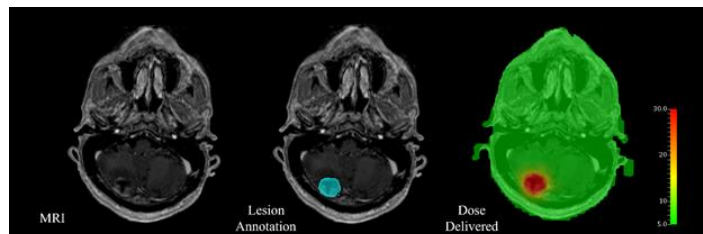


Brain-TR-GammaKnife | Brain Tumor Recurrence Prediction after Gamma Knife Radiotherapy from MRI and Related DICOM-RT: An Open Annotated Dataset and Baseline Algorithm

Summary

Here we release a brain cancer MRI dataset with the companion Gamma Knife treatment planning and follow-up data for the purpose of tumor recurrence prediction. The dataset consists of 47 subjects. A



A total of 244 lesions were collected with annotations. The dataset contains original patient MRI images (in DICOM format), radiation therapy structure data (in DICOM and NRRD format), code, and clinical information. First, dose MRI images were resampled to original MRI spacing via a linear transformation. Second, each region in each patient's MRI was extracted and cropped out; note that one patient may have multiple lesions and/or multiple imaging sessions. Third, the corresponding radiation dose information was cropped out to the resampled aligned lesion mask. In this way, each lesion MRI is paired with its radiation dose MRI. The release of this dataset is expected to contribute to the development of automated brain tumor recurrence prediction algorithms.

Introduction

Prediction and identification of tumor recurrence are critical for brain cancer treatment design and planning. Stereotactic radiation therapy delivered with Gamma Knife has been developed as one of the common treatment approaches in addition to others by delivering radiation that accurately to the tumor while minimizing toxicity to nearby healthy tissues. Artificial intelligence (AI)-based research has shown

great potential in brain tumor MRI analysis recently with its effective data-driven feature extraction and recognition capabilities. Here we release a brain cancer MRI dataset with the companion Gamma Knife treatment planning and follow-up data for the purpose of tumor recurrence prediction. The dataset contains original patient MRI images, radiation therapy data, and additional clinical information. The included MRI series are T1 MPRAGE with Gadolinium contrast acquired on a 1.5 T Siemens Magnetom scanner for Gamma Knife treatment for brain metastasis. This public dataset retrospectively involved a total of 47 brain cancer patients, organized based on patient-level, course-level, and lesion-level, respectively. Patient-level keeps all the data related to a specific patient altogether. Each course-level treatment data has been stored for each patient. Among all the patients, 17 patients received more than 1 course of treatment, 7 received more than 2 courses, and 1 patient received up to 8 courses in total. Regarding to the lesion-level, a total of 244 lesions are collected with annotations. 221 lesions are stable, and 23 are recurrence, which makes the dataset imbalance from the lesion-level perspective. Lesion annotations are provided, and inclusive preprocessing steps have been specified to simplify the usage of this dataset. The release of this dataset is expected to contribute to future development of automated brain tumor recurrence prediction algorithms and promote the clinical implementations associated with AI-based computer-aided diagnosis field.

Data Records and Population

This dataset was collected retrospectively under IRB-approval (2017-0266) from a clinical database of patients treated for brain metastases with Gamma Knife radiation therapy at the University of Mississippi Medical Center (UMMC). The dataset consisted of 47 brain cancer with 21 males and 26 females who developed brain metastases for various primary tumors and then underwent Gamma Knife stereotactic radiation therapy in an effort to control the identified brain lesions. A total of 244 lesions are collected with annotations. Additionally, these patients participated in treatment follow-up during which the response to treatment could be adequately assessed to identify recurring lesions and those that responded to treatment by stopping growth or even shrinking. A keyed spreadsheet stored various other relevant clinical information as well as how many treatment sessions were used to deliver the dose for each target. It should be noted that each patient may have undergone more than one treatment course as they developed new lesions and/or treated lesions recurred. For each patient, the treatment targets were identified and delineated through collaboration between radiation oncologists and neurosurgeons with validation from neuroradiologists.

Image Acquisition

In preparation for the dataset, an MRI was acquired on a 1.5 T Siemens Magnetom scanner. The included MRI series are T1 MPRAGE with Gadolinium contrast and a voxel size of $1 \times 1 \times 1$ mm which acts as the primary planning dataset for Gamma Knife treatment for brain metastases. For each patient, the treatment planning DICOM data was collected in the MRI dataset with its accompanying RTStruct, which indicates the lesions that were targeted using a given MRI. The RTDose information provides how much and where dose was deposited for each respective lesion/target. The treatment sessions were recorded to describe the dose delivery for a target represented in the RTDose as it may have been delivered in as few as 1, but maybe as many as 5 treatment sessions. Additionally, each patient may have undergone more than one treatment course as they developed new lesions and/or treated lesions recurred. The dose units are in Gray (Gy, AKA J/kg) which is typical in radiation therapy. The dose was calculated using the TMR10 algorithm in the GammaPlan® (version 11) software which does not consider tissue heterogeneity in the calculation but is consistent with previous Gamma Knife treatment practices across many institutions. Treatment was ultimately delivered on the Gamma Knife Icon® at the same institution. After collection, all DICOM data (MRI images, RTStruct, and RTDose) was fully anonymized removing all protected health information and treatment-related dates. The anonymization process included mapping them to identifiers that match the clinical information in the database spreadsheet.

Data Preprocessing and Usage

In this dataset, both original and preprocessed data are available for both clinical and engineering researchers. For preprocessed images, we specify a series of preprocessing steps that have been conducted. First, dose MRI images were resampled to the same spacing as the original patient MRI via a linear transformation process because the raw MRI and dose files share not only different dimensions but incompatible actual spacing. Second, each lesion region in each patient's MRI was extracted and cropped out based on the patient's lesion mask. It is noted that one patient may have multiple lesions. Third, the corresponding radiation dose information was also cropped out using the same lesion mask because the masks are shared through the same coordinate after resampling. In this way, each lesion MRI is paired with its radiation dose MRI. For any AI model training or test, the voxel values are recommended to be normalized to the range of 0 to 1 in each cropped lesion and dose MRI pair. Resizing is another necessary preprocessing step to keep

the cropped training and test samples in the same dimension. In our dataset, we kept the original dimension of the cropped lesions. Since the dataset is imbalanced in terms of lesion-level, data augmentation techniques are suggested to be employed before model training.

The aim of this dataset is to facilitate the development and evolution of automated tumor recurrence identification. Further potential applications based on this dataset include radiation treatment planning, evaluating and refining, and automated dose delivery planning. AI research community may take advantage of this public data resource to achieve promising results and promote the implementation of artificial intelligence-assist tumor recurrence prediction. Different preprocessing procedures in terms of various focuses and implementation can be used, and more advanced AI frameworks and techniques are expected to be developed.