

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

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

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

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

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

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

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

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

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

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

\*Data were missing in 35 patients; <sup>†</sup>Data were missing in 271 patients; <sup>‡</sup>Data were missing in 1 patient; <sup>§</sup>Data were missing in 129 patients; <sup>‡</sup>Fisher's exact test was used.