
MICCAI
MICCAI2022
Singapore
25th International Conference on
Medical Image Computing and
Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore
www.miccai2022.org

MICCAI2022 Singapore
25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore
MICCAI

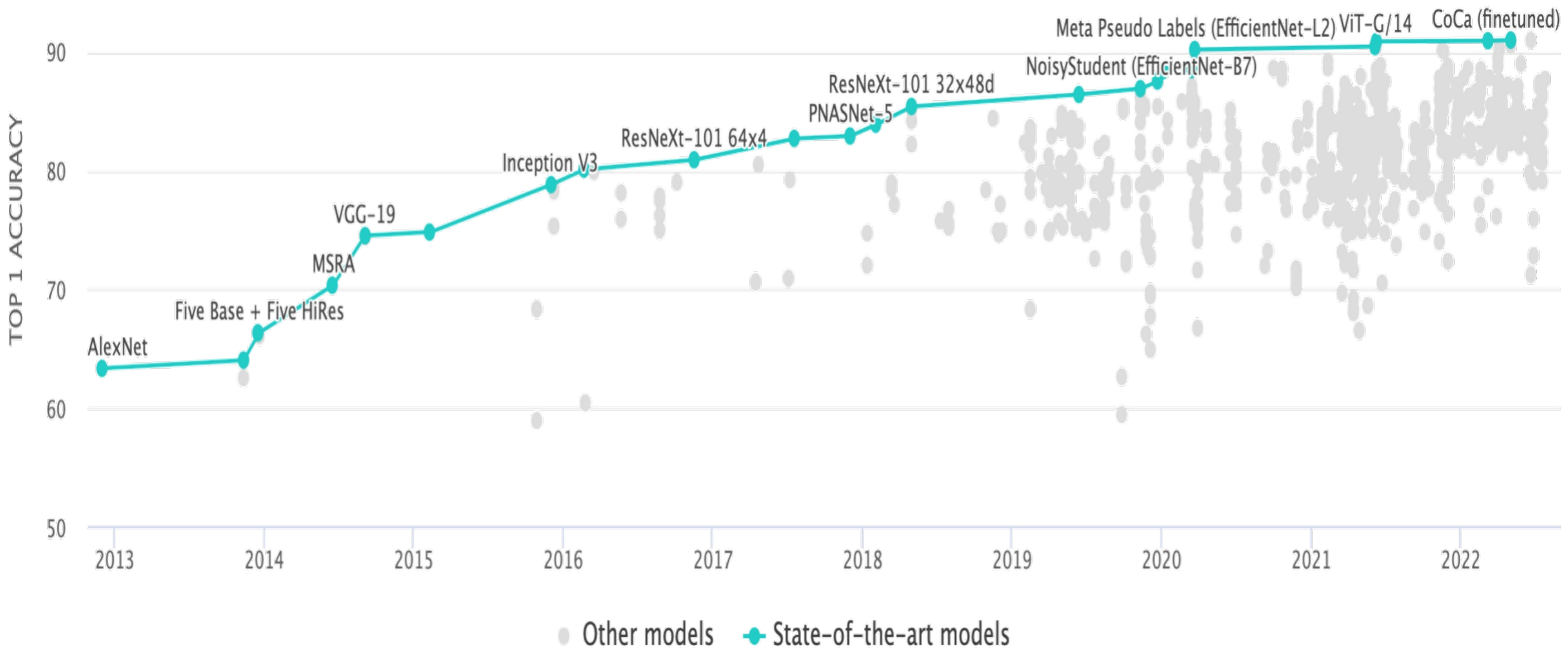
We create generalizable domain-aware loss and provide pretrained 3D models!

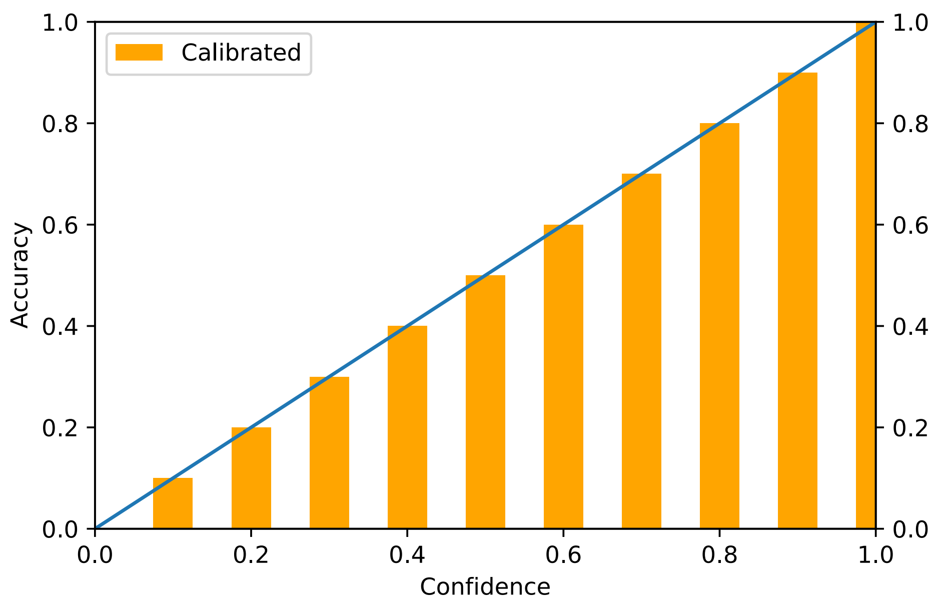
DOMINO: Domain-aware Model Calibration in Medical Image Segmentation

Skylar E. Stolte[1], Kyle Volle[2], Aprinda Indahlastari[3,4], Alejandro Albizu[3,5], Adam J. Woods[3,4,5], Kevin Brink[6], Matthew Hale[2], and <u>Ruogu Fang</u>[1,3,7]

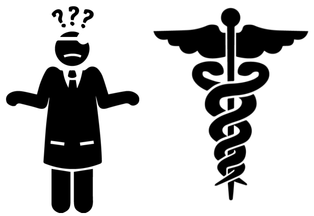
[1]J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida (UF), USA
[2]Department of Mechanical and Aerospace Engineering, Herbert Wertheim College of Engineering, UF, USA
[3]Center for Cognitive Aging and Memory, McKnight Brain Institute, UF, USA
[4]Department of Clinical and Health Psychology, College of Public Health and Health Professions, UF, USA
[5]Department of Neuroscience, College of Medicine, UF, USA
[6]United States Air Force Research Laboratory, Eglin Air Force Base, Florida, USA
[7]Department of Electrical and Computer Engineering, Herbert Wertheim College of Engineering, UF, USA

