# Integrating Explainable AI with Federated Based Deep Learning for Accurate and Transparent Lung Cancer Classification


**Submitted By**

**MD RAYHAN KHAN**

221-35-831

Department of Software Engineering

Daffodil International University

**Supervised by**

**MR. MD. SHOHEL ARMAN**

Assistant Professor

Department of Software Engineering

Daffodil International University


A thesis submitted in partial fulfillment of the requirement for the degree of

Bachelor of Science in Software Engineering

Fall-2025

Integrating Explainable AI with Federated Based Deep Learning for Accurate and Transparent Lung Cancer Classification

MD RAYHAN KHAN

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

## APPROVAL

This thesis titled on **"Integrating Explainable AI with Federated Based Deep Learning for Accurate and Transparent Lung Cancer Classification"**, submitted by **Md Rayhan Khan (ID: 221-35-831)** 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. Imran Mahmud**
**Professor & Head**
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

---

**Internal Examiner 1**

**Afsana Begum**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---

**Internal Examiner 2**

**Md. Shohel Arman**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---

**Internal Examiner 3**

**Nadira Islam**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

---

**External Examiner**

**Md Manowarul Islam**
**Professor**
Department of Computer Science and Engineering
Jagannath University, Bangladesh

# DAFFODIL INTERNATIONAL UNIVERSITY

**DECLARATION OF THESIS AND COPYRIGHT**

| | | |
|---|---|---|
| Author's Full Name | : | MD Rayhan Khan |
| Date of Birth | : | 04 July, 2002 |
| Title | : | Integrating Explainable AI with Federated Based Deep Learning for Accurate and Transparent Lung Cancer Classification |
| Academic Session | : | Fall 2025 |

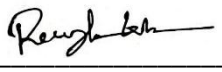
I declare that this thesis is classified as:

| | | |
|---|---|---|
| ☑ | CONFIDENTIAL | (Contains confidential information under the Official Secret Act 1997) * |
| ☑ | RESTRICTED | (Contains restricted information as specified by the organization where research was done) * |
| ☐ | OPEN ACCESS | I agree that my thesis to be published as online open access (Full Text) |

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

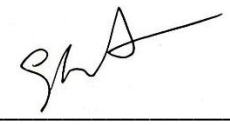
Certified by:



(Student's Signature)

221-35-831

Student ID
Date: 27, November 2025



(Supervisor's Signature)

Mr. Md. Shohel Arman

Name of Supervisor
Date: 27, November 2025

NOTE: * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

# SUPERVISOR's DECLARATION

I hereby declare that I have checked this thesis and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science

(Supervisor's Signature)

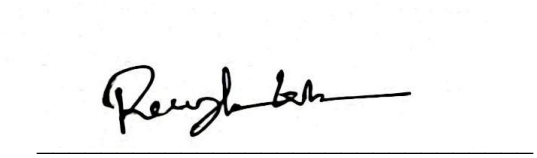
Full Name : Mr. Md. Shohel Arman

Position : Assistant Professor

Date : 27 November 2025

# STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.



(Student's Signature)

Full Name : MD Rayhan Khan

ID Number : 221-35-831

Date : 27 November 2025

# Integrating Explainable AI with Federated Based Deep Learning for Accurate and Transparent Lung Cancer Classification

MD Rayhan Khan

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

November, 2025

# Acknowledgement

I have always liked exploring new ideas and figuring out how things work, and this naturally led me to Machine Learning and its use in lung cancer image analysis. Working with these images and trying to build models and Federated pipelines that can support cancer detection feels valuable to me, and I am grateful to Almighty for giving me the opportunity, strength, and patience to continue this work.

I am deeply grateful to my parents for always standing beside me. Their support, prayers, and constant encouragement have helped me reach this stage of my life and complete this research.

I would also like to thank Dr. Imran Mahmud, Head of the Department of Software Engineering, and all my teachers at the department for guiding me throughout my studies and helping me build up the knowledge needed for this work. I am thankful to Daffodil International University for providing a learning environment where I could grow both academically and personally.

