SarcomaAI: Playbook for Establishing a Pipeline for Data Gathering, Processing, and Multimodal Federated Learning

Jonathan A. Hubermann, Anthony Bozzo

September 11, 2024

Version 0.1

Contents

Introduction

This document serves as a comprehensive guide for institutions participating in the SarcomaAI project, which aims to develop advanced machine learning multimodal models through federated learning to enhance the understanding and treatment of soft tissue sarcomas in orthopaedic oncology.

The document outlines the essential steps for establishing an effective data pipeline, beginning with the ethical considerations and the process of obtaining Research Ethics Board (REB) approval. It then addresses the technical aspects of data extraction, preprocessing, and storage, ensuring that data is de-identified, standardized, and securely managed. Key areas of focus include the transition from clinical imaging systems, such as PACS, to specialized imaging informatics platforms, the role of manual and automatic segmentation in preparing imaging data, and the significance of precise segmentation for the development of robust machine learning models.

The final section discusses the implementation of a deep learning multimodal model within a federated learning framework, which facilitates collaborative model training across multiple institutions while preserving the privacy of patient data. Each component of this guide is designed to integrate seamlessly, providing a clear and structured approach to establishing a secure and efficient pipeline for the SarcomaAI project, ultimately advancing personalized medicine in the treatment of sarcomas.

1 Obtaining REB Approval

Research Ethics Board (REB) approval is crucial for ensuring that the data collection process complies with institutional ethical standards and regulations. This step involves preparing and submitting the necessary documentation, addressing any concerns, and obtaining official approval before data collection begins. The information in this document outlining the SarcomaAI pipeline can serve as a helpful resource when drafting the REB proposal.

Institutional requirements on data safety, deidentification and storage may vary. However, our approved REB includes key details that help describe the project, data safety measures enacted.

We will share our approved REB to serve as a template for other institutions to use in their application. After adjusting for differing institutional requirements, this will help avoid common pitfalls and ultimately expedite the approval.

After REB approval is obtained, a *collaborative research agreement* will need to be signed between institutions. There is already one in place for this study between Memorial Sloan Kettering Cancer Center and McGill University, and we will share this as a template as well.

Information: Institutions may have differences in policies and protocols that are important to keep in mind when preparing to join the SarcomaAI project. Possible institutional differences include:

- Some institutions may have "internal deidentification services" for their image data, some do not.
- Differences in requirements for identifying patients and their clinical variables.
  - i.e., is an MRN de-identified enough or do anonymized study IDs need to be created.

As more institutions are onboarded, additional pearls and considerations will be added.

2 Data Extraction

Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are forms of volumetric data, meaning they represent three-dimensional structures. These images can be conceptualized as a series of two-dimensional slices that, when compiled, form a comprehensive 3D representation of an anatomical region. Due to this complexity, such data is not typically stored as simple image files within a folder. Instead, they are organized within specialized 3D volumetric data formats, the most widely used being Digital Imaging and Communications in Medicine (DICOM) and Neuroimaging Informatics Technology Initiative (NIfTI) formats. Other formats, such as NRRD (Nearly Raw Raster Data) and Analyze, are also utilized, though less commonly.

These formats are not merely containers for image data; they are robust frameworks that encapsulate extensive metadata, which provides crucial contextual information. This metadata can include patient demographics, acquisition parameters (such as the type of imaging modality and settings used), and spatial orientation data, which is essential for accurate interpretation and analysis of the images. The inclusion of such metadata is vital as it ensures the integrity and usability of the data for clinical and research purposes.

However, the hierarchical nature of these volumetric data formats, combined with their embedded metadata, makes direct navigation through a conventional file system cumbersome and

impractical. To address this, specialized systems such as Picture Archiving and Communication Systems (PACS) have been developed. These systems are designed to organize, store, and facilitate access to these complex datasets, ensuring that the valuable information they contain can be efficiently managed and retrieved.

The following subsections outline the various methods for exporting this data from an institution's systems and provide guidance on how to effectively store it, ensuring both the integrity of the data and its accessibility for subsequent processing and analysis.

2.1 PACS-to-PACS Data Transfer

Large institutions, such as health centers and universities, typically store their CT and MRI scans using PACS. These systems are essential for managing, storing, and retrieving medical images, providing an organized framework for handling large volumes of complex data. Prominent examples of PACS software include GE Healthcare's Centricity, Philips IntelliSpace, Sectra PACS, and Intelerad's IntelePACS, which are widely recognized and used across healthcare institutions.

A common and efficient method of extracting and transferring medical imaging data involves direct PACS-to-PACS transfer. In this approach, the scans stored in the institution's PACS are transferred directly to a locally-hosted PACS set up specifically for the institution's involvement in the SarcomaAI project. This method ensures that the original copies are never directly interacted with or potentially modified, thereby preserving their integrity and ensuring their safe involvement in the project by separating them from the copies involved in the project's course. PACS-to-PACS transfer also leverages the DICOM protocol, ensuring that the integrity of both the image data and its associated metadata is preserved during transfer.

Information: The DICOM (Digital Imaging and Communications in Medicine) protocol is a comprehensive standard used in medical imaging for storing, transmitting, and exchanging images and associated data. The protocol has two primary aspects: **image data and metadata** and **communication protocols**. The first aspect involves the organization and management of image data along with its accompanying metadata, which is essential for accurate interpretation and integration across various healthcare systems. This aspect is structured into four hierarchical levels:

1. **Patient Level**: This level stores patient information, including name, ID, and demographic details, which is crucial for identifying and associating the medical images with the correct individual.
2. **Study Level**: Here, data pertaining to a specific imaging study, such as the study date, time, and modality used (e.g., MRI, CT), is stored. This ensures that all images and metadata from a particular study are grouped together.
3. **Series Level**: This level organizes images within a study into series, each corresponding to a specific imaging sequence or protocol. For example, in an MRI study, there might be different series for T1-weighted and T2-weighted images.
4. **Image Level**: Finally, the image level contains the actual image data and its related attributes, such as image dimensions, pixel data, and orientation. Each image in a series is individually referenced and can be accessed or processed as needed.

The second aspect of the DICOM protocol focuses on the communication protocols that facilitate the exchange of medical images and information between devices and systems. This ensures that different imaging equipment, PACS, and healthcare information systems can interoperate seamlessly, allowing for efficient and reliable transfer of imaging data across various platforms.

