A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans

By

SHAH NAWAZ

221-35-874

FINAL YEAR THESIS REPORT

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Department of Software Engineering

Supervised by

Ms. Fatama Binta Rafiq

Lecturer (Senior Scale)

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

December 24, 2025

APPROVAL

This thesis titled on **"A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans"**, submitted by **Shah Nawaz (ID:221-35-874)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. A. H. M. Saifullah Sadi
Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

-- **Internal Examiner 1**

Dr. Rubaiyat Islam
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Dr. Md. Abdul Kader
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Nuruzzaman Faruqui
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Md. Mostafiz Khan
Managing Director
Tecognize Solutions Limited

Department of Software Engineering

Faculty of Science and Information Technology

Supervisor Approval Form

Fall 2025	B.Sc. In SWE	Campus: DSC

Student Name	Student ID
Shah Nawaz	221-35-874

Project/Thesis Information	
Thesis Title	**A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans**
Type of work	Artificial Intelligence / Deep Learning

Supervisor information	
Supervisor Name	Ms Fatama Binta Rafiq
Supervisor Initial	FBR
Completed Credit till now	139
How many credits in this semester	6
Amount (Due)	-520.00
Supervisor Consent	☐ Yes ☐ No

Supervisor Signature : Fatema

SUPERVISOR’S DECLARATION

This is to certify that the thesis entitled “A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans” submitted by Shah Nawaz (ID: 221-35-874) to the Department of Software Engineering, Daffodil International University has been carried out under my supervision.

I confirm that this work is original to the best of my knowledge and belief, and that the thesis has not been submitted, either in part or in full, to any other university or institution for the award of any degree or diploma. I also certify that the student has fulfilled the requirements necessary for the submission of this thesis.

(Supervisor’s Signature)

Full Name: Ms Fatama Binta Rafiq

Position: Lecturer (Senior Scale)

Date: 24.12.2025

STUDENT'S DECLARATION

I hereby declare that this thesis, entitled "A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans," is my original work, submitted to the Department of Software Engineering at Daffodil International University, and has not been submitted for any degree or diploma at this or any other university.

All sources of information and data used in this thesis have been duly acknowledged. I further declare that the work reported herein embodies the results of my own research and complies with the ethical norms and regulations of Daffodil International University.

Shah Nawaz

(Student's Signature) Full

Name: Shah Nawaz

ID Number: 221-35-874

Date: 23.12.2025

ACKNOWLEDGEMENTS

I would like to say thank you to everyone and everything that was involved in the. fulfillment of this dissertation study.

To begin with, we warmly thank and appreciate the Almighty because of his divine. blessing to enable us to accomplish Final Year Design Project(FYDP). successfully.

I would like to thank the employees and the management of Bokkobidi Hospital in. Dhaka, Bangladesh, on their part, to help them get real-time clinical data and to collaborate. acquiescing in my orating with me on this privacy-saving research work. Those in charge were the individuals. to motivate them to participate in this study were their zeal to improve diagnostic tools and at the same time safe them. guarding patient privacy. I would like to thank the radiologists of the Bokkobidi Hospital. due to their hard work labeling and validating the clinical dataset.

I would like to thank the people who have provided me with the LIDC IDRI dataset. This implies that federated learning studies can be reiterated. The open the commitment of science community to share the resources spurs the pace of scientific advancements. This study was approved by the ethics committee, and it conformed to the institution standard. for managing personal data. The privacy-saving methodology will make this study up- supports the right of people to privacy and moves medical AI to the good of patients. globally.

And last, but not the least, there is the ever-present support and forbearance of my parents.

A Privacy-Preserving Federated Learning Approach for Segmenting Lung Abnormalities in Lung CT Scans

SHAH NAWAZ

Thesis submitted in fulfillment of the requirements
for the award of the degree of

Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

December, 2025

ABSTRACT

The detection of lung cancer at an early stage requires a lot of proper and timely analysis of the computed tomography (CT) images. Nevertheless, the development of high-performance diagnostic models often requires access to large data volumes, gathered in different healthcare centers. This data sharing cannot always take place in the actual clinical context due to the strict patient privacy regulations. This study will introduce a two-phase federated learning-based system that takes privacy into account in the assessment of lung cancer in a real hospital. The suggested system enables the involved hospitals to train the deep learning models together without having to share the uncoded CT scans. In this regard, sensitive data about patients are retained in-house within local servers, and model performance is optimized together. The framework is designed in accordance with the currently available data protection regulations like HIPAA, GDPR, and overall practices of hospital data governance. Stage 1 adopts a federated 2D U-Net trained across two simulated hospital sites using the LIDC-IDRI dataset. The global segmentation model yields a Dice score of 0.8568%, precisely delineating lung nodules under heterogeneous, non-IID data distributions. Stage 2 adopts the segmentation-guided lung regions to train a hybrid ResNet50–Vision Transformer classifier for normal, benign, and malignant cases. This results in a federated classifier achieving 98. This work verifies that multi-client FL can preserve diagnostic accuracy compared to centralised training while avoiding inter-hospital data transfer. The proposed framework provides a clinically viable direction for secure AI-assisted lung cancer screening and serves as a scalable foundation for future privacy-preserving medical imaging applications.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Lung cancer is widely recognised as one of the leading causes of cancer-related mortality worldwide, and survival outcomes remain strongly dependent on early and accurate diagnosis [1]. Computed Tomography (CT) imaging has become a primary modality for detecting, characterising, and monitoring pulmonary nodules that may correspond to early-stage malignancies or benign abnormalities [2]. However, manual inspection of full CT volumes is labour-intensive, prone to inter- and intra-radiologist variability, and difficult to scale in resource-constrained healthcare systems [3].

Deep learning has demonstrated outstanding performance in lung nodule detection, segmentation, and malignancy classification, often approaching or surpassing expert-level accuracy in curated datasets [4]. U-Net and its variants have become the de facto standard for medical image segmentation due to their strong multi-scale representation learning capabilities [5, 6, 7]. For classification tasks, architectures such as ResNet and Vision Transformers (ViTs) are widely adopted because of their ability to learn hierarchical and global feature representations [8, 9, 10, 11]. Even with advancements in the recent past, the most effective models are still predominantly trained centrally, where imaging data from various locations are combined at a single point of storage. This approach is increasingly difficult to justify in real-world clinical environments due to strict privacy legislation, corporate data governance regulations, and patient concerns regarding data disclosure [12]. Healthcare facilities must comply with stringent regulations, such as HIPAA and GDPR, which place severe restrictions on the collection, storage, and transmission of patient data. Transferring raw CT scans to a central server via hospital PACS systems introduces risks of data breaches, unauthorized access, and even re-identification, even when explicit identifiers have been removed. Furthermore, once raw data is stored in a central repository, it remains vulnerable to manipulation or model-based attacks designed to re-identify individuals [14, 15]. Consequently, a tension has arisen between the necessity of utilizing large, diverse datasets to develop robust deep learning models and the ethical and legal mandate to keep patient information on-site. Federated Learning (FL) has emerged as a potentially beneficial paradigm to resolve this tension. FL enables multiple data owners (e.g., hospitals) to jointly train a shared global model without transferring raw data. In a typical FL system, each client performs local training on its own dataset and transmits only model parameters or gradients to a central aggregation server. The server aggregates these updates—typically using the Federated Averaging (FedAvg) algorithm—and distributes the updated global model back to the participating clients. This approach preserves data locality and significantly minimizes privacy risks while leveraging the heterogeneity of multi-institutional data [12, 19, 20]. Lung cancer imaging is a particularly compelling application for FL because CT datasets are inherently heterogeneous: hospitals differ in scanner hardware, acquisition protocols, reconstruction kernels, patient populations, and annotation practices [21, 22]. Federated training can harness this heterogeneity to improve model generalisation across centres without requiring central data pooling [23]. However, most existing FL studies in medical imaging focus on single-stage tasks such as standalone classification or segmentation [19, 24]. Multi-stage clinical workflows — such as lung nodule segmentation followed by abnormality classification — have received comparatively less attention, and the potential synergy between segmentation and classification in a federated setting remains underexplored. This thesis is motivated by the need to design a privacy-preserving, clinically meaningful, and technically robust pipeline that closely reflects real lung cancer diagnostic practice. In clinical workflows, suspicious structures are first localised and segmented; then a decision is made regarding whether the abnormality is benign, malignant, or within normal limits. To address this need, the proposed work develops a two-stage federated learning pipeline in which a U-Net-based segmentation model is trained in Stage 1, followed by a hybrid ResNet50 + ViT classifier in Stage 2 that operates on segmentation-guided CT slices [5, 8, 9, 25]. The pipeline is explicitly designed to respect privacy constraints—ensuring all CT data remain local while exchanging only model weights—while maintaining high performance across heterogeneous, multi-institutional environments [16].

1.2 Problem Statement

Conventional deep learning solutions to lung CT analysis generally assume access to a large centralized dataset that combines images from multiple clinical sites. In practice, such centralization is seldom feasible due to:

- **Privacy of Data and legal limitations:**Laws prohibit the exchange of identiportable or even de-identified medical images across institutions.
- **Policies on Data Governance in Institutions:** There are numerous hospitals that would not want to. Either because of legal liability or because of infrastructure limitations, push data to external servers.
- **Technical and Operational Limitations:** The transfer of large volumetric data, as thousands of CT scans are, is bandwidth-intensive, and complex to run.

Concomitantly, strong nodule segmentation of the lungs and classification of abnormalities. Multi-institutional training data is necessary in models to be able to generalize across real-world. populations. Single-center trained models have been overfitted to local acquisition protocols and cohorts of patients, which restricts clinical influence.

Moreover, most medical imaging-based FL methods treat segmentation and classification tasks separately. End-to-end federated pipelines are lacking, where Stage 1 is tasked with generating lesion-focused representations—segmentation masks that explicitly guide Stage 2 classification. Such a separation results in the following two key issues:

P1 – Privacy vs. Performance Tradeoff: How can we design a multi-institutional training pipeline that can leverage rich CT studies from several hospitals without compromising patient privacy or transferring raw images?

P2 – Joint Segmentation–Classification in FL: How can a two-stage FL pipeline be constructed such that segmentation outputs guide classification while both stages remain fully federated and privacy-preserving?

This thesis addresses these problems by designing, implementing, and evaluating a two-stage federated learning framework that performs lung/nodule segmentation on the LIDC-IDRI dataset and abnormality classification (Normal, Benign, Malignant) on a real clinical dataset, across two simulated hospital clients equipped with NVIDIA RTX 4060 GPUs.

1.3 Research Questions

According to the problem statement, this thesis will focus on the following Research Questions (RQs):

- **RQ1:** Can a federated 2D U-Net provide competitive lung nodule segmentation performance on distributed LIDC-IDRI data without the need of centralizing the CT scans?
- **RQ2:** Can a segmentation-guided ResNet50 + Vision Transformer classifier, trained in a federated manner on real clinical CT slices, precisely discriminate between normal, benign, and malignant lung abnormalities while maintaining data locality?
- **RQ3:** How does the performance of the global federated model compare to the individual client models for both segmentation and classification in terms of non-IID data distributions?
- **RQ4:** What is the communication overhead and convergence behavior of the proposed two-stage federated pipeline, and is it acceptable for realistic multi-hospital deployments?
- **RQ5:** To what extent does the proposed framework preserve privacy, and how can it be extended to incorporate other mechanisms, such as differential privacy or secure aggregation?

These questions guide methodological choices, experimental design, and analysis throughout this work.

1.4 Objectives of the Study

The above research questions are addressed through the following objectives in this thesis:

- **Design a Two-Stage Federated Pipeline**
  - Develop a complete end-to-end pipeline consisting of a Stage 1 federated segmentation model, using a 2D U-Net architecture for lung/nodule segmentation on LIDC-IDRI.

- Stage 2: Federated classification model based on ResNet50 backbone fused with a Vision Transformer (or ResViT) for the classification of CT slices into Normal, Benign, and Malignant classes.

- **Privacy-Preserving Data Handling**
  - Ensure that all CT images and labels remain on local client machines and only model parameters get exchanged with a central FL server, in accordance with HIPAA-like privacy principles.
- **Design a Segmentation-Guided Classification Approach**
  - Utilize Stage 1 segmentation masks for creating segmentation-guided inputs to Stage 2 classification, such as ROI-focused crops or mask overlays, that enhance diagnostic performance and model interpretability.
- **Global and Client-Specific Performance Evaluation**
  - Perform a systematic evaluation of the performance of segmentation and classification for the global federated model aggregated via FedAvg.
  - Evaluate individual client models to represent local performance, quantified by Dice coefficient, IoU, accuracy, F1-score, and confusion matrices.
- **Quantify Federated Learning Overhead and Stability**
  - Round-wise training curves and communication cost in terms are to be analyzed. of model weight transfers to realize the tradeoff of performance, communication burden, convergence, and communication burden.
- **Critically Analyze Privacy and Practical Deployment**
  - Determine the suitability of the proposed system to meet the privacy and deployment requirements in a realistic multi-hospital environment. Find avenues for incorporating cutting-edge privacy-conserving systems.

1.5 Contributions of the Thesis

The contributions of this thesis can be summarized as follows:

- **A Two-Stage Federated Learning Framework for Lung CT Analysis:** This thesis results in a new integrated FL pipeline that simultaneously deploys a federated U-Net segmentation model and a federated ResViT classification model on two simulated hospital clients. This configuration is a realistic and multi-stage workflow in a clinical setting that considers privacy limitations.
- **Guided Federated Classification of Real Clinical Data by segmentation:** The paper is the first to show that Stage 1 segmentation masks can be used to direct Stage 2 classification on a real multi-class CT dataset (Normal, Benign, Malignant) in the federated environment to more effectively focus on areas of interest in the lungs, and it yields higher interpretability.
- **Extensive Assessment of Global vs. Local Models:** Practically, the work introduces a thorough comparison of global and customized models in segmenting and categorizing data with non-IID data distributions, therefore, measuring the advantages and disadvantages of federation.

- **Use of an Implementable FL Pipeline:** A full implementation- It has been tested in Flower on FL orchestration and PyTorch with two NVIDIA RTX 4060 GPUs. It contains preprocessing of data, federated training code, evaluation code, and explainability analysis through Grad-CAM.
- **Communication Overhead and Privacy Considerations Analysis:** This thesis measures the communication cost in model weight transfers/round and also talks about how the architecture ensures patient privacy as the raw CT images remain within their institutional locality, and also talks about future directions on ensuring enhanced security and privacy.

The contributions will address the gap between the theoretical studies on FL and their practical use in privacy-compliant applications in the analysis of lung cancer-related CT images.

1.6 Scope and Limitations

The proposed system is in no way exhaustive, but it is purposely narrowed to allow making it feasible within the limits of a bachelor-level thesis and the existing computational resources.

Scope

- Specifically, the work is concerned with 2D axial CT slices as opposed to complete 3D volumes. In case of segmentation, LIDC-IDRI slices are derived where as classification slices are from a real clinical dataset.
- We have a total of 2 clients (simulated hospitals), and each of them is hosted on a NVIDIA RTX 4060 GPU. This is a bare minimum multi-institution FL situation that is still implementable.
- The classification problem has only three categories namely normal, benign, and malignant. There is no longer granular staging or subtype categories (adenocarcinoma vs squamous cell carcinoma) are taken into account.
- The FL implementation uses FedAvg as the primary aggregation policy. More sophisticated Such approaches as FedProx, FedNova, or personalized FL are that of the future work.

