Ministère de l'Enseignement Supérieur et de la Recherche Scientifique Université de Carthage Institut National des Sciences Appliquées et de Technologie

Comparative Study of Brain Metastasis Segmentation Datasets and Models

Brain Metastases

Réalisé par:

Mariem Makni

Contents

Introduction

Brain metastases (BMs) are the most prevalent form of intracranial tumors in adults, significantly impacting the lives of cancer patients. An estimated 98,000 to 170,000 cases occur annually in the United States alone, making BMs a frequent complication of cancer and a common type of brain tumor [1]. Studies indicate that 20% to 40% of patients diagnosed with a primary cancer will develop secondary tumors in the brain, marking it as a major site for metastatic spread [2].

The management of brain metastases poses substantial challenges in clinical settings, particularly due to the complexity of their treatment which includes radiotherapy, surgery, and advanced drug therapies. The advancements in imaging techniques, particularly MRI, play a crucial role in the diagnosis and management of BMs. However, the manual tasks involved in the diagnosis and treatment planning, such as lesion identification and segmentation, are time-consuming and susceptible to human error.

To address these challenges, machine learning (ML) and deep learning (DL) techniques are being developed to automate these critical tasks, potentially increasing accuracy and efficiency in clinical workflows. Nonetheless, the progress in this field is hampered by the scarcity of publicly available datasets specifically curated for brain metastases, which is crucial for training and validating these AI models.

Chapter 1

Datasets for Brain Metastasis Segmentation

Table 1.1 presents a comparison of several pivotal datasets used in brain metastasis segmentation research. These datasets encompass a diverse range of imaging techniques and clinical data sourced from multiple institutions, enhancing the robustness and applicability of research outcomes.

The "Brain metastases MRI dataset" [3], "BrainMetShare" [4], and "UCSF-BMSR" [5] provide extensive MRI data along with validated segmentations, which are crucial for the development of segmentation algorithms. These datasets are openly accessible, and I have successfully downloaded them.

The "Pretreat-MetsToBrain-Masks" dataset, while publicly available, has presented unique challenges due to issues with the IBM Aspera download protocol. Nevertheless, it offers a comprehensive array of clinical and imaging data, including detailed segmentations of both major and subtle lesions.

The "BraTS 2023" [6] dataset provides MRI studies with semi-automatic and expert-drawn segmentations. However, accessing this dataset **requires official permission**. Similarly, the "BraTS 2024" [7] dataset can only be accessed by **participating in the challenge**.

BraTS 2024 challenge Deadlines:

- Mid-June to July 31: Validation phase (Segmentation file submissions). The exact start date of submissions will depend on the challenge (as it includes several challenges).
- July 31: Hard deadline for the submission of short papers in CMT, reporting methods and results based on training and validation data.

Table 1.1: Comparison of Brain Metastasis Segmentation Datasets

Dataset Name	**Year**	**Source**	**Number of Patients**	**Access Status**	**Key Features**
Brain metastases MRI dataset	2023	Scientific Data	75	Open Access	Comprehensive clinical and radiomic data, 637 high-resolution MRIs, semi-automatic segmentations
BrainMetShare	2023	Stanford	156	Open Access	156 whole brain MRIs, multi-modal sequences, radiologist-validated segmentations
BraTS 2023 Challenge Data	2023	MICCAI BraTS	Multiple	Requires Permission	Multi-parametric MRI, advanced segmentations, focus on varying lesion sizes
BraTS 2024 Challenge Data	2024	MICCAI BraTS	Multiple	Requires Participation	Expanded dataset, includes additional external datasets, refined segmentations by neuroradiologists
Pretreat-MetsToBrain-Masks	2023	IBM-Aspera-Connect	200	Public, issues with Aspera IBM	200 patients, clinical and qualitative/quantitative imaging information, sub-centimeter lesions, manual 3D segmentations
UCSF-BMSR	2023	The University of California San Francisco	412	Public, UCSF provided, non-commercial usage	Multimodal brain MRI, expert annotations, detailed segmentation

Chapter 2

Selection and Compliance of Dataset

Numerous datasets exist in the field of Brain Metastases. However, integrating these datasets to ensure compatibility and uniformity across segments and labels is resource-intensive and challenging. In this context, the BraTS METS datasets, which aggregate data from various sources into a unified format, emerge as the most suitable options. They guarantee consistency and compatibility, making them optimal for comprehensive research.

