Supplementary Material 1. Methods and Results: AI quantification system

Supplementary Methods

Overview. The AI quantification system applied to each SWI scan in this study comprises two complementary 3D nnU-Net components, each trained, monitored, and externally validated on cohorts whose contributing institutions are mutually disjoint. The first component is a multi-label fully convolutional encoder–decoder that produces voxelwise maps of intraparenchymal hemorrhage and cortical superficial siderosis (cSS) from a skull-stripped MRI volume; cSS is separated from intraparenchymal hemorrhage by a cerebrospinal-fluid (CSF) overlap rule using an auxiliary 2D CSF segmentation network. Modality-matched weights for this multi-label network were trained for SWI and for GRE separately. The second component is a dedicated 3D nnU-Net for cerebral microbleed (CMB) detection on SWI, followed by a three-stage anatomical-, geometric-, and vessel-aware post-processing pipeline that suppresses CMB mimics (calcification, flow voids, vessel cross-sections). In the present manuscript, hemorrhage and cSS volumes are taken from the multi-label network, and CMB counts are taken from the dedicated CMB pipeline. Supplementary Method Figure 1 shows the architecture and CSF-guided logic of the multi-label network. The present manuscript uses only the SWI inputs and the SWI external validation; the GRE weights, GRE training data, and GRE external-validation are described here only for completeness of the software product and did not enter any analytical pipeline of this manuscript.

1. Multi-label hemorrhage and cortical superficial siderosis model

1.1 Network architecture. The multi-label segmentation network is built on the nnU-Net 3D framework. The encoder applies two (1, 3, 3) and five (3, 3, 3) convolutional layers, each followed by instance normalization, leaky-ReLU activation, and a strided downsampling step. The decoder mirrors the encoder with transposed-convolution blocks that take the skip connection from the corresponding encoder level as additional input; each decoder block is composed of five (3, 3, 3) convolutional layers followed by a final (1, 3, 3) convolutional layer with instance normalization and leaky-ReLU activation. A final softmax layer produces voxelwise multi-label segmentation. Modality-specific weights were trained for SWI and for GRE separately.

1.2 Preprocessing. Skull stripping was performed with SynthStrip. The skull-stripped volume was cropped to the non-zero region, resampled to the per-modality median voxel spacing, intensity-normalized by z-score, and split into 3D patches with 50% overlap. Per-modality target voxel spacings (mm) and patch sizes (voxels) were (1, 0.544, 0.544) and (80, 160, 192) for SWI; (4, 0.430, 0.430) and (28, 224, 256) for GRE.

1.3 Training data and ground-truth annotation. Training data comprised 1,152 SWI scans from 12 Korean academic medical centers and 1,441 GRE scans from a separately constituted set of 12 Korean academic medical centers. Acquisition characteristics of the training cohorts spanned three scanner manufacturers (Philips, Siemens, GE) at 1.5T and 3.0T (SWI predominantly 3.0T, GRE evenly distributed between 1.5T and 3.0T), and a range of in-plane resolutions and slice thicknesses (Supplementary Method Table 3). Voxelwise reference masks were annotated as a multi-label (hemorrhage, cSS, cerebral microbleed) by a vascular neurologist with more than 15 years of experience. Among the training set, 335/1,152 SWI scans (29.1%) contained at least one hemorrhage lesion and 77/1,152 (6.7%) contained at least one cSS lesion; corresponding GRE training positivity was 494/1,441 (34.3%) for hemorrhage and 82/1,441 (5.7%) for cSS.

1.4 Training procedure. The objective combined Dice loss and binary cross-entropy. Optimization used Adam with initial learning rate 3×10^{-4} and weight decay 3×10^{-5}, paired with the ReduceLROnPlateau scheduler. Training was capped at 1,000 epochs with early stopping after 50 epochs without improvement in validation loss; each epoch comprised 250 minibatches at batch size 2 (500 patches per epoch). A positive-to-negative patch ratio of 1:2 was enforced by oversampling. Standard nnU-Net augmentation was applied (elastic deformation, scaling, rotation, mirroring, and gamma correction). Final convergence occurred at epoch 442 (SWI multi-label network) and epoch 433 (GRE multi-label network). Implementation used Python 3.10.4, PyTorch 2.0.1, CUDA 12.0, on a single NVIDIA A6000 GPU; each epoch required approximately 90 s. Following the default nnU-Net protocol, training fold and validation fold were not partitioned in the traditional sense; the entire training set was also used to compute validation loss for the early-stopping criterion, with overfitting controlled by random per-epoch patch sampling, oversampling, and on-the-fly augmentation.

