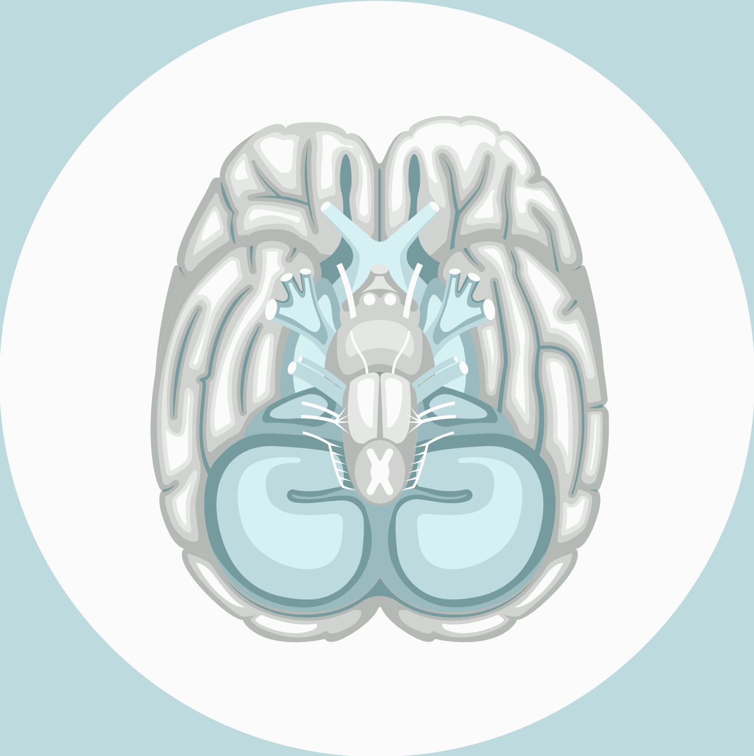


Comparative Study of Machine Learning and Deep Learning Multi-Class Segmentation Methods

Elizabeth Nemeti and Shaniah Reece

Introduction and State of the Literature

MRI: non-invasive imaging for brain tumor diagnosis

Detection via "segmentation":

- *manual* - radiologists annotate images
- *automated* - ML/DL classifies pixels

Figure. Radiologist manually segmenting MRI images with software.

Limitations of Manual Segmentation

1. Image quality
2. Experience/opinion of the analyst
3. Sheer volume of images in large dataset
4. Morphology: small size of early tumors
5. Human error + inconsistency

Advantages of Automated Segmentation

1. Efficiency and speed in large datasets
2. Quantitative + qualitative evaluation metrics
3. Early tumor detection
4. Speedy analysis of complex underlying patterns

Limitations

1. Image quality (noise, blur, size)
2. Class imbalances
3. Computational resource constraints
4. Smaller unlabeled datasets

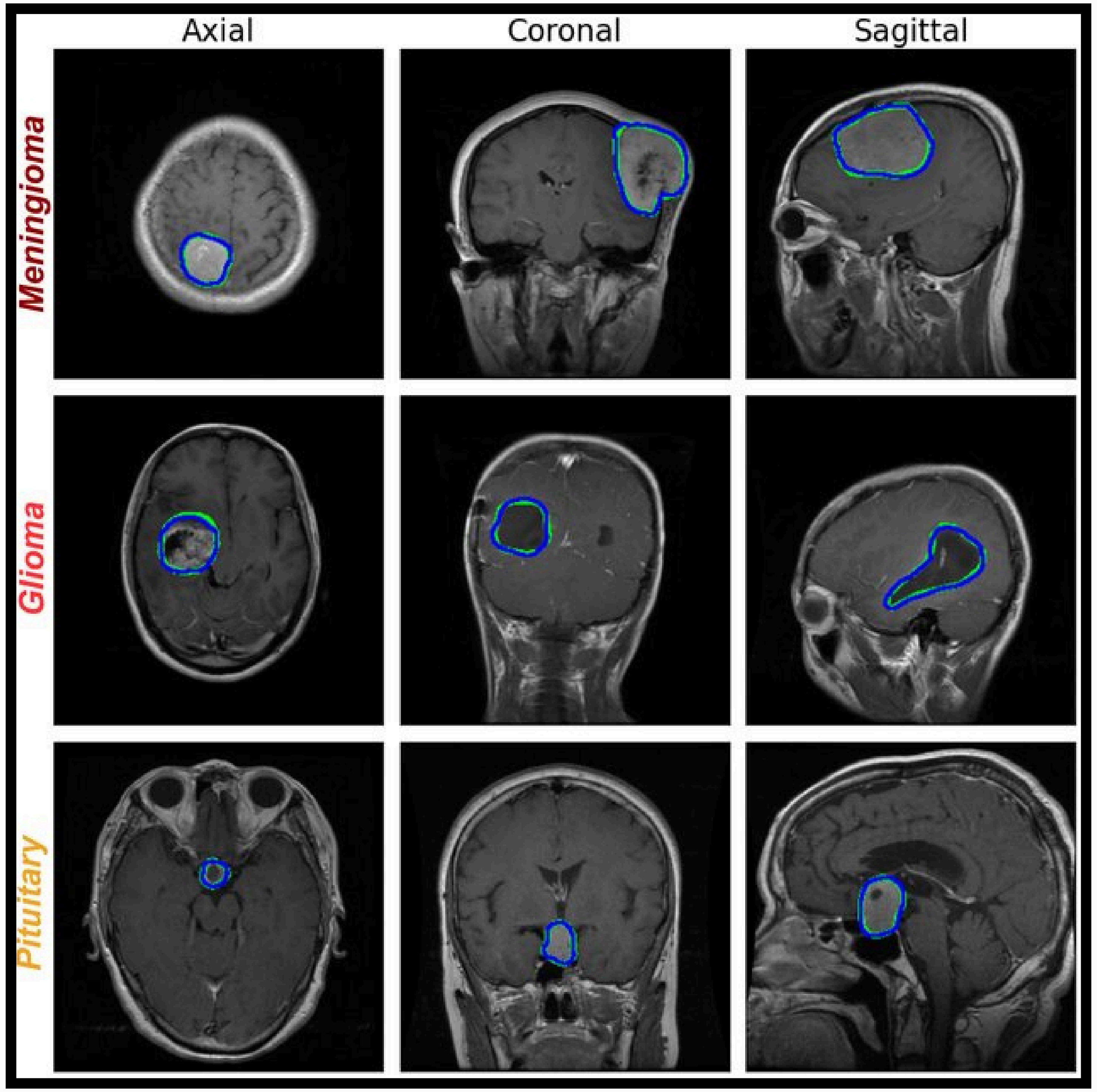


Figure. (Maas et al., 2021) Network Segmentation - High DICE Scores: Three brain tumor classes in the three different planes from our local test set. Green delineations are the radiologist annotations (ground truth) and blue borders are the network's predicted multi-class segmentations correctly classified.

The Shift from ML to DL in the Literature

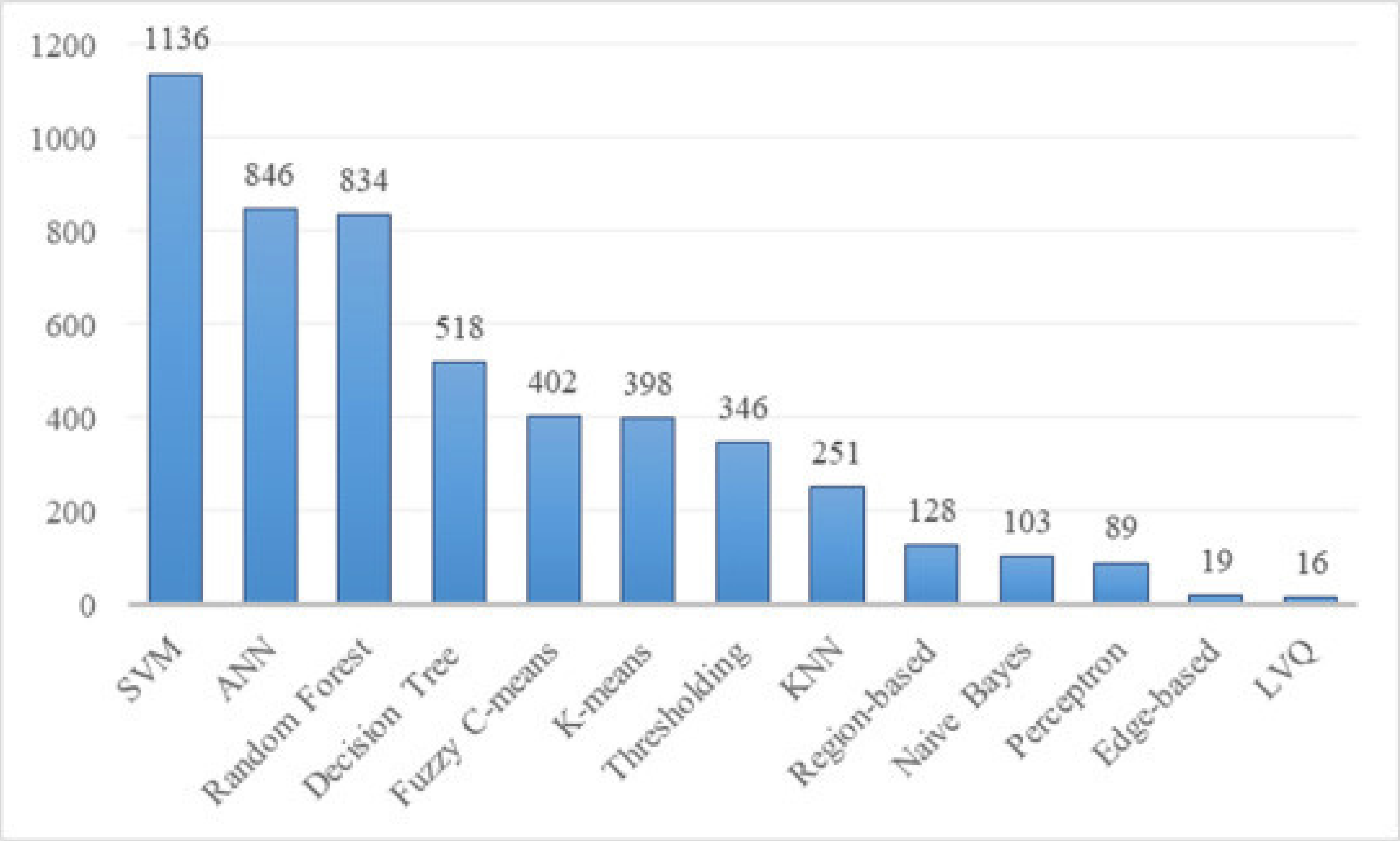


Figure. Ranjbarzadeh et al. 2023
Number of publications between 2015 and 2022 for brain tumor segmentation using supervised and unsupervised learning techniques

