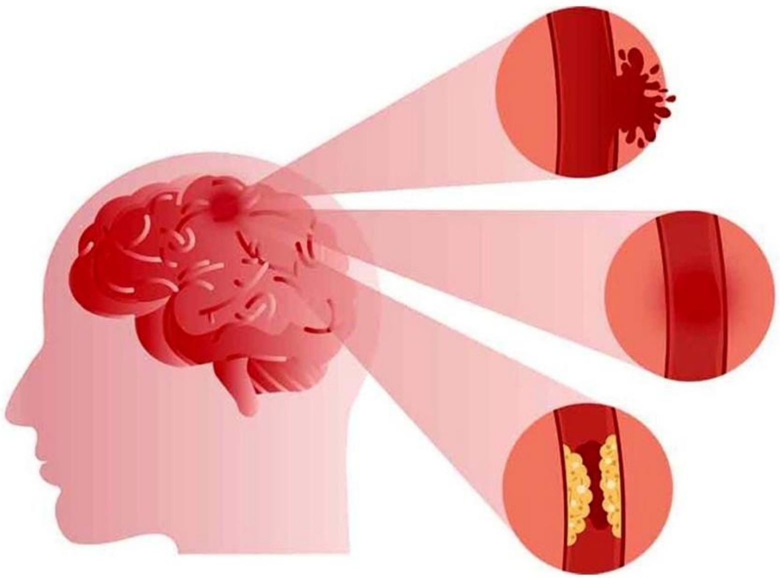
Medical/Bio Research Topics Ⅱ: Week 06 (10.10.2025)

Hands-on AI Segmentation Model Development (1): Data and Prediction Problem

인공지능 분할 모델 개발 실습 (1): 데이터 및 예측 문제

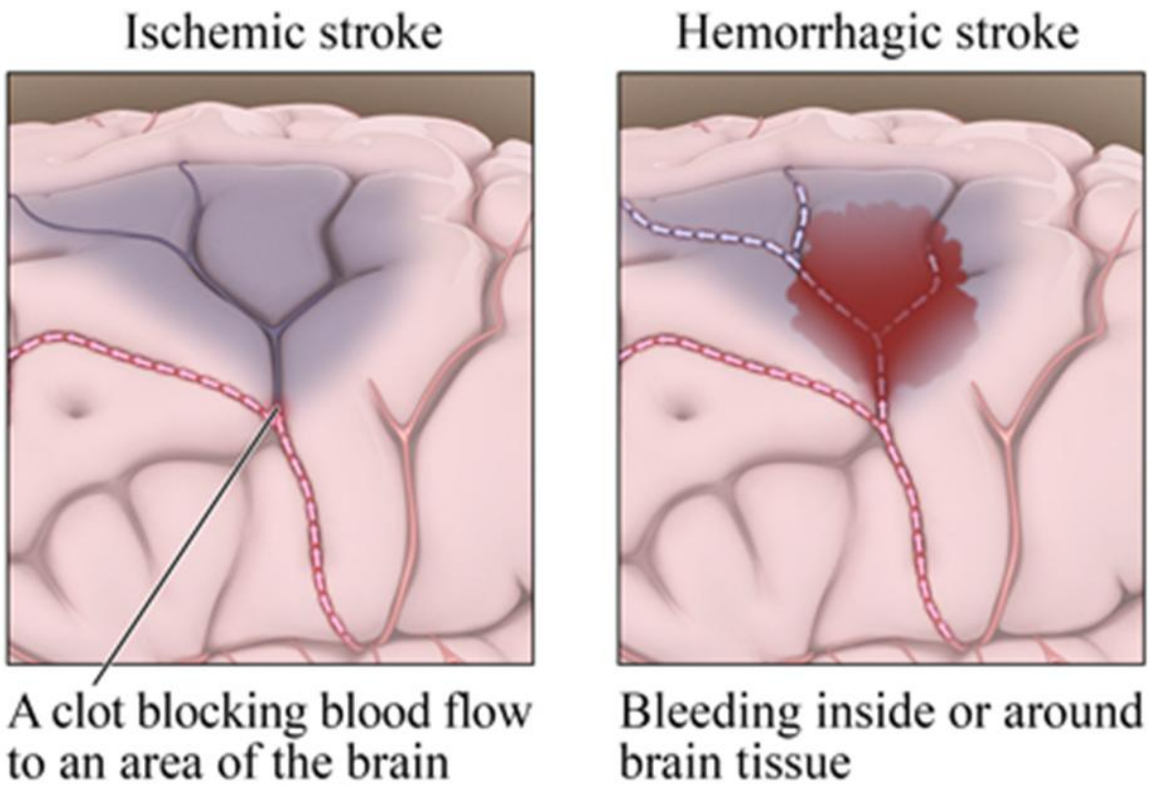
Stroke

- Medical condition in which poor blood flow to the brain causes cell death



[https://mewarhospitals.com/stroke-causes-symptoms-and-treatment/]

- Two types of stroke
  - Ischemic ('ischein' (to restrain) + 'haima' (blood)) stroke
    - Most common type of stroke
    - State where blood supply to a specific area is reduced or blocked
    - The brain cannot get oxygen and nutrients from the blood, so that brain cells begin to die within minutes
  - Hemorrhagic ('haima' (blood) + 'rhegnynai' (to burst forth)) stroke
    - State where a blood vessel has ruptured, causing bleeding
    - The leaked blood results in pressure on brain cells, damaging them

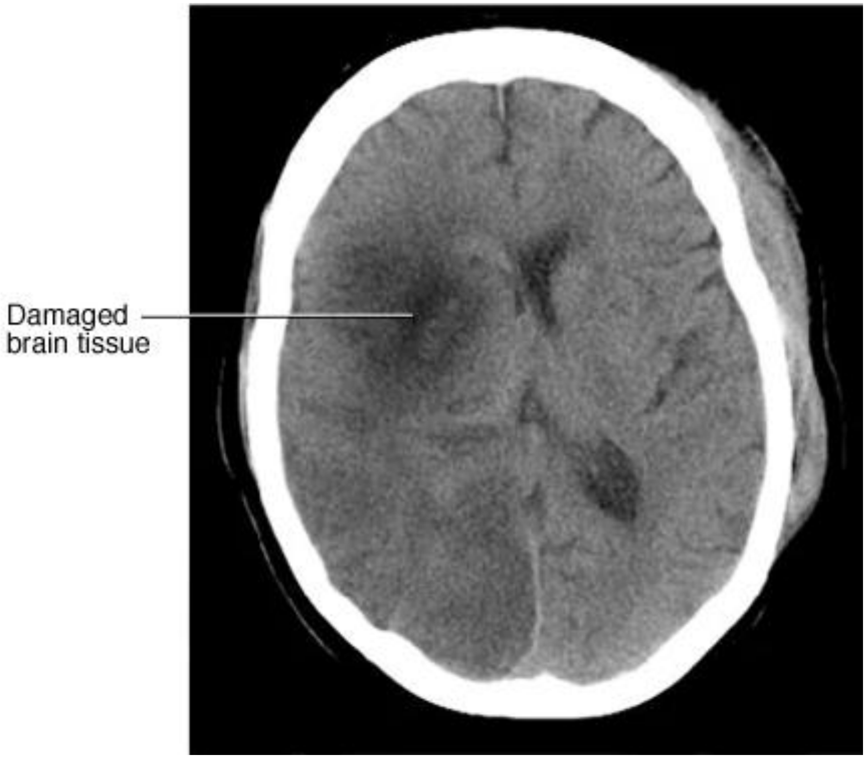


[https://myhealth.alberta.ca/Health/Pages/conditions.aspx?hwid=tp12720]

Ischemic vs. Hemorrhagic Stroke

- Medical emergency
  - Signs and symptoms
    - Trouble speaking and understanding what others are saying
    - Paralysis or numbness of the face, arm, or leg
    - Problems seeing in one or both eyes
    - Headache
    - Trouble walking
  - Early treatment can reduce brain damage and other complications

- Diagnosis
  - Determines the type of stroke
  - Rules out other possible causes of symptoms
  - Tests
    - Physical exam
    - Blood tests
    - CT
    - MRI

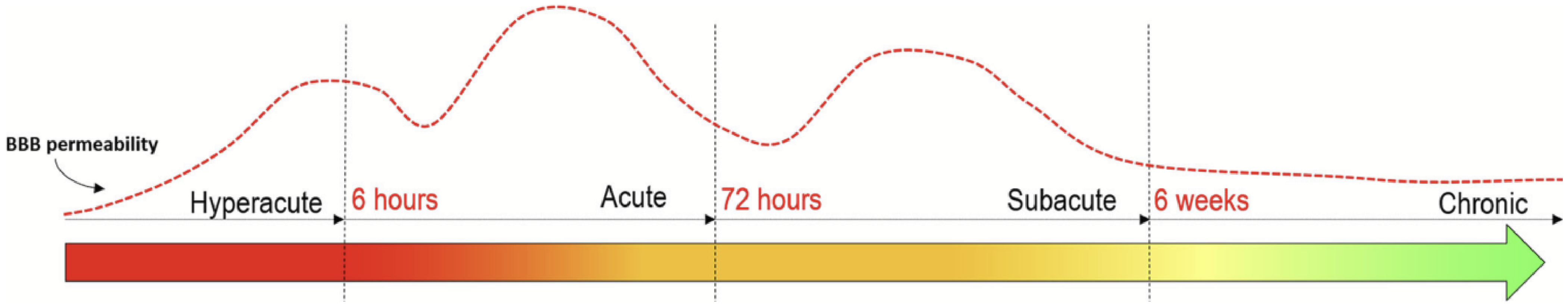


[https://www.mayoclinic.org/diseases-conditions/stroke/diagnosis-treatment/drc-20350119]

CT Scan of Brain Tissue Damaged by Stroke

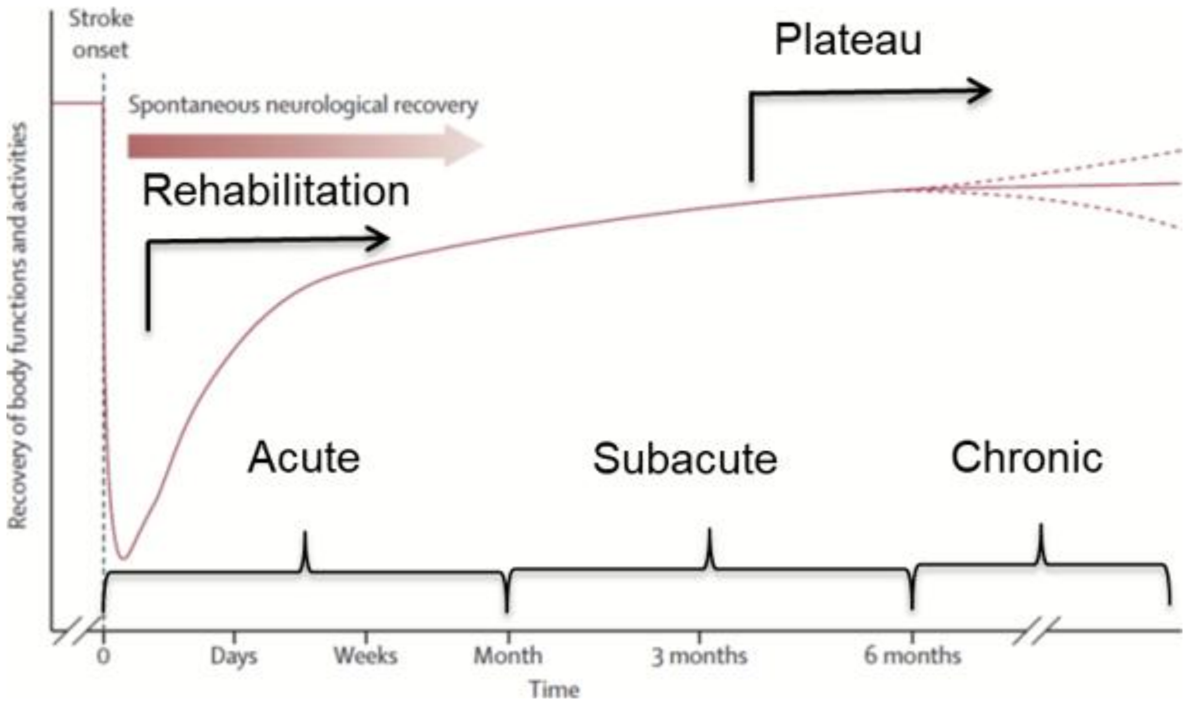
- Emergency treatment
  - Depends on the type of stroke
  - Ischemic stroke
    - Intravenous injection of recombinant tissue plasminogen activator (TPA) to dissolve the blood clot
      - Usually given through a vein in the arm within 4.5 hours of symptom onset
    - Endovascular therapy to directly remove the blood clot
      - Mechanical thrombectomy recommended within 6 hours of symptom onset
  - Hemorrhagic stroke
    - Surgery to remove the blood and relieve pressure on the brain
    - Endovascular therapy to cause blood to clot

- Stages of stroke
  - Acute phase: hours to days after onset
  - Subacute phase: days to weeks
  - Chronic phase: weeks onwards



[Bernardo-Castro et al., 2020]

- Rehabilitation therapy
  - For most stroke survivors depending on the area of the brain involved and the amount of tissue damaged
  - Focuses on helping to recover as much function as possible and return to independent living
  - May begin before discharge and continue after discharge in a rehabilitation unit, as an outpatient, or at home
  - After getting proper treatment during stroke attacks, most of the neurological recovery happens within 3-6 months
    - Most commonly, a stroke recovery plateau occurs around 3-6 months after stroke, in which little or no gains in function happen



[Sharif et al., 2022]

Stroke Recovery Plateau

- Proportional recovery rule
  - The degree of natural recovery up to a stroke recovery plateau is proportional to initial functional impairment [Winters et al., 2015]

$$\underbrace{\Delta \mathrm{FMA}-\mathrm{UE}_{\mathrm{predicted}}}_{\text{Predicted recovery}} = 0.7 \cdot (\underbrace{66-\mathrm{FMA}-\mathrm{UE}_{\mathrm{initial}}}_{\text{Potential recovery}}) + 0.4$$

