Project Proposal

Client
Dr Vera Chung, vera.chung@sydney.edu.au

Tutor
Muhammad atif Iqbal, muhammadatif.iqbal@sydney.edu.au

Group members

Rajiv Mehta	rmeh0608	540771859
Ngai Chun Fu	ngfu0299	540363470
David Cortés Sánchez	vcor0924	500623734
Eshan Arora	earo0293	540425550
Mohamad Akmal bin Abu Bakar	mbin0316	530432780
HanWen Tian(Simon)	htia0294	540244065

Abstract

Federated Learning (FL) has been studied as an approach to deal with privacy concerns in machine learning, especially in domains such as healthcare, where sharing data is often not even legal. This project proposes a multi-modal federated learning framework with the objective of enhancing machine learning model performance in medical imaging, with a particular focus on the issues produced by modality and class heterogeneity across clients. Initially, the framework is designed to work with clients possessing identical modalities and classes, before involving varied modalities and overlapping or distinct class sets across clients. The project seeks to evaluate the effectiveness of federated learning in medical image classification tasks, comparing its performance with centralized models in terms of accuracy and robustness, and analyzing performance in local and global scale. By utilizing publicly available medical image datasets, the research aims to explore how federated learning can perform collaborative model training while preserving patient privacy.

Contents

1 Introduction

In a world where security and privacy are at the forefront of minds, federated learning has become an increasingly important methodology as it enables machine learning models to maintain data privacy, empowering organizations to harness collective insights without compromising sensitive information. This is particularly important in industries such as the medical field where hospitals often due to legal requirements and patient confidentiality, data cannot be shared between hospitals easily making the process of training machine learning models difficult, halting advancements within this industry. Furthermore, the challenge is compounded by the diversity of data modalities such as different types of imaging, electronic health records and genetic data, all requiring specialised models. This study aims to tackle the challenges faced by the medical industry as the authors propose a new multi-modal federated machine learning model framework.

2 Literature Review

Several reviews have been published to cover the topic of Federated Learning in diverse task contexts and research areas. One of them is *'A Review of Federated Learning Methods in Heterogeneous Scenarios'* [1]. The authors highlight the need for distributed collaborative training to address the problem of data scarcity. However, this approach introduces complexity and heterogeneity in Federated Learning scenarios, which affect the efficiency and accuracy of models trained in this setting. This work aims to fill the gap created by the lack of comprehensive and specific reviews on the heterogeneity of FL. The study provides a definition of Federated Learning and establishes the background of convergence theory. Specifically, the authors define five forms of non-IID distribution that influence the convergence process in a Federated Learning setting: label skew, feature skew, same label with different features, different labels with the same features, and quantity skew. They categorize the heterogeneous challenges in Federated Learning into three types: device heterogeneity, data heterogeneity, and model heterogeneity. Additionally, they discuss various approaches to addressing these challenges and evaluate their effectiveness. However, they fall short in providing a detailed comparison of experimental data.

In terms of heterogeneity context, data heterogeneity is a special interest for this project. Specifically, modality heterogeneity has been seen as one of the big problems in convergence and accuracy. *'Multimodal Federated Learning: A Survey'* by Che et. al. dug deep into this topic [2]. The authors conducted a literature review on multimodal federated learning, following a well-defined article selection process with specific criteria. Their proposed approach distinguishes between Multimodal Federated Learning (MFL) and traditional Federated Learning, introducing the concepts of modality combination and modality heterogeneity. To facilitate comparisons between the reviewed studies, they classified MFL research into four categories: horizontal, vertical, transfer, and hybrid MFL, expanding on the initial classification of MFL into congruent and incongruent types. Additionally, they identified common tasks in Multimodal Federated Learning, compiled a benchmark of datasets for MFL, and discussed potential research directions and challenges in the field. Some of these challenges include modality heterogeneity, missing modalities, data complexity, large-scale pre-trained models, privacy concerns, and weakly supervised learning.

'An efficient federated learning method based on enhanced classification-GAN for medical image classification' by Liu et.al explores the issues faced by the medical industry through the lack of the labeled data and privacy concerns causing image classification models to struggle to accurately classify medical images [3]. To address this, the authors propose a federated learning generative

adversarial network implemented with blockchain to improve security of the model while also addressing the issues of reduced number of labels present in medical imaging. The authors evaluated this model using the 'Covid-19 Radiography Database' and 'ChestCOVID', which contain lung images from COVID-19-positive cases, normal lungs, viral pneumonia, and other conditions. These datasets, comprising 900 lung images, were divided into two new independent subsets, derived from prior COVID-19 studies [4] [5] [6]. The proposed FEDBG was then compared to models proposed in previous studies through training time, precision, F1_Score and synthetic image quality which was then evaluated through an ablation study. The results found that on both training sets, the precision, recall and F1_Score to be above 95 for all categories placing it higher then baseline models. It was also found that the training speed appeared to approximately 27–38% faster and overall accuracy increase 0.9–2%, therefore indicating that the proposed FEDBG is an optimal choice in medical imaging classification.