My special thanks go to my supervisor, Md. Shohel Arman, for his time, guidance, and constructive feedback. His continuous supervision and support made it possible for me to stay on track and complete this thesis.

Lastly, I want to thank my batchmates and friends from DIU for their cooperation, late-night discussions, and constant motivation. Their presence made the whole journey easier and more enjoyable, and I truly appreciate the support they have given me.

# Abstract

Lung cancer is still the most common cause of cancer death, and the best way to improve survival is to move the diagnosis from late to early-stage disease. However, high-performing models are often trained on small, separate datasets and work like "black boxes," which makes it hard for institutions to work together and for patients to trust the models. We show how to combine Federated Learning (FL) and Explainable AI (XAI) in a way that makes lung cancer detection accurate, private, and clear. We trained six deep learning backbones on decentralized data using the FedAvg algorithm, which means we didn't have to move patient images off-site. Our framework introduces federated explainability: clients create local Grad-CAM visualizations and a quantitative faithfulness metric (Deletion AUC) at the same time. The central server combines these scores to keep track of the model's trustworthiness on a global, round-by-round basis. Our findings indicate that accuracy and transparency can be attained concurrently. The DenseNet-121 and HVR-18 (Hybrid ViT–ResNet-18, MLP) models that were suggested were the best-performing ones, with validation F1-scores of 0.9678 and 0.9677, respectively. The HSD-121 (Hybrid Swin-T + DenseNet-121, MLP) model also did very well, with a validation F1-score of 0.9555. DenseNet-121 had a Deletion AUC of 0.36 for federated explainability, while HVR-18 and HSD-121 had scores of about 0.38 and 0.33, respectively. This means that Grad-CAM-based explanations were reliable for all three architectures. We saw a strong positive relationship between the global F1-score and the aggregated Deletion AUC during training. Local heatmaps provided additional validation that models incrementally acquired the ability to concentrate on diagnostically pertinent features. These results confirm a new framework in which privacy, accuracy, and interpretability increase together, providing a clear path for creating and keeping an eye on reliable clinical AI in real-world, multi-institutional settings.

# Table of Contents

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

## Introduction

Lung cancer is still a big health problem in many regions of the world. It is still one of the most common reasons for cancer deaths in the world [1]. Quantitative evaluations by the World Health Organization (WHO) and other international health organizations underscore its alarming prevalence, with millions of new cases documented each year [2]. In this therapeutic context, timely and accurate diagnosis is unequivocally recognized as the principal factor in improving patient prognosis and long-term survival rates [3]. For a long time, the best way to find out what was wrong with someone was to look at medical photos. The approach normally starts with CT scans to locate and describe lung nodules. Then, histology slides are looked at to determine out the specific subtype [4]. But it could be hard to trust other people to read photos. It takes a lot of time and effort to look at pictures of wounds and illnesses, and you need extensive training [5]. This strategy slows down the clinical workflow and has a lot of differences amongst observers, which indicates that various specialists could come to different conclusions [6]. Sadly, this lack of clarity, together with mental fatigue from intense workloads, can lead to delays or mistakes in diagnosis, which can have a direct impact on how well patients do.

## Background:

Because of these clinical issues, Artificial Intelligence (AI), especially Deep Learning (DL), has become a game-changing tool for looking at medical images [7]. Convolutional Neural Networks (CNNs) and other advanced architectures have shown a lot of promise for detecting complicated visual patterns. Attention-based techniques [4] and ensemble models (8) are two new ideas that have made things even better.

Deep learning algorithms can learn on their own and find complex, high-dimensional patterns in X-ray images that most people can't perceive because they are too little or too hard to see.

There is a lot of proof that these models work in the actual world. Many studies have demonstrated that DL models can be more accurate for several essential diagnostic tasks.

These include not only the initial identification of lung nodules but also the classification of cancer subtypes (3) and even the prediction of patient responses to certain therapies (9). Deep learning models have shown that they can do at least as well as, and maybe better than, experienced human experts on a number of tests [1].
There is a lot of evidence that these models work in the actual world. Many studies have demonstrated that DL models can be more accurate for a number of essential diagnostic tasks. These include not only the initial identification of lung nodules but also the classification of cancer subtypes [3] and even the prediction of patient responses to certain therapies [9]. Numerous assessments [1] have shown that deep learning models can perform at levels that are equal to or better than those of experienced human professionals.