Technical Information: A common method for performing PACS-to-PACS transfers of DICOMs data is by utilizing DCMTK, a suite of open-source libraries and applications designed to manage DICOM data. On Ubuntu, this can be installed as a package, offering key commands such as `movescu` (for moving DICOM data), `findscu` (for querying DICOMs), and others, facilitating efficient data transfer. Visit https://dcmtk.org/ for more information.

To facilitate efficient participation in the SarcomaAI project, it is recommended to host a standalone PACS system specifically intended for the project. This setup is particularly crucial when the institution plans to contribute a large number (more than 100) of scans. Hosting a dedicated PACS ensures seamless retrieval and management of these datasets, optimizing the workflow for both data transfer and subsequent processing.

Technical Information: Orthanc is a leading open-source PACS server that can be used to host a dedicated PACS for the SarcomaAI project. It provides robust functionality for managing and distributing medical images and is highly customizable to fit specific project needs. For more details and documentation, visit https://www.orthanc-server.com/.

To facilitate PACS-to-PACS transfer, it is essential to gain authorized access to the institution's PACS system. Each PACS is identified by three key components: the Application Entity Title (AET), the IP address, and the port number. These components are unique to the PACS and define how it can be accessed remotely. To enable a secure and seamless data transfer, the institution's system administrators—typically from the IT department responsible for managing the PACS—must whitelist the local PACS that has been set up for the SarcomaAI project. This whitelisting process involves granting the local PACS permission to communicate with the institution's PACS using the specified AET, IP address, and port number.

Normally, this access information must be requested by the individual responsible for the project's approval. This request should include any required documentation that aligns with the institution's policies and protocols for data handling. It is important to ensure that all access and data transfer procedures comply with institutional guidelines to maintain data security and integrity throughout the process.

This method of PACS-to-PACS transfer is necessary when contributing a large number of scans, as it ensures that the data is organized, secure, and readily accessible for analysis and machine learning applications.

2.2 Manual Data Export

In circumstances where direct extraction from the PACS system is not feasible due to institutional constraints or the need to manage smaller datasets, manual export may be employed. Institutional

limitations may include policies that restrict the dissemination of critical PACS access information, such as the Application Entity Title (AET), IP address, and port number, to safeguard system security and control over data access. Additionally, certain institutions may impose restrictions on non-standard data interactions during high-traffic periods to ensure system stability and availability. These limitations can introduce unforeseen challenges and hinder timely data extraction.

Furthermore, the establishment of a project-specific, local PACS may pose a significant technological burden, particularly if the institution lacks the necessary resources or technical expertise to support such an implementation for the SarcomaAI project.

In such cases, and when the dataset size is modest (fewer than 300 scans), manual export may represent a feasible alternative. However, this process is inherently repetitive and labor-intensive, as many PACS systems typically permit the export of only one patient's data at a time. Moreover, the exported scans are usually transferred directly to the local file system, necessitating manual organization and management. This approach also circumvents the intended advantages of the DICOM communication protocol, which is designed to facilitate standardized and efficient data exchange. Despite these limitations, manual export remains a viable option when other methods are not practical.

Technical Information: In InteleViewer, the web-based CT/MRI viewer provided by the IntelePACS system, manual export can be accomplished by selecting the *File* > *Export* option from the top menu.

3 Data Preprocessing

The following subsections outline the sequential steps in the pipeline that are essential for performing the necessary data preprocessing and preparation. These steps ensure that the data is properly anonymized and has the necessary corrective actions applied, such as N4 bias correction and z-score normalization, to optimize the results in the final federated learning stage. This preparation is crucial for ensuring the quality and consistency of the data before it is stored and used in subsequent analysis.

3.1 Anonymization Script

In the context of medical imaging, particularly when using the DICOM standard, a substantial amount of personal identifying information (PII) is embedded within the metadata associated with each image file. This metadata exists at various levels within the DICOM hierarchy, starting at the patient level and continuing through the study, series, and instance levels. Examples of identifying information at the patient level include the Patient Name (tagged as (0010,0010) in the DICOM standard), Patient ID, Birth Date, and Gender. As these identifiers are embedded within the metadata, it is crucial to remove or anonymize them to protect patient privacy and comply with ethical and legal requirements.

The anonymization process can be automated through the use of scripted procedures, such as those written in Python. By programmatically modifying or removing these identifying fields, the integrity of the data can be maintained while ensuring that patient confidentiality is preserved. A typical approach involves using a Python script to locate and either replace the sensitive information with blank or generic values or entirely remove the fields from the DICOM files. This automated method ensures consistency and efficiency, especially when dealing with large datasets.

Technical Information: The pydicom package in Python is a powerful tool for handling DICOM files. It allows users to read, modify, and write DICOM datasets programmatically. For instance, the following lines of code demonstrate how to anonymize the Patient Name and Patient ID fields in a DICOM file by setting them to blank values:

```
dataset.PatientName = ""
dataset.PatientID = ""
```

These lines of code target the Patient Name field, corresponding to the DICOM tag (0010,0010), and the Patient ID field, corresponding to the DICOM tag (0010,0020), removing the patient's name and ID to effectively anonymize these pieces of information. By iterating over all relevant fields, this method can be applied to remove or anonymize all identifying information. For more details on the pydicom package, visit https://pydicom.github.io/. Additionally, see https://dicom.innolitics.com/ciods/mr-image for more on the DICOM standard for MRIs.

The SarcomaAI pipeline, initially developed for implementation at the McGill University Health Centre (MUHC) and Memorial Sloan Kettering Cancer Center (MSKCC), includes a pre-configured script that automates this anonymization process. This script can be provided to institutions participating in the SarcomaAI project and, with minor modifications to account for the specific DICOM conventions used by the institution, can be employed to achieve comprehensive data anonymization with minimal additional development effort. This approach ensures that all participating sites can meet the necessary privacy and confidentiality standards while contributing data to the project.

3.2 N4 Bias Correction

N4 bias correction is a widely recognized technique used to address and correct intensity non-uniformity MR images. This non-uniformity, often referred to as bias or field inhomogeneity, can arise from various sources, including imperfections in the magnetic field or differences in tissue properties, which can lead to variations in signal intensity across the image. Such variations can obscure critical anatomical details and introduce artifacts that complicate image interpretation and subsequent analysis.

