

1 Deep learning-based automatic classification of ischemic stroke subtype 2 using diffusion-weighted images

3

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

9

10 Affiliations

- 11 1. Department of Neurology, Dongguk University Ilsan Hospital, Goyang, Republic of
12 Korea
- 13 2. Artificial Intelligence Research Center, JLK Inc., Seoul, Republic of Korea
- 14 3. Department of Neuroradiology and Imaging Physics, The University of Texas M.D.
15 Anderson Cancer Center, Houston, USA
- 16 4. Department of Neurology, Korea University Guro Hospital, Seoul, Republic of Korea
- 17 5. Department of Neurology, Seoul National University Bundang Hospital, Seongnam,
18 Republic of Korea
- 19 6. Department of Neurology, Samsung Medical Center, Sungkyunkwan University,
20 Seoul, Republic of Korea
- 21 7. Department of Neurology, Asan Medical Center, Seoul, Republic of Korea
- 22 8. Department of Neurology, Chonnam National University Hospital, Gwangju, Republic
23 of Korea
- 24 9. Department of Neurology, Dong-A University Hospital, Busan, Republic of Korea
- 25 10. Department of Radiology, Seoul National University Bundang Hospital, Seongnam,
26 Republic of Korea
- 27 11. Department of Radiology, Korea University Guro Hospital, Korea University College
28 of Medicine, Seoul, South Korea.
- 29 12. National Priority Research Center for Stroke, Goyang, Republic of Korea

30

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~~NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.~~

32 Dong-Eog Kim, MD, PhD.

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

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

37 E-mail: kdongeog@duih.org

38

39 Word Count: 5,987

40 Running title: Subtype classification by deep learning

41

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

43

44 **Acknowledgements:** The authors appreciate the contributions of all members of the
45 Comprehensive Registry Collaboration for Stroke in Korea to this study. This study was
46 supported by the Multiministry Grant for Medical Device Development
47 (KMDF_PR_20200901_0098), the National Priority Research Center Program Grant (NRF-
48 2021R1A6A1A03038865), and the Basic Science Research Program Grant (NRF-
49 2020R1A2C3008295) of National Research Foundation, funded by the Korean government.

50 **Abstract**

51

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

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

66 **RESULTS:** In the training-and-validation datasets, the mean age was 68.0 ± 12.5 (61.1% male).
67 In internal testing, compared with the experts, the DWI-only algorithm and the DWI+AF
68 algorithm respectively achieved moderate (65.3%) and near-strong (79.1%) agreement. In
69 external testing, both algorithms again showed good agreements (59.3-60.7% and 73.7-
70 74.0%, respectively). In the clinical trial dataset, compared with the expert consensus,
71 percentage agreements and Cohen's kappa were respectively 58.1% and 0.34 for the DWI-
72 only algorithm vs. 72.9% and 0.57 for the DWI+AF algorithm. The corresponding values
73 between experts were comparable (76.0% and 0.61) to the DWI+AF algorithm.

74 **CONCLUSIONS:** Our deep learning algorithm trained on a large dataset of DWI (both with or
75 without AF information) was able to classify ischemic stroke subtypes as accurately as a
76 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

89 Introduction

90

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

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

111 To date, a few previous studies have developed automated systems for classifying
112 stroke subtypes using deep learning algorithms and DWI.^{10,11} However, no study has
113 externally validated these algorithms, which is critically important given the low inter-rater
114 reliability in the classification of stroke subtypes.¹² In the present multi-center study, we
115 enrolled about 6,500 patients with acute ischemic stroke. Using 2,489 patients' DWI data
116 with and without information on the presence of AF, we developed a deep learning
117 algorithm to classify stroke subtypes. We then externally validated the deep learning
118 algorithm on a new set of 3,384 patients, using three temporally and regionally different
119 datasets. In addition, we compared stroke subtype classifications by the deep learning

120 algorithm vs. neurovascular experts. Finally, we outlined practical applications of the deep
121 learning-based stroke subtype classification for cardioembolism risk stratification based
122 solely on initial DWI assessments, for use when AF information is not available or becomes
123 available after continuous ECG monitoring (for days ~ years).^{13,14}

