Dong-A University Hospital, Busan, Republic of Korea
10. Department of Radiology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
11. Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, South Korea.
12. National Priority Research Center for Stroke, Goyang, Republic of Korea

Correspondence to:

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Dong-Eog Kim, MD, PhD.

Professor of Neurology, Dongguk University College of Medicine & Chair of National Priority Research Center for Stroke.

Address: Department of Neurology, Dongguk University Ilsan Hospital, 27, Dongguk-ro, Ilsandong-gu, Goyang, South Korea.

E-mail: kdongeog@duih.org

Word Count: 5,987

Running title: Subtype classification by deep learning

Disclosure: Wi-Sun Ryu, Hoyoun Lee, and Dongmin Kim are employees of JLK Inc.

Acknowledgements: The authors appreciate the contributions of all members of the Comprehensive Registry Collaboration for Stroke in Korea to this study. This study was supported by the Multiministry 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.

Abstract

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

METHODS: Model training, validation, and internal testing were done in 2,988 patients with acute ischemic stroke from three stroke centers by using U-net for infarct segmentation and EfficientNetV2 for stroke subtype classification. Experienced vascular neurologists (n=5) determined stroke subtypes for external test datasets, while establishing a consensus for clinical trial datasets using the TOAST classification. Infarcts on DW images were automatically segmented using an artificial intelligence solution that we recently developed, and their masks were fed as inputs to a deep learning algorithm (DWI-only algorithm). Subsequently, another model was trained, with the presence or absence of AF included in the training as a categorical variable (DWI+AF algorithm). These models were tested: a) internally against the opinion of the labeling experts, b) against fresh external DWI data, and also c) against clinical trial DWI data acquired at a later date.

RESULTS: In the training-and-validation datasets, the mean age was 68.0 ± 12.5 (61.1% male). In internal testing, compared with the experts, the DWI-only algorithm and the DWI+AF algorithm 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 algorithm vs. 72.9% and 0.57 for the DWI+AF algorithm. The corresponding values between experts were comparable (76.0% and 0.61) to the DWI+AF algorithm.

CONCLUSIONS: Our deep learning algorithm trained on a large dataset of DWI (both with or without AF information) was able to classify ischemic stroke subtypes as accurately as a consensus of stroke experts.

77 **Non-standard Abbreviations and Acronyms**

78	DWI	diffusion-weighted MRI
79	LAA	large artery atherosclerosis
80	CE	cardioembolism
81	SVO	small vessel occlusion
82	TOAST	The Trial of Org10172 in Acute Stroke
83	ECG	electrocardiography
84	AF	atrial fibrillation
85	AI	artificial intelligence
86	NIHSS	National Institute of Health Stroke Scale
87	ESUS	Embolic Stroke with Undetermined Source
88	RCT	randomized clinical trial

Introduction

Studies have shown that the volume¹ and pattern² of ischemic lesions on diffusion-weighted MRI (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 cardio-embolic (CE) strokes were associated with a ~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 (CT/MRI), cardiac evaluation (electrocardiography, 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 electrocardiography (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 vs. neurovascular experts. Finally, we outlined practical applications of the deep learning-based stroke subtype classification for cardioembolism 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

Datasets 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) within 7 days of symptom onset. A total of 1,516 patients were excluded for the following reasons: (1) unavailable or poor-quality DWI (n = 342), (2) other causes of strokes such as arterial dissection, moyamoya disease, cancer-related stroke, etc. (n = 241), and (3) undetermined cause of stroke (n = 933). The remaining 2,998 patients' data were used for training, validation, and internal test, using random sub-setting in ratio of 7:2:1 (Figure S1). The institutional review board of Dongguk University Hospital approved the study protocol (IRB No. 2017-09-017), 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.

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 enrolled from the

Chonnam National University Hospital. After excluding 491 patients, 824 were finally included.

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 enrolled from Korea University Guro Hospital. After excluding 290 patients, 641 were finally included.