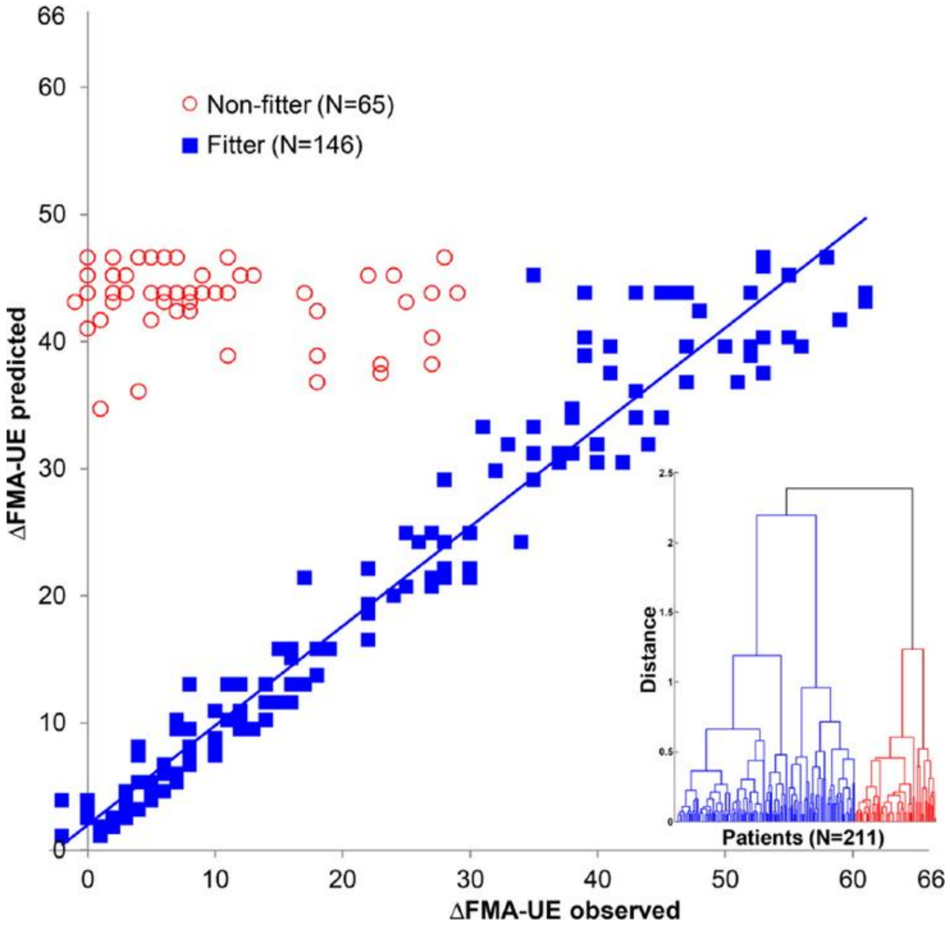
$$\approx 0.7 \cdot (\text{maximal potential recovery})$$

FMA-UE, Fugl-Meyer assessment of the upper extremity

  - Applied to different functional domains including upper and lower limb motor, aphasia, and neglect

$$\Delta\text{FMA-UE}_{\text{observed}} = \text{FMA-UE}_{\text{6months}} - \text{FMA-UE}_{\text{initial}}$$

$$\Delta\text{FMA-UE}_{\text{predicted}} = 0.7\times\left(66 - \text{FMA-UE}_{\text{initial}}\right) + 0.4$$

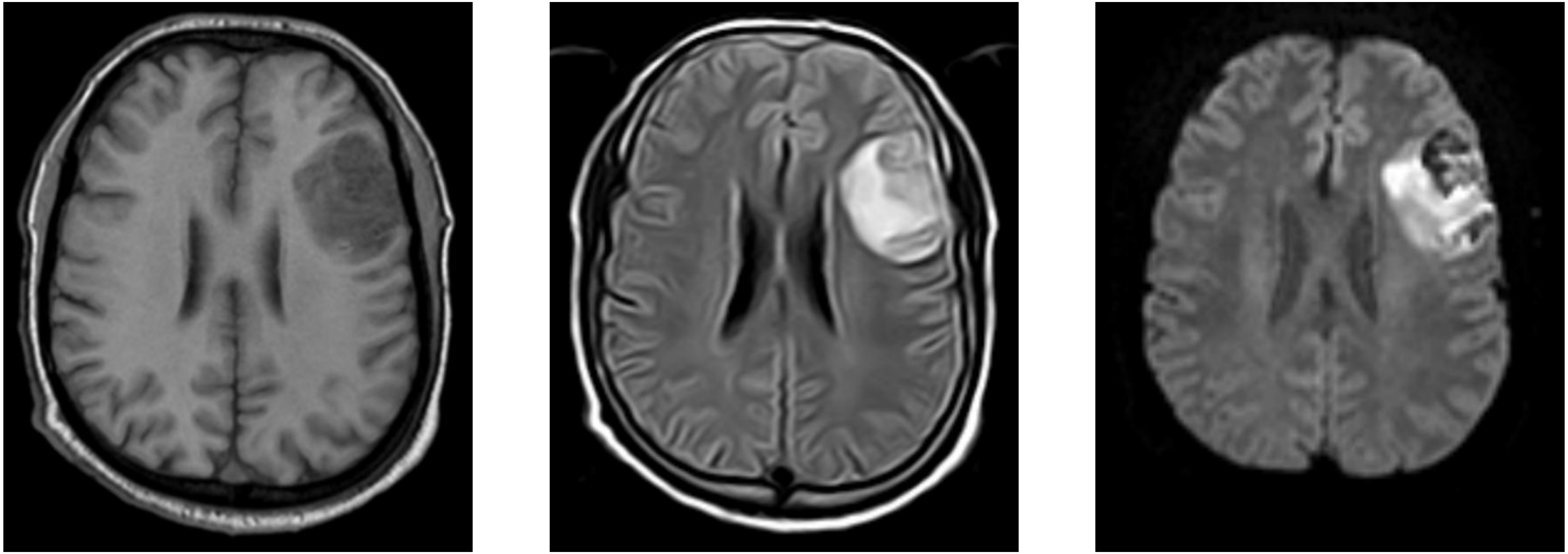


[Winters et al., 2015]

Proportional Motor Improvement in the Upper Limb

Stroke Lesion

- Ischemic lesion
  - Acute ischemic lesion
  - Subacute/chronic infarct (permanent tissue damage)
- Hemorrhagic lesion
  - Intracerebral hemorrhage (ICH)
  - Subarachnoid hemorrhage (SAH)

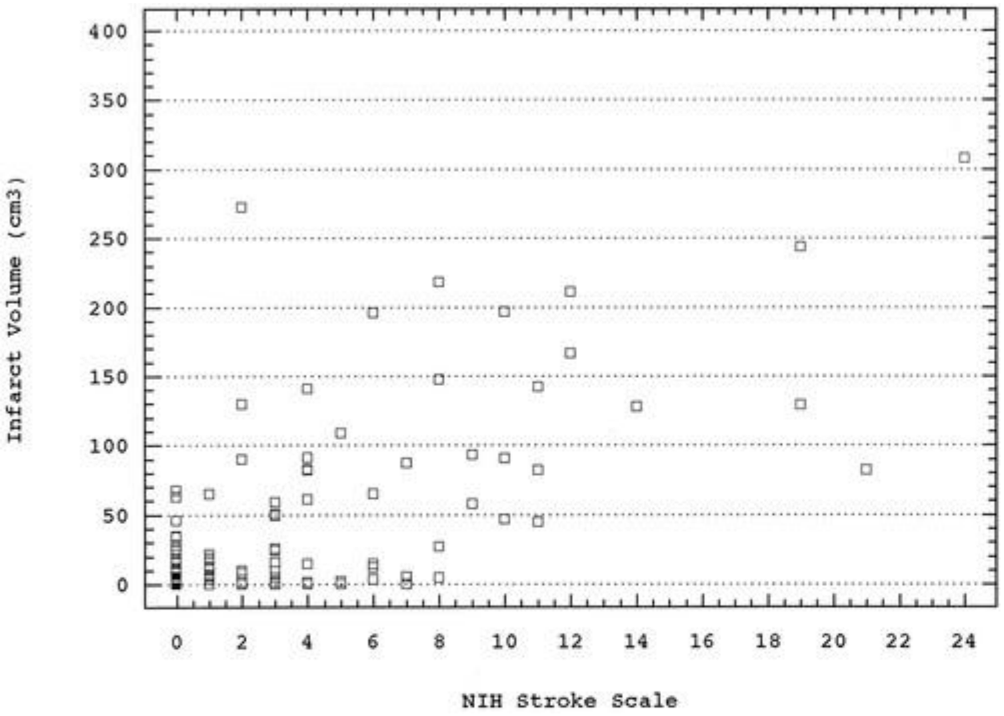


T1-weighted FLAIR Diffusion-weighted

[https://www.mayoclinic.org/diseases-conditions/stroke/diagnosis-treatment/drc-20350119]

Stroke Lesion Displayed as Altered Signals in MRI

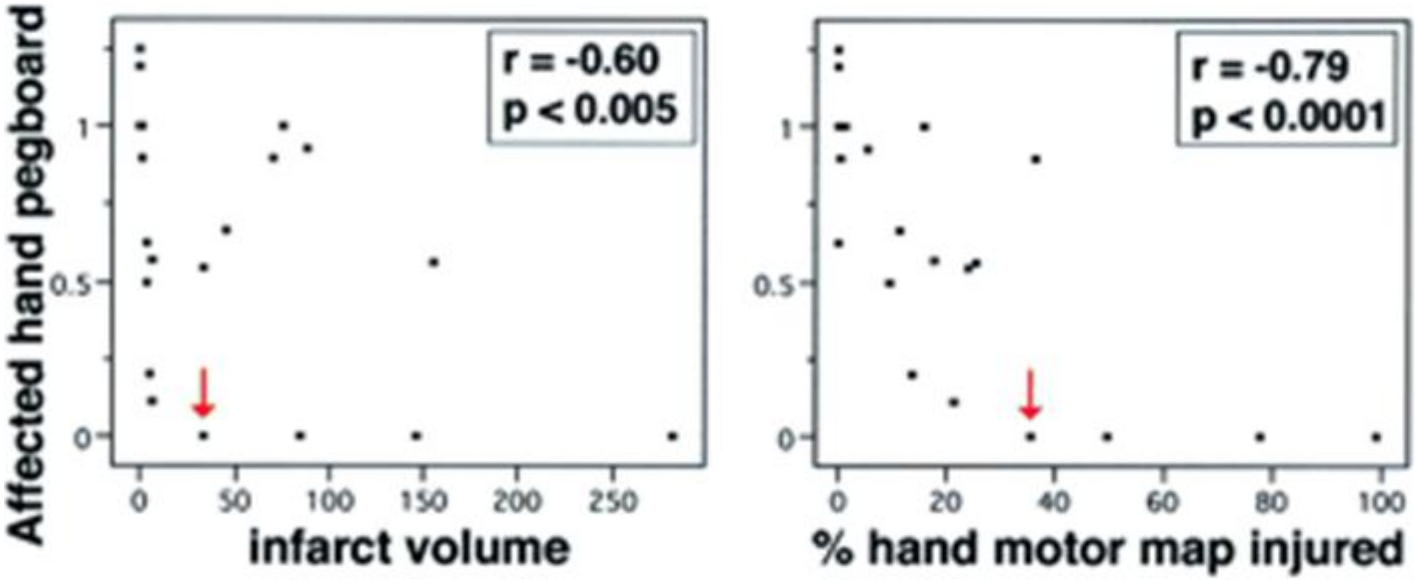
- Brain-behaviour relationship in stroke rehabilitation
  - Lesion size
    - Lesion volume correlates with clinical outcome



[Saver et al., 1999]

– Lesion location

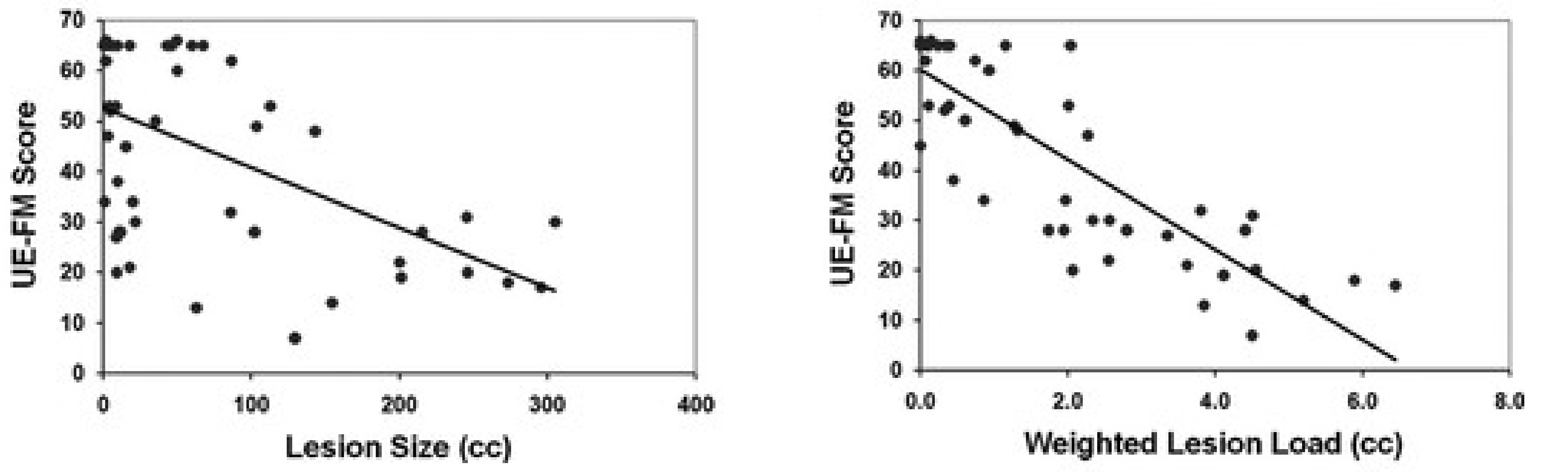
- Motor performance correlates with the fraction of hand motor map injured more strongly than with lesion volume



[Crafton et al., 2003]