124

125

126 **Methods**

127 **Participants**

128 ***Datasets for training, validation, and internal testing***

129 From May 2011 to March 2014, we consecutively enrolled 4,514 patients with acute
130 ischemic stroke, who were admitted to three university hospitals (Dongguk University
131 Hospital, Seoul National University Bundang Hospital, and Dong-A University Hospital)
132 within 7 days of symptom onset. A total of 1,516 patients were excluded for the following
133 reasons: (1) unavailable or poor-quality DWI (n = 342), (2) other causes of strokes such as
134 arterial dissection, moyamoya disease, cancer-related stroke, etc. (n = 241), and (3)
135 undetermined cause of stroke (n = 933). The remaining 2,998 patients' data were used for
136 training, validation, and internal test, using random sub-setting in ratio of 7:2:1 (Figure S1).
137 The institutional review board of Dongguk University Hospital approved the study protocol
138 (IRB No. 2017-09-017), and patients or their legal proxies provided a written informed
139 consent.

140

141 ***Datasets for external testing***

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

144 **External test dataset 1** From May 2011 to March 2014, 2,787 patients with acute ischemic
145 stroke who were admitted within 7 days of symptom onset were consecutively enrolled
146 from Chonnam National University Hospital. After excluding 868 patients, 1,919 were finally
147 included.

148 **External test dataset 2** From October 2021 to August 2022, 1,315 patients with acute
149 ischemic stroke who were admitted within 7 days of symptom onset were enrolled from the

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

152 **External test dataset 3** From March 2021 to April 2022, 931 patients with acute ischemic
153 stroke who were admitted within 7 days of symptom onset were enrolled from Korea
154 University Guro Hospital. After excluding 290 patients, 641 were finally included.

155

156 ***Clinical trial datasets***

157 A pivotal clinical trial was conducted to assess the efficacy of deep learning algorithms in
158 comparison to a standard reference established through expert consensus, and to measure
159 the level of agreement between the deep learning algorithm and the consensus as well as
160 among the experts themselves. From March 2016 to May 2017, 1,701 patients who met the
161 following inclusion criteria were enrolled from the two stroke centers (Dongguk University
162 Hospital and Seoul National University Bundang Hospital): 1) age between 20 and 95 years,
163 2) patients with acute ischemic stroke who visited the hospitals within 7 days after symptom
164 onset, and 3) patients who underwent DWI. According to the prespecified exclusion criteria,
165 we excluded 612 patients according to the following reasons: (1) inadequate or poor-quality
166 DWI (n = 148), other causes of strokes (n = 114), and undetermined causes of strokes (n =
167 350), leaving 900 patients' data for clinical testing.

168

169 **Clinical data collection**

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

173

174 **Imaging acquisition and infarct segmentation**

175 For the training data, brain MRIs were performed on 1.5 Tesla (n = 2,471) or 3.0 Tesla (n =
176 527) MRI systems. The DWI protocol was as follows: b-values of 0 and 1000 s/mm², TE
177 (echo time) 50-99 ms, TR (repetition time) 2400-9000 ms, voxel size 1×1×3~5 mm³, interslice
178 gap of 0–2 mm, and slice thickness of 3–7mm. Using a validated 3D U-net algorithm, we
179 automatically segmented infarct lesions on DW images.^{16,17}

180

181 **Ischemic stroke subtype classification**

182 For the datasets for training and validation, internal testing, and external test datasets 1-3,
183 stroke subtypes were determined by experienced vascular neurologists at each hospital,
184 using a validated MRI-based classification system built on the TOAST criteria (details
185 provided in supplementary materials).⁷ Briefly, the modified TOAST classification is
186 composed of the following five steps: 1) consideration of other determined etiologies of
187 stroke; 2) screening for SVO on DWI; 3) consideration of relevant artery stenosis or
188 occlusion; 4) consideration of recanalization status after thrombolytic therapy; and 5)
189 consideration of follow-up recanalization status without thrombolytic therapy. For the
190 clinical trial dataset, stroke subtypes were determined through consensus among three
191 experienced vascular neurologists (J-.W.Chung, C.K.Kim, and D-.E.Kim).