Clinical trial datasets

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 prespecified exclusion criteria, we excluded 612 patients according to the following reasons: (1) inadequate or poor-quality DWI (n = 148), other causes of strokes (n = 114), and undetermined causes of strokes (n = 350), leaving 900 patients' data 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 1000 s/mm², TE (echo time) 50-99 ms, TR (repetition time) 2400-9000 ms, voxel size 1×1×3~5 mm³, interslice gap of 0–2 mm, and slice thickness of 3–7mm. Using a validated 3D U-net algorithm, we automatically segmented infarct lesions on DW images.^{16,17}

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 materials).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 thrombolytic therapy; and 5) consideration of follow-up recanalization status without thrombolytic therapy. For the clinical trial dataset, stroke subtypes were determined through consensus among three experienced vascular neurologists (J.-W.Chung, C.K.Kim, 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 validated 3D U-net algorithm.^{16,17} The segmented infarct masks from raw DW images were stacked and condensed into three 2D X, Y, Z-axis images to ensure consistent data input regardless of the number of slices (Figure S2). These condensed 2D X, Y, Z-axis images were resized to 256 x 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).

For the training, we utilized the EfficientNet v2,¹⁹ 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 Figure S2. 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 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 ~9 hours using a hardware system comprising Intel Xeon Silver 4314 @2.40GHz, 640GB 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 five years and independently reviewed the brain MRI and patients' data. Information provided to the reviewers included age, sex, the presence of AF, DW images, and MR 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 ANOVA 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. In addition,

we calculated the sensitivity, specificity, positive predictive value, and negative predictive value for each subtype (LAA, SVO, and CE). To examine the clinical implications of artificial intelligence (AI) prediction of cardioembolism using DWI, participants in each dataset were stratified into ten groups based on the probability of having cardioembolism estimated by deep learning algorithm. The trend of the observed frequency of cardioembolic stroke, as determined by experts, was examined using a Wilcoxon-type test for trend.²⁰ All the statistical analyses described above were performed using STATA 16.0 (STATA Corp., Texas, USA), and a p-value < 0.05 was considered statistically significant.

Results

Baseline characteristics

In the training and validation datasets, the mean 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 (NIHSS) 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 vs. DWI plus AF data

In the internal test dataset (Figure 1), the percentage agreement between the DWI-only algorithm and stroke experts was 65.3% (95% CI: 60.0–70.6%). After incorporating the information regarding the presence of AF (DWI+AF algorithm), the percentage agreement was increased to 79.1% (95% CI: 74.6–83.6%).

In the external test datasets (Figure 1), both algorithms again showed good agreements. The DWI-only algorithm achieved 58.1 ~ 60.7% levels of agreements (Table 2, Figure S3). 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.

In the clinical trial dataset (Figure 1), 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.

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

DWI-based prediction of cardioembolism

When we divided subjects into deciles of the expected CE probability (estimated by the DWI-only algorithm; Table S1), the observed frequency of the CE subtype determined by experts increased with a nearly linear fashion (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.

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.²¹ Another study reported that a deep learning algorithm to classify stroke subtypes using DWI showed an average accuracy of 81.9%.²² However, these

investigations 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.^{2,23,24} 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, 24} SVO stroke could be distinguished from other stroke subtypes based on distinctive morphological properties.²³ Thus, our deep learning algorithm trained on extensive DWI data may infer morphological and geometrical patterns associated with stroke etiologies.

Guidelines for secondary prevention of stroke underscore a tailored therapeutic approach based on stroke subtypes,^{25,26} recommending strict blood pressure management for SVO strokes,²⁷ intensive antiplatelet and lipid-lowering therapy for LAA strokes,²⁸⁻³¹ and anticoagulant therapy for CE strokes.³² However, a quarter of strokes are classified as embolic stroke with undetermined source (ESUS).³³ Repeated failures of randomized clinical trials (RCTs) to compare the effectiveness of antiplatelets and direct oral anticoagulants in preventing stroke in patients with ESUS³⁴⁻³⁶ 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.^{33,37} 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 one 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).³⁵ 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.^{35,38} 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.³⁵ However, 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,³⁹ 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.³⁵ Thus, future ESUS trials for direct oral anticoagulants may have to exclude strokes due to carotid atherosclerosis.⁴⁰ Our deep learning algorithms, which effectively classify stroke subtypes using DW images with or without AF data, would facilitate these research, such as by improving eligibility assessments.