Information: N4 Bias Correction standardizes the intensity values across MR images by correcting for low-frequency intensity non-uniformities, often caused by inhomogeneities in the magnetic field or variations in the radiofrequency (RF) coil sensitivity. This process enhances both the contrast and uniformity of the images, making subtle anatomical features more discernible. The corrected images exhibit consistent intensity levels, which is crucial for ensuring the reliability and accuracy of quantitative measurements, such as those used in radiomics analysis. These measurements might include texture analysis, voxel-based morphometry, or other features that depend on intensity values being consistent across the image.

For orthopaedic oncology, where precise imaging is paramount for diagnosing and planning the treatment of bone and soft tissue tumors, minimizing these biases is crucial. This standardization is essential for ensuring that the images are reliable and that the quantitative measurements taken from them are accurate, which is particularly important when these images are used in a federated learning model where consistency across datasets is critical.

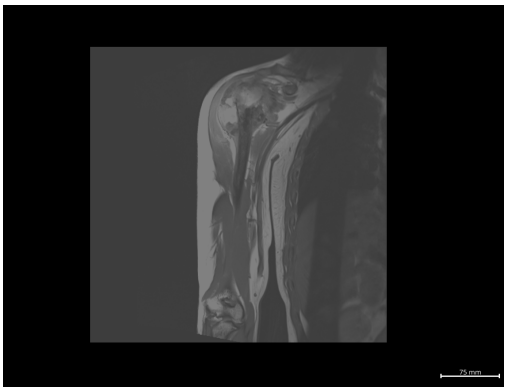

Figure 1: MR Image Before N4 Bias Correction

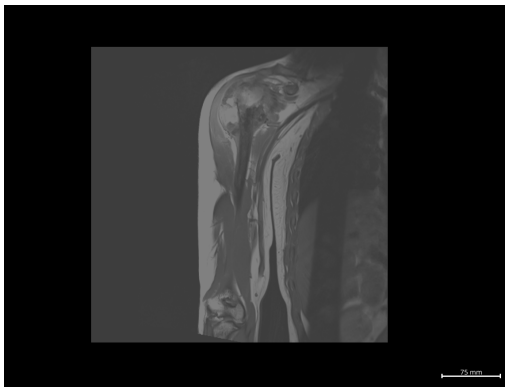

Figure 2: MR Image After N4 Bias Correction

To the naked eye, the images before and after N4 bias correction may appear nearly identical. However, the nuanced corrections made by this algorithm can significantly enhance the performance of neural network models, particularly when applied to hundreds or thousands of images. These subtle adjustments improve the uniformity of the dataset, leading to more consistent and reliable outcomes in machine learning applications.

The SarcomaAI project pipeline has developed a Python script specifically designed to implement N4 bias correction as part of the preprocessing workflow. This script has been optimized for use in this pipeline and can be shared with collaborating institutions, ensuring that all participating sites can uniformly prepare their imaging data for the federated learning model.

Technical Information: The SimpleITK package in Python provides a robust toolset for medical image analysis, including a built-in function for N4 bias correction. SimpleITK's implementation of the N4 algorithm allows for efficient and automated correction of intensity non-uniformity in MR images. Here is an example of how to use the N4 bias correction function in SimpleITK:

```
import SimpleITK as sitk

# Load the MR image
image = sitk.ReadImage('path_to_image.nii', sitk.sitkFloat32)

# Perform N4 bias correction
corrector = sitk.N4BiasFieldCorrectionImageFilter()
corrected_image = corrector.Execute(image)

# Save the corrected image
sitk.WriteImage(corrected_image, 'corrected_image.nii')
```

3.3 Z-score Normalization

Z-score normalization is a statistical method used to standardize data by converting it into a distribution where the mean is zero and the standard deviation is one. In the context of volumetric

medical imaging data, such as MR images, Z-score normalization plays a crucial role in preparing the data for advanced analysis and machine learning.

Information: The Z-score normalization algorithm works by adjusting each pixel or voxel intensity value within an image according to the formula:

$$Z = \frac{X - \mu}{\sigma}$$

where X is the original intensity value, μ is the mean intensity of the image, and σ is the standard deviation of the intensity values. This transformation ensures that the resulting data has a mean of zero and a standard deviation of one, effectively eliminating any bias caused by variations in intensity scales across different images. This standardization is essential for consistent and accurate analysis across datasets.

In volumetric data, this step is particularly important because it allows for the comparison of images across different patients, scanners, and institutions, all of which may have varying intensity distributions due to differences in imaging protocols, equipment calibration, or patient anatomy. By standardizing the intensity values, Z-score normalization ensures that the data is consistent and comparable, which is essential for reliable analysis.

Moreover, in a federated learning environment where datasets from multiple institutions are combined to train machine learning models, Z-score normalization becomes even more vital. The standardization ensures that the input data fed into the federated learning algorithms is homogeneous, regardless of its origin. This homogeneity is crucial for improving the generalizability of the models and for minimizing the risk of overfitting to the idiosyncrasies of a particular dataset.

The SarcomaAI pipeline includes a pre-configured Python script designed to perform Z-score normalization on volumetric data. This ensures that the data from any participating institution is prepared in a manner that is optimal for downstream analysis within the SarcomaAI framework.

4 Data Storage

After data preprocessing is complete, the de-identified data needs to be stored in a manner that supports further analysis and machine learning tasks. While the initial extraction of data may involve using a SarcomaAI project-specific local PACS, this system is primarily intended for clinical workflows and is not always optimal for the advanced data management and processing required in research contexts. Therefore, transitioning the data to more specialized storage systems, such as a cloud-based imaging informatics platforms, is recommended for the next stages of the pipeline.

4.1 Imaging Informatics Platforms vs. PACS

In the SarcomaAI pipeline, while PACS is used initially for storing extracted imaging data due to its integration with healthcare and clinical workflows, the complexity and demands of research tasks require transitioning to an imaging informatics platform for subsequent stages. Imaging informatics platforms like FlyWheel, XNAT (Extensible Neuroimaging Archive Toolkit), and LORIS (Longitudinal Online Research and Imaging System) are designed specifically for research environments, offering advanced capabilities that extend beyond the scope of PACS. These platforms are essential for managing complex metadata, supporting sophisticated analysis pipelines, facilitating data annotation, and enabling secure data sharing among research collaborators.

