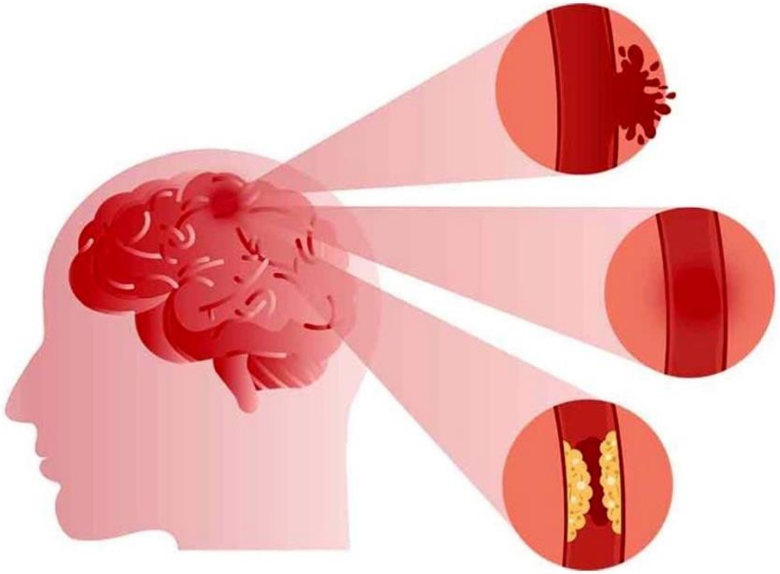
**Medical/Bio Research Topics Ⅱ: Week 06 (12.10.2023)**

# Lesion segmentation artificial intelligence models (1): data and prediction problem

## (병변 분할 인공지능 모델 개발 연습 (1): 데이터 및 예측 문제)

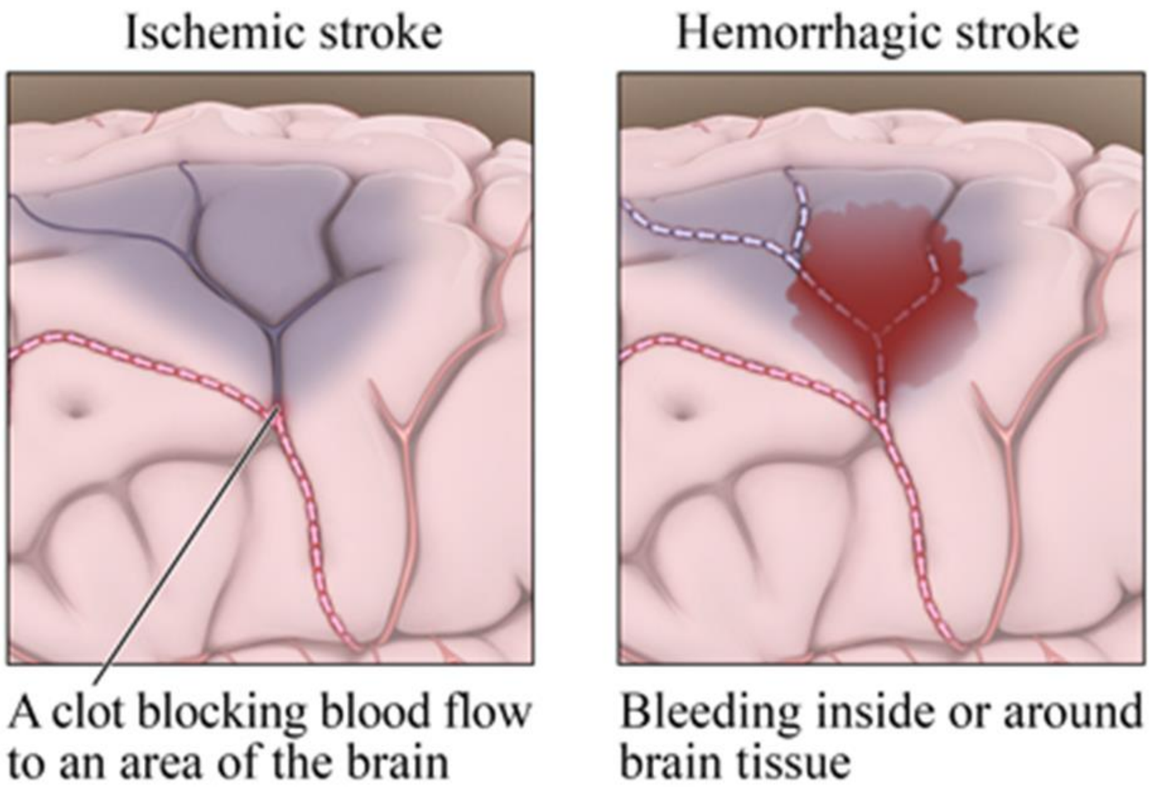
# Stroke

- 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 stroke
    - Most common type of stroke
    - Caused by interrupted or reduced blood flow to the brain
    - The brain cannot get oxygen and nutrients from the blood, so that brain cells begin to die within minutes
  - Hemorrhagic stroke
    - Caused by bleeding in the brain
    - The leaked blood results in pressure on brain cells, damaging them

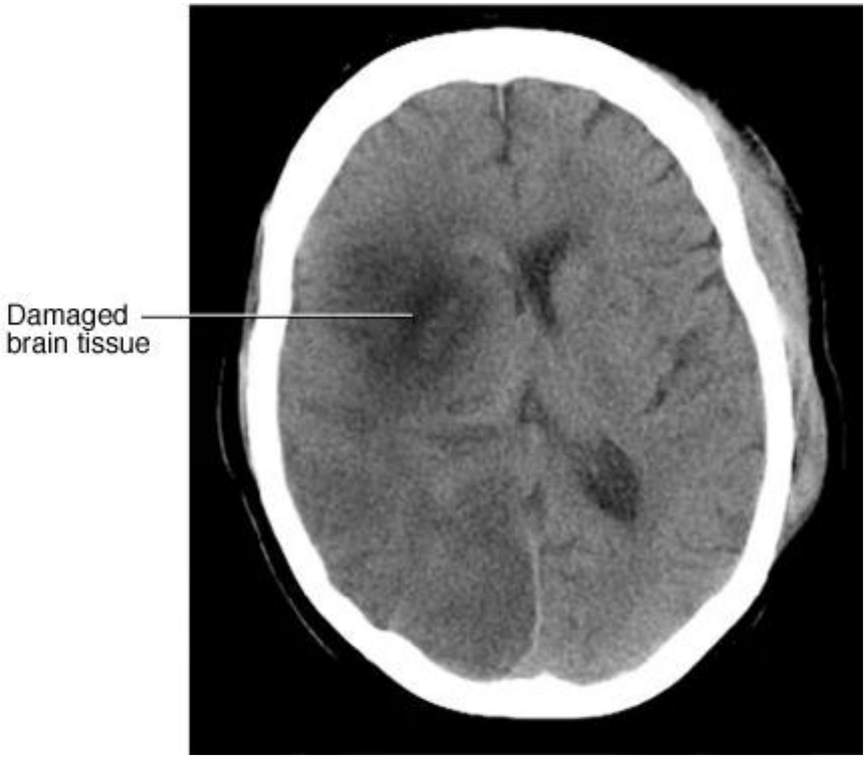



[https://myhealth.alberta.ca/Health/Pages/conditions.aspx?hwid=tp12720]

**Ischemic vs. hemorragic 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
    - Computerized tomography (CT)
    - MRI




