# (Redone) Project Proposal
# A Comparative Study of Machine Learning and Deep Learning Multi-Class Segmentation Methods

Elizabeth Nemeti Shaniah Reece

## I. Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique [1], [18], [15]. Its depth and detail make it the most common and well-established imaging technique in brain tumor diagnosis. Despite this benefit, manual annotation or "segmentation" of tumor regions by radiologists is a laborious and time-consuming task [16]. Due to the sheer volume of the image sets, unclear morphology of brain tumor regions, and small size of early-stage tumors, manual segmentation can produce inaccurate results [9]. Additionally, for a patient's further treatment options, quantification of the tumor area is needed [13]. Hence, automated techniques are sought after for accurate diagnosis.

Automatic segmentation computationally categorizes pixels to distinguish between brain tissue, image background, and tumors [12]. It can be accomplished using machine learning and deep learning methods to aid in the fast and accurate detection of these tumors. These methods have the advantage of exploiting large datasets to learn the patterns and characteristics indicative of tumors, making them more feasible than manual approaches [11]. While they can exhibit high accuracy, they are still liable to generating inaccurate conclusions due to issues with generalization across diverse tumor types [10], variability in imaging protocols, and the requirement for annotated data. Thus, refinement and evaluation are necessary for effective segmentation techniques [11].

While deep learning models often outperform machine learning models in brain tumor segmentation, these results are not enough to exclude machine learning models [6]. For example, while research shows that deep learning models can be effective for complex image segmentation tasks, there are still instances when machine learning models can be useful. These include working with small datasets, enhancing interpretability and computational efficiency, and handling noise and outliers [6].

### A. Novelty and Justification

Profiling the state of the art for segmentation reveals a strong shift towards deep learning methods, as evidenced by the significant increase in publication rates illustrated in Figure 2 over the last few years. While deep learning methods consistently outperform machine learning methods, they are hindered by the requirement for large annotated datasets, interpretability issues, and substantial computational resources. In a 2021 study utilizing our selected dataset [2] (see also Section I-C), Maas et al. developed a U-CNN, an automatic multiclass segmentation tool [11]. Although their model yielded promising results, they identified the following areas for future work: 1) conducting a comparison of the network model to simpler architectures to enhance understanding of the underlying information extracted from medical images, and 2) employing datasets covering a wider variety of samples and instances of rare occurrences, as well as more diversified cohorts.

We intend to build upon their work and address these specific gaps by undertaking the following aims:

**Aim 1**: 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.

**Aim 2**: 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.

**Aim 3**: To address the labeling challenge in medical imaging, where labels are often sparse, noisy, or inconsistent [19], we will:

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

These aims are specifically designed to address and expand on the areas of future work outlined by Maas et al. [11] and Zhou et al. [19], ensuring that our research contributes effectively to filling these identified gaps.

### B. Lit Review and Model Selection

Segmentation in medical image processing can be employed using manual, semi-automatic, or fully automatic approaches. However, automation has become a necessity as radiologists require efficient techniques for accurate diagnosis using large data sets [9]. Constant improvement of medical image segmentation remains at the forefront [12]. In recent years, applications of DL approaches have been significantly greater than ML techniques [10]. We choose to compare an SVM and CNN due to their state-of-the-art nature and strong prominence in current publications that explore our same task of brain tumor segmentation. As seen in Figures 1 and 2 from Ranjbarzadeh et al.'s 2019 comprehensive review [16], SVM and CNN are the most widely used ML and DL tools for brain tumor segmentation as of late 2022. Their popularity stems from excellent performance; for example, SVM indicated high accuracy rates over 90 percent of the time. Although FCM was only the fifth most popular approach, we also include it in our research due to its implementation of fuzzy membership, making it particularly useful for pixel classification and hence image segmentation [12]. Including FCM also allows us to compare an unsupervised approach with our two supervised approaches,

SVM and CNN. However, as FCM is only a clustering method, to obtain a segmentation outline, we will incorporate an Active Contour (AC) algorithm. To select our CNN, we opted for "one of the key contributions that emerged from the medical imaging community, the U-NET architecture" [9], developed specifically for biomedical image segmentation and proven to robustly execute segmentation tasks [17].