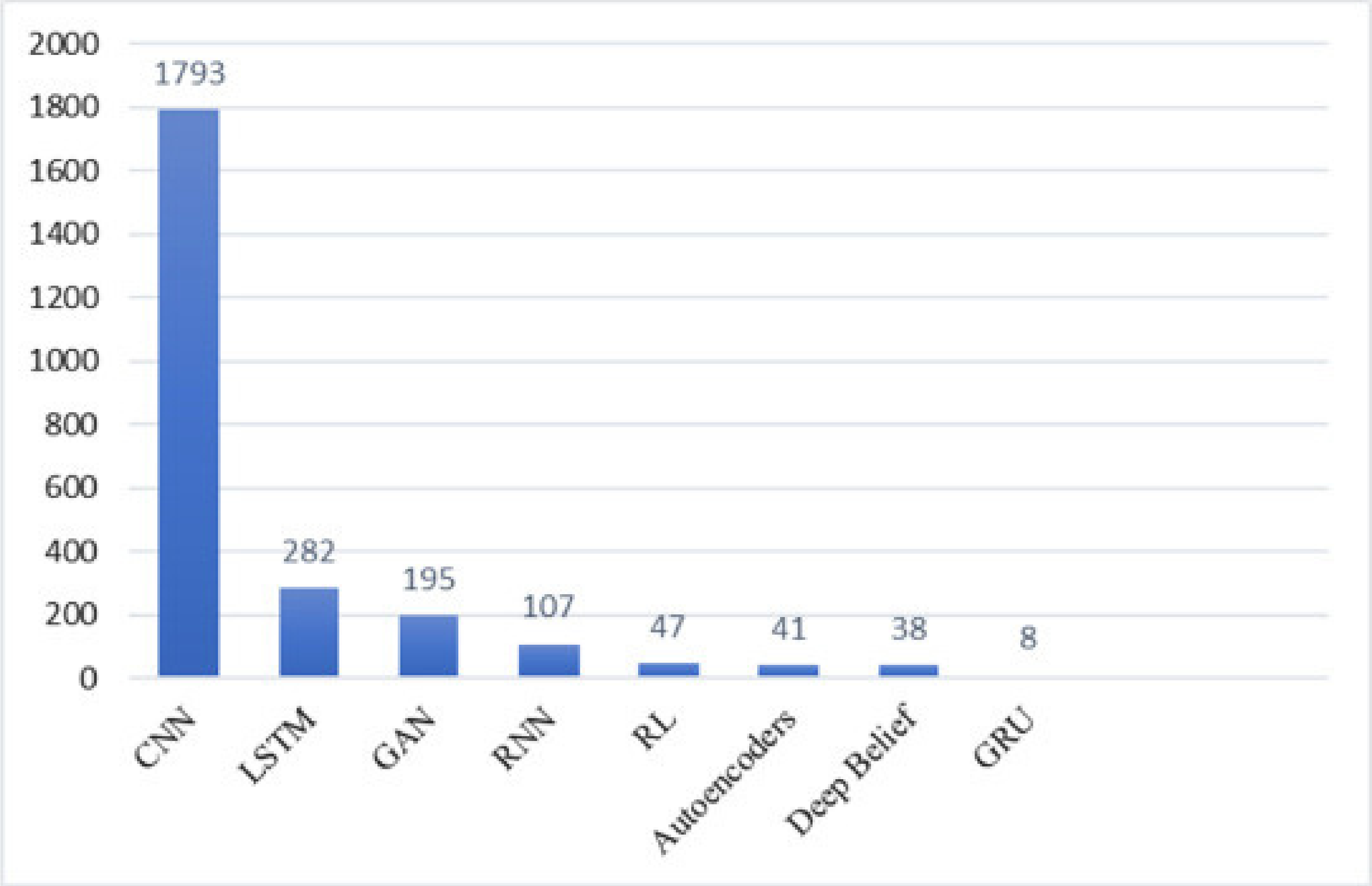
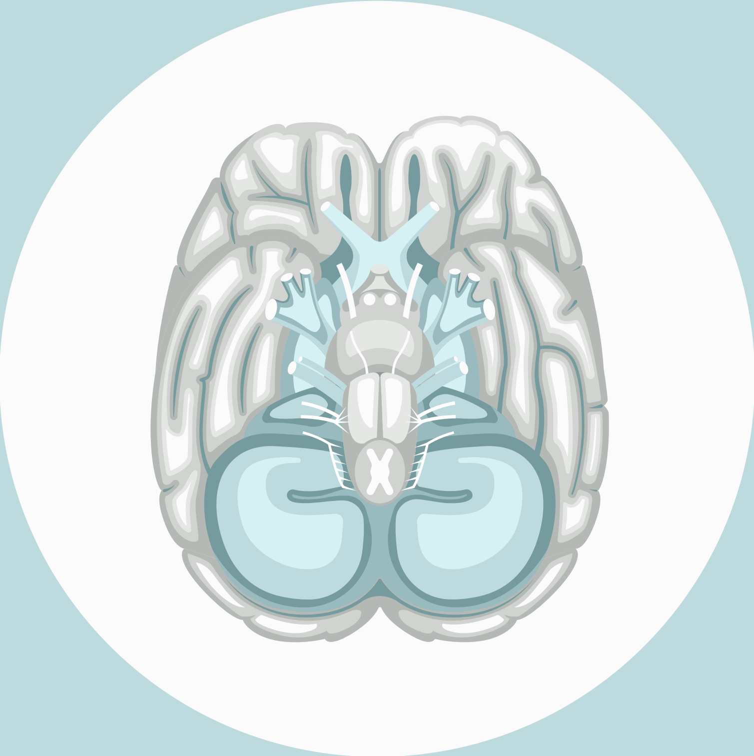


Figure. Ranjbarzadeh et al. 2023
Number of publications between 2015 and 2022 for brain tumor segmentation using DL models



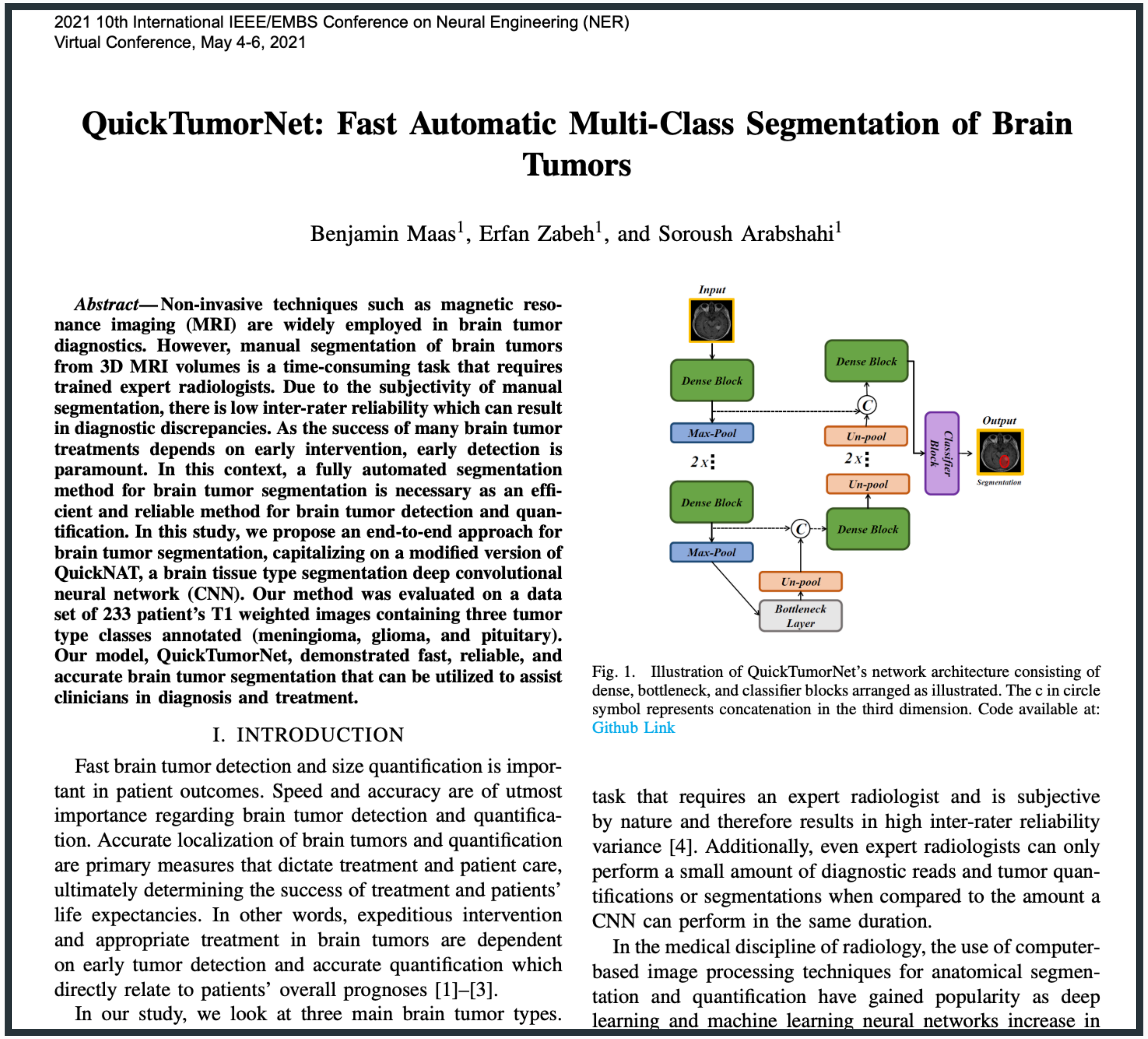
A Comparative Study of Machine Learning and Deep Learning Multi-Class Segmentation Methods

Aims/Novelty

Mass et al. 2021:

- Automated U-based deep CNN
- Used our same dataset
- Same multiclass segmentation task

We specifically address and expand upon the areas of future work outlined by Maas et al. 2021.