192

193 **Development of a deep learning algorithm for ischemic stroke subtype classification**

194 Brain DWIs were preprocessed by (1) skull stripping using the Gaussian blur and Otsu's
195 threshold,¹⁸ (2) applying N4 bias field correction using the SimpleITK library, and (3)
196 performing image signal normalization. After the preprocessing, infarct areas on DWI were
197 automatically segmented using the validated 3D U-net algorithm.^{16,17} The segmented infarct
198 masks from raw DW images were stacked and condensed into three 2D X, Y, Z-axis images
199 to ensure consistent data input regardless of the number of slices (Figure S2). These
200 condensed 2D X, Y, Z-axis images were resized to 256 x 256 pixels using bilinear
201 interpolation. Thus, the training data for the algorithm was comprised of: three 2D images
202 representing X, Y, Z-axis projections of segmented infarct area, and a label (LAA, SVO, and
203 CE).

204

205 For the training, we utilized the EfficientNet v2,¹⁹ a new family of convolutional networks
206 that have faster training speed and better parameter efficiency, while adding a
207 `global_average_pooling2d` layer to minimize overfitting by reducing the total number of
208 parameters. In addition, we incorporated a sequence of one inner dense layer with dropout
209 layers. In total, a 30% dropout rate was randomly chosen to avoid overfitting. Finally, one
210 output dense layer contained 3 output units for multi-class (LAA, SVO, and CE) classification,
211 which were designated as the DWI-only based subtype classification. The details of the

212 layers, their order in the proposed model, and the output shape of each layer are presented
213 in Figure S2. The total number of parameters was 52,862,199.

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

218 For all the procedures, including preprocessing and model development, we used
219 Python 3.7.9 and 3.8.13, PyTorch 1.12.0, Torchvision 0.13.0, pandas 1.2.4, NumPy
220 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
221 2.1.2. Each model was trained for ~9 hours using a hardware system comprising Intel Xeon
222 Silver 4314 @2.40GHz, 640GB RAM, and NVIDIA Quadro RTX A6000 with 48GB GDDR6.

223

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

225 For the clinical trial dataset, we first assessed the inter-observer agreement of stroke
226 subtype classification between two experts (J-W Chung and J-S Lim, board-certified
227 neurologists with more than 5-year experience in both stroke practice and research), who
228 had served as stroke neurologists at least five years and independently reviewed the brain
229 MRI and patients' data. Information provided to the reviewers included age, sex, the
230 presence of AF, DW images, and MR or CT angiography. Based on the aforementioned
231 ischemic stroke subtype-classification system,⁷ they independently determined etiologies
232 (i.e., LAA, SVO, or CE). In cases of disagreement between the two reviewers, a third reviewer
233 (D-E Kim) served as the tie-breaker. When the final consensus on stroke subtype was
234 undetermined or other determined stroke, the case was excluded from the analysis. The
235 experts' consensus classifications were compared with the deep learning algorithm's
236 classifications.

237

238 **Statistical analysis**

239 The baseline characteristics between datasets were compared using the ANOVA or Kruskal-
240 Wallis test for continuous variables and Chi-square test for categorical variables, as
241 appropriate. To compare the subtype classifications made by experts and those made by
242 deep learning algorithms, we used percentage agreement and Cohen's kappa. In addition,

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

251

252

253 **Results**

254 ***Baseline characteristics***

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

260

261 ***Deep learning prediction of stroke subtype using DWI data only vs. DWI plus AF data***

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

266 In the external test datasets (Figure 1), both algorithms again showed good
267 agreements. The DWI-only algorithm achieved 58.1 ~ 60.7% levels of agreements (Table 2,
268 Figure S3). The DWI+AF algorithm again showed higher agreements, ranging from 73.7% to
269 74.0%, with Cohen's Kappa ranging from 0.57 to 0.59. In addition, the accuracy of stroke
270 subtype classification reached 0.83.

271 In the clinical trial dataset (Figure 1), the percentage agreements and Cohen's
272 kappa were respectively 58.1% (95% CI: 54.9–61.3%) and 0.34 (0.29–0.39) for the DWI-only

273 algorithm, and the values were 72.9% (95% CI: 69.1–76.7%) and 0.57 (0.51–0.62) for the
274 DWI+AF algorithm, respectively.