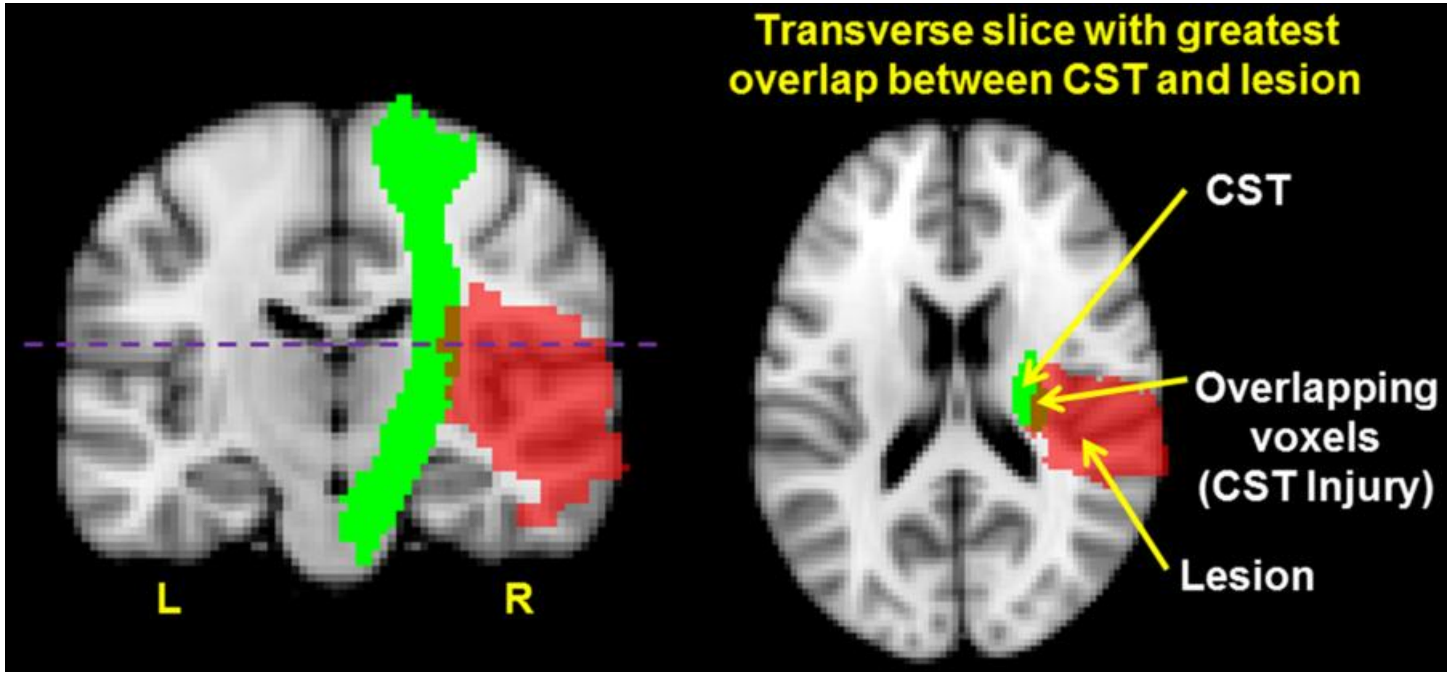
– Lesion load: lesion overlap with extant brain structures

- Motor impairment correlates with the proportion of the corticospinal tract injured more strongly than with lesion volume



[Zhu et al., 2010]

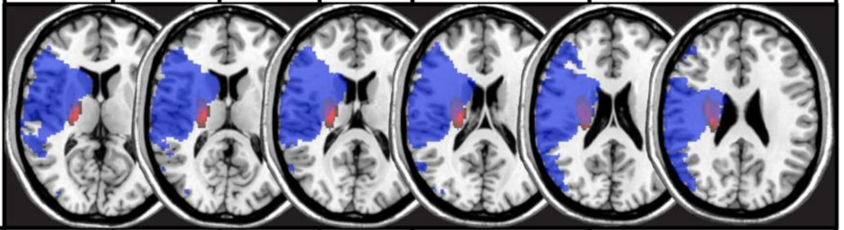
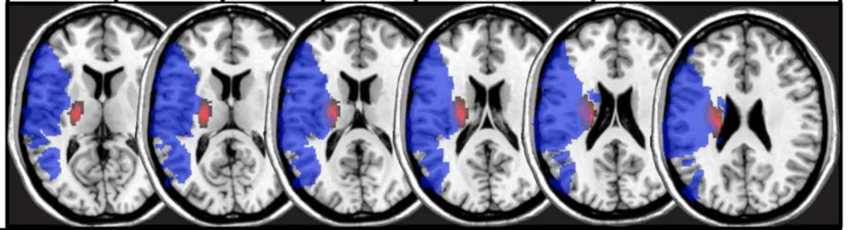
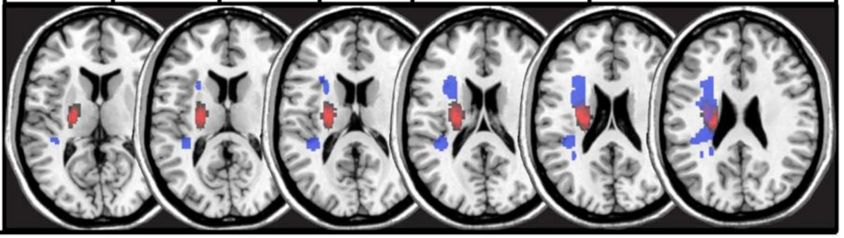
$$\text{CST Injury} = \left(\frac{\text{Number of overlapping voxels between the CST and lesion for the transverse slice}}{\text{Total number of CST voxels for the transverse slice}}\right) \times 100\%$$



[Lam et al., 2020]

Computation of Corticospinal Tract Lesion Load

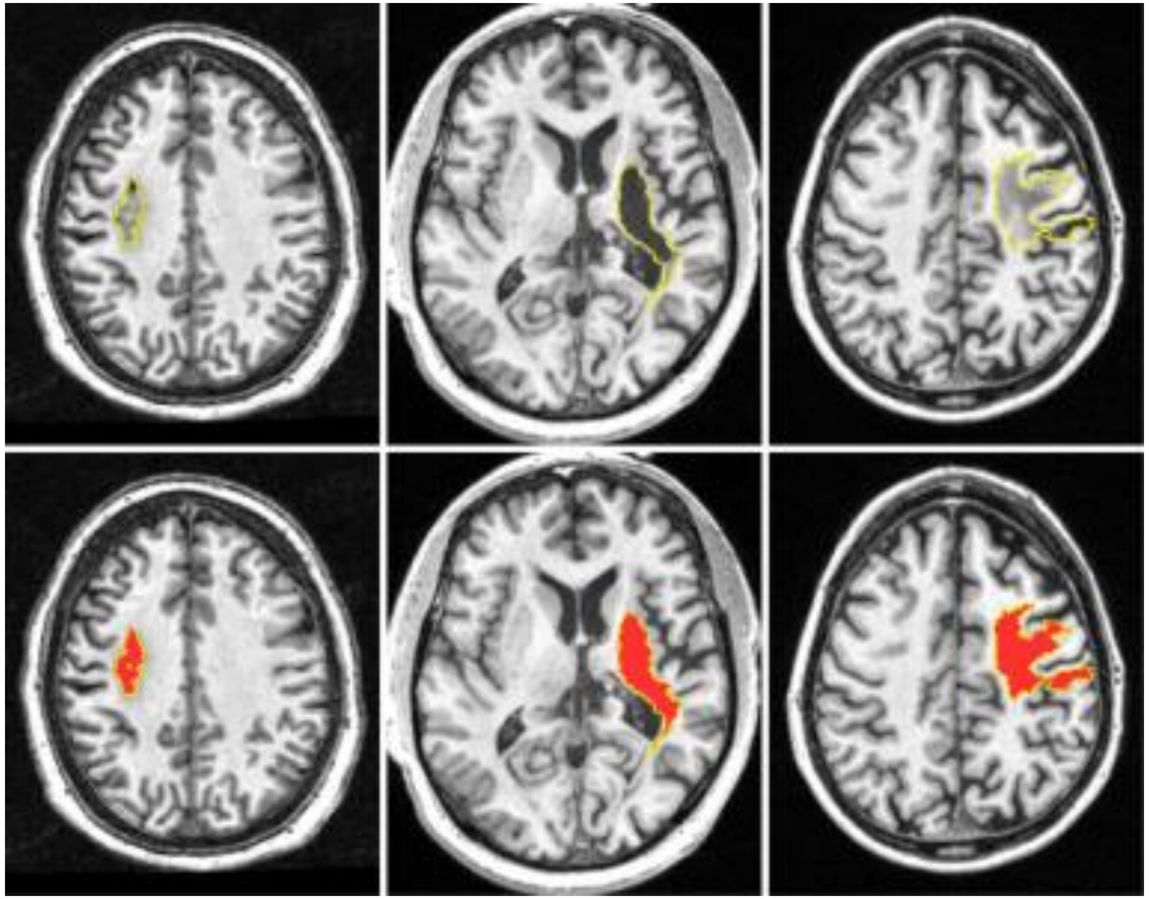
- Corticospinal tract lesion load can predict motor outcome

Patients	FM-UE		NIHSS		Lesion Size (cc)	Weighted Lesion Load (cc)
	Pre	Post	Pre	Post		
A	8	8	18	11	149	9.19
						
B	11	65	13	1	143.81	4.38
						
C	8	12	18	6	20.01	7.45
						

[Feng et al., 2015]

Lesion Segmentation

- Critical in stroke rehabilitation research
  - For the quantification of lesion burden
  - For accurate image processing
- Still faces challenges and difficulties primarily due to variations of lesions in terms of shape, size, and location
- Manual segmentation remains the gold standard, but it is time-consuming, subjective, and requires neuroanatomical expertise



[Wu et al., 2023]

Variations in Stroke Lesions

- Anatomical Tracings of Lesions After Stroke (ATLAS) R2.0 dataset [https://fcon_1000.projects.nitrc.org/indi/retro/atlas.html]
  - Primarily led by the Mark and Mary Stevens Neuroimaging and Informatics Institute at the University of Southern California (USC)
  - Released in 2021 by expanding upon and replacing ATLAS R1.2 released in 2018
  - Largest dataset of its kind
  - Intended to be a resource for the scientific community to develop more accurate lesion segmentation algorithms
  - Derived from diverse, multi-site data from 44 research cohorts worldwide

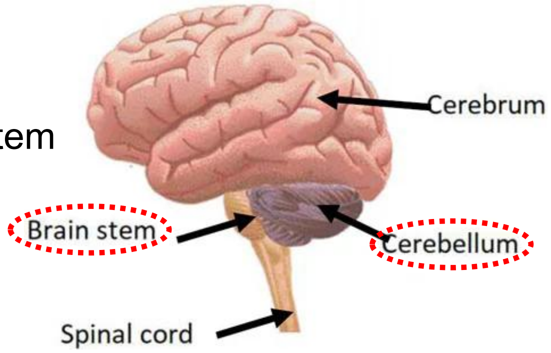
- Includes T1-weighted MRI scans and manually segmented lesion masks (*n* = 1,271)
  - Training and test sets derived from 33 research cohorts
    - Samples from each research cohort are randomly assigned to either training or test sets so that they have similar compositions
    - Training set (*n* = 655): publicly released T1-weighted MRI scans and lesion masks
    - Test set (*n* = 300): publicly released T1-weighted MRI scans and hidden lesion masks
  - Generalizability set derived from 11 new cohorts
    - To test the performance of trained algorithms on completely unseen data
    - Generalizability set (*n* = 316): completely hidden T1-weighted MRI scans and lesion masks from separate cohorts

– T1-weighted MRI data
  - Collected on 1.5 Tesla and 3 Tesla MR scanners
    – Each cohort was collected on a single scanner using the same parameters except for 2 cohorts
  - High-resolution with the voxel size of 1 mm^3 or higher

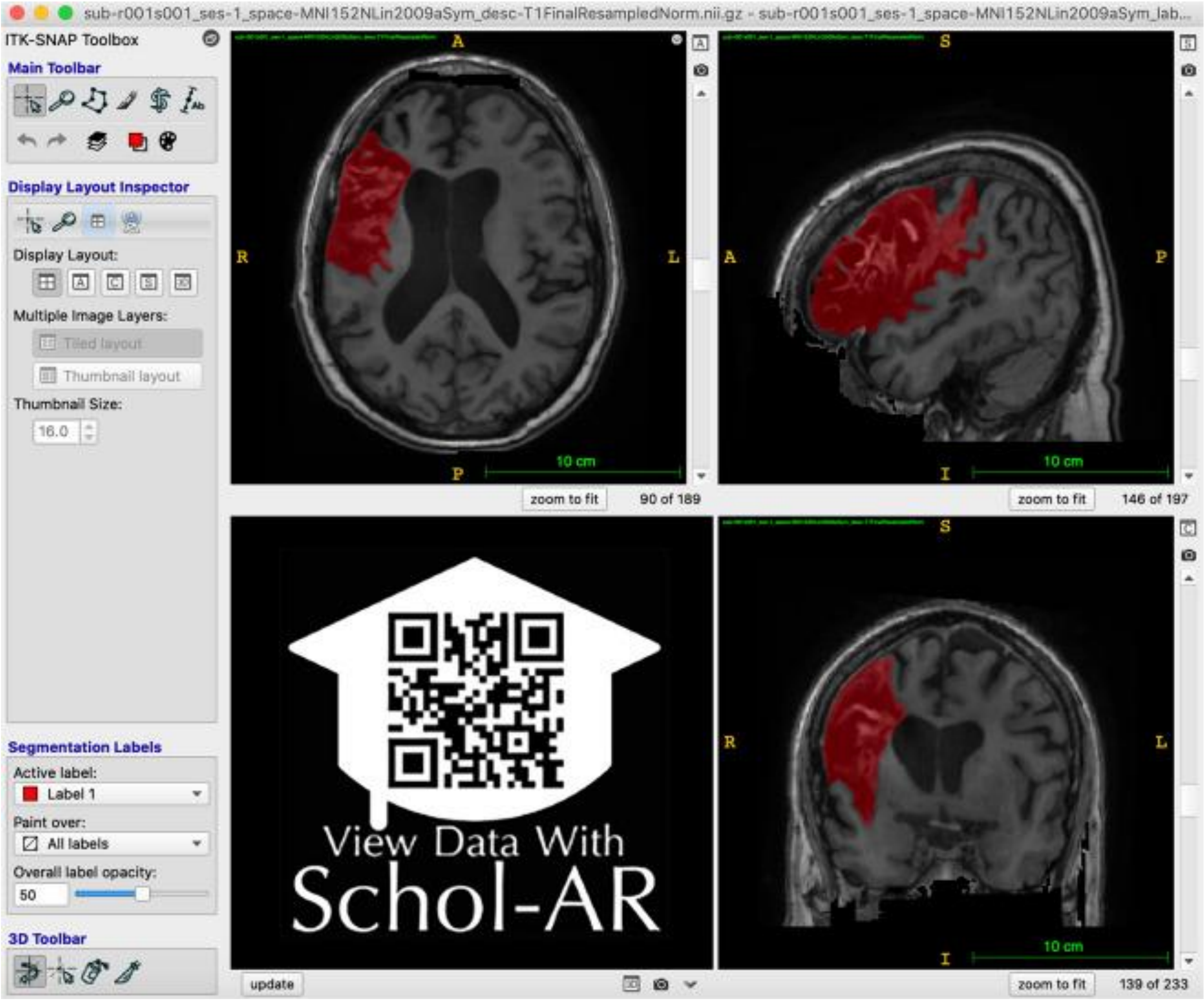
– Lesion masks [Liew et al., 2022]
  - Number of lesions and lesion location were manually recorded