2.1 BraTS METS 2023

The BraTS METS 2023 dataset features a robust selection of cases from prestigious institutions, all of which underwent rigorous annotation and vetting processes. Contributors include NYU, NCI, Duke, WashU, and Yale University. The dataset details are as follows:

- **NYU**: 164 cases in total, consisting of 107 cases finalized without corrections, 57 cases needing corrections, 46 cases excluded due to absence of lesions or incomplete scans, and 22 cases excluded post-treatment.
- **NCI**: 32 cases in total, with 21 finalized without corrections, 11 needing corrections, and 1 excluded due to absence of lesions.
- **Duke**: 24 cases, all finalized with 14 without corrections and 10 needing corrections.
- **WashU**: 28 cases in total, including 16 finalized without corrections, 12 needing corrections, and 1 excluded post-treatment.
- **Yale**: 80 cases, all finalized with 44 without corrections and 36 needing corrections.

2.2 BraTS METS 2024

The BraTS METS 2024 dataset builds on the 2023 dataset by including post-treatment changes and other instances. It offers a comprehensive collection of pre-treatment brain metastases multi-parametric magnetic resonance imaging (mpMRI) scans, ensuring data uniformity crucial for developing and validating advanced machine learning models. New external sources include:

- 164 **NYU** cases with **all 4 sequences in BraTS format**, requested directly from NYU (data agreement reviewed and confirmed to be compatible with our research use).
- **UCSF-BMSR** MRI Dataset with full BraTS format annotations on 324 cases, **available as Training Dataset 2**. Additional cases lacking the full format are not utilized due to compatibility requirements.

2.3 Compliance with Data Use Agreement

Evaluating the model's ability to generalize is a crucial aspect of research, particularly in the medical field where it has garnered significant attention. This evaluation typically involves using train/test splits from the initial dataset and subsequently testing the model on additional datasets. Such an approach helps ensure that the model generalizes well across various institutions and equipment, accounting for differences in data quality. The integration of BraTS datasets with other data sources for such research purposes is permissible, subject to specific conditions as outlined in the respective data use agreements [8,9]:

- **Explicit Documentation**: Clearly document the use of the BraTS dataset alongside other datasets in research methodologies. Specify the contributions of each dataset to the training, validation, or testing phases of the study.
- **Citing and Acknowledgments**: Ensure that all datasets, including BraTS, are properly cited according to the official challenge guidelines. General acknowledgments should be made if the datasets were used as part of a challenge or personal research project.
- **Compliance with Embargo and Publication Policies**: Adhere to embargo periods and publication restrictions specified by the BraTS organizers. For the 2024 dataset, this involves waiting until the official release of the joint overview paper before publishing comprehensive challenge results. The 2023 dataset is already available [10] for immediate academic dissemination.

Chapter 3

Models for Brain Metastasis Segmentation

While further study is essential to provide a thorough comparison and detailed analysis of the most effective models for brain metastasis segmentation, a review of current benchmarks [11–13] highlights several models. In my future research, I am particularly interested in exploring lesser-known or innovative models, including those based on transformer architectures, which have yet to be widely applied in this domain.

Model Descriptions:

- **U-Net:** Known for its symmetric architecture, U-Net is originally designed for biomedical image segmentation and excels in precise localization.
- **SynthSeg:** This model is robust to variations in imaging conditions, segmenting brain scans across different modalities without retraining, thanks to its domain randomization technique.
- **OM-Net + CGAp:** Enhances feature extraction and focuses attention on relevant areas through its Context Gating Attention Pooling mechanism, improving segmentation accuracy.
- **Semantic Genesis:** Uses self-supervised learning from unlabeled 3D medical images for pre-training, which can be fine-tuned for specific tasks, enhancing adaptability and efficiency.
- **NVDLMED:** Optimized for medical image segmentation, incorporating various network design innovations to improve performance on volumetric data.
- **Cascaded Anisotropic CNNs:** Processes images at multiple scales with a cascading architecture, effectively detecting and segmenting small or irregularly shaped lesions.