2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)
Virtual Conference, May 4-6, 2021

QuickTumorNet: Fast Automatic Multi-Class Segmentation of Brain Tumors

Benjamin Maas[1], Erfan Zabeh[1], and Soroush Arabshahi[1]

Abstract—**Non-invasive techniques such as magnetic resonance imaging (MRI) are widely employed in brain tumor diagnostics. However, manual segmentation of brain tumors from 3D MRI volumes is a time-consuming task that requires trained expert radiologists. Due to the subjectivity of manual segmentation, there is low inter-rater reliability which can result in diagnostic discrepancies. As the success of many brain tumor treatments depends on early intervention, early detection is paramount. In this context, a fully automated segmentation method for brain tumor segmentation is necessary as an efficient and reliable method for brain tumor detection and quantification. In this study, we propose an end-to-end approach for brain tumor segmentation, capitalizing on a modified version of QuickNAT, a brain tissue type segmentation deep convolutional neural network (CNN). Our method was evaluated on a data set of 233 patient's T1 weighted images containing three tumor type classes annotated (meningioma, glioma, and pituitary). Our model, QuickTumorNet, demonstrated fast, reliable, and accurate brain tumor segmentation that can be utilized to assist clinicians in diagnosis and treatment.**

Fig. 1. Illustration of QuickTumorNet's network architecture consisting of dense, bottleneck, and classifier blocks arranged as illustrated. The c in circle symbol represents concatenation in the third dimension. Code available at: Github Link

I. INTRODUCTION

Fast brain tumor detection and size quantification is important in patient outcomes. Speed and accuracy are of utmost importance regarding brain tumor detection and quantification. Accurate localization of brain tumors and quantification are primary measures that dictate treatment and patient care, ultimately determining the success of treatment and patients' life expectancies. In other words, expeditious intervention and appropriate treatment in brain tumors are dependent on early tumor detection and accurate quantification which directly relate to patients' overall prognoses [1]–[3].

In our study, we look at three main brain tumor types.

task that requires an expert radiologist and is subjective by nature and therefore results in high inter-rater reliability variance [4]. Additionally, even expert radiologists can only perform a small amount of diagnostic reads and tumor quantifications or segmentations when compared to the amount a CNN can perform in the same duration.

In the medical discipline of radiology, the use of computer-based image processing techniques for anatomical segmentation and quantification have gained popularity as deep learning and machine learning neural networks increase in

Aims

To address the need for comparing CNNs with simpler models to enhance interpretability, we will:

- ***Compare a U-NET (CNN) with both an SVM and FCM across multiple performance metrics and qualitative visual results.***

To address the need for utilizing more varied/diversified datasets in training, we will:

- ***Implement data augmentation for U-NET and benchmark all three models with the well-established BRaTs 2018 dataset.***

To address the labeling challenge in medical imaging (frequent noisy/inconsistent labels) we will:

- ***Assess the effectiveness of supervised (i.e. UNET and SVM) versus unsupervised segmentation methods (i.e. FCM).***

Selected Dataset

- 3064 MRI images
- TI-weighted contrast-enhanced
- Great qualty: 512 x 512 pixels
- 233 patients
- Labeled: radiologist annotated tumors

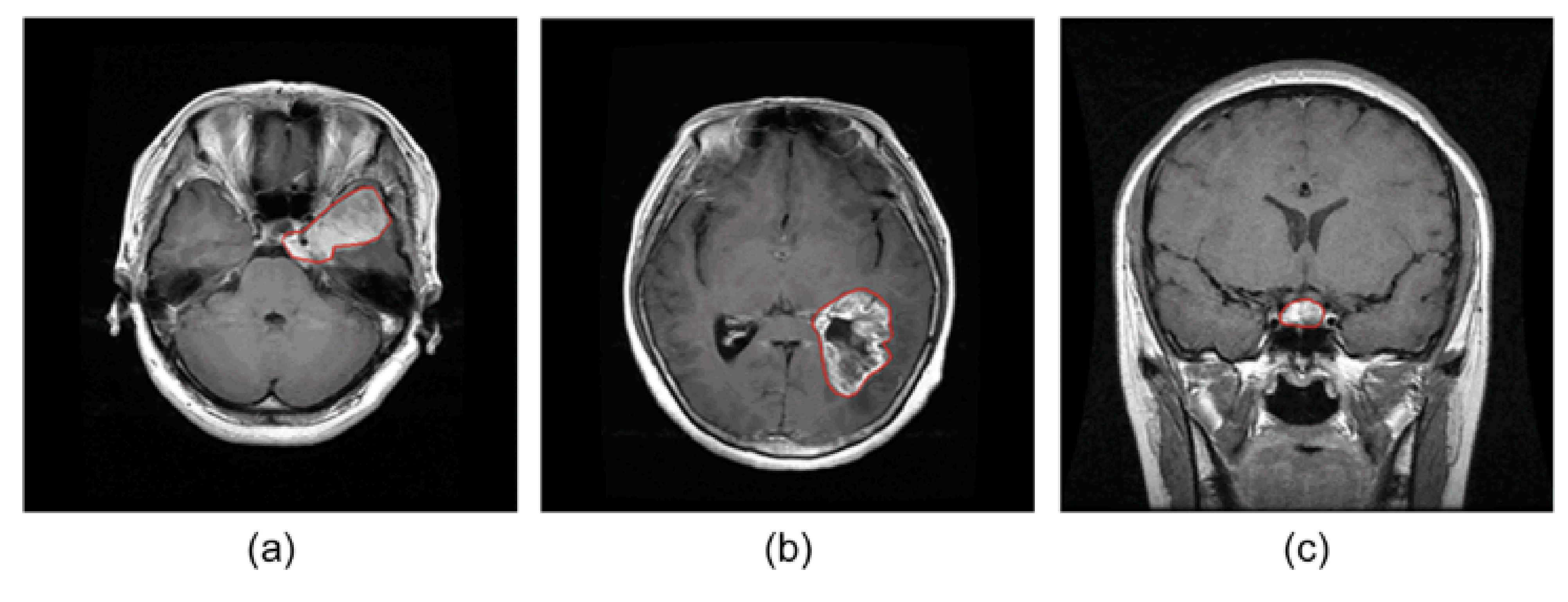


Fig. 3. Illustrations of three typical brain tumors: (a) meningioma; (b) glioma; and (c) pituitary tumor. Red lines indicate the tumor border. [3]

Cheng, Jun (2017). brain tumor dataset. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1512427.v5

Handling Class Imbalance

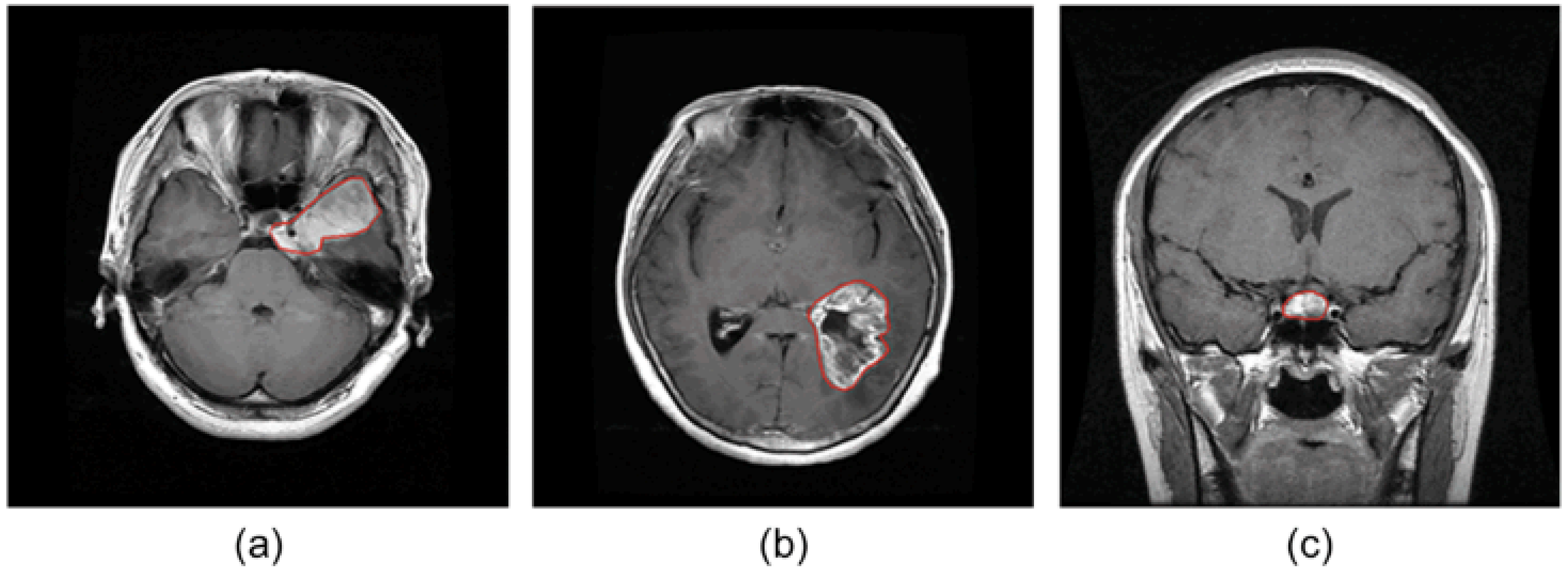


Fig. 3. Illustrations of three typical brain tumors: (a) meningioma; (b) glioma; and (c) pituitary tumor. Red lines indicate the tumor border. [3]

Original Data Distribution:
(a) meningioma 708 images
(b) glioma 1426 images
(c) pituitary tumor 930 images