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. 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 MRA,⁴¹ 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.

In conclusion, our deep learning algorithm trained on a large dataset of DWI and AF information was able to classify ischemic stroke subtypes as accurate as 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.

Author Contribution: W-.S.Ryu, and D-.E. Kim had full access to all study data and take responsibility for data integrity and the accuracy of the data analysis. W-.S.Ryu, D.S. and D-.E. Kim contributed study's concept and design and drafted the manuscript. K.-J.Lee, C.K.Kim, B.J.Kim, J-.W.Chung, J-.S.Lim, J-.T.Kim, D-.H.Kim, and J-.K.Cha contributed to the acquisition, analyzing, and interpretation of data. W-.S.Ryu contributed statistical analysis and data visualization. All authors contributed to the critical revisions of the manuscript for important intellectual content.

References

1. Albers GW. Diffusion-weighted MRI for evaluation of acute stroke. *Neurology*. 1998;51:S47-49.
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-Base Cohort Study. *Stroke*. 2020;51:2435-2444.
5. Ryu WS, Schellingerhout D, Hong KS, Jeong SW, Jang MU, Park MS, Choi KH, Kim JT, Kim BJ, Lee J, 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, Marsh EE, 3rd. 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, Park TH, Park SS, Cho YJ, Hong KS, 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;8:118426-118432.
11. Zhang S, Wang J, Pei L, Liu K, Gao Y, Fang H, Zhang R, Zhao L, Sun S, Wu J, 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, Chilukuri V, Armstrong SB, Horner RD. 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, Siddiqui M, Coutts SB, Jeerakathil T, Smith EE, 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, Rymer MM, Thijs V, Rogers T, Beckers F, 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, Cha JK, Kim DH, Nah HW, Han MK, et al. Case characteristics, hyperacute treatment, and outcome information from the clinical research center for stroke-fifth division registry in South Korea. *J Stroke*. 2015;17:38-53.
16. Noh Y-G, Ryu W-S, Schellingerhout D, Park J, Chung J, Jeong S-W, Gwak D-S, Kim BJ, Kim J-T, Hong K-S. 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*. 2023:2023.2007. 2002.23292150.
17. Ryu WS, Kang YR, Noh YG, Park JH, Kim D, Kim BC, Park MS, Kim BJ, Kim JT. 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 Transactions on Systems, Man, and Cybernetics*. 1979;9:62-66.
19. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. Paper/Poster presented at: International conference on machine learning; 2021;
20. Cuzick J. A Wilcoxon-type test for trend. *Statistics in medicine*. 1985;4:87-90.
21. 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.
22. Park S, Kim B-k, Han M-K, Hong J-H, Yum KS, Lee D-I. Deep Learning for Prediction of Mechanism in Acute Ischemic Stroke Using Brain MRI. *Research Square*. 2021.
23. 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.
24. Kang DW, Kwon SU, Yoo SH, Kwon KY, Choi CG, Kim SJ, Koh JY, Kim JS. Early recurrent ischemic lesions on diffusion-weighted imaging in symptomatic intracranial atherosclerosis. *Arch Neurol*. 2007;64:50-54.
25. Kleindorfer DO, Towfighi A, Chaturvedi S, Cockcroft KM, Gutierrez J, Lombardi-Hill D, Kamel H, Kernan WN, Kittner SJ, Leira EC, 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.