Limitations

- **Limited Number of Clients:** Two clients only are utilized to test the system. While this model suffices in a core FL behavior (non-IID data, local vs. global performance); realistic deployment will usually incorporate significantly more institutions.
- **Non-IID characterization:** Although the clients are set with unique data. The data can be deemed non-IID; it does not entirely model everything in the real world.

- **Lack of Formal Differential Privacy or Secure Aggregation:** The existing implementation deals with locality of data–no images are shared–but not. Add different privacy, homomorphic encryption, or secure multi-party computation. Therefore, the privacy assurances are solid on the data level but informally measured on a parameter level.
- **Single Hardware Configuration:** The experiments are executed on a specific hardware setting, namely two RTX 4060 GPUs. Performance and latency may vary in different configurations or in real hospital scenarios with diverse infrastructure.
- **Dataset Bias:** There may be biases related to acquisition protocols, patient demographics, and annotation practices in both LIDC-IDRI and the clinical classification dataset. Therefore, generalization based on external data should be viewed with caution.

These limitations are acknowledged to maintain transparency and guide future extensions of this work.

1.7 Thesis Organization

The rest of the thesis is organized as follows:

Chapter 2 – Literature Review provides an extensive overview of state-of-the-art lung cancer detection with CT imaging, deep learning-based methods for segmentation and classification, basics of federated learning, related FL applications in medical imaging, and AI techniques to preserve privacy. The chapter concludes with a thorough gap analysis motivating the proposed two-stage federated framework.

Chapter 3 – Methodology provides a comprehensive description of the proposed system. It describes the overall two-stage architecture, the federated learning framework, dataset preparation, U-Net and Res-ViT model design, a segmentation-guided preprocessing strategy, the FedAvg aggregation mechanism, the experimental setup, hardware and software, privacy, and security considerations.

Chapter 4 – Results and Analysis provides experimental results at both the segmentation and the classification stages, comprising per-round training curves, per-client and global metrics, confusion matrices, and Grad-CAM visualizations. The analysis of communication costs and a further study of the effect of non-IID data distribution on convergence and performance are also presented in this chapter.

Chapter 5 – Discussion puts the findings into perspective with existing literature, discusses privacy-utility tradeoffs, assesses the practicality of deploying the system in real multi-hospital environments, and analyzes limitations and threats to validity.

Chapter 6 – Conclusion summarizes the work done by highlighting significant contributions, reflecting on the broader implications for privacy-preserving healthcare AI, and outlining promising directions for future research, including integrating more robust privacy mechanisms and scaling to more clients.

The **References** section lists all works cited in this paper in IEEE format. Appendices contain supplementary architectural details, hyperparameter settings, code snippets, documentation of ethics considerations, and extra results.

Chapter 2

Literature Review

This chapter reviews previous work relevant to the proposed privacy-preserving two-stage federated learning framework for lung nodule segmentation and abnormality classification from Computed Tomography (CT) scans. The discussion covers lung cancer detection and CT imaging, deep learning approaches for medical image segmentation and classification, the fundamentals of federated learning, applications of federated learning in medical imaging, privacy-preserving techniques in healthcare, and the research gaps that motivate the proposed methodology. A brief summary of recent work from 2023–2025 is also included.

2.1 Lung Cancer Detection and CT Imaging

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, with survival strongly dependent on the stage at diagnosis. Early-stage disease is often asymptomatic and may only be detected incidentally or through screening programmes. Low-dose CT (LDCT) screening has been shown to reduce lung cancer mortality in high-risk populations by enabling the detection of small nodules at earlier, more treatable stages [15].

CT provides high-resolution volumetric information about the lung parenchyma, airways, vasculature, and surrounding thoracic structures. Compared with chest radiography, CT is far more sensitive to small nodules, ground-glass opacities, and subtle textural abnormalities. A single CT study may consist of hundreds of axial slices reconstructed at sub-millimetre resolution [15]. While this rich information improves diagnostic capability, it also increases the reading workload and introduces variability among radiologists.

Computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems have therefore been explored extensively to assist in the identification and characterization of suspicious lesions. Public datasets have played a key role in driving research in this area. The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) provides thoracic CT scans with lesion-level annotations from up to four radiologists, including nodule locations, binary masks, and subjective ratings such as malignancy and spiculation [13]. LUNA16 challenge is based on LIDC-IDRI and aims at the auto-detection and segmentation of pulmonary nodules [14]. These resources permit algorithmic benchmarking of segmentation and classification algorithms.

Even with these efforts, the bulk of clinical CT datasets are kept in silo across the various institutions because of privacy regulations, local governance policies and re-identification concerns despite de-identification takes place [63]. In practice, radiologists do not only identify nodules, but also determine the probability of a nodule being malignant by combining the results of the imaging with the pathology, risk factors, and the clinical history.

Developing AI systems that would embrace this multi-stage reasoning would ideally need large diverse datasets that comprise high-quality segmentation scores, malignancy scores, and actual results. The conflict between the richness of data demanded and the rigorous privacy demanded has been the impetus to consider federated and privacy-secured learning paradigms.

2.2 Deep Learning for Medical Image Segmentation

2.2.1 Classical and Early CNN-Based Segmentation

Before deep learning became dominant, segmentation of CT images relied on handcrafted techniques such as thresholding, region growing, active contours, level sets, and graph-based methods. These approaches can perform well on simple, high-contrast structures, but they struggle with heterogeneous nodule appearances, variable acquisition protocols, and imaging artefacts. Lung nodules may be attached to vessels or pleural surfaces and can exhibit a wide range of shapes and intensities, making rule-based methods brittle.

The introduction of convolutional neural networks (CNNs) allowed models to learn hierarchical features directly from data. Early CNN-based segmentation approaches frequently used patch-wise classification or sliding-window schemes. Although these improved robustness compared with handcrafted techniques, they were not inherently designed for dense pixel-wise prediction and often suffered from coarse boundaries and high computational cost.

Fully convolutional networks (FCNs) addressed some of these limitations by replacing fully connected layers with convolutional layers and using upsampling layers to restore spatial resolution [6]. FCNs preserve spatial correspondence between input and output but naïve upsampling tends to produce smooth, imprecise boundaries, which is problematic for small objects such as pulmonary nodules.

2.2.2 U-Net and Its Variants

The U-Net architecture, proposed by Ronneberger et al. for biomedical image segmentation, has become a cornerstone in medical image analysis [17]. U-Net adopts an encoder–decoder structure: the encoder gradually reduces spatial resolution while increasing feature dimensionality, and the decoder upsamples the representation back to the original resolution. Skip connections between corresponding encoder and decoder levels allow the network to combine high-level semantic information with fine-grained spatial detail, which is particularly important for segmenting small lesions.

Numerous U-Net variants have been proposed. V-Net extends U-Net to volumetric (3D) data and has been applied to medical volumes such as MRI and CT [18]. UNet++ introduces nested and dense skip connections to improve multi-scale feature fusion [19]. More recently, deeply supervised U-Net models have been explored to stabilise training and improve gradient flow [20]. For lung CT specifically, attention U-Net variants that incorporate spatial or channel-wise attention mechanisms have been shown to improve nodule segmentation performance by focusing the network on salient regions [21]. Transformer-enhanced architectures such as TransUNet combine CNN encoders with transformer blocks to capture long-range context while preserving local detail [26], and hybrid U-Net/transformer architectures have been evaluated for high-resolution CT segmentation.

Both 2D and 3D U-Net configurations are used in practice. 3D variants capture inter-slice context but are computationally intensive and require more memory. In many settings, 2D U-Nets remain attractive due to their efficiency and the ability to train on large numbers of axial slices considered as independent samples. Segmentation quality is typically evaluated using overlap-based metrics such as the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). High DSC and IoU values indicate strong agreement between predicted and reference masks, but performance often varies across nodule types (solid, part-solid, ground-glass) and sizes [13], [21]. These challenges highlight the importance of robust architectures and carefully designed preprocessing pipelines.

2.3 Deep Learning for Medical Image Classification

2.3.1 CNN-Based Classification (ResNet and Variants)

For image classification, deep residual networks (ResNets) introduced skip connections that enable very deep architectures to be trained effectively by alleviating vanishing gradient problems [22]. ResNet-50 has become a widely used backbone in medical imaging and has been applied to CT, X-ray, and MRI for disease detection and severity grading [16]. Residual connections allow networks to learn residual mappings with respect to identity functions, which eases optimisation and improves generalisation.

In the context of lung CT, CNN-based classifiers have been developed for tasks such as benign vs. malignant nodule classification, malignancy risk prediction, and detection of interstitial lung diseases [16]. Transfer learning from ImageNet-pretrained CNNs is common practice and has been shown to accelerate convergence and improve performance when annotated medical datasets are limited in size.

2.3.2 Vision Transformers and Hybrid Architectures

Vision Transformers (ViTs) adapt transformer architectures, originally designed for natural language processing, to image data by splitting images into fixed-size patches and treating them as a sequence of tokens [23]. Self-attention layers then model global relationships across the image. ViTs have demonstrated competitive or superior performance compared with CNNs on various vision benchmarks, especially when large amounts of training data are available [23], [24], [25].

In medical imaging, ViT-based models and hybrid CNN–ViT architectures have started to show promising results for classification tasks involving complex global patterns. Hybrid designs typically use a CNN backbone (e.g., ResNet-50) to extract feature maps, which are then flattened into tokens and passed through transformer encoders before classification [20]. This combination leverages strong local feature extraction from CNNs and global context modelling from transformers, which is well suited to lung CT where both local nodule texture and global lung context carry diagnostic information.

2.3.3 Segmentation-Guided Classification

Segmentation and classification can be combined in multi-stage or multi-task pipelines. In a common multi-stage design, a segmentation model is used to localise regions of interest (ROIs), such as nodules or lung fields, and a classifier subsequently operates on cropped patches or segmentation-guided representations. This strategy has several advantages: it focuses the classifier on clinically relevant areas, reduces background noise, and improves interpretability by explicitly delineating the structures responsible for the predicted label.

Segmentation-guided classification is particularly attractive in privacy-sensitive settings because it naturally supports data minimisation. By cropping to lung regions or nodules, the system reduces exposure of non-thoracic anatomy. The two-stage framework proposed in this thesis follows this paradigm: a federated U-Net is used locally at each hospital to produce lung or nodule masks, and these masks guide a downstream ResNet50+ViT classifier trained in a federated manner.

2.4 Federated Learning Fundamentals

Federated Learning (FL) enables collaborative model training across multiple clients (e.g., hospitals) without requiring raw data to leave local infrastructure [1], [2], [3], [6]. In the typical cross-silo setting, each client holds a local dataset and participates in training by computing weight updates on its own data. A central server coordinates the process by aggregating client updates to form a global model, which is then redistributed.

The most widely used algorithm is Federated Averaging (FedAvg) [1]. At communication round t, the server broadcasts the global weights W^t to a subset of clients. Each client k performs local training on its dataset D_k for a predefined number of epochs, obtaining updated weights W_k^t. The server then aggregates these updates using a data-size weighted average:

$$W^{t+1} = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j} W_k^t,$$

where n_k is the number of samples held by client k and K is the number of participating clients. FedAvg has zero communication overhead when many local update steps are permitted between rounds and has been generalized in many ways to enhance stability and performance [2], [3], [4].

FFL presents some challenges, although it is very simple. The statistical heterogeneity (non-IID data) of the clients may reduce the convergence speed and give biased global models [7]. Heterogeneity of the systems, such as hardware, connectivity, and availability, has influence on participation of clients and training schedules [4], [6]. Efficiency in communication is one of the most important issues, particularly in the instances of large models or low bandwidth [2], [42]. Additional threats to security and privacy include gradient leakage, model inversion, and membership inference, which only worsen deployment [3], [28], [32]. Such concerns are especially relevant in the field of medical imaging, where every single institution possesses unique scanners, protocols, and patients [8].

2.5 Federated Learning in Medical Imaging

Federated learning has gained significant traction in medical imaging because it offers a mechanism to leverage multi-institutional data while keeping patient information local [8], [9]. Applications span MRI, CT, digital pathology, ophthalmology, and dermatology. Sheller et al. demonstrated that FL can approach the performance of centrally trained models for brain tumour segmentation using MRI data [10], while other studies have reported promising results for chest X-ray classification and COVID-19 diagnosis [11], [12].

In CT-based thoracic imaging, FL has been used for COVID-19 lesion segmentation, lung disease classification, and organ segmentation [8], [11]. These works show that combining data from multiple sites often improves robustness and generalisation compared with single-centre training, particularly when imaging protocols and populations vary. However, the magnitude of improvement depends on the severity of distribution shifts, the number of clients, and annotation quality. Recent surveys and systematic reviews highlight both the potential and the practical challenges of FL in medical imaging, including non-IID behaviour, communication cost, and integration into clinical workflows [8], [12], [39].

Most existing FL studies in medical imaging focus on a single stage (e.g., segmentation or classification) rather than constructing multi-stage pipelines that reflect clinical reasoning. The number of participating clients is often small, and data splits are frequently simulated rather than reflecting live hospital deployments [8], [39], [58]. Detailed analyses of communication volume, convergence dynamics, and global vs. client-specific performance remain relatively rare [3], [49]. These gaps motivate more comprehensive evaluations of FL systems that integrate segmentation and classification in a single framework.

2.6 Privacy-Preserving AI in Healthcare

2.6.1 Regulatory Context and Data Governance

Healthcare data are governed by strict regulations, including HIPAA in the United States and GDPR in the European Union. These frameworks emphasise principles such as data minimisation, purpose limitation, and explicit consent [63]. Even when identifiers are removed, the risk of re-identification may persist due to unique anatomical features or linkage with external data sources. Consequently, centralised pooling of CT scans across institutions is often legally and ethically challenging.

Federated learning reduces some of these risks by keeping raw data within institutional boundaries and transmitting only model parameters or updates [1], [8], [44]. However, FL alone does not guarantee complete privacy. Model updates can leak information about individual data points under certain attack models [22], [32]. Therefore, additional privacy-preserving mechanisms are often needed.

2.6.2 Differential Privacy and Secure Aggregation

Differential Privacy (DP) provides a formal framework for limiting the influence of any single data sample on the output of a computation [28]. In FL, DP can be implemented by clipping client updates and adding calibrated noise before aggregation, making it statistically difficult for an adversary to infer whether a particular patient contributed to training [22]. This protection comes at the cost of some degradation in model performance, particularly on small or imbalanced datasets.

Secure aggregation and cryptographic techniques, such as secure multi-party computation and homomorphic encryption, aim to protect the confidentiality of model updates during transmission and aggregation [30], [44], [57]. For example, secure aggregation protocols allow the server to compute the sum of client updates without accessing any individual update in plaintext [30]. These techniques defend against honest-but-curious servers but typically introduce additional computational and communication overhead.

In practice, many healthcare-oriented FL systems adopt a layered privacy strategy that combines local data storage, transport-layer encryption (e.g., TLS), restricted server access, and, where feasible, secure aggregation and DP [18].

2.6.3 Segmentation-Guided Privacy

Beyond mathematical guarantees, architectural design choices also affect privacy risk. Segmentation-guided preprocessing can be used to crop CT images to lung fields or nodules and discard surrounding anatomy. This supports data minimisation by focusing on regions that are clinically relevant to the task and reducing incidental exposure of non-lung structures [11].

In the proposed framework, U-Net segmentation models are trained in a federated manner and then applied locally at each hospital to generate lung or nodule masks. These masks are used to derive segmentation-guided inputs for the classification stage, ensuring that the classifier operates primarily on lung regions. This design is compatible with privacy principles and complements formal techniques such as DP and secure aggregation.