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

**CT scan of brain tissue damaged by stroke**

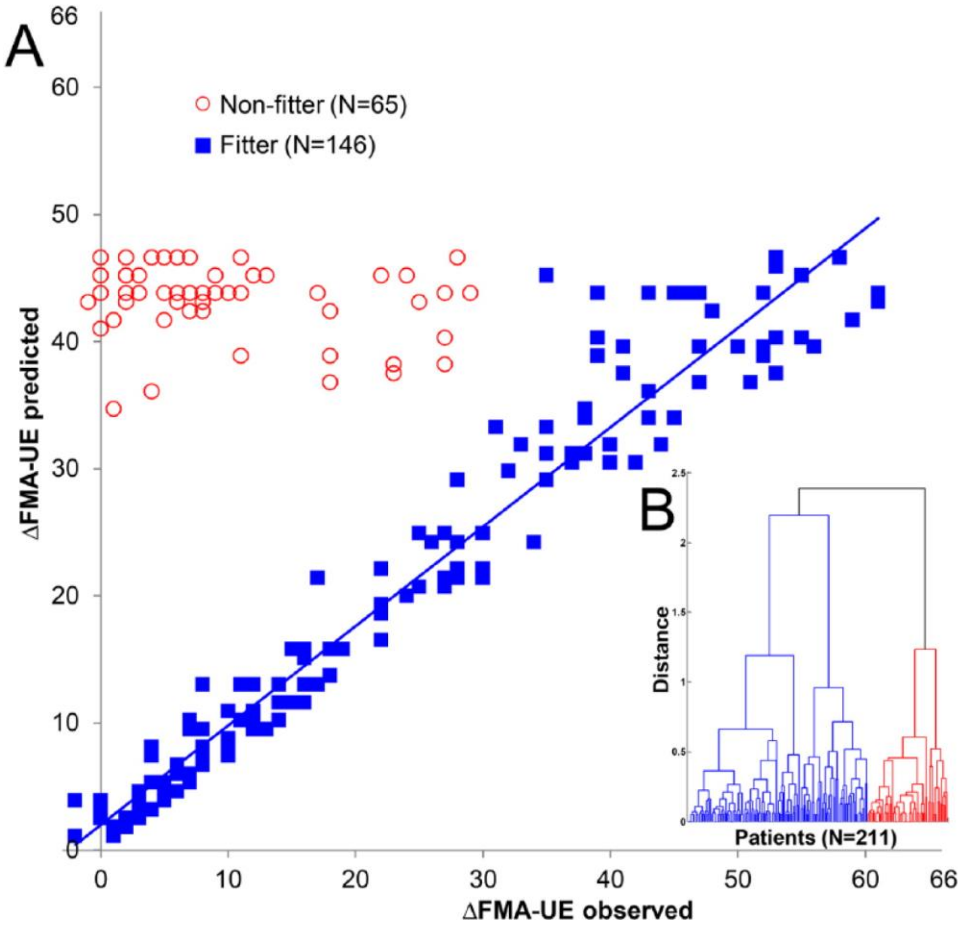
- Emergency treatment
  - Depends on the type of stroke
  - Ischemic stroke
    - Intravenous injection of recombinant tissue plasminogen activator (TPA) to dissolve the blood clot
      - Usually given through a vein in the arm within the first three hours
    - Endovascular therapy to directly remove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

- 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

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

$$\Delta \mathrm{FMA}-\mathrm{UE}_{\text{predicted}} = 0.7 \cdot (66 - \mathrm{FMA} - \mathrm{UE}_{\text{initial}}) + 0.4$$
$$\approx 0.7 \cdot (\text{maximal potential recovery})$$

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

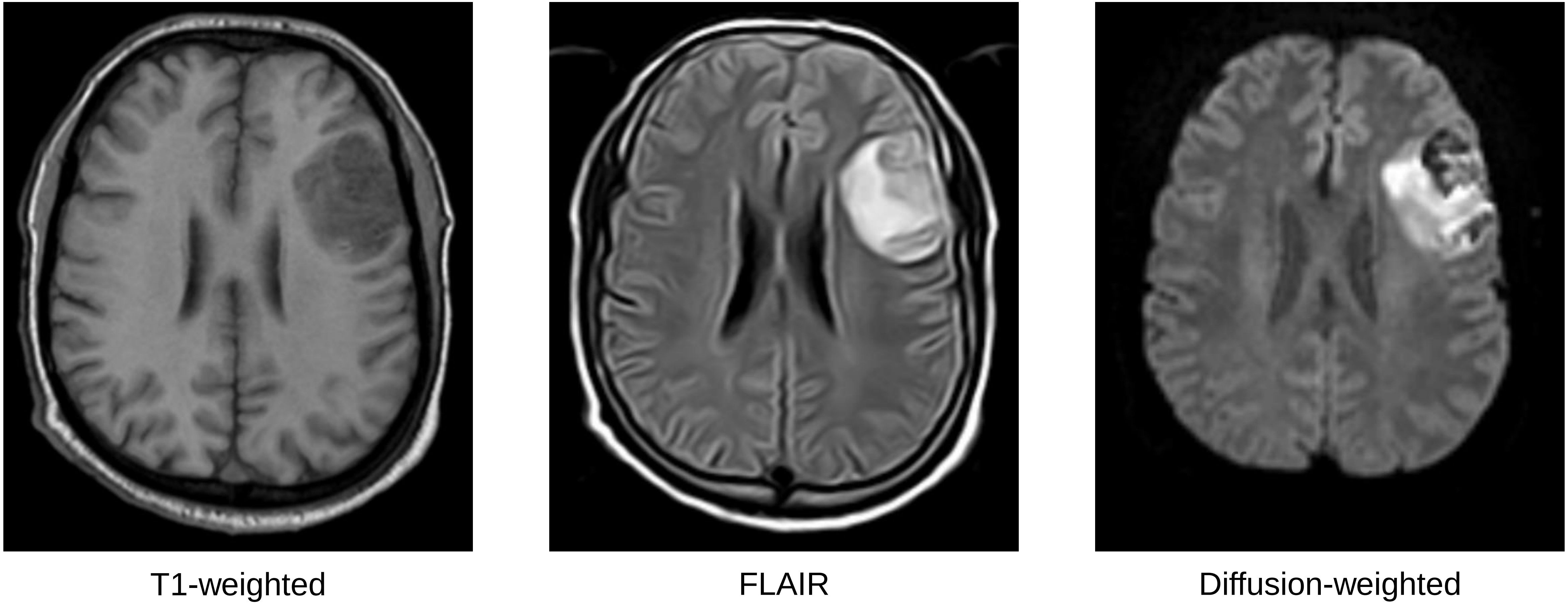



[Winters et al., 2015]