Goal: High Accuracy + Good Calibration





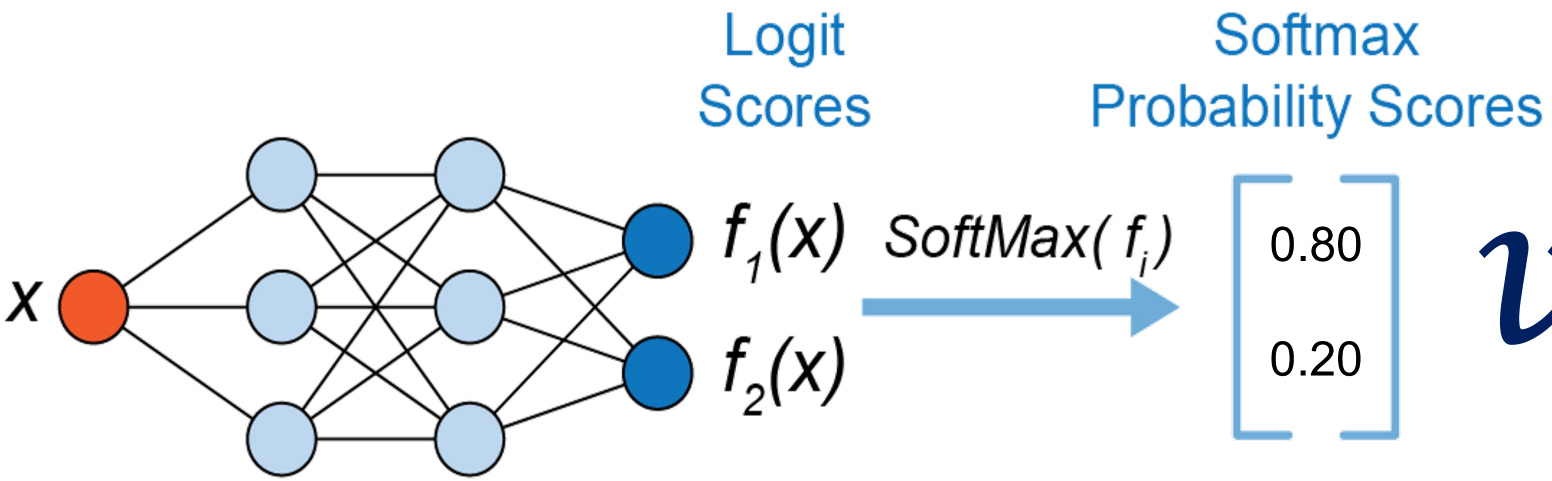
Confidence vs. Accuracy Vital in medicine!



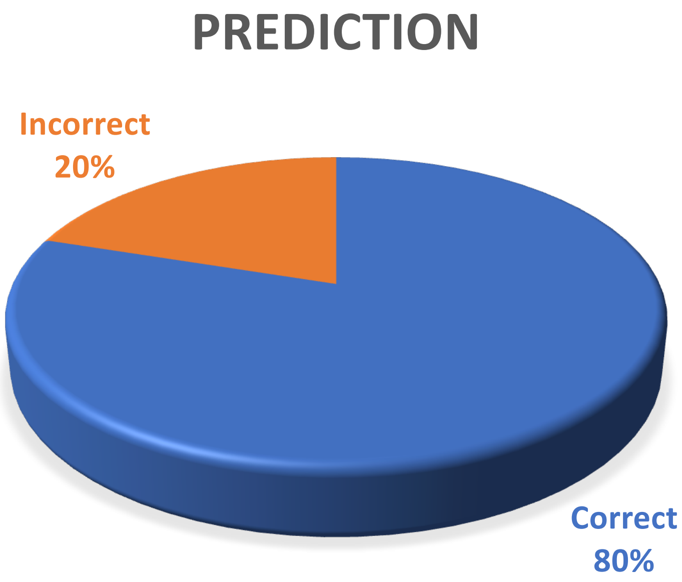
https://paperswithcode.com/sota/image-classification-on-imagenet

Deep Learning Calibration

Confidence vs. Accuracy



vs.