1.5 CSF-guided post-processing. A separately trained 2D nnU-Net produces a voxelwise CSF mask for each scan (SWI target voxel spacing 1 × 0.313×0.313 mm, patch 448×512; GRE 4×0.450×0.450 mm, patch 320×384; convergence at epoch 830 for SWI and epoch 672 for GRE). Each connected component output by the multi-label network is classified by its fractional overlap with the CSF mask: lesions with more than 50% voxelwise overlap with CSF are labelled as cSS, consistent with the cortical-surface anatomical location of cSS, while lesions with 50% or less overlap are labelled as intraparenchymal hemorrhage. The CMB label from the multi-label network is passed through directly and is not subject to CSF-overlap reclassification; in the present study, however, CMB counts are taken from the dedicated CMB model described in §2, not from the multi-label network's CMB channel.

1.6 Inference. At inference the test volume is decomposed into overlapping 3D patches under the same target-spacing scheme as training. Each patch is passed through the network with test-time augmentation by mirroring; predictions are aggregated by averaging in overlapping regions to produce the volume-level multi-label segmentation, after which the CSF-guided rule in §1.5 is applied.

2. Dedicated cerebral microbleed detection model

2.1 Architecture and preprocessing. The CMB detection model is a 3D nnU-Net trained specifically for CMB on skull-stripped SWI volumes. Target voxel spacing was 0.327 × 0.327 × 2.0 mm and patch size 32 × 224 × 256 voxels. Preprocessing followed the same nnU-Net pipeline as the multi-label network (crop to non-zero region, resampling to target spacing, z-score intensity normalization, patch extraction with 50% overlap).

2.2 Training data and ground-truth annotation. The CMB model was trained on 932 SWI scans from 9 Korean academic medical centers. Field strength was 3.0T for 99.6% of training scans. Ground-truth CMB masks were annotated by the same vascular neurologist (more than 15 years of experience) following STandards for ReportIng Vascular changes on nEuroimaging (STRIVE) descriptors. The 9 institutions contributing to CMB training are mutually disjoint from the 3 institutions contributing to external validation of this model.

2.3 Training procedure. Training used the same loss (Dice + binary cross-entropy), optimiser (Adam, initial learning rate 3×10^{-4}, weight decay 3×10^{-5}), early-stopping criterion (patience 50 epochs), positive:negative patch oversampling (1:2), and augmentation scheme as the multi-label network. Final convergence was reached at epoch 247 with batch size 2.

2.4 CMB-specific mimics-filtering post-processing. The raw CMB candidate set produced by the network is filtered through three sequential stages designed to suppress the principal CMB mimics on SWI:

(i) Anatomy-based filtering. A separately trained anatomical segmentation model labels the ventricles, pineal gland, choroid plexus, globus pallidus, midline plane, and brain periphery; CMB candidates overlapping these calcification- or flow-prone structures are removed.

(ii) Geometric filtering. Candidates with high eccentricity are excluded; only approximately spherical lesions are retained, consistent with the STRIVE morphological criteria for true microbleeds.

(iii) Vessel-tracking filter. Frangi vesselness filters followed by connected-component analysis identify linear hypointense structures spanning multiple slices on SWI; candidates classified as continuous vascular structures are removed, retaining only isolated spherical lesions.