The transition to an imaging informatics platform is crucial for effectively managing the data as it progresses from the preprocessing stages of the pipeline through further analysis and processing, and ultimately for machine learning applications. Platforms like XNAT, FlyWheel, and LORIS provide the necessary tools to organize and process the data, ensuring that it is ready for in-depth research and collaborative efforts. Unlike traditional file management systems on a local computer, which lack the ability to handle the complexity and scale of such data, these platforms offer robust solutions that maintain data integrity and streamline workflow. This shift not only supports the technical requirements of advanced data management but also enhances the overall efficiency and security of the research process, making it feasible to manage and analyze large datasets that would be impractical with simpler storage methods.

Technical Information: XNAT is an enterprise-grade, open-source imaging informatics platform built on an Apache Tomcat server, designed to facilitate the management and analysis of research imaging data. It offers seamless integration with the DICOM standard, enabling efficient importation of imaging datasets while ensuring robust anonymization protocols to safeguard patient privacy. The platform is equipped with comprehensive security features, allowing for precise control over data access. XNAT's integrated search and reporting capabilities further enhance its utility, providing researchers with the tools to efficiently query imaging data alongside associated clinical or research metadata, thus streamlining the research workflow. Additionally, XNAT supports the execution of complex data processing pipelines, making it particularly advantageous for large-scale studies that demand substantial computational resources. Its modular architecture permits extensive customization, enabling adaptation to the specific needs of diverse research projects.

The SarcomaAI project has successfully developed an Amazon Web Services (AWS) stack that securely deploys XNAT, leveraging cloud infrastructure to ensure that the platform is scalable, secure, and accessible to all project collaborators. The following image illustrates the AWS architecture used for this deployment:

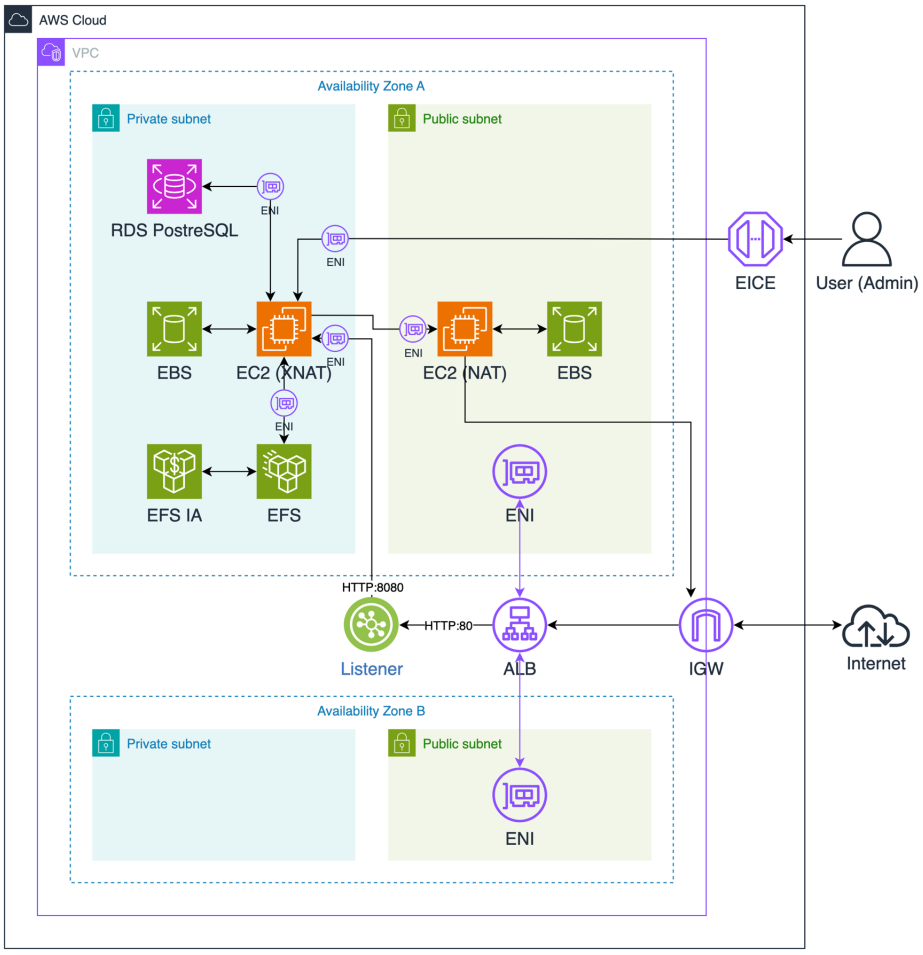

Figure 3: AWS Architecture Diagram for Secure XNAT Deployment

4.2 Cloud-Based vs. Local Storage

When deploying an imaging informatics platform, institutions have the option of hosting the system locally or in the cloud. While local deployment may offer direct control over the infrastructure, it also requires substantial resources for maintenance, scalability, and security.

Information: Modern cloud providers, such as Amazon Web Services (AWS), offer a compelling alternative with enhanced security, scalability, and accessibility. Data centers operated by these providers are secured with state-of-the-art measures, often exceeding the security capabilities of most institutions. This includes physical security measures like 24/7 surveillance, biometric access controls, and digital protections like advanced encryption protocols for data at rest and in transit. Additionally, cloud providers offer customizable deployment options, allowing institutions to select the region in which their compute resources are hosted. For example, the McGill University Health Centre (MUHC) deploys its resources in the *ca-central-1* region, located in Montreal, Quebec. This choice keeps the data geographically close to the institution while entrusting a corporation with billions of dollars in infrastructure to manage

the technical and security aspects.

Cloud-based storage also facilitates easier collaboration across institutions, allowing data to be accessed from anywhere with the necessary permissions. The scalability of cloud services ensures that storage and computational resources can be adjusted according to the project's needs without the need for significant upfront investment in hardware.

However, deploying and managing cloud-based solutions does require some technical expertise. Institutions should consider involving personnel with experience in cloud infrastructure, such as a Master's or PhD student with AWS solutions experience, or collaborating with the institution's IT department, to ensure that the deployment is optimized for the project's requirements and that best practices in security and data management are followed.