# Problem Statement

It is very important for doctors to be able to quickly and accurately find lung cancer in medical imaging, yet deep learning (DL) is still not widely used because of two problems that are closely related: data access and model transparency. High-capacity deep learning models can perform well on CT and similar modalities, but they frequently don't do well when applied to data from places other than where they were developed unless they were trained on big, diverse, multi-institution datasets [13; 14]. It is sometimes impossible to centralize this kind of data because of privacy laws, governance, and operational issues that keep institutions apart, which limits the quantity and variety needed for strong external validity [15; 16]. Many accurate architectures, such as contemporary attention and hybrid designs, function as "black boxes," which makes it hard for clinicians to trust them and use them safely in high-stakes diagnostic processes [17; 18]. While new research combines explainable AI (XAI) with improved backbones, best practices that provide accurate, useful explanations and protect privacy on a large scale are still developing [19; 20].

We are working on a way to learn from distributed medical data while yet keeping patient information private and making sure that the information is clinically useful. Our specific goal is to create and test a federated, explainable method for analyzing lung cancer that

(i) trains together across sites without combining private health information,

(ii) Maintains good diagnostic performance despite differences between institutions, and

(iii) Creates rationales that can be checked against clinical reasoning to promote safe deployment at several centers [15; 14]. The proposed direction seeks to bridge the practical divide between attractive research outcomes and reliable real-world application by integrating privacy-preserving collaboration with interpretable modeling [16; 17].

# Research Gaps

Centralized deep learning has clearly pushed the field of computer-assisted lung cancer diagnosis forward. Prior work on CNNs and related architecture has identified useful design choices for backbones, data preprocessing, transfer learning, and assembling [24, 25]. These systems often report very strong performance on benchmark datasets. However, most of these studies are trained and evaluated in tightly controlled, single-site settings. As a result, they still face major issues with generalizability, external validation, and routine deployment in real hospitals [2, 14]. When models are only tested on homogeneous cohorts, it remains unclear whether the reported gains will hold across different scanners, institutions, and patient populations.

Lesion-focused segmentation networks, such as U-Net–style architectures, have helped by improving nodule localization and enabling earlier detection that more closely matches radiologists' expectations [20]. Even so, these methods usually assume that all data can be pooled into a single centralized repository. In practice, that assumption is often unrealistic because of privacy regulations, institutional policies, and ethical concerns. Many hospitals

cannot or will not share raw CT scans, which keeps models siloed and limits the benefit of centralized deep learning [11, 12, 15]. Explainability introduces another layer of unmet need.

Hybrid DL+XAI models for example, CNN/transformer pipelines that include saliency or attention maps—have shown that it is possible to add interpretability without completely sacrificing accuracy (19). Yet, in most cases, explainability is treated as a side product: a local, post-hoc visualization presented after training a centralized model. There are very few frameworks in which interpretability is a core design goal, is evaluated systematically, and is used to guide model selection—especially when the data come from multiple, heterogeneous sites.

Empirical work in lung cancer imaging shows that federated learning can reach, and in some cases even exceed, the performance of conventional centralized models. It has been used to improve multi-site classification and to support tasks such as predicting treatment response in NSCLC cohorts [9, 22]. Even so, several important gaps remain.

First, most federated learning studies still focus mainly on accuracy and privacy. Only a small number of works try to make the decision process itself understandable, and there is still no widely adopted framework for lung CT that combines explainability, privacy protection, and high performance in a single, coherent design [10, 19].

Second, current research usually evaluates only a limited set of backbone architectures at a time. A careful, side-by-side comparison of modern model families—such as ConvNeXt- or FocalNet-inspired CNNs and recurrent or hybrid transformer models—within one unified FL pipeline is still missing [24, 25]. This makes it hard to know which architectures are truly best suited for decentralized lung cancer CT under realistic, non-IID data distributions.