**Proportional motor improvement in the upper limb**

# Stroke Lesion: Cerebral Infarct

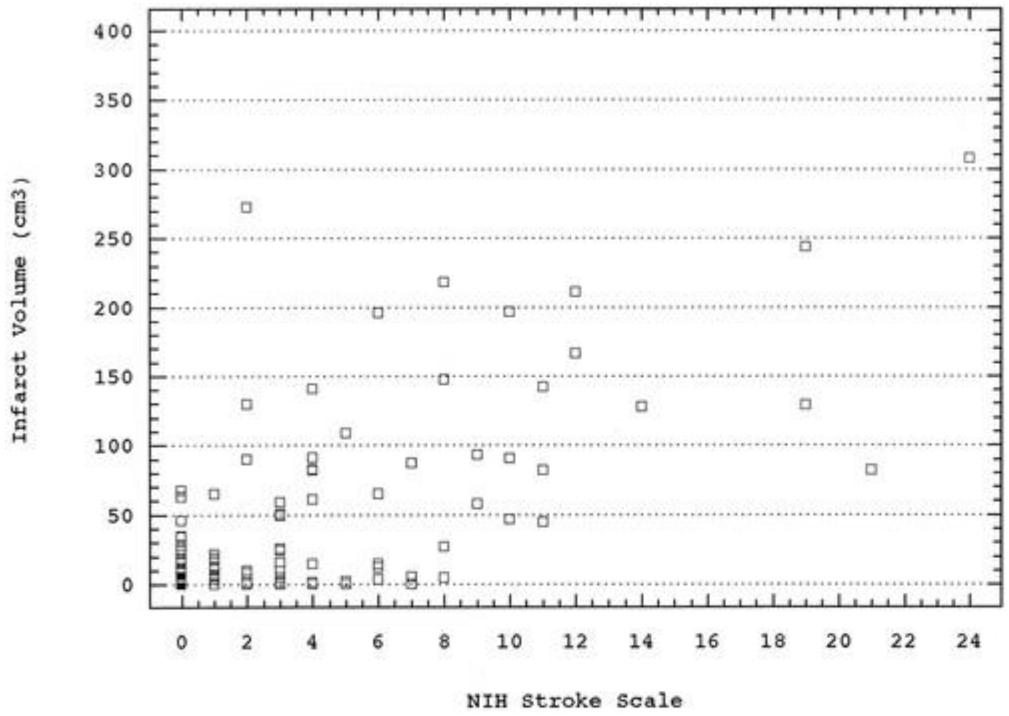
- Cerebral infarction
  - Pathological process that results in an area of ischemic necrosis (cellular death) in the brain
  - Manifests clinically as ischemic stroke
- Cerebral infarct
  - Lesion resulted from cerebral infarction
  - The ischemic part of the lesion remains after stroke, with permanent damage to the affected area of the brain




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

**Cerebral infarct displayed as altered signals in MRI**

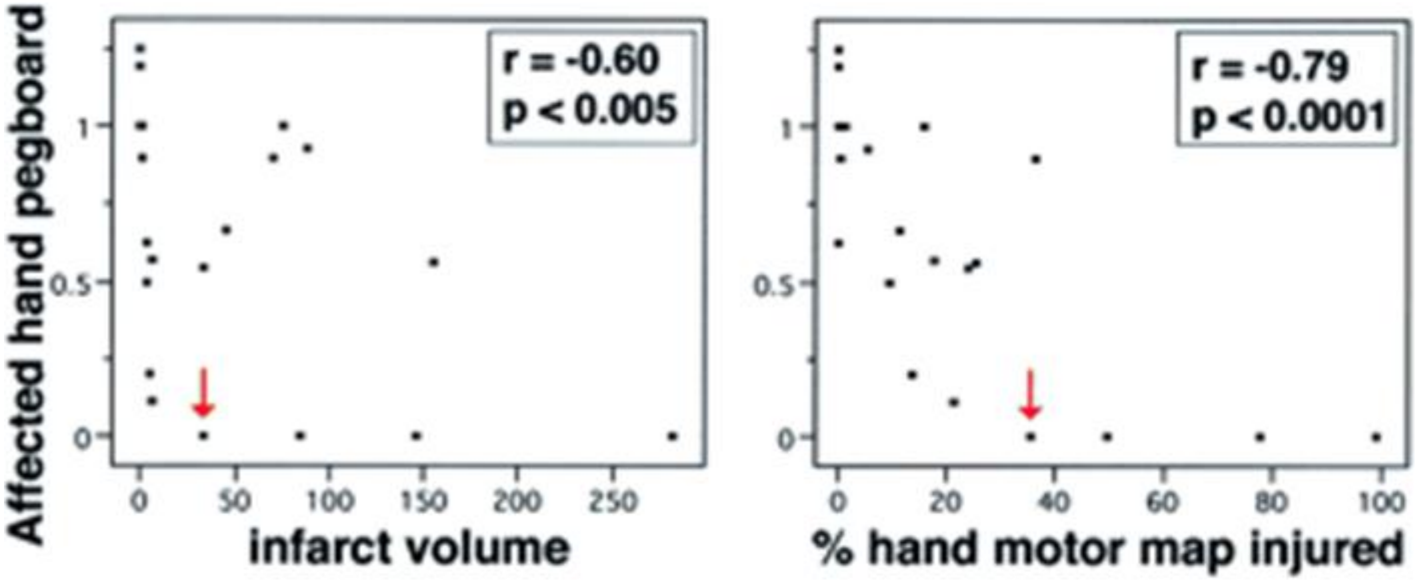
- Brain-behaviour relationship in stroke rehabilitation
  - Lesion size
    - Infarct volume correlates with clinical outcome [Saver et al., 1999]




[Saver et al., 1999]

– Lesion location

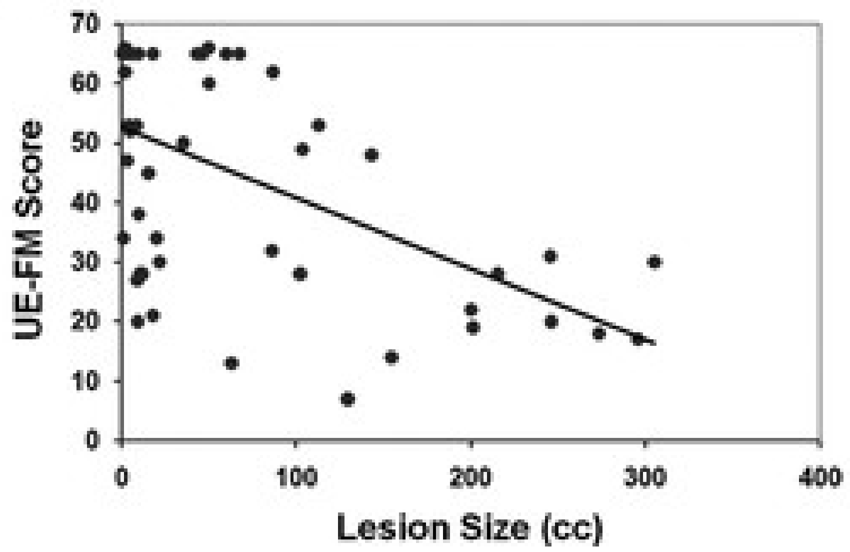
- Motor performance correlates with the fraction of hand motor map injured more strongly than with infarct volume [Crafton et al., 2003]

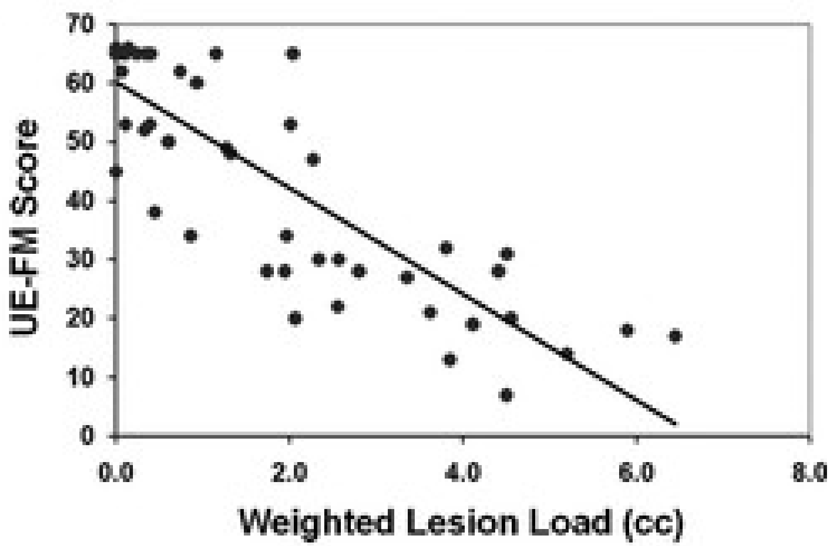



[Crafton et al., 2003]