5 Image Data Segmentation

Image segmentation is a fundamental process in medical imaging, wherein an image is partitioned into distinct segments or regions, each corresponding to different anatomical structures or tissue types. This process is critical for isolating areas of interest within the image, such as tumors, organs, or other relevant anatomical features, thereby facilitating more detailed analysis and interpretation. In the context of soft tissue sarcoma research within orthopaedic oncology, accurate segmentation is particularly vital for delineating tumor boundaries, which is essential for both diagnostic and therapeutic planning. Moreover, precise segmentation is necessary for training the multi-modal SarcomaAI neural network, which uses the delineated tumor margins to identify patterns in pathology across thousands of patients, ultimately providing prognostic information.

When dealing with volumetric data, such as MRIs, the complexity of image segmentation increases due to the three-dimensional nature of the data. Segmentation in this context involves the precise identification and delineation of structures across multiple image slices, ensuring consistency and accuracy throughout the entire volume. This is crucial not only for clinical assessments but also for research purposes, particularly in projects like SarcomaAI.

Traditionally, image segmentation has been performed manually by experts, who meticulously annotate images on a slice-by-slice basis. While this manual approach can yield highly accurate results, it is also exceptionally time-consuming and labor-intensive. Recent technological advancements have introduced novel approaches to segmentation, including automatic segmentation using pre-trained neural network models. These models offer the potential to rapidly and accurately segment images, thereby reducing the burden on experts and ensuring consistent results across large datasets.

5.1 Manual Segmentation

Manual segmentation is a critical process that involves the precise delineation of anatomical structures within imaging data, typically performed using specialized software integrated into imaging informatics platforms such as XNAT, FlyWheel, or LORIS. These platforms often utilize viewers like the Open Health Imaging Foundation (OHIF) viewer, which facilitates accurate annotation directly on the imaging slices.

During manual segmentation, the expert carefully traces the boundaries of the tumor or other structures of interest on each slice of the MRI, often using a cursor or, for greater precision, a tablet with a pen or stylus. The resulting annotations create a mask associated with each slice, effectively isolating the region of interest. This mask is crucial for further analysis or for training

neural networks. The accuracy of these segmentations is paramount, as reducing segmentation errors directly improves the performance of machine learning models trained on this data. Precise delineation enhances the neural network's ability to learn from the data, resulting in more reliable and accurate prognostic predictions.

The time required for manual segmentation can be considerable. Depending on the complexity of the case and the number of slices involved, segmenting a single tumor in a full MRI scan may take between 15 minutes and one hour. For institutions managing large datasets, this time investment can quickly become substantial. It is therefore essential to plan and allocate sufficient time for this task, ensuring that the dataset is fully prepared for participation in the SarcomaAI project according to the project schedule and any other operational constraints.

Technical Information: The Open Health Imaging Foundation (OHIF) viewer is a powerful, open-source tool for visualizing and annotating medical images, widely used in conjunction with platforms like XNAT. It supports various imaging modalities, including MRI and CT, and is optimized for both desktop and tablet interfaces, allowing for precise manual segmentation.

OHIF's capabilities include multi-planar reformatting, which enables segmentation across different planes (axial, coronal, sagittal) simultaneously, ensuring consistency in volumetric data. The viewer also integrates robust DICOM support, ensuring that segmentations are stored in a standardized format suitable for further analysis.

The viewer is user-friendly yet powerful, making it ideal for detailed image analysis in both research and clinical settings. The SarcomaAI pipeline easily integrates OHIF with XNAT using the available plugin, and more information including installation instructions are available at https://wiki.xnat.org/xnat-ohif-viewer.

Given the substantial time commitment required for manual segmentation, institutions with limited expert resources may find automatic segmentation to be a more practical solution for participating in the SarcomaAI project. Automatic segmentation can significantly reduce the time required for data preparation, enabling the processing of larger datasets with greater efficiency.

5.2 Automatic Segmentation

In addition to manual segmentation, a promising and increasingly effective approach is automatic segmentation, which utilizes neural network models pre-trained on manually annotated datasets of similar MR images to perform tumor delineation. These models are trained on extensive datasets where each image is paired with a corresponding segmentation mask, enabling the network to learn how to accurately identify and segment tumors in new, unannotated images.

Notable datasets that have contributed significantly to the advancement of automatic segmentation include the Brain Tumor Segmentation (BraTS) challenge datasets, which offer comprehensive annotations for brain tumors, and the ISLES challenge datasets, focused on ischemic stroke lesion segmentation. These resources have been pivotal in developing a range of automatic segmentation models, which can be adapted for various tumor types and imaging modalities.

Notable datasets that have contributed significantly to the advancement of automatic segmentation include the Brain Tumor Segmentation (BraTS) challenge datasets, which provide comprehensive annotations for brain tumors, and the ISLES challenge datasets, focused on ischemic stroke lesion segmentation. These resources have been pivotal in developing a range of automatic segmentation models, which can be adapted for various tumor types and imaging modalities. In addition to these datasets, several tools and platforms now offer integrated auto-segmentation capabilities.

For instance, AWS provides Amazon SageMaker Ground Truth, a machine learning service that can be adapted for medical image segmentation tasks, and Meta has recently released their improved Segment Anything Model 2 (SAM 2). Furthermore, many online and offline annotation tools, such as 3D Slicer, ITK-SNAP, and commercial platforms like NVIDIA Clara, incorporate built-in auto-segmentation tools with varying degrees of accuracy. These tools can be employed directly within the annotation workflow, providing a practical option for quickly generating segmentation masks, although their performance can vary depending on the complexity of the task and the quality of the underlying models.

Information: In the context of image segmentation, particularly with volumetric data (3D images), it is essential to understand the distinction between instance segmentation and semantic segmentation.

- **Semantic segmentation** involves labeling each voxel (the 3D equivalent of a pixel) in the image with a class label, such as "tumor," "healthy tissue," or "bone." This method does not differentiate between individual objects of the same class; for example, all tumor regions are labeled as "tumor" without distinguishing between separate tumors in the same image.
- **Instance segmentation**, on the other hand, goes a step further by identifying and delineating individual instances of objects within the same class. In the context of MRIs, this would mean not only identifying tumor tissue but also distinguishing between separate tumor masses within the same scan.

For most medical imaging tasks, particularly in 3D MRI segmentation, semantic segmentation is employed and typically sufficient. This voxel-based approach accurately identifies and classifies regions of interest, such as tumor tissue, without the need to distinguish between multiple instances of the same class.