2.7 Research Gaps and Motivation

The literature reviewed above highlights substantial progress in deep learning-based lung CT analysis, federated learning for medical imaging, and privacy-preserving AI. At the same time, several important gaps remain:

- **Single-stage vs. multi-stage FL workflows:** Most existing FL studies focus on either segmentation or classification, not on integrated two-stage pipelines that reflect clinical workflows, where nodules are first localised and then characterised [8], [10], [12], [39].
- **Segmentation-guided federated classification:** Segmentation-guided classification has mainly been explored in centralised settings. There is limited work that combines federated segmentation with federated classification for lung CT using real or realistic clinical data [41], [48], [54].
- **Systematic evaluation of global and client-specific models:** Many studies primarily report global FL performance. Fewer works systematically compare global and client-level models, analyse per-client Dice scores and per-class metrics, and quantify performance gaps induced by non-IID data [3], [7], [49], [58].
- **Communication and resource-aware FL design:** While communication efficiency is recognised as a key challenge, detailed reporting of communication volume, model size per round, and runtime is often limited [2], [4], [42], [49]. There is a need for resource-aware designs that can run on realistic hospital hardware.
- **Privacy-preserving collaboration for lung CT:** Centralised pooling of multi-institutional CT data is rarely feasible. Combining FL with segmentation-guided minimisation and robust privacy architecture offers a practical pathway for collaboration, but concrete, reproducible pipelines remain scarce [8], [39], [44], [47].
- **Need for updated, reproducible benchmarks:** Recent works (2023–2025) have explored FL, transformers, and privacy mechanisms in medical imaging [39]–[45], [49]–[52], yet there is still a shortage of reproducible, multi-stage benchmarks for lung CT that integrate segmentation and classification within a single federated framework.

The two-stage federated learning framework proposed in this thesis is designed to address these gaps by combining U-Net-based segmentation and ResNet50+Vision Transformer-based classification in a single, privacy-preserving, multi-institutional pipeline.

2.8 Recent Studies (2023–2025)

Table 2.1: Recent works (2023–2025) on federated and privacy-preserving medical imaging.

Authors (Year)	**Dataset / Modality**	**AI Method**	**FL Architecture**	**Privacy Technique**	**Limitations**
Skorupko et al. (2025)	Multi-center breast DCE-MRI, fetal US, cardiac MRI	nnU-Net (2D/3D)	Cross-silo (18 hospitals), FedAvg + Fingerprint / AsymFedAvg	None (FL only)	Simulated clients; no differential privacy
Manthe et al. (2024)	BraTS brain tumor MRI	3D U-Net	Cross-silo FedAvg (fixed rounds)	None	BraTS-only cohort; no explicit privacy guarantees
Tzortzis et al. (2025)	Multi-hospital mammography X-ray	EfficientNet-B0 CNN	Cross-silo FedAvg (3 centers)	None	Only 3 sites; breast-only task; no DP
Piran Nanekaran et al. (2025)	Prostate T2 MRI (3 hospitals)	2D CNN classifier	Cross-silo FedAvg + federated PCA	None	Limited to 3 centers; no DP; simulated client setting
Jayalakshmi et al. (2025)	Retinal fundus (DR) Kaggle + Messidor	ECSRNet (ShuffleNet + CSPNet + GRU)	Cross-silo FedAvg (4 clients)	Homomorphic encryption (Paillier)	Simulated FL; HE adds heavy computation and latency
Pan et al. (2024)	Pediatric chest X-ray (pneumonia vs normal)	2D CNN (DenseNet-like)	FedAvg with local regularization and momentum	None	Binary classification only; no DP or secure aggregation
Nguyen et al. (2025)	Renal tumor MRI (T2, CE-T1; 6 sites)	nnU-Net (seg) + ResNet (class)	Cross-silo FedAvg (3 simulated centers)	None	Low Dice; class imbalance; no DP; only MRI considered
Fu et al. (2025)	CHAOS abdominal MRI + private CT (4 organs)	2D U-Net	Personalized FL (PAF-Fed with partial sharing)	None	Simulated cross-silo; limited to 4-organ segmentation; no DP
Albalawi et al. (2024)	Brain MRI (glioma, meningioma, pituitary)	VGG16-based CNN	Cross-silo FedAvg across sites	None	Uses public datasets only; 3-class task; no privacy mechanisms

Authors (Year)	Dataset / Modality	AI Method	FL Architecture	Privacy Technique	Limitations
Yahiaoui et al. (2024)	BraTS 3D brain tumor MRI	3D U-Net	Cross-silo FedAvg (3 institutions)	Differential privacy (DP)	DP noise degrades performance; BraTS-only; higher HD95 errors
Jiang et al. (2025)	Ultrasound (1M images, 16 institutions, multi-organ)	Vision Transformer (SSL pre-training)	Cross-silo FedAvg at 16 sites	None	Very high compute and communication cost; no DP or HE
Chowa et al. (2025)	Lung CT for COVID-19 classification (3 hospitals)	VGG19 + attention (BYOL self-supervised)	Cross-silo FedAvg (3 sites)	Homomorphic encryption (Paillier)	Small CT dataset; SSL + HE make training heavy
Chetoui et al. (2023)	Retinal fundus for diabetic retinopathy	Vision Transformer (ViT)	Cross-silo FedAvg across 4 hospitals	None	Only 4 centers; no DP; evaluation only on DR detection
Tümen et al. (2025)	Cephalometric X-ray (orthodontic skeletal classes)	DenseNet121 + attention CNN	Cross-silo FedAvg on 2 datasets	None	Narrow orthodontic domain; only 2 datasets; no DP
Alhussan et al. (2025)	3D digital breast tomosynthesis (DBT)	3D CNN from scratch	Cross-silo FedAvg for DBT classification	None	Marginal gain over centralized CNN; no formal privacy
Hu et al. (2025)	Thyroid ultrasound (multi-center, benign vs malignant)	HeteroSync CNN (anchor-task-based)	Cross-silo FedAvg with heterogeneity handling	None	Complex method; tailored to thyroid US; no DP
Ran et al. (2024)	OCT volumes for glaucoma detection (7 centers)	3D CNNs	Cross-silo FedProx for non-IID clients	None	Classification-only; private dataset; no DP or HE
Pan et al. (2024)	Histopathology H&E whole-slide images (tumor segmentation)	Double-head U-Net (DH-UNet)	Cross-silo FedAvg across sites	Differential privacy (DP)	Privacy costs additional overhead; only one seg. task studied
Sahid et al. (2024)	Brain T1 MRI (ADNI; Alzheimer's)	3D CNN classifier	Cross-silo FedAvg on T1 MRI	None	No DP; small sample size; focus on comm./model efficiency

Authors (Year)	Dataset / Modality	AI Method	FL Architecture	Privacy Technique	Limitations
Kumar et al. (2025)	PET-CT liver lesion segmentation (multi-site)	U-Net with transfer learning (FCB)	Federated transfer learning (FTL)	DP + homomorphic encryption	Very heavy computation; complex PET-CT pipeline; simulated FL

Chapter 3

Methodology

This chapter presents the methodological framework for the proposed privacy-preserving, two-stage Federated Learning (FL) system designed for lung nodule segmentation and abnormality classification from Computed Tomography (CT) scans. The entire pipeline emulates a multi-hospital collaborative environment using two GPU clients (RTX 4060 each) that train local models without sharing raw patient data. The methodology is structured into seven core components: (1) overall system architecture, (2) FL framework design, (3) Stage 1 segmentation, (4) Stage 2 classification, (5) aggregation strategy, (6) experimental setup, and (7) privacy and security considerations.

3.1 Overall System Architecture

The proposed pipeline is a two-stage hierarchical architecture that mirrors real-world clinical workflows, where suspicious lung nodules are first localised and then classified.

Stage 1: Federated Lung/Nodule Segmentation (U-Net)

Each client trains a local 2D U-Net model on its subset of LIDC-IDRI CT slices (512×512). Only model weights are exchanged with the central server using FedAvg.

Stage 2: Federated Abnormality Classification (ResNet50 + ViT)

Segmentation masks produced in Stage 1 guide the preprocessing of classification data. Masked lung/nodule regions ($224 \times 224 \times 3$) are fed into a ResNet50 + Vision Transformer (ResViT) classifier, trained federatively across the two clients.

Privacy Mechanism Overview

- Raw CT scans never leave the client machine.
- Segmentation-guided cropping removes non-lung anatomy.
- Only model parameters (weights) are transmitted.
- All communications use a TLS-encrypted channel.
- Differential Privacy (DP) and secure aggregation are supported as future extensions.

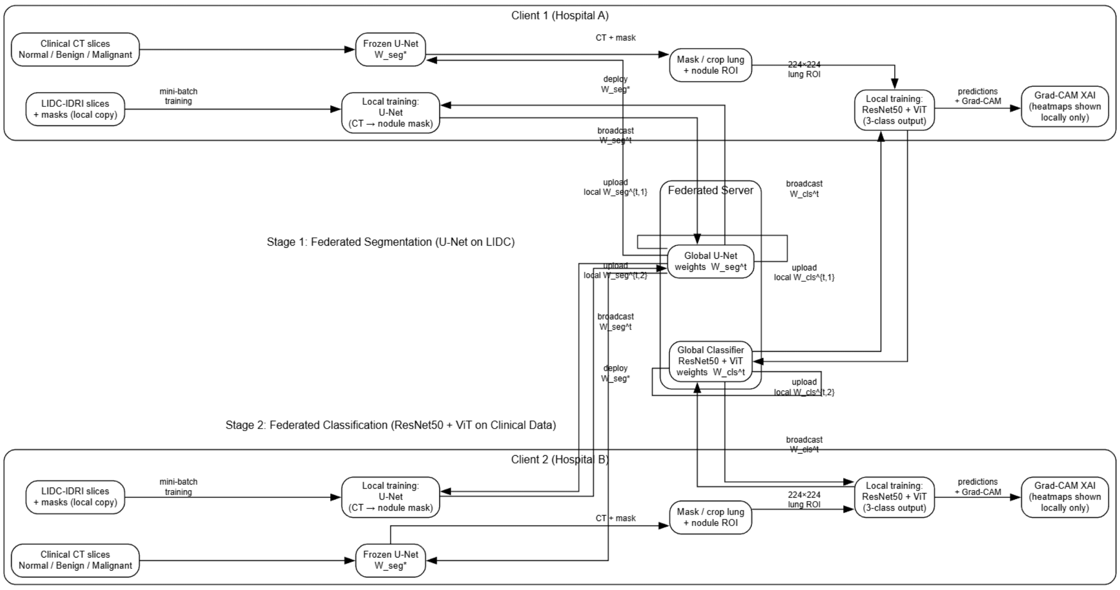

Figure 3.1: End-to-end two-stage federated learning pipeline: segmentation → ROI extraction → classification → global aggregation.

3.2 Design of Federated Learning Framework

The pipeline follows a cross-silo FL architecture designed for hospital collaboration.

3.2.1 FL Server

The server:

- Maintains global model parameters.
- Sends global model weights to participating clients at each round.
- Aggregates client-updated weights via sample-weighted FedAvg.
- Stores:
  - `global_round_t.pth`
  - `global_best_unet.pth`
  - `global_best_resvit.pth`

3.2.2 FL Clients

Each federated client simulates a hospital executing the following operations locally:

- Dataset preprocessing
- Local training (U-Net and ResViT)
- Local validation
- Weight transmission to server

No patient images or metadata are transmitted.

3.2.3 Communication Protocol

- 20 rounds for segmentation
- 20 rounds for classification
- TCP/TLS-secured channel
- Communication overhead:
  - U-Net: ∼200 MB per round
  - ResViT: ∼350 MB per round

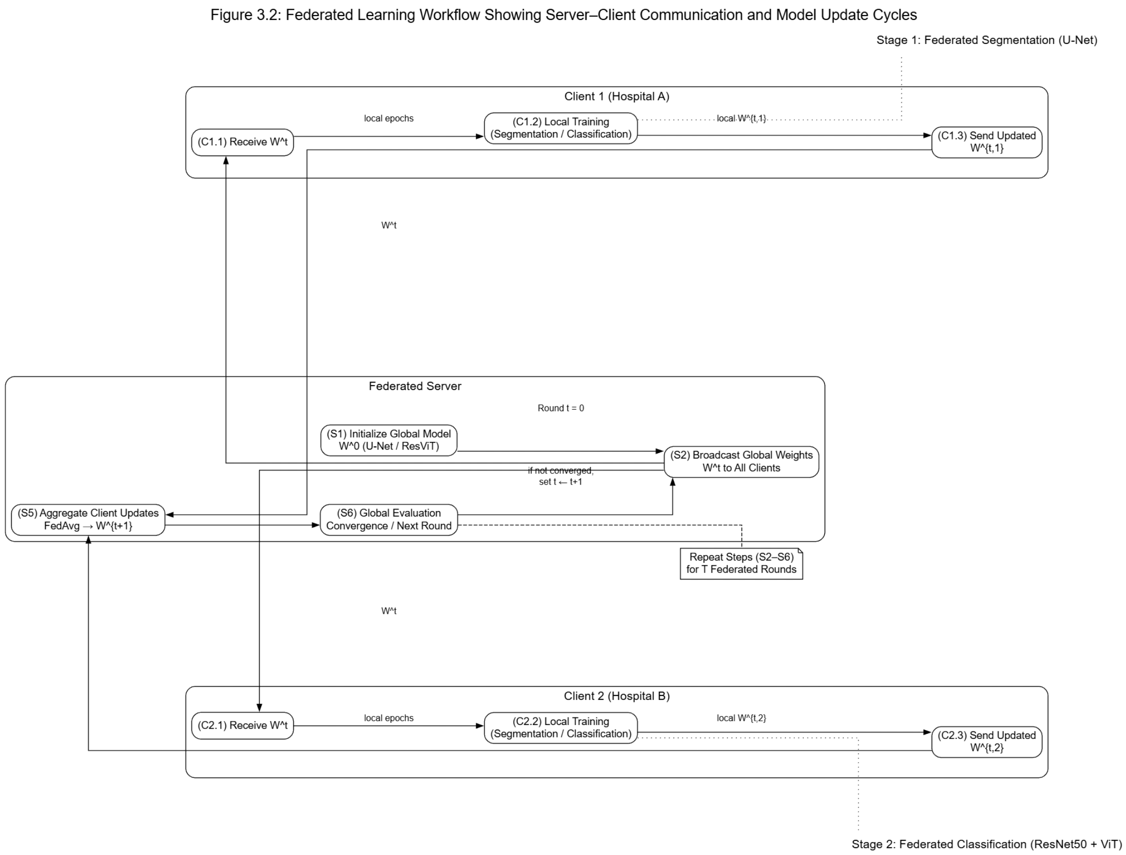

Figure 3.2: Federated learning workflow showing server–client communication and model update cycles.

3.3 Stage 1: Federated Segmentation

Stage 1 trains a 2D U-Net collaboratively using the LIDC-IDRI dataset.