26. Dawson J, Bejot Y, Christensen LM, De Marchis GM, Dichgans M, Hagberg G, Heldner MR, Milionis H, Li L, Pezzella FR, 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.
27. Group SPSS, Benavente OR, Coffey CS, Conwit R, Hart RG, McClure LA, Pearce LA, Pergola PE, Szychowski JM. Blood-pressure targets in patients with recent lacunar stroke: the SPS3 randomised trial. *Lancet*. 2013;382:507-515.
28. Jing J, Meng X, Zhao X, Liu L, Wang A, Pan Y, Li H, Wang D, Johnston SC, Wang Y, Wang Y. 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.
29. Gao Y, Chen W, Pan Y, Jing J, Wang C, Johnston SC, Amarenco P, Bath PM, Jiang L, Yang Y, et al. Dual Antiplatelet Treatment up to 72 Hours after Ischemic Stroke. *N Engl J Med*. 2023;389:2413-2424.
30. 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.
31. 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.
32. 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.
33. 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.
34. Poli S, Keller T, Martus P, Poli K, Ziemann U, Geisler T, Committee S, investigators A. The ATTICUS Randomized Controlled Trial-Subgroup Analyses. *Stroke*. 2023;54:A31-A31.
35. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI, Berkowitz SD, Swaminathan B, Lavados P, Wang Y, Wang Y, et al. Rivaroxaban for Stroke Prevention after Embolic Stroke of Undetermined Source. *N Engl J Med*. 2018;378:2191-2201.
36. Diener HC, Sacco RL, Easton JD, Granger CB, Bernstein RA, Uchiyama S, Kreuzer J, Cronin L, Cotton D, Grauer C, et al. Dabigatran for Prevention of Stroke after Embolic Stroke of Undetermined Source. *N Engl J Med*. 2019;380:1906-1917.
37. Kamel H, Navi BB, Parikh NS, Merkle AE, Okin PM, Devereux RB, Weinsaft JW, Kim J, Cheung

490 JW, Kim LK, et al. Machine Learning Prediction of Stroke Mechanism in Embolic Strokes of
 491 Undetermined Source. *Stroke*. 2020;51:e203-e210.

492 38. Veltkamp R, Pearce LA, Korompoki E, Sharma M, Kasner SE, Toni D, Ameriso SF, Mundl H,
 493 Tatlisumak T, Hankey GJ, et al. Characteristics of Recurrent Ischemic Stroke After Embolic
 494 Stroke of Undetermined Source: Secondary Analysis of a Randomized Clinical Trial. *JAMA*
 495 *Neurol*. 2020;77:1233-1240.

496 39. Kim AS, Kamel H, Bernstein RA, Manchanda M, Caprio FZ. Controversies in Stroke: Should
 497 Patients With Embolic Stroke of Undetermined Source Undergo Intensive Heart Rhythm
 498 Monitoring With an Implantable Loop Recorder? *Stroke*. 2022;53:3243-3247.

499 40. Hong KS. Non-Vitamin K Antagonist Oral Anticoagulants in Medical Conditions at High Risk of
 500 Thromboembolism beyond Atrial Fibrillation. *J Stroke*. 2019;21:259-275.

501 41. Lee LJ, Kidwell CS, Alger J, Starkman S, Saver JL. Impact on stroke subtype diagnosis of early
 502 diffusion-weighted magnetic resonance imaging and magnetic resonance angiography.
 503 *Stroke*. 2000;31:1081-1089.

504 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 value
Age	68.0 ± 12.5	68.2 ± 12.9	68.7 ± 12.0	70.2 ± 12.4	67.5 ± 12.0	68.6 ± 12.4	0.23
Sex, men	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 ^a	4 (2 – 9)	4 (2 – 9)	4 (2 – 10)	2 (1 – 5)	3 (1 – 5)	4 (2 – 7)	< 0.001 ^b
Prestroke mRS score 0~2 ^a	2,346 (87.3%)	274 (88.1%)	1,667 (87.9%)	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

505 ^aData were missing in 472 and 1 patients of the external test dataset-3 and clinical trial dataset, respectively. ^bKruskal-Wallis test was used. NIHSS=National
506 Institute of Health Stroke Scale; mRS=modified Rankin Scale ; LAA=large artery atherosclerosis; SVO=small vessel occlusion; CE=cardioembolism.