As part of the SarcomaAI project, the Memorial Sloan Kettering Cancer Center (MSKCC) is currently developing an automatic segmentation model tailored to the specific requirements of sarcoma imaging. This model is being trained on the MSKCC patient dataset, which has been meticulously annotated to provide high-quality training data. Following rigorous validation, this automatic segmentation tool will be applied to the McGill University Health Centre (MUHC) dataset, enhancing the model's generalizability and performance across diverse patient populations and imaging conditions.

The incorporation of automatic segmentation into the SarcomaAI pipeline is anticipated to substantially reduce the time and resources required for data preparation, thereby facilitating the efficient processing of large datasets and enhancing the overall scalability and impact of the project.

6 Deep Learning Multimodal Model with Federated Learning

The deep learning multimodal model developed for the SarcomaAI project represents the culmination of the comprehensive data pipeline outlined in this playbook, designed to support advanced AI-driven insights into soft tissue sarcomas. This model is implemented using federated learning, a method that enables collaborative training across multiple institutions while preserving patient data privacy. Detailed discussion on federated learning follows later in this section.

The steps in the data pipeline, including data extraction, preprocessing of volumetric data, seg-

mentation, and secure data storage, are all critical to ensuring the quality and consistency of the data used to train the model. Deep learning is a subset of machine learning that involves the use of artificial neural networks with many layers to model complex patterns and relationships in data, making it particularly powerful for tasks such as image recognition. The accuracy of the model is heavily dependent on the quality of the input data, where accurate annotations through segmentation provide the critical information the model needs to learn and identify intricate patterns. This robust data preparation process ensures that the model is trained on datasets that are both high-quality and representative, ultimately leading to more reliable and generalizable prognostic predictions.

6.1 Multimodal Model Architecture

The SarcomaAI model is a multimodal neural network, representing an advanced application of deep learning techniques designed to enhance predictive accuracy by integrating diverse data sources.

Information: Multimodal learning is a deep learning technique in which a model is trained on multiple types of data, known as modalities, to enhance its predictive capabilities. Each modality—such as text, images, or numerical data—provides unique insights. When these diverse data sources are integrated, they offer a more comprehensive and nuanced understanding of the problem, leading to predictions that are more accurate than those derived from a single data type.

A critical component of multimodal learning is the model's ability to effectively combine insights from each modality. This is typically achieved by processing each data type through separate pathways, or subnetworks, which then converge later in the model to generate a unified prediction. Gradient blending, a technique used in multimodal models, merges the contributions of each modality by blending the gradients during training. This enables the model to assign appropriate weight to each modality based on its relevance to the specific prediction task, ensuring that the final predictions leverage the strengths of each data type and enhancing the model's overall performance.

In the SarcomaAI project, multimodal learning is employed to integrate clinical and imaging data, leveraging the unique strengths of each to generate more robust and accurate predictions. Clinical data provides essential contextual information about the patient and their diagnosis, while imaging data offers detailed insights into the tumor and surrounding tissues. Through the combination of these modalities, the model is able to deliver comprehensive prognostic predictions that are more effective than those derived from a single data source.

