

Diffusion-Weighted Imaging-Alone Endovascular Thrombectomy Triage in Acute Stroke: Simulating Diffusion-Perfusion Mismatch Using Machine Learning

Yoon-Chul Kim,^a Woo-Keun Seo,^{b,c} In-Young Baek,^b Ji-Eun Lee,^b Ha-Na Song,^b Jong-Won Chung,^b Chi Kyung Kim,^d Kyungmi Oh,^d Sang-il Suh,^e Oh Young Bang,^b Gyeong-Moon Kim,^b Jeffrey L. Saver,^f David S. Liebeskind^f

^aResearch Institute for Future Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

^bDepartment of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

^cDepartment of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Korea

^dDepartment of Neurology, Korea University College of Medicine, Seoul, Korea

^eDepartment of Radiology, Korea University College of Medicine, Seoul, Korea

^fDepartment of Neurology, University of California in Los Angeles, Los Angeles, CA, USA

Dear Sir:

Diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI) in magnetic resonance imaging (MRI) are used to estimate ischemic core and penumbra volumes in patients with acute ischemic stroke.¹ Core size and target mismatch are important indicators in selecting patients for endovascular thrombectomy (EVT).² However, PWI has several drawbacks such as additional scan time, side effects of gadolinium-based contrast agents, and potential inaccuracy in post-processing.^{3,4} Here, we present a method for DWI-alone EVT triage based on machine learning (ML).

The eligibility criterion for inclusion in this study was acute ischemic stroke due to large vessel occlusion within 24 hours from the last known well and available DWI, between January 1, 2011, and December 31, 2019. Demographic information, history of vascular risk factors, initial National Institutes of Health Stroke Scale scores, modified treatment in cerebral infarction (mTICI) scores, 90-day modified Rankin Scale (mRS) scores, and time interval from the last known well to MRI were obtained and compared between the model development and external validation cohorts (Supplementary Table 1). We used prospectively maintained stroke registry databases from the

Samsung Medical Center, Korea University Guro Hospital (KUGH), and University of California Los Angeles Medical Center. Each registry was approved by the Institutional Review Board of each center.

MRI scan parameters are provided in Supplementary Table 2, and detailed information on image processing and ML is provided in the Supplementary Methods. Image features included DWI lesion volume,⁵ lesion sizes in the inner and outer brain regions, and distributions of apparent diffusion coefficient (ADC) ratios (Supplementary Table 3).^{6,7} The features were used as input to ML classifiers. Alternatively, one could consider using deep convolutional neural networks, which learn feature extraction processes from a large training dataset, but we adopted handcrafted features because the size of training data (e.g., the number of patients with unsuccessful recanalization [UR] and the number of patients treated without EVT) was small for deep learning model training.

A flowchart of the EVT triage is shown in Figure 1A. As for the criterion of EVT decision, we defined the gain of thrombectomy (GoT) as a likelihood ratio:

$$\text{GoT} = \frac{\text{Probability of good clinical outcome in case of EVT}}{\text{Probability of good clinical outcome in case of no EVT}} = \frac{P(\text{G}|\text{EVT})}{P(\text{G}|\text{noEVT})}$$

