Appendix 1 Calculation methods of performance metrics

True positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs)

To automatically count TPs, TNs, FPs, and FNs, computer sentences were used to determine whether the annotated lesion segmented by the reader was the same as the lesion of reference standard. The centroid of each annotated lesion was marked, and the corrected coordinates were calculated based on the image registration. If the coordinates of the annotated lesion were within the region of interests that were segmented on the ground-truth FLAIR images, this lesion was regarded as correct, otherwise, it was regarded as incorrect. All the annotated lesions were judged and recorded automatically. Next, a neuroradiologist with 9 years of experience evaluated the records and manually corrected the inappropriate, as necessary.

Formulas of the readers' performance metrics

Abbreviations: PPV, positive predictive value; NPV, negative predictive value.

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Sensitivity = \frac{TN}{TN+FP}$$

$$PPV = \frac{TP}{TP+FP}$$

$$NPV = \frac{TN}{TN+FN}$$

$$F\ measure = \frac{2 \times PPV \times Sensitivity}{PPV + Sensitivity}$$

Table S1 Individual reader performance with CT and synthetic MRI

Performance Metrics	Reader 1		Reader 2		Reader 3	
	CT (n=53)	Synthetic MRI (n=53)	CT (n=53)	Synthetic MRI (n=53)	CT (n=53)	Synthetic MRI (n=53)
Patient detection						
True positive	3	33	37	46	17	44
True negative	3	2	0	0	3	1
False positive	0	1	3	3	0	2
False negative	47	17	13	4	33	6
Sensitivity (%)[†]	6 (3/50) [2, 18]	66 (33/50) [51, 78]	74 (37/50) [59, 85]	92 (46/50) [80, 97]	34 (17/50) [22, 49]	88 (44/50) [75, 95]
Specificity (%)[†]	100 (3/3) [31, 100]	67 (2/3) [13, 98]	0 (0/3) [0, 69]	0 (0/3) [0, 69]	100 (3/3) [31, 100]	33 (1/3) [2, 88]
PPV (%)[†]	100 (3/3) [31, 100]	97 (33/34) [83, 100]	93 (37/40) [79, 98]	94 (46/49) [82, 98]	100 (17/17) [77, 100]	96 (44/46) [84, 99]
NPV (%)[†]	6 (3/50) [2, 18]	11 (2/19) [2, 35]	0 (0/13) [0, 28]	0 (0/4) [0, 60]	8 (3/36) [2, 24]	14 (1/7) [1, 58]
F measure	0.50	0.66	0.56	0.58	0.65	0.67
Detection time (s)*	8 (6, 12)	14 (6, 35)	31 (14, 47)	14 (9, 24)	6 (5, 8)	4 (3, 7)
Self-confidence*	100 (100, 100)	100 (89, 100)	73 (64, 89)	73 (61, 92)	100 (95, 100)	100 (96, 100)
Lesion detection						
True positive	12	111	29	49	27	102
False positive	55	154	39	37	54	95
False negative	528	429	511	491	513	438
Sensitivity (%)[†]	2 (12/540) [1, 4]	21 (111/540) [17, 24]	5 (29/540) [4, 8]	9 (49/540) [7, 12]	5 (27/540) [3, 7]	19 (102/540) [16, 23]
PPV (%)[†]	18 (12/67) [10, 30]	42 (111/265) [36, 48]	43 (29/68) [31, 55]	57 (49/86) [46, 67]	33 (27/81) [23, 45]	52 (102/197) [45, 59]
F measure	0.04	0.33	0.10	0.16	0.09	0.30

*, data are the median, with interquartile ranges in parentheses. [†], data in parentheses are the numerator and denominator, with 95% CI in brackets. True negatives were not counted at the lesion level. Reader 1, a CT technician with 1.5 years of experience; Reader 2, a second-year radiology resident; Reader 3, a junior radiologist with 4 years of experience. PPV, positive predictive value; NPV, negative predictive value; MRI, magnetic resonance imaging.