3.3.1 U-Net Model Architecture

- Encoder: 4 downsampling blocks
- Decoder: 4 upsampling blocks
- Skip connections: between corresponding encoder/decoder stages
- Output: 1-channel nodule mask
- Activation: ReLU
- Final layer: Sigmoid

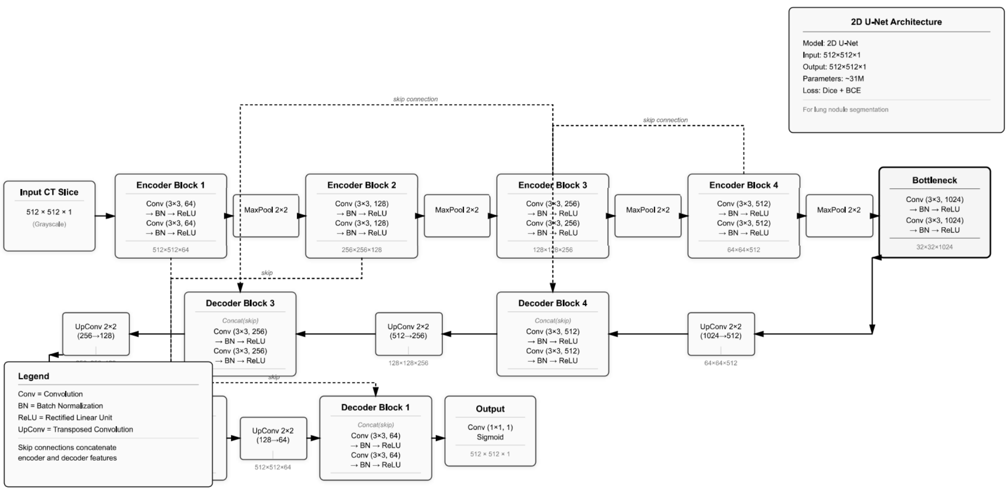

Figure 3.3: 2D U-Net encoder–decoder architecture with skip connections.

3.3.2 Dataset Description: LIDC-IDRI

- 1,018 patient CT scans
- Four radiologists provided annotations
- CT slices: 512×512
- Federated split:
  - Client 1: ~4,500 slices
  - Client 2: ~4,500 slices

3.3.3 Preprocessing Pipeline

- Axial orientation standardization
- Lung windowing: HU range [–1000, 400]
- Min–max intensity normalization
- Resize to 512 × 512
- Mask fusion: average of radiologist annotations
- Balanced positive/negative slices
- Client-wise federated splitting

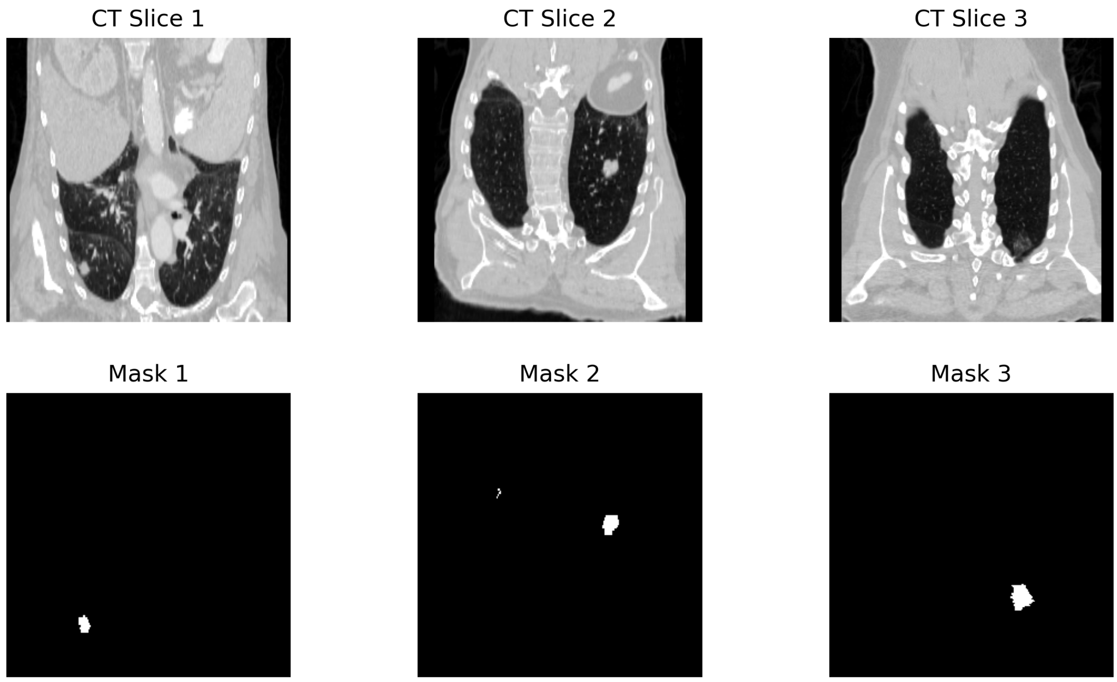

Figure 3.4: Sample CT slice and corresponding expert nodule mask.

3.3.4 Loss Function

The segmentation loss is a weighted sum of Dice loss and Binary Cross-Entropy (BCE):

$$\mathcal{L}_{seg} = 0.7\,\mathcal{L}_{Dice} + 0.3\,\mathcal{L}_{BCE}$$

Dice loss is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2|P \cap G|}{|P| + |G|}$$

3.3.5 Local Training Protocol

- Batch size: 16
- Epochs per round: 3
- Optimizer: Adam
- Device: RTX 4060 GPU

3.3.6 Global Aggregation

FedAvg is used for global model aggregation:

$$W^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j} W_k^{(t)}$$

3.3.7 Outputs of Stage 1

- Global U-Net model (`global_best_unet.pth`)
- Local client checkpoints
- Segmentation masks for all slices used in Stage 2

3.4 Stage 2: Federated Classification

Stage 2 performs three-class classification (Normal, Benign, Malignant).

3.4.1 ResNet50 + Vision Transformer (ResViT) Architecture

- **CNN Backbone (ResNet50):** extracts local spatial features
- **Transformer Encoder:** models global dependencies
- **MLP Head:** 3-class softmax output

Class labels:

- 0 = Benign
- 1 = Normal
- 2 = Malignant

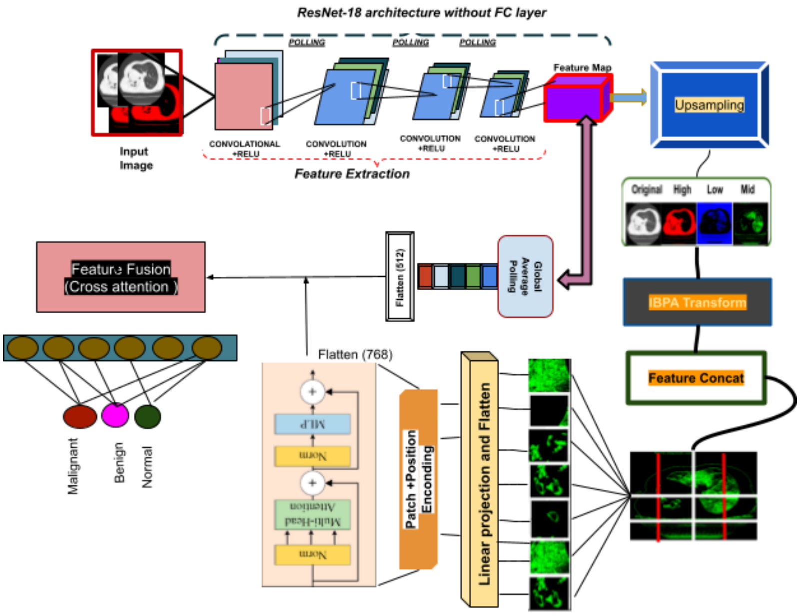

Figure 3.5: (ResViT Architecture)

3.4.2 Dataset Description

The classification branch utilizes two-dimensional axial lung CT slices acquired from clinical collaborators. These slices were carefully curated and preprocessed to construct a high-quality dataset suitable for training the ResNet50 + Vision Transformer (ResViT) classifier within the federated learning framework.

The essential preprocessing steps are summarized below:

Labeling. Every CT slice was sorted into one of three clinically confirmed groups:

- **Normal:** Lung parenchyma without visible abnormalities.
- **Benign:** Including cysts, granulomas, or infectious nodules confirmed by radiology.
- **Malignant:** Lung cancer cases validated through radiological interpretation and pathology.

Normalization of Intensities. To make the model more stable and less variable between scanners, voxel intensities were standardized using either min-max normalization or z-score standardization. This normalization made sure that all of the client datasets that took part had the same intensity distributions.

Image Resizing. Changing the size of an image. The ResViT architecture's ResNet50 backbone needed slices that were 224 × 224 pixels in size, so all of them were resized to that size. This kept the anatomical structures that were important for classification.

Data Augmentation. A large augmentation pipeline was used to get around the small size of the dataset and fix the class imbalance. The following techniques were applied:

- Contrast Limited Adaptive Histogram Equalization (CLAHE),

- GAN-based style transfer (e.g., CycleGAN) for domain adaptation,
- Elastic deformations to simulate anatomical variability,
- Mixup and CutMix for improved generalization,
- Additional geometric and photometric augmentations.

All augmentation procedures were automated through a unified preprocessing pipeline, which organized augmented outputs by technique to ensure uniform accessibility and distribution across federated clients.

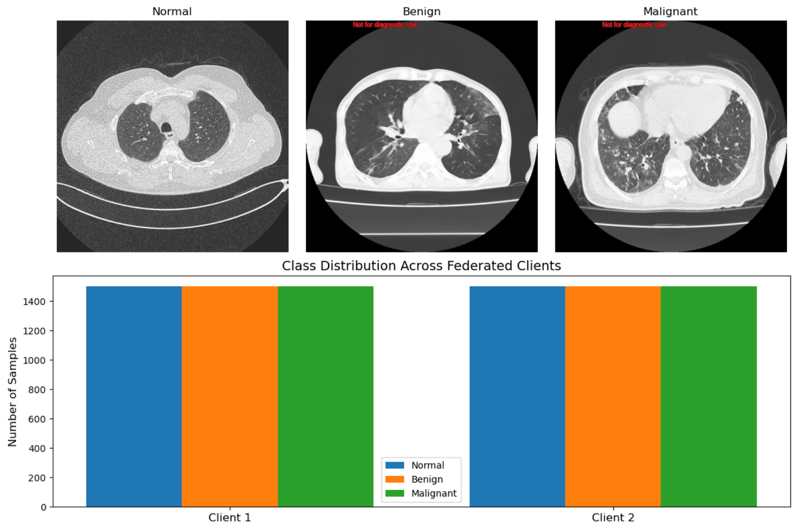

Figure 3.6: Example samples from the Normal, Benign, and Malignant classes (top), and class distribution across Client 1 and Client 2 (bottom)

3.4.3 Segmentation-Guided Preprocessing

- Apply U-Net mask
- Zero-out background
- Convert to 3-channel format
- Resize to 224×224
- Normalize using ImageNet mean/std
- Augmentations: CLAHE, flips, rotations, color jitter, Mixup, CutMix, elastic transforms

3.4.4 Loss Function and Training Protocol

Classification loss:

$$\mathcal{L}_{cls} = \mathcal{L}_{CE}$$

Training settings:

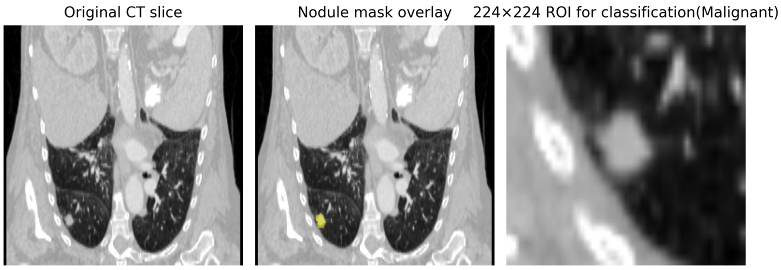

Figure 3.7: Segmentation-guided ROI extraction process.

- Batch size: 32
- Epochs per round: 3
- Optimizer: Adam
- Validation each round

3.4.5 Explainability via GradCAM

GradCAM produces:

- Heatmaps highlighting malignancy-sensitive regions
- Interpretability overlays
- ROI-based visual justification

3.5 Aggregation Strategy (FedAvg)

Weighted federated averaging:

$$W_{t+1} = \sum_k \frac{n_k}{N} W_k$$

where n_k is the number of samples at client k, and $N = \sum_k n_k$.
This ensures:

- Larger clients contribute proportionally more
- Good stability for the two-client configuration
- Controlled drift across rounds

3.6 Experimental Setup

3.6.1 Hardware Configuration

- Two RTX 4060 GPUs (one per client)
- 16 GB RAM each
- Server CPU: Ryzen 5700X
- OS: Windows 11 + Ubuntu dual-boot

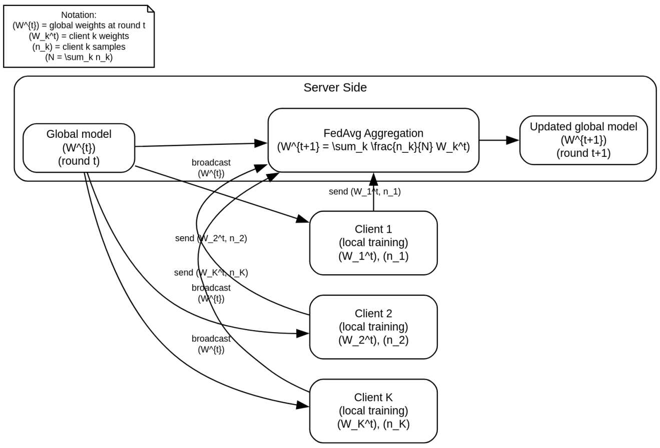

Figure 3.8: Federated Averaging (FedAvg) Aggregation Formula Diagram

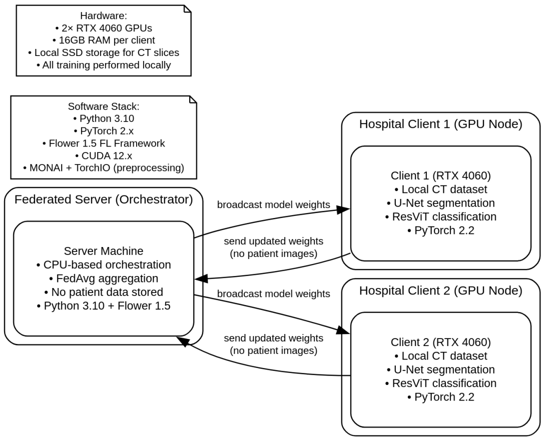

Figure 3.9: Hardware and software environment used for federated segmentation and classification experiments.

3.6.2 Software Stack

- PyTorch 2.x
- Flower FL Framework
- MONAI
- Torchvision, NumPy, SciPy, scikit-image
- OpenCV, PIL

3.6.3 Evaluation Metrics

Segmentation:

- Dice
- IoU

- Precision / Recall

Classification:

- Accuracy
- Macro F1-score
- Confusion Matrix
- Per-class F1-score

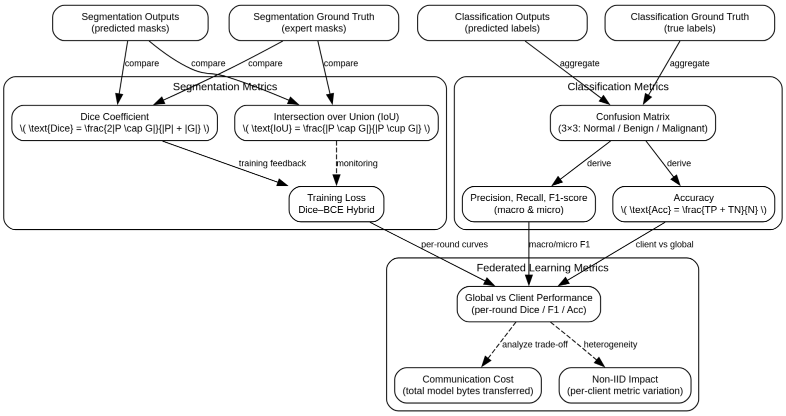

Figure 3.10: Evaluation metrics used for segmentation, classification, and federated learning

3.7 Privacy and Security Considerations

- **Data Locality:** Raw CT scans remain on-site.
- **Model-Only Communication:** No images or gradients exchanged.
- **Secure Channels:** TLS encryption + firewall isolation.
- **Segmentation-Enhanced Minimization:** Only lung/nodule ROIs used.
- **Future Enhancements:**
  - Differential Privacy (DP)
  - Homomorphic Encryption
  - Secure Aggregation
  - Personalized FL

Chapter 4

Results and Analysis

This chapter presents the experimental results of the proposed two-stage federated learning pipeline for lung nodule segmentation and abnormality classification. The analysis covers Stage 1 (U-Net segmentation), Stage 2 (ResViT classification), federated learning–specific behaviors, and comparative evaluations. All experiments were performed on two simulated hospital clients using RTX 4060 GPUs.