– Lesion load: lesion overlap with extant brain structures

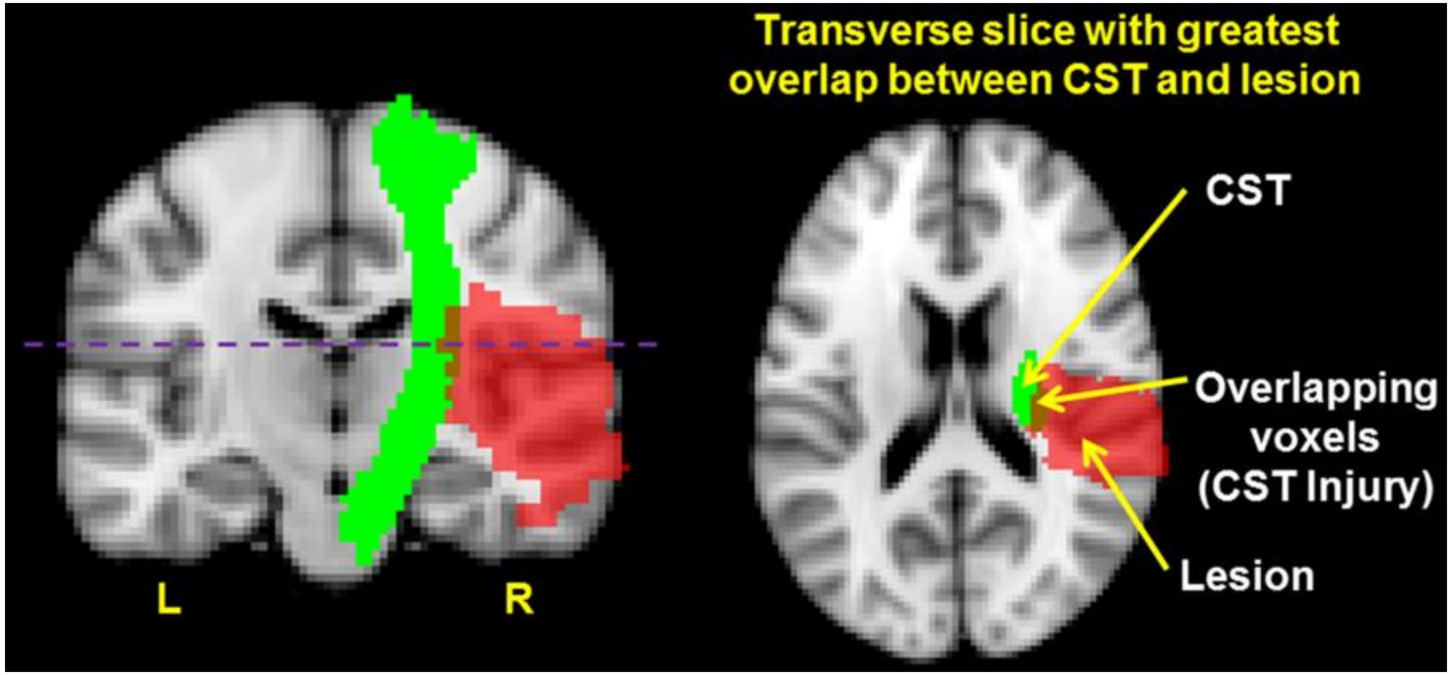
- Motor impairment correlates with the proportion of the corticospinal tract injured more strongly than with infarct volume [Zhu et al., 2010]




[Zhu et al., 2010]

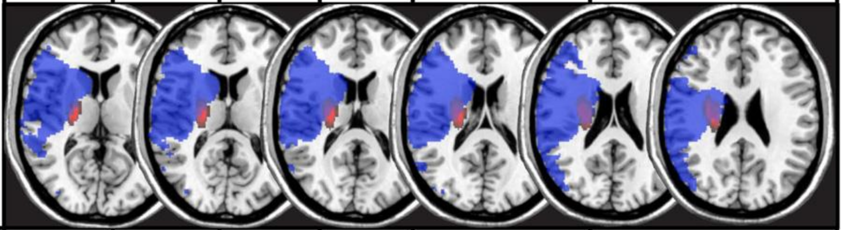
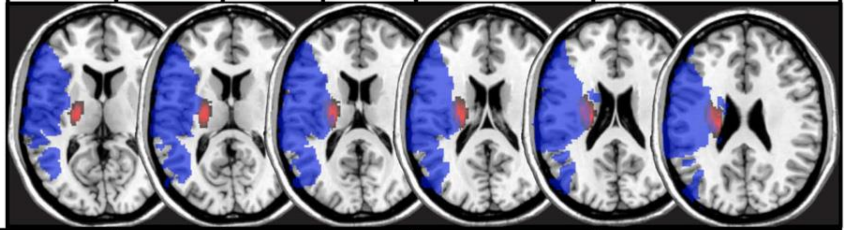
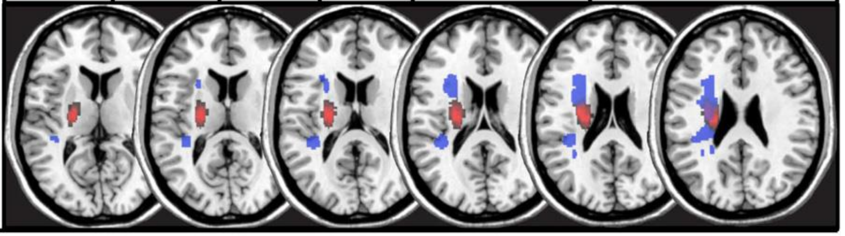
$$\text{CST Injury} = \left(\frac{\text{Number of overlapping voxels between the CST and lesion for the transverse slice}}{\text{Total number of CST voxels for the transverse slice}}\right) \times 100\%$$




[Lam et al., 2020]

**Computation of a corticospinal tract lesion load**

- Corticospinal tract lesion load can predict motor outcome [Feng et al., 2015]

| 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
- Manual segmentation remains the gold standard, but it is time-consuming, subjective, and requires neuroanatomical expertise

- Anatomical Tracings of Lesions After Stroke (ATLAS) v2.0 dataset [https://fcon_1000.projects.nitrc.org/indi/retro/atlas.html]
  - Released in 2021 by expanding upon and replacing ATLAS v1.2 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