275 Alluvial plots for the five datasets (Figure 2) showed that additional information
276 regarding the presence of AF on ECG changed the categorization of stroke subtype by the
277 DWI-only algorithm from CE to LAA more often (22.1~38.2%) than from LAA to CE
278 (13.7~16.2%) or from SVO to CE (4.2~7.2%). There was no reclassification from CE to SVO.

279

280 ***DWI-based prediction of cardioembolism***

281 When we divided subjects into deciles of the expected CE probability (estimated by the
282 DWI-only algorithm; Table S1), the observed frequency of the CE subtype determined by
283 experts increased with a nearly linear fashion (Figure 3), showing good agreement. A similar
284 trend was observed in all external test datasets. In the 8th, 9th, and 10th decile groups,
285 approximately 40-70% of subjects were shown to have CE strokes. Furthermore, in the
286 clinical trial dataset, there was a strong correlation between the expected probability and
287 observed frequency.

288

289

290 **Discussion**

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

299 To date, few studies have developed deep learning algorithms to classify stroke
300 subtypes. According to a study that exclusively utilized electronic medical records, deep
301 learning algorithms demonstrated moderate agreement (kappa = 0.57) when compared
302 with expert decisions.²¹ Another study reported that a deep learning algorithm to classify
303 stroke subtypes using DWI showed an average accuracy of 81.9%.²² However, these

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

312 Studies have demonstrated that stroke subtypes are closely associated with the
313 pattern and extent of ischemic lesions.^{2,23,24} Cardioembolic strokes were associated with
314 corticosubcortical single lesions, multiple lesions in anterior and posterior circulations, and
315 multiple lesions in multiple cerebral circulations ($P = 0.008$).² LAA stroke lesions were
316 located more frequently in the same vascular territory than CE strokes.^{23, 24} SVO stroke
317 could be distinguished from other stroke subtypes based on distinctive morphological
318 properties.²³ Thus, our deep learning algorithm trained on extensive DWI data may infer
319 morphological and geometrical patterns associated with stroke etiologies.

320 Guidelines for secondary prevention of stroke underscore a tailored therapeutic
321 approach based on stroke subtypes,^{25,26} recommending strict blood pressure management
322 for SVO strokes,²⁷ intensive antiplatelet and lipid-lowering therapy for LAA strokes,²⁸⁻³¹ and
323 anticoagulant therapy for CE strokes.³² However, a quarter of strokes are classified as
324 embolic stroke with undetermined source (ESUS).³³ Repeated failures of randomized clinical
325 trials (RCTs) to compare the effectiveness of antiplatelets and direct oral anticoagulants in
326 preventing stroke in patients with ESUS³⁴⁻³⁶ have highlighted the need for new biomarkers
327 or tools to identify people at high risk of CE stroke. A few machine learning algorithms using
328 clinical and echocardiography data have demonstrated promising results in identifying
329 individuals with an increased risk of AF within ESUS subjects.^{33,37} However, these algorithms
330 relied on extensive data input such as patients' demographics, vascular risk factors,
331 comorbidities, vital signs, laboratory results, and echocardiographic findings. The
332 comprehensive data requirement poses a challenge in real-world scenarios, where data
333 acquisition varies and resources are often limited. Our deep learning algorithm identified CE

334 strokes based solely on DWI, suggesting its potential clinical utility in predicting an occult
335 cardioembolic source in ESUS without additional clinical and laboratory data.

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

347 Including AF information changed the DWI-only algorithm-based original
348 categorization of stroke subtype in about 20% of cases, which highlights the importance of
349 detecting AF. In the NAVIGATE ESUS (New Approach Rivaroxaban Inhibition of Factor Xa in a
350 Global Trial Versus ASA to Prevent Embolism in Embolic Stroke of Undetermined Source)
351 trial, rivaroxaban failed to show superiority over aspirin in preventing recurrent ischemic
352 stroke (4.7% per year in both groups).³⁵ It was suggested that the eligibility assessment may
353 not have effectively identified strokes due to embolism and that AF was not a major cause
354 of recurrent stroke.^{35,38} Indeed, AF was identified in only 3% of the patients at a median
355 follow-up of 5 months, although systematic screening for arrhythmia was not performed
356 during the trial.³⁵ However, the role of AF in patients with ESUS, whether it is the underlying
357 cause of the index stroke or not, and its effect on stroke recurrence remain unclear,³⁹
358 requiring further investigations. In the NAVIGATE ESUS trial, about two-thirds of carotid
359 plaques were present in the carotid artery ipsilateral to the index stroke, showing a strong
360 trend of a higher risk of recurrent ischemic stroke.³⁵ Thus, future ESUS trials for direct oral
361 anticoagulants may have to exclude strokes due to carotid atherosclerosis.⁴⁰ Our deep
362 learning algorithms, which effectively classify stroke subtypes using DW images with or
363 without AF data, would facilitate these research, such as by improving eligibility
364 assessments.

