



Automated Prediction of Proximal Middle Cerebral Artery Occlusions in Noncontrast Brain Computed Tomography

Pyeong Eun Kim, MSc*; Hyojung Yang, BS*; Dongmin Kim¹, PhD; Leonard Sunwoo², MD, PhD; Chi Kyung Kim³, MD, PhD; Beom Joon Kim⁴, MD, PhD; Joon-Tae Kim⁵, MD, PhD; Wi-Sun Ryu⁶, MD, PhD; Ho Sung Kim⁷, PhD

BACKGROUND: Early identification of large vessel occlusion (LVO) in patients with ischemic stroke is crucial for timely interventions. We propose a machine learning–based algorithm (JLK-CTL) that uses handcrafted features from noncontrast computed tomography to predict LVO.

METHODS: We included patients with ischemic stroke who underwent concurrent noncontrast computed tomography and computed tomography angiography in seven hospitals. Patients from 5 of these hospitals, admitted between May 2011 and March 2015, were randomly divided into training and internal validation (9:1 ratio). Those from the remaining 2 hospitals, admitted between March 2021 and September 2021, were designated for external validation. From each noncontrast computed tomography scan, we extracted differences in volume, tissue density, and Hounsfield unit distribution between bihemispheric regions (striatocapsular, insula, M1–M3, and M4–M6, modified from the Alberta Stroke Program Early Computed Tomography Score). A deep learning algorithm was used to incorporate clot signs as an additional feature. Machine learning models, including ExtraTrees, random forest, extreme gradient boosting, support vector machine, and multilayer perceptron, as well as a deep learning model, were trained and evaluated. Additionally, we assessed the models' performance after incorporating the National Institutes of Health Stroke Scale scores as an additional feature.

RESULTS: Among 2919 patients, 83 were excluded. Across the training (n=2463), internal validation (n=275), and external validation (n=95) datasets, the mean ages were 68.5 ± 12.4 , 67.6 ± 13.8 , and 67.9 ± 13.6 years, respectively. The proportions of men were 57%, 53%, and 59%, with LVO prevalences of 17.0%, 16.4%, and 26.3%, respectively. In the external validation, the ExtraTrees model achieved a robust area under the curve of 0.888 (95% CI, 0.850–0.925), with a sensitivity of 80.1% (95% CI, 72.0–88.1) and a specificity of 88.6% (95% CI, 84.7–92.5). Adding the National Institutes of Health Stroke Scale score to the ExtraTrees model increased sensitivity (from 80.1% to 92.1%) while maintaining specificity.

CONCLUSIONS: Our algorithm provides reliable predictions of LVO using noncontrast computed tomography. By enabling early LVO identification, our algorithm has the potential to expedite the stroke workflow.

GRAPHIC ABSTRACT: A [graphic abstract](#) is available for this article.

Key Words: artificial intelligence ■ atrial fibrillation ■ computed tomography angiography ■ deep learning ■ ischemic stroke ■ predictive value of tests

Correspondence to: Wi-Sun Ryu, MD, PhD, Artificial Intelligence Research Center, JLK Inc, 5, Teheran-ro 33-gil, Gangnam-gu, Seoul, 06141, Republic of Korea. Email wisunryu@jlkgroup.com

*PE. Kim and H. Yang contributed equally.

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.123.045772>.

For Sources of Funding and Disclosures, see page 1617.

© 2024 JLK Inc. *Stroke* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](#) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

AF	atrial fibrillation
AUC	area under the curve
CTA	computed tomography angiography
DL	deep learning
HU	Hounsfield units
LVO	large vessel occlusion
ML	machine learning
NCCT	noncontrast computed tomography
NIHSS	National Institutes of Health Stroke Scale
NWU	net water uptake
ROC	receiver operating characteristic
STARD	Standards for Reporting of Diagnostic Accuracy Studies

Large vessel occlusion (LVO), defined as the occlusion of large, proximal cerebral arteries, has a substantial clinical impact on ischemic strokes, with LVO accounting for up to 46% of ischemic strokes.¹ A computed tomography or magnetic resonance angiography is commonly used to diagnose LVO. However, these modalities have limitations, including restricted accessibility, procedural delays, and the risk of contrast agent-induced anaphylaxis.² Given the importance of early recognition

of LVO on timely interventions and improved functional outcomes, there remains an unmet medical need for an accurate and rapid LVO diagnosis.

A few studies have reported that several features on noncontrast computed tomography (NCCT) are indicators of LVO, such as hyperdense arterial signs,^{3,4} sulcal effacement,⁵ and loss of gray-white matter differentiation.⁶ While NCCT scans are highly accessible and offer fast acquisition, achieving an accurate diagnosis of LVO from these scans requires expertise in stroke, which is difficult to find in developing countries or hospitals in excessive demand. Previous studies exploring automatic LVO detection using artificial intelligence encountered limitations such as a lack of robust external validation and interpretability in clinical usage.^{3,4,6} To date, there have been a limited number of attempts aimed at extracting handcrafted features that hold clinical relevance to LVO, particularly about ischemic density changes and the presence of clots.

This study used modified 4 Alberta Stroke Program Early Computed Tomography Score regions (striatocapsular, insula, M1–M3, and M4–M6)⁷ as a mean of dividing the brain and extracted features of asymmetry in density, volume, and Hounsfield units (HU) distribution between left and right hemispheres from each region, along with the existence of clot sign (Figure 1). Using the handcrafted features, we developed the machine learning (ML) algorithm (JLK-CTL) to predict LVO in NCCT. This model was externally validated using multicenter data.