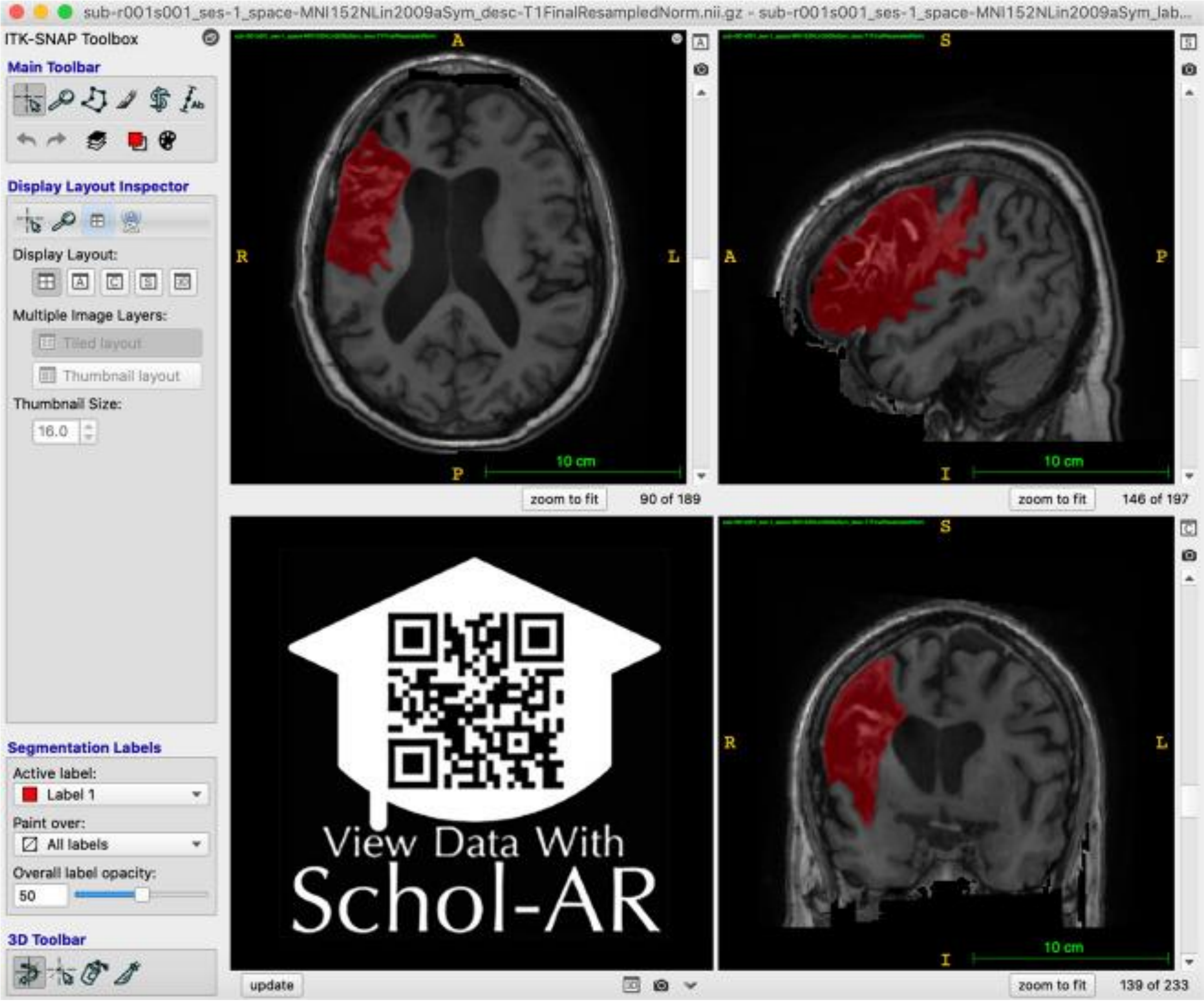
- Includes T1-weighted MRI scans and manually segmented lesion masks (*n* = 1,271)
  - Training and test datasets derived from 33 research cohorts
    - Samples from each research cohort are randomly assigned to either training or test datasets so that they have similar compositions
    - Training dataset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

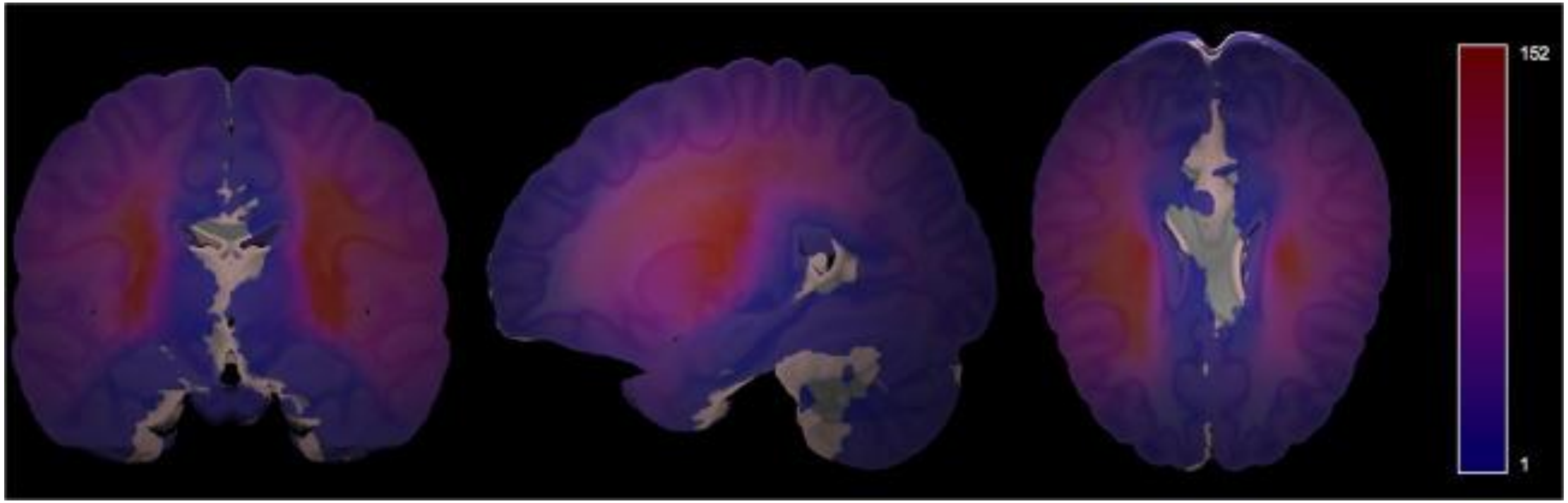
Located In either cerebellum or brainstem

- 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]