3.1 Transformer-Based Models in Medical Image Segmentation

Recent advancements in machine learning have introduced transformer-based models as a promising approach for medical image segmentation, particularly for 3D brain metastases. These models offer significant improvements in capturing global contextual information compared to traditional Convolutional Neural Networks (CNNs). Despite their potential, transformer models typically require substantial computational resources and large datasets to achieve optimal performance. This section briefly discusses two notable transformer-based models adapted for 3D medical image segmentation:

SegFormer3D

Introduced by Perera et al., SegFormer3D is an efficient transformer designed for 3D medical image segmentation. It addresses the challenges of large-scale transformer models by employing a hierarchical structure that processes multi-scale volumetric features with significantly reduced computational demands. SegFormer3D utilizes an all-MLP (Multi-Layer Perceptron) decoder to aggregate local and global attention features efficiently, producing highly accurate segmentation masks with fewer parameters and reduced computational complexity. This model demonstrates the feasibility of lightweight transformers in medical imaging, providing a balance between performance and resource efficiency [14].

Swin Unet3D

Developed by Cai et al., Swin Unet3D combines the robust feature extraction capabilities of Vision Transformers with traditional convolutional techniques. This model is specifically designed to handle the segmentation of 3D medical images by employing a hybrid architecture that learns both global and local dependencies effectively. Swin Unet3D integrates a parallel structure of convolution and transformer modules, allowing it to capture intricate details necessary for precise tumor segmentation. The model has been validated on several benchmark datasets, showing competitive performance and offering a new paradigm in medical image analysis [15].

Bibliography

[1] "National center for biotechnology information." https://www.ncbi.nlm.nih.gov/books/NBK470246/, 2023.

[2] "Medscape." https://emedicine.medscape.com/article/1157902-overview, 2022.

[3] B. Ocaña-Tienda *et al.*, "A comprehensive dataset of annotated brain metastasis mr images with clinical and radiomic data," *Scientific Data*, vol. 10, p. 208, 2023.

[4] "Brainmetshare dataset." https://stanfordaimi.azurewebsites.net/datasets/ae0182f1-d5b6-451a-8177-d1f39f019962, 2021.

[5] "The university of california san francisco brain metastases stereotactic radiosurgery mri dataset." https://imagingdatasets.ucsf.edu/dataset/1, 2023.

[6] "Brats 2023: Adult glioma." https://www.med.upenn.edu/sbia/brats2023/data.html, 2023.

[7] "Brats 2024 challenge data." https://www.synapse.org/Synapse:syn53708249/wiki/627504, 2024.

[8] "Data conditions for use, brats 2023." https://www.synapse.org/Synapse:syn51156910/wiki/622360.

[9] "Data conditions for use, brats 2024." https://www.synapse.org/Synapse:syn53708249/wiki/627508.

[10] "The brain tumor segmentation (brats-mets) challenge 2023: Brain metastasis segmentation on pre-treatment mri." https://arxiv.org/pdf/2306.00838, 2023.

[11] "Brain tumor segmentation benchmarks, papers with code." https://paperswithcode.com/task/brain-tumor-segmentation.

[12] "Brain segmentation on brain mri segmentation, papers with code." https://paperswithcode.com/sota/brain-segmentation-on-brain-mri-segmentation.

[13] "Tumor segmentation, papers with code." https://paperswithcode.com/task/tumor-segmentation.

[14] "Shehan perera, pouyan navard, alper yilmaz, "segformer3d: an efficient transformer for 3d medical image segmentation." https://arxiv.org/abs/2404.10156, 2024.

[15] "Yimin cai, yuqing long, zhenggong han, et al., "swin unet3d: a three-dimensional medical image segmentation network combining vision transformer and convolution, bmc medical informatics and decision making." https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02129-z, 2023.