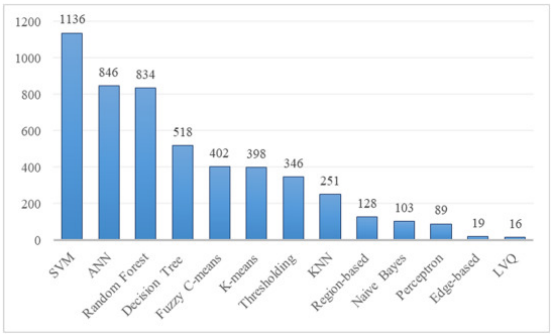


Fig. 1. Number of publications between 2015 and 2022 for brain tumor segmentation using supervised and unsupervised learning techniques. [16]

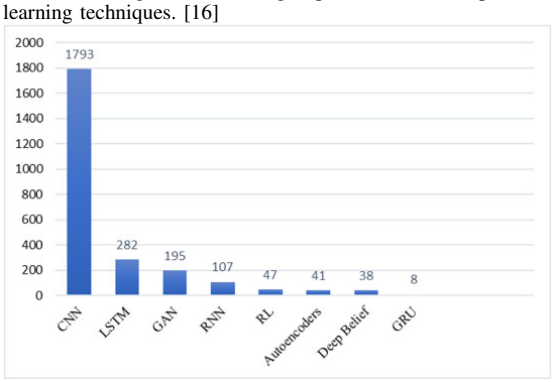


Fig. 2. Number of publications between 2015 and 2022 for brain tumor segmentation using DL models. [16]

### C. Selected Dataset: Brain Tumor Classification MRI Images

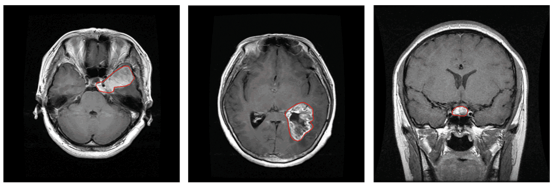


Fig. 3. Illustrations of three typical brain tumors: (a) meningioma; (b) glioma; and (c) pituitary tumor. Red lines indicate the tumor border. [3]

We have selected a comprehensive brain tumor dataset acquired from Nanfang Hospital in Guangzhou and General Hospital at Tianjin Medical University, China, between 2005 and 2010. Patient records and information are anonymized and de-identified. This dataset consists of 3064 T1-weighted contrast-enhanced MRI slices from 233 patients, encompassing three types of brain tumors: 708 meningiomas, 1426 gliomas, and 930 pituitary tumors. We will utilize the complete dataset in our study. The images boast an in-plane resolution of 512×512 with a pixel size of 0.49×0.49 mm², a slice thickness of 6 mm, and a slice gap of 1 mm. Each tumor border was delineated by three radiologists. This is how, in terms of image processing, the "ground truth annotation masks" or, in terms of machine learning, the "labels" were acquired. Organized in MATLAB's .mat format, each file contains structured data with several fields: cjdata.label for tumor type, cjdata.PID for patient ID, cjdata.image for image data, cjdata.tumorBorder for coordinates of the tumor borders, and cjdata.tumorMask for the binary tumor areas. Figure 3 contains an example of each tumor type with its ground truth annotation from the dataset. As such, the dataset is primed for robust analysis and the segmentation task. For benchmarking, we will be using the dataset from the well-recognized BraTS 2018 Multimodal Brain Tumor Segmentation Challenge, which includes training and validation data.

**Link to Dataset on Figshare**:
https://figshare.com/articles/dataset/brain_tumor_dataset/1512427?file=7953679

**Link to Benchmark Dataset**:
https://www.med.upenn.edu/sbia/brats2018/registration.html

## II. Methods Plan

### A. Proposed Methods

Our two supervised models UNET and SVM will follow the high-level pipeline illustrated in the Figure 4.