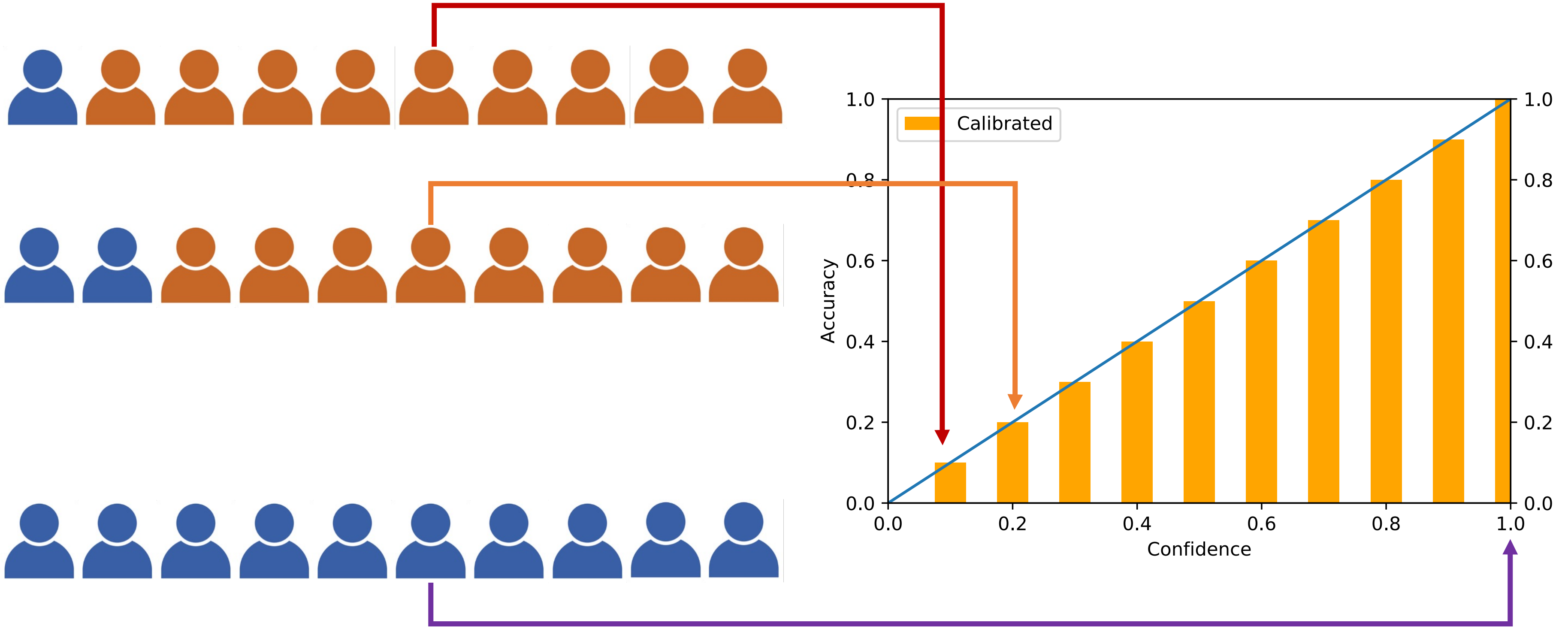
Calibration and Trustworthiness

- Confidence = 0.1
- Accuracy = 1/10 = 0.1

- Confidence = 0.2
- Accuracy = 2/10 = 0.2

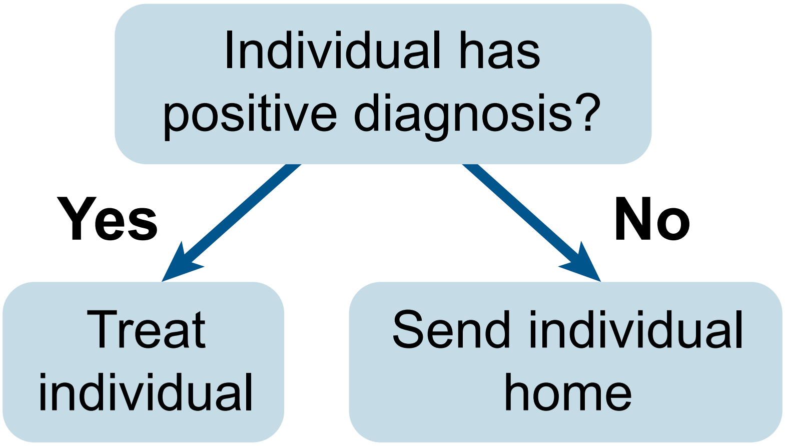
…

- Confidence = 1.0
- Accuracy = 10/10 = 1.0

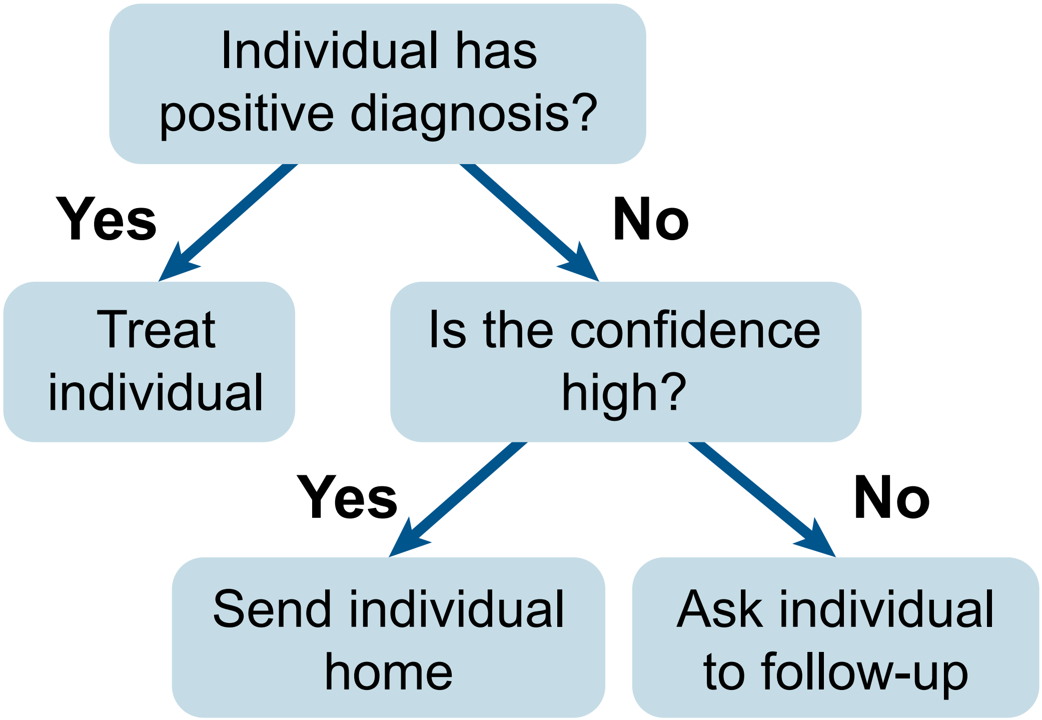


Calibration in Medical Triage

Uncalibrated Model Decision



Calibrated Model Decision

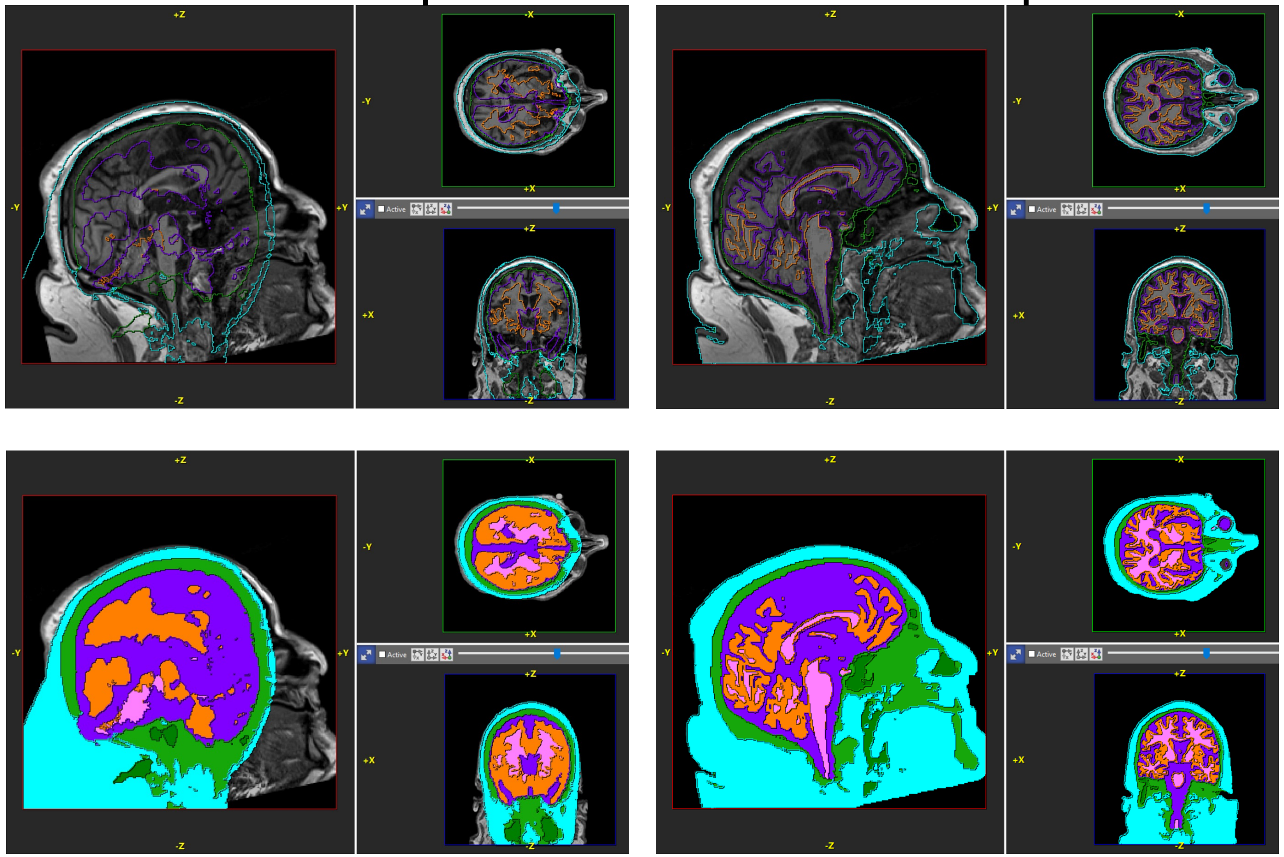


Uncertainty in Segmentation

- Out-of-Distribution (OOD) data:
  - Rarer tissues (e.g., air, blood, tumor)
  - Quality differences (i.e., ringing artifact)
  - Anatomical differences
- Pixel-wise classification:
  - More data
- Most uncertain at tissue boundaries:
  - Partial volume effects

ROAST Output:

Our Output:



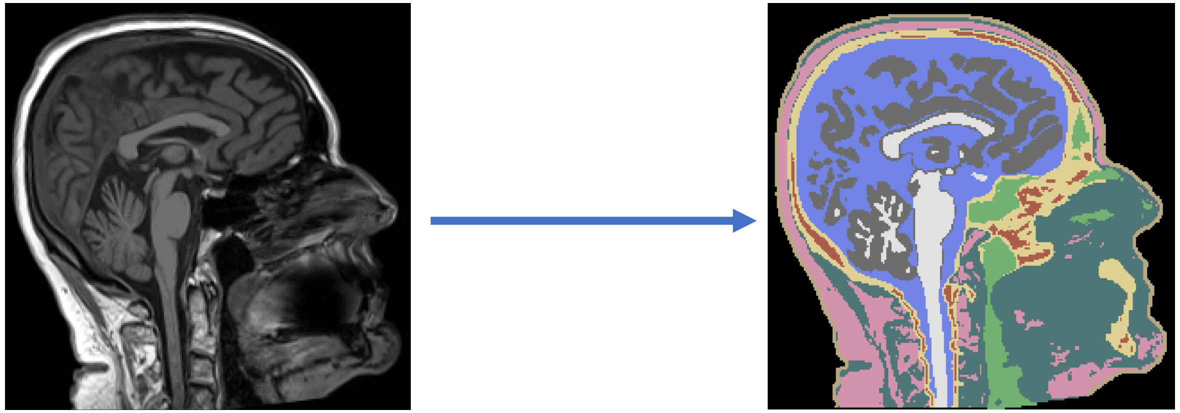
Dataset and Network

Data:

- Augmenting Cognitive Training in Older Adults (ACT) trial (NCT02851511)
- NIH-R01 Phase III clinical trial
- Cognitive training + tDCS for cognitive improvement
- 379 participants
- University of Florida (UF) and University of Arizona (AZ)
- 65-85 years of age