Third, surveys across breast, lung, and prostate imaging highlight broader open issues around fairness, standardized benchmarks, and long-term robustness. Many FL studies still lack rigorous cross-site validation and do not fully examine how performance changes across different patient subgroups or over time [10, 11]. Personalized and lifelong FL strategies have been proposed to address these concerns [26], but concrete, lung-focused implementations that maintain stable performance while the data distribution evolves are still rare. Finally, there is an operational gap between algorithm design and real-world clinical workflows. While several papers describe orchestration, security, and compliance aspects of multi-institutional FL deployments at a high level [12, 15], they often stop short of integrating these ideas with day-to-day radiology practice—for example, providing explanations that radiologists can easily interpret, handling strongly non-IID client data, or supporting continuous model monitoring and updates.

Taken together, these gaps indicate that the field needs to move beyond (i) purely centralized CNN pipelines and (ii) accuracy-only FL studies. There is a clear opportunity to develop an *explainable, privacy-preserving federated learning framework for lung cancer CT* that (a) leverages diverse, modern backbones, (b) remains robust under realistic, non-IID multi-site conditions, and (c) treats interpretability as a measurable, first-class objective rather than an optional add-on.

## Objectives

The main objective of this study is to develop a privacy-preserving federated learning framework for lung cancer CT image classification that allows multiple institutions to collaborate without sharing raw imaging data. Within this framework, we aim to systematically compare several modern backbone architectures, including CNN-based and transformer-inspired models, under realistic non-IID, multi-site data distributions. A further objective is to integrate explainable AI directly into the federated workflow by combining qualitative visualization methods such as Grad-CAM++ heatmaps with quantitative faithfulness metrics, so that the reasoning of the models over lung lesions can be critically evaluated. In addition, we seek to assess the proposed framework in terms of both accuracy and robustness through cross-site validation and comparisons with centralized baselines. Finally, we aim to design a privacy-preserving federated learning pipeline that achieves accurate lung cancer detection while providing transparent, explainable AI–based reasoning.

## Motivation

This research is driven by the necessity to simultaneously tackle the two predominant obstacles hindering the clinical implementation of deep learning for lung cancer: scalable privacy-preserving data access and reliable model transparency. Federated Learning (FL) immediately addresses the data-silo issue by facilitating collaboration among multiple institutions without the need to share raw pictures.

Recent surveys, systems articles, and cancer applications demonstrate that FL is both practical and successful [21]. FL has empirically matched or exceeded centralized baselines in heterogeneous client distributions and has been verified on lung cancer tasks, including classification and treatment-response prediction [22]. These results indicate that a meticulously designed FL pipeline can leverage the diversity of multisite datasets while adhering to institutional and regulatory requirements [10]. Nevertheless, privacy alone is inadequate for clinical adoption. Clinicians need to know why a system marks a lesion or suggests a class. Reviews always say that explainability is necessary for trust and safe use [17]. Recent efforts integrating deep learning with explainable AI, saliency techniques, and hybrid architectures demonstrate potential for producing clinically interpretable justifications [23]. Additionally, studies focused on FL are increasingly underscoring that interpretability must align with privacy considerations and be scalable across several sites [10]. Based on this research and what we've done thus far, we are working on a federated, explainable way to diagnose lung cancer that:

I. trains multiple high-capacity backbones collaboratively via FedAvg on decentralized data without pooling protected health information (PHI);
II. standardizes client-side preprocessing and evaluation to remain robust under non-IID, cross-institution variation;

III. generates client-local explanations (Grad-CAM++) and computes a quantitative faithfulness score (Deletion AUC) alongside predictions;
IV. combines these local faithfulness scores on the server side to create a global trust metric that can be tracked with performance (F1) throughout training;
V. gives governance hooks, round-wise alerts when trust metrics drop, auditable heatmap evidence, and criteria for early stopping or rollback; and
VI. checks deployability by comparing backbones and reporting accuracy–explainability trade-offs at scale, allowing for networkwide oversight instead of site-specific post hoc checks.