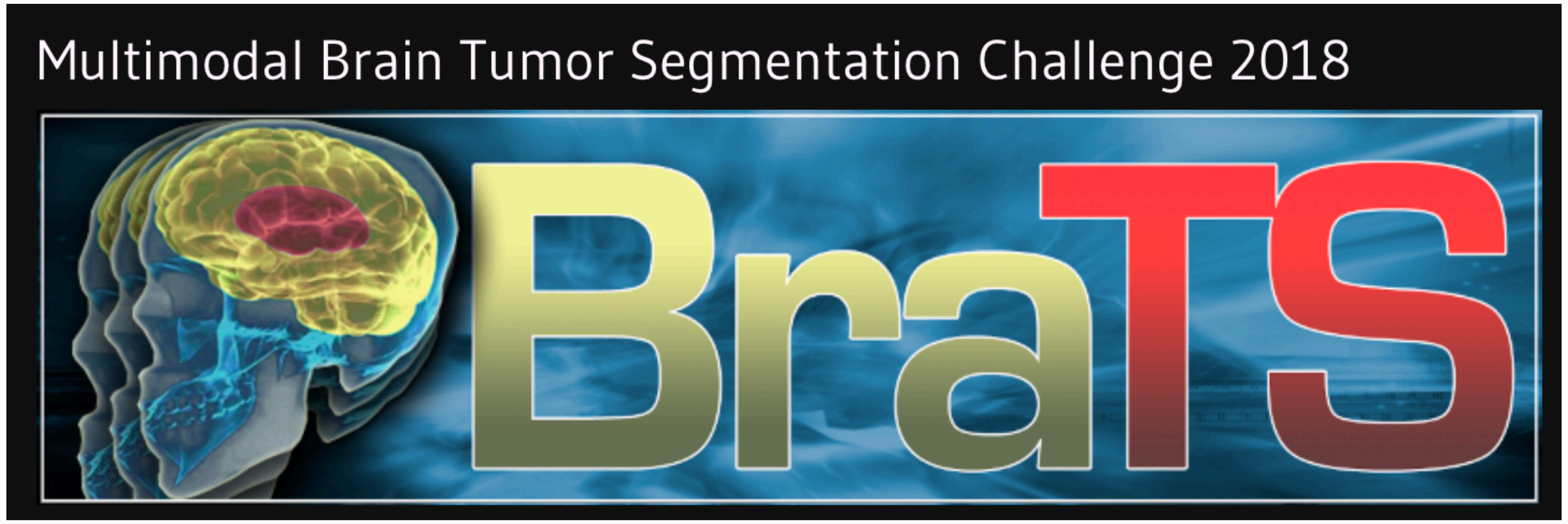
Balancing algorithm

New Data Distribution:
(a) 708 images
(b) 708 images
(c) 708 images

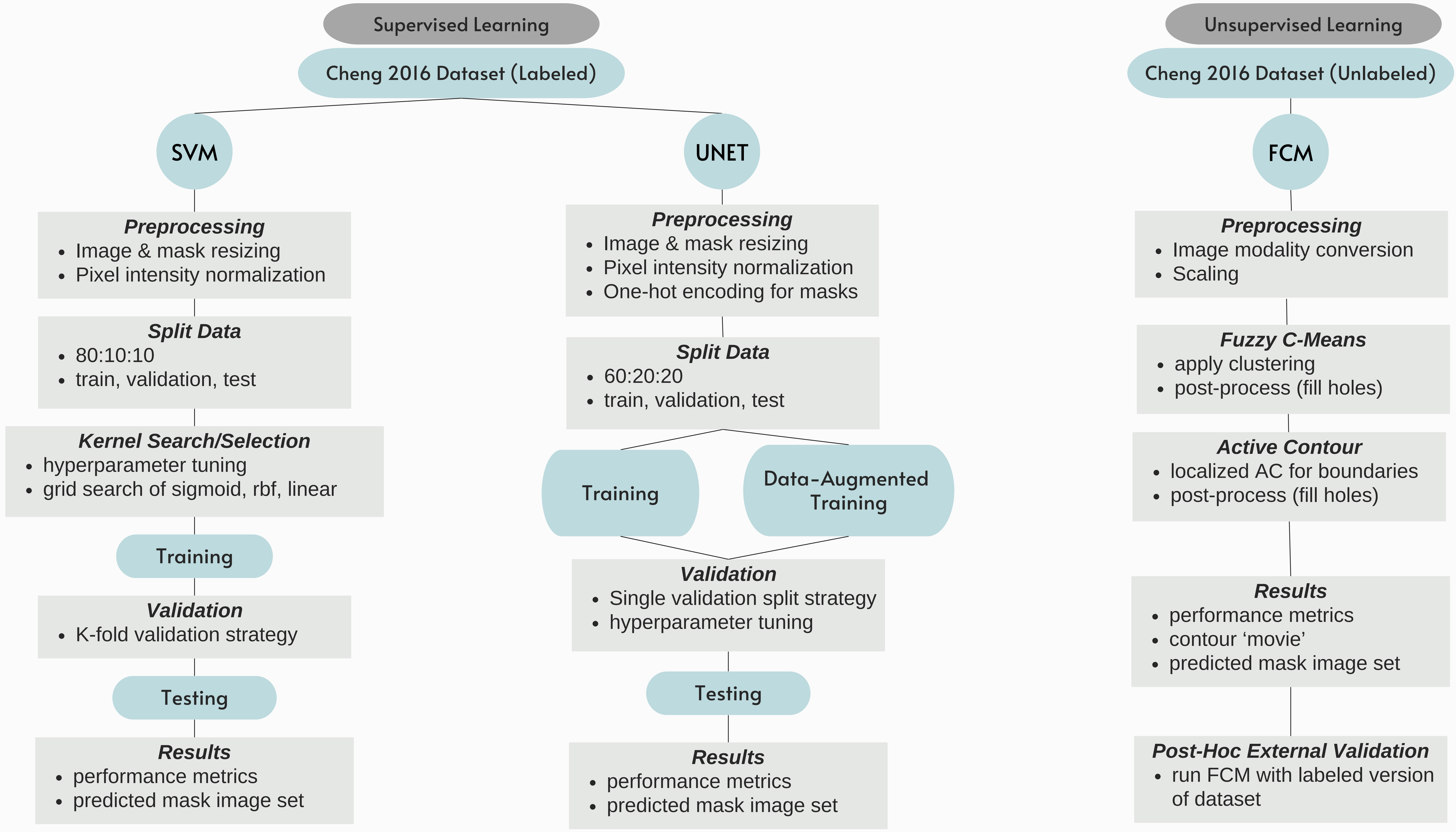
Benchmark Dataset

(enforces better model generalizability on unseen data)

- sourced from 19 institutions/diverse scanners
- 3gb worth of multimodal MRI images
- 4 modalities
- Labeled: radiologist annotated tumors



https://www.med.upenn.edu/sbia/brats2018/data.html



Supervised Learning
Cheng 2016 Dataset (Labeled)
SVM
Preprocessing
• Image & mask resizing
• Pixel intensity normalization
Split Data
• 80:10:10
• train, validation, test
Kernel Search/Selection
• hyperparameter tuning
• grid search of sigmoid, rbf, linear
Training
Validation
• K-fold validation strategy
Testing
Results
• performance metrics
• predicted mask image set
UNET
Preprocessing
• Image & mask resizing
• Pixel intensity normalization
• One-hot encoding for masks
Split Data
• 60:20:20
• train, validation, test
Training
Data-Augmented Training
Validation
• Single validation split strategy
• hyperparameter tuning
Testing
Results
• performance metrics
• predicted mask image set
Unsupervised Learning
Cheng 2016 Dataset (Unlabeled)
FCM
Preprocessing
• Image modality conversion
• Scaling
Fuzzy C-Means
• apply clustering
• post-process (fill holes)
Active Contour
• localized AC for boundaries
• post-process (fill holes)
Results
• performance metrics
• contour 'movie'
• predicted mask image set
Post-Hoc External Validation
• run FCM with labeled version of dataset

	UNET	SVM	FCM + AC
Evaluation Metrics	Sensitivity Specificity Accuracy Precision Confusion Matrix Jaccard Index (IoU) Dice Similarity	Sensitivity Specificity Accuracy Precision Confusion Matrix Jaccard Index Dice Similarity	Silhouette Score Davies - Bouldin Calinski-Harabasz