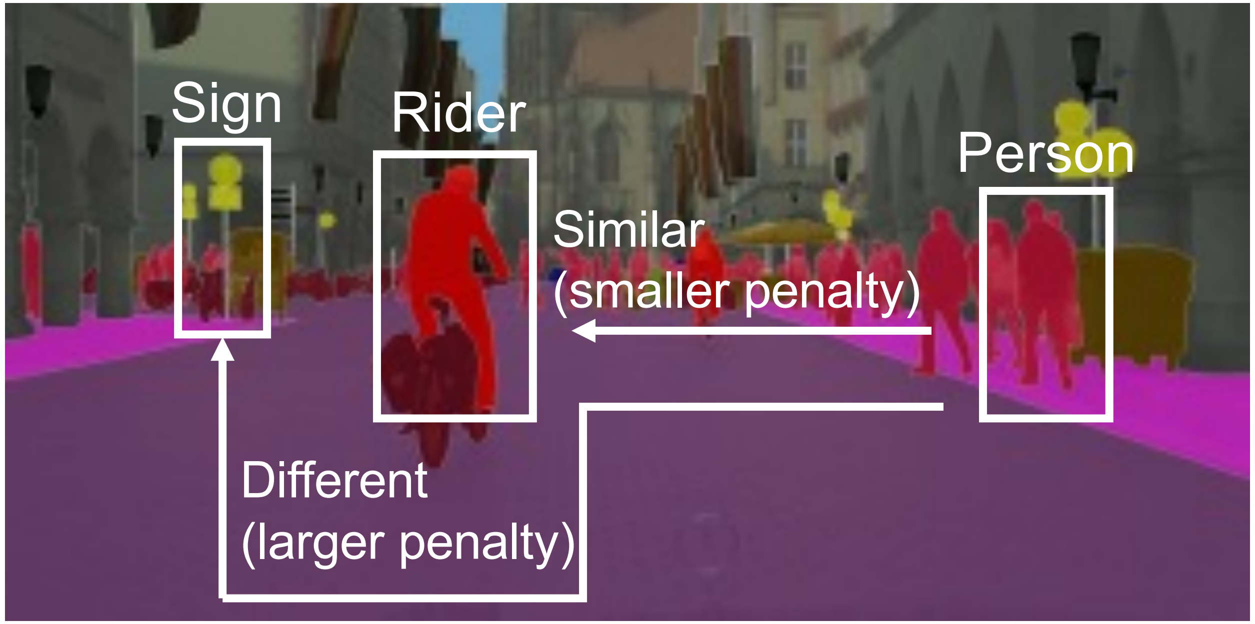
Two Deep Learning Frameworks:

- **U-Net Transformer (UNETR)**
- DeepLab-v3+



Loss Concept

- Estimate meaningful uncertainty
- Uncalibrated loss:
  - Arbitrarily increase confidence
  - Equally confuse classes
- Proposed loss:
  - Similarity is assigned based on task



Example: Task-oriented = Avoid hitting human

Different task could be detecting number of vehicles (person and sign would be closer)

Loss Derivation

$$L(y,\hat{y}) = \ell(y,\hat{y}) + \beta(y' * W * \hat{y})$$

ℓ = Cross entropy (CE), DiceCE, etc.

β = constant

True Class: Person

$$y' = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.0 & 0.9 & 2.2 \\ 0.7 & 0.0 & 2.0 \\ 1.9 & 1.7 & 0.0 \end{bmatrix}$$

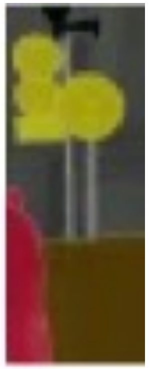
$$y' * W = \begin{bmatrix} 0.7 & 0.0 & 2.0 \end{bmatrix}$$

Multiply $(y' * W)$ by softmax vector to get penalty

True Label: Person

Rider

Person



Sign

Model	Softmax Output	Loss
Model 1 (BEST = CORRECT)	$\hat{y} = \begin{bmatrix} 0.25 \\ 0.50 \\ 0.25 \end{bmatrix}$	CE + $(0.25 * 0.7) + (0.50 * 0.0) + (0.25 * 2.0) =$ 1.00 + 0.675
Model 2 (2nd Best)	$\hat{y} = \begin{bmatrix} 0.90 \\ 0.10 \\ 0.00 \end{bmatrix}$	CE + $(0.90 * 0.7) + (0.10 * 0.0) + (0.0 * 2.0) =$ 3.32 + 0.630
Model 3 (3rd Best)	$\hat{y} = \begin{bmatrix} 0.45 \\ 0.10 \\ 0.45 \end{bmatrix}$	CE + $(0.45 * 0.7) + (0.10 * 0.0) + (0.45 * 2.0) =$ 3.32 + 1.215

DOMINO-CM

- Steps:
  1. Generate confusion matrix using uncalibrated model (UNETR-Base)
  2. Normalize confusion matrix by class
  3. Subtract matrix from identity matrix
  4. Set diagonal values to all zeroes
  5. Scale by a constant (e.g., 3)

Machine learned

	Background	White Matter	Grey Matter	Eyes	CSF	Air	Blood	Cancellous	Cortical	Skin	Fat	Muscle
Background	0	3	3	3	3	3	3	3	3	3	3	3
White Matter	3	0	2.8	3	3	3	3	3	3	3	3	3
Grey Matter	3	2.9	0	3	2.8	3	3	3	3	3	3	3
Eyes	3	3	3	0	2.9	3	3	3	3	2.9	2.9	2.8
CSF	3	3	2.7	3	0	3	3	3	2.8	3	3	3
Air	3	3	3	3	3	0	3	3	2.7	3	3	2.9
Blood	3	3	3	3	2.5	3	0	2.9	2.8	3	2.4	0.91
Cancellous	3	3	3	3	3	3	3	0	2.4	3	3	2.8
Cortical	3	3	3	3	2.9	2.9	3	2.9	0	3	3	2.7
Skin	2.6	3	3	3	3	3	3	3	3	0	3	2.9
Fat	3	3	3	3	3	3	3	3	3	2.9	0	2.9
Muscle	3	3	3	3	3	3	3	3	2.9	2.9	2.9	0

DOMINO-CM

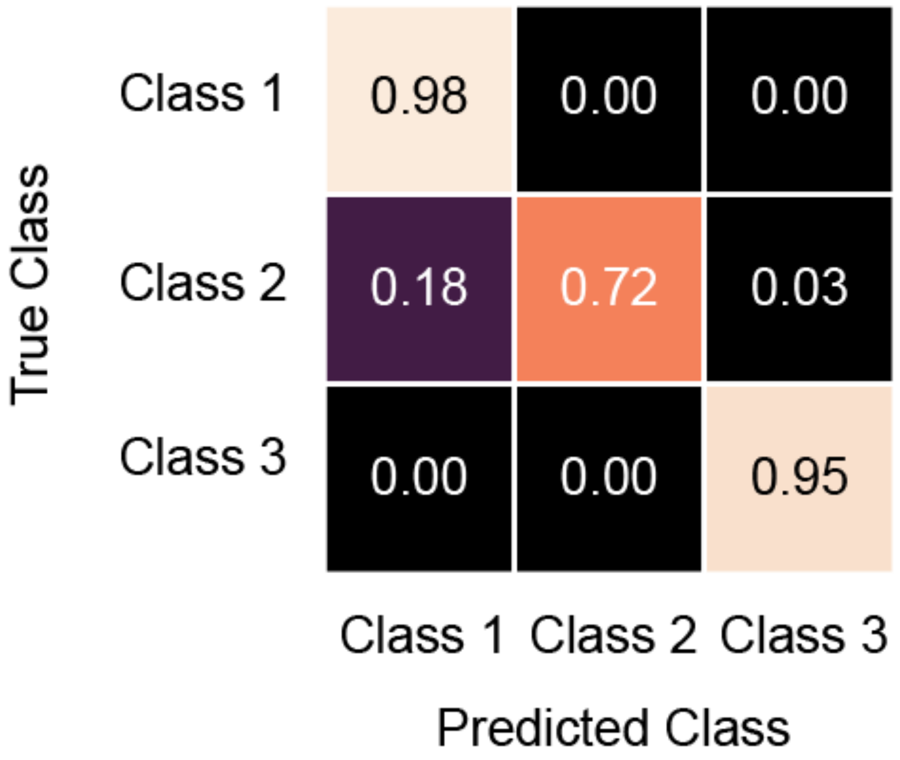
Machine learned

- Steps:
  1. Generate confusion matrix using uncalibrated model (UNETR-Base)
  2. Normalize confusion matrix by class
  3. Subtract matrix from identity matrix
  4. Set diagonal values to all zeroes
  5. Scale by a constant (e.g., 3)