## Summary

This chapter introduces lung cancer as a major global health burden where early, accurate CT-based diagnosis is vital but still heavily limited by manual image reading, inter-observer variability, and clinical workload. It explains how centralized deep learning and modern backbones (CNNs, attention, ensembles, segmentation networks) have improved detection and subtype classification, yet still struggle with generalizability, external validation, data-sharing constraints, and "black-box" decision making. The research gap is framed around the lack of an explainable, privacy-preserving federated learning (FL) framework for lung cancer CT that (i) works across non-IID, multi-institution data, (ii) systematically compares diverse architectures, and (iii) treats interpretability as a core, measurable objective. The study therefore aims to design a federated pipeline (using FedAvg) that lets multiple sites train high-capacity models without sharing raw images, while integrating XAI through Grad-CAM++ and a quantitative faithfulness metric (Deletion AUC) that is aggregated server-side as a global "trust" signal alongside F1. The motivation centers on closing the gap between promising DL research and real-world deployment by combining privacy, robustness, and transparent, auditable explanations that clinicians can rely on for safer multi-center lung cancer diagnosis.

# CHAPTER 2
# LITERATURE REVIEW

## Introduction

I review the existing body of work related to automated lung cancer diagnosis, with a particular focus on deep learning, federated learning, and explainable AI. By systematically examining prior studies, I aimed to understand which model architectures, training strategies, and datasets have been used for lung cancer CT analysis, how well these approaches perform in practice, and where they fall short. This includes centralized CNN-based pipelines, attention and hybrid transformer models, lesion-focused segmentation networks such as U-Net, and more recent efforts that apply federated learning to multi-site medical imaging and combine it with saliency-based explanation methods.

Through this review, I not only gained a broad overview of the technical landscape but also identified several recurring limitations: restricted generalizability due to single-center data, privacy constraints that prevent large-scale data sharing, and a lack of transparent, trustworthy explanations for model decisions. These gaps directly motivate the direction of my own research. Building on the strengths of earlier work while addressing their shortcomings, my study proposes an explainable, privacy-preserving federated learning framework for lung cancer CT that compares multiple modern backbones under non-IID, multi-institution conditions and treats interpretability as a measurable, central objective rather than an afterthought.

## Previous Literature:

I looked over the most important publications that have transformed how lung cancer is automatically identified using several methodologies. I look at both old and new research on centralized deep learning models, segmentations, and attention-based architectures in neural networks. For example, federated and explainable AI are two new ideas. I concentrate on their methodologies, datasets, and primary findings. These publications provide the technological foundation and justification for the paradigm proposed in this thesis.

Deep learning has made it easier for computers to discover lung cancer by helping them comprehend the best CNN architectures, data features, and transfer learning methodologies [13]. Using single backbones is one way to make it easier to tell the difference between chest CT scans. You can also use CNNs on purpose to pick out and combine characteristics. This shows that using different models is important for good categorization [24]. Lesion-focused segmentation networks, especially U-Net, have improved localization and made it easier to find lung nodules in CT scans early on, which is more in line with what doctors see in real life [20]. Architectures that use sparsity and attention have made it easier to group different types of non-small cell lung cancer by producing representations that are smaller and more unique [18]. ConvNeXt and FocalNet are two new convolutional backbones built for thoracic CT that keep

improving the bar for accuracy in similar data pipelines [25]. A recent, comprehensive investigation consolidates these findings and indicates persistent issues regarding generalizability, external validation, and therapeutic use [14]. To mitigate trust concerns, hybrid DL+XAI pipelines, such as DCNN–ViT–GRU with saliency, have demonstrated the integration of interpretable logic while maintaining accuracy [19]. Surveys on federated learning for medical imaging demonstrate how decentralized training may protect privacy and extract insights from multi-site cohorts in non-IID contexts [16]. Theoretically, numerous groups can communicate without violating regulations or compromising their privacy. For example, systems that employ encrypted and auditable federated training [15] in lung oncology have proven that working together on federated techniques can make multi-site categorization better without needing to share raw data [22]. Federation has helped clinical goals in more ways than only discovering people with NSCLC. It has also helped us make educated guesses about how people from different areas could respond to treatment. This shows that it is helpful in therapy [9]. Decentralized ensembles have provided precise diagnoses for multi-order lung cancer, illustrating that several model types remain beneficial in a federated system [8]. CNNs that are made just for federated contexts do better on lung tests. This demonstrates that the combination of decentralized optimization and attention approaches is more effective [4]. When done correctly, diversity, aggregation, and assessment show that federated learning can perform just as well as or even better than centralized baselines [2].