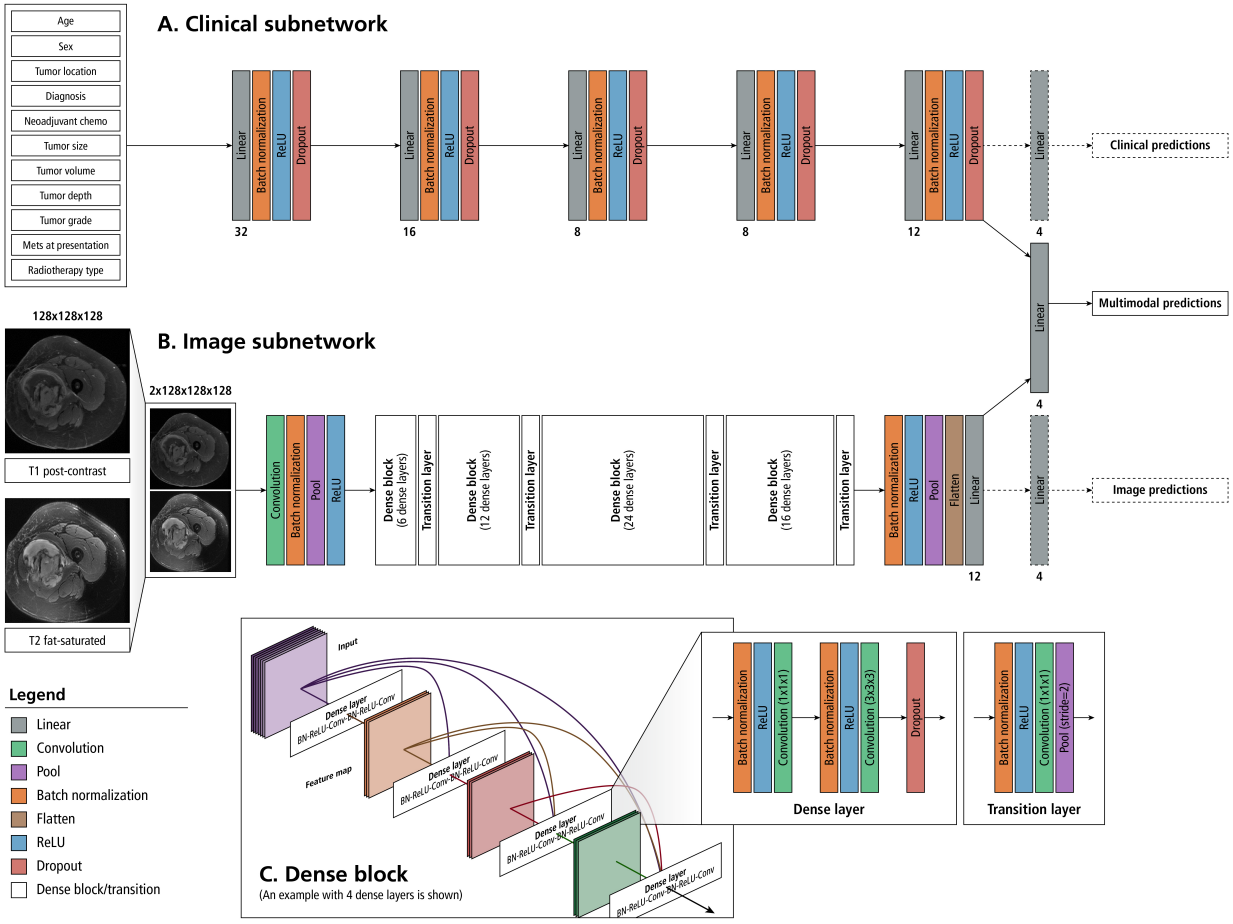

Figure 4: SarcomaAI Multimodal Neural Network Architecture

- **Clinical Subnetwork:** This subnetwork analyzes 11 key clinical variables: age, sex, tumor location, diagnosis, neoadjuvant chemotherapy, tumor size, tumor volume, tumor depth, tumor grade, metastases at presentation, and radiotherapy type. The network is designed to distill complex clinical information into meaningful predictions. It utilizes a combination of techniques to ensure accurate and reliable outcomes: linear activation maintains consistency in processing the data, batch normalization stabilizes learning by adjusting and scaling inputs, ReLU (Rectified Linear Unit) activation introduces non-linearity to help the model capture complex patterns, and dropout helps prevent overfitting and ensures the model generalizes well to new patient data. The final output layer generates predictions that are crucial for assessing patient outcomes.

- **Image Subnetwork:** The image subnetwork is specifically designed to process volumetric imaging data, taking as input one T1-weighted and one T2-weighted MRI scan. These scans are pre-processed using the methods outlined in this playbook and resized to 128x128x128 voxels to standardize the input data. This subnetwork employs a sequence of dense blocks and transition layers to extract high-level features from the imaging data. Dense blocks facilitate the efficient flow of information through the network, capturing important features while reducing the model's complexity. Transition layers downsample the data, improving

computational efficiency without sacrificing critical information. The image subnetwork then generates predictions based on these extracted features.

The outputs from the clinical and image subnetworks are combined to produce the final multimodal predictions. This integration of clinical and imaging data allows the model to utilize the complementary strengths of each modality, leading to more accurate and reliable predictions. The SarcomaAI model can perform both classification tasks and time-to-event modeling, particularly for predicting survival and distant metastasis.

The results and performance of the SarcomaAI model, as demonstrated by C-Index results, exceed those of previous models, showcasing the effectiveness of this multimodal approach in the context of soft tissue sarcoma prognosis.

6.2 Federated Learning

Federated learning (FL) is a decentralized approach to training machine learning models that allows multiple institutions to collaborate without sharing raw patient data. This method is particularly valuable in the SarcomaAI project, where aggregating data from diverse sources can significantly enhance the model's predictive capabilities. By enabling institutions to keep patient data within their own secure environments, federated learning mitigates privacy concerns while still allowing for the development of robust, generalized models.

Information: Federated learning involves a coordinated interaction between two key types of servers:

- **Central server:** The central server, or federated server, functions as the primary coordinator and model aggregator for the federated learning project. It is responsible for controlling and managing the workflow among all client servers and itself, such as initiating training tasks and monitoring progress, as well as storing the global model and managing the synchronization of updates from all participants. Importantly, the central server does not have direct access to the raw data. Instead, it collects locally trained model updates from each client server and aggregates these updates to refine the global model.
- **Client server:** The client servers are hosted within each participating institution and are responsible for training the model locally using the institution's data. The client servers handle all computations, including data preprocessing, model training, and updating model weights based on the local dataset. After training, only the updated weights are transmitted to the central server. This setup ensures that the raw patient data remains securely within the institution's environment and is never exposed or shared externally, while still contributing to the global model.

The federated learning process commences with the central server initializing a global model, which is distributed to each participating institution's client server. This model maintains a standardized architecture across all institutions, ensuring consistency in the learning process. Each client server autonomously trains the model on its local dataset, with the model weights being adjusted based on the specific characteristics of the data. Upon completion of the local training, the updated model weights are transmitted back to the central server. The central server aggregates these updates from all participants to refine the global model further. This iterative process is repeated over multiple cycles, progressively enhancing the model's accuracy and performance.

A fundamental advantage of federated learning is that it obviates the need to transfer raw patient data between institutions. All data remains securely within the institution's own computing infrastructure, which is under the full control of the institution. External collaborators receive only the aggregated model updates, and these model weights transmitted to the central server cannot be reverse-engineered to reconstruct the original data. This ensures the robust protection of patient privacy, making federated learning an ideal methodology for sensitive medical research endeavors such as the SarcomaAI project.

Technical Information: NVFlare is the federated learning platform employed by the SarcomaAI project, developed by NVIDIA, a global leader in AI and machine learning technologies. NVFlare is designed to enable secure, scalable, and privacy-preserving federated learning, making it a trusted choice for medical research and other sensitive applications. The platform offers advanced encryption to safeguard data during transmission, along with flexible deployment options that can accommodate various institutional needs. NVFlare also supports complex, large-scale collaborations across multiple institutions, ensuring that each participant can contribute to the model's training without compromising data privacy. For more information, see https://developer.nvidia.com/flare

The central server for the SarcomaAI project is managed by a lead researcher at MSKCC who is responsible for ensuring the technical coordination and management of the federated learning process. It is recommended for each participating institution to designate a researcher with experience in server infrastructure and a basic understanding of federated learning principles and machine learning. This individual will be responsible for ensuring the client server remains active and operational, applying any necessary updates or modifications as directed by the SarcomaAI project team. While the setup and ongoing management of the client server are not overly complex, having a dedicated resource to oversee these tasks is essential for smooth participation in the federated learning process. It is also important to note that client servers can be deployed on either local physical hardware or cloud-based platforms, providing flexibility based on institutional requirements.

6.3 Local vs. Cloud-Based Compute Resources

Models can be run on either local or cloud-based compute resources, with both options offering distinct advantages depending on the institution's existing infrastructure and goals. Institutions that have access to on-site, machine learning-oriented computing resources, including sufficient Graphics Processing Unit (GPU) capacity, may opt to leverage these existing resources by deploying the client server locally. Additionally, institutions may choose local deployment to avoid the costs associated with cloud providers or to adhere to institutional policies that favor maintaining data and computational processes on-premises. However, when using local resources, it is critical to ensure that the hardware is capable of handling the high graphical and computational demands of model training. In federated learning, a significant imbalance in computational capacity between institutions can create a bottleneck, where underpowered resources at one participant can slow down the aggregation process at the central server, causing delays across the entire system.