DOMINO-CM

Machine learned

- Steps:
  1. Generate confusion matrix using uncalibrated model (UNETR-Base)
  2. Normalize confusion matrix by class
  3. Subtract matrix from identity matrix
  4. Set diagonal values to all zeroes
  5. Scale by a constant (e.g., 3)



DOMINO-CM

Machine learned

- Steps:
    1. Generate confusion matrix using uncalibrated model (UNETR-Base)
    2. Normalize confusion matrix by class
    3. Subtract matrix from identity matrix
    4. Set diagonal values to all zeroes
    5. Scale by a constant (e.g., 3)

DOMINO-CM

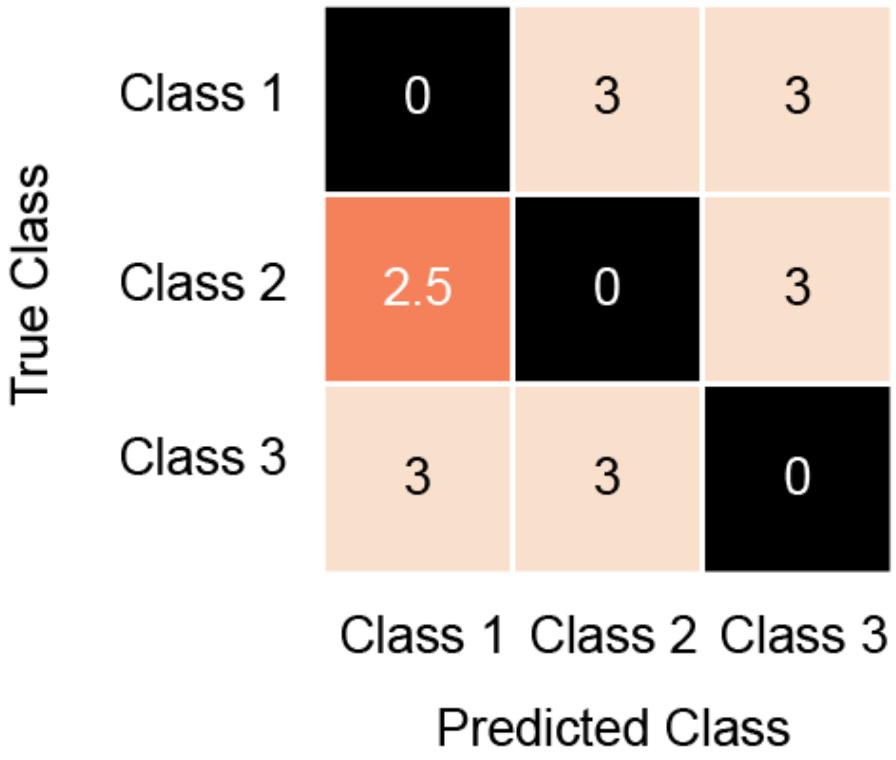
Machine learned

- Steps:
  1. Generate confusion matrix using uncalibrated model (UNETR-Base)
  2. Normalize confusion matrix by class
  3. Subtract matrix from identity matrix
  4. Set diagonal values to all zeroes
  5. Scale by a constant (e.g., 3)

DOMINO-CM

Machine learned

- Steps:
    1. Generate confusion matrix using uncalibrated model (UNETR-Base)
    2. Normalize confusion matrix by class
    3. Subtract matrix from identity matrix
    4. Set diagonal values to all zeroes
    5. Scale by a constant (e.g., 3)



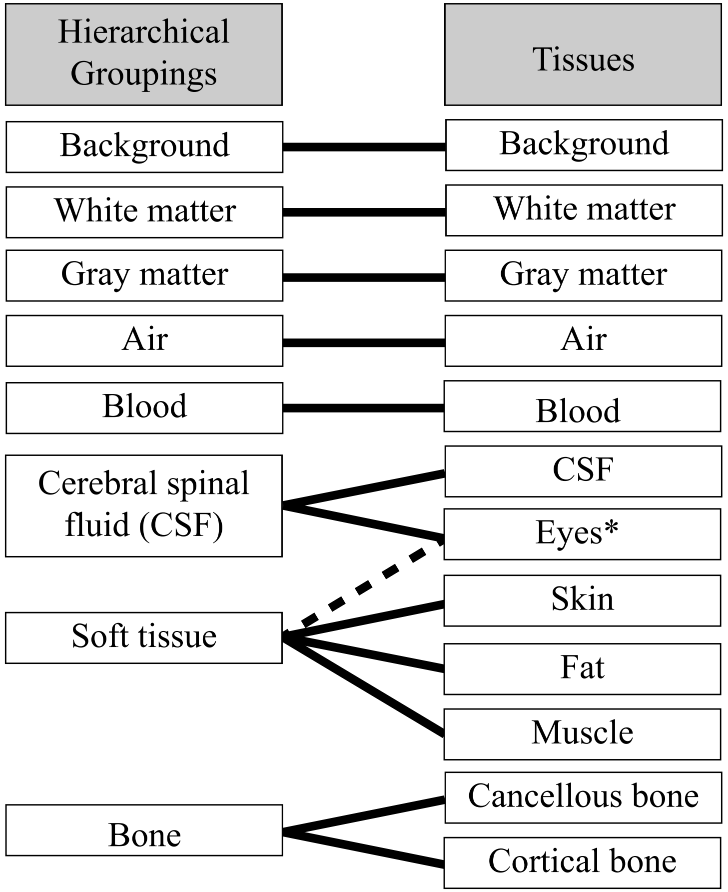
DOMINO-CM

- Steps:
  1. Generate confusion matrix using uncalibrated model (UNETR-Base)
  2. Normalize confusion matrix by class
  3. Subtract matrix from identity matrix
  4. Set diagonal values to all zeroes
  5. Scale by a constant (e.g., 3)

	Background	White Matter	Grey Matter	Eyes	CSF	Air	Blood	Cancellous	Cortical	Skin	Fat	Muscle
Background	0	3	3	3	3	3	3	3	3	3	3	3
White Matter	3	0	2.8	3	3	3	3	3	3	3	3	3
Grey Matter	3	2.9	0	3	2.8	3	3	3	3	3	3	3
Eyes	3	3	3	0	2.9	3	3	3	3	2.9	2.9	2.8
CSF	3	3	2.7	3	0	3	3	3	2.8	3	3	3
Air	3	3	3	3	3	0	3	3	2.7	3	3	2.9
Blood	3	3	3	3	2.5	3	0	2.9	2.8	3	2.4	0.91
Cancellous	3	3	3	3	3	3	3	0	2.4	3	3	2.8
Cortical	3	3	3	3	2.9	2.9	3	2.9	0	3	3	2.7
Skin	2.6	3	3	3	3	3	3	3	3	0	3	2.9
Fat	3	3	3	3	3	3	3	3	3	2.9	0	2.9
Muscle	3	3	3	3	3	3	3	3	2.9	2.9	2.9	0