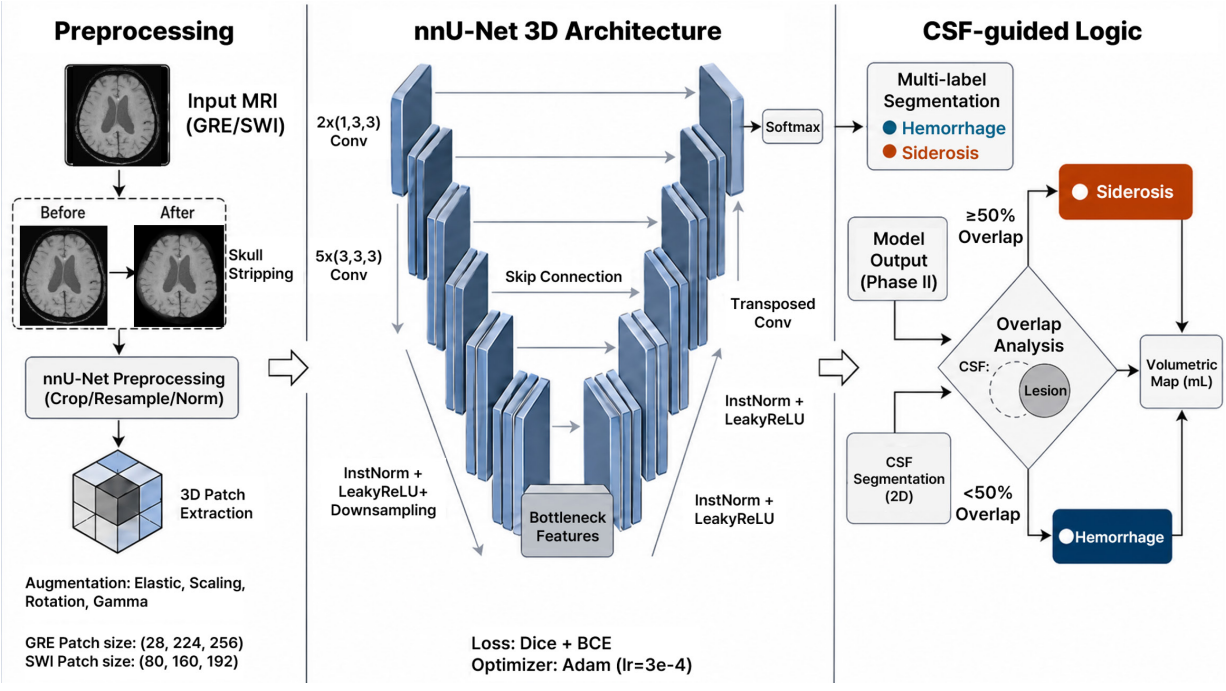
2.5 Inference. Inference uses the same overlapping-patch test-time-augmentation scheme as the multi-label network, after which the three-stage mimics filter is applied to produce the final per-scan CMB count.

3. Confirmation of training–validation independence

For each model and each modality, the set of institutions contributing scans to the external validation cohort is mutually disjoint from the set of institutions contributing scans to the corresponding training cohort. For the multi-label hemorrhage/cSS model, the 2 SWI external-validation institutions are not represented in the 12-institution SWI training pool, and the 2 GRE external-validation institutions are not represented in the 12-institution GRE training pool. For the dedicated CMB model, the 3 external-validation institutions are not represented in the 9 CMB training institutions. Where a given institution contributed scans to one modality's training pool and a different modality's external-validation pool, no scan is shared because SWI and GRE are physically distinct acquisitions: this constitutes an additional cross-modality independence check rather than a contamination concern. No patient or scan appears in both a training set and the external validation set used to evaluate the same network.

Supplementary Results

External validation cohorts comprised 216 SWI scans (33 hemorrhage-positive, 5 cSS-positive) from 2 institutions for the multi-label SWI model; 310 GRE scans (123 hemorrhage-positive, 9 cSS-positive) from 2 institutions for the multi-label GRE model; and 180 SWI scans from 3 institutions for the dedicated CMB model. Performance is summarized in Supplementary Method Table 1; cohort composition is summarized in Supplementary Method Table 2; scanner and acquisition heterogeneity for the multi-label cohorts is summarized in Supplementary Method Table 3.



Supplementary Method Figure 1. Architecture and pipeline of the multi-label AI quantification model for intraparenchymal hemorrhage and cortical superficial siderosis (cSS).
Left (Preprocessing): the input MRI (SWI or GRE) is skull-stripped (before/after thumbnails shown) and passed through the standard nnU-Net preprocessing chain — cropping to the non-zero region, resampling to the per-modality median voxel spacing, z-score intensity normalization, and 3D patch extraction. Standard nnU-Net augmentation (elastic deformation, scaling, rotation, gamma correction) is applied during training; per-modality patch sizes are (28, 224, 256) for GRE and (80, 160, 192) for SWI. **Middle (nnU-Net 3D architecture):** a fully convolutional encoder–decoder. The encoder applies two (1, 3, 3) and five (3, 3, 3) convolutional layers, each followed by instance normalization, leaky-ReLU activation, and a strided downsampling step, reducing the input to bottleneck features. The mirrored decoder uses transposed convolutions with skip connections from the corresponding encoder level; each decoder block is followed by instance normalization and leaky-ReLU activation. A final softmax layer produces