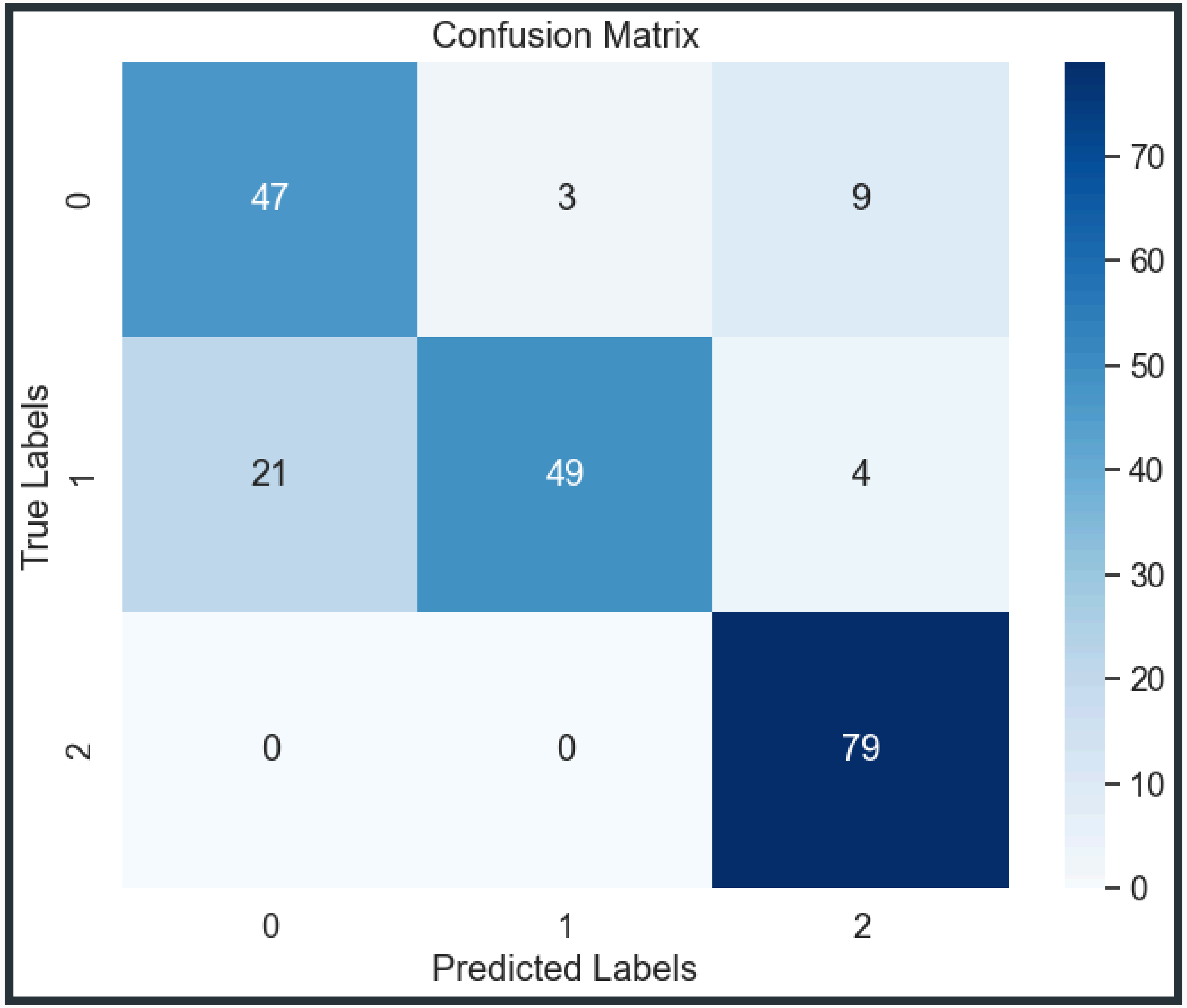
Preliminary Results

SVM	UNET	FCM
Accuracy: 0.83 Precision: 0.84 Recall: 0.83 Jaccard Index: 0.71 F1 Score: 0.83 Dice Similarity: 0.79 Specificity: 0.86	Accuracy: 0.99 Precision: 0.01 Recall: 0.99 Specificity: 0.99 IoU: 0.016 Dice: 0.031	Silhouette Score Davies - Bouldin Calinski-Harabasz

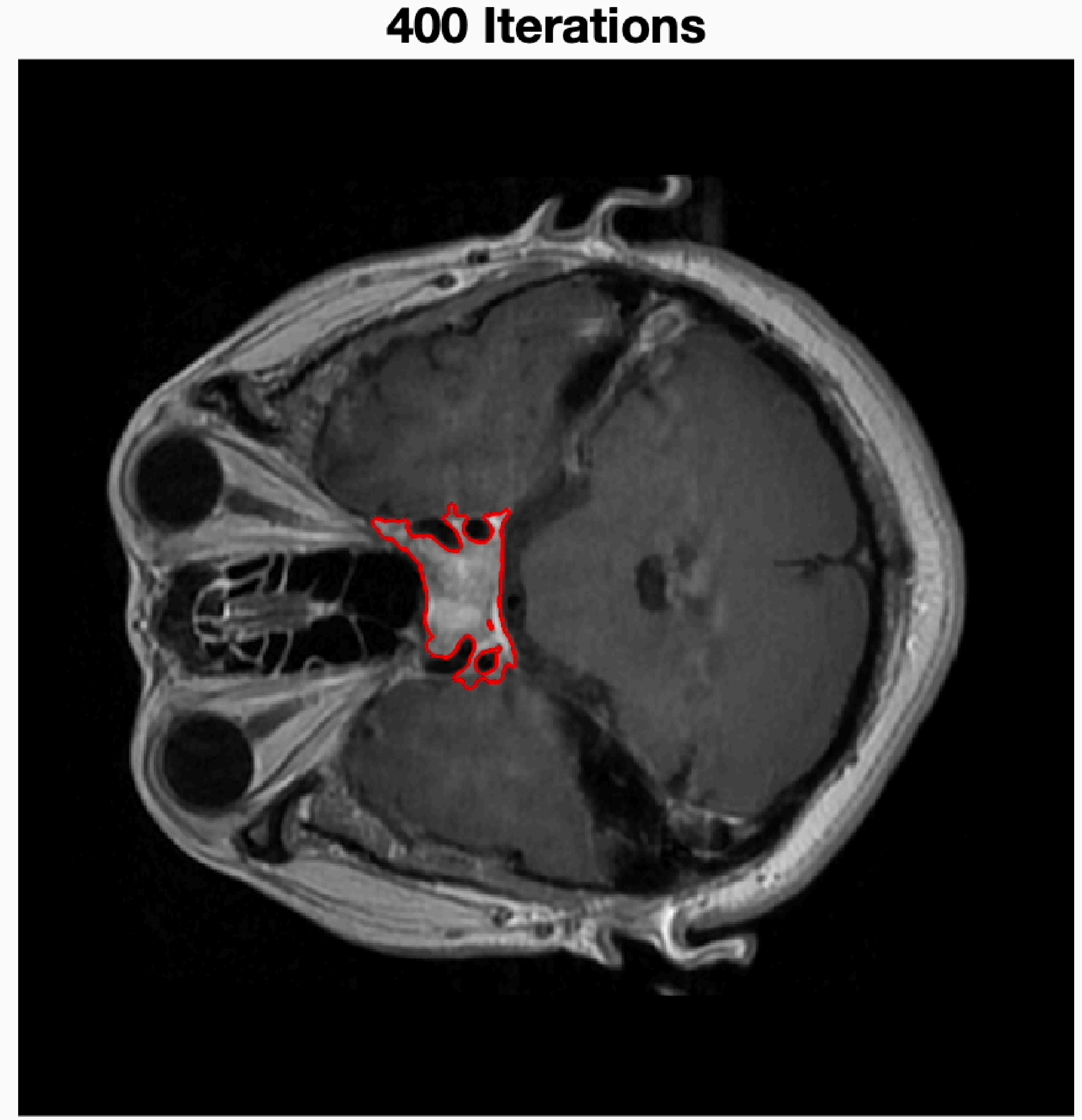
Preliminary Results

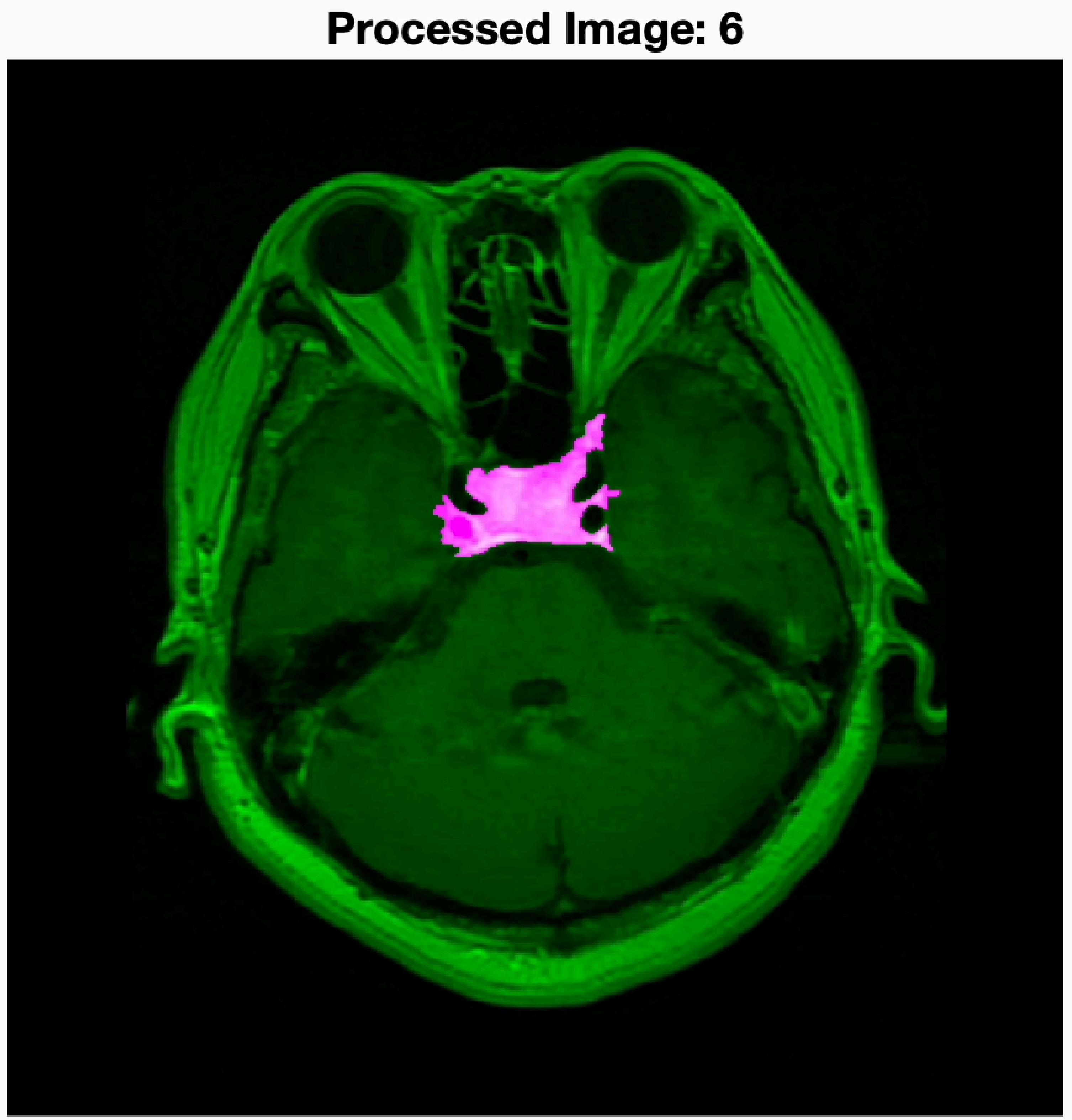
RBF	Linear	Sigmoid	Average
Accuracy: 0.83 Precision: 0.84 Recall: 0.83 Jaccard Index: 0.71 F1 Score: 0.83 Dice Similarity: 0.79S pecificity: 0.86	Accuracy : 0.81 Precision : 0.81 Recall: 0.81 Jaccard Index : 0.69 F1 Score : 0.81 Dice Similarity: 0.74 Specificity: 0.79	Accuracy: 0.53 Precision: 0.52 Recall: 0.53 Jaccard Index: 0.36 F1 Score: 0.52 Dice Similarity: 0.45 Specificity: 0.41	Accuracy: 0.53 Precision: 0.52 Recall: 0.53 Jaccard Index: 0.36 F1 Score: 0.52 Dice Similarity: 0.45 Specificity: 0.41