'Active Learning Based Federated Learning for Waste and Natural Disaster Image Classification' by Ahmed et.al explores how federated learning can benefit from the training of unlabeled data from clients using active learning modeling techniques [7]. To achieve this, the authors used two datasets the first a collection of natural disasters related images that were sourced from social media platforms. This collection contained 7000 images with 8 different categories, 5000 split into smaller training datasets with 2450 being used to testing with the remainder remaining unlabeled for validation. The second dataset was used as a benchmark dataset coming from a previous paper *'CNN-RNN: A large-scale hierarchical image classification framework'* by Guo et.al which contained 6 categories of different types of waste containing 2527 images [8]. These 2 datasets were then used in the authors proposed AL-based framework which could be split into three main components starting with the feature extraction where they used a pre-trained ResNet model as the experiment isn't focused on how features are initially identified and therefore, should not impact experimental results. This is then followed with AL where one of the small training sets is used plus samples from the unlabeled image pool through different *'Uncertainty Sampling and Query by Comittee* parameters. This is done for multiple clients where an LSTM is then used to push the data into a federated model where a *FedAvg* is used to train all five clients. The experiment found that the AL-based models achieved approximately an 86% accuracy with QBC vote Entropy which closely matched the accuracy of the fully manually labeled baseline while greatly exceeding the loosely labeled dataset. AL-based models achieved up to 90.2% accuracy in federated learning, compared to 91.3% for the fully labeled baseline and 86.2% for the loosely labeled dataset. When testing the number of clients it was worth noting that the increased clients caused a 6.1% decline in accuracy from 2-8 clients due to the fewer training samples, though more samples and AL could help mitigate this, overall showcasing the effectiveness of AL and Federated modeling.

Alam,in their study of *"Enhancing Image Classification with Federated Learning: A Comparative Study of VGG16 and MobileNet on CIFAR-10"*,investigates the use of Federated Learning (FL) across different research areas and tasks specifically in machine with limited computational resources[9]. Alam (2024) provided an analysis of FL techniques for image classification using the CIFAR-10 dataset, employing popular deep learning architectures such as VGG16 and MobileNet. Initial accuracies without FL were established at 74.5% for VGG16 and 70.8% for MobileNet. Further experiments evaluated how productive various FL algorithms are.Among them are FedAvg,FedProx,FedMA,and FedPAQ and they are measured in terms of their accuracy and data privacy. The author reported that while FedAvg ensure robust privacy, this had led to reduced accuracy of 71.1% for VGG16 and 67.5% for MobileNet. This highlights performance issues in distributed

scenarios that arent identical. In contrast, algorithms such as FedProx, FedMA, and FedPAQ significantly improved accuracy.FedMA delivered highest accuracy improvements which increased the VGG16 performance to 76.3% while MobileNet to 73.1%. FedProx contributed by restricting local model updates to match to the global model, and FedMA improved accuracy through strategic matching and averaging of layers. The author emphasized that algorithm selection critically impacts the effectiveness of FL frameworks which suggests the need for balancing privacy with performance for practical scalable software for consumer use(Alam,2024).

Adnan et al., in their groundbreaking study of *"Federated learning and Differential Privacy For Medical Image Analysis"*investigate the use of Federated Learning (FL) to classify the different type of lung cancers from histopathology images [10]. They utilised Whole Slide Images(WSIs) of lung cancers, specifically targeting two cancer subtypes: LUAD and LUSC.Initially, they extracted image patches from the WSI and used DenseNet to derive feature vectors. Multiple Instance Learning (MIL) was then applied to classify these slides. In their initial experiment, the authors simulated clients to evaluate FL performance under different data conditions, including IID and non-IID scenarios.They discovered that FL significantly outperformed individual hospitals in training alone and nearly matched centralized training accuracy when the number of clients was limited. In a followup experiment involving real hospitals,the model implemented Differential Privacy (DP)) which resulted in a strong privacy outcome. However, due to this there was also a slight decrease in accuracy on external hospitals, mostly due to domain differences among datasets.The study concluded that FL combined with DP can effectively maintain privacy while also achieving high accuracy. Thus this proposed framework is essential to be studied for medical institutions that intend to collaborate without directly sharing patient data (Adnan et al,2022).