https://doi.org/10.18700/jnc.260038

voxelwise multi-label segmentation, indicated in the upper-right legend as hemorrhage (blue) and cortical superficial siderosis (orange). Training uses Dice + binary cross-entropy loss and the Adam optimizer at an initial learning rate of 3×10^{-4}. **Right (CSF-guided logic):** the multi-label network output is jointly evaluated with an auxiliary 2D cerebrospinal-fluid (CSF) segmentation mask in a per-component overlap analysis (diamond node, with CSF and a candidate lesion shown schematically inside). Connected components with ≥ 50 % CSF overlap are classified as cortical superficial siderosis (orange); components with < 50 % CSF overlap are classified as intraparenchymal hemorrhage (dark blue). Both classes are exported as a per-lesion volumetric map in milliliters. Cerebral microbleed counts used in this study are produced by a separate dedicated 3D nnU-Net described in §2 of Supplementary Methods and are not generated through the pathway depicted in this figure. BCE, binary cross-entropy; CSF, cerebrospinal fluid; cSS, cortical superficial siderosis; GRE, gradient-recalled echo T2*; MRI, magnetic resonance imaging; SWI, susceptibility-weighted imaging.

Supplementary Method Table 1. External validation performance of the AI quantification system

Component (model)	Modality	N scans	Dice similarity coefficient	Sensitivity	Specificity
Hemorrhage and cortical superficial siderosis (multi-label)	SWI	216	0.859	0.970	0.885
Hemorrhage and cortical superficial siderosis (multi-label)	GRE	310	0.930	0.951	0.957
Cerebral microbleed (dedicated, with mimics filter)	SWI	180	—	0.913	0.842

Sensitivity and specificity for the multi-label hemorrhage/cSS model are at the lesion level against the vascular-neurologist multi-label reference; the Dice similarity coefficient is the voxelwise mean over the hemorrhage and cortical superficial siderosis foreground classes combined. Performance for the dedicated CMB model is reported at the subject level against the STRIVE-graded CMB reference. The CMB model outputs CMB counts and was not optimized for voxelwise Dice; hence the Dice column is not applicable. cSS, cortical superficial siderosis; GRE, gradient-recalled echo T2*; SWI, susceptibility-weighted imaging.

Supplementary Method Table 2. Composition of the training and external validation cohorts

Component (model)	Modality	Set	Scans (N)	Institutions (N)	Hemorrhage-positive, n (%)	cSS-positive, n (%)
Hemorrhage/cSS (multi-label)	SWI	Training	1,152	12	335 (29.1%)	77 (6.7%)
Hemorrhage/cSS (multi-label)	SWI	External validation	216	2	33 (15.3%)	5 (2.3%)
Hemorrhage/cSS (multi-label)	GRE	Training	1,441	12	494 (34.3%)	82 (5.7%)
Hemorrhage/cSS (multi-label)	GRE	External validation	310	2	123 (39.7%)	9 (2.9%)
CMB (dedicated)	SWI	Training	932	9	—	—
CMB (dedicated)	SWI	External validation	180	3	—	—

Institutions contributing scans to each training cohort and the corresponding external validation cohort are mutually disjoint within each row's modality (see §3 of Supplementary Methods). Hemorrhage-positive and cSS-positive denote scans with at least one annotated lesion of the respective class in the reference mask; per-lesion positivity columns are not applicable to the dedicated CMB model, which is reported only at subject level for CMB count. CMB, cerebral microbleed; cSS, cortical superficial siderosis; GRE, gradient-recalled echo T2*; SWI, susceptibility-weighted imaging.

Supplementary Method Table 3. Acquisition characteristics of the multi-label hemorrhage/cSS training and external validation cohorts

Characteristic	SWI training (N=1,152)	SWI external (N=216)	GRE training (N=1,441)	GRE external (N=310)
Manufacturer				
Philips	609	66	763	0
Siemens	502	150	277	0
GE	41	0	364	0
Unknown	0	0	37	310
Field strength				
1.5 T	5	23	705	307
3.0 T	1,147	193	736	3
Slice thickness (mm)				
<2	91	26	0	0
2 to <3	762	190	0	0
3 to <4	97	0	304	0
4 to <5	190	0	8	76
5 to <6	12	0	1,126	234
≥6	0	0	3	0

Entries are number of scans. Acquisition heterogeneity for the dedicated CMB model: training (n = 932) was predominantly 3.0 T (99.6%) SWI; external validation (n = 180) was acquired at the 3 validation institutions and is not enumerated here. cSS, cortical superficial siderosis; GRE, gradient-recalled echo T2*; SWI, susceptibility-weighted imaging.