Quantitative Results: SVM

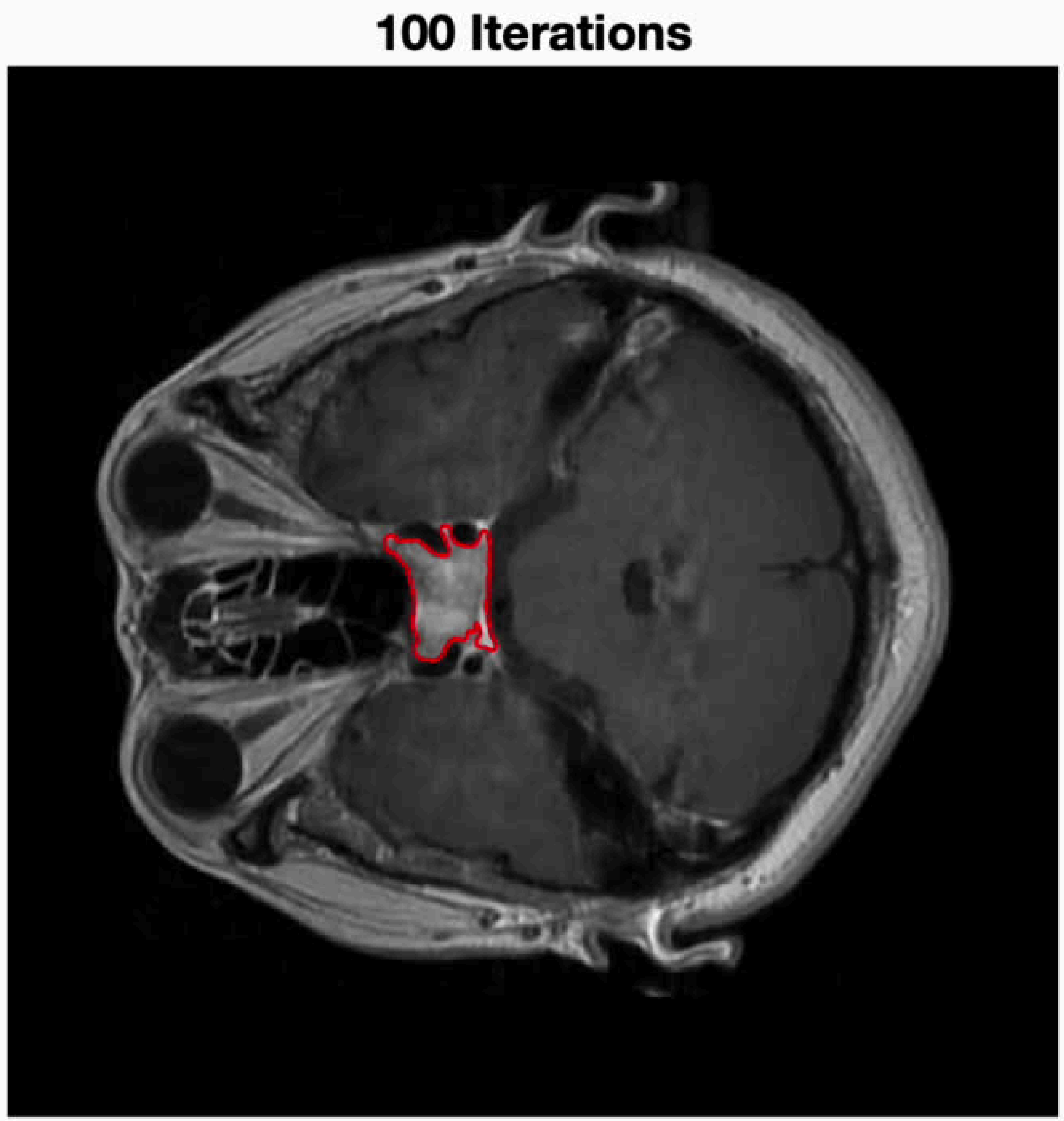


Qualitative Results: FCM (img)

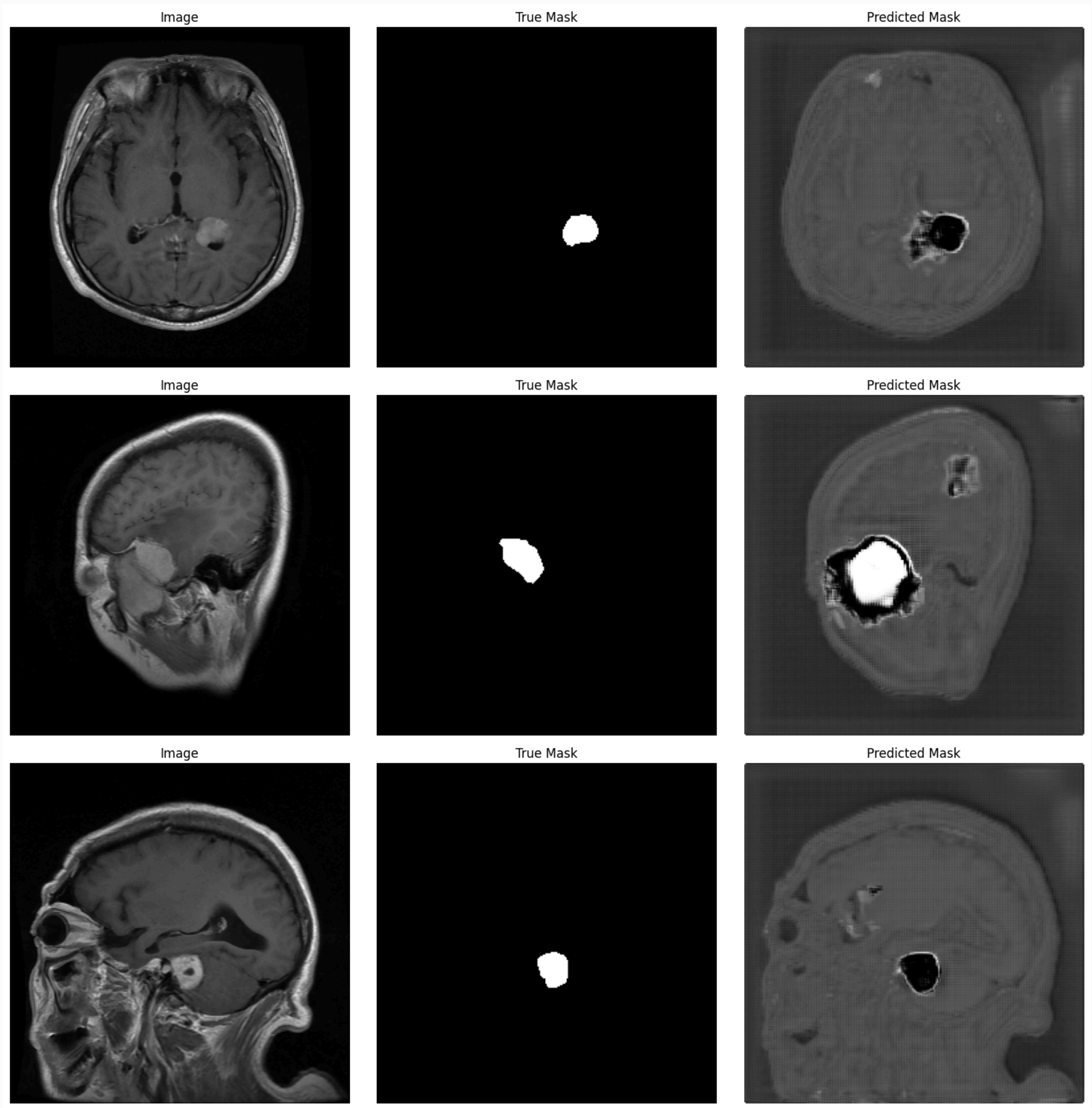




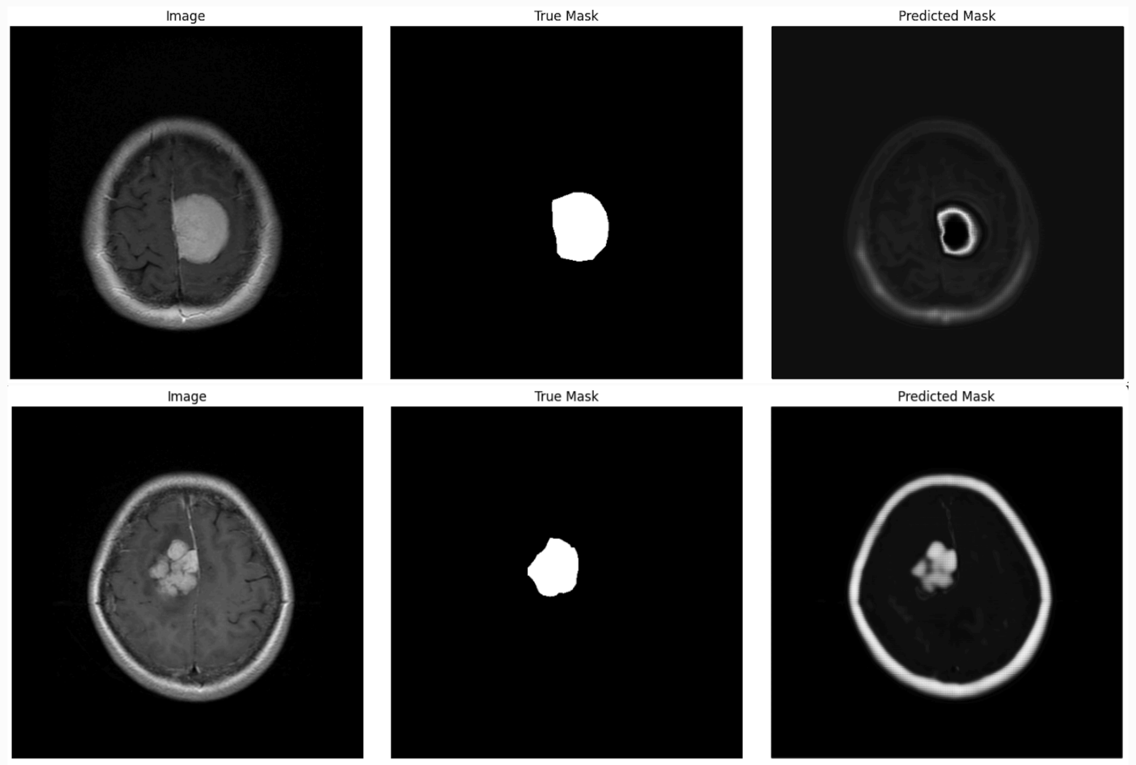
Qualitative Results: FCM (mov)