## Summary:

The literature review shows that deep learning has significantly advanced automated lung cancer diagnosis, especially through CNNs, attention-based and hybrid architectures, and lesion-focused segmentation networks like U-Net. These models achieve strong performance on CT-based classification and detection tasks, but most of the system trained in centralized, single-site settings, which limits generalizability, external validation, and real-world deployment. Recent works introduce explainable AI and demonstrate that saliency and attention mechanisms can provide more interpretable decisions, while federated learning studies in lung and other cancers indicate that decentralized training can protect privacy and still match or surpass centralized baselines on multi-site data. However, open challenges remain around fairness, benchmarking, robustness over time, and the lack of an explicitly explainable, privacy-preserving FL framework tailored to lung cancer CT—precisely the gap this thesis aims to address.

# CHAPTER 3
# METHODOLOGY

## Introduction:

We provide privacy-preserving, explainable training of lung-image classifiers across several hospitals using a synchronous federated learning process, as illustrated in Figure 1. We used four clients in our tests, but the procedure can easily be expanded to include more sites without changing the protocol. Each client adjusts a backbone on images that don't have any identifying information. To make sure that outputs are the same across architectures, we replace the task head with a single classifier that includes Flatten → Dense+ReLU → Dense+softmax for multiclass prediction. We chose DenseNet-121 as our main model because it has the best validation performance and is stable across many sites. HVR-18 (Hybrid ViT–ResNet-18, MLP) is our second baseline since it is easy to compute. Training happens in rounds, and rigorous privacy rules mean that no raw photos or identifiers can leave a hospital at any moment. The coordinator sets up the global model and the total number of communication rounds when the server starts up. The server sends the current global weights θr to all clients at the start of round r. Each client loads these weights and then trains on its own data for a set number of epochs. It then calculates scalar metrics (loss, accuracy, macro-F1) and uses Grad-CAM on a tiny validation probe to summarize explainability along with a lightweight faithfulness score (Deletion AUC). After finishing, the client only uploads the modified model weights and scalar summaries, never images or heatmaps, and then waits for aggregation. After the local client finishes uploading in round r, it stays idle and waits for more responses from the server (the end of aggregation and the following broadcast of θr+1). The server works in a several synchronized ways, meaning that it only adds up data after getting updates from all clients round by round wise. The server waits for the slower client to send the update before moving on to the next step in the aggregation process. After getting all the updates, the server uses sample-size–weighted averaging (FedAvg) to make new global weights θr+1, records the round's performance and transparency metrics, and sends the revised model to all clients again. Each client quickly swaps out its local weights for θr+1 and starts the next cycle of local training. This cycle of broadcasting, sending local trains, uploading, waiting, aggregating, and redistributing goes on until all intended rounds are done.

The server keeps track of round-wise curves (loss, accuracy, macro-F1) and combines clients's redistribute cycle repeats until all intended rounds are done. The server keeps track of round-wise curves (loss, accuracy, macro-F1) and combines clients' Deletion-AUC summaries to keep an eye on transparency trends without looking at any pictures. The server saves both the best checkpoint chosen by validation macro-F1 and the final global model at the end of training. The end result is a useful, verifiable workflow that lets people work together without sharing data, with the same classifier heads across backbones, updates that happen at the same time for all hospitals, and explainability signals that become better as performance improves.