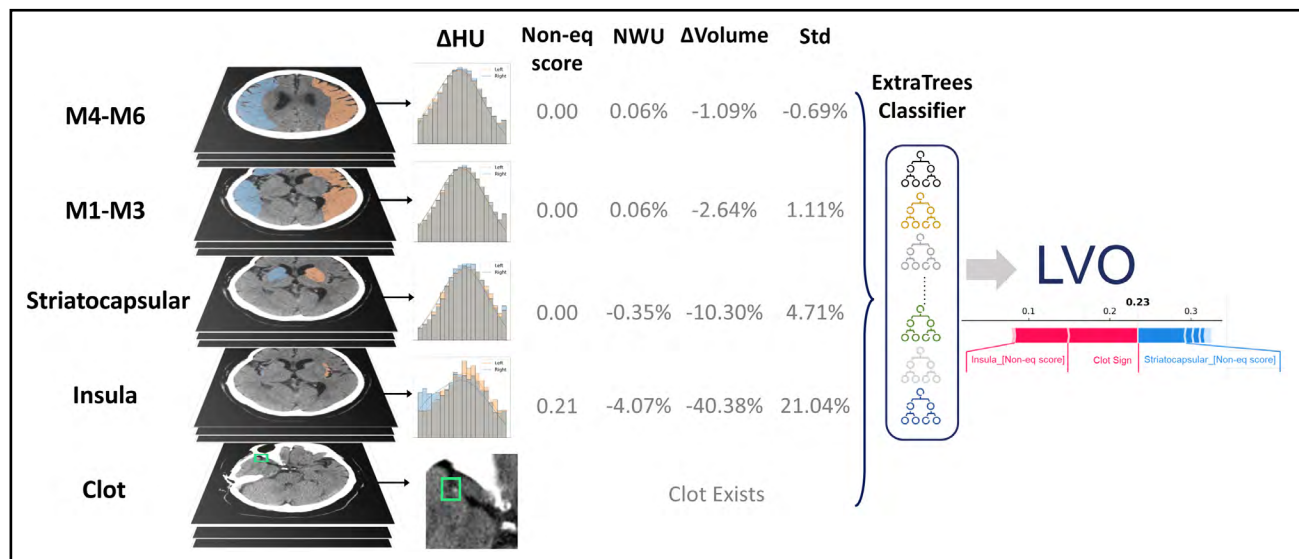


Figure 1. Schematic explanation for JLK-CTL.

By aligning a standard template with regional masks (striatocapsular, insula, M1–M3, and M4–M6, modified from Alberta Stroke Program Early Computed Tomography Score) onto noncontrast computed tomography scan, we calculated nonequivalence scores (non-eq score) indicating mean differences in Hounsfield units (HU) between corresponding regions on both sides. Furthermore, relative differences in tissue density, volume, and HU distribution between the right and left regions were calculated. These are indicated as net water uptake (NWU), change in volume (Δ Volume), and SD of HU distribution (Std), respectively. The existence of clot sign was included as an additional feature, encompassing 17 features. Outputs of ExtraTrees are represented as large vascular occlusion (LVO) scores. We offer a bar graph that illustrates factors that contribute to an increase (shown in red) or decrease (indicated in blue) in the LVO score.

METHODS

The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions. The study was designed to align with STARD (Standards for Reporting of Diagnostic Accuracy Studies).⁸

Study Participants

For the training and internal validation datasets, we retrospectively collected data of patients with ischemic stroke (admitted between May 2011 and March 2015) across 5 comprehensive stroke centers, using the following criteria: (1) age ≥ 18 years, (2) admission within 7 days of symptom onset, and (3) concurrent computed tomography angiography (CTA) and NCCT evaluations for LVO assessment. For the external validation dataset, patients with ischemic stroke (admitted between March 2021 and September 2021) were retrospectively collected from 2 comprehensive stroke centers, adhering to the aforementioned criteria. In all cases, patients underwent both NCCT and CTA to evaluate LVO at the discretion of the attending physician. Patients were excluded based on the following criteria: (1) the presence of brain tumors, intracranial hemorrhages, ventriculoperitoneal shunts, or extraventricular drainage and (2) NCCT scans indicating errors in the skull-stripping process (Figure S1). The institutional review boards of all participating centers approved the study and waived informed consent due to the anonymous and retrospective study design.

Clinical Data Collection

Using a standardized protocol,⁹ we collected demographic data, medication history, and details about vascular risk factors. Stroke subtypes were determined by consensus among experienced neurologists at each participating center, using a validated magnetic resonance imaging-based algorithm.¹⁰ Admission National Institutes of Health Stroke Scale (NIHSS) scores were evaluated by certified (<http://www.stroke-edu.or.kr/>) physicians in each hospital.

Identification of LVO

Vascular status dynamically varies throughout the hyperacute stroke phase¹¹; hence, we used CTA, concurrently performed with NCCT, to determine LVO. We defined LVO as a complete contrast filling defect in the intracranial carotid artery or middle cerebral artery M1/M2 segment. An experienced neurologist diagnosed the presence, side, and location of LVO in all cases. Subsequently, we compared these diagnoses with the stroke registry data, which had been independently verified by attending vascular neurologists at each hospital. If the decision was discordant, a consensus was made. All images were analyzed in a central image laboratory.

Data Preprocessing

Alberta Stroke Program Early Computed Tomography Score Region Segmentation

As preprocessing, skull stripping was performed on individual images using an in-house algorithm. Affine transformation and

symmetrical image normalization deformation methods were conducted using ANTS software.¹² The resulting transformation was applied to the Alberta Stroke Program Early Computed Tomography Score mask,⁷ labeled by an experienced neurologist, using nearest-neighbor interpolation.¹² The modified mask consists of 4 regions: insula, M1 to M3, M4 to M6, and internal capsule-caudate-lentiform (striatocapsular) regions.

To minimize confounding effects of outliers such as artifacts, hemorrhage, chronic infarcts, and cerebrospinal fluid spaces, additional preprocessing was conducted. A windowing with a center value of 40 HU and a width of 40 HU was applied, followed by the removal of voxels in each region that fell outside the 10 HU margin from the median value.

Clot Sign Segmentation

Clot signs were detected on NCCT images using a convolutional neural network with a U-Net architecture.¹³ The U-Net model was trained on 520 patients, validated on 60 patients, and externally validated on 51 patients. Details on training and validation processes are in the [Supplemental Material \(Supplemental Methods; Figure S2\)](#).

Feature Extraction

Nonequivalence Score (Mean HU Differences)

A statistical test on the mean HU difference between left and right regions was conducted using a 2 one-sided t test.¹⁴ It analyzes whether the observed mean HU difference is statistically significant considering the predefined equivalence margin of ± 1 HU. The margin was defined to detect subtle changes of HU in the acute phase, as HU density linearly decreases -0.4 HU/h in patients with ischemic stroke.¹⁵ From the test, a nonequivalence score ranging from 0 to 1 was derived. Lower scores indicate that the observed difference between the 2 hemispheres' regions is more likely within the margin.

Net Water Uptake

The relative difference in densities of the left and right parts of each region, represented as net water uptake (NWU), was measured as shown in equation (1).¹⁶

$$NWU_i = 1 - \frac{\text{mean}(HU_i^L)}{\text{mean}(HU_i^R)} \dots \quad \text{Equation (1)}$$

where i represents modified Alberta Stroke Program Early Computed Tomography Score regions

SD Ratio

The relative difference in the SD of the left region to that of its contralateral regions was calculated by equation (2).

$$STD_i = 1 - \frac{SD(HU_i^L)}{SD(HU_i^R)} \dots \quad \text{Equation (2)}$$

where i represents modified Alberta Stroke Program Early Computed Tomography Score regions

Volumetric Ratio

The relative difference in volume of the left and right parts of each region was calculated through equation (3). This score was included to represent the change in volume of the ischemic brain due to swelling.

$$\Delta V_i = 1 - \frac{\text{Volume}(i^L)}{\text{Volume}(i^R)} \dots \quad \text{Equation (3)}$$

where i represents modified Alberta Stroke Program Early Computed Tomography Score regions

Presence of Clot Sign

The existence of the clot sign was determined by using the automated clot sign segmentation model. Binary scores of 0 or 1 were appended as features, depending on the predicted clot mask.