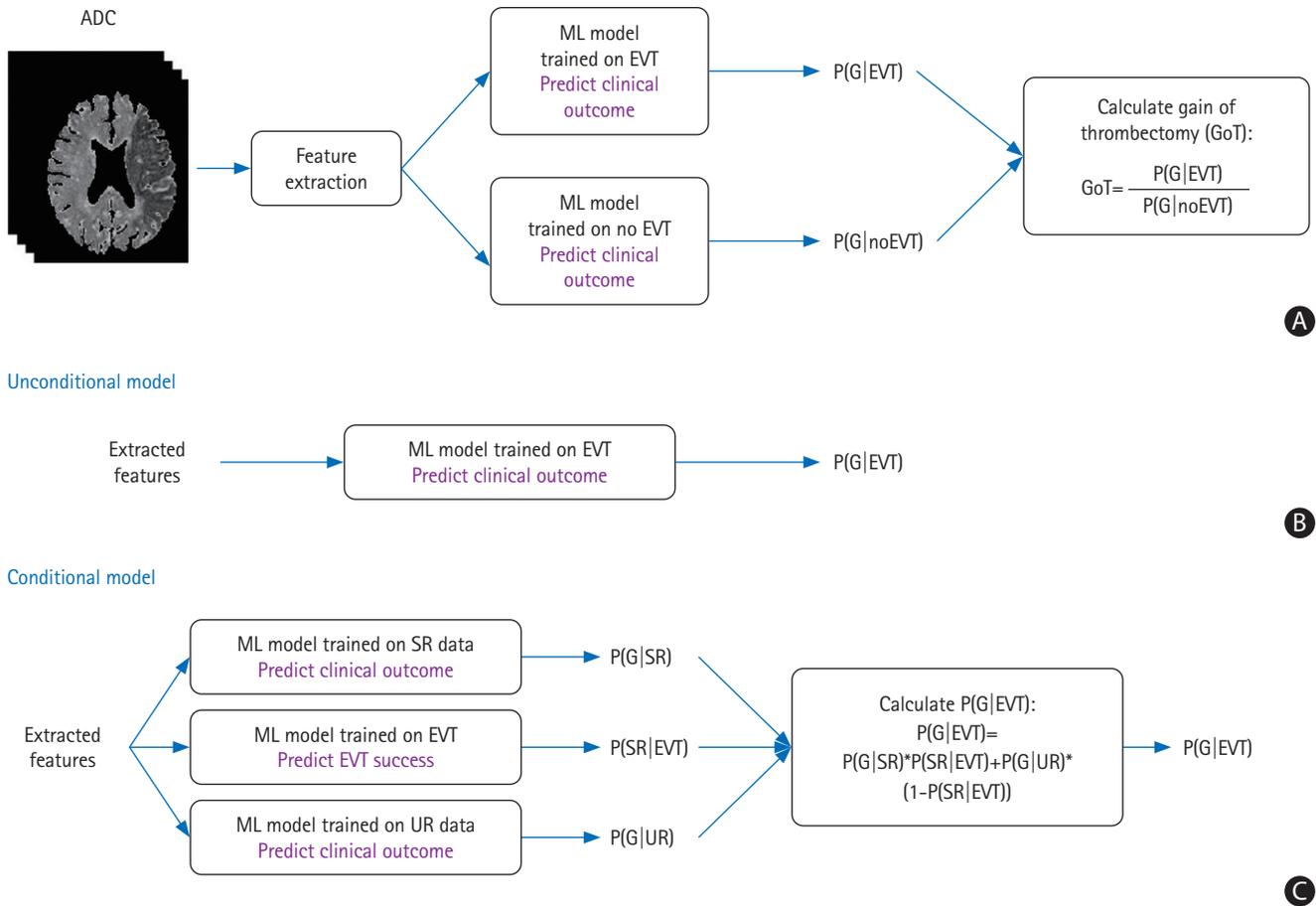


Figure 1. (A) A block diagram of the proposed endovascular thrombectomy (EVT) triage method. Feature extraction is performed on apparent diffusion coefficient (ADC) images. Gain of thrombectomy (GoT) is calculated as a ratio of probability of good outcome in case of EVT to probability of good outcome in case of no EVT. (B, C) Two methods of calculating P(G|EVT). (B) A simple “unconditional” method. All EVT patient data are used to build a model. (C) A “conditional” method where successful recanalization (SR) and unsuccessful recanalization (UR) patient data are separately used to build machine learning (ML) models. In addition, an extra ML model for predicting the probability of EVT success is built. The output probability scores are combined to calculate P(G|EVT). G, good clinical outcome.

Where $p(G|EVT)$ can be further expressed as $p(G|EVT)=p(SR|EVT) \times p(G|SR)+p(UR|EVT) \times p(G|UR)$. UR was defined as the mTICI of 0–2a, and successful recanalization (SR) was defined as the mTICI of 2b–3. As shown in Figure 1B and C, we investigated two approaches: (1) calculate $p(G|EVT)$ regardless of any conditions on the success/failure of recanalization, referred to as “unconditional” model; and (2) calculate $p(G|EVT)$ by reflecting the results of $p(SR|EVT)$, $p(G|SR)$, and $p(G|UR)$, referred to as “conditional” model. Cross-validation results demonstrated similar performance in the logistic regression (LR), random forest (RF), and support vector machine (SVM) classifiers, and $P(G|UR)$ showed low mean F1 values (i.e., 0.23 for LR, 0.24 for RF, and 0.18 for SVM), probably because of the inherent difficulty of predicting infarct growth in patients with UR (Supplementary Table 4).