The UNET architecture (see Figure 5) will comprise an encoder, bottleneck, decoder, output layer, and take the 512x512 MRI images as input. The encoder, serving as the contracting path, will include 4 blocks of x2 convolutional layers with ReLU activation, batch normalization, and max-pooling operation to downsample the feature maps. Dropouts will be utilized to prevent overfitting, with a rate of 0.3 following max-pooling. The bottleneck will consist of x2 convolutional layers without pooling layers. The decoder will constitute the expansive path, defined by 4 blocks containing a transposed convolution layer, a skip connection, and two standard convolutions. Dropout will be included again at the same rate to aid learning. The model will be trained for 50, 100, 150, and 200 epochs, with the best estimated performance anticipated at 150, as explored in Das and Das 2023 [4].

The SVM will incorporate feature scaling and PCA for dimension reduction. We will conduct a kernel search to compare linear, RBF (Gaussian), and sigmoid kernels to determine the best parameters using random search or grid search (if time permits). Parameters will include the SVM kernel, controlling regularization strength, and others.

The FCM+AC architecture will involve parameters such as fuzziness, number of clusters, window size, length penalty, iterations, and epsilon. Fuzziness controls membership allowance, clusters indicate the regions to be segmented, window size represents the area around each pixel that AC considers when adjusting contours, length penalty penalizes the contour length making it more jagged/smooth, iterations denote the number of iterations AC is allowed, and epsilon serves as the convergence criterion [8]. Initially, these parameters will be set

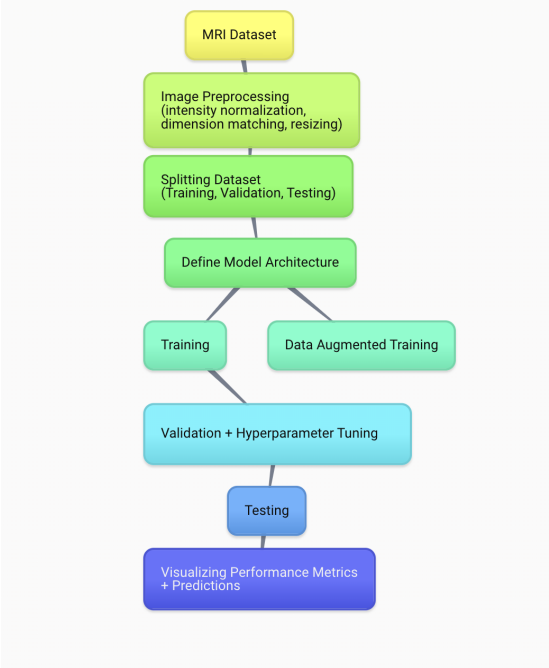


Fig. 4. Illustration of our high-level ML/DL pipeline. (Data augmentation should also connect to validation) (our figure)

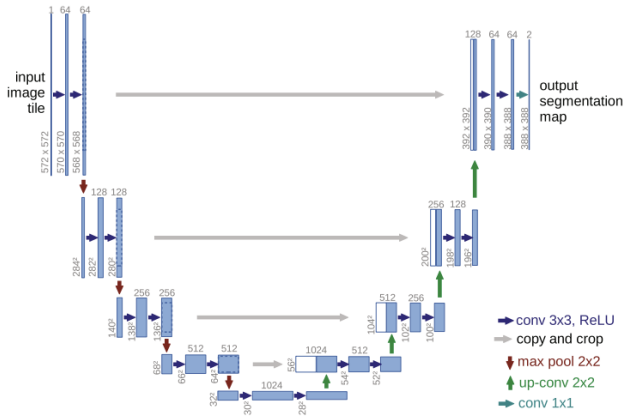


Fig. 5. The network structure of U-net from Feng and Tang 2024 [5]

to 3, 3, 7, 0.000001, 400, and 0.3 respectively before tuning. As it is a clustering method, it will be applied to the entire dataset without splitting it into training and testing first [7].