# Proposed Framework:

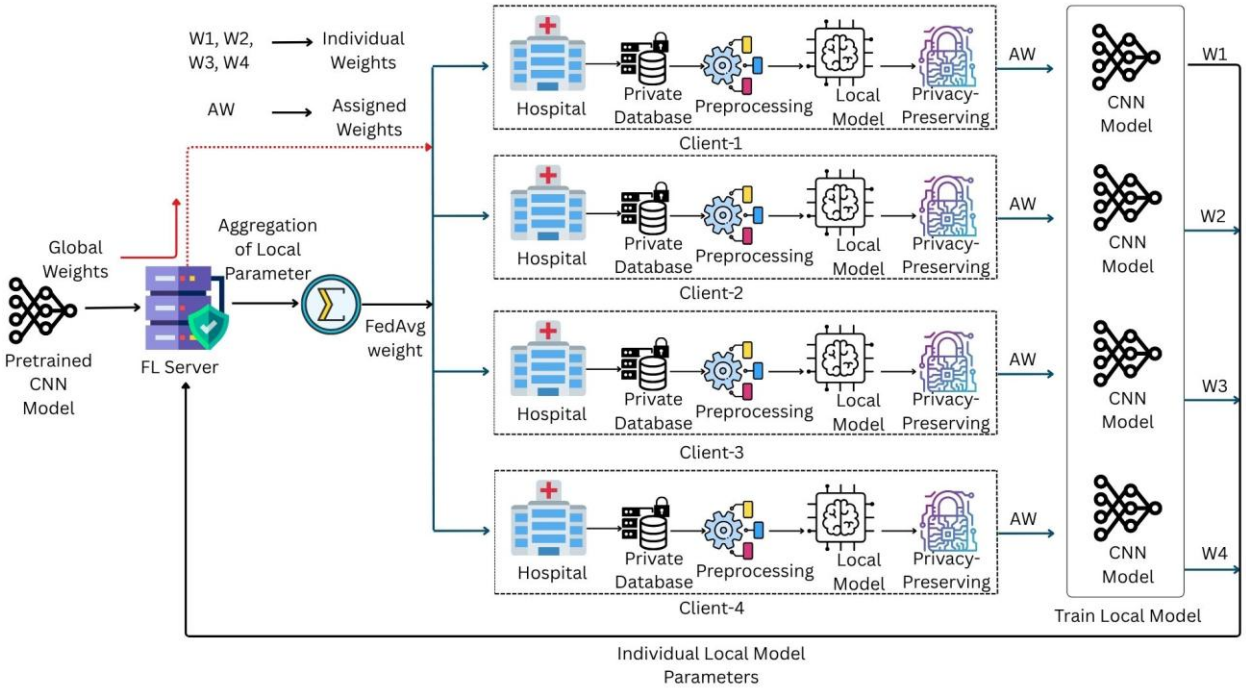



Figure 1: Synchronous FL workflow for lung Cancer

# Dataset Splitting:

Each client does the split on its own dataset: 80% for training, 10% for validation, and 10% for testing. This keeps privacy preserving and shows how things are spread out in federated learning. Splits are done by patients to keep data from leaking between partitions, and when practicable, they are done by class to reduce imbalance. Images are standardized to 224 x 224 grayscale (1 channel) with normalization that is always the same. Training employs light, label-preserving augmentation, whereas validation and testing are deterministic. The server just gets per-split counts and scalar metrics like loss, accuracy, and macro-F1. It doesn't get any raw images or IDs. When 10% divisions don't divide evenly, the counts are rounded and changed so that train+validation+test equals the total for the site.

| Client | Benign | Malignant | Normal |
|---|---|---|---|
| Client-1 | 800 | 845 | 800 |
| Client-2 | 1000 | 1050 | 1000 |
| Client-3 | 600 | 700 | 600 |
| Client-4 | 1000 | 1050 | 1050 |
| Totals | 3400 | 3645 | 3450 |

Table 1: Per-Client Distribution Of Lung CT Image Counts by Class

| Client | Train | Test | Validation |
|---|---|---|---|
| Client-1 | 1956 | 244 | 245 |
| Client-2 | 2440 | 305 | 305 |
| Client-3 | 1520 | 190 | 190 |
| Client-4 | 2480 | 310 | 310 |
| Totals | 8396 | 1049 | 1050 |