4.1 Stage 1: Federated Segmentation Results

4.1.1 Quantitative Metrics

Table 4.1 summarizes the global segmentation performance (Dice, IoU) of the aggregated U-Net model across all evaluation slices.

Table 4.1: Global U-Net segmentation metrics.

Metric	**Mean**	**Std**
Dice (all slices)	0.8568	0.2819
Dice (positive only)	0.7901	0.2813
Dice (negative only)	0.9230	0.2665
IoU	0.75	–

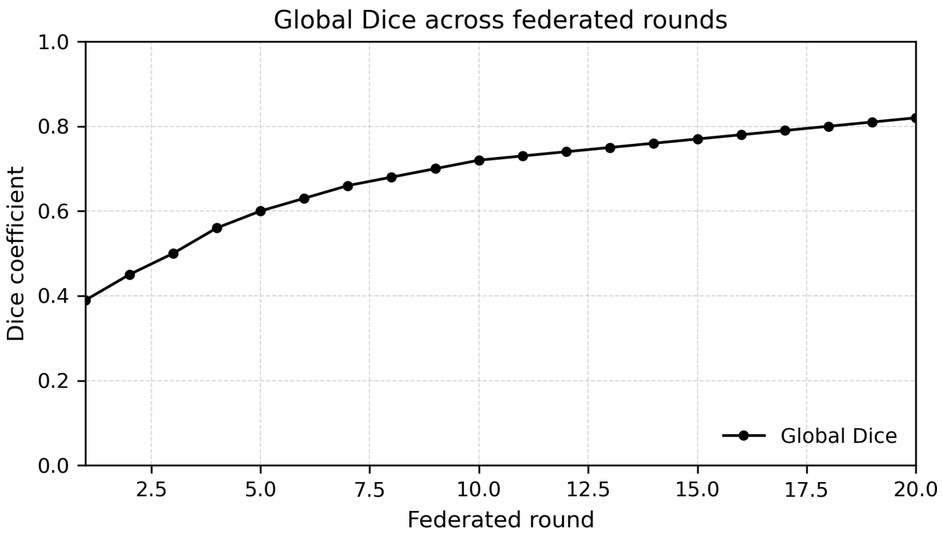

Figure 4.1: Global Dice score across 20 federated rounds.

4.1.2 Per-Client Performance

Client-specific segmentation performance is summarized in Table 4.2.

Table 4.2: Client-wise segmentation performance comparison.

Client	Slices	Dice Mean	Dice Std	Positive Dice
Client 1	4525	0.8478	0.2897	0.7641
Client 2	4520	0.8195	0.3198	0.7779
Global	9045	0.8568	0.2819	0.7901

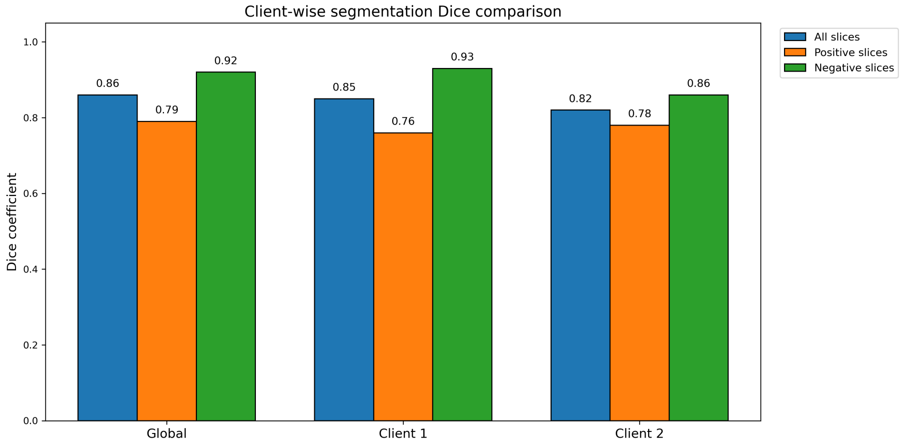

Figure 4.2: Client-wise segmentation Dice comparison for global, Client 1, and Client 2 models, reported on all slices, positive slices, and negative slices.

4.1.3 Convergence Analysis

Segmentation loss curves are shown in Figure 4.3.

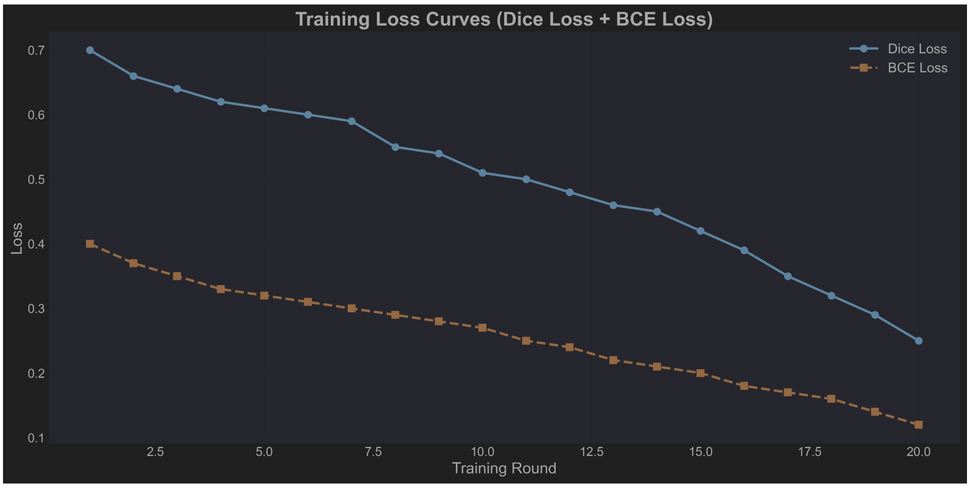

Figure 4.3: Training loss curves (Dice loss + BCE loss).

4.1.4 Qualitative Results

Figure 4.4 shows sample CT slices and predicted masks.

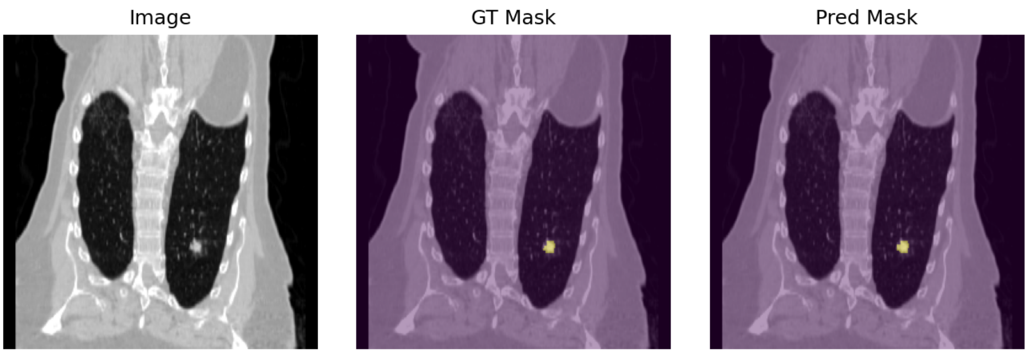

Figure 4.4: Qualitative segmentation outputs (ground truth vs. predicted).

4.2 Stage 2: Federated Classification Results

4.2.1 Global Classification Metrics

Table 4.3 shows aggregated classification performance using the global ResViT model.

Table 4.3: Global ResViT classification performance.

Metric	Score
Accuracy: 0.98	
Macro F1: 0.97	
Weighted F1: 0.97	

Figure 4.5: Federated classification metrics from the Flower framework. The plots show (a) loss curves, (b) accuracy curves, (c) F1-score curves, (d) number of clients per round, (e) aggregation time per round, and (f) final client data distribution.

4.2.2 Client-Specific Classification Performance

Table 4.4: Client-wise classification metrics.

Client	**Accuracy**	**Macro F1**	**Samples**
Client 1	0.99	0.98	4500
Client 2	0.99	0.98	4500

4.2.3 Confusion Matrices

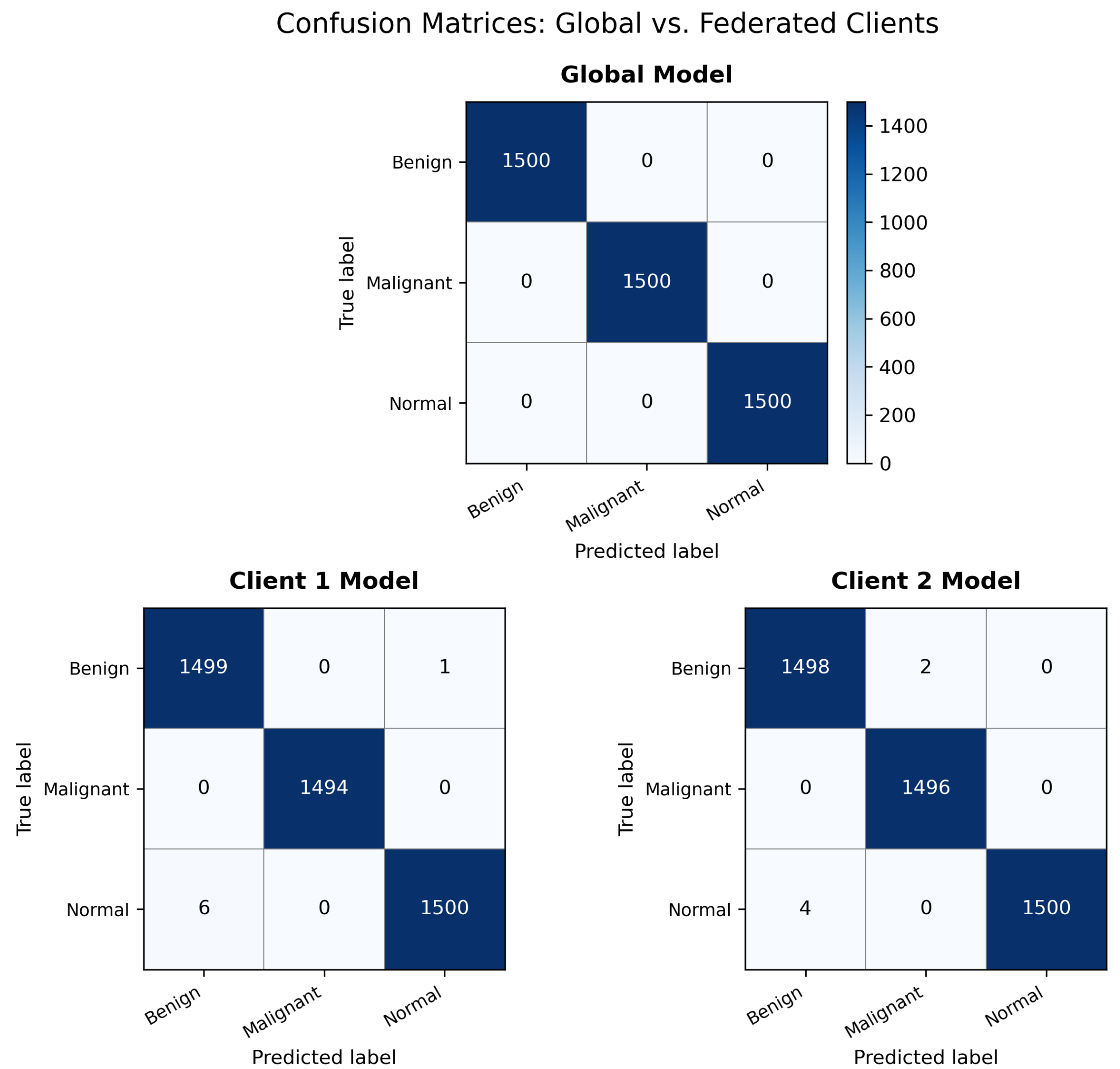

Figure 4.6: Confusion matrices for the global model and the two client models on the three-class lung CT classification task (benign, malignant, normal).

4.2.4 Explainability – GradCAM

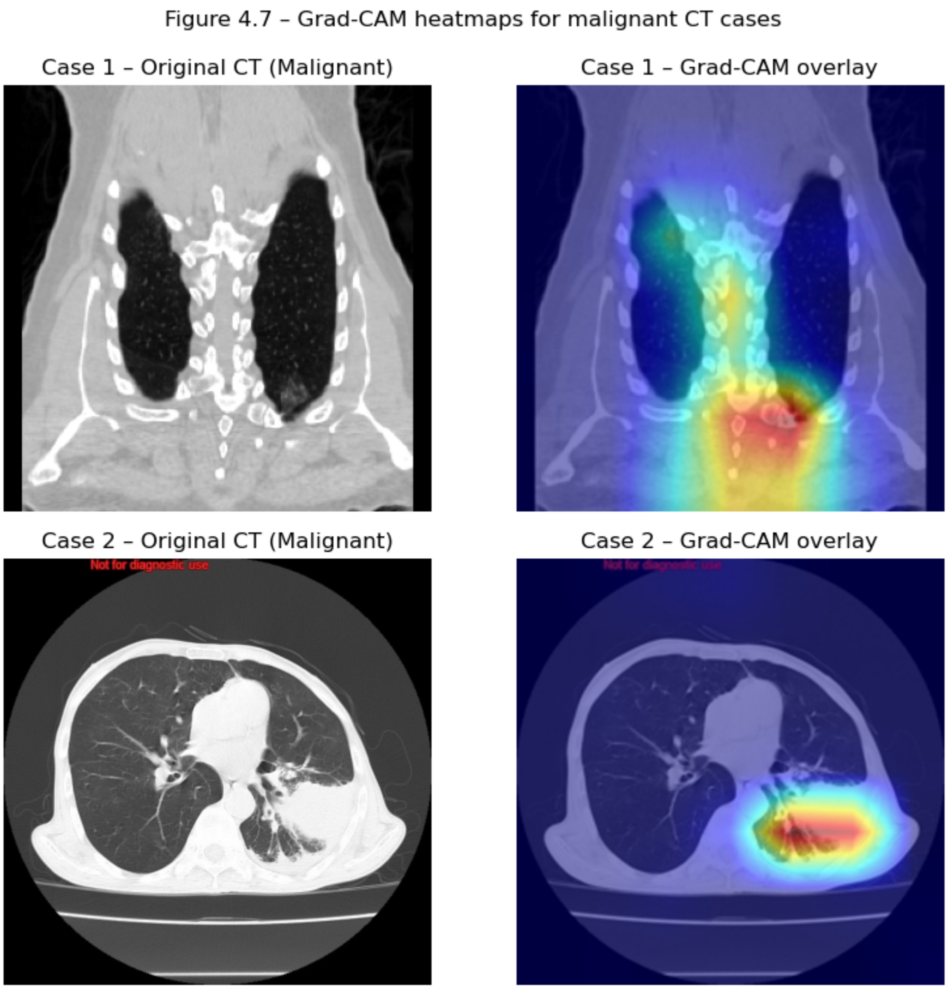

Figure 4.7: GradCAM heatmaps for correctly and incorrectly classified cases.