A receiver operating characteristic analysis was performed with a favorable outcome of the mRS score of ≤ 2 at 90 days as

the dependent variable. The external validation of patients with EVT showed that the areas under the curve (AUCs) of favorable outcome prediction with the ML method were higher than those with the PWI–DWI mismatch ratio (AUC=0.656) (Figure 2). The AUCs were not statistically significantly different between the unconditional and conditional models in the LR and SVM classifiers.

The GoT×EVT interaction in the LR model was statistically significant ($P<0.01$) in the unadjusted model.⁸ We compared the distributions of clinical outcomes between the high and low GoT groups. For the EVT-treated group, the high GoT group showed a larger proportion of low mRS scores than the low GoT group (Figure 3A), indicating that GoT may help predict clinical outcomes prior to treatment. For the no EVT group, the high GoT group showed a smaller proportion of low mRS scores than the low GoT group (Figure 3B).

While this study showed that features extracted from DWI

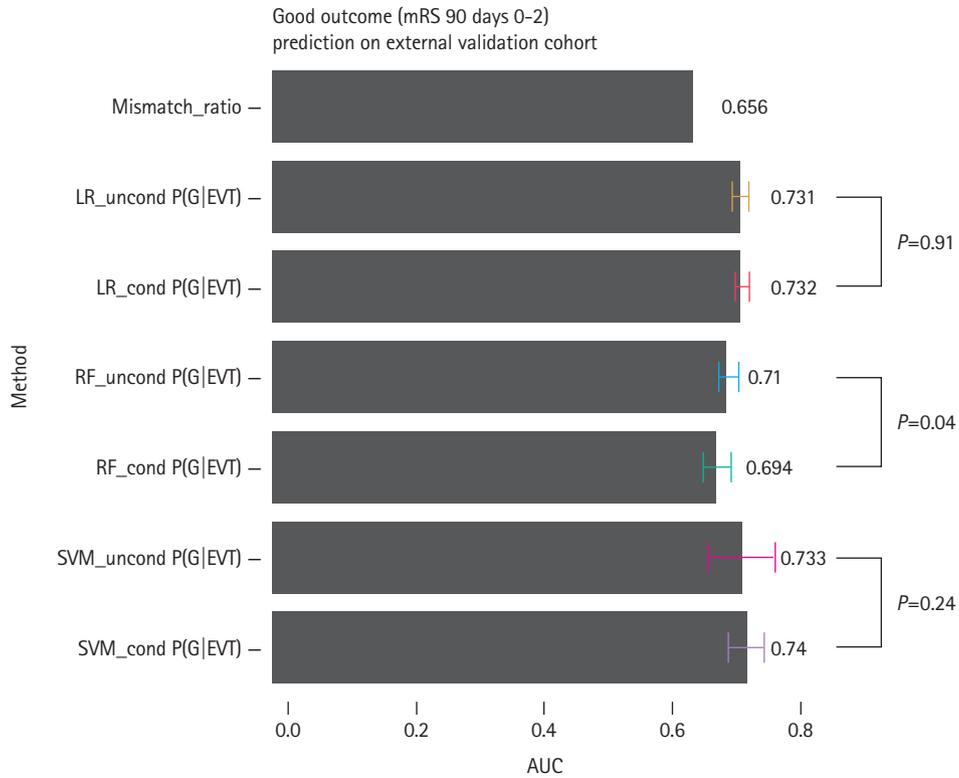


Figure 2. The receiver operating characteristic area under the curve (AUC) values for the mismatch ratio and machine learning methods in the external validation cohort with endovascular thrombectomy (EVT) for the prediction of good clinical outcome (i.e., modified Rankin Scale [mRS] score at 90 days, ≤ 2). LR, logistic regression; G, good clinical outcome; RF, random forest; SVM, support vector machine.