**Visualization of the lesion overlap across all subjects (*n* = 955) overlaid on the MNI template**

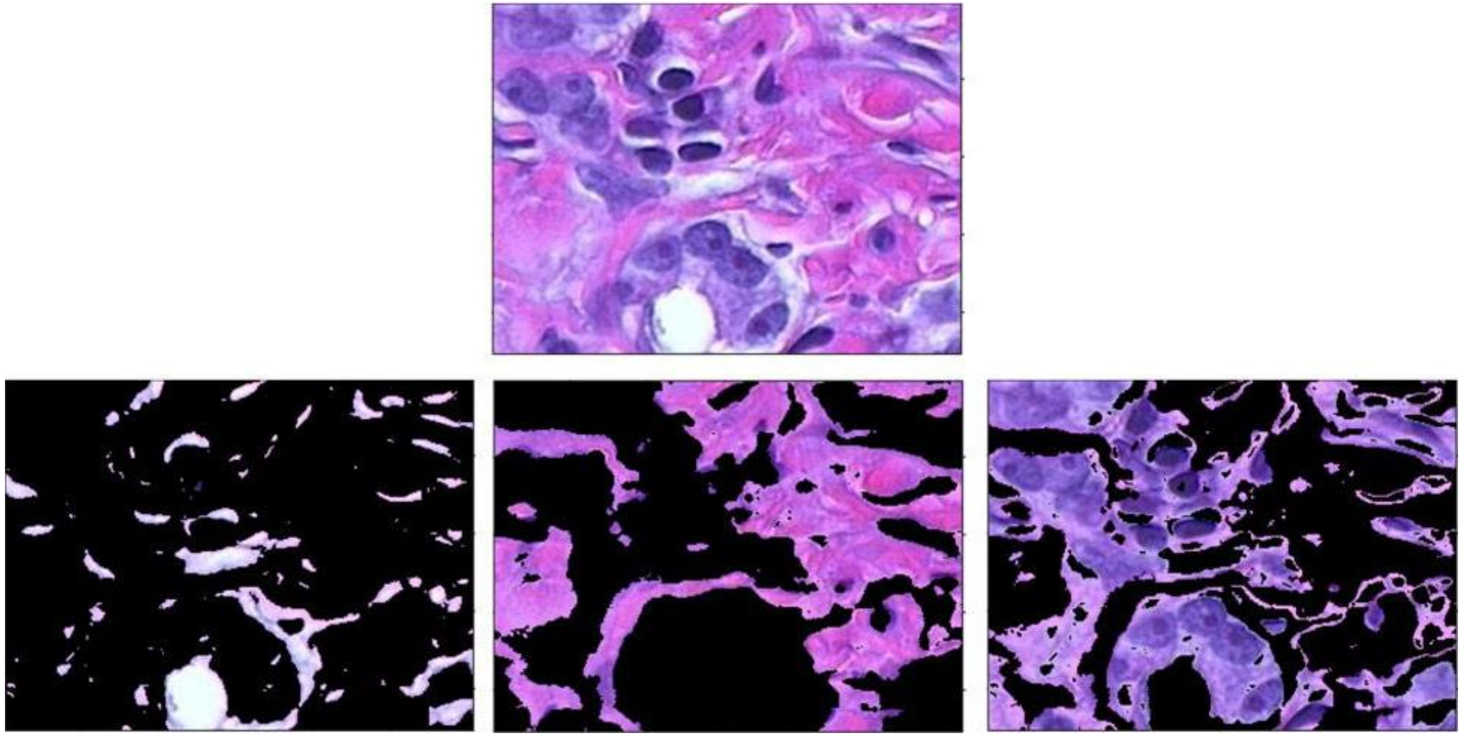
# Image Segmentation

- Technique in digital image processing and analysis to partition an image into multiple parts or areas, often based on the characteristics of the pixels in the image
  - Involves converting an image into a collection of regions of pixels that are represented by a mask or a labeled image
- A common application in medical imaging is to detect and label pixels in an image or voxels of a 3D volume 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 grayscale 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

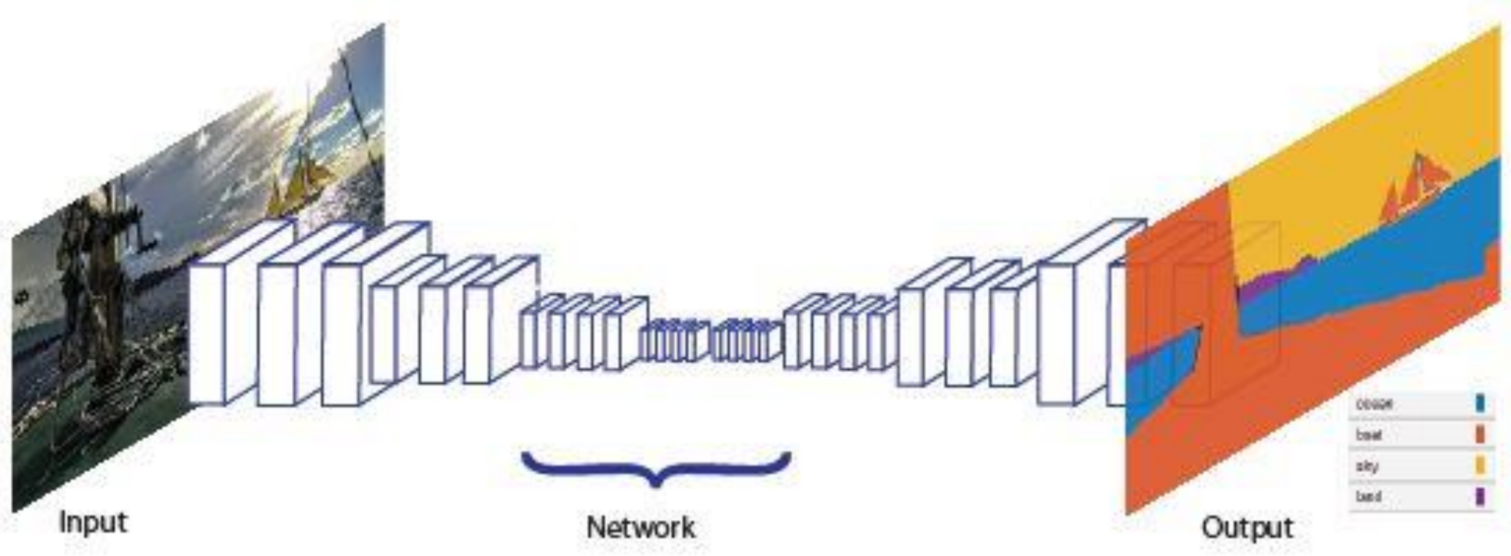


– Graph-based segmentation
  - Enables to segment an image into foreground and background areas

– Region growing
  - Examines neighbouring pixels of initial seed points and determines iteratively whether the pixel neighbours should be added to the area

# Deep Learning for Image Segmentation

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




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

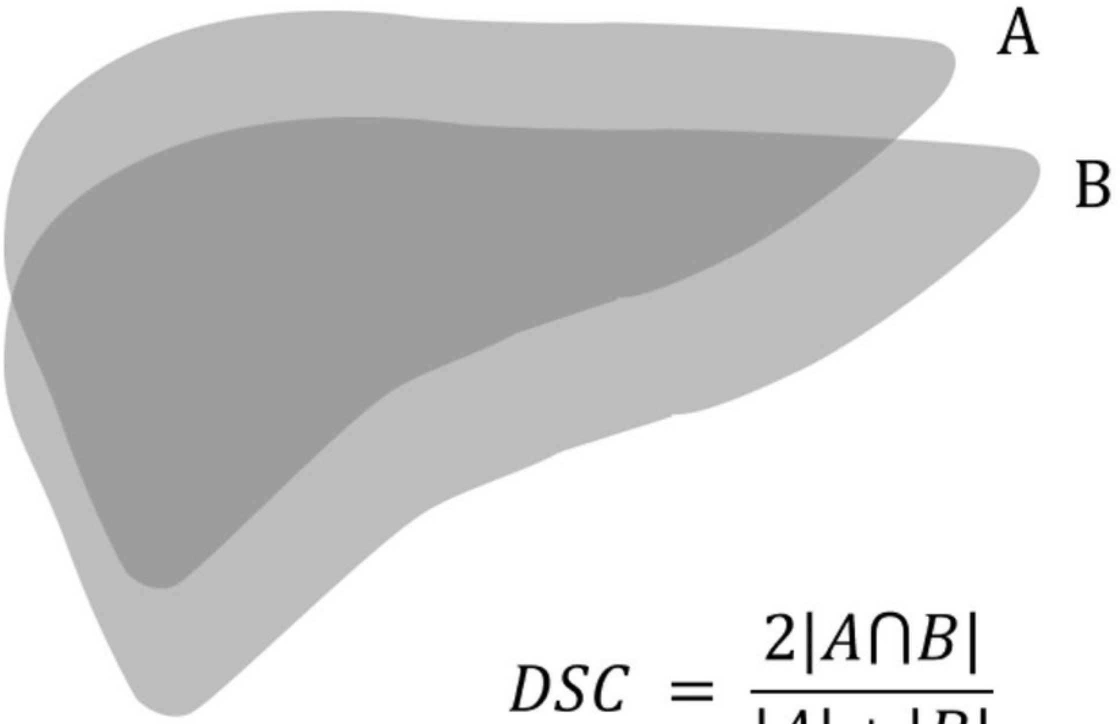
- Deep learning for lesion segmentation
  - Semantic segmentation that classifies each voxel as lesioned or non-lesioned
  - Patch-wise segmentation
    - Takes a small patch around a voxel as the input and traverses the entire volume by repeatedly taking patches
    - Redundant calculations caused by 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 (lesion areas are much smaller than non-lesion areas in the brain)