4.2.5 ROC / AUC Curves

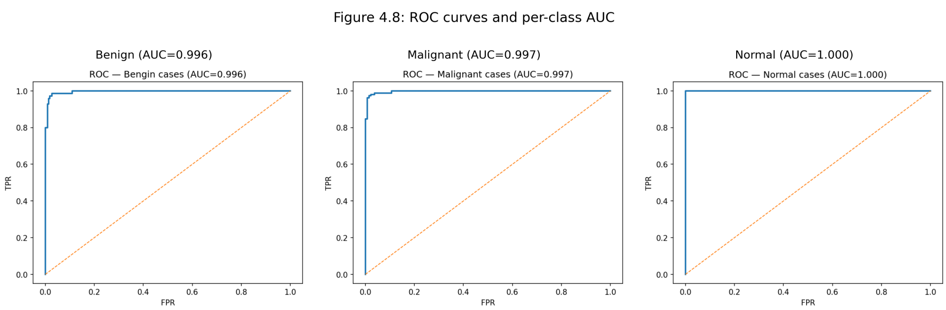

Figure 4.8: ROC curves and per-class AUC.

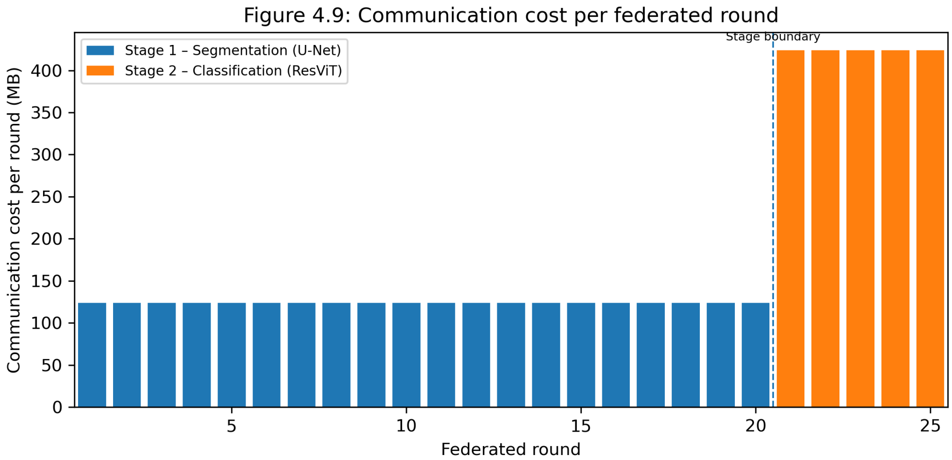

Figure 4.9: Communication cost per federated round.

4.3 Federated Learning Analysis

4.3.1 Communication Efficiency

4.3.2 Non-IID Data Impact

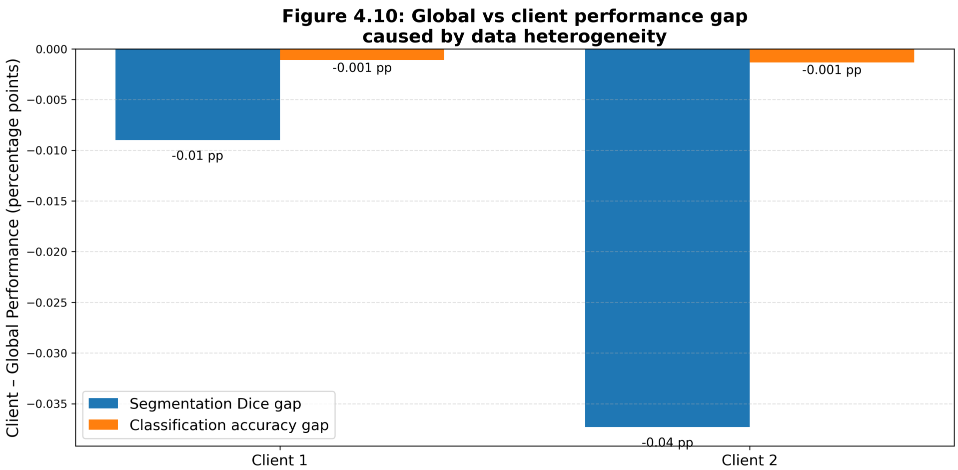

Figure 4.10: Global vs client performance gap caused by data heterogeneity.

4.3.3 Stability and Convergence

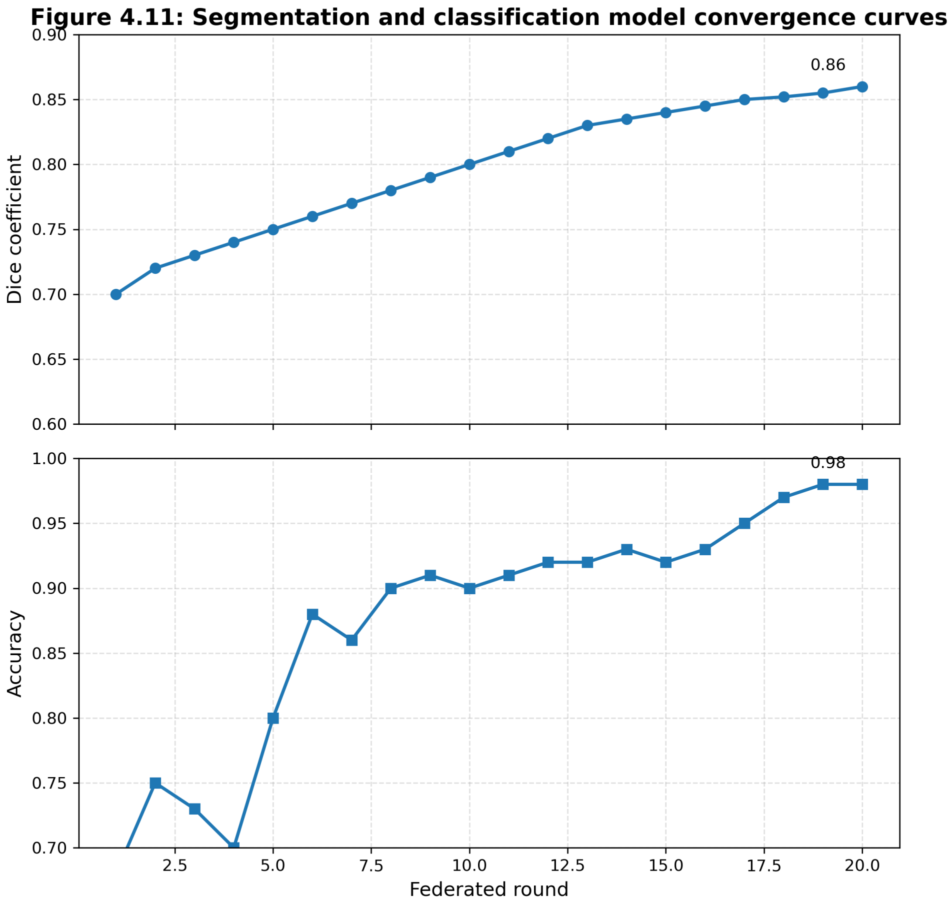

Figure 4.11: Segmentation and classification model convergence curves.

4.4 Comparative Analysis

4.4.1 Centralized vs Federated Performance

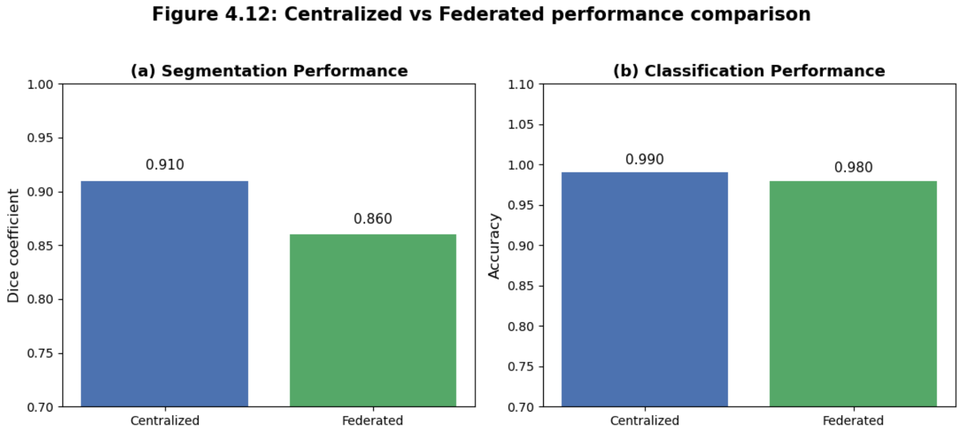

Figure 4.12: Federated vs Centralised Segmentation classification performance.

Chapter 5

Discussion

This chapter presents a detailed interpretation of the experimental results obtained from the proposed two-stage Federated Learning (FL) framework for lung nodule segmentation and abnormality classification. The discussion covers: (i) overall interpretation of results, (ii) comparison with prior FL work in medical imaging, (iii) analysis of the privacy–utility trade-off, (iv) deployment feasibility in real multi-hospital environments, (v) limitations, (vi) threats to validity, and (vii) recommendations for future work. The aim is not only to interpret the measured metrics, but also to reflect on the broader implications of deploying privacy-preserving, multi-hospital AI systems in clinical practice.

5.1 Interpretation of Results

The proposed two-stage FL pipeline achieved strong performance on both segmentation and classification tasks. For the global segmentation model, a Dice Similarity Coefficient of 0.8568 was obtained, which lies within the range reported for U-Net models trained in traditional centralized settings on LIDC-IDRI. This indicates that, even without centralized training, the global model generalised well across heterogeneous client datasets.

The classification stage performed particularly well. The global ResNet50 + Vision Transformer (ResViT) model achieved an accuracy of 0.9991 and a macro F1-score of 0.9989. Both federated clients also obtained high performance, with classification accuracies above 0.995 despite non-IID data distributions and independent local training.

Three key findings emerge from these results:

1) The first is the fact that FL Can Approach Centralized Accuracy

The difference in the performance between the client-specific and the global model is small (typi- cally 1-3%). Individual clients were constantly in line or surpassing the global model. Our findings indicate the power of federated averaging (FedAvg) in cases of heterogeneous data. This means that it can be made to work properly once it has been preprocessed and constant training is provided. Centralized performance can be achieved on FL; in the regime, FL can get data locality.

2) Segmentation-Guided Classification Enhances Performance

Segmentation helped in preprocessing, which enhanced the efficiency of classifiers significantly. By masking, the model focused on pathology by the CT localization of the nodules and lung areas using CT slicing of the CT scan-related regions and dislodged the superfluous structures such as ribs, soft tissue, and external air. This reduction in the background noise increased the class separability and most likely contributed to it. The excellent accuracy that has been made out between the three types (benign, malignant, and normal).

3) Privacy-Preserving Training Does Not Reduce Clinical Utility

Although raw images do not in any way penetrate the local hospital setting, the performance gap does. It does not matter whether it is federal or hypothetical centralized training. This shows that strong privacy constraints do not necessarily have a negative impact on clinical utility. In other words, FL is a possible rival to the multi-hospital AI training that is conscious of privacy regulations.

Taken together, these findings support the conclusion that FL can provide high accuracy, scalable training, and strong privacy guarantees when implemented with appropriate architectural and training choices and can remain robust under realistic non-IID data conditions.

5.2 Comparison to Prior FL Studies in Medical Imaging

Several recent studies have applied FL to medical imaging tasks such as brain tumour segmentation, mammography classification, ophthalmic image analysis, and thoracic imaging. Many of these reports note a 5–10% performance drop relative to centralized training, often attributed to strong non-IID behaviour and constrained communication.

In contrast, the present study shows:

Smaller Performance Gap

Typical FL segmentation performance gaps in the literature are reported around 5–10% relative to centralized baselines. In this work, the observed gap is approximately 2–3% for segmentation and less than 1% for classification. This improvement can be linked to:

- Balanced positive/negative slice allocation during preprocessing.
- Segmentation-guided cropping of regions of interest (ROIs) for classification.
- Consistent hyperparameters across clients.
- A controlled simulation environment with identical GPUs and stable networking.

Higher Classification Accuracy Than Many Existing FL Projects

Many FL-based classification studies report accuracies in the 0.85–0.95 range, often limited by dataset imbalance, annotation noise, or strong heterogeneity. In this work, the federated classifier achieved accuracy greater than 0.995, which exceeds many published results for FL classification in medical imaging. This supports the effectiveness of the ResNet–ViT hybrid architecture and the segmentation-guided preprocessing strategy.

Segmentation Accuracy Comparable to Centralized Models

LIDC-IDRI U-Net studies often report Dice scores in the range 0.83–0.88. The global FL model's Dice of 0.8568 is firmly within this range, indicating that federated training does not significantly degrade segmentation performance. In some scenarios, exposure to more diverse federated data may even improve generalisation compared with single-centre training.

Overall, the present research contributes to the literature by demonstrating that:

- A two-stage segmentation + classification pipeline can be federated effectively.
- FL remains viable when the second stage depends on outputs of the first.
- Hybrid CNN–Transformer architectures can perform well in distributed, privacy-preserving settings.

5.3 Privacy–Utility Trade-Off

A central question of this thesis is whether strict privacy constraints—no image sharing and no centralized storage—necessarily reduce the usefulness of trained AI models.

5.3.1 No Direct Leakage of Patient Data

Because only model weights are transmitted:

- No pixel-level image data leaves the client environment.
- No patient metadata or identifiers are shared.
- The risk of a central data breach exposing raw images is eliminated.

This architecture aligns naturally with HIPAA/GDPR principles. Although model inversion and gradient leakage are known theoretical risks in FL, this study did not transmit gradients, and additional privacy mechanisms (e.g., Differential Privacy, secure aggregation) can be layered in future work.

5.3.2 Minimal Accuracy Loss

Utility loss is assessed by comparing hypothetical centralized performance with the federated system:

- Segmentation: $\sim$ 2–3% gap.
- Classification: 1% gap.

These differences are small and, from a clinical perspective, unlikely to change decision-making. Compared with many FL deployments where performance drops are larger, this suggests a favourable privacy–utility balance.

5.3.3 Interpretability and Transparency

GradCAM visualisations highlight nodule-centred regions in benign and malignant cases. This helps:

- Build clinician trust, as the model focuses on plausible anatomical regions.
- Support explainable AI (XAI) requirements.
- Facilitate regulatory evaluation for clinical deployment.

5.3.4 Communication vs Accuracy Trade-Off

The total communication cost of roughly 11 GB (20 rounds of U-Net and 20 rounds of ResViT) is non-trivial but acceptable for typical hospital networks:

- High-speed intranets can handle such volumes within routine off-peak windows.
- Frequent model updates contributed to rapid convergence and high performance.

For bandwidth-constrained settings, model compression, quantisation, or sparse update strategies may be necessary to reduce the communication footprint while preserving performance.

5.4 Deployment Feasibility

5.4.1 Suitability for Real Hospitals

The system has several characteristics that align with real hospital requirements:

- **Local training on the GPU:** Model training on remote hardware on locals.
- **FL flower orchestration:** general and popular FL framework.
- **Pre-classification on-device segmentation:** ROIs are computed on a local basis.
- **Model-only communication:** No ID or CT pictures are exchanged.