Machine learned

DOMINO-HC Expert learned



	Background	White Matter	Grey Matter	Eyes	CSF	Air	Blood	Cancellous	Cortical	Skin	Fat	Muscle
Background	0	3	3	3	3	3	3	3	3	3	3	3
White Matter	3	0	3	3	3	3	3	3	3	3	3	3
Grey Matter	3	3	0	3	3	3	3	3	3	3	3	3
Eyes	3	3	3	0	2.4	3	3	3	3	2.7	2.7	2.7
CSF	3	3	3	2.4	0	3	3	3	3	3	3	3
Air	3	3	3	3	3	0	3	3	3	3	3	3
Blood	3	3	3	3	3	3	0	3	3	3	3	3
Cancellous	3	3	3	3	3	3	3	0	2.1	3	3	3
Cortical	3	3	3	3	3	3	3	2.1	0	3	3	3
Skin	3	3	3	2.7	3	3	3	3	3	0	2.1	2.1
Fat	3	3	3	2.7	3	3	3	3	3	2.1	0	2.1
Muscle	3	3	3	2.7	3	3	3	3	3	2.1	2.1	0

Top-N Accuracy

- 11 tissue accuracy scores show superiority of the DOMINO-CM method
- 6 tissue model also shows that DOMINO-CM is superior

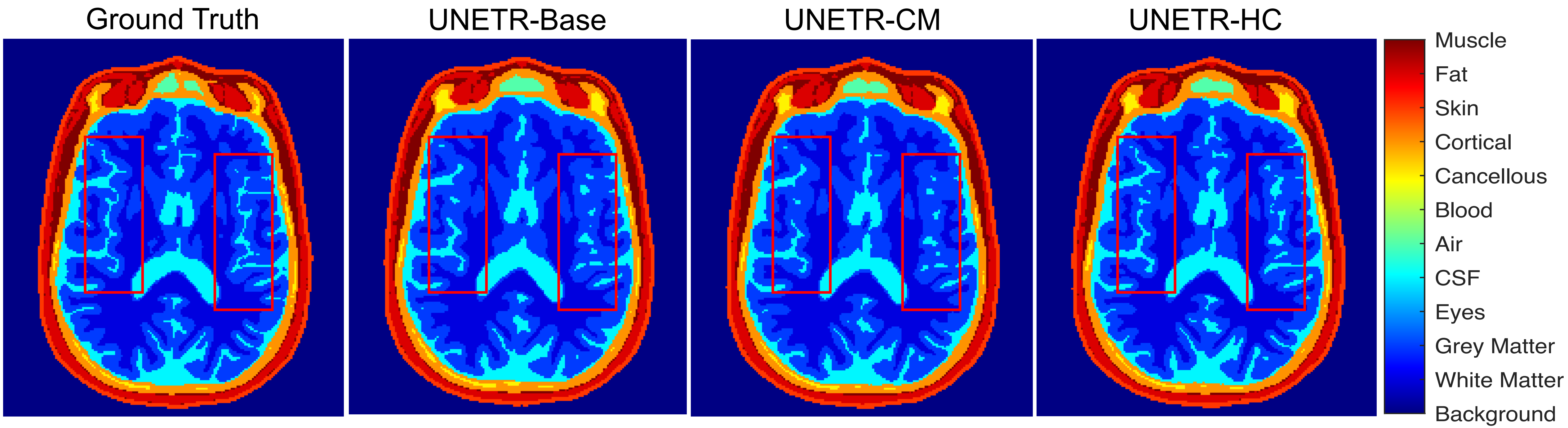
Method	Top-1	Top-2	Top-3
UNETR-Base	0.876	0.979	0.990
UNETR-HC	0.891	0.984	0.993
UNETR-CM	**0.895**	**0.986**	**0.996**

Method	Top-1	Top-2	Top-3
HEADRECO	0.905	0.977	0.983
UNETR-Base	0.913	0.993	0.998
UNETR-HC	0.924	0.995	0.998
UNETR-CM	**0.928**	**0.996**	**0.999**