Thus, by extracting 4 features from each of the 4 regions and including the presence of a clot sign, 17 features were identified in each NCCT scan.

Model Training and Evaluation

Among the study population, patients from 5 hospitals were randomly split in a 9:1 ratio into training and internal validation datasets. Patients from the remaining 2 hospitals were selected as external validation datasets.

We trained 5 ML classification models, including support vector machine,¹⁷ ExtraTrees,¹⁸ random forest,¹⁹ extreme gradient boosting,²⁰ and multilayer perceptron,²¹ and EfficientNetV2S, a deep learning (DL) model.²² Each model was trained and validated on the same dataset. Subsequently, receiver operating characteristic curves were generated for each model, and the area under the curve (AUC) was computed to quantitatively assess model performance. Also, we assessed the model's performance depending on onset to image, NIHSS scores, and the presence of atrial fibrillation (AF).

Optimal threshold values of each model maximizing specificity and sensitivity were determined from the internal validation dataset without bootstrapping. Furthermore, the discriminative power of NIHSS scores on LVO prediction was analyzed by incorporating NIHSS scores into the best model as a continuous variable (JLK-CTL+). Details on model training and evaluation were provided in the [Supplemental Material, Table S1](#), and [Figure S3](#).

Feature Importance

To interpret the classification results of ML models, mean decrease in impurity²³ and permutation importance²⁴ were calculated to explain the contribution of each feature to predictions of the model. For the interpretation of an individual prediction case, we implemented Shapley additive explanations.²⁵ Mean decrease in impurity and permutation importance were used to understand the overall importance of features with respect to all cases in datasets, whereas Shapley additive explanations were used to explicitly explain the contribution of features on a single case prediction. The details of feature importance are presented in the [Supplemental Methods](#).

Statistical Analysis

Data were presented as mean±SD, median (interquartile range), or number (percentage). For the comparison of baseline characteristics between training, internal, and external validation datasets, we used ANOVA or Kruskal-Wallis test for continuous variables and the χ^2 test or Fisher exact test for categorical variables as appropriate. To compare AUC among models, we used the DeLong test.²⁶ All statistical analyses were performed

using STATA, version 16.0 (StataCorp, College Station, TX), and $P<0.05$ was considered statistically significant.

RESULTS

Baseline Characteristics

After excluding 83 patients following exclusion criteria ([Figure S1](#)) from 2919 collected patients, 2833 patients with ischemic stroke from seven hospitals were divided into training (n=2463) and internal (n=275) and external validations (n=95). The mean ages of subjects in the training, internal, and external validation datasets were 68.5±12.4, 67.6±13.8, and 67.9±13.6 years, respectively ([Table 1](#)). Distributions of sex and vascular risk factors were comparable between datasets. LVO was more frequent in the external validation dataset than in other datasets (17.0%, 16.4%, and 26.0% in training, internal, and external validation datasets, respectively; $P=0.045$). The slice thickness of NCCT scans ranged from 3.0 to 5.0 mm.

Performance Comparison of Models in Predicting LVO

ML models achieved an AUC of 0.825–0.865 in the internal validation and 0.864–0.896 in the external validation without significant differences ([Figure 2A](#) and [2B](#)). The DL model exhibited a superior AUC of 0.951 in the internal validation dataset compared with the ExtraTrees model ($P=0.026$). However, in the external validation dataset, its performance was significantly compromised, showing worse AUC compared with ML models (all $P<0.001$). Of note, a marked disparity in the specificity of the DL model was observed between the internal and external validation datasets ([Table 2](#)).

ML models exhibited similar performances. However, in the internal validation dataset, the ExtraTrees model demonstrated the highest AUC compared with other models; in the external validation dataset, only the ExtraTrees model exhibited a sensitivity >80% with a specificity of 89%. Also, consistency of the performance of the ExtraTrees model between the internal and external validation datasets regardless of the selected threshold was observed ([Figure S4](#)). Thus, we selected the ExtraTrees model as our best performing model for further analysis (JLK-CTL). The whole result of JLK-CTL in the external validation dataset is presented in [Figure S5](#).

Better Performance of Imaging Feature Compared With NIHSS Score

The best NIHSS score cutoff value for LVO prediction was 11. Compared with the model solely using NIHSS score (AUC, 0.760 and 0.711 in the internal

Table 1. Baseline Characteristics of Datasets

	Training* (n=2463)	Internal validation (n=275)	External validation (n=95)	P value
Age, y	68.5±12.4	67.6±13.8	67.9±13.6	0.03
Sex, men	1385 (57.2%)	144 (53.3%)	56 (59.0%)	0.44
Large vessel occlusion	419 (17.0%)	45 (16.4%)	25 (26.3%)	0.057
Large vessel occlusion, location				
Intracranial ICA	147 (35.1%)	19 (42.2%)	3 (12.0%)	0.033
MCA-M1	406 (96.9%)	43 (95.6%)	24 (96.0%)	0.87
MCA-M2	333 (79.5%)	40 (88.9%)	21 (84.0%)	0.29
Slice thickness, mm				<0.001
3	27 (1.1%)	4 (1.5%)	0 (0.0%)	
4	920 (37.4%)	109 (39.6%)	0 (0.0%)	
4.8	1020 (41.4%)	109 (39.6%)	0 (0.0%)	
5	496 (20.1%)	53 (19.3%)	95 (100.0%)	
Onset to image, h	16.1 (4.8–48.0)	13.6 (4.9–44.4)	7.7 (2.2–56.9)	0.16†
Admission NIHSS score	4 (2–8)	4 (2–10)	6 (3–14)	<0.001†
Stroke subtype				<0.001
LAA	878 (35.7%)	98 (35.6%)	40 (42.1%)	
SVO	382 (15.5%)	36 (13.1%)	7 (7.4%)	
CE	494 (20.1%)	59 (21.5%)	22 (23.2%)	
Undetermined	503 (20.4%)	62 (22.5%)	12 (12.5%)	
Other determined	45 (1.8%)	10 (3.6%)	10 (10.4%)	
Transient ischemic attack	52 (2.1%)	10 (3.6%)	5 (5.2%)	
Prior stroke	539 (21.9%)	62 (22.6%)	27 (28.4%)	0.45
Coronary artery disease	233 (9.5%)	29 (10.6%)	14 (14.7%)	0.33
Hypertension	1646 (66.8%)	189 (68.7%)	64 (67.4%)	0.71
Diabetes	738 (30.0%)	76 (27.6%)	32 (33.7%)	0.59
Hyperlipidemia	565 (22.9%)	62 (22.6%)	20 (21.1%)	0.74
Atrial fibrillation	468 (19.0%)	50 (18.2%)	28 (29.5%)	0.09

Data are represented as mean±SD, median (interquartile range), or number (percentage). CE indicates cardioembolism; ICA, internal carotid artery; LAA, large artery atherosclerosis; MCA, middle cerebral artery; NIHSS, National Institutes of Health Stroke Scale; and SVO, small vessel occlusion.