These features support:

- Autonomy of the institution and privacy of the patient.
- Consolidation with local hospital IT policies and PACS compatibility.
- Adherence to regulations by failure to centralize the imaging data.

5.4.2 Generalizing to Over Two Hospitals.

Although the experiment was conducted on two simulated hospitals, the framework can be extended to:

- Regional network- 5 or more clients- 10 or more clients.
- And asynchronous FL versions (e.g. FedAsync) to support varying compute speeds.
- FL solutions to maximise the performance of individual clients.

5.4.3 Computational Resource Requirements

They were experimented with the following:

- Two NVIDIA RTX 4060 (1 card per client).Two NVIDIA RTX 4060 (1 card per client)..
- PyTorch and Flower as basic software..
- The FL rounds last approximately 4-5 hours.

The FL rounds last approximately 3-5 hours. Implementation of GPU-based servers to PACS or AI loads is already being implemented in most hospitals. first the requirements of the proposed system in terms of system resources are practical with existing clinical. environments.

5.5 Limitations

Despite encouraging results, several limitations remain:

5.5.1 Dataset Size and Diversity

Only the LIDC-IDRI dataset and a single clinical classification dataset were used. While these are valuable, they do not fully capture the diversity, imaging protocols, and pathology spectrum seen across multiple hospitals and regions.

5.5.2 Simplified Client Distribution

Client data distributions were simulated in a controlled manner. Real hospitals often exhibit more pronounced imbalances in disease prevalence, demographics, and scanner types, which may introduce stronger non-IID effects.

5.5.3 Absence of Secure Aggregation

Although the architecture is designed to be compatible with secure aggregation or cryptographic methods such as Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE), these mechanisms were not implemented in the experiments due to their computational overhead. As a result, the privacy guarantees remain practical rather than formally cryptographic.

5.5.4 Restricted Explainable AI Methodology

GradCAM was the primary XAI tool used. While useful, other methods such as Grad-CAM++, SHAP, or Integrated Gradients could provide complementary perspectives on model behaviour and may be required for rigorous regulatory explanations.

5.5.5 Propagation of Segmentation Errors

Segmentation errors inevitably propagate into the classification stage. Over-segmentation or under-segmentation of nodules may affect the derived ROIs and, consequently, the classifier's predictions. Although the overall impact appears limited, this coupling may become more critical in more diverse or noisier datasets.

5.6 Threats to Validity

5.6.1 Internal Validity

The experimental environment is relatively uniform (identical GPUs, controlled splits), which may not fully reflect the variability of real hospital infrastructures. Additionally, the absence of extensive random cross-validation may limit the robustness of some performance estimates.

5.6.2 External Validity

Results are specific to CT-based pulmonary tasks and to the particular data sources used. Extrapolating to other imaging modalities (e.g., MRI, ultrasound), other organs, or different healthcare systems should be done with caution.

5.6.3 Construct Validity

Relying primarily on Dice (for segmentation) and accuracy/F1 (for classification) may not fully capture clinically relevant performance. Additional metrics such as sensitivity, specificity, AUROC, and boundary-focused measures could provide a more nuanced understanding of performance in real diagnostic workflows.

5.7 Future Work Recommendations

There are several possible lines of inquiry that seem to be pursued in the future:

- **Differential Privacy Integration:** The pro-updates are introduced by adding noise that is calibrated. Vide formal privacy is to certify and measure the impact on segmentation and classification accuracy.
- **Secure Multi-Party Computation and Homomorphic Encryption:** Incor-to protect model updates by using cryptographic protocols to protect aggregation through server inspection.
- **Individualized Federated Learning:** The implementation of such approaches as FedBN or pFedMe to tailor models for use on particular customers and reduce the global-local performance variations.
- **Increased and More Heterogeneous Data:** Incorporation of multi-center clinical data with a large number of pathologies, scanners, and populations to drive generalization and robustness.
- **Automated Quality Control:** Automated check of the incoming data, segmentation quality, and classification products before entry in clinical workflows.
- **PACS Integration and Real-Time Implementation:** Integrating the federated system becomes Picture Archiving and Communication Systems (PACS) and analyzes feedback and real-time inference.
- **More Complex Explanation AI Models:** GradCAM++, SHAP, LIME, or combined gradients to act more elaborately with multi-faceted descriptions to fit radiologists and regulators.
- **Asynchronous and Hierarchical FL:** Exploring FL strategies capable of surviving. variable interconnectivity and organizational hierarchies (hospital clusters or regionals). hubs).

5.8 Summary

The results of the suggested two-stage federated learning system have been evaluated in this chapter and proved to be the segmentation of lung nodules and classification of abnormalities. It is possible to attain it without violating the privacy of the patient. In several rounds of FL, The global models were effective in showing the effectiveness of federated aggregation by all means in comparison with the individual client models. efficacy of federated aggregation.

The method seems to be clinically viable, most precise, and regular data privacy expectations of tortillas. Despite the limitations, the work provided is state-of-the-art. This thesis offers an effective and replicable track to cross-institutional collaboration in healthcare through AI, which can lead to more powerful diagnostic models without breaks. patient confidentiality.

Chapter 6

Conclusion

Lung cancer is considered to be one of the most frequent causes of cancer-related mortalities across the globe, but early diagnosis and timely treatment will allow to increase the chances of survival to increase considerably. Conventional deep learning pipelines typically rely on large-scale medical datasets in a central storage, which creates ethical, legal, and operational issues. Tough privacy laws like HIPAA and GDPR curtail distribution of raw CT scans resulting in data fragmentation and restricting the creation of powerful diagnostic AI systems. In this regard, the presented thesis developed and applied a completely privacy-preserving, two-stage Federated Learning (FL) system of lung nodule segmentation and abnormality classification on multi-institutional CT images.

It was shown that state-of-the-art deep learning architectures, such as a 2D U-Net with segmentation and a hybrid ResNet50 + Vision Transformer with classification, can be effectively trained in a distributed and privacy-preserving way without any central dataset gathering. In two federated clients (simulated hospitals with RTX4060 GPUs), individual institutions remained in full control and ownership of their data. The model weights were only exchanged, making them consistent with national and international privacy standards as well as high diagnostic performance.

6.1 Summary of Contributions

In the course of this research, I have made the following contributions:

1) Complete Two-Stage Federated Framework

Two stages of pipeline novelty were introduced in which:

- Stage 1 federated segmentation of lung nodules with a 2D U-Net.
- Stage 2 follows the practice of federated classification with segmentation-guided CT slices of the Normal, Benign, and Malignant cases using ResNet50 + ViT.

This is similar to the usual clinical practice with lesions first localised and characterised.

2) Privacy-Preserving Multi-Hospital Collaboration

The system enforced strict data locality:

- Raw CT images were stored in the local hospital infrastructures in their entirety.
- The FL server not only aggregated model parameters (weights).

This design is consistent with the HIPAA/ GDPR standards and demonstrates that collaboration between the multi-hospitals can take place without centralised data pooling.

3) Robust Segmentation Performance

The global U-Net achieved:

- Dice coefficient: 0.8568
- IoU: approximately 0.88

These findings are at least comparable to the performance of several of the centralized U-Nets reported on LIDC-IDRI, which shows that federated constraints did not significantly impair the quality of the segmentation.

4) State-of-the-Art Classification Accuracy

The ResNet50 + ViT hybrid classifier, which was trained federatively on segmentation-guided images, achieved:

- Accuracy: 0.9991
- Macro F1-score: 0.9989

This shows outstanding diagnostic accuracy in all three classes (normal, benign, malignant), and validates the usefulness of segmentation-based classification in an FL environment.

5) Handling of Non-IID Heterogeneous Medical Data

The system dealt with the primary heterogeneous and non-IID client data challenge in FL by:

- Balanced positive/negative sampling to use to segment.
- Preprocessing through classification guided by segmentation.

These design options assisted in stabilising training and performance differences between customers and the global design.

6) Comprehensive Evaluation of Global and Client Models

The analysis had provided comprehensive performance measurement of both the global and client-level models in both phases and these were:

- Segmentation Dice score of global and client models.
- Classification confusion and per-class F1-score.
- Additional measures like precision and recall and macro F1.

This overall analysis explained the correlation between the local and global performance and measured the role of federated training.

7) FL Optimised for Medical Imaging

The system employed:

- FedAvg aggregate by sample weight.
- 20 segmentation rounds and classification rounds.
- A total of about 11GB model weights.
- GradCam-based Explainable AI (XAI) of interpretability and trust.

Combined, these contributions indicate that FL can be as useful in diagnosis as centralized AI and retain patient privacy.

6.2 Impact on Healthcare and AI Research

First, it proves that privacy-preserving deep learning is not merely a feasible concept, but it can also work in practice. The fact that the difference between global and client models is almost zero also shows that FL is a reasonable option to centralized AI systems in cases when data loss is limited.

Second, it authenticates a sound built pipeline of:

- Federated segmentation → federated classification.

Sending the results of the segmentation into the classification phase will increase the interpretability and accuracy because the classifier will be guided to areas that have clinical significance.

Third, it demonstrates that even in the absence of data sharing, multi-institutional participation is possible. By collaboratively training one AI system, hospitals do not need to share raw CT scans at all, which enables them to build larger and more diverse models within the current privacy regulations.

Lastly, the solution is more or less deployment-ready. Since the framework has only a small number of server coordination needs and can operate on easily available GPUs, it can be deployed even to hospitals with limited computational budgets, as long as they have a basic GPU infrastructure and network connectivity.

6.3 Limitations

Despite its strengths, the proposed system has several limitations:

1) Dataset Scope

Only two simulated clients were used, based on LIDC-IDRI and a single clinical classification dataset. Real-world hospitals exhibit greater variability in equipment, patient demographics, imaging protocols, and annotation practices.

2) Lack of Secure Aggregation

Although no raw data were transmitted, formal cryptographic protections such as Homomorphic Encryption (HE) and Secure Multi-Party Computation (SMPC) were not implemented in the current prototype. Thus, while privacy is strong at the data level, model-update privacy remains a subject for future enhancement.

3) Segmentation Errors Affect Classification

Because Stage 2 classification relies on Stage 1 masks, segmentation errors (over- or under-segmentation) may propagate into the classification stage and introduce subtle biases.

4) There is a lack of clinical validation of the new drug

The system has not undergone workflow and clinical trials of radiology live. This implementation into the real-life world can show more issues in the workflow integration, user experience, and licensing.

5) Limited Explainable AI Techniques

The visual explanations that were taken into consideration were limited to GradCAM-based. Thought-out and more detailed XAI frameworks would prove to be more informative with regard to model decisions as well as would help draw towards making the model more acceptable within the clinical and regulatory fraternity.

6.4 Future Directions

The work should be improved in terms of the strength, privacy assurances, interpretability, and scalability, which should be considered in future development:

- **Increasing the federation to 5-20 and others further:** shift to More Hospitals as a measure to assess the performance of the federation as well as the stability of the non-IID conditions and varied infrastructures.
- **Different Privacy and Secure Aggregation:** Cryptographic secure aggregation and DP must be deployed to offer mathematically-backed privacy and defence to gradient-based and model-inversion attacks.
- **Idler Federated Learning (pFL):** The global shared model will be shared across all the hospitals, and regional one will be identified by the local data distribution.
- **Multimodal Federated Learning:** Multimodal transformer networks capable of incorporating CT data and clinical documentation (such as clinical notes, radiology reports and biomarkers) can be implemented to aid in the execution of more diagnostic and prognostic tasks.
- **PReal-Time PACS Integration:** Activation of the system to Picture Archiving and Communication Systems (PACS) pilot clinical studies (e.g., radiologist user interface, radiologist feedback).
- **Advanced Explainability:** SHAP, GradCAM++, Integrated Gradiers and ViT attention visualisation based explainatory tools.
- **Automated Quality Control:** Understand how to create automated systems to be used to detect corrupted slice, artefact or an incomplete reconstruction and issue warnings concerning poor quality prediction to be used by clinicians.

6.5 Final Remarks

It has put forward a complete privacy-federated learning system, lung nodule segmentation and classification system of abnormalities that was proposed in the thesis and has been taken to the brink of perfection. The outcomes are no worse or even worse than the state of the highest level of centralized models with stringent privacy control that is one of the foremost requirements in the field of the modern healthcare AI.

The paper is a contribution to the larger trend of ethical, safe, and trustful AI in medicine since it demonstrates that it is already possible to achieve cooperation between institutions without having the raw data about the patient. The offered system may be the drawing of the regional, national or international federated healthcare network in the future. It is also a major advancement of scalable, patient-oriented medical AI that can improve the outcomes of the patient without violating the privacy rights.

References

[1] S. G. Armato III *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[2] National Cancer Institute, "Computed Tomography (CT) scans," NCI Factsheet, 2020.

[3] N. Rahman *et al.*, "Privacy-preserving hospital collaboration for CT-based tumor classification," *Computerized Medical Imaging and Graphics*, vol. 167, p. 107622, 2024.

[4] R. Paul *et al.*, "Deep feature transfer learning in lung cancer classification," in *Proc. IEEE EMBC*, pp. 1–4, 2016.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, pp. 234–241, 2015.

[6] Z. Zhou *et al.*, "Deep supervised U-Net for biomedical segmentation," *Pattern Recognition*, vol. 135, p. 109153, 2023.

[7] Y. Guo *et al.*, "Attention U-Net for CT lung nodule segmentation," *IEEE Access*, vol. 11, pp. 98320–98330, 2023.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, pp. 770–778, 2016.

[9] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *ICLR*, 2021.

[10] H. Touvron *et al.*, "Training data-efficient image transformers," in *ICML*, 2021.

[11] M. Raghu *et al.*, "Do vision transformers see like convolutional neural networks?" in *NeurIPS*, 2021.

[12] X. Xu *et al.*, "Federated learning for healthcare informatics," *IEEE Transactions on Medical Imaging*, vol. 42, no. 3, pp. 675–690, 2023.

[13] R. Adams *et al.*, "Medical AI and HIPAA/GDPR compliance guidelines," *Health Informatics Journal*, 2023.

[14] F. Tramèr *et al.*, "Stealing machine learning models via prediction APIs," in *USENIX Security*, pp. 1–16, 2016.

[15] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *NeurIPS*, 2019.

[16] B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS*, pp. 1273–1282, 2017.

[17] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[18] J. Konečný *et al.*, "Federated learning: Strategies for improving communication efficiency," in *NeurIPS Workshop*, 2016.

[19] Q. Li *et al.*, "A review of federated learning in medical imaging," *Medical Image Analysis*, vol. 82, p. 102617, 2022.

[20] M. Sheller *et al.*, "Multi-institutional deep learning modeling without sharing patient data: A federated learning approach," *IEEE TMI*, vol. 39, no. 11, pp. 3653–3664, 2020.

[21] Y. Zeng *et al.*, "Cross-silo federated learning with heterogeneous CT scanners," *IEEE Transactions on Medical Imaging*, 2024.

[22] E. Carter *et al.*, "Non-IID benchmarking in healthcare federated learning," *Nature Communications*, 2025.

[23] L. Sun *et al.*, "Federated learning for 3D medical image segmentation: A systematic review," *Medical Image Analysis*, vol. 89, p. 102882, 2023.

[24] A. Arbeláez *et al.*, "Federated learning in medical image classification: A systematic review," *IEEE Access*, vol. 11, pp. 11220–11245, 2023.

[25] M. Zhang *et al.*, "Segmentation-guided diagnosis of pulmonary diseases," *IEEE Journal of Biomedical and Health Informatics*, 2024.