- Lesion segmentation performance
  - Dice similarity coefficient (DSC, Sørensen–Dice index or Dice coefficient) [Dice, 1945]
    - $F_1$ score that is a harmonic mean of precision and recall
      - Precision (true positive value (TPV)) = TP/(TP + FP)
      - Recall (true positive rate (TPR)) = TP/P = TP/(TP + FN)
      - $F_1$ score = 2/((1/precision) + (1/recall)) = 2TP/(2TP + FP + FN)
    - Has become the most broadly used tool in the validation of image segmentation algorithms
  - Positive predictive value (PPV) [Altman and Bland, 1994]
    - Proportion of positive predictions that are true positives
      - PPV = TP/PP

| | | 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) $= \frac{TPR + TNR}{2}$ | $F_1$ score $= \frac{2 PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$ | 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**

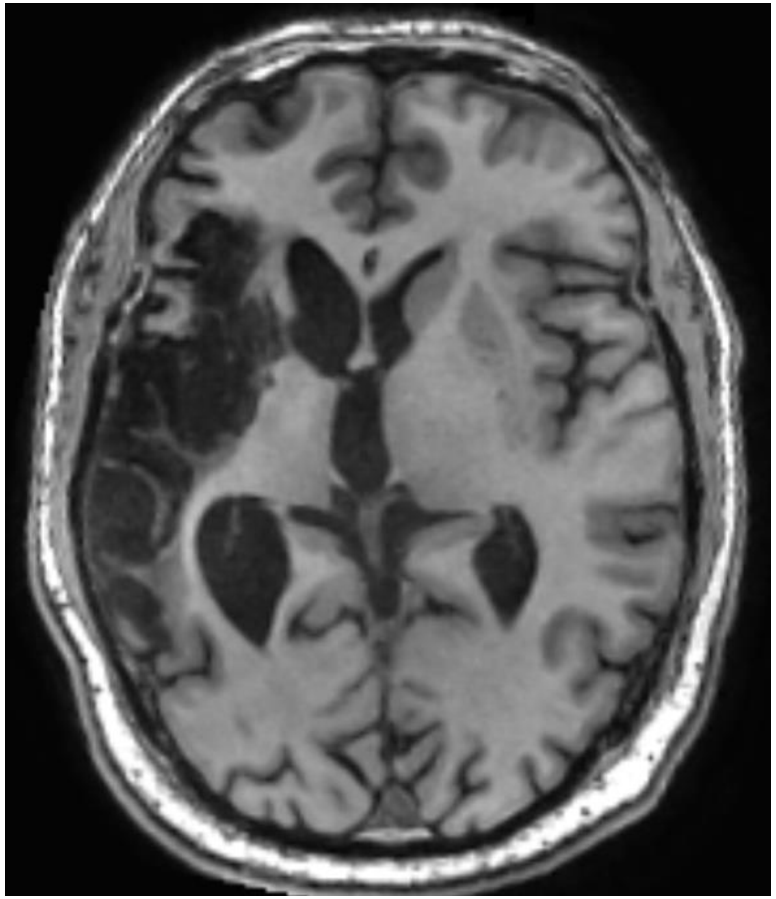
| | | Predicted condition | | | |
|---|---|---|---|---|---|
| | Total population = P + N | **Positive (PP)** | **Negative (PN)** | Informedness, bookmaker informedness (BM) = TPR + TNR − 1 | Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$ |
| **Actual condition** | **Positive (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) $= \frac{TPR + TNR}{2}$ | $F_1$ 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]

# PPV in a confusion matrix

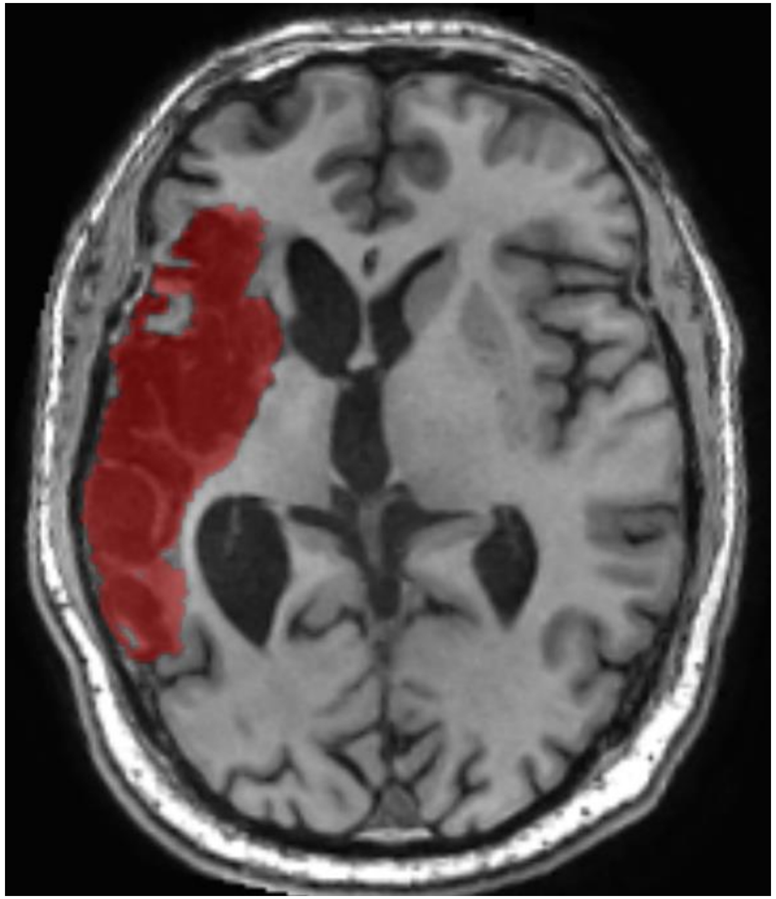
# Dataset

- ATLAS v2.0 training dataset (*n* = 655)
  - Training dataset: *n* = 600
    - T1-weighted MRI scans: train/brain001-600.nii.gz
    - Lesion masks: train/lesionmask001-600.nii.gz
  - Test dataset: *n* = 55
    - T1-weighted MRI scans: test/brain001-055.nii.gz
    - Lesion masks: hidden