507 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%)	74.0% (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)

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

Figure Legends

Figure 1. Confusion matrix for deep learning algorithm vs. expert classification of stroke subtype using diffusion-weighted images and atrial fibrillation information

For each stroke subtype, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were evaluated. The average value of each statistic was shown in the last column. LAA=large artery atherosclerosis; SVO=small vessel occlusion; CE=cardioembolism; Avg=average.

Figure 2. Alluvial plot depicting changes of stroke subtype classification after using atrial fibrillation (AF) data in addition to diffusion-weighted images (DWIs)

Numbers indicates the number of patients in each stroke subtype. LAA=large artery atherosclerosis; SVO=small vessel occlusion; CE=cardioembolism.

Figure 3. Proportions of stroke subtypes determined by experts in each decile of increasing cardioembolic (CE) probability that was estimated by the DWI-only based deep learning algorithm

Using diffusion-weighted images 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 experts. 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.

Internal test dataset						
		Prediction			Avg.	
		LAA	SVO	CE		
Experts	LAA	106	22	14		
	SVO	14	60	1		
	CE	11	3	80		
Percentage agreement		79.1				
Sensitivity		0.75	0.80	0.85	0.80	
Specificity		0.85	0.89	0.93	0.89	
PPV		0.81	0.71	0.84	0.79	
NPV		0.80	0.93	0.94	0.89	
Accuracy		0.80	0.87	0.91	0.86	

External test dataset 1						
		Prediction			Avg.	
		LAA	SVO	CE		
Experts	LAA	678	179	187		
	SVO	66	151	4		
	CE	61	7	586		
Percentage agreement		73.7				
Sensitivity		0.65	0.68	0.90	0.74	
Specificity		0.85	0.89	0.85	0.86	
PPV		0.84	0.45	0.75	0.68	
NPV		0.67	0.96	0.94	0.86	
Accuracy		0.74	0.87	0.87	0.82	

External test dataset 2						
		Prediction			Avg.	
		LAA	SVO	CE		
Experts	LAA	291	87	56		
	SVO	50	95	9		
	CE	9	3	224		
Percentage agreement		74.0				
Sensitivity		0.67	0.62	0.95	0.75	
Specificity		0.85	0.87	0.89	0.87	
PPV		0.83	0.51	0.78	0.71	
NPV		0.70	0.91	0.98	0.86	
Accuracy		0.75	0.82	0.91	0.83	

External test dataset 3						
		Prediction			Avg.	
		LAA	SVO	CE		
Experts	LAA	210	49	41		
	SVO	54	165	3		
	CE	17	3	99		
Percentage agreement		73.9				
Sensitivity		0.70	0.74	0.83	0.76	
Specificity		0.79	0.88	0.92	0.86	
PPV		0.75	0.76	0.69	0.73	
NPV		0.75	0.87	0.96	0.86	
Accuracy		0.75	0.83	0.90	0.83	

Clinical trial dataset						
		Prediction			Avg.	
		LAA	SVO	CE		
Experts' consensus	LAA	300	123	57		
	SVO	19	110	2		
	CE	0	2	136		
Percentage agreement		72.9				
Sensitivity		0.63	0.84	0.99	0.82	
Specificity		0.93	0.80	0.90	0.88	
PPV		0.94	0.47	0.70	0.70	
NPV		0.58	0.96	1.00	0.85	
Accuracy		0.73	0.81	0.92	0.82	

Internal test dataset



External test dataset 1



External test dataset 2



DWIs
only

DWIs
+ AF data

External test dataset 3



DWIs
only

DWIs
+ AF data

Clinical trial dataset



DWIs
only

DWIs
+ AF data