365 Our study has limitations. First, stroke experts typically determine ischemic stroke
366 etiology by using clinical, angiographic, and laboratory data in a comprehensive manner.
367 The validity of relying solely on DWI and AF information could be questioned. An earlier
368 study demonstrated that TOAST diagnoses without DWI matched final diagnoses in 48%,
369 improving to 83% after DWI alone and to 94% after DWI plus MRA,⁴¹ indicating that DWI
370 features has a major impact on classification accuracy enhancement. Second, although we
371 validated the algorithm using multiple external datasets of Korean stroke populations,
372 further investigation is required for multi-ethnic populations.

373 In conclusion, our deep learning algorithm trained on a large dataset of DWI and AF
374 information was able to classify ischemic stroke subtypes as accurate as stroke experts. The
375 AI algorithm, which performed well with the minimal data input in three different external
376 test datasets and a multi-center clinical trial dataset, could be useful for stroke management
377 by less experienced physicians or general practitioners.

378

379

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

387

388 References

- 389 1. Albers GW. Diffusion-weighted MRI for evaluation of acute stroke. *Neurology*. 1998;51:S47-
390 49.
- 391 2. Kang DW, Chalela JA, Ezzeddine MA, Warach S. Association of ischemic lesion patterns on
392 early diffusion-weighted imaging with TOAST stroke subtypes. *Arch Neurol*. 2003;60:1730-
393 1734.
- 394 3. Chin YY, Sakinah H, Aryati A, Hassan BM. Prevalence, risk factors and secondary prevention of
395 stroke recurrence in eight countries from south, east and southeast asia: a scoping review.
396 *Med J Malaysia*. 2018;73:90-99.
- 397 4. Flach C, Muruet W, Wolfe CDA, Bhalla A, Douiri A. Risk and Secondary Prevention of Stroke
398 Recurrence: A Population-Base Cohort Study. *Stroke*. 2020;51:2435-2444.
- 399 5. Ryu WS, Schellingerhout D, Hong KS, Jeong SW, Jang MU, Park MS, Choi KH, Kim JT, Kim BJ,
400 Lee J, et al. White matter hyperintensity load on stroke recurrence and mortality at 1 year
401 after ischemic stroke. *Neurology*. 2019;93:e578-e589.
- 402 6. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE, 3rd.
403 Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical
404 trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35-41.
- 405 7. Ko Y, Lee S, Chung JW, Han MK, Park JM, Kang K, Park TH, Park SS, Cho YJ, Hong KS, et al. MRI-
406 based Algorithm for Acute Ischemic Stroke Subtype Classification. *J Stroke*. 2014;16:161-172.
- 407 8. Mamoli A, Censori B, Casto L, Sileo C, Cesana B, Camerlingo M. An analysis of the costs of
408 ischemic stroke in an Italian stroke unit. *Neurology*. 1999;53:112-116.
- 409 9. Jung WS, Seo KD, Suh SH. National Trends in Medical Costs and Prognosis of Acute Ischemic
410 Stroke Patients in Endovascular Thrombectomy Era: Analysis Using Medical Claim Data in
411 Korea. *Neurointervention*. 2022;17:152-160.
- 412 10. Fang G, Xu P, Liu W. Automated ischemic stroke subtyping based on machine learning
413 approach. *IEEE Access*. 2020;8:118426-118432.
- 414 11. Zhang S, Wang J, Pei L, Liu K, Gao Y, Fang H, Zhang R, Zhao L, Sun S, Wu J, et al. Interpretable
415 CNN for ischemic stroke subtype classification with active model adaptation. *BMC Med
416 Inform Decis Mak*. 2022;22:3.
- 417 12. Goldstein LB, Jones MR, Matchar DB, Edwards LJ, Hoff J, Chilukuri V, Armstrong SB, Horner
418 RD. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in
419 Acute Stroke Treatment (TOAST) criteria. *Stroke*. 2001;32:1091-1098.
- 420 13. Buck BH, Hill MD, Quinn FR, Butcher KS, Menon BK, Gulamhusein S, Siddiqui M, Coutts SB,
421 Jeerakathil T, Smith EE, et al. Effect of Implantable vs Prolonged External Electrocardiographic