The paper *'FEDMM: Federated Multi-Modal Learning with Modality Heterogeneity in Computational Pathology'* by Peng et.al proposed a different framework for federated learning where instead of a focus on single-modal feature extraction the authors propose a model *FedMM* to train models on different data modalities to obtain the benefit of federated learning while preserving data privacy [11]. The FedMM framework utilities a dynamic loss function for training each client model. This is because in the training phase, a global 'pseudo-label' prototype is used which allows the updating of federated feature extractors in the absence of labels caused by privacy constraints. This global prototype acts as a proxy for a given class and is determined by averaging the embeddings within the same class and modality across all clients, noting that each modality has the same number of prototypes as number of classes. The aim is then to minimise the loss functions by using a mix of L2 and BCE loss. To test the effectiveness of their framework, FedMM was evaluated on two datasets, *TCGA-NSCLC* which contains data from patiends with non-small cell lung cancer with two subtypes. The other *TCGA-RCC* included data from patients with renal cell carcinoma (most prevalent type of kidney cancer), with three subtypes. These datasets had two different modalities and two different classes for each set. To test these, the whole slide images were cropped and then had their features extracted using a ResNet-34 model, pre-trained on ImageNet, which then attention pooling is used to aggregate patch-level representations. For Copy Number Variation (CNV) data, preprocessing is done using GISTIC2, followed by feature extraction with a Self-Normalizing Network (SNN) that consists of two hidden layers. The final fully connected layer generates the CNV feature representation. Three clients were used for each test with 100 global rounds used for training and the experiment repeated 20 times. From this study, they found that FedMM surpasses local training and Multi-FedAvg baselines by AUC values of a 2.79% increase and 0.065% increase respectively with

the TCGA-NSCLC dataset. The TCGA-RCC dataset backed this with an increase of AUC values of 2.97% and 7.96% respectively showcasing the practical effectiveness of the FedMM framework.

To deal with high modality and task diversity, *'High-Modality Multimodal Transformer: Quantifying Modality and Interaction Heterogeneity for High-Modality Representation Learning'* was proposed [12]. The authors focus on contexts with a high number of diverse tasks and modalities, proposing a single model capable of adapting to different modality configurations and tasks. To achieve this, they first introduce an approach for quantifying heterogeneity using transfer learning between modalities. This approach enables the reuse of certain model components to process modalities that share similarities and allows for the integration of useful information from different modalities to leverage their interactions. In practice, this involves parameter sharing across similar modalities that exhibit common features or interactions. Their experiment includes a single-model setup with 10 modalities and 15 prediction tasks across five different research areas. HighMMT demonstrates strong overall performance while using only one-tenth of the parameters required for task-specific models.

The paper*"Federated Learning for Enhanced Medical Image Analysis"* by Sanaa Lakrouni et.al explores how Federated Learning(FL) enables collaborative model training across multiple medical institutions while preserving data privacy [13]. The authors emphasize that medical datasets are always different due to variations in imaging equipment, scanning protocols, and etc. This non-IID (non-independent and identically distributed) data distribution poses a significant challenge in Federated Learning, as it can lead to performance degradation. In order to duel with that, the study suggested methods such as Vision Transformers (ViTs) and Self-Supervised Learning (SSL) to enhance FL's robustness in handling this non-IID medical image data. The results from the authors' research indicate that the techniques above can significantly improve classification performance compared to traditional Federated Learning models. Another significant contribution in the field is the integration of Generative Adversarial Networks (GANs) into FL frameworks. Some studies suggested using FL-GAN architectures to address the scarcity of labeled medical images while maintaining data security. These models generate synthetic medical images, which can supply for real datasets and improve model generalization. Blockchain technology is also incorporated in some studies to further secure the environments for Federated Learning, ensuring data integrity and make it hard to be attacked. Privacy-preserving techniques have been another critical aspect in recent FL research. Differential Privacy (DP) and Secure Multi-Party Computation (SMPC) have been explored as methods to enhance the security of FL models. Studies suggest that adding noise to model updates or encrypting data during aggregation can help prevent potential privacy leaks while maintaining model performance. However, a key challenge remains in balancing privacy with model accuracy, as excessive noise can degrade learning efficiency. Also, some studies talk about a novel FL framework designed to handle multi-modal medical imaging data. Unlike conventional FL models that focus on single-modal data, this approach utilizes dynamic loss functions and global pseudo-label prototypes to train models across diverse data modalities. The authors can see from the evaluation on cancer datasets that this method significantly outperforms standard FL techniques in terms of classification accuracy and robustness to data variations. What's more, research on Active Learning (AL) combined with FL has shown the ability of reducing the reliance on fully labeled datasets. By leveraging uncertainty-based sampling techniques, AL-FL models selectively request labels for the most informative data points, minimizing annotation costs while maintaining the accuracy. In conclusion, the paper covers the recent advancements in federated learning for medical image analysis which mainly focused on addressing challenges related to data heterogeneity, privacy, and limited