*Data were missing in 44, 47, 75, 49, 109, 47, 47, 47, 47, and 47 for age, sex, onset to image, admission NIHSS, stroke subtype, history of prior stroke, coronary artery disease, hypertension, diabetes, hyperlipidemia, and atrial fibrillation, respectively.

†The Kruskal-Wallis test was used.

and external validation datasets, respectively), both models that used imaging (JLK-CTL) or imaging with NIHSS score (JLK-CTL+) showed better performance in predicting LVO (both $P<0.001$; Figure 2C and 2D;

Table 2). Also, on a predefined threshold, incorporating NIHSS scores increased the sensitivity of the model compared with the model using imaging features only. To be specific, to achieve the same sensitivity of JLK-CTL+, the model only using imaging features largely sacrificed specificity (Table 2). This indicates that embedding NIHSS scores increased sensitivity without significantly sacrificing specificity.

Subgroup Analysis: Onset to Image (≤ 6 Versus >6 Hours), NIHSS Score (≤ 11 Versus >11), AF (With AF Versus Without AF), and Exhibition of Arterial Stenosis

When subjects were divided into 2 groups by onset to image of 6 hours, AUCs were comparable between 2 groups in internal and external validation datasets ($P=0.86$ and $P=0.07$, respectively; Figure 3A and 3B). In internal and external datasets, no significant differences in specificity and sensitivity were observed between 2 groups (Table 3). Adding NIHSS scores to the model increased AUC and sensitivity without sacrificing specificity (Figure S6A and S6B). Stratification by onset to image of 4.5, 4, 3, and 2 hours showed similar results (Figure S7; Tables S2 and S3).

After stratified by NIHSS score of 11, no difference in AUC was observed between 2 groups in the internal and external validation datasets ($P=0.21$ and $P=0.08$, respectively; Figures 3C and 3D; Table 3). In addition, AUCs were similar between groups with and without AF groups in the internal and external validation datasets ($P=0.49$ and $P=0.40$, respectively; Figure S8). The sensitivity tended to be higher in AF-positive subgroups (Table S4).

In the external validation dataset, 6 subjects had relevant stenosis without LVO. LVO scores were not different between subjects with and without relevant stenosis; however, results for those with LVO were markedly different compared with subjects with relevant stenosis but without LVO (Figure S9).

Feature Importance Analysis

Analysis of both mean decrease in impurity and permutation importance metrics revealed that the 3 most influential features in predicting LVO were the nonequivalence scores in the striatocapsular region, the nonequivalence scores in the insula region, and the existence of clot sign (Figure S10). Although not as important as these 3 features, nonequivalence scores in M1 to M3 and M4 to M6 regions and NWU in striatocapsular and insula regions were also identified as important factors in the model's prediction.

DISCUSSION

In this study, we developed an ML algorithm to predict LVO in NCCT using handcrafted imaging features