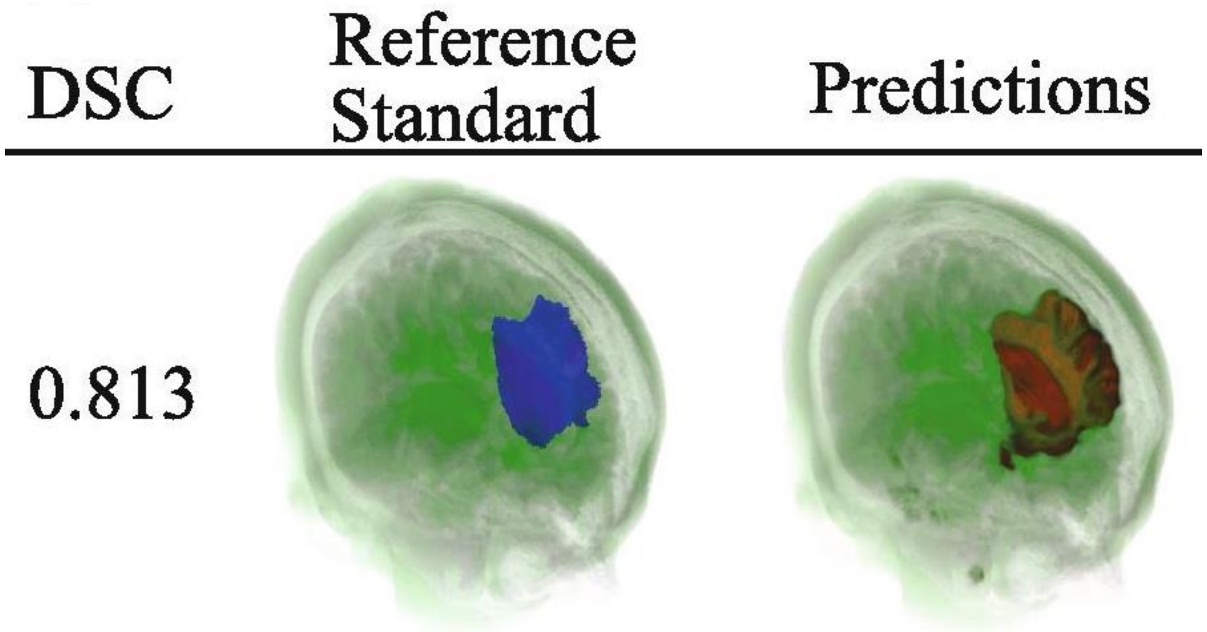
train/brain002.nii.gz

train/lesionmask002.nii.gz



**Example pair of a T1-weighted MRI scan and a lesion mask**

- Lesion segmentation performance
  - Average DSC for the test dataset
    - Ranges from 0 to 1, with 1 indicating a perfect match between the two sets of data and 0 indicating no overlap




[Tomita et al., 2020]

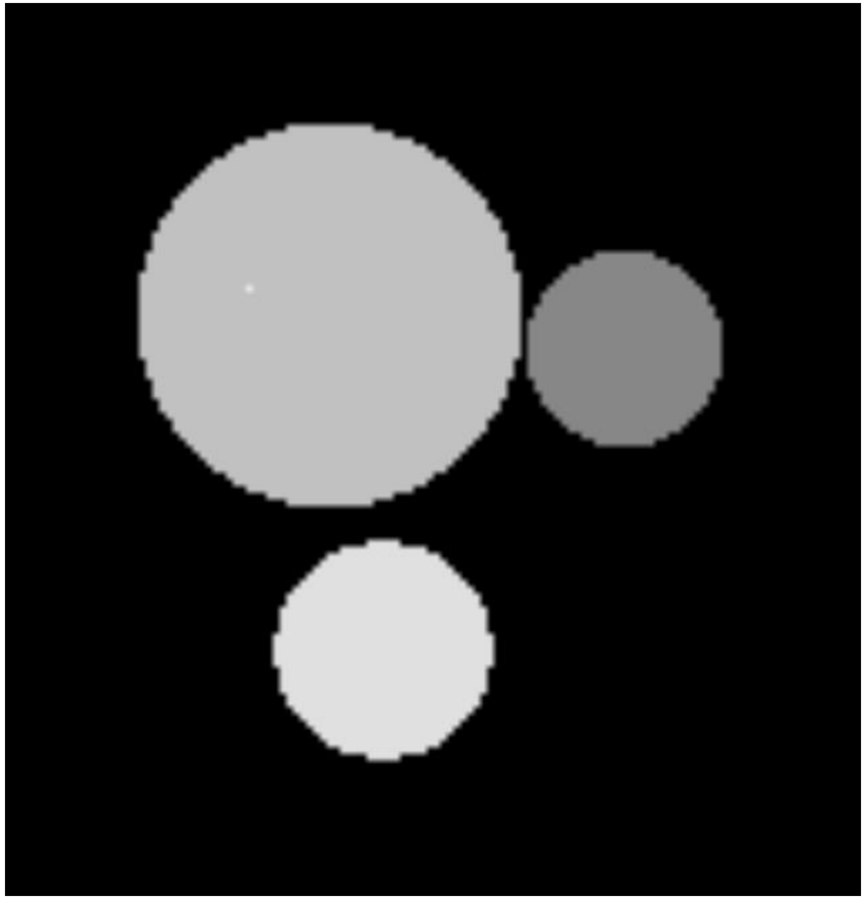
| Article | Method | Reported Dice | Code Publicly Available | *n* | Validation Method | Input size 2D/3D (H, W, D) |
|---|---|---|---|---|---|---|
| | | | | | **Cross-validation** | |
| Basak *et al.*, 2021 | DFENet | 0.546 | no | 229 | 5-fold cross-validation | 2D 192, 192 or 3D 192, 192, 4 |
| Hui *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]

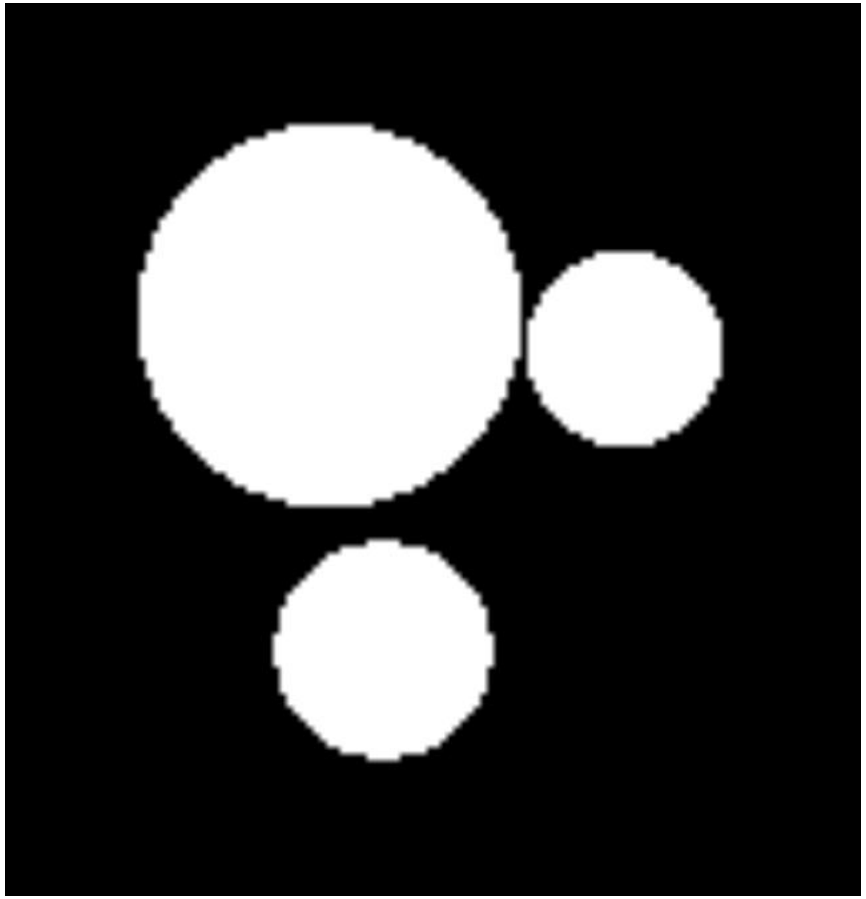
# Performance of lesion segmentation using ATLAS v1.2

# Demo Dataset

- Generated images and masks
  - Training dataset: $n$ = 40
    - Images: demo/im0-39.nii.gz
    - Masks: demo/seg0-39.nii.gz



demo/im2.nii.gz



demo/seg2.nii.gz

**Example pair of an image and a mask**