Table 2: Per-Client Dataset Partition Using An 80/10/10

# Federated Learning:

Federated learning (FL) lets a lot of clients (like hospitals, institutions, or edge devices) work together to train a shared model without sending raw data to a central server. We use a synchronous FedAvg protocol (see Fig. 2 for an overview). At the start of round r, the server sets up the global model and sends its weights θr to all clients. Each client loads θr, trains on its own de-identified data for a set number of epochs, and then calculates scalar summary (loss, accuracy, macro-F1) as well as an explainability faithfulness score (Deletion AUC) based on Grad-CAM probes on the client side. Clients only send the modified model weights (or deltas) and scalar summaries, not pictures or heatmaps. The server follows a barrier-synchronized pattern: it waits for updates from all clients in round r, uses sample-size–weighted averaging (FedAvg) to get new global weights θr+1, logs performance and transparency indicators (like the aggregated Deletion AUC), and sends θr+1 back out. By doing this a set number of times, you may make a global model that uses the knowledge of everyone involved while still keeping their anonymity. It also gives you a way to check the model's trustworthiness at the network level.

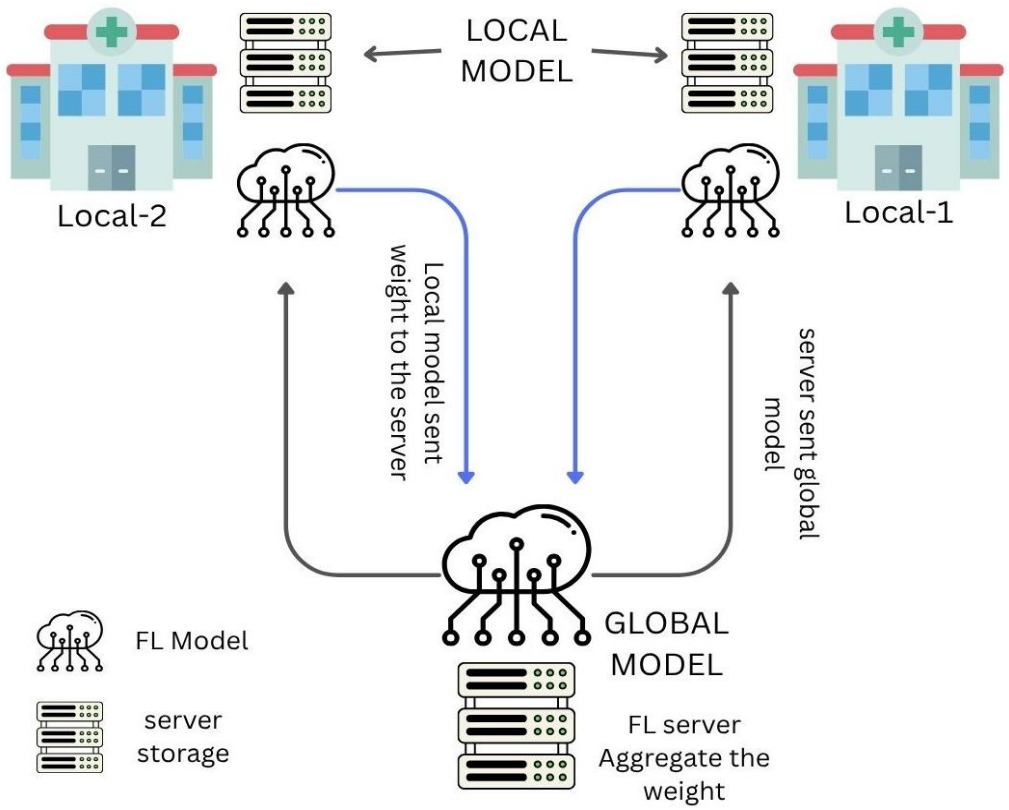



Figure 2: Federated Workflow Client & Server

# Experimental Setup:

In this portion we will talk about how we created and tested our lung cancer classification models in Federated Pipeline within details process. First, we talk about how we made the dataset better and added to it. Next, we talk about the fine-tuning process and the general training settings that all backbones employ. Then we talk about how we built and set up each model, like HVR-18, DenseNet-121, HSD-121, ResNet-50, MobileNetV3-Large, and the Custom CNN. This will help you understand what each model does and how they are different from each other in our framework.

## Dataset Preprocessing and Augmentation:

Images are all saved in folders with names that match the class and are loaded straight from disk when the program runs. The FL pipeline loads the photos directly from the