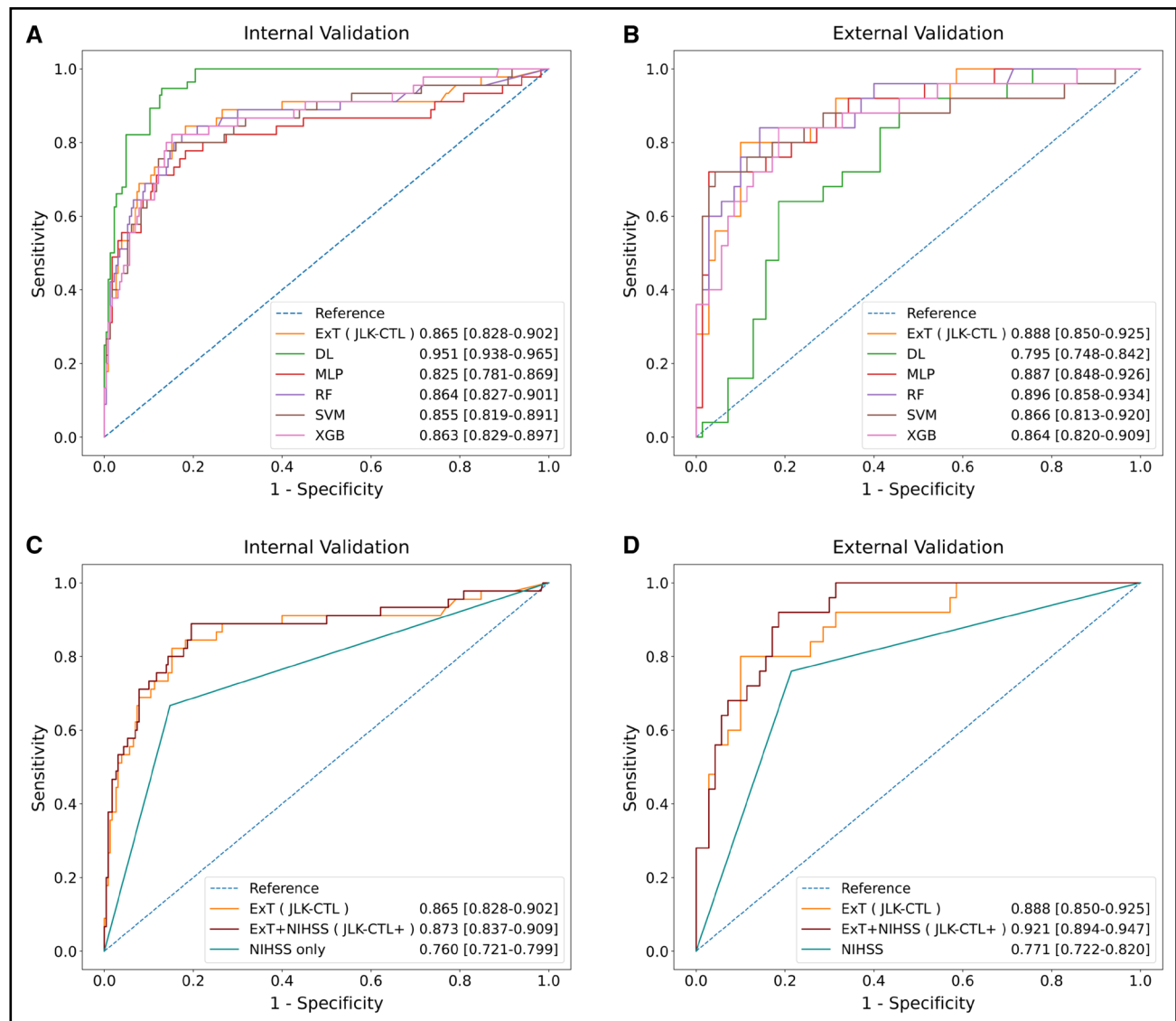


Figure 2. Receiver operating characteristic (ROC) curve of machine learning models and a deep learning (DL) model.

A and **B**, Comparisons of the ROC curves of the DL model with those of machine learning models in the internal and external validation datasets, respectively. **C** and **D**, ROC curves of National Institutes of Health Stroke Scale (NIHSS) model, imaging-only model, and imaging plus NIHSS model in internal and external validation datasets, respectively. ExT indicates ExtraTrees; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine; and XGB, extreme gradient boosting.

including mean HU difference (nonequivalence scores), NWU, SD of HU, volumetric ratio, and presence of clot sign. Our model showed robust performance in detecting LVO in both internal and external validation datasets. In addition, compared with the DL model, our model showed better results in the external validation dataset. Adding NIHSS scores to the model improved predictive performance. Feature importance analysis indicated nonequivalence scores in the striatocapsular and insular regions, along with the presence of a clot sign, as crucial features for predicting LVO in NCCT.

Despite promising results of DL models in a variety of diseases,²⁷ they have an inherent overfitting problem due to their high complexity, which reduces model

performance on unseen data.²⁸ As indicated by our results, ML algorithms, which have less complex architectures, have a more robust performance when applied to unseen data compared with the DL model. Furthermore, our model allowed the visualization of the tree nodes and important features in the individual case (Figure S11), which is difficult in a DL model.²⁹ Such visualization provided insight into how the classifier used the features during LVO prediction and, thus, may increase the credibility of results.

A previous study³ demonstrated high sensitivity (83%) and specificity (71%) of LVO prediction in NCCT from 1453 patients with stroke in whom 57% had an LVO on CTA. However, unlike our approach, the algorithm in the

Table 2. Performance Metrics of Models in Internal and External Validation Datasets

Model	Dataset	Sensitivity	Specificity	PPV	NPV
JLK-CTL (ExtraTrees)	Internal	0.824 (0.765–0.883)	0.848 (0.825–0.872)	0.514 (0.454–0.574)	0.961 (0.947–0.975)
	External	0.801 (0.720–0.881)	0.886 (0.847–0.925)	0.720 (0.636–0.803)	0.925 (0.893–0.956)
Deep learning	Internal	0.872 (0.823–0.920)	0.863 (0.840–0.886)	0.560 (0.500–0.620)	0.971 (0.960–0.983)
	External	0.722 (0.636–0.809)	0.701 (0.646–0.755)	0.459 (0.377–0.542)	0.878 (0.835–0.920)
Multilayer perceptron	Internal	0.780 (0.718–0.842)	0.818 (0.792–0.843)	0.454 (0.397–0.512)	0.950 (0.935–0.966)
	External	0.759 (0.671–0.846)	0.828 (0.783–0.873)	0.616 (0.532–0.701)	0.904 (0.868–0.941)
Random forest	Internal	0.825 (0.767–0.883)	0.827 (0.802–0.852)	0.481 (0.422–0.540)	0.960 (0.947–0.974)
	External	0.760 (0.674–0.847)	0.858 (0.816–0.900)	0.661 (0.575–0.747)	0.908 (0.873–0.943)
Support vector machine	Internal	0.779 (0.716–0.842)	0.865 (0.843–0.888)	0.530 (0.467–0.593)	0.953 (0.938–0.967)
	External	0.718 (0.627–0.809)	0.929 (0.898–0.959)	0.786 (0.702–0.871)	0.901 (0.867–0.935)
Gradient boosting	Internal	0.824 (0.767–0.881)	0.848 (0.825–0.871)	0.513 (0.453–0.572)	0.961 (0.948–0.975)
	External	0.761 (0.674–0.847)	0.831 (0.787–0.876)	0.622 (0.537–0.707)	0.905 (0.869–0.941)
JLK-CTL* (ExtraTrees)	Internal	0.890 (0.841–0.938)	0.726 (0.696–0.755)	0.387 (0.337–0.436)	0.971 (0.959–0.984)
	External	0.801 (0.720–0.881)	0.770 (0.717–0.822)	0.559 (0.478–0.641)	0.914 (0.878–0.950)
JLK-CTL+ (ExtraTrees+NIHSS score)	Internal	0.890 (0.841–0.938)	0.804 (0.778–0.830)	0.469 (0.412–0.526)	0.974 (0.963–0.986)
	External	0.921 (0.866–0.976)	0.815 (0.767–0.863)	0.645 (0.566–0.724)	0.966 (0.943–0.989)
NIHSS score	Internal	0.668 (0.593–0.743)	0.852 (0.829–0.876)	0.468 (0.402–0.534)	0.930 (0.912–0.948)
	External	0.756 (0.672–0.840)	0.786 (0.736–0.836)	0.564 (0.480–0.647)	0.899 (0.861–0.936)

NIHSS indicates National Institutes of Health Stroke Scale; NPV, negative predictive value; and PPV, positive predictive value.
*JLK-CTL threshold shifted to have the same internal validation sensitivity as JLK-CTL+.

previous study used outputs of a DL-based model as features, which is prone to overfitting and hard to interpret^{30,31}; the algorithm was not evaluated on completely new datasets (external validation dataset in our study).

Intriguingly, our model demonstrated similar performance in patients with onset to image ≤ 4.5 and >4.5 hours. In the external validation dataset, patients with onset to image ≤ 4.5 hours had higher NIHSS scores compared with those with onset to image >4.5 hours (median, 9 versus 5; $P=0.01$) although having a comparable frequency of LVO (29% versus 25%; $P=0.83$). This implies that patients with rapidly evolving stroke are more likely to suffer from more severe symptoms, seek medical attention earlier, and exhibit prominent features associated with ischemic stroke on NCCT, even if they arrive within the first 4.5 hours.