	Subjects with One Lesion			Subjects with Multiple Lesions		
	Left	Right	Other	Unilateral	Bilateral	Other
Training data (n = 655)	173 (26.4%)	187 (28.5%)	46 (7.0%)	47 (7.2%)	121 (18.5%)	81 (12.4%)
Testing data (n = 300)	88 (29.3%)	95 (31.7%)	23 (7.7%)	16 (5.3%)	43 (14.3%)	35 (11.7%)

Located In either
cerebellum or brainstem

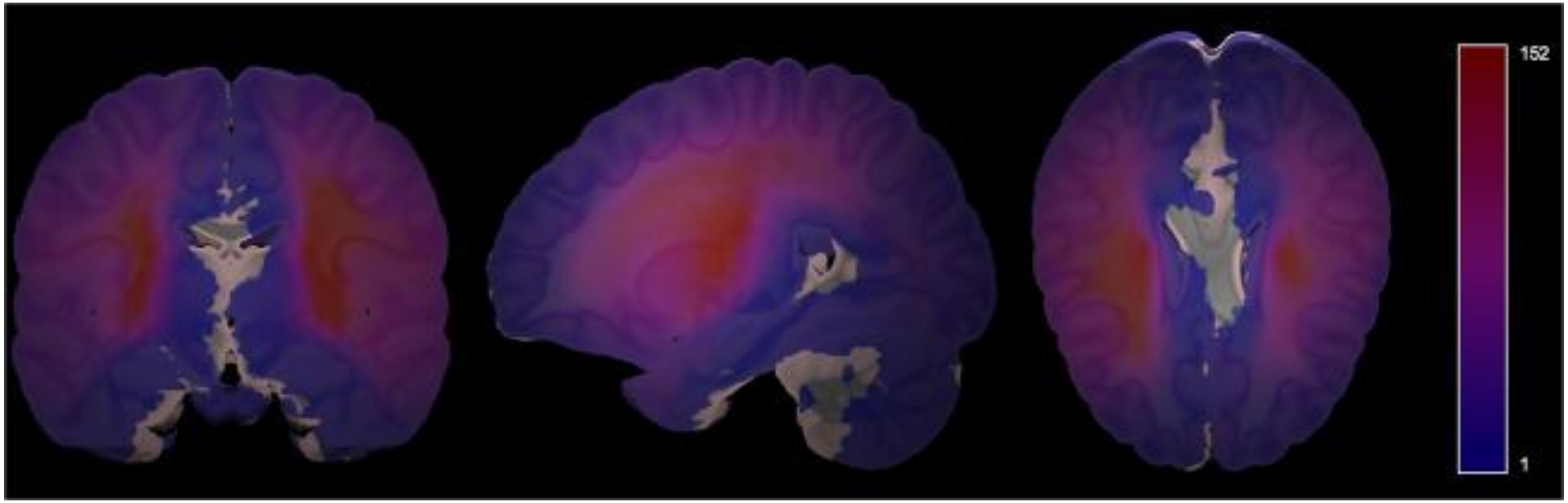


– Lesion identification and manual tracing
  - By using ITK-SNAP [http://www.itksnap.org/]
  - White matter hyperintensities of presumed vascular origin and perivascular spaces were excluded from lesion masks as much as possible
  - All identified lesions for each subject were reviewed for quality control by two additional trained raters



[Liew et al., 2022]

Manual Lesion Segmentation in ITK-SNAP



[Liew et al., 2022]

Lesion Overlap across All Subjects (*n* = 955) Overlaid on the MNI Template Brain

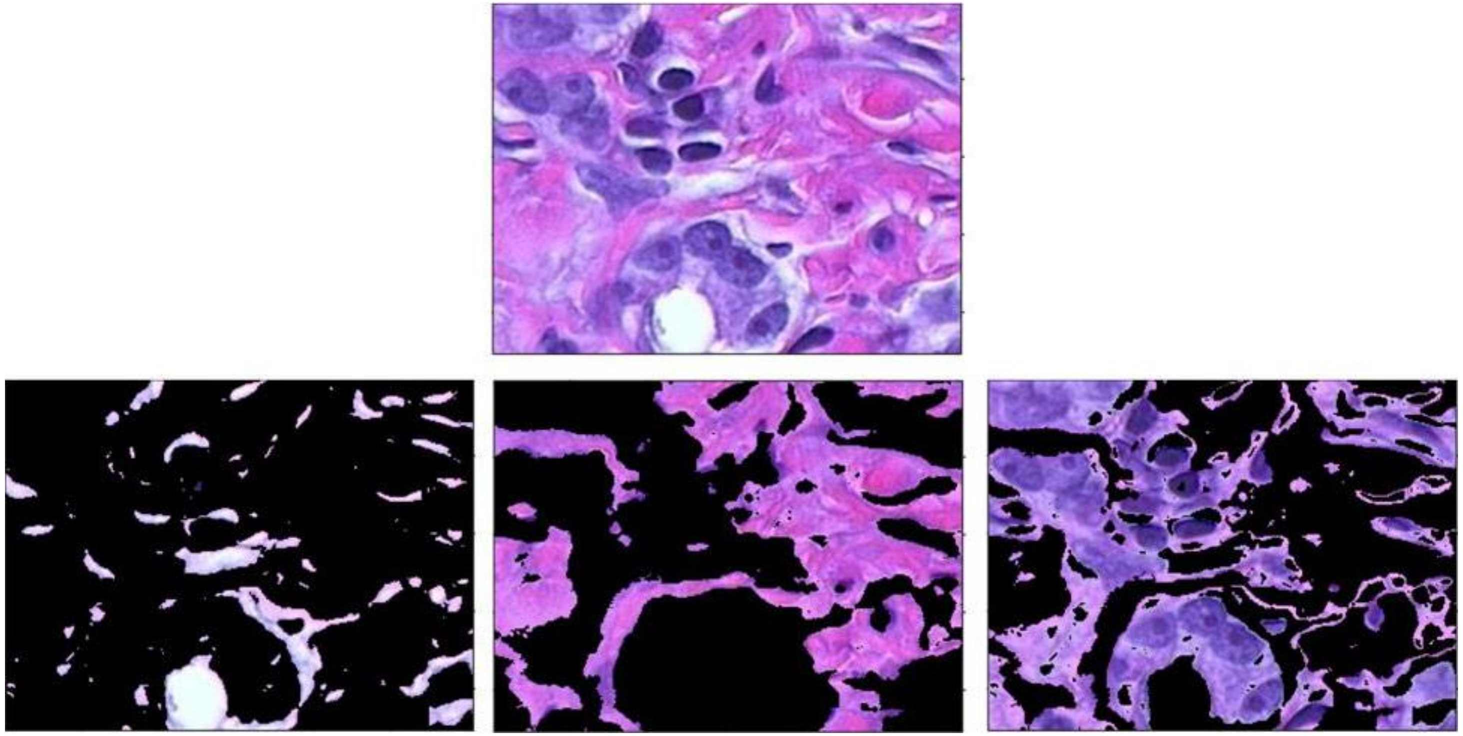
Image Segmentation

- Technique in digital image processing and analysis to partition an image into multiple parts or areas, often based on the characteristics of the pixels/voxels in the image
  - Involves converting an image into a collection of regions of pixels/voxels that are represented by a mask or a labeled image
- A common application in medical imaging is to detect and label pixels/voxels that represent an abnormality in the brain or other organs

- Algorithms and techniques [https://www.mathworks.com/discovery/image-segmentation.html]
  - Developed over the years using domain-specific knowledge to effectively solve segmentation problems in specific application areas such as medical imaging, automated driving, video surveillance, and machine vision
  - Thresholding
    - Performs thresholding on a greyscale or color image to create a binary image

– Clustering

- Creates a segmented labeled image using a specific clustering algorithm such as K-means clustering
- For example, to distinguish between tissue types in an image of body tissue stained with hematoxylin and eosin



– Graph-based segmentation

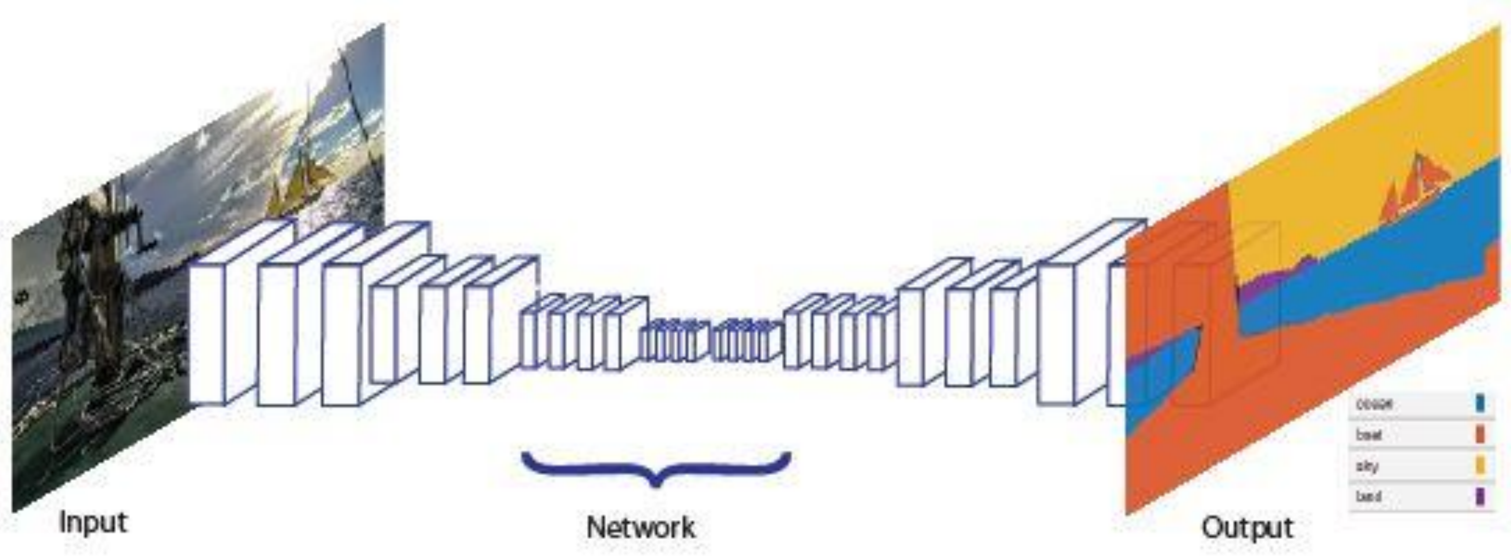
- Enables to segment an image into foreground and background areas

– Region growing

- Examines neighboring pixels/voxels of initial seed points and determines iteratively whether the pixel neighbors should be added to the area

Deep Learning-based Image Segmentation

- Associates every pixel/voxel of an image with a class label by using neural networks



[https://www.mathworks.com/discovery/image-segmentation.html]

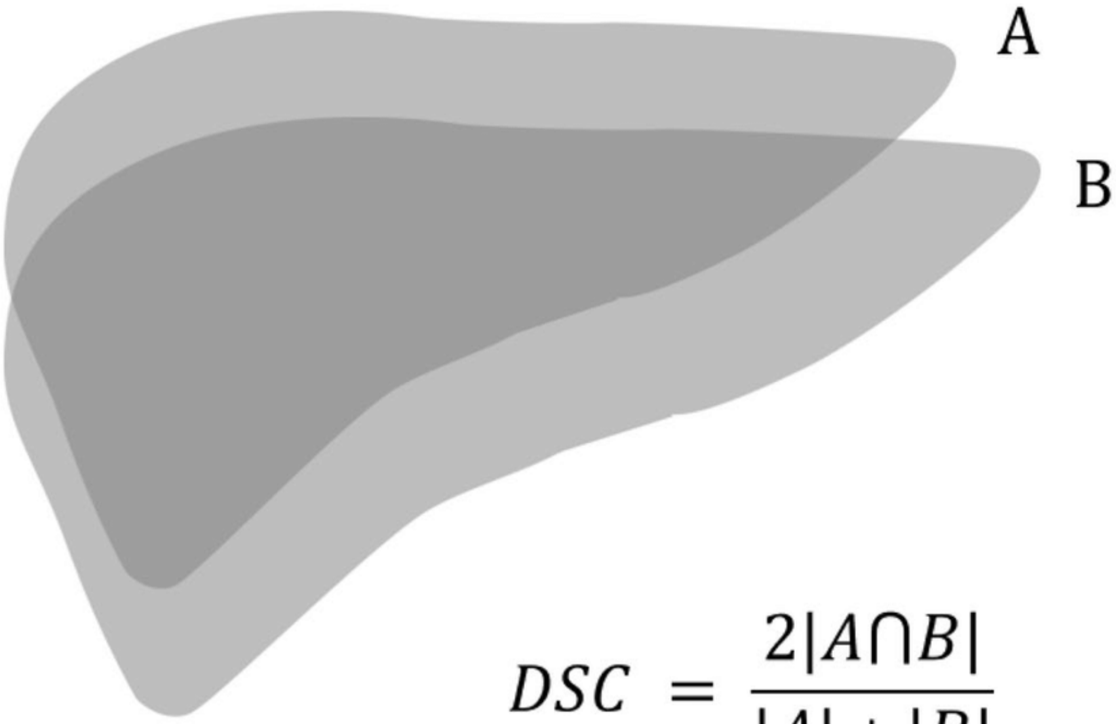
- Leverages the power of deep learning algorithms to analyze image features at various scales, offering improved accuracy and efficiency compared to traditional methods
- Processes the entire image in smaller sections vs. holistically
  - Patch-wise segmentation
    - Takes a small patch around a voxel as the input and traverses the entire volume by repeatedly taking patches
    - Redundant calculations caused by overlapping patches decreases computational efficiency
  - Semantic-wise segmentation
    - Takes the entire volume or a large patch as the input
    - Prone to overfitting during training due to class imbalance