The UNET will be validated using a Single validation split strategy (80:20) to reduce computation time and resources, whereas K-fold validation will be more feasible for the SVM. Since FCM+AC is unsupervised, K-fold validation cannot be utilized as there is no ground truth for comparison. However, the dataset can still be manually split into subsets to evaluate performance and consistency across subsets.

### *B. Evaluation Metrics and Success Criteria*

The performance assessment of each of our three segmentation strategies can be achieved using various metrics. The most popular and widely used performance measures for brain tumor segmentation tasks include Sensitivity (Recall), Specificity, Accuracy, Precision, the Confusion matrix, Jaccard Index (IoU), and Dice Similarity [16]. We will evaluate our chosen models with their appropriate metrics as indicated in recent publications:

TABLE I
SELECTED METRICS PER MODEL

| **UNET** | **SVM** | **FCM+AC** |
|---|---|---|
| Sensitivity | Sensitivity | Silhouette Score |
| Specificity | Specificity | Davies-Bouldin |
| Accuracy | Accuracy | Calinski-Harabasz |
| Precision | Precision | |
| Confusion Matrix | Confusion Matrix | |
| Jaccard Index (IoU) | Jaccard Index (IoU) | |
| Dice Similarity | Dice Similarity | |
| | F1 Score | |

The Dice and Jaccard indices are most appropriate for the UNET [19], as these metrics measure the overlap in pixels of the predicted masks and the ground truth annotations. Although similar to Dice, SVM will use the F1 score as it is primarily a classification metric and SVM is primarily a classifier [**?**]. However, since we have engineered the SVM to segment by classifying pixels, it will also share the Dice and IoU metrics, as well as Sensitivity, Specificity, Accuracy, and Precision. Finally, FCM+AC must be effectively evaluated with metrics focusing on clustering such as Silhouette Score, which indicates correct assignment, Davies-Bouldin Index, indicating similarity between clusters, and Calinski-Harabasz index, measuring the ratio of inter- to intra-cluster dispersion [14].

The shared success criteria across the models will be evaluated in three stages. First, visual comparison will serve as a qualitative measure for the final predicted masks compared to the ground truth for each model. Second, we will compare the models according to their shared metrics. For the two supervised models, we can directly compare their sensitivity, specificity, accuracy, prediction scores, the confusion matrix, Dice, and IoU. Since FCM is unsupervised and operates with clustering, it has metrics that are useful for its own evaluation, as shown in Table 1, but not directly comparable to those of the supervised models. For a direct comparison, we must additionally map the final clusters to ground truth annotations from our dataset.

## III. PITFALLS AND MITIGATION STRATEGIES

The U-NET architecture, while powerful, is prone to overfitting, particularly when dealing with small to moderately sized datasets. To counter this, we will implement dropout layers and regularization techniques to enhance generalization across various data subsets. Moreover, to address U-NET's sensitivity to noise, we will enhance the preprocessing stage. Additionally, considering the high memory requirements, we will adopt separable convolutions to reduce memory demands. Fine-tuning hyperparameters suitable to the hardware, adjusting batch size, or downsizing image dimensions are alternative strategies to manage memory usage.

The SVM algorithm may encounter challenges with nonlinear separability. In response, we will explore different

kernel tricks such as RBF, sigmoid, or polynomial kernels to transform the data into linearly separable spaces. Furthermore, scalability issues arising from high-dimensional data will be mitigated by applying PCA to reduce dimensions and improve training efficiency. Addressing class imbalances is crucial, and we will ensure an equal number of samples per class, given the slight disproportionality in the full dataset.

FCM's accuracy may be compromised by noise and outliers, necessitating robust preprocessing techniques for effective clustering. To mitigate the risk of falling into local minima, we will run the FCM algorithm multiple times with varying initializations and hyperparameters, utilizing metrics like Calinski-Harabasz to optimize outcomes. To handle class imbalances, we will consider manual class balancing or adjust FCM's sensitivity to cluster size variations. Determining the optimal cluster count is another challenge, which we plan to tackle by leveraging metrics like the silhouette score or Davies-Bouldin index.

## REFERENCES


[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, François 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.