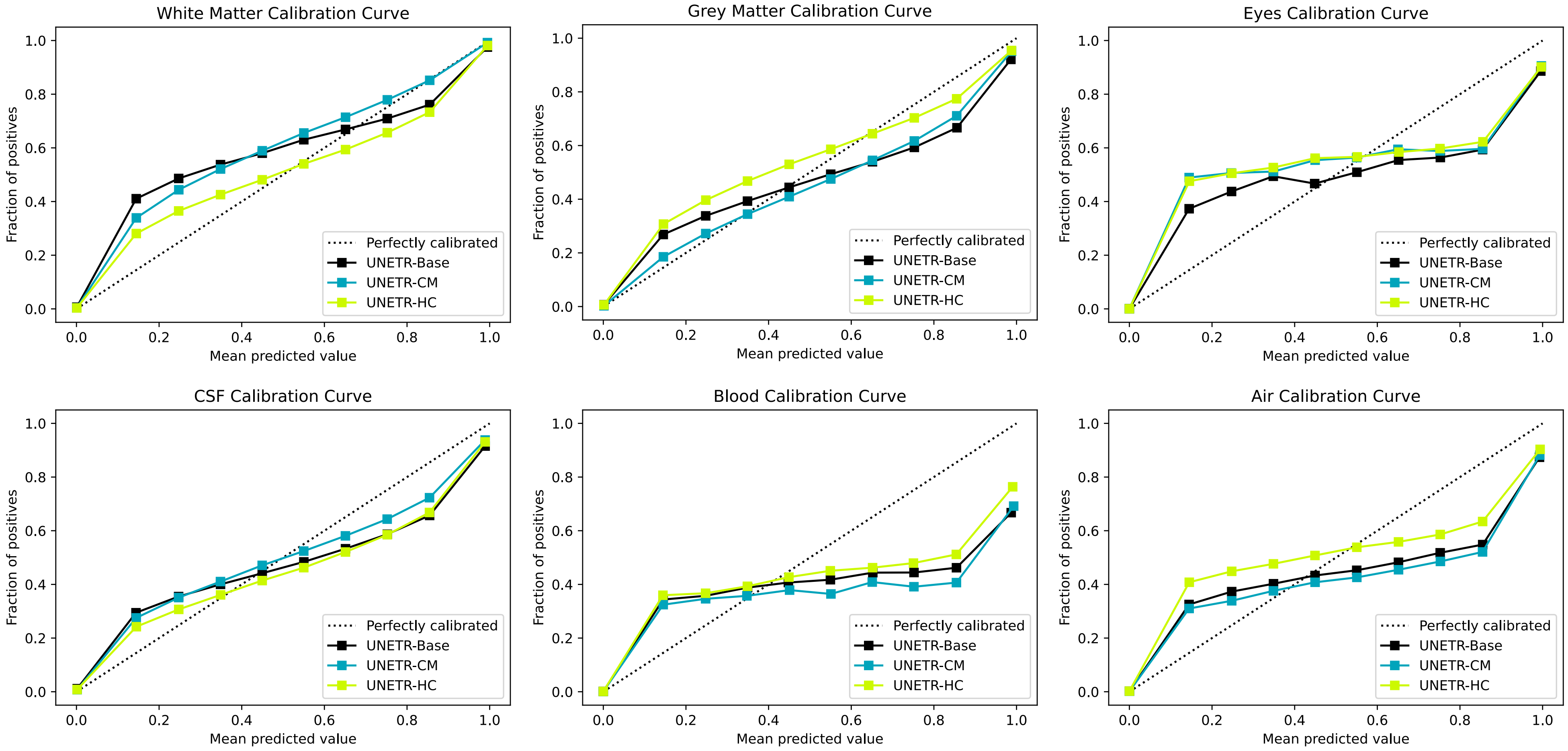
Better Regional Accuracy →

Visual Results on 11 tissues

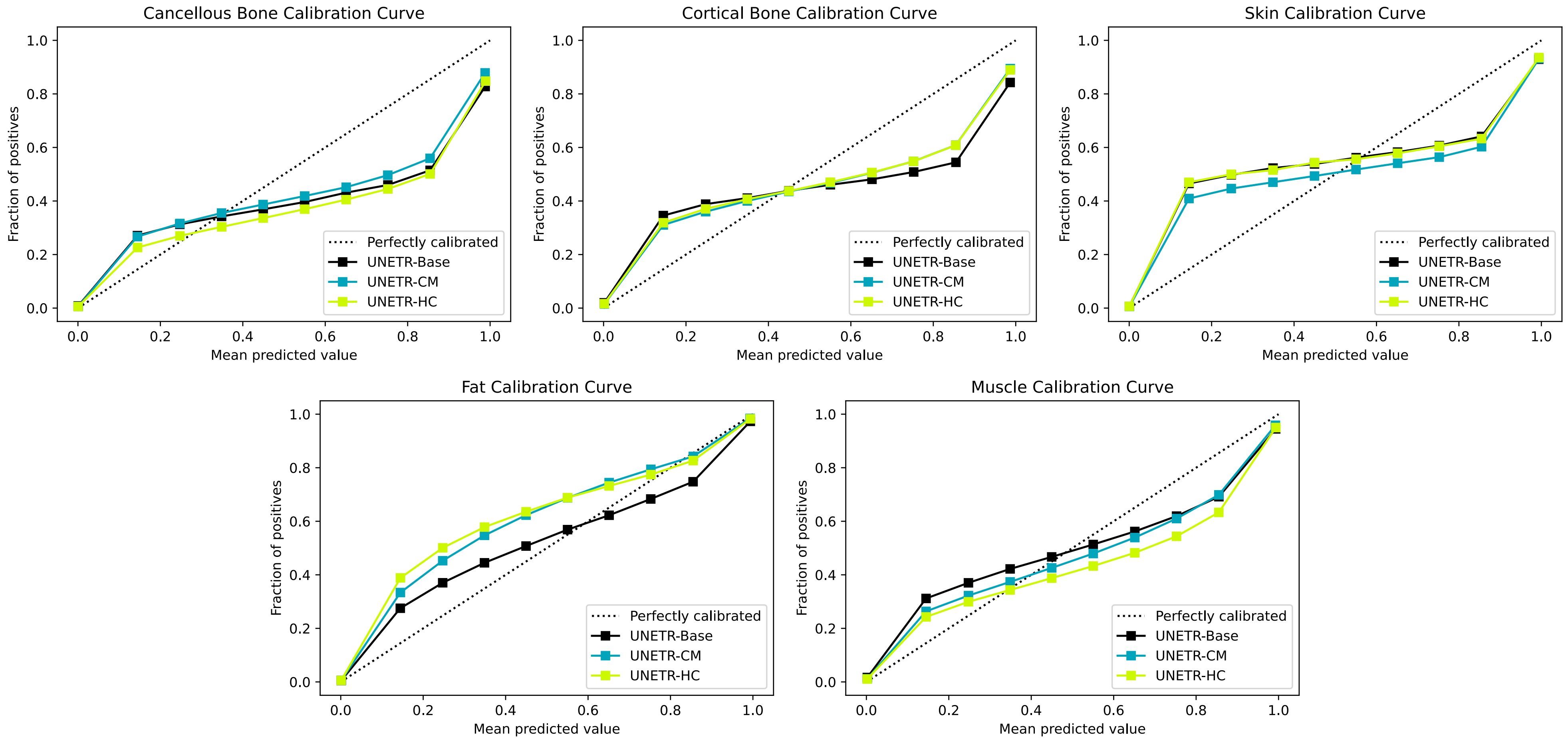
Better Boundary Detection



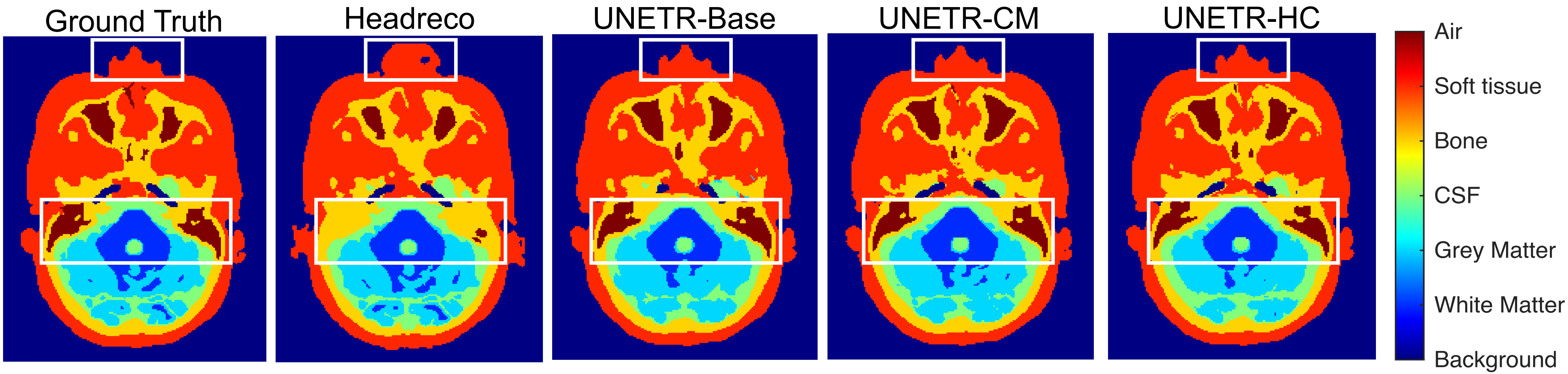
Model Calibration Results (11 tissue)



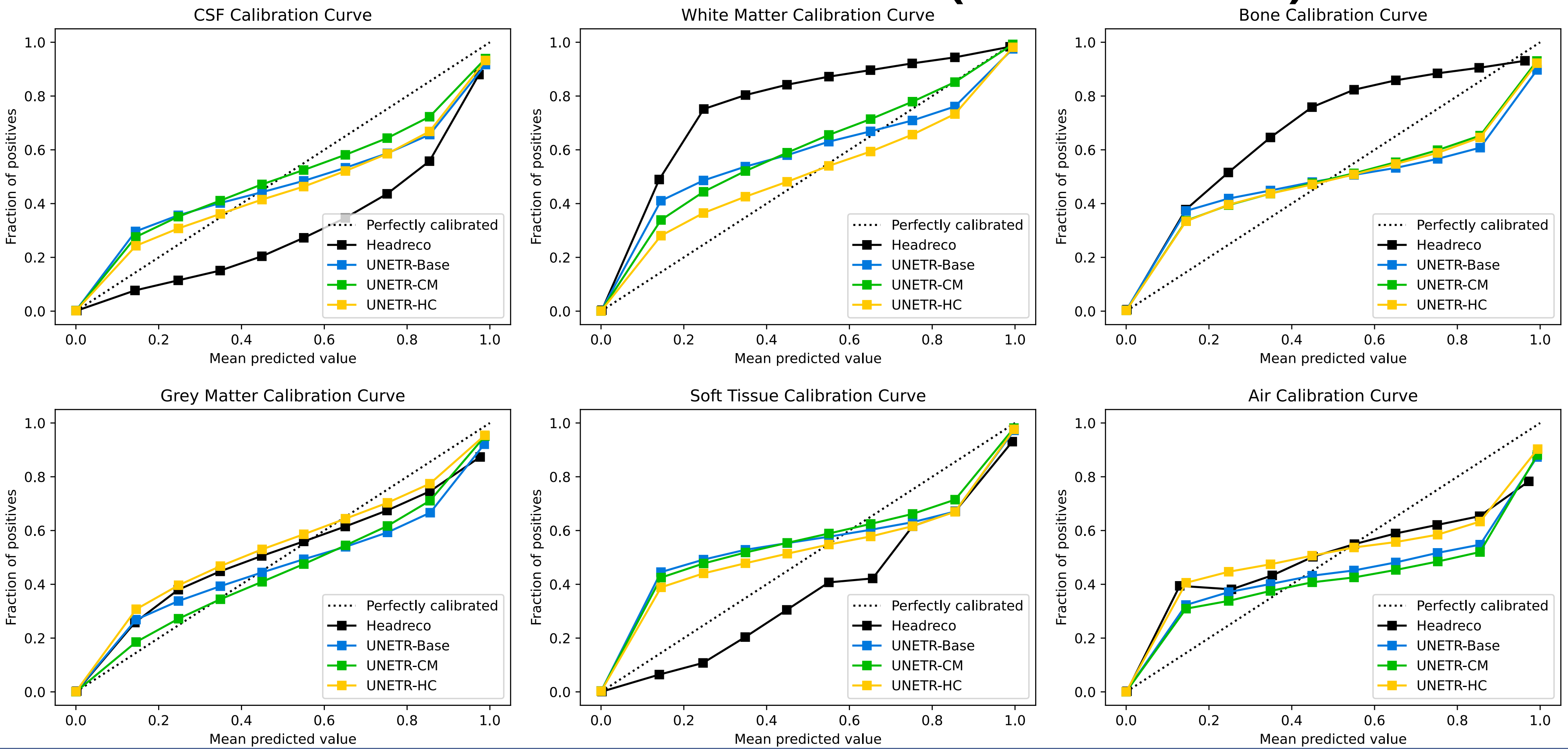
Model Calibration Results (11 tissue)