- Segmentation performance
  - Dice similarity coefficient (DSC, Dice-Sørensen coefficient or Dice coefficient) [Dice, 1945]
    - 2 * |X ∩ Y| / (|X| + |Y|), where X and Y are the predicted and ground truth segmentations
    - Measures the overlap between predicted and ground truth segmentations
    - F_1 score that is a harmonic mean of precision and recall
      - Precision (True Positive Value (TPV)) = TP / (TP + FP)
      - Recall (sensitivity) = TP / (TP + FN)
      - F_1 score = 2 / ((1 / precision) + (1 / recall)) = 2TP / (2TP + FP + FN)
    - Range: 0 (no overlap) to 1 (perfect overlap)
    - Sensitive to both false positives and false negatives

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP), hit**	**False negative (FN), type II error, miss, underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP), type I error, false alarm, overestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	$\boxed{F_1 \text{ score} = \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

[https://en.wikipedia.org/wiki/Confusion_matrix]

DSC or F_1 Score in a Confusion Matrix



$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

DSC: Dice similarity coefficient

[Lee et al., 2018; https://www.mathworks.com/help/images/ref/dice.html]

Computation of DSC

- Intersection over Union (IoU, Jaccard Index)
  - $|X \cap Y| / |X \cup Y|$
  - Measures the overlap ratio of the intersection to the union of predicted and ground truth segmentations
  - Range: 0 (no overlap) to 1 (perfect overlap)
  - Stricter than DSC by penalizing errors more heavily
- Mean Intersection over Union (mIoU)
  - Average of IoU scores for all classes
  - Provides an overall measure of segmentation quality across multiple classes
  - Range: 0 (no overlap) to 1 (perfect overlap)
  - Useful for multi-class segmentation tasks

- Hausdorff distance
  - max(h(X,Y), h(Y,X)), where h(X,Y) = max(min(d(x,y))) for x in X, y in Y and h(Y,X) = max(min(d(y,x))) for y in Y, x in X
  - Measures the maximum distance between the boundaries of predicted and ground truth segmentations
  - Range: 0 to ∞ (lower is better)
  - Sensitive to outliers, useful for evaluating boundary accuracy
- Average Surface Distance (ASD)
  - Average of distances between surfaces of predicted and ground truth segmentations
  - Measures the average error in boundary delineation
  - 0 to ∞ (lower is better)
  - Less sensitive to outliers than Hausdorff Distance

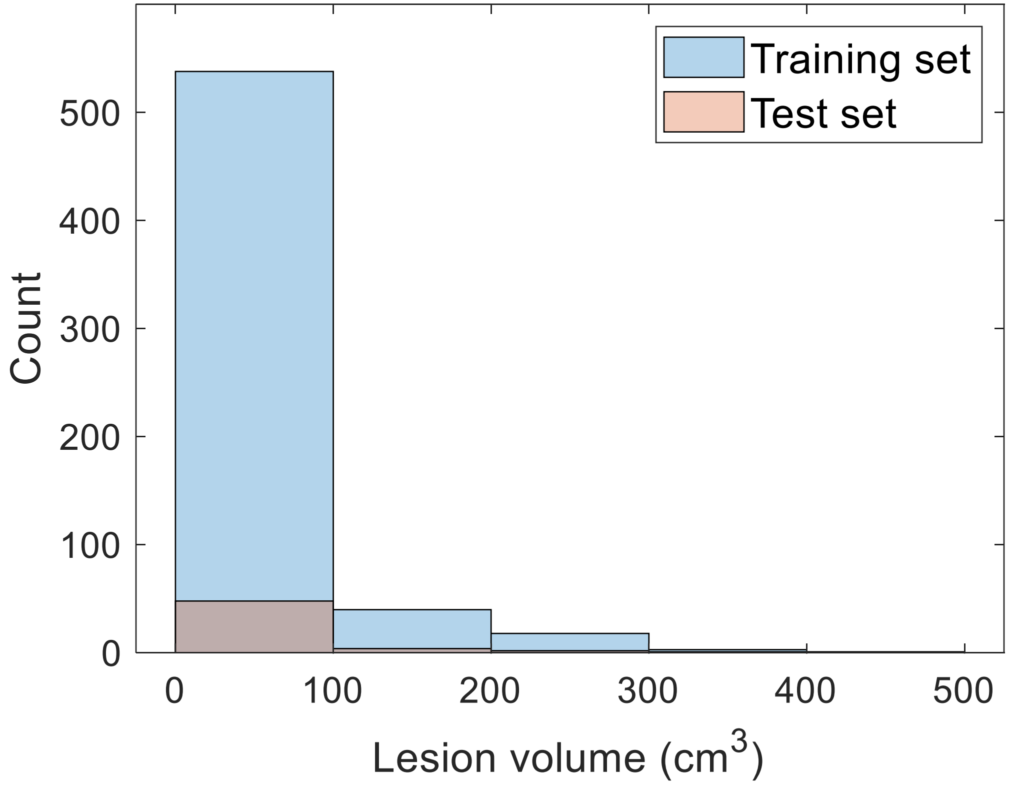
- Accuracy
  - (Correctly Classified Elements) / (Total Elements) = (TP + TN) / (Total Elements)
  - Measures the proportion of elements correctly classified across all classes
  - Range: 0 (completely incorrect classification) to 1 (perfect classification)
- Sensitivity and specificity
  - Sensitivity = TP / (TP + FN), specificity = TN / (TN + FP)
  - Measure the model's ability to correctly identify positive and negative cases
  - Range: 0 (complete failure to detect positive/negative cases) to 1 (perfect detection of positive/negative cases)
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
  - Measures the model's ability to distinguish between classes
  - Range: 0.5 (random guessing) to 1 (perfect classification)

- Deep learning-based lesion segmentation
  - Specific application of image segmentation to medical images by targeting abnormal tissues or pathological regions
  - Challenges
    - Class imbalance (lesions often small compared to healthy tissue)
    - Variabe in lesion shape, size, and location
    - Artifacts and noise common in medical images
  - Data considerations
    - Often works with 3D volumetric images (CT, MRI scans)
    - Requires expert annotations, which can be costly and time-consuming
    - Employs data augmentation to efficiently use limited training data
    - May benefit from multi-modal data integration

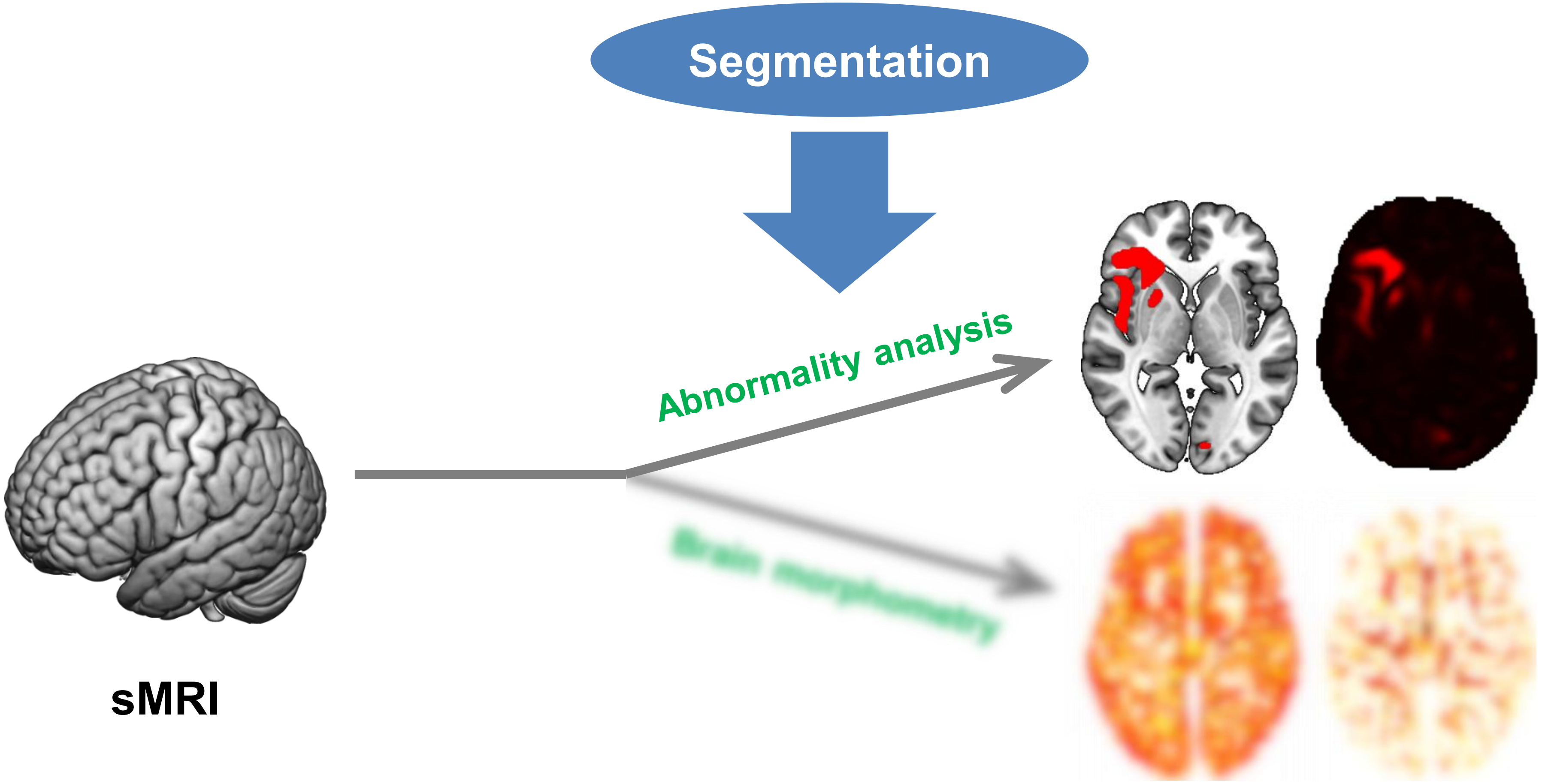
- Performance metrics
  - Usually uses domain-specific metrics like DSC and Hausdorff distance
  - Emphasizes both quantitative accuracy and clinical relevance
- Specialized architectures
  - Uses common segmentation models such as U-Net and its variants
    - Ability to capture both local and global context
    - Skip connections that preserve fine details, crucial for precise lesion boundaries
  - Designed to handle medical imaging specificities

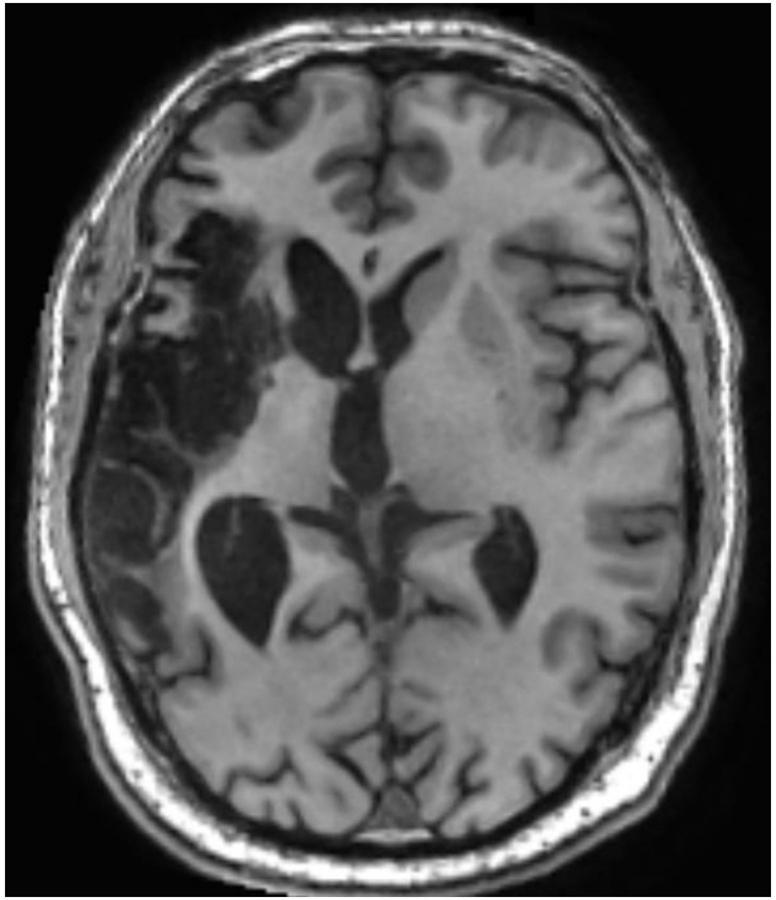
Dataset

- ATLAS R2.0 dataset for training (n = 655)
  - Training set: n = 600
    - T1-weighted MRI scans: train/Brain/001-600.nii.gz
    - Lesion masks: train/Lesion/001-600.nii.gz
  - Test set: n = 55
    - T1-weighted MRI scans: test/Brain/001-055.nii.gz
    - Lesion masks: hidden