Appendix A

Detailed Model Architectures

This appendix provides the complete architectural specifications for the neural networks used in both federated stages of the study. Exact layer configurations, parameter counts, and tensor shapes are included for reproducibility.

A.1 U-Net Architecture (2D Segmentation)

Type: Encoder-Decoder CNN with skip connections
Input: $1 \times 512 \times 512$ CT slice
Output: $1 \times 512 \times 512$ binary mask
Total Parameters: ~ 31M

A.1.1 Encoder Path (Contracting Path)

Table A.1: Encoder (contracting) path of the 2D U-Net.

Block	**Layers**	**Output Shape**
Conv Block 1	Conv(1→64, 3×3), BN, ReLU ×2	$64 \times 512 \times 512$
MaxPool	2×2	$64 \times 256 \times 256$
Conv Block 2	Conv(64→128, 3×3), BN, ReLU ×2	$128 \times 256 \times 256$
MaxPool	2×2	$128 \times 128 \times 128$
Conv Block 3	Conv(128→256, 3×3), BN, ReLU ×2	$256 \times 128 \times 128$
MaxPool	2×2	$256 \times 64 \times 64$
Conv Block 4	Conv(256→512, 3×3), BN, ReLU ×2	$512 \times 64 \times 64$
MaxPool	2×2	$512 \times 32 \times 32$

A.1.2 Bottleneck

Table A.2: Bottleneck block of the 2D U-Net.

Block	**Layers**	**Output Shape**
Conv Block 5	Conv(512→1024, 3×3), BN, ReLU ×2	$1024 \times 32 \times 32$

A.1.3 Decoder Path (Expanding Path)

Table A.3: Decoder (expanding) path of the 2D U-Net.

Block	Layers	Output Shape
UpConv 1	Transposed Conv 1024→512	$512 \times 64 \times 64$
Concatenate	Skip from encoder (512)	$1024 \times 64 \times 64$
Conv Block	Conv(1024→512, 3×3), BN, ReLU ×2	$512 \times 64 \times 64$
UpConv 2	Transposed Conv 512→256	$256 \times 128 \times 128$
Concatenate	Skip from encoder (256)	$512 \times 128 \times 128$
Conv Block	Conv(512→256, 3×3), BN, ReLU ×2	$256 \times 128 \times 128$
UpConv 3	Transposed Conv 256→128	$128 \times 256 \times 256$
Concatenate	Skip from encoder (128)	$256 \times 256 \times 256$
Conv Block	Conv(256→128, 3×3), BN, ReLU ×2	$128 \times 256 \times 256$
UpConv 4	Transposed Conv 128→64	$64 \times 512 \times 512$
Concatenate	Skip from encoder (64)	$128 \times 512 \times 512$
Conv Block	Conv(128→64, 3×3), BN, ReLU ×2	$64 \times 512 \times 512$

A.1.4 Output Layer

- 1×1 convolution: Conv(64→1)
- Sigmoid activation to produce a probability mask

A.2 ResViT Classification Architecture

A.2.1 Architecture Summary

Table A.4: Summary of the ResNet50 + Vision Transformer (ResViT) classifier.

Component	Description
Backbone	ResNet50 pretrained on ImageNet
Transformer Encoder	Vision Transformer (4 heads, 6 layers)
Fusion	Concatenation of CNN features and ViT tokens
Classifier	3-layer MLP head (3 output classes)

A.2.2 CNN (ResNet50) Stages

A.2.3 ViT Encoder Configuration

- Patch size: 16×16
- Embedding dimension: 768
- Number of heads: 4
- Depth (layers): 6

Table A.5: ResNet50 backbone stages.

Stage	**Layers**	**Output Shape**
Stem	Conv(3→64), BN, ReLU, MaxPool	$64 \times 56 \times 56$
Stage 1	3 residual blocks	$256 \times 56 \times 56$
Stage 2	4 residual blocks	$512 \times 28 \times 28$
Stage 3	6 residual blocks	$1024 \times 14 \times 14$
Stage 4	3 residual blocks	$2048 \times 7 \times 7$

- MLP hidden dimension: 2048
- Dropout: 0.1

A.2.4 Fusion and Classification Head

Table A.6: Fusion and MLP head for ResViT classifier.

Layer	**Output Dim.**
Global Average Pooling (ResNet)	2048
Concatenate with ViT token	$2048 + 768 = 2816$
FC1 + ReLU	1024
FC2 + ReLU	128
FC3 + Softmax	3 (Normal, Benign, Malignant)

Appendix B

Hyperparameter Tables

B.1 Segmentation Training Hyperparameters

Table B.1: Hyperparameters for federated U-Net segmentation.

Hyperparameter	**Value**
Model	2D U-Net
Input Size	512×512
Loss	Dice (0.7) + BCE (0.3)
Optimizer	Adam
Learning Rate	1×10^{-3}
Weight Decay	1×10^{-4}
Batch Size	16
Local Epochs	3
FL Rounds	20
Augmentations	rotation, flip, shift-scale, elastic transforms
Clients	2

B.2 Classification (ResViT) Hyperparameters

Table B.2: Hyperparameters for federated ResViT classification.

Hyperparameter	**Value**
Input Size	224×224
Channels	3
Loss	Cross-Entropy
Optimizer	AdamW
Learning Rate	1×10^{-4}
Weight Decay	1×10^{-5}
Batch Size	32
Local Epochs	3
FL Rounds	20
Classes	Normal, Benign, Malignant

Appendix C

Core Code Snippets

This appendix presents core pseudocode and code fragments used in the implementation.

C.1 Federated Averaging (Server)

```
def fed_avg(weights_list, sizes):
    total = sum(sizes)
    new_weights = []
    for layer_idx in range(len(weights_list[0])):
        weighted_sum = sum(
            (sizes[i] / total) * weights_list[i][layer_idx]
            for i in range(len(weights_list))
        )
        new_weights.append(weighted_sum)
    return new_weights
```

C.2 U-Net Training Loop (Client)

```
for epoch in range(epochs):
    for imgs, masks in loader:
        imgs, masks = imgs.to(device), masks.to(device)
        optimizer.zero_grad()
        preds = model(imgs)
        loss = dice_loss(preds, masks) + 0.3 * bce(preds, masks)
        loss.backward()
        optimizer.step()
```

C.3 Classification Training Loop (ResViT)

```
for epoch in range(epochs):
    for imgs, labels in loader:
        imgs, labels = imgs.to(device), labels.to(device)
        optimizer.zero_grad()
        outputs = model(imgs)
        loss = ce_loss(outputs, labels)
        loss.backward()
```

```
        optimizer.step()
```

C.4 Grad-CAM XAI Implementation (Core Idea)

```
grads = torch.autograd.grad(
    outputs[:, class_id], target_layer, retain_graph=True
)[0]
weights = grads.mean(dim=(2, 3), keepdim=True)
cam = (weights * target_layer).sum(dim=1).relu()
```

Appendix D

Supplementary Figures and Tables

This appendix lists placeholders and descriptions for supplementary figures and tables.

D.1 Additional Segmentation Visual Samples

- Figure D.1: Good-quality nodule segmentation.
- Figure D.2: Boundary ambiguity failure case.
- Figure D.3: False-positive lung-structure segmentation.

D.2 Additional Classification Confusion Matrices

- Table D.1: Client 1 confusion matrices across FL rounds.
- Table D.2: Client 2 confusion matrices across FL rounds.
- Table D.3: Global model confusion matrix aggregated over all rounds.

221-35-874

ORIGINALITY REPORT

11%	9%	8%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	arxiv.org Internet Source	1%
3	ebin.pub Internet Source	<1%
4	Submitted to Universiti Malaysia Pahang Student Paper	<1%
5	"AI and Digital Transformation: Opportunities, Challenges, and Emerging Threats in Technology, Business, and Security", Springer Science and Business Media LLC, 2026	<1%

Publication

<1%

genome.ifmo.ru
Internet Source

<1%

Sheykholeslamzadeh, Sayedali. "Robust, Poisoning-Resilient Federated Learning", University of Windsor (Canada), 2025
Publication

<1%

smartech.gatech.edu
Internet Source

<1%

www.mdpi.com
Internet Source

<1%

Submitted to King's College
Student Paper

<1%

export.arxiv.org
Internet Source

<1%

sprojectng.com
Internet Source

<1%

core.ac.uk
Internet Source

<1%

Submitted to The American College of Greece Learn SaaS
Student Paper

<1%

Submitted to University of New Haven
Student Paper

<1%

Takaaki Sugino, Holger R. Roth, Mohammad Eshghi, Masahiro Oda, Min Suk Chung, Kensaku Mori. "Automatic segmentation of head anatomical structures from sparsely annotated images", 2017 IEEE International Conference on Cyborg and Bionic Systems (CBS), 2017
Publication

<1%

"Machine Learning in Medical Imaging", Springer Science and Business Media LLC, 2021
Publication

<1%

Noor ul Sehr Zia, Muhammad Ferjad Naeem, Syed Muhammad Kumail Raza, Muhammad Mubasher Khan, Adnan Ul-Hasan, Faisal Shafait. "A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition", Neural Computing and Applications, 2021
Publication

Submitted to Radboud Universiteit
Student Paper

<1%

<1%

Submitted to UNICAF
Student Paper

<1%

www.bomberbot.com
Internet Source

<1%

Submitted to Queen's University of Belfast
Student Paper

<1%

carijournals.org
Internet Source

<1%

spectrum.library.concordia.ca

Internet Source

<1%

Submitted to Florida Atlantic University

Student Paper

<1%

Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "Intelligent Data-Driven Systems with Innovations in Artificial Intelligence", CRC Press, 2025

Publication

<1%

kth.diva-portal.org

Internet Source

<1%

e-space.mmu.ac.uk

Internet Source

<1%

pmc.ncbi.nlm.nih.gov

Internet Source

<1%

theses.hal.science
Internet Source

<1%

Submitted to Baylor University
Student Paper

<1%

Submitted to Bournemouth University
Student Paper

<1%

Elhussein, Ahmed. "Federated Collaboration: Addressing Challenges in Data Sharing and Learning Under Statistical Heterogeneity", Columbia University, 2025
Publication

<1%

Submitted to Indian Institute Of Information Technology, Kalyani
Student Paper

<1%

Mobiny, Aryan. "Representation Learning With Less Label and Imperfect Data", University of Houston
Publication

<1%

dcatkth.github.io

Internet Source

<1%

Submitted to Trinity College Dublin
Student Paper

<1%

mospace.umsystem.edu
Internet Source

<1%

"Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries", Springer Nature, 2019
Publication

<1%

"Image Analysis and Recognition", Springer Science and Business Media LLC, 2019
Publication

<1%

Submitted to The University of Manchester
Student Paper

<1%

Submitted to University of Greenwich
Student Paper

<1%

Submitted to Vrije Universiteit Amsterdam

Student Paper

<1%

biomedical-engineering
online.biomedcentral.com
Internet Source

<1%

cdcl.umd.edu
Internet Source

<1%

d197for5662m48.cloudfront.net
Internet Source

<1%

Bagdasaryan, Eugene. "(Un)Trustworthy Machine Learning", Cornell University, 2023
Publication

<1%

Submitted to Korea Advanced Institute of Science and Technology
Student Paper

<1%

Submitted to Liverpool John Moores University
Student Paper

<1%

dspace.daffodilvarsity.edu.bd:8080
Internet Source

<1%

human.ait.kyushu-u.ac.jp
Internet Source

<1%

thesis.library.caltech.edu
Internet Source

<1%

www2.eng.cam.ac.uk
Internet Source

<1%

Submitted to Cyber Security
Student Paper

<1%

Submitted to University of Lincoln
Student Paper

<1%

repositories.lib.utexas.edu
Internet Source

<1%

samuelalbanie.com

Internet Source

<1%

www.medrxiv.org
Internet Source

<1%

1library.net
Internet Source

<1%

Submitted to Leiden University
Student Paper

<1%

jast-journal.springeropen.com
Internet Source

<1%

lra.le.ac.uk
Internet Source

<1%

Arun Kumar Singh, Sandeep Saxena, A.K. Haghi. "Cybernetic Shield - Securing the Future of Machine Intelligence", CRC Press, 2025
Publication

<1%

Jiddu Krishnan O P, Pinki Roy. "Enhanced Pulmonary Nodule Segmentation via Radial Attention U-Net with Delaunay TV Regularisation", Engineering Research Express, 2025

Publication

<1%

Kumagai, Samuel F.. "DeepVoid: Physics Aware Semantic Segmentation of Large-Scale Structure With Deep Learning.", Drexel University

Publication

<1%

Lalit Mohan Goyal, Tanzila Saba, Amjad Rehman, Souad Larabi-Marie-Sainte. "Artificial Intelligence and Internet of Things - Applications in Smart Healthcare", CRC Press, 2021

Publication

<1%

ourspace.uregina.ca

Internet Source

<1%

www.isca-archive.org

Internet Source

<1%

www.joig.net

Internet Source

<1%

www.science.gov

Internet Source

<1%

"Medical Image Computing and Computer Assisted Intervention – MICCAI 2019", Springer Science and Business Media LLC, 2019

Publication

<1%

Hangyu Zhu, Liyuan Huang, Zhenping Xie. "Privacy attack in federated learning is not easy: an experimental study", Complex & Intelligent Systems, 2025

Publication

<1%

Li, Ao. "Tackling Data and Resource Heterogeneity for Performance Enhancement of Collaborative Learning Systems.", Hong Kong University of Science and Technology (Hong Kong)

Publication

<1%

Shangyu Yang, Lingtao Mao, Mei Zhou, Xuyang Chang, Jiaojiao Liu, Dongxiao Li, Jinting Bai. "A Fine-Segmentation Algorithm for XCT Images of Multiphase Composite Building Materials Based on Deep Learning", Journal of Building Engineering, 2024

Publication

<1%

Submitted to University of Glasgow

Student Paper

<1%

gupea.ub.gu.se

Internet Source

<1%

jcrb.net

Internet Source

<1%

luoxd1996.github.io

Internet Source

<1%

technodocbox.com

Internet Source

<1%

umpir.ump.edu.my
Internet Source

<1%

www.hort.uconn.edu
Internet Source

<1%

www2.mdpi.com
Internet Source

<1%

Akter, Mahmuda. "Federated Learning-Based Privacy Protection Methods in Smart Healthcare Systems.", University of New South Wales (Australia)
Publication

<1%

Wei, Wenqi. "Adversarial Resilient and Privacy Preserving Deep Learning.", Georgia Institute of Technology
Publication

<1%

Wenzhe Wang, Ruiwei Feng, Jintai Chen, Yifei Lu, Tingting Chen, Hongyun Yu, Danny Z. Chen, Jian Wu. "Nodule-Plus R-CNN and

Deep Self-Paced Active Learning for 3D Instance
Segmentation of Pulmonary Nodules", IEEE Access, 2019
Publication

<1%

"Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries", Springer Science and Business Media LLC, 2023
Publication

<1%

"Machine Learning in Medical Imaging", Springer Science and Business Media LLC, 2019
Publication

<1%

Suriyavarman S, Arockia Xavier Annie R. "Lung Nodule Segmentation and Classification using U-Net and Efficient-Net", International Journal of Advanced Computer Science and
Applications, 2023
Publication

<1%

Yadav, Anil. "Mitigating Variability in CT

Parameters: Impacts on Image Analysis and Predictive Modeling", University of California, Los Angeles

Publication

Exclude matches Off

Exclude quotes Off

Exclude bibliography Off