- 422 Monitoring on Atrial Fibrillation Detection in Patients With Ischemic Stroke: The PER DIEM
423 Randomized Clinical Trial. *JAMA*. 2021;325:2160-2168.
- 424 14. Sanna T, Diener HC, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, Rymer MM, Thijs V,
425 Rogers T, Beckers F, et al. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med*.
426 2014;370:2478-2486.
- 427 15. Kim BJ, Park JM, Kang K, Lee SJ, Ko Y, Kim JG, Cha JK, Kim DH, Nah HW, Han MK, et al. Case
428 characteristics, hyperacute treatment, and outcome information from the clinical research
429 center for stroke-fifth division registry in South Korea. *J Stroke*. 2015;17:38-53.
- 430 16. Noh Y-G, Ryu W-S, Schellingerhout D, Park J, Chung J, Jeong S-W, Gwak D-S, Kim BJ, Kim J-T,
431 Hong K-S. Deep learning algorithms for automatic segmentation of acute cerebral infarcts on
432 diffusion-weighted images: Effects of training data sample size, transfer learning, and data
433 features. *medRxiv*. 2023:2023.2007. 2002.23292150.
- 434 17. Ryu WS, Kang YR, Noh YG, Park JH, Kim D, Kim BC, Park MS, Kim BJ, Kim JT. Acute Infarct
435 Segmentation on Diffusion-Weighted Imaging Using Deep Learning Algorithm and RAPID MRI.
436 *J Stroke*. 2023;25:425-429.
- 437 18. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on*
438 *Systems, Man, and Cybernetics*. 1979;9:62-66.
- 439 19. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. Paper/Poster presented at:
440 International conference on machine learning; 2021;
- 441 20. Cuzick J. A Wilcoxon-type test for trend. *Statistics in medicine*. 1985;4:87-90.
- 442 21. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating Ischemic Stroke Subtype
443 Classification Using Machine Learning and Natural Language Processing. *J Stroke Cerebrovasc*
444 *Dis*. 2019;28:2045-2051.
- 445 22. Park S, Kim B-k, Han M-K, Hong J-H, Yum KS, Lee D-I. Deep Learning for Prediction of
446 Mechanism in Acute Ischemic Stroke Using Brain MRI. *Research Square*. 2021.
- 447 23. Cheng B, Knaack C, Forkert ND, Schnabel R, Gerloff C, Thomalla G. Stroke subtype
448 classification by geometrical descriptors of lesion shape. *PloS one*. 2017;12:e0185063.
- 449 24. Kang DW, Kwon SU, Yoo SH, Kwon KY, Choi CG, Kim SJ, Koh JY, Kim JS. Early recurrent ischemic
450 lesions on diffusion-weighted imaging in symptomatic intracranial atherosclerosis. *Arch*
451 *Neurol*. 2007;64:50-54.
- 452 25. Kleindorfer DO, Towfighi A, Chaturvedi S, Cockroft KM, Gutierrez J, Lombardi-Hill D, Kamel H,
453 Kernan WN, Kittner SJ, Leira EC, et al. 2021 Guideline for the Prevention of Stroke in Patients
454 With Stroke and Transient Ischemic Attack: A Guideline From the American Heart
455 Association/American Stroke Association. *Stroke*. 2021;52:e364-e467.

- 456 26. Dawson J, Bejot Y, Christensen LM, De Marchis GM, Dichgans M, Hagberg G, Heldner MR,
457 Milionis H, Li L, Pezzella FR, et al. European Stroke Organisation (ESO) guideline on
458 pharmacological interventions for long-term secondary prevention after ischaemic stroke or
459 transient ischaemic attack. *Eur Stroke J*. 2022;7:I-II.
- 460 27. Group SPSS, Benavente OR, Coffey CS, Conwit R, Hart RG, McClure LA, Pearce LA, Pergola PE,
461 Szychowski JM. Blood-pressure targets in patients with recent lacunar stroke: the SPS3
462 randomised trial. *Lancet*. 2013;382:507-515.
- 463 28. Jing J, Meng X, Zhao X, Liu L, Wang A, Pan Y, Li H, Wang D, Johnston SC, Wang Y, Wang Y. Dual
464 Antiplatelet Therapy in Transient Ischemic Attack and Minor Stroke With Different Infarction
465 Patterns: Subgroup Analysis of the CHANCE Randomized Clinical Trial. *JAMA Neurol*.
466 2018;75:711-719.
- 467 29. Gao Y, Chen W, Pan Y, Jing J, Wang C, Johnston SC, Amarenco P, Bath PM, Jiang L, Yang Y, et al.
468 Dual Antiplatelet Treatment up to 72 Hours after Ischemic Stroke. *N Engl J Med*.
469 2023;389:2413-2424.
- 470 30. Lee M, Ovbiagele B, Saver JL. Intensive Medical Management to Prevent Large and Small
471 Artery Atherothrombotic Stroke: Time to Expand the Horizon. *JAMA*. 2021;326:217-218.
- 472 31. Kim JS. Role of Blood Lipid Levels and Lipid-Lowering Therapy in Stroke Patients with
473 Different Levels of Cerebral Artery Diseases: Reconsidering Recent Stroke Guidelines. *J Stroke*.
474 2021;23:149-161.
- 475 32. Sharma M, Cornelius VR, Patel JP, Davies JG, Molokhia M. Efficacy and Harms of Direct Oral
476 Anticoagulants in the Elderly for Stroke Prevention in Atrial Fibrillation and Secondary
477 Prevention of Venous Thromboembolism: Systematic Review and Meta-Analysis. *Circulation*.
478 2015;132:194-204.
- 479 33. Hart RG, Catanese L, Perera KS, Ntaios G, Connolly SJ. Embolic Stroke of Undetermined
480 Source: A Systematic Review and Clinical Update. *Stroke*. 2017;48:867-872.
- 481 34. Poli S, Keller T, Martus P, Poli K, Ziemann U, Geisler T, Committee S, investigators A. The
482 ATTICUS Randomized Controlled Trial-Subgroup Analyses. *Stroke*. 2023;54:A31-A31.
- 483 35. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI, Berkowitz SD, Swaminathan B,
484 Lavados P, Wang Y, Wang Y, et al. Rivaroxaban for Stroke Prevention after Embolic Stroke of
485 Undetermined Source. *N Engl J Med*. 2018;378:2191-2201.
- 486 36. Diener HC, Sacco RL, Easton JD, Granger CB, Bernstein RA, Uchiyama S, Kreuzer J, Cronin L,
487 Cotton D, Grauer C, et al. Dabigatran for Prevention of Stroke after Embolic Stroke of
488 Undetermined Source. *N Engl J Med*. 2019;380:1906-1917.
- 489 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.

509 **Figure Legends**

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

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

516

517

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

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

522

523

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

527 Using diffusion-weighted images only, a deep learning algorithm estimated
528 probabilities of CE stroke. Then, the probabilities of every case were categorized into
529 deciles in each dataset. Bars indicate observed frequency of each stroke subtype
530 determined by experts. Note that the proportion of CE stroke diagnosed rises
531 proportionally with the estimated CE probability, suggesting that both human experts
532 and the AI, are examining the same underlying entity. LAA=large artery atherosclerosis;
533 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

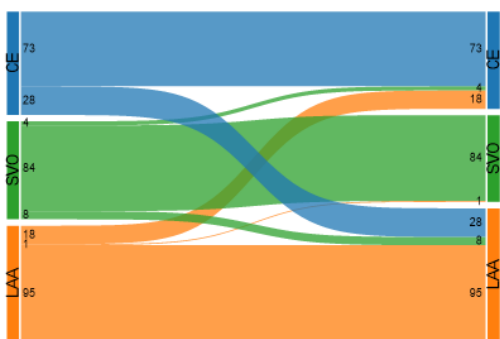
External test dataset 3

		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



Clinical trial dataset



DWIs only

DWIs + AF data

DWIs only

DWIs + AF data