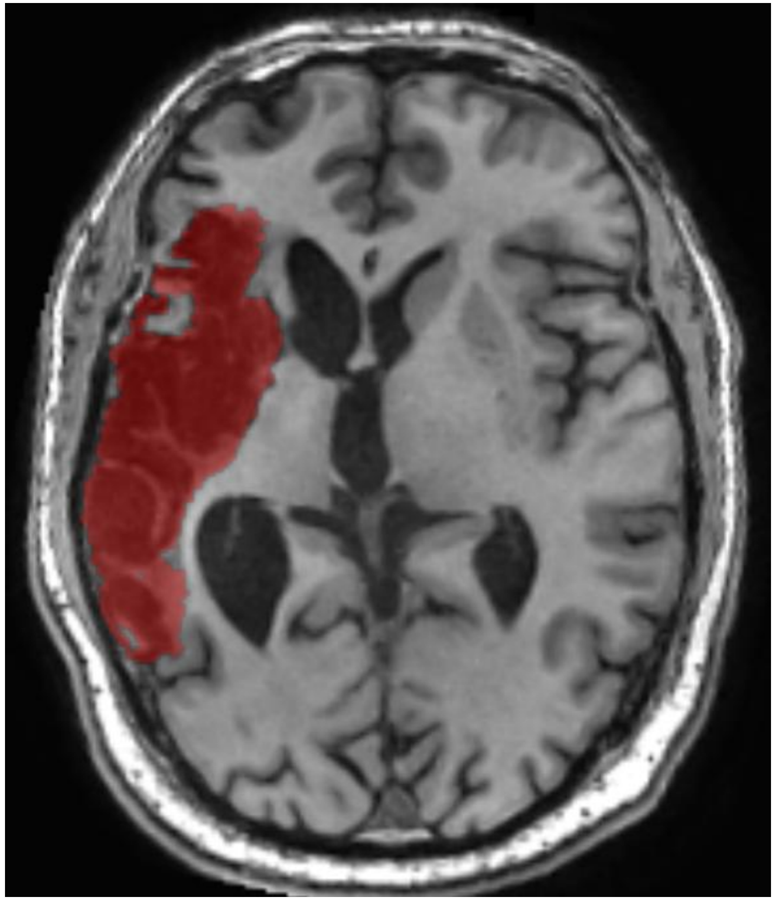
Distribution of Lesion Volume


Segmentation
Abnormality analysis
sMRI



Brain

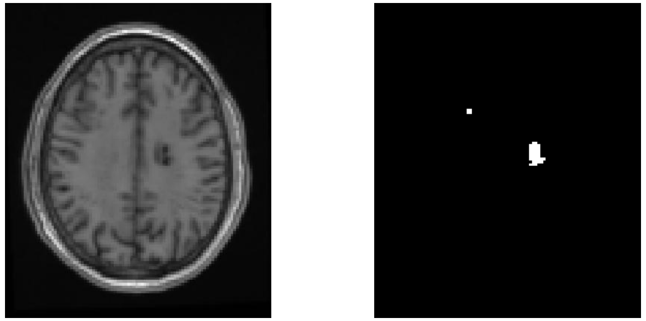
Lesion



Lesion-overlaid brain

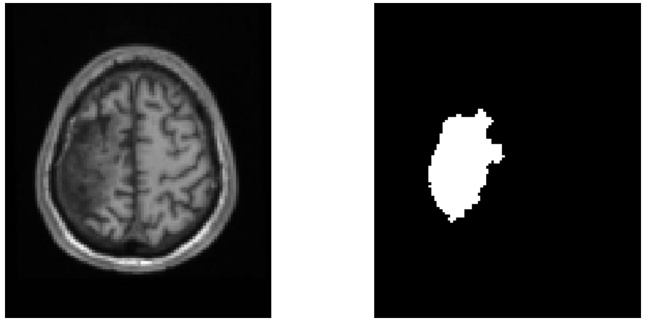
Example Pair of a T1-weighted MRI Scan and a Lesion Mask

001: 2.594 cm^3

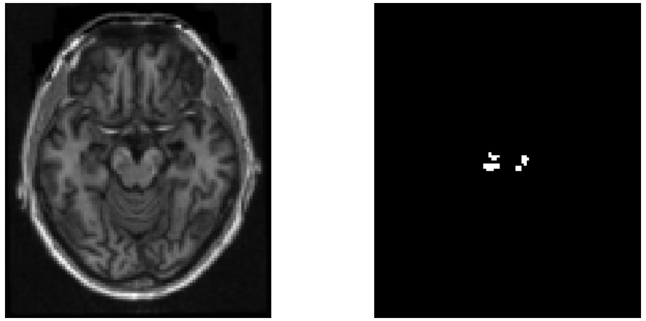


002: 61.096 cm^3

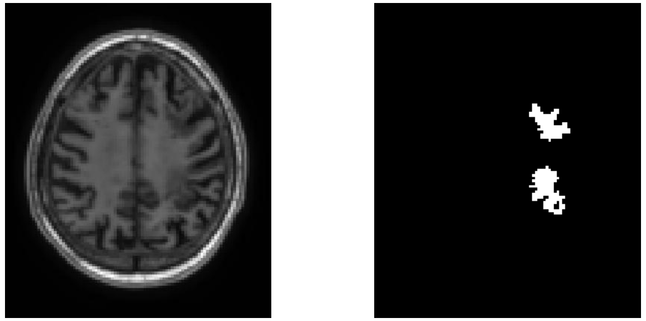
003: 139.979 cm^3



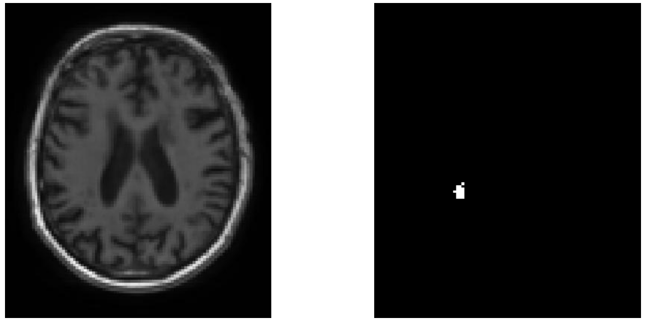
004: 3.145 cm^3

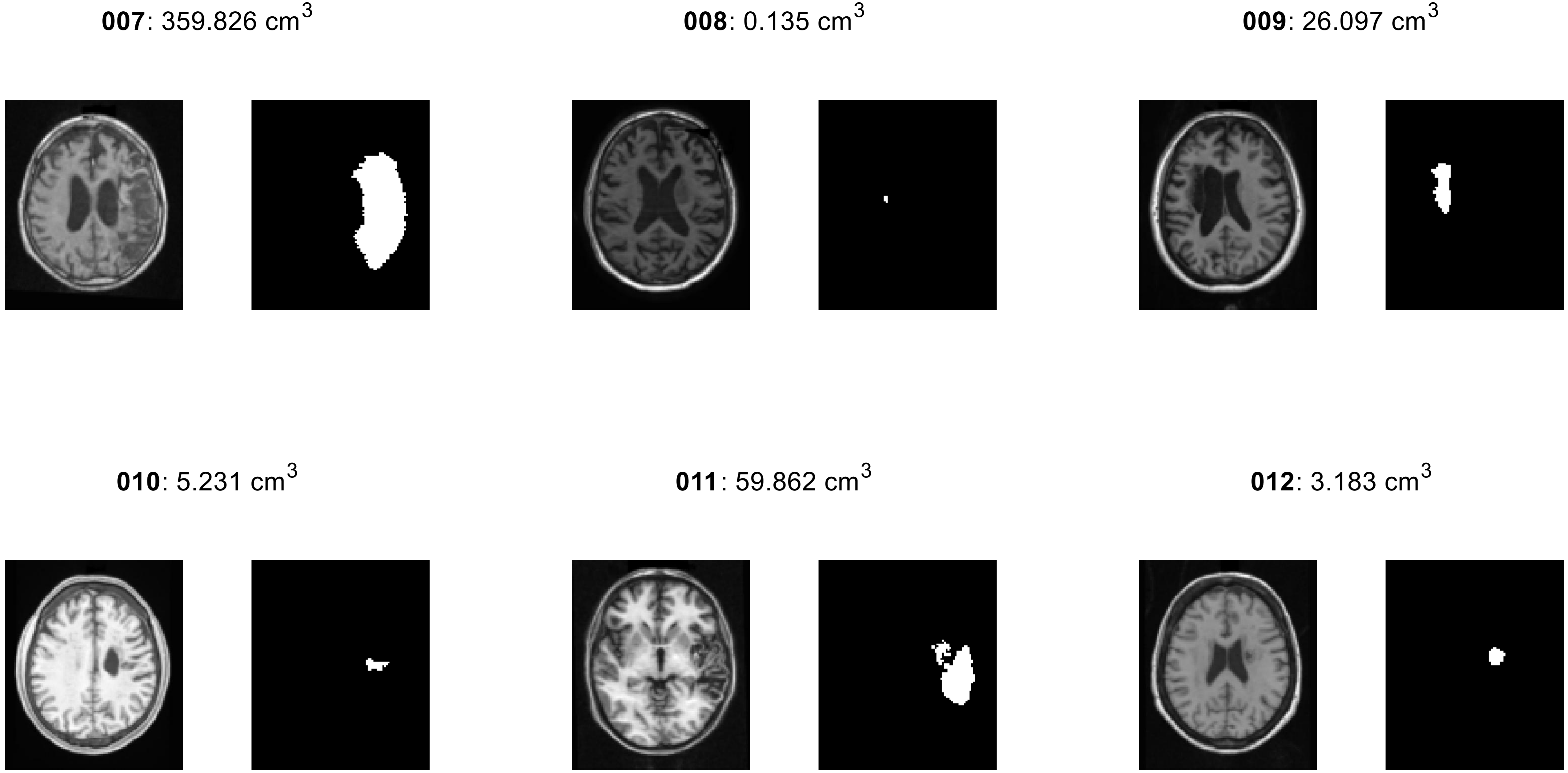


005: 17.424 cm^3

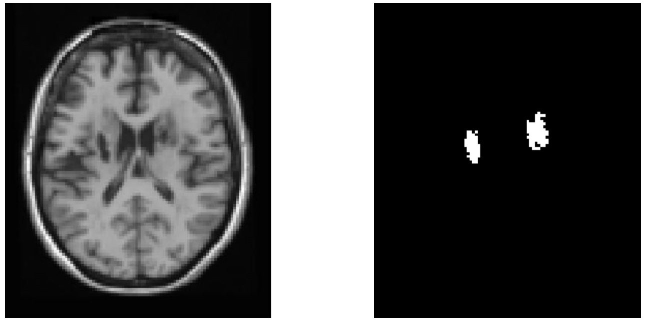


006: 0.915 cm^3

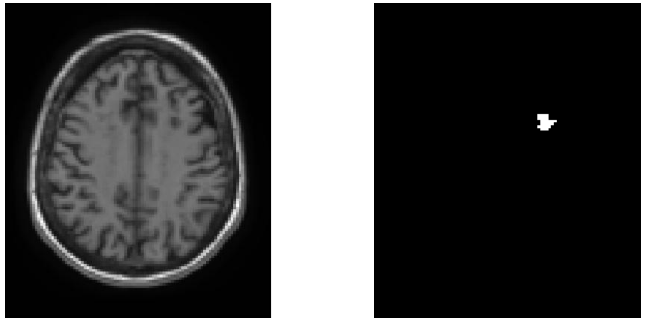



007: 359.826 cm^3
008: 0.135 cm^3
009: 26.097 cm^3
010: 5.231 cm^3
011: 59.862 cm^3
012: 3.183 cm^3