In the external validation dataset, LVO scores were comparable between patients without either LVO or stenosis and patients with stenosis. This finding may be attributed to the small number of patients with stenosis in our study and the variability in infarct patterns associated with stenosis. To explore the association between predicted LVO scores and arterial stenosis more comprehensively, further research involving a broader range of stenosis degrees is warranted.

Feature importance analysis revealed that nonequivalence scores across all regions, clot signs, and NWU in the striatocapsular and insular regions are significant attributes. These results align with our understanding of LVO. The importance of density changes (nonequivalence scores and NWU) in insular and striatocapsular

regions reflects the loss of gray-white matter differentiation, occurring at the margins of the insula and lentiform nucleus, respectively.⁷ This raises concerns about performance in cases involving middle cerebral artery bifurcation or M2-middle cerebral artery occlusion. However, in such cases, we observed that HU differences in the insular and M1 to M3 regions emerged as important features for predicting LVO (Figure S12). Additionally, the model's reliance on the presence of a clot corresponds with the previously established significance of clot signs in detecting LVO.³² Moreover, recent studies have associated NWU with functional outcome,³³ hemorrhagic transformation,³⁴ and early neurological deterioration¹⁶ following ischemic stroke. Although NWU was not as significant as HU differences and clot signs in our study, our findings are consistent with these recent studies.

In the emergency room, NCCT scans are routinely performed on patients presenting with neurological symptoms. Approximately 20% of ischemic stroke cases were initially misdiagnosed on arrival at the emergency room, which leads to a delay in the proper assessment and treatment.³⁵ Furthermore, while endovascular treatment is applicable to 10% of patients with acute ischemic stroke and arriving at the hospital within 24 hours,³⁶ the actual number of patients receiving endovascular treatment is much lower. This discrepancy may be partially attributed to the underdiagnosis of LVO in the emergency department. A recent study³⁷ showed that using an automated platform, compared with radiology residents, halves the time required to process imaging in patients suspected of having LVO. From this

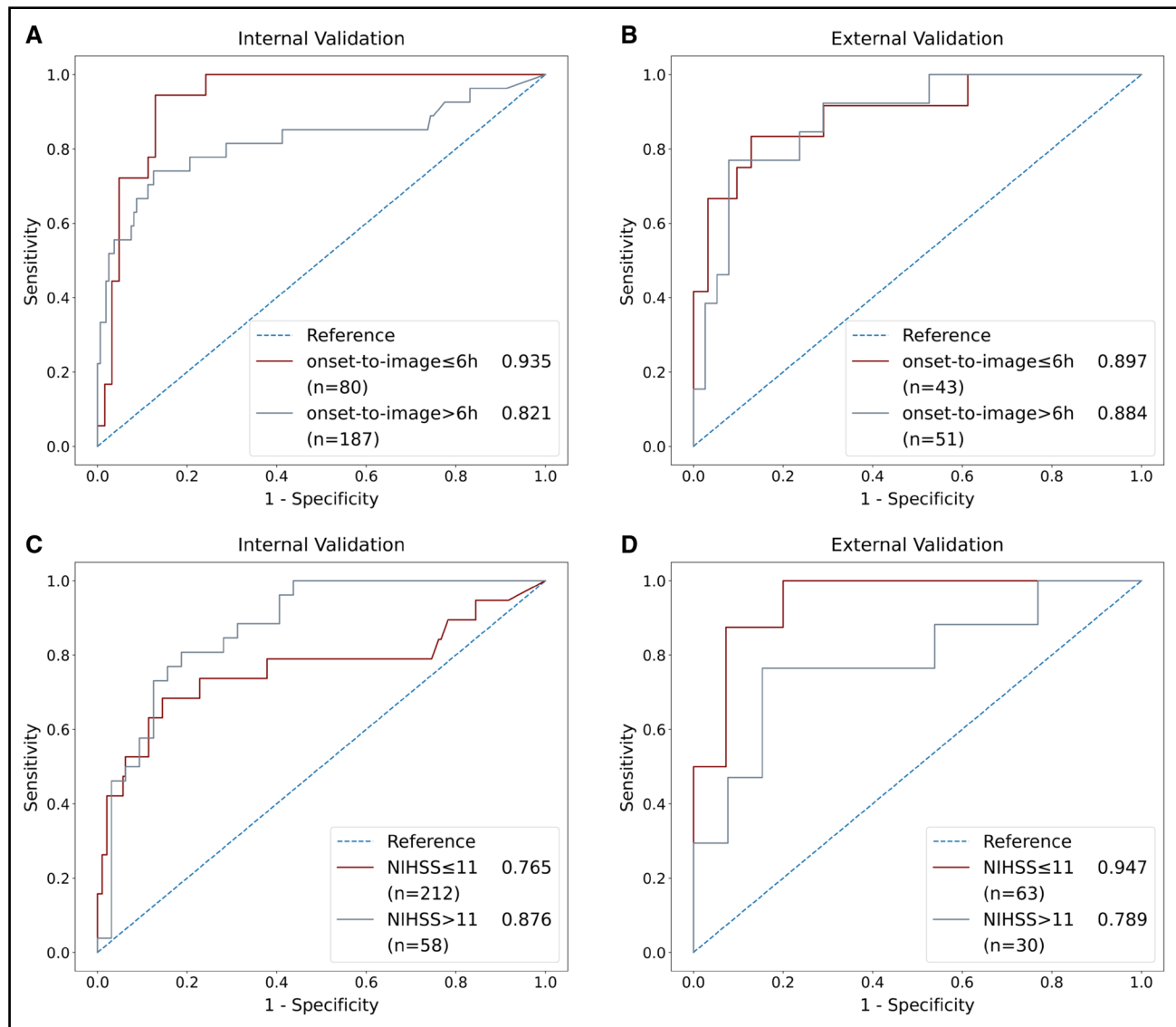


Figure 3. Subgroup analysis stratified by onset to image (≤6 vs >6 hours) and National Institutes of Health Stroke Scale score (NIHSS score ≤11 vs >11) using JLK-CTL.

A and **B**, Internal and external validations after stratified by onset to image. Red and green lines indicate the receiver operating characteristic curve of the onset-to-image ≤6-hour group and >6-hour group, respectively. **C** and **D**, Internal and external validations after stratified by NIHSS score. Red and green lines indicate the receiver operating characteristic curve of the NIHSS score ≤11 group and >11 group, respectively.

perspective, our model has the potential to enhance LVO detection rates and accelerate the workflow for patients eligible for endovascular treatment, thereby improving stroke outcomes.