In addition to hardware capacity, network stability is also an important consideration when opting for local deployment. An unstable or low-bandwidth internet connection can create further delays, particularly during the transmission of model updates to the central server. Institutions deploying locally must ensure that both their compute resources and network infrastructure are

capable of maintaining the necessary speed and reliability to support continuous and efficient model training, without causing interruptions to the federated learning process.

Cloud platforms such as AWS offer advanced GPU capabilities and flexible infrastructure, enabling institutions to meet the computational demands of model training without the need for dedicated local resources. Deploying the client server on cloud-based infrastructure allows institutions to offload security, maintenance, and scaling concerns to the cloud provider, which may be particularly advantageous for those with limited IT support. Moreover, a cloud-based setup seamlessly integrates with cloud-based dataset storage. Hosting both the client server and data storage within the same cloud environment ensures that storage resources are directly and exclusively accessible to the institution's own client server, enhancing efficiency, security, and reducing latency while simplifying data access management.

The figure below shows a cloud-based central server and three client servers—one deployed locally on-site and two cloud-based—demonstrating the flexibility of federated learning, allowing institutions to choose their preferred setup while collaborating seamlessly.

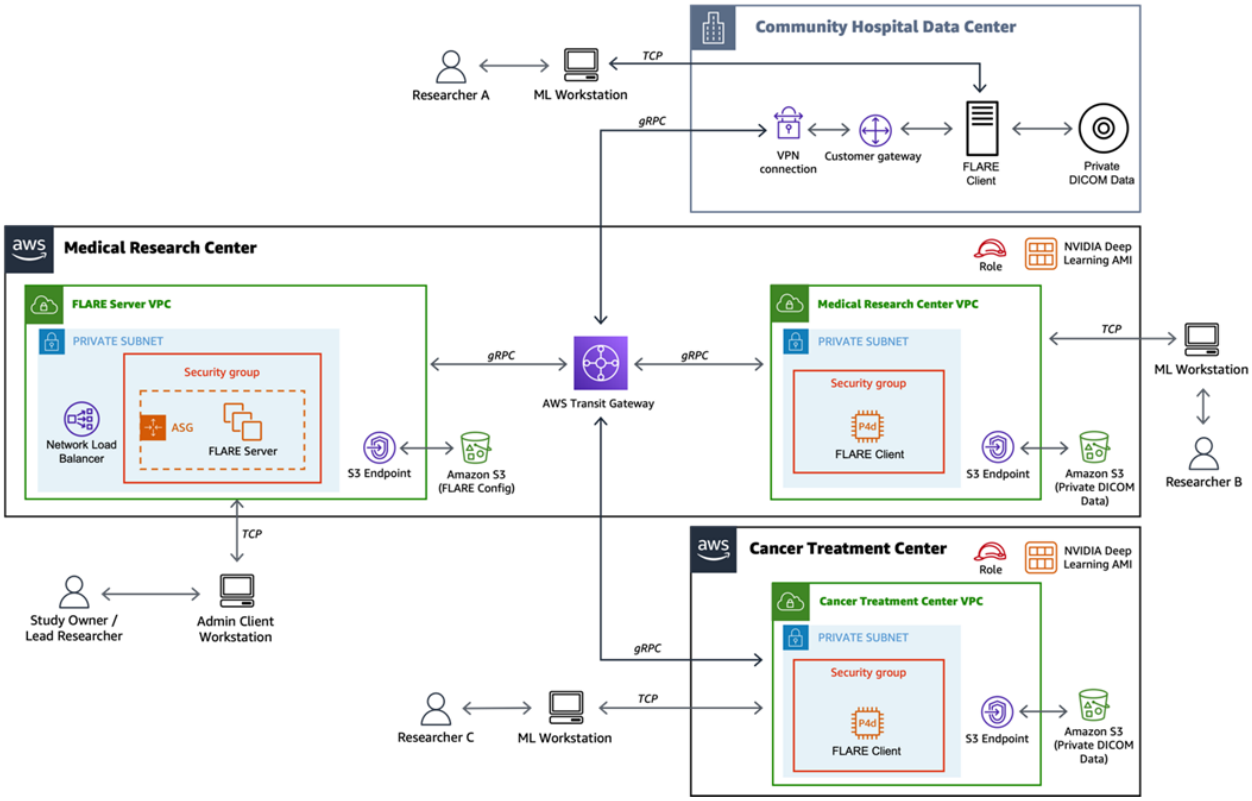

Figure 5: AWS Architecture Diagram for NVFlare Deployment

Technical Information: The SarcomaAI project has developed easy-to-use template files that simplify the deployment of an NVFlare client server on AWS. These templates are designed to ensure institutions can quickly set up the required infrastructure while adhering to the security and performance standards needed for federated learning.

Key AWS resources involved in the deployment include a Virtual Private Cloud (VPC)

with a private subnet, where the virtual machine hosting the client server resides. The private subnet ensures the server has no direct internet access, enhancing security. A security group is configured to allow only inbound connections from a private S3 bucket containing the DICOM data, enforcing strict access control. Communication between the VPC and the central server is facilitated via a secure gRPC protocol, routed through an AWS Transit Gateway, ensuring reliable and secure coordination throughout the federated learning process.

Glossary:

- **Virtual Private Cloud (VPC):** A private, isolated network within AWS used to deploy and manage cloud resources.
- **Private Subnet:** A segment of a VPC that restricts internet access to its resources for enhanced security.
- **Virtual Machine (VM):** A cloud-based server that runs the NVFlare client server for federated learning.
- **Security Group:** A virtual firewall that controls inbound and outbound traffic for AWS resources, allowing specific connections.
- **S3 Bucket:** A secure, scalable storage service used to store DICOM data, accessible only by the designated resources.
- **gRPC Protocol:** A secure, high-performance communication protocol used for exchanging data between the client and central servers.
- **AWS Transit Gateway:** A networking service that connects VPCs and on-premises networks, enabling secure communication across different environments.

It should be mentioned that other cloud providers, such as Google Cloud and Microsoft Azure, offer equivalent resources that can be used to deploy the client server.

Institutions should carefully evaluate their available resources and consult with the SarcomaAI project team to verify that their chosen setup is suitable. Whether opting for local or cloud-based deployment, it is crucial to ensure that the infrastructure is capable of supporting the demands of the federated learning process.