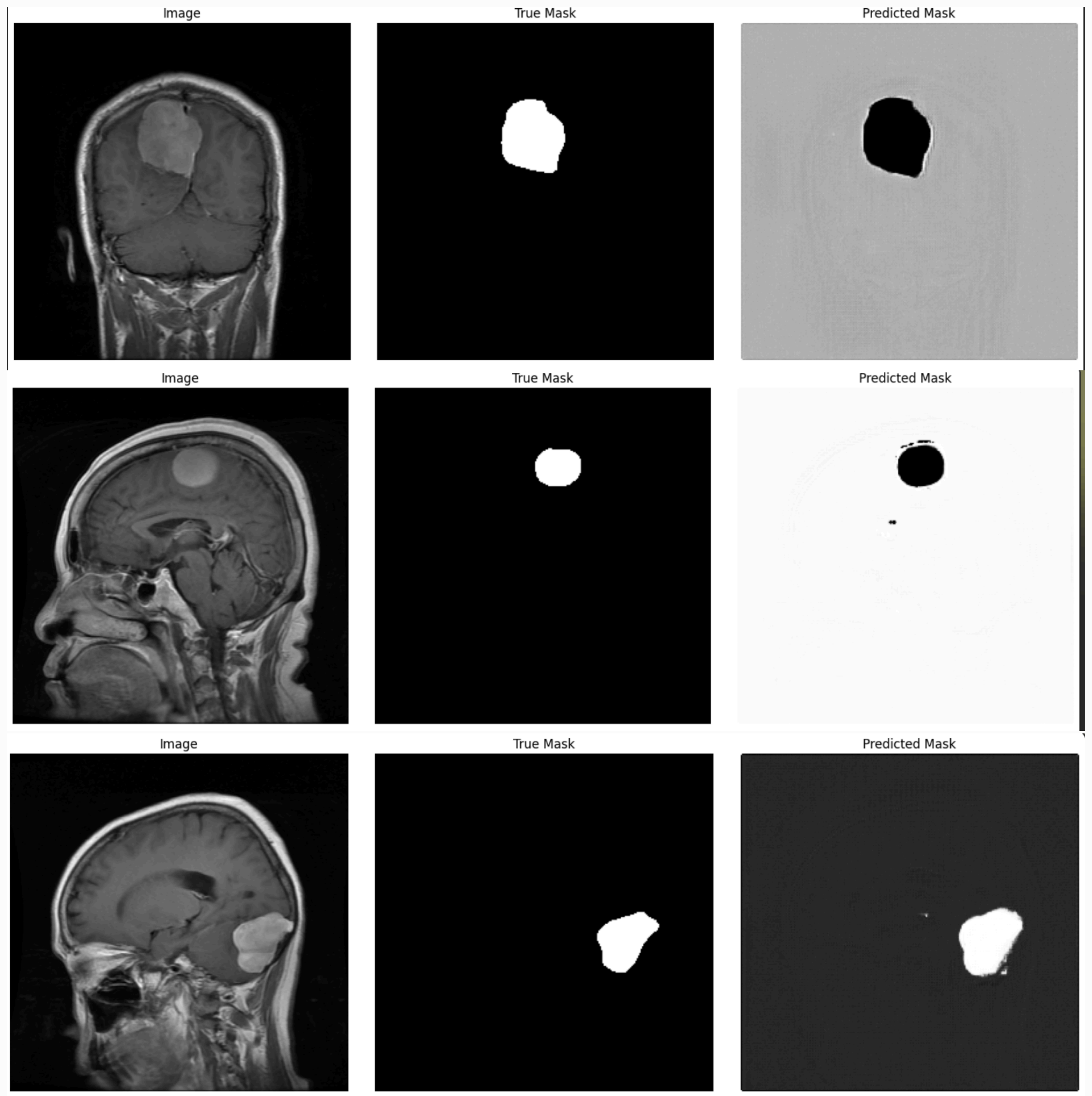
Qualitative Results: UNET



epochs max = 100
early stop = 46
patience = 20
balanced dataset
256x256 resolution
batch = 2



epochs max = 10
early stop = N/A
patience = 10
Imbalanced dataset
512x512 resolution
batch = 1



epochs max = 50
early stop = 17
patience = 10
balanced dataset
256x256 resolution
batch = 1

Comparing UNET Results to Prior Studies

Table 2: Comparison of Our U-Net Model Performance with Literature on BRaTS 2018

Method	Dice Comp	Dice Core	Dice Enh	Precision Comp	Precision Core	Precision Enh
Our U-Net (512x512, 100 epochs, subset)	3.18%	-	-	1.66%	-	-
Our U-Net (256x256, subset, 50 epochs)	1.66%	-	-	1.42%	-	-
Our U-Net (256x256, balanced, 50 epochs)	3.09%	-	-	1.61%	-	-
Our U-Net (256x256, balanced, 100 epochs)	2.97%	-	-	1.56%	-	-
3D UNet [225]	86%	82%	76%	7.01	5.63	5.60
Deep CNN [226]	88%	79%	78%	5.50	6.90	2.93
3D UNet [227]	87%	77%	71%	6.50	8.31	4.14
Contour-aware 3D CNN [228]	89%	79%	72%	8.05	7.50	5.20
S3D-UNet [229]	84%	78%	69%	9.20	7.70	4.50
HITU-Net [230]	88%	89%	82%	7.53	8.81	4.43
Auto-encoder Regularization [231]	88%	81%	77%	5.90	4.80	3.80
Cascaded UNet [232]	88%	78%	72%	-	-	-
CNN [233]	87%	77%	78%	6.55	27.05	15.90
CNN+Test-time Augmentation [234]	88%	80%	75%	5.97	6.71	4.16

Discussion

Oversegmentation

High recall suggests the model is sensitive to tumor presence but may over-segment, capturing more than the tumor region.

Performance Gap

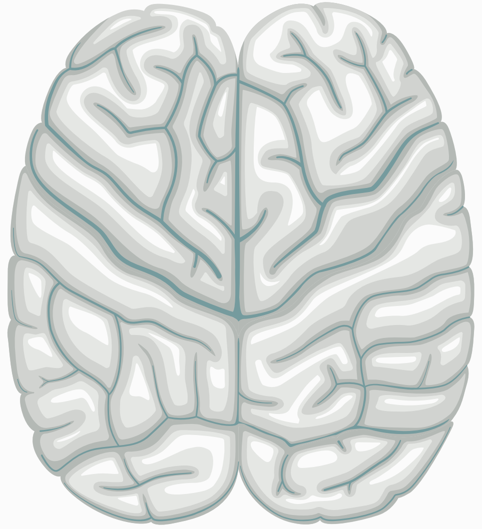
E.g. U-Net Dice Coefficient significantly lags behind the literature benchmarks, indicating a need for model and training process evaluation

Balanced data

The relatively better performance of our balanced dataset runs suggests that addressing class imbalance is crucial for improving our model's segmentation ability.

Further training/complexity

Model showed early stopping in some cases, which might have prevented it from fully converging to the best possible performance.



Metrics Per Tumor Type

- measuring performance per tumor type, per model

Using only High Quality Image

- capacity to use only 512x512 or above

Overcoming Computational Resources

- connecting to Cluster's NVIDIA GPU

Flag unlabeled or unrecognized tumors

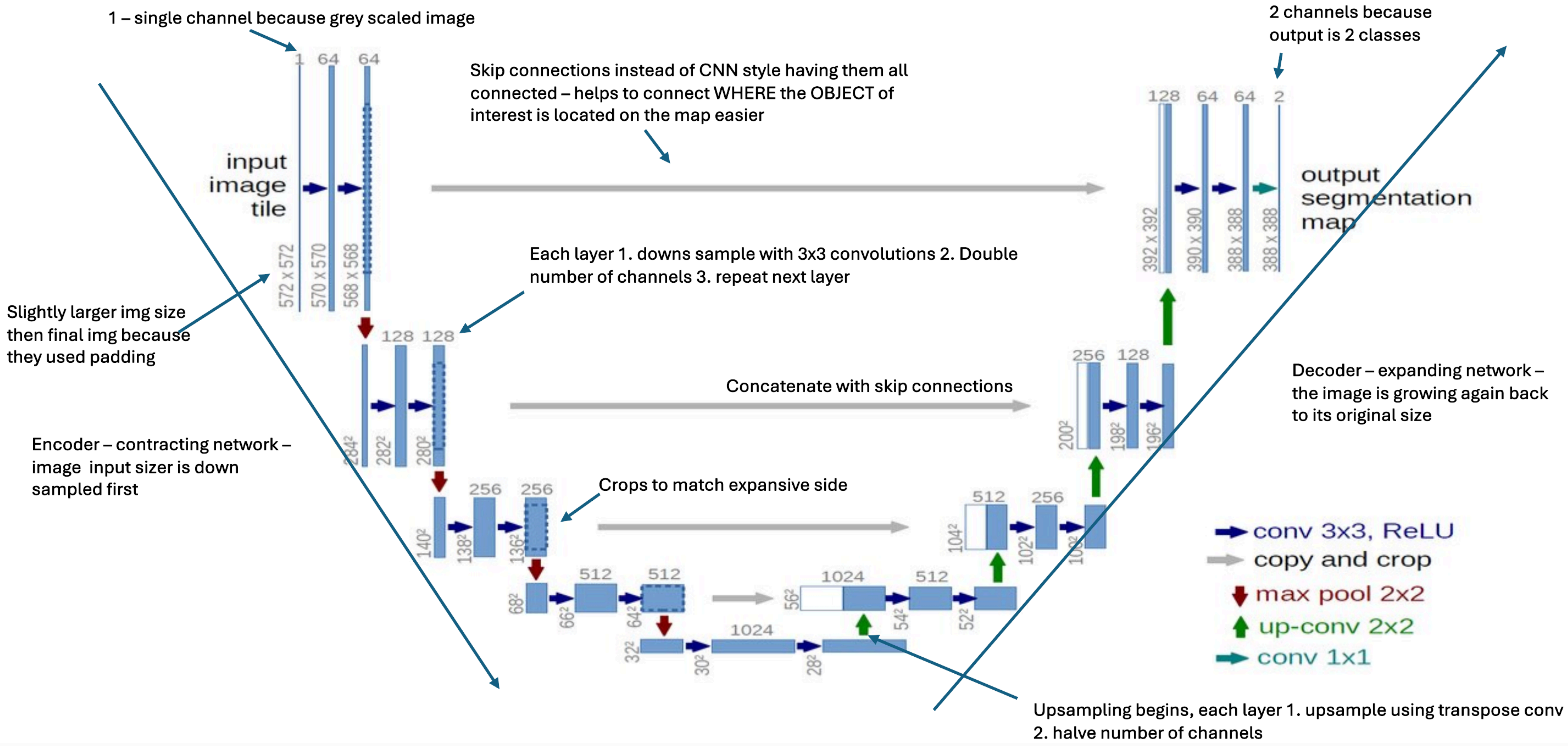
- continue addressing label issue with varied datasets