Our study has strengths, such as large training data labeled through concurrent CTA and available external validation. Nonetheless, a few limitations should be acknowledged. First, our model was trained and validated on the Korean population of patients with stroke. To validate our algorithm in other ethnicities, additional research is required despite our ML model utilized comprehensive imaging features. Second, additional clinical data may improve the performance of the model. However, as information requirements increase, clinical utility

decreases. Third, our model showed low positive predictive value, indicating high false positive rates. The best threshold for streamlining the workflow while minimizing unnecessary further examination should be explored in future studies. Additionally, a clot sign prediction algorithm trained on NCCT with a slice thickness ranging from 3 to 5 mm struggles to identify tiny clots. Nevertheless, the utilization of thinner slice thickness may be less feasible in routine clinical settings. Moreover, the small sample size prevents us from adequately evaluating the algorithm's performance in distinguishing LVO from stenosis. Most importantly, as 96% of patients in datasets had M1 occlusion, our study did not reflect the typical distribution of occluded vessels in patients with acute

Table 3. Subgroup Analysis of JLK-CTL and JLK-CTL+ on Onset to Image (≤6 Versus >6 Hours) and NIHSS Score (≤11 Versus >11)

Model	Dataset	Subgroup	Sensitivity	Specificity	PPV	NPV
JLK-CTL (image only)	Internal	≤6 h (n=80)	0.943 (0.887–0.999)	0.872 (0.831–0.914)	0.683 (0.586–0.780)	0.981 (0.963–0.999)
		>6 h (n=187)	0.744 (0.659–0.829)	0.830 (0.800–0.860)	0.424 (0.352–0.497)	0.951 (0.932–0.969)
	External	≤6 h (n=43)	0.834 (0.725–0.942)	0.871 (0.813–0.930)	0.718 (0.597–0.839)	0.930 (0.883–0.977)
		>6 h (n=51)	0.765 (0.647–0.883)	0.898 (0.848–0.947)	0.721 (0.597–0.844)	0.917 (0.874–0.961)
JLK-CTL+ (image+NIHSS Score)	Internal	≤6 h (n=80)	1.000 (1.000–1.000)	0.824 (0.775–0.873)	0.625 (0.529–0.720)	1.000 (1.000–1.000)
		>6 h (n=187)	0.812 (0.738–0.887)	0.786 (0.754–0.818)	0.390 (0.325–0.455)	0.961 (0.945–0.978)
	External	≤6 h (n=43)	0.916 (0.836–0.996)	0.772 (0.695–0.850)	0.614 (0.497–0.730)	0.959 (0.920–0.998)
		>6 h (n=51)	0.920 (0.843–0.996)	0.845 (0.785–0.904)	0.671 (0.556–0.785)	0.968 (0.939–0.998)
JLK-CTL (image only)	Internal	≤11 (n=212)	0.636 (0.525–0.748)	0.887 (0.864–0.910)	0.358 (0.275–0.442)	0.961 (0.947–0.975)
		>11 (n=58)	0.963 (0.927–1.000)	0.597 (0.508–0.685)	0.661 (0.583–0.740)	0.952 (0.905–1.000)
	External	≤11 (n=63)	0.871 (0.742–0.999)	0.928 (0.894–0.963)	0.639 (0.486–0.792)	0.980 (0.961–0.999)
		>11 (n=30)	0.766 (0.662–0.869)	0.772 (0.654–0.889)	0.809 (0.708–0.910)	0.721 (0.599–0.843)

NIHSS indicates National Institutes of Health Stroke Scale; NPV, negative predictive value; and PPV, positive predictive value.

stroke due to LVO. This would limit the generalizability of our model in real clinical settings. In future studies, analysis of the model based on locations of occlusions is required.

In conclusion, we developed and validated an automated algorithm predicting LVO using NCCT. As the time window for endovascular treatment of LVO strokes is anticipated to increase over time, automated detection of LVO using NCCT may enhance stroke workflow and functional outcomes after LVO strokes.

ARTICLE INFORMATION

Received November 9, 2023; final revision received March 27, 2024; accepted April 11, 2024.

Affiliations

Artificial Intelligence Research Center, JLK Inc, Seoul, Republic of Korea (P.E.K., H.Y., D.K., W.-S.R.). Department of Computer Science and Technology, University of Cambridge, United Kingdom (H.Y.). Department of Radiology, Seoul National University College of Medicine, Republic of Korea (L.S.). Department of Radiology (L.S.) and Department of Neurology (B.J.K.), Seoul National University Bundang Hospital, Seongnam, Republic of Korea. Department of Neurology, Korea University Guro Hospital, Seoul, Republic of Korea (C.K.K.). Department of Neurology, Chonnam National University Hospital, Gwangju, Republic of Korea (J.-T.K.). USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles (H.S.K.).

Acknowledgments

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

Sources of Funding

This study was supported by the Multiminsty Grant for Medical Device Development (grant KMDF_PR_20200901_0098).

Disclosures

P.E. Kim, D. Kim, and Dr Ryu are employees of JLK Inc. The other authors report no conflicts.

Supplemental Material

Supplemental Methods
Tables S1–S4
Figures S1–S12
References 38, 39

REFERENCES

- Morsi RZ, Elfil M, Ghaith HS, Aladawi M, Elmashad A, Kothari S, Desai H, Prabhakaran S, Al-Mufti F, Kass-Hout T. Endovascular thrombectomy for large ischemic strokes: a living systematic review and meta-analysis of randomized trials. *J Stroke*. 2023;25:214–222. doi: 10.5853/jos.2023.00752
- Abdalkader M, Siegler JE, Lee JS, Yaghi S, Qiu Z, Huo X, Miao Z, Campbell BCV, Nguyen TN. Neuroimaging of acute ischemic stroke: multimodal imaging approach for acute endovascular therapy. *J Stroke*. 2023;25:55–71. doi: 10.5853/jos.2022.03286
- Olive-Gadea M, Crespo C, Granes C, Hernandez-Perez M, Perez de la Ossa N, Laredo C, Urria X, Carlos Soler J, Soler A, Puyalto P, et al. Deep learning based software to identify large vessel occlusion on noncontrast computed tomography. *Stroke*. 2020;51:3133–3137. doi: 10.1161/STROKEAHA.120.030326
- Weyland CS, Papanagiotou P, Schmitt N, Joly O, Bellot P, Mokli Y, Ringleb PA, Kastrup A, Mohlenbruch MA, Bendszus M, et al. Hyperdense artery sign in patients with acute ischemic stroke-automated detection with artificial intelligence-driven software. *Front Neurol*. 2022;13:807145. doi: 10.3389/fneur.2022.807145
- Potter CA, Vagal AS, Goyal M, Nunez DB, Leslie-Mazwi TM, Lev MH. CT for treatment selection in acute ischemic stroke: a code stroke primer. *Radiographics*. 2019;39:1717–1738. doi: 10.1148/rq.2019190142
- Wolman DN, van Ommen F, Tong E, Kauw F, Dankbaar JW, Bennink E, de Jong H, Molvin L, Wintermark M, Heit JJ. Non-contrast dual-energy CT virtual ischemia maps accurately estimate ischemic core size in large-vessel occlusive stroke. *Sci Rep*. 2021;11:6745. doi: 10.1038/s41598-021-85143-3
- Pexman JW, Barber PA, Hill MD, Sevrick RJ, Demchuk AM, Hudon ME, Hu WY, Buchan AM. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. *AJNR Am J Neuroradiol*. 2001;22:1534–1542.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, Vet HCW, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277:826–832. doi: 10.1148/radiol.2015151516
- 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. doi: 10.5853/jos.2015.17.1.38
- 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. doi: 10.5853/jos.2014.16.3.161
- Ren Y, Churilov L, Mitchell P, Dowling R, Bush S, Yan B. Clot migration is associated with intravenous thrombolysis in the setting of acute ischemic stroke. *Stroke*. 2018;49:3060–3062. doi: 10.1161/STROKEAHA.118.022751
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54:2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

13. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for bio-medical image segmentation. Paper/poster presented at: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18; 2015.
14. Seaman MA, Serlin RC. Equivalence confidence intervals for two-group comparisons of means. *Psychol Methods*. 1998;3:403–411. doi: 10.1037//1082-989x.3.4.403
15. Kucinski T, Väterlein O, Glauche V, Fiehler J, Klotz E, Eckert B, Koch C, Röther J, Zeumer H. Correlation of apparent diffusion coefficient and computed tomography density in acute ischemic stroke. *Stroke*. 2002;33:1786–1791. doi: 10.1161/01.STR.0000019125.80118.99
16. Ng FC, Yassi N, Sharma G, Brown SB, Goyal M, Majoie CB, Jovin TG, Hill MD, Muir KW, Saver JL, et al; for the HERMES Collaborators. Correlation between computed tomography-based tissue net water uptake and volumetric measures of cerebral edema after reperfusion therapy. *Stroke*. 2022;53:2628–2636. doi: 10.1161/strokeaha.121.037073
17. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst*. 1998;13:18–28. doi: 10.1016/j.ajodo.2023.08.003
18. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63:3–42. doi: 10.1007/s10994-006-6226-1
19. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
20. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Paper/poster presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016.
21. Kruse R, Mostaghim S, Borgelt C, Braune C, Steinbrecher M. Multi-layer perceptrons. In: *Computational Intelligence: A Methodological Introduction*. Springer; 2022:53–124.
22. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. Paper/poster presented at: International Conference on machine learning; 2019.
23. Louppe G, Wehenkel L, Suter A, Geurts P. Understanding variable importances in forests of randomized trees. *Adv Neural Inf Process Syst*. 2013;26.
24. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26:1340–1347. doi: 10.1093/bioinformatics/btq134
25. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using Shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584. doi: 10.1016/j.cmpb.2021.106584
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845. doi: 10.2307/2531595
27. Shen D, Wu G, Suk H. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221–248. doi: 10.1146/annurev-bioeng-071516-044442
28. Rice L, Wong E, Kolter Z. Overfitting in adversarially robust deep learning. Paper/poster presented at: International Conference on Machine Learning; 2020.
29. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Mark*. 2021;31:685–695. doi: 10.1007/s12525-021-00475-2
30. Puig J, Pedraza S, Demchuk A, Dainis IEJ, Termes H, Blasco G, Soria G, Boada I, Remollo S, Banos J, et al. Quantification of thrombus Hounsfield units on noncontrast CT predicts stroke subtype and early recanalization after intravenous recombinant tissue plasminogen activator. *AJNR Am J Neuroradiol*. 2012;33:90–96. doi: 10.3174/ajnr.A2878
31. Kuo KH, Chang FC, Lai YJ, Pan YJ. Hyperdense artery sign, clot characteristics, and response to intravenous thrombolysis in Han Chinese people with acute large arterial infarction. *J Stroke Cerebrovasc Dis*. 2016;25:695–701. doi: 10.1016/j.jstrokecerebrovasdis.2015.11.031
32. Tomsick T, Brott T, Chambers A, Fox A, Gaskill M, Lukin R, Pleatman C, Wiot J, Bourekas E. Hyperdense middle cerebral artery sign on CT: efficacy in detecting middle cerebral artery thrombosis. *AJNR Am J Neuroradiol*. 1990;11:473–477.
33. Brooks G, Meyer L, Elsayed S, McDonough R, Bechstein M, Faizy TD, Sporns P, Schon G, Minnerup J, Kniep HC, et al; I-LAST Investigators. Association between net water uptake and functional outcome in patients with low ASPECTS brain lesions: results from the I-last study. *Neurology*. 2023;100:e954–e963. doi: 10.1212/WNL.00000000000021601
34. Xu T, Yang J, Han Q, Wu Y, Gao X, Xu Y, Huang Y, Wang A, Parsons MW, Lin L. Net water uptake, a neuroimaging marker of early brain edema, as a predictor of symptomatic intracranial hemorrhage after acute ischemic stroke. *Front Neurol*. 2022;13:903263. doi: 10.3389/fneur.2022.903263
35. Tan S, Tang C, Ng JS, Ng C, Kooor J, Gupta A, Goh R, Bacchi S, Jannes J, Kleinig T. Delays in the diagnosis of ischaemic stroke presenting with persistent reduced level of consciousness: a systematic review. *J Clin Neurosci*. 2023;115:14–19. doi: 10.1016/j.jocn.2023.07.009
36. Vanacker P, Lambrou D, Eskandari A, Mosimann PJ, Maghraoui A, Michel P. Eligibility and predictors for acute revascularization procedures in a stroke center. *Stroke*. 2016;47:1844–1849. doi: 10.1161/STROKEAHA.115.012577
37. Temmen SE, Becks MJ, Schalekamp S, van Leeuwen KG, Meijer FJA. Duration and accuracy of automated stroke CT workflow with AI-supported intracranial large vessel occlusion detection. *Sci Rep*. 2023;13:12551. doi: 10.1038/s41598-023-39831-x
38. Mattle HP, Arnold M, Georgiadis D, Baumann C, Nedeltchev K, Benninger D, Remonda L, Büdingen Cv, Diana A, Pangalu A, et al. Comparison of intraarterial and intravenous thrombolysis for ischemic stroke with hyperdense middle cerebral artery sign. *Stroke*. 2008;39:379–383. doi: 10.1161/STROKEAHA.107.492348
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg VS. Machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>