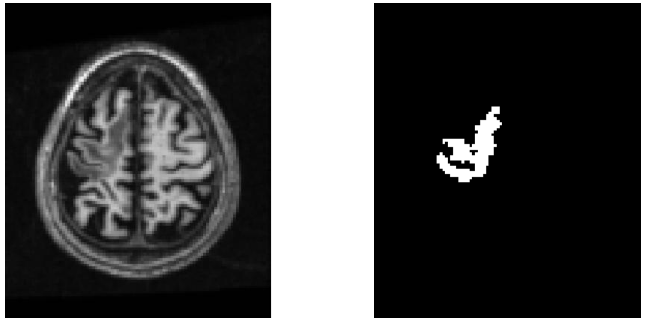
013: 9.897 cm^3



014: 1.156 cm^3

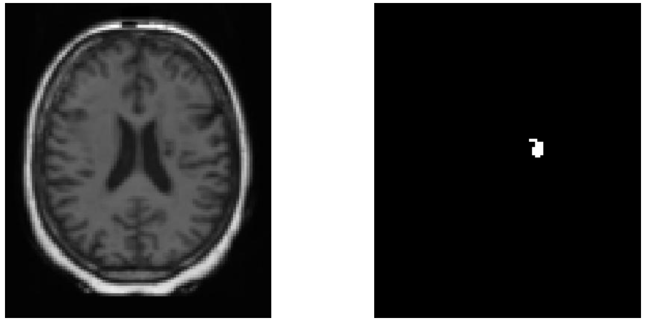


015: 36.535 cm^3



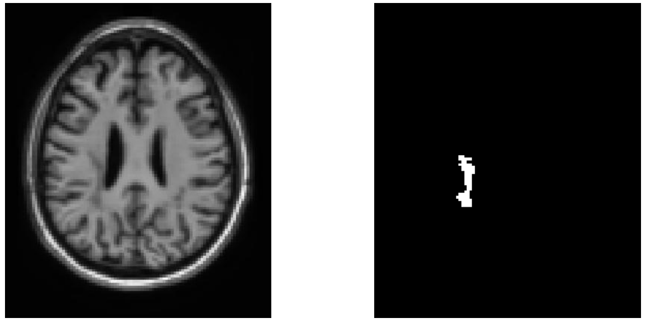
016: 22.883 cm^3

017: 1.038 cm^3



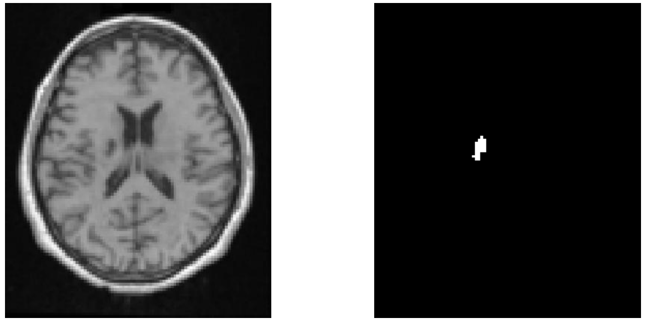
018: 1.214 cm^3

019: 7.469 cm^3



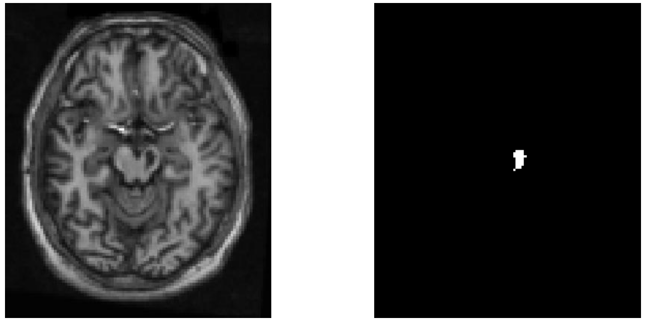
020: 0.375 cm^3

021: 1.180 cm^3

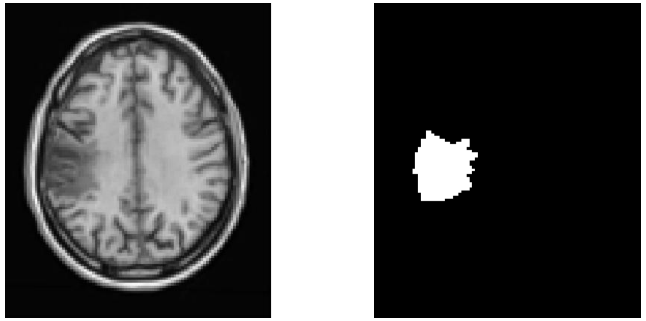


022: 34.330 cm^3

023: 0.726 cm^3

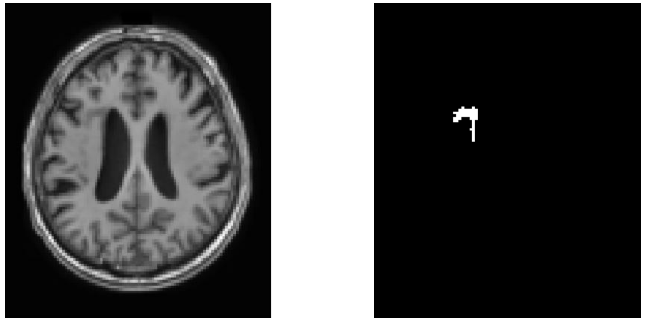


024: 67.076 cm^3

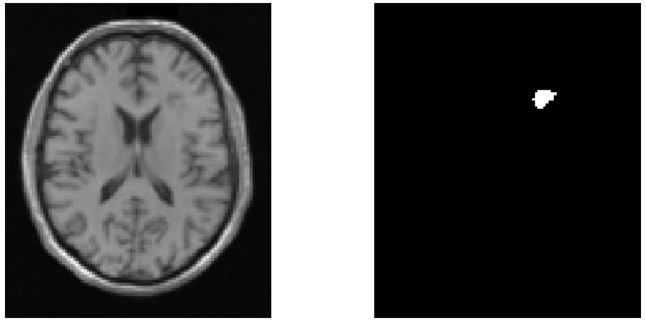


025: 7.209 cm^3

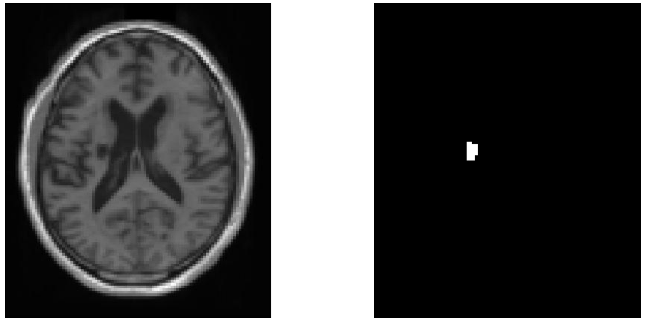
026: 2.441 cm^3



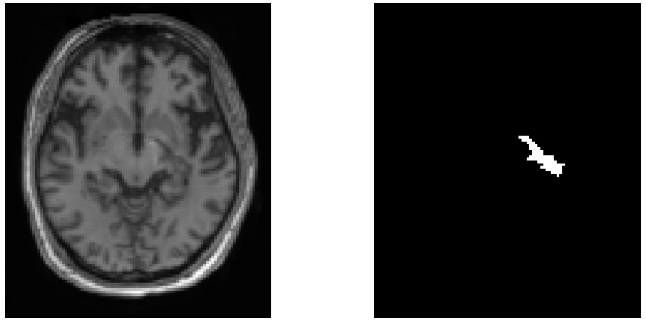
027: 2.810 cm^3



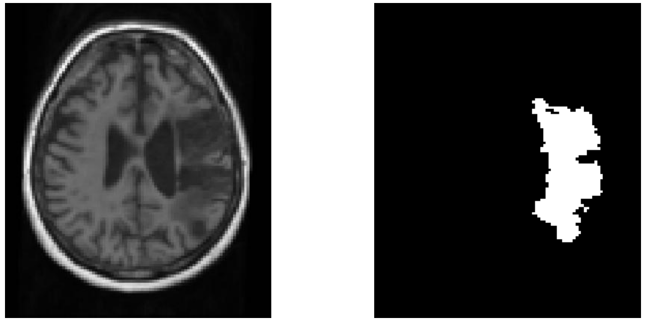
028: 1.015 cm^3



029: 5.848 cm^3

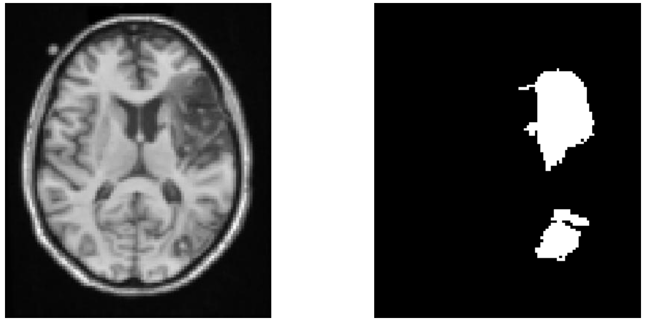


030: 141.278 cm^3



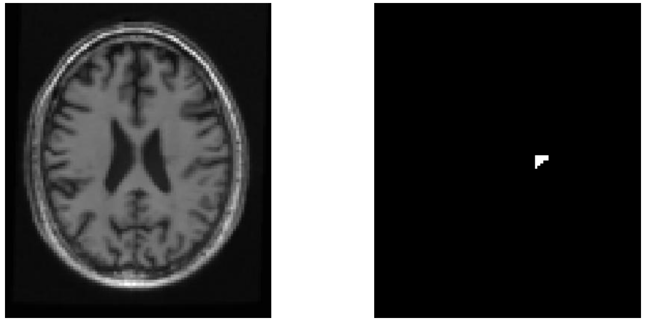
031: 1.701 cm^3

032: 123.248 cm^3

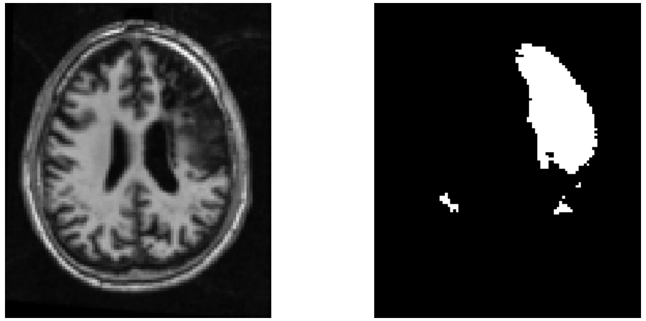


033: 4.102 cm^3

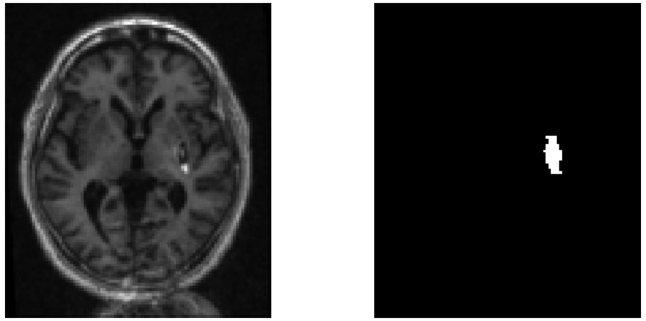
034: 1.573 cm^3



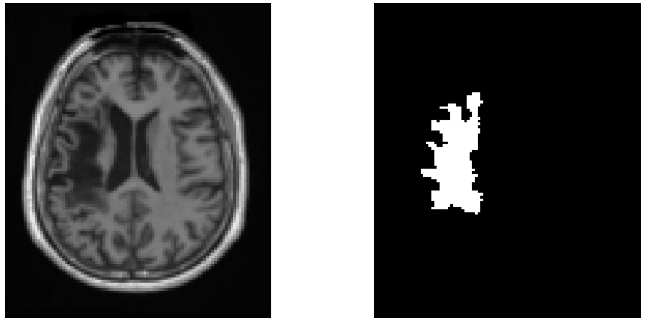
035: 222.005 cm^3



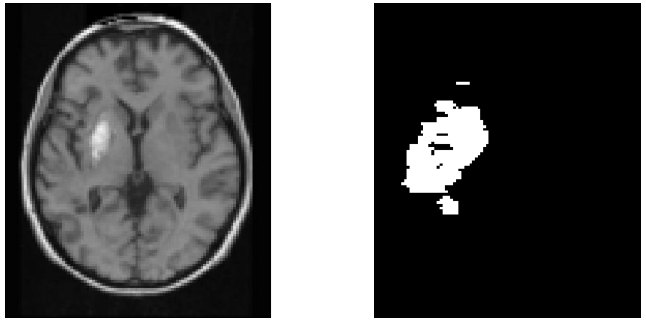
036: 8.119 cm^3



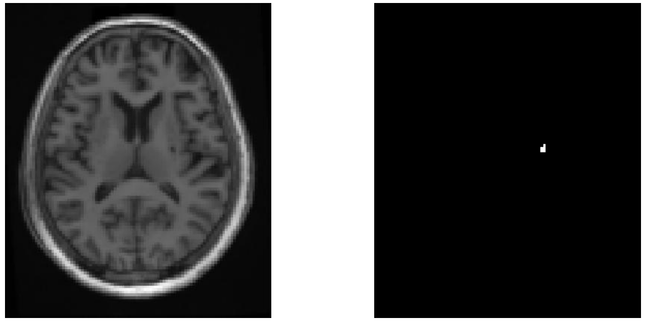
037: 66.057 cm^3



038: 130.660 cm^3

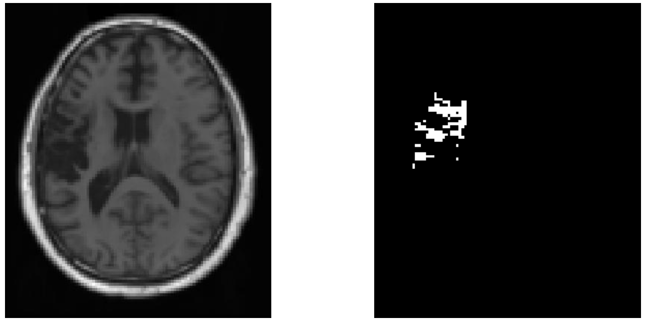


039: 0.089 cm^3

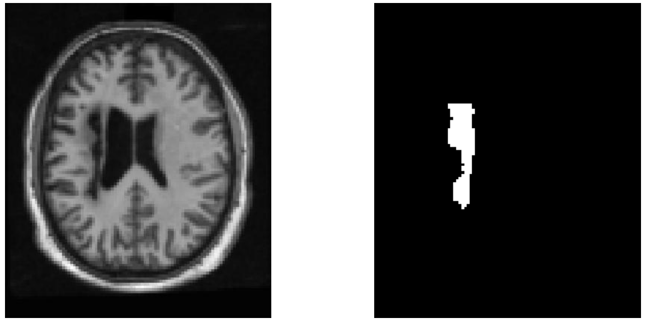


040: 1.806 cm^3

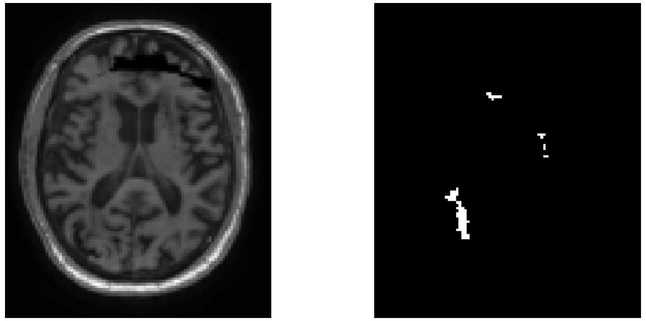
041: 19.158 cm^3



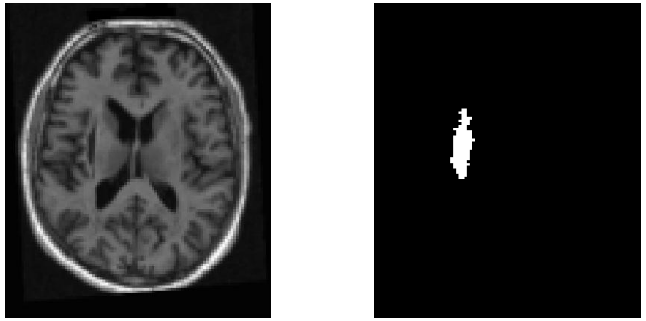
042: 40.818 cm^3



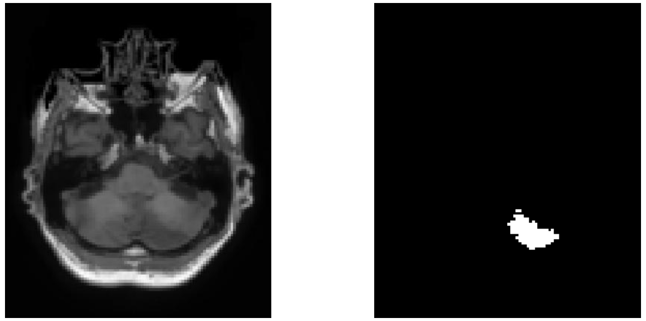
043: 6.161 cm^3



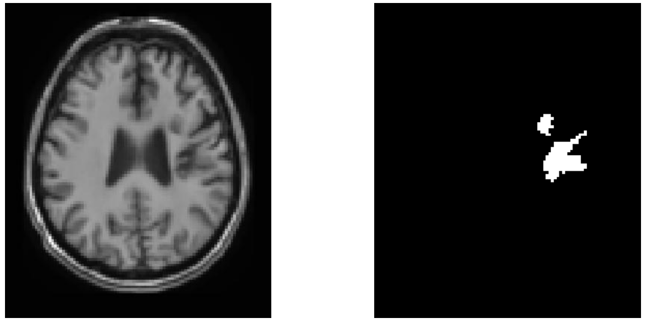
044: 14.618 cm^3



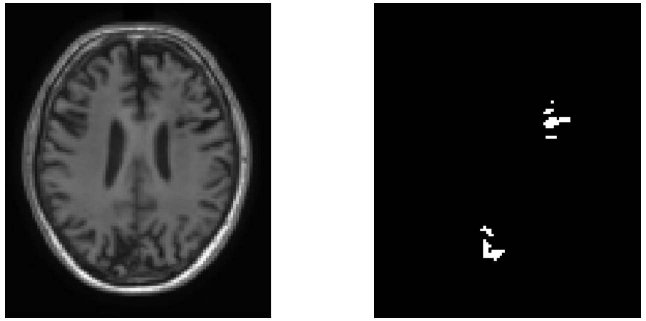
045: 13.653 cm^3



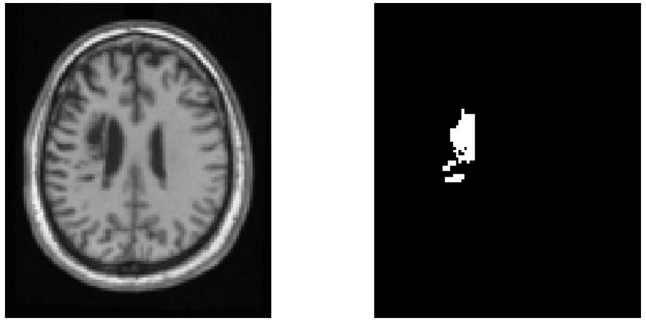
046: 16.615 cm^3



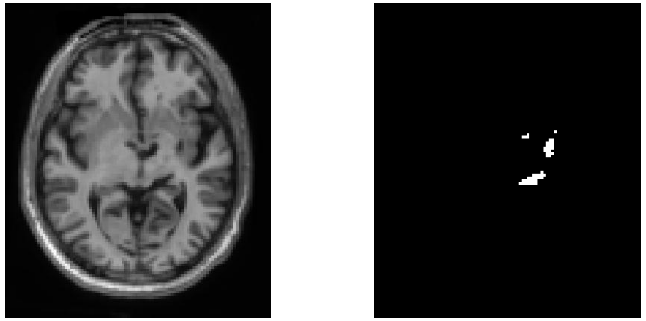
047: 5.247 cm^3



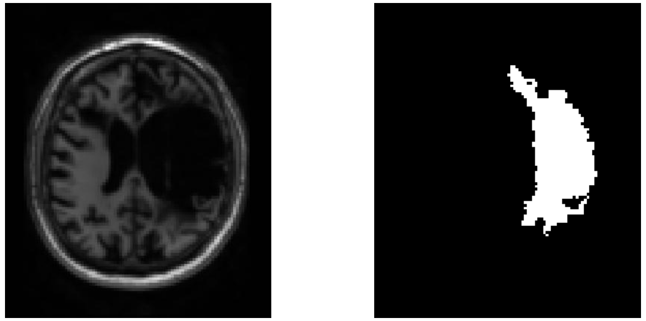
048: 19.138 cm^3



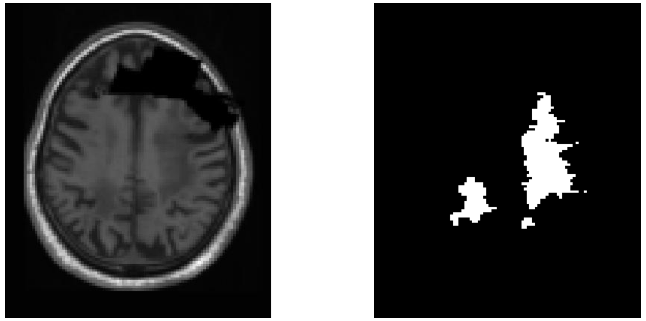
049: 6.826 cm^3



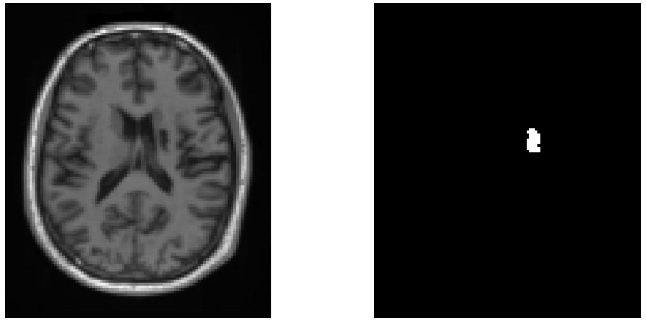
050: 281.558 cm^3



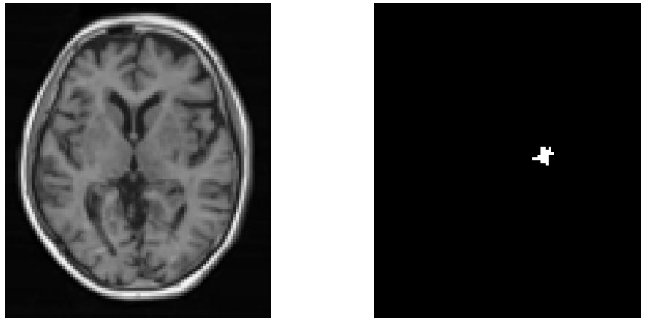
051: 69.392 cm^3



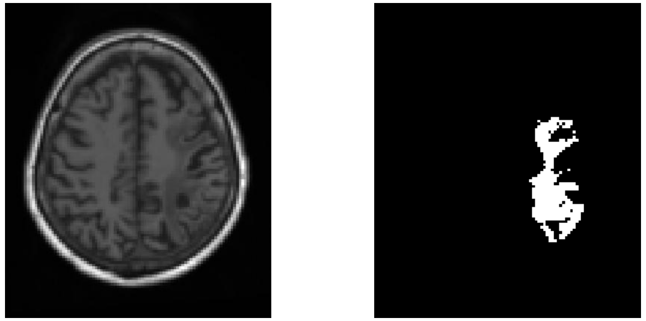
052: 1.201 cm^3



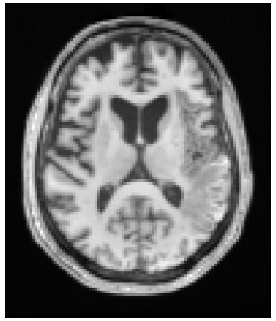
053: 2.790 cm^3

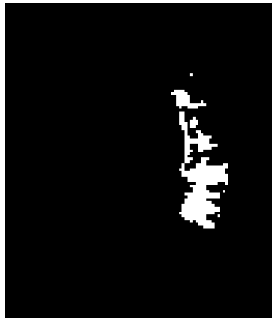


054: 52.789 cm^3



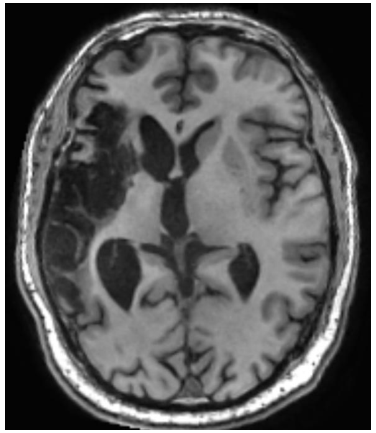
055: 47.632 cm^3





- Raw T1-weighted image and lesion mask