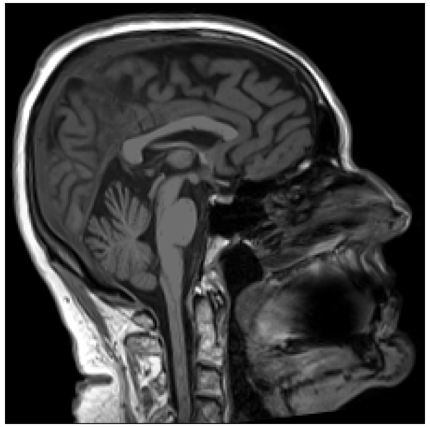
Visual Results on 6 tissues



Model Calibration Results (6 tissue)



Conclusions



- Developed two DOMINO methods:
  - DOMINO-CM – better at regional accuracy
  - DOMINO-HC – better at boundary detection
- Model improves calibration without sacrificing accuracy
- Promising results generalizable to public Cityscapes dataset

Source: https://www.cityscapes-dataset.com/

Acknowledgements

BME & SMILE Lab
- Ruogu Fang, PhD (thesis advisor)
- Kyle See, BS
- Ziqian Huang, MS

MAE & Air Force Research Laboratory
- Kevin Brink, PhD
- Matthew Hale, PhD
- Kyle Volle, PhD

Center for Cognitive Aging and Memory
- Adam J. Woods, PhD
- Aprinda Indahlastari, PhD
- Alejandro Albizu, BS

NVIDIA
- Huiwen Ju, PhD
- Kaleb Smith, PhD

Funding Support

NIH NIA RF1AG071469
NIH NIA R01AG054077
NSF IIS 1908299
NSF-AFRL INTERN Supplement 2130885.

NSF
NIH
National Institutes of Health
AIR FORCE RESEARCH LABORATORY

MICCAI2022 Singapore
25th International Conference on Medical Image Computing and Computer Assisted Intervention
September 18-22, 2022
Resorts World Convention Centre Singapore

MICCAI

DOMINO: Domain-aware Model Calibration in Medical Image Segmentation

Poster Session 7

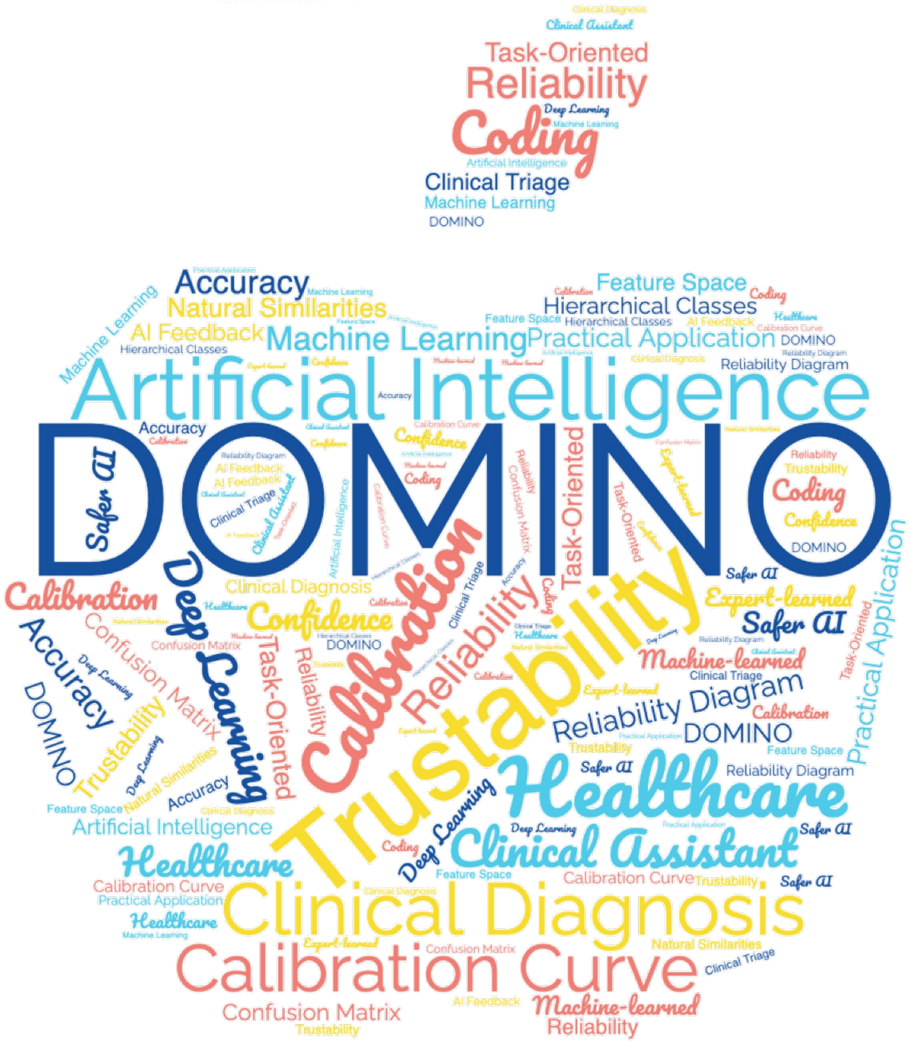
Paper ID: 1693

Image Segmentation, Registration & Reconstruction III

Wednesday, September 21, 2022, 10:30-11:30 AM

Pre-trained models on 11 tissue segmentation from T1 MRIs

github.com/lab-smile/DOMINO


Task-Oriented
Reliability
Coding
Clinical Triage
Machine Learning
DOMINO
Accuracy
Natural Similarities
AI Feedback
Machine Learning
Feature Space
Hierarchical Classes
Practical Application
Reliability Diagram
Artificial Intelligence
DOMINO
Calibration
Deep Learning
Confidence
Clinical Diagnosis
Trustability
Reliability
Task-Oriented
Confusion Matrix
Accuracy
Expert-learned
Safer AI
Machine-learned
Reliability Diagram
Healthcare
Clinical Assistant
Calibration Curve
Clinical Diagnosis
Calibration Curve
Confusion Matrix

Thank you!

Backup Slides