labeled data. Techniques such as Vision Transformers, self-supervised learning, generative adversarial networks, blockchain security, and active learning have significantly contributed to enhancing the effectiveness of FL in healthcare applications.

3 Project Definition

3.1 Project Questions

- Federated learning can achieve better or equal performance compared to the centralized alternative.
- Different but overlapping sets of classes between clients can improve overall performance in Federated Learning with unimodal clients.
- Multi-modality, with different modalities and classes (partially overlapping) for each client, increases performance in Federated Learning contexts (highly heterogeneous conditions) due to shared knowledge between modalities and classes.

3.2 Aims and Objectives

Aims:

- Develop and implement the core concept of the project - Establish a strong theoretical and practical foundation through research and data collection in phase 1.
- Optimize and test the proposed solution - refine and iterate on the solution, conducting small-scale testing to ensure feasibility and practicality in phase 2.
- Deliver and evaluate the final project outcomes - implement the solution, assess effectiveness, and propose improvements for long-term sustainability.

Objectives:

- Design and implement a federated learning framework that enables decentralized model training without compromising data privacy.
- Optimize the performance of the federated learning system through iterative testing and feedback loops.
- Evaluate the accuracy, efficiency, and security of the federated model compared to traditional centralized approaches.
- Deploy the federated learning solution in a real or simulated environment and analyze its performance based on key metrics.

3.3 Scope

- Implementation Of Federated Learning: Design a federated learning framework to enable model training across multiple devices or clients without sharing the raw data.
- Optimization: Testing and optimizing the model to improve it's accuracy, efficiency and scalability.Also get a comparison with the traditional centralized machine model to access advantages.
- Real-World Applications: Identifying and applying federated learning in a relevant domain, specifically in healthcare area in this research.

- Security and Privacy: Address key challenges related to data privacy, security risks, and compliance with regulations such as GDPR in order to protect it against attacks and data breaches.

4 Methodology

4.1 Data Collection and Datasets

For the current phase, the authors are using the Lung and Colon Cancer Histopathological Image Dataset (LC25000). Our experiments focus on testing the accuracy of image classification on medical image data with the same modality and class for each federated learning client. In later phases, the authors will modify the input data to include different modalities and classes for each client, with overlapping classes. Other medical image datasets will also be introduced and combined for testing the model and the federated learning framework's capability in future stages.

The LC25000 data set that is used in the current phrase is a whole slide images dataset with 25000 color images. 1250 HIPAA compliant and validated images are expanded to 25000 images by left right rotations and horizontal vertical flips, creating a larger dataset that allows the machine learning model to learn from a wide variety of lesion.[14] We choose the LC25000 dataset because it contains a diverse and extensive collection of high-quality, HIPAA-compliant images, ensuring that sensitive patient details are removed or anonymized. This enables our machine learning model to learn from different types of lesion, making it robust and reflective of real-life medical scenarios.

4.2 Exploratory Analysis

1. Exploration of directory structure: After the process of data collection, next step is to inspect the folders. This was done using Python code in a Jupyter Notebook. The code goes through each folder and subfolder to count how many images there are in each of the folders. This would help to show how the images are grouped, usually categorised based on type of diseases or organs involved. In the context of our project,the organs involved are lung and colon and the diseases involved would be cancer. Knowing the folder structure is important as it help us to it make sure images are labelled correctly. It also assists us in finding mistakes or problem with the dataset in the early stages of our project. Counting the images in each folder can also tell us if some classes have more images than others. This would help in informing whether transformation to balance the data before moving to the next step.
2. Class Distribution Analysis: Next, the authors will perform class distribution analysis where to look at how many images there are in each class. For this step, in Python code a count of the number of image that each class has by using recursion method. After finishing the counting, a Matplotlib library was used to generate the bar chart. The bar chart would help us to visualise clearly and show if image between classes are the same. If one class has too many images and other has only few, it is called imbalance. This is important for later steps as a model can learn from imbalanced data which would lead to inaccurate outcome in its prediction. To fix imbalance the authors might have to add more images to small classes or remove some images from bigger ones.
3. Viewing Sample Images: After examining the dataset structure, the authors looked at random sample images. The authors will look through the images from different folders, then

the authors convert images from BGR to RGB format. RGB format would make the image colours look more visible to the naked eyes when the authors plot them. Plotting these images help us check their quality. We can see if images are labelled correctly or clearly. This step let us to identify problem such as wrong labels or corrupted image files. Finding these problems early on would help us fix before doing further analysis. Viewing images also show us how dataset looks visually. These steps may include actions such as image resizing, augmentation, or normalization. This is done to help us in producing a model that is more robust and reliable.