  - Raw T1-weighted image in native space
  - Lesion mask in native space

Raw T1-weighted image



Lesion mask

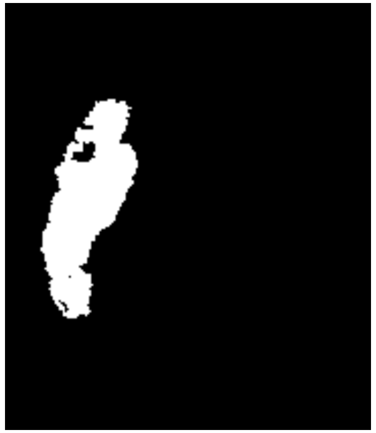


Image specifications:

Dimensions: 98 × 116 × 94

Voxel size: 2.0 mm × 2.0 mm × 2.0 mm

Raw T1-weighted Image and Lesion Mask

- Segmentation label map
  - Lesion mask
- Lesion segmentation performance
  - Mean DSC for the test set (n = 55)
    - Average of the overlap between predicted and manually annotated lesion masks across the test set
    - Ranges from 0 to 1

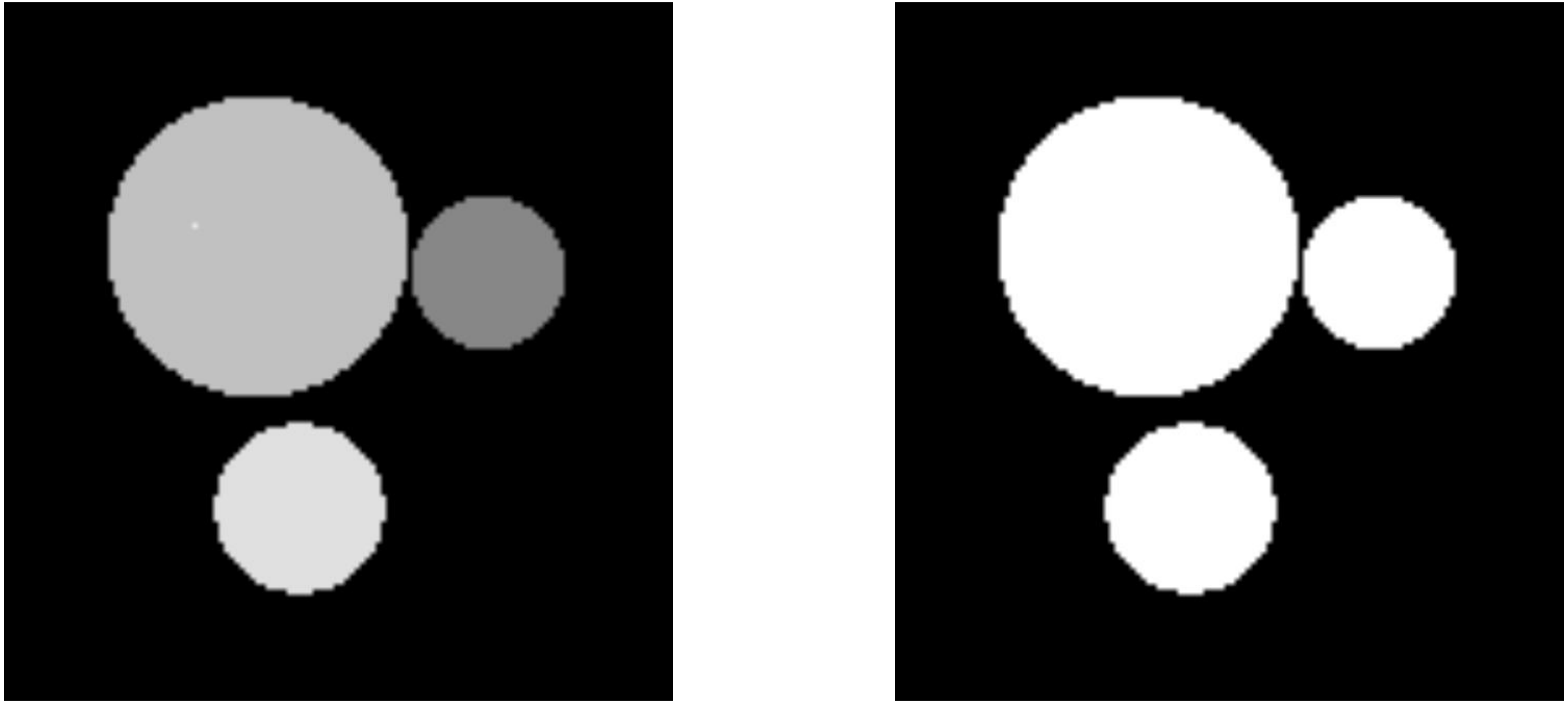
Article	Method	Reported Dice	Code Publicly Available	*n*	Validation Method	Input size 2D/3D (H, W, D)
					Cross-validation	
Basak *et al.*, 2021	DFENet	0.546	no	229	5-fold cross-validation	2D 192, 192 or 3D 192, 192, 4
Hui *et al.*, 2020	PSPF and U-Net	0.593	no	239	6-fold cross-validation	2D 176, 176
Lu *et al.*, 2020	EDCL w/ 3D Unet	0.148 (0.584)**	no	239	5-fold cross-validation	3D 64, 64, 64
Qi *et al.*, 2019	X-Net	0.487	yes	229	5-fold cross-validation	2D 192, 224
Zhang *et al.*, 2020	MI-UNet	0.567	no	229	5-fold cross-validation	2D 233, 197 or 3D 49, 49, 49
					One hold-out Train, Validation, Test	
Chen *et al.*, 2018	U-Net/GMM*	0.500/0.170	no	220	unclear/0, 0, 100 (%)	2D 128, 128 or 256, 256
Chen *et al.*, 2020	VAE*/GMVAE*	0.110/0.120	no	220	0, 0, 100/0, 0, 100 (%)	2D 200, 200
Kervadec *et al.*, 2020	Enet	0.474	yes	229	203, 26, 0	unclear
Liu *et al.*, 2019	MSDF-Net	0.558	no	229	160, 69, 0	2D 224, 177
Paing *et al.*, 2021	3D U-Net	0.668	no	239	60, 20, 20 (%)	3D 197, 233, 189
Qi *et al.*, 2020	U-Net	0.518	no	229	120, 40, 69	2D 224, 192
Sahayam *et al.*, 2020	MUDCap3	0.670	no	229	160, 69, 0	3D 256, 256, 256
Tomita *et al.*, 2020	3D-ResU-Net	0.640	yes	239	76, 11, 13 (%)	3D 144, 172, 168
Wang *et al.*, 2020	CPGAN	0.617	no	239	129, 40, 60	2D 256, 256
Xue *et al.*, 2020	U-Net (9 paths)	0.540	yes	54	0, 0, 54	3D 192, 224, 192
Yang *et al.*, 2019	CLCI-Net	0.581	yes	220	55, 18, 27 (%)	2D 224–233, 176–197
Zhou *et al.*, 2019	D-Unet	0.535	no	229	80, 20, 0 (%)	2D 192, 192 or 3D 192, 192, 4

[Liew et al., 2022]

Performance of Lesion Segmentation Using ATLAS R1.2

Demo Dataset

- Simulated images and labels
  - Training dataset: $n = 40$
    - Images: Image/0-39.nii.gz
    - Masks: Label/0-39.nii.gz



Image

Label

Example Pair of an Image and a Label Mask