Future Work

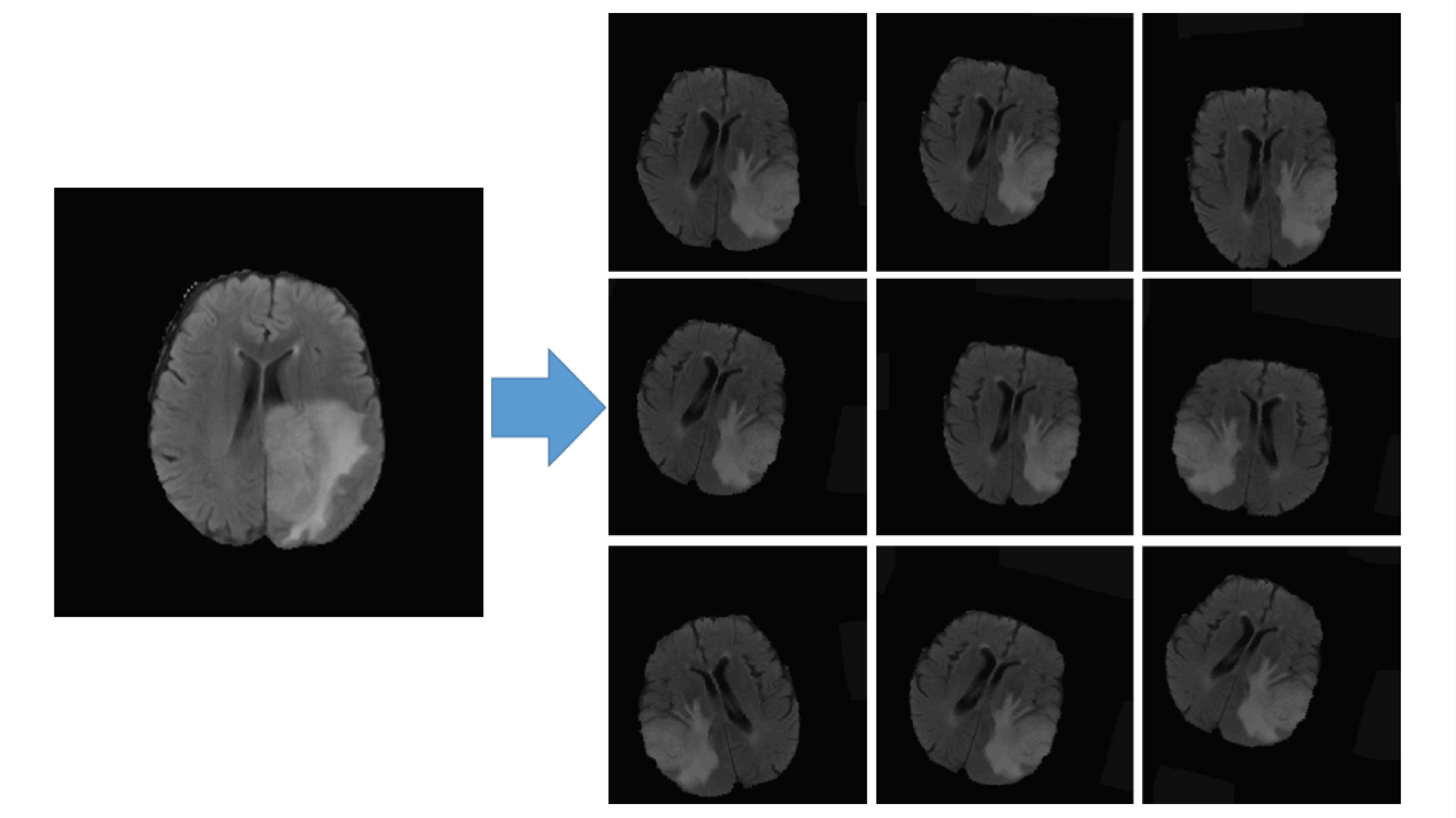
[1] Magnetic resonance imaging (mri).
[2] Jun Cheng. Brain tumor dataset. figshare. Dataset, 1512427(5), 2017.
[3] Jun Cheng, Wei Huang, Shuangliang Cao, Ru Yang, Wei Yang, Zhaoqiang Yun, Zhijian Wang, and Qianjin Feng. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLOS ONE, 10(10):e0140381, October 2015.
[4] Niharika Das and Sujoy Das. Epoch and accuracy based empirical study for cardiac mri segmentation using deep learning technique. PeerJ, 11:e14939, 2023.
[5] Guang Feng and Chong Tang. Portrait semantic segmentation method based on dual modal information complementarity. Applied Sciences, 14(4):1439, 2024.
[6] Chris Giordano, Meghan Brennan, Basma Mohamed, Parisa Rashidi, Franc¸ois Modave, and Patrick Tighe. Accessing artificial intelligence for clinical decision-making. Frontiers in digital health, 3:645232, 2021.
[7] Hong Huang, Fanzhi Meng, Shaohua Zhou, Feng Jiang, and Gunasekaran Manogaran. Brain image segmentation based on fcm clustering algorithm and rough set. IEEE Access, 7:12386–12396, 2019.
[8] Wen-Liang Hung, Miin-Shen Yang, and De-Hua Chen. Parameter selection for suppressed fuzzy c-means with an application to mri segmentation. Pattern Recognition Letters, 27(5):424–438, 2006.
[9] Haichun Li, Ao Li, and Minghui Wang. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. Computers in Biology and Medicine, 108:150–160, May 2019.
[10] Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. Knowledge and Information Systems, 64(12):3197–3234, 2022.
11] Benjamin Maas, Erfan Zabeh, and Soroush Arabshahi. Quicktumornet: Fast automatic multi-class segmentation of brain tumors. In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), pages 81–85, 2021.
[12] Mamta Mittal, Lalit Mohan Goyal, Sumit Kaur, Iqbaldeep Kaur, Amit Verma, and D. Jude Hemanth. Deep learning based enhanced tumor segmentation approach for mr brain images. Applied Soft Computing, 78:346–354, May 2019.
[13] Geethu Mohan and M. Monica Subashini. Mri based medical image analysis: Survey on brain tumor grade classification. Biomedical Signal Processing and Control, 39:139–161, January 2018.
[14] Samina Naz, Hammad Majeed, and Humayun Irshad. Image segmentation using fuzzy clustering: A survey. In 2010 6th international conference on emerging technologies (ICET), pages 181–186. IEEE, 2010.
[15] Wynton B. Overcast, Korbin M. Davis, Chang Y. Ho, Gary D. Hutchins, Mark A. Green, Brian D. Graner, and Michael C. Veronesi. Advanced imaging techniques for neuro-oncologic tumor diagnosis, with an emphasis on pet-mri imaging of malignant brain tumors. Current Oncology Reports, 23(3):34, February 2021
16] Ramin Ranjbarzadeh, Annalina Caputo, Erfan Babaee Tirkolaee, Saeid Jafarzadeh Ghoushchi, and Malika Bendechache. Brain tumor segmentation of mri images: A comprehensive review on the application of artificial intelligence tools. Computers in Biology and Medicine, 152:106405, January 2023.
[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pages 234–241. Springer, 2015.
[18] S. Srakocic. Can an mri detect cancer? what it can and can't detect, Mar 2023.
[19] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5):820–838, 2021.

Supplementary Slides

UNET Architecture



Data Augmentation



- Rotation, scaling, or elastic deformations might introduce unrealistic changes in medical images
- Recognizing augmented features rather than true pathological features
- Over-application -> loss of important clinical details
- Misalign labels with the images

FCM Methods

Fuzziness parameter > 1

- allows for "softer clustering" -> data points can belong to **multiple clusters** with varying degrees of membership
- Name comes from here – fuzzier boundary? -> data belongs to multiple clusters

As fuzziness approaches 1

- tends for "hard clustering" e.g. k-means -> each data point is assigned to **only one cluster**

- ***The FCM objective function quantifies how good a particular clustering solution is***
- ***Goal is to minimize the objective function via adjusting cluster centers and membership degrees for each point***

$$O_f = \sum_{i=1}^{n} \sum_{j=1}^{c=3} U_{ij}^{m} \left\| x_i - y_j \right\|^2$$

n	number of data points
c	number of clusters
O_f	objective function to be minimized
x_i	i-th data point
y_j	centroid of the j-th cluster
u_{ij}	degree of membership of x_i in the j-th cluster
m	weighting exponent on each fuzzy membership
$\|\|x_i - y_j\|\|$	Euclidean distance between the data point x_i and the cluster center y_j

Active Contour Methods

applies active contour model on each slice of the 2D image

- method for detecting and delineating object boundaries
- refines the contour output further by filling holes
- a low length penalty = risk of overfitting to noise/irrelevant details from the scan