4. Image Size and Corrupt File Detection: In this section, image sizes and corrupt files were checked. The authors opened each image individually and record the width and height of every image. By doing this, minimum, maximum, and average dimensions of the images were inferred. Knowing these sizes would inform us in deciding if standardisation or resizing of image sizes are needed. Resize images make analysis easier as all images would be uniform. At the same time, the code checks if some images fail to open and those images that are unable to be opened are flagged as corrupted. Finding corrupt images prevents issues from arising later in analysis or modeling stages. This step helps make sure our dataset integrity.
5. Colour Distribution Analysis: Final step of the EDA process is checking the colour distribution of the images. We implement plot histograms in RGB and HSV colour spaces to achieve this. RGB histogram shows the distribution of pixel intensity for red, green, and blue colour channels. HSV histogram gives additional information about hue, saturation, and value channels. We look at histograms of random images from each class to find unusual colour patterns. This process enables us to further understand if the images were captured with different lighting conditions or imaging methods. Obtaining information about these colour issues help us decide if images need preprocessing steps. Consistent images help machine learning models become more accurate and reliable in later analysis

4.3 Proposed Framework

The methodology will be divided into three distinct phases, each addressing the project questions. The critical points relate to Federated Learning performance, introducing different classes for each client, and similarly, introducing different modalities for each client. These three critical points will be introduced incrementally in each phase. In this way, the three phases are dependent on each other.

Flower Framework will be used to host both the clients and the central server locally. This framework has been chosen since the entire exercise is a simulation. In the real world, the clients could be hospitals and private radiology.

Performance will be evaluated using classic evaluation metrics for this setting: (1) Accuracy and ROC AUC of centralized versus Federated Learning, (2) Accuracy ROC AUC between each client and the global model, and (3) Class-specific accuracy. Accuracy is selected as the standard metric of performance evaluation and ROC AUC as a more comprehensive and robust metric for more inconsistent scenarios.

In the initial phase, the team will focus on exploring possible options to improve the performance of the federated learning framework. In later stages, the team will increase the complexity and try to simulate real-life scenarios by incorporating additional modalities and classes for each client in the framework. This is shown in Figure 1 and outlined below.

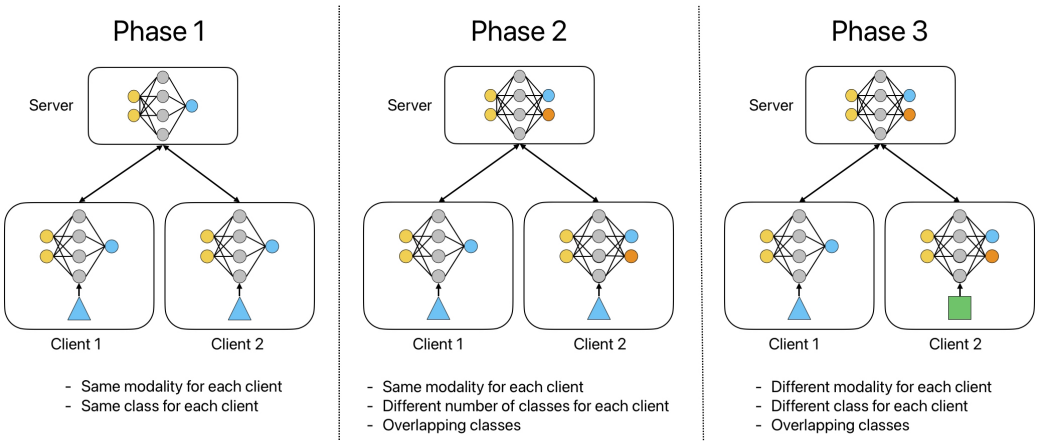

Figure 1: Project Phases

4.3.1 Phase 1

In Phase 1 (current phrase), the proposed framework is constructed with all clients possessing the same modality and image classes. Our objectives in this phase are to establish the federated framework's structure and to explore suitable models and aggregation methods. The team plans to experiment with machine learning models for image classification, using a convolutional neural network (CNN) as the primary model. We will also explore lightweight versions of VGGNet and ResNet, which are models built on the CNN architecture.

In our framework, the authors are using centralised federated learning, which means in our framework, the authors have a central server to orchestrate the client nodes in the learning process. For the federated learning aggregation methods, the authors will use the federated averaging as the federated learning aggregation method to aggregates these clients' model parameter by computing the weighted average.

In this phase, the author will design experiments to test the robustness of the models and federated learning method using selected evaluation metrics. We will evaluate the model's performance on both Independent and Identically Distributed (IID) and Non-IID data inputs.

The DirichletPartitioner from the Flower library will be used for data partitioning across different clients. This partitioner allows us to control data heterogeneity by adjusting the alpha parameter, enabling the simulation of both IID and Non-IID conditions. A higher alpha value results in the uniform data distribution among different clients, while a lower alpha value creates imbalanced distribution.

The authors aim to have 2-3 clients trained locally with at least 3 iterations for the framework to observe the improvement of different combination of models and aggregation methods.

4.3.2 Phase 2

This phase consists of a Federated Learning experiment with clients that still share the same number of modalities, but different classes (labels). To evaluate this, a basic setting will be used: two or three clients with same modality, WSI. The set of labels for each client will be overlapped to explicitly set a relation between clients and meet conditions for multi-task learning.

Public available datasets of medical images will be used in this experiment. The alternatives consist of splitting one or more datasets into two clients with different classes but the same modality, or using different datasets and assigning them to a each client. These datasets must have overlapping classes. Whole slide images (WSI) will be considered for this phase. Classic techniques for analyzing data will be used because the data type is WSI: (1) color deconvolution: This technique can separate color channels into their respective components for more precise analysis, and (2) Image Histogram: This will be used to analyze the histogram of pixel intensities.

Since the only difference with respect to the previous phase is the class difference between clients, a single model will be used for both local and global tasks. The difference lies in the upstream section of the classifier, which changes depending on the classification task performed by each client. This means that local models are trained with local data, but global aggregation is performed only for the downstream layers. This model is selected because the team will be dealing with the same modality and overlapping labels, so there must be some common features and representations. Basic CNN and pre-trained architectures will be tested.

4.3.3 Phase 3

For the final deliverable (Phase 3), the proposed framework would be built in a way that each client classifies medical images irrespective of the modality. The scope of this project is limited to one disease. Ideally, the client will be hosting data with different modalities and different classes. The classes can overlap with other clients. Hence, the final deployed aggregated model would enable the clients to classify multiple-modality images for one disease, as of now the decided disease is Lung and Colon Cancer.

5 Resources

5.1 Hardware and Software

Throughout the project, no hardware devices will be deployed. The tutor advised the authors to consider the project as a simulation exercise. Hence, to deploy clients and a central server virtually, the authors will be using the Flower Framework. This framework works well in a Python Environment. The only hardware, in use, will be the laptop hosting the local environment for performing the Federated Learning.

In terms of software, the authors will leverage VSCodium ver. 1.98.2. A Python ver. 3.12.7 environment will be set up to deploy the Flower Framework. Flower will help to perform Federated Learning i.e. deploy clients and a central server. The environment will be leveraged to perform image classification on different modalities like US, CT, XRAY, and MRI. Various other libraries, like PyTorch and NumPy, will be used to supplement the Flower Framework.

5.2 Materials

Since the project is a simulation exercise, no external materials will be used. No virtual machines, to represent the clients nodes, or a server will be deployed. The authors will be using the Flower Framework, which deploys the clients and central server virtually in real-time, that is when the client and server apps are initiated. Essentially, the framework, in use, will save on a lot of processing power since hosting data on each client node, and then training the model would require a decent amount of CPUs.

For Phase 1, the authors will not be storing the data locally as it requires a lot of RAM to pre-process and then export the dataset to each client node. Instead, the authors will import the data in real-time directly through in-built APIs. For Phase 2 and 3, the authors will be using multiple datasets. Hence, the authors will be pre-processing and joining it into one dataset. This dataset will be hosted on cloud and pulled to the local environment in real-time i.e. when the clients nodes and central server are initiated. Therefore, for Phase 2 and 3, the cloud service will be the external material.

Besides this, no other external materials will be used throughout the project.

5.3 Roles and Responsibilities

Table 1 describes each role in the project and the team members assigned to each role. This outlines the distribution of roles until the moment of this proposal. It is worth noting that since the team is following some Agile principles, roles can be performed by any team member and it will be flexible during the project.

Role Description	**People**
Project Manager – Oversees the entire project, ensuring deadline and team coordination.	Rajiv
Data Scientist – Responsible for data analysis, feature engineering, and model development.	Eshan, Ngai Chun, David
Software Engineer – Implements and maintains the code, including model deployment.	Eshan, David, Ngai Chun
Data Engineer – Designs and maintains data pipelines, ensuring efficient data storage and retrieval.	Akmal, Rajiv
Reporting and Documentation	Rajiv, Ngai Chun, David, Akmal, Eshan, Simon

Table 1: Roles and Assigned People

6 Expected Outcome

6.1 Project Deliverables

To analyse the effectiveness of the proposed multi-modal federated learning framework, the authors have split the project into the four main phases as seen in Figure 1:

1. '*Phase 1*' of the project acts as a trial for the authors to get better acquainted with federated learning models and design the basic framework that will reach the minimal viable product of phase 2. This phase consists of obtaining one dataset where there is only one type of image modality i.e. 'X-Rays' and only one singular class. This singular dataset is then split randomly into multiple clients where each client is trained locally with a machine learning model, where parameters and results are outputted into a server which parameter transformation is applied and trained back to each model. This deliverable acts as a basis for our MVP model as seen in *Phase 2*.

2. The second phase of the project '*Phase 2*', is considered the minimum viable product (MVP) for this project. In this deliverable, the framework is changed so that each client will have a different dataset with a different number of classes (diseases) between them. This means that potentially one dataset might include one type of disease with other not including it, with the modality still remaining constant. The models results will then be tested against the baseline of phase 1 to test the effectiveness of the proposed framework.
3. '*Phase 3*' further expands on the MVP where not just the datasets and classes are different but also the modalities of the datasets. This means that clients 1 and 2 could have X-Rays and CT-scans while client 3 could have X-Rays and MRI scans as an example. This deliverable adds another level to the model, more accurate to the issue faced by professionals within the industry.
4. '*Phase 4*' aims to demonstrate the findings and results of the first three phases and the project overall. A final report and a presentation to the client. The report will be written as a research paper with a video presentation by the group members plus 'Questions and Answers' session to answer any questions on the study for the client to have a full understanding of the study.

6.2 Implications

Overall, the aim of the project deliverables is to produce a model that could be potentially used in the advancement of modeling practices within areas of business that require confidentiality and security. This study would also outline the effectiveness of a federated model where different modalities are present within the training data and give an indication on whether or not it is an effective method of prediction modeling and if the increase or decrease in accuracy is worth bundling the different modalities together instead of implementing different models for each type of class and modality. Furthermore, these findings could encourage wider use of federated-learning in real world applications and lays the groundwork for further research into federated learning optimization, addressing open challenges such as communication efficiency, system heterogeneity, and etc.

7 Proposed Schedule and Key Milestones

To achieve the set out project deliverables of the four phases over the 15 week period, the Gantt Chart as seen in Figure 2 outlines the estimated project timeline:

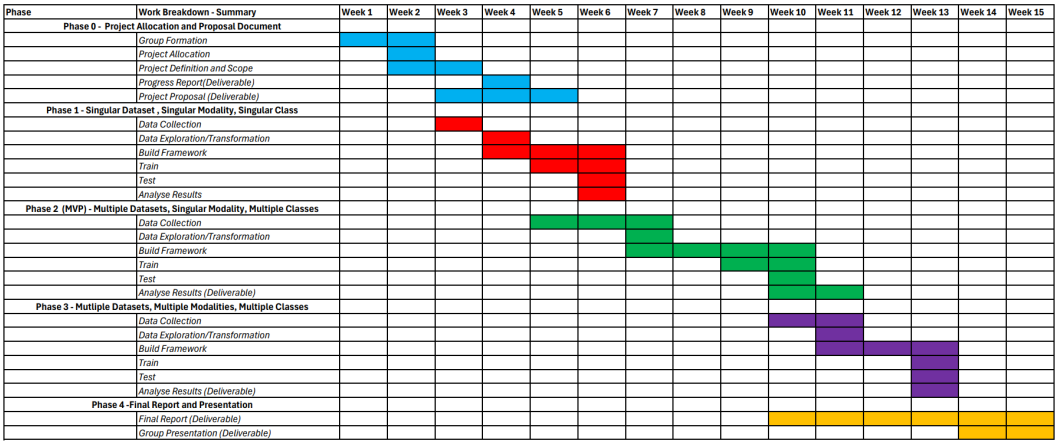

Figure 2: Project Expected Timeline

The Gantt Chart is split into the 4 main phases of the project, with an additional 'Phase 0' which occurred prior to the start of the project and included aspects such as group formation and project definition. Phase 1 to 3, all have a similar work breakdown summary, including:

1. *Data Collection*: This involves searching for potential datasets that can be used for the various phases of the project. For this study, it is important that this task is done in an efficient matter as it could act as a potential blocker for the remaining of the phase but also could impact the amount of time required for 'data exploration and transformation.
2. *Data Exploration and Transformation*: Once the data has been shortlisted, exploration of the dataset needs to be performed and necessary transformations such as image cropping will need to be performed.
3. *Build Model/Framework*: While the previous step is occurring, the building of the proposed framework as well as adjustments to later phases will occur. This will take the longest amount of time as debugging bugs can act as a blocker for later phases.
4. *Train and Test*: Training and Testing the model will then happen, giving approximately two weeks per model to allow for optimisation of the model and achieve higher accuracies and better results.
5. *Analysis*: The final step of Phases 1-3 is to analyse the results and prepare them for when they will be inserted into the Final Report, as seen in Phase 4.

References

[1] J. Pei, W. Liu, J. Li, L. Wang, and C. Liu, "A review of federated learning methods in heterogeneous scenarios," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 3, pp. 5983–5999, 2024.

[2] L. Che, J. Wang, Y. Zhou, and F. Ma, "Multimodal federated learning: A survey," *Sensors*, vol. 23, no. 15, p. 6986, 2023.

[3] W. Liu, Y. Zheng, Z. Xiang, Y. Wang, Z. Tian, and W. She, "An efficient federated learning method based on enhanced classification-gan for medical image classification: An efficient federated learning method," *Multimedia Systems*, vol. 31, no. 1, 2025.

[4] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can ai help in screening viral and covid-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[5] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images," *Pattern Recognition Letters*, vol. 138, pp. 638–643, 2020.

[6] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. A. Maadeed, S. M. Zughaier, M. S. Khan, and M. E. H. Chowdhury, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Computers in Biology and Medicine*, vol. 132, 2021.

[7] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, "Active learning based federated learning for waste and natural disaster image classification," *IEEE Access*, vol. 8, pp. 208 518–208 531, 2020.

[8] Y. Guo, Y. Liu, E. M. Bakker, Y. Guo, and M. S. Lew, "Cnn-rnn: A large-scale hierarchical image classification framework," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 10 251–10 271, Apr 2018.

[9] E. E. Alam, "Enhancing image classification with federated learning: A comparative study of vgg16 and mobilenet on cifar-10," *arXiv*, 2024.

[10] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," *Scientific Reports*, vol. 12, p. 1953, 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-05539-7

[11] Y. Peng, J. Bian, and J. Xu, "Fedmm: Federated multi-modal learning with modality heterogeneity in computational pathology," 2024. [Online]. Available: https://arxiv.org/abs/2402.15858

[12] P. P. Liang, Y. Lyu, X. Fan, J. Tsaw, Y. Liu, S. Mo, D. Yogatama, L.-P. Morency, and R. Salakhutdinov, "High-modality multimodal transformer: Quantifying modality & interaction heterogeneity for high-modality representation learning," *arXiv*, 2023. [Online]. Available: https://arxiv.org/abs/2203.01311

[13] S. Lakrouni, S. Bah, M. Sebgui, N. Gupta, A. Castañeda, and C. Enea, "Federated learning for enhanced medical image analysis," *Networked Systems*, vol. 14783, pp. 157–170, 2024. [Online]. Available: DOIordirectlinkifavailable

[14] A. A. Borkowski, M.-M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (lc25000)," *arXiv preprint arXiv:1912.12142v1*, 2019.

8 Appendix

8.1 Acknowledgement of AI Usage

Part A: Have you used AI tools in the completion of this assignment? If your answer to this part is "No", you can leave the following Part B and Part C as blank.

Yes

Part B: What automated writing or generative AI tools you have used in the completion of this assignment? Clearly state the name(s) of the tool(s) and including a link to each tool.

Microsoft Co-Pilot: https://copilot.microsoft.com/

Part C: How have you used automated writing or generative AI tools in the assessment?

For this project proposal, the team had utilised the Microsoft Co-Pilot GPT offered by the University of Sydney. Through this tool the team had used the engine to fix up the grammatical errors within the document as well as paraphrase certain sections such as the literature review and methodology sections. Co-Pilot was also used for a base understanding on federated learning and the flower-framework understanding where parts were rephrased for this document.