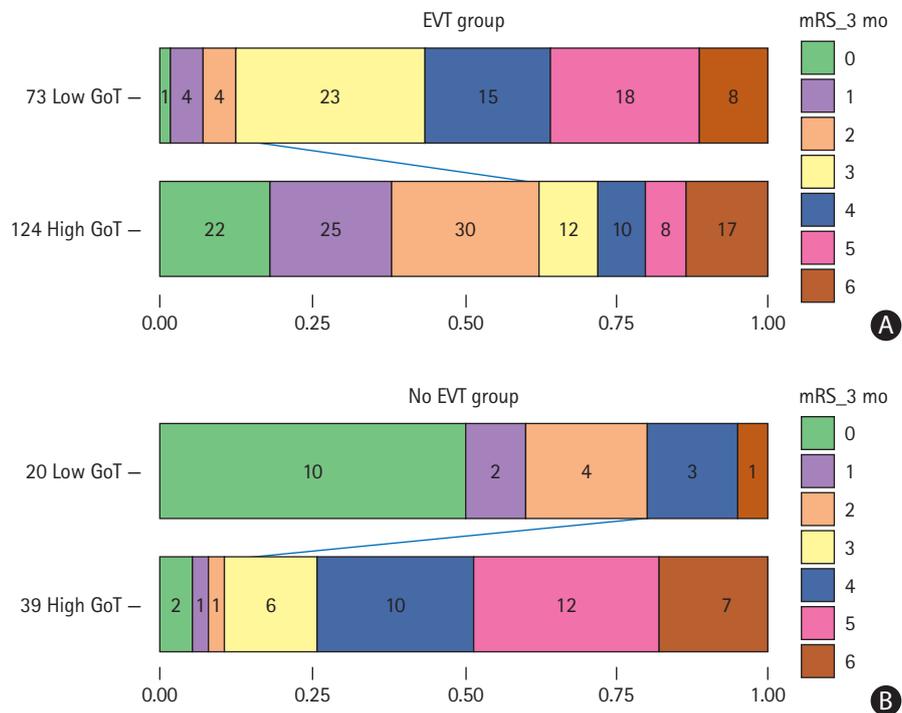


Figure 3. Distributions of the modified Rankin Scale (mRS) scores at 90 days for the (A) endovascular thrombectomy (EVT) and (B) no EVT groups, when the conditional random forest models were used to estimate gain of thrombectomy (GoT). In the EVT group, the high GoT group shows a larger proportion of low mRS scores at 90 days than the low GoT group. In the no EVT group, the low GoT group shows a larger proportion of low mRS scores at 90 days than the high GoT group.

data are robustly predictive of treatment benefit with EVT, additional imaging information could potentially further improve predictive performance. Additional non-perfusion approaches to improve the estimation of $P(\text{SR}|\text{EVT})$ include considering imaging features available from either T2*-weighted gradient echo (GRE) images or time-of-flight (TOF) magnetic resonance angiography. Clot characteristics estimated by GRE can help predict the success or failure of reperfusion intervention. Intracranial vessel tortuosity estimated by TOF image analysis would also be helpful in predicting reperfusion success.⁹

Our study has limitations. First, the sample size was small, and this was a retrospective study. In particular, the number of no EVT subjects and UR subjects was relatively fewer than that of SR subjects. Second, the development/validation data split was not fully randomized. Data from one institution (KUGH) were only used for model development because the PWI data were not available from KUGH for this study. As such, we could not use the data for external validation in the PWI-DWI mismatch ratio calculation. Third, we did not compare our ML methods with an ML-based method using both DWI and PWI data for feature extraction. Fourth, the presented method is based on image features extracted from ADC maps; however, it does not take into account the temporal dynamics of ADC values. The time interval between symptom onset and MRI scan time may be an important factor in outcome prediction.

In conclusion, we presented an ML-based method that takes features derived from DWI images as input and guides patient selection for EVT. The concept of GoT was introduced for EVT triage based on the predictions of clinical outcomes. Although this is a preliminary study with a small sample size, the presented ML method may have the potential to select patients for EVT and predict clinical outcomes with DWI data alone.

Supplementary materials

Supplementary materials related to this article can be found online at <https://doi.org/10.5853/jos.2021.02817>.

References

1. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. *J Magn Reson Imaging* 2010;32:1024-1037.
2. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Orte-

ga-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med* 2018;378:708-718.

3. Calamante F. Arterial input function in perfusion MRI: a comprehensive review. *Prog Nucl Magn Reson Spectrosc* 2013;74:1-32.
4. Deutschmann H, Hinteregger N, Wießpeiner U, Kneihsl M, Fandler-Höfler S, Michenthaler M, et al. Automated MRI perfusion-diffusion mismatch estimation may be significantly different in individual patients when using different software packages. *Eur Radiol* 2021;31:658-665.
5. Kim YC, Lee JE, Yu I, Song HN, Baek IY, Seong JK, et al. Evaluation of diffusion lesion volume measurements in acute ischemic stroke using encoder-decoder convolutional network. *Stroke* 2019;50:1444-1451.
6. Oppenheim C, Grandin C, Samson Y, Smith A, Duprez T, Marsault C, et al. Is there an apparent diffusion coefficient threshold in predicting tissue viability in hyperacute stroke? *Stroke* 2001;32:2486-2491.
7. Rosso C, Hevia-Montiel N, Deltour S, Bardinet E, Dormont D, Crozier S, et al. Prediction of infarct growth based on apparent diffusion coefficients: penumbral assessment without intravenous contrast material. *Radiology* 2009;250:184-192.
8. Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA stroke study. *Neurology* 2000;55:1649-1655.
9. Kim BJ, Kim SM, Kang DW, Kwon SU, Suh DC, Kim JS. Vascular tortuosity may be related to intracranial artery atherosclerosis. *Int J Stroke* 2015;10:1081-1086.

Correspondence: Woo-Keun Seo

Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea

Tel: +82-2-3410-0799

Fax: +82-2-3410-0630

E-mail: mcastenosis@gmail.com

<https://orcid.org/0000-0002-4004-8434>

Received: August 11, 2021

Revised: November 1, 2021

Accepted: November 1, 2021

This work was supported by the National Research Foundation of Korea (NRF), funded by the Korean government (grant numbers 2019R1A2C2008788 and 2018R1D1A1B07042692).

The authors have no financial conflicts of interest.

Supplementary Methods

Image processing

Digital Imaging and Communications in Medicine (DICOM) data for the patients were de-identified using DICOM Anonymizer Pro (NeoLogica S.R.L., Cairo Montenotte, Italy) and converted to the Neuroimaging Informatics Technology Initiative format using dcm2nii (McCausland Center for Brain Imaging, Columbia, SC, USA).

Brain tissue extraction was automatically performed slice-by-slice (i.e., center out slice ordering) by taking the diffusion-weighted images ($b=1,000 \text{ sec/mm}^2$) and using level set segmentation in which the initial level set function is an outer circle enclosing the brain tissue. We excluded the cerebrospinal fluid regions by thresholding apparent diffusion coefficient (ADC) values greater than $1,200 \times 10^{-6} \text{ mm}^2/\text{sec}$ and setting them to 0. The midsagittal plane in the axial brain slice was automatically estimated after determining the optimal values of the translation and rotation parameters. The midline helped correct the head tilt in the axial brain images and identified the contralateral and lesion hemispheres. The ADC ratio has been reported to predict the final infarct volume better than the absolute ADC value. Accordingly, we identified the "core" and "core+penumbra" regions based on the ADC ratio thresholds of 0.7 and 0.85, respectively.

First, the diffusion-weighted imaging (DWI) lesion was automatically detected and segmented using a trained U-Net model. With this process, two features (i.e., DWI lesion volume and right or left brain side of the lesion) were extracted. The center of the brain tissue mask was determined, and a polygon enclosing the brain tissue mask was obtained. Radial lines connecting each vertex in the polygon and the center were drawn, and the vertices dividing the inner/outer regions in an axial brain were obtained to determine the inner and outer regions.

Once the inner and outer regions were identified in the lesion hemisphere, the following features were extracted for each region. The largest lesion size in each slice was identified after automatic labeling of the binary lesion mask and finding the area of the largest four-connected labels. The five largest lesion areas were chosen as features. In addition, a histogram of the ADC ratio values in 15 bins was obtained from 0 to 0.9. In each bin, the lesion volume (i.e., voxel count multiplied by the voxel size) was calculated. Hence, the total number of features was $42=2+5 \times 2+15 \times 2$. The feature names are listed in the Supplementary Table 3.

Machine learning

Machine learning (ML) models were developed using logistic

regression (LR), random forest, and support vector machine classifiers. Five models were developed: (1) a model that predicts the probability of good outcome from data with successful recanalization (SR) (i.e., modified treatment in cerebral infarction [mTICI] 2b–3); (2) a model that predicts the probability of a good outcome from data with unsuccessful recanalization (i.e., mTICI 0–2a); (3) a model that predicts the probability of SR from endovascular thrombectomy (EVT) treatment; (4) a model that predicts the probability of a good outcome from data with EVT treatment; and (5) a model that predicts the probability of good outcome from data without EVT treatment. Fivefold cross-validation was performed using a randomized search for hyperparameter optimization of each classifier in Scikit-learn.

For external validation, we used baseline DWI images from 54 patients with EVT and 18 patients without EVT, who were from either University of California Los Angeles (UCLA) or Samsung Medical Center (SMC). The perfusion-weighted imaging (PWI) and DWI lesion volumes were estimated using RAPID software for the UCLA data and a custom software tool for the SMC data. The mismatch ratio was calculated as the ratio of the PWI lesion volume to the DWI lesion volume. All 42 features were used as inputs for ML predictions. For each ML method, we obtained 10 different train/validation splits using 10 different random seeds and performed 10 five-fold cross-validations to 10 different models, which were tested on an external validation cohort to calculate 10 area under the curve (AUC) values.

Statistical analysis

Statistical analysis was conducted using R version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Descriptive demographics and clinical and radiological data are shown as mean \pm standard deviations or numbers or median and interquartile range. A two-sample unpaired Student's *t*-test was performed to determine if the development and validation data were significantly different in terms of age, National Institutes of Health Stroke Scale score, lesion volume, and onset-to-magnetic resonance imaging time. For the external validation data, we calculated the receiver operating characteristic (ROC) AUC values for the comparison of the prediction of good outcome. The Mann-Whitney *U* test was used to test the statistical significance of AUC differences between unconditional and conditional ML models. Statistical significance was set at $P < 0.05$.

An interaction term of the ML-estimated gain of thrombectomy (GoT) and EVT (binary variable) was used for a LR model with an mRS score ≤ 2 at 90 days as the outcome binary vari-

able. The GoT×treatment interaction was considered significant if the *P*-value was <0.05. We divided the patients treated with EVT into two groups (high and low GoT groups) based on the threshold in the ML-estimated GoT. The cutoff point was ob-

tained by calculating the Youden index from the ROC curve, where the binary functional outcome was the dependent variable. The same dichotomization scheme was used for the patients treated without EVT.

Supplementary Table 1. Patient characteristics for machine learning development and external validation

Characteristic	Model development (n=184)	External validation (n=72)	P
Age (yr)	69±13	72±13	0.11
Male sex	105 (57)	35 (49)	0.23
Hypertension	118 (64)	50 (69)	0.29
Diabetes mellitus	45 (24)	22 (31)	0.30
Dyslipidemia	40 (22)	28 (39)	0.02
Atrial fibrillation	93 (51)	35 (49)	0.88
TOAST classification			
Large artery atherosclerosis	45 (24)	13 (18)	-
Cardioembolism	107 (58)	42 (58)	-
Other determined etiology	13 (7)	10 (14)	-
Undetermined etiology	19 (10)	7 (10)	-
NIHSS score at baseline	10 (6–17)	14 (9–18)	0.06
Last known well to MRI time (min)	111 (72–235)	154 (100–320)	0.23
Pre-treatment diffusion lesion volume (mL)	14 (5–33)	24 (12–55)	0.02
EVT, number	143	54	-
EVT, mTICI=0–2a	48 (34)	7 (13)	-
EVT, mTICI=2b–3	95 (66)	47 (87)	-
No EVT, number	41	18	-

Values are presented as mean±standard deviation, number (%), or median (interquartile range).

TOAST, Trial of Org 10172 in Acute Stroke Treatment; NIHSS, National Institutes of Health Stroke Scale; MRI, magnetic resonance imaging; EVT, endovascular thrombectomy; mTICI, modified treatment in cerebral infarction.

Supplementary Table 2. Diffusion-weighted imaging scan parameters across multiple institutions

Field strength	SMC (n=126)		KUGH (n=76)		UCLA (n=54)	
	3T	1.5T	3T	3T	1.5T	3T
Manufacturer	Philips Achieva	Siemens Avanto	Siemens Trio	Siemens Prisma	Siemens Avanto	Siemens Trio
Repetition time (sec)	3.0	3.7	3.8	3.5	4.1	4.3
Echo time (ms)	81–88	82	91	69	89	78
No. of slices	22	20–22	20	22	26	26
Image matrix size	256×256	192×192	384×384	250×250	130×130	256×256
Pixel spacing (mm ²)	0.94×0.94	1.20×1.20	0.60×0.60	0.96×0.96	1.69×1.69	0.86×0.86
Spacing between slices (mm)	6.5	8	7.5	7	6	6
b-Values (s/mm ²)	0, 1,000	0, 1,000	0, 1,000	0, 1,000	0, 500, 1,000	0, 500, 1,000

SMC, Samsung Medical Center; KUGH, Korea University Guro Hospital; UCLA, University of California Los Angeles Medical Center.

Supplementary Table 3. List of features

Feature characteristics	No. of features	Note
DWI_vol	1	DWI lesion volume (mL) calculated by the U-Net model.
Lesion_R	1	Binary variable. 1=lesion dominance in the right hemisphere. 0=lesion dominance in the left hemisphere.
Outer_cluster_area	5	1st, 2nd, 3rd, 4th, 5th largest areas in the outer brain "core" lesions.
Outer_ADCr_interval	15	Volume in each ADC ratio interval in the outer brain region: (0.00–0.06), (0.06–0.12), (0.12–0.18), (0.18–0.24), (0.24–0.30), (0.30–0.36), (0.36–0.42), (0.42–0.48), (0.48–0.54), (0.54–0.60), (0.60–0.66), (0.66–0.72), (0.72–0.78), (0.78–0.84), (0.84–0.90)
Inner_cluster_area	5	1st, 2nd, 3rd, 4th, 5th largest areas in the inner brain "core" lesions.
Inner_ADCr_interval	15	Volume in each ADC ratio interval in the inner brain region: (0.00–0.06), (0.06–0.12), (0.12–0.18), (0.18–0.24), (0.24–0.30), (0.30–0.36), (0.36–0.42), (0.42–0.48), (0.48–0.54), (0.54–0.60), (0.60–0.66), (0.66–0.72), (0.72–0.78), (0.78–0.84), (0.84–0.90)

DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

Supplementary Table 4. Cross-validation scores for machine learning models in the model development cohort

Model	LR		RF		SVM		
	Accuracy mean (SD)	F1 mean (SD)	Accuracy mean (SD)	F1 mean (SD)	Accuracy mean (SD)	F1 mean (SD)	
Conditional P(G EVT)	P(SR EVT)	0.66 (0.01)	0.82 (0.04)	0.69 (0.05)	0.80 (0.03)	0.66 (0.01)	0.80 (0.01)
	P(G SR)	0.64 (0.05)	0.69 (0.07)	0.66 (0.08)	0.71 (0.07)	0.59 (0.09)	0.72 (0.03)
	P(G UR)	0.73 (0.04)	0.23 (0.29)	0.67 (0.09)	0.24 (0.32)	0.73 (0.04)	0.18 (0.22)
Unconditional P(G EVT)	P(G EVT)	0.59 (0.08)	0.64 (0.06)	0.66 (0.07)	0.64 (0.03)	0.60 (0.06)	0.65 (0.07)
Unconditional P(G noEVT)	P(G noEVT)	0.73 (0.09)	0.18 (0.22)	0.73 (0.15)	0.63 (0.16)	0.68 (0.10)	0.48 (0.25)

LR, logistic regression; RF, random forest; SVM, support vector machine; SD, standard deviation; G, good clinical outcome; EVT, endovascular thrombectomy; SR, successful recanalization; UR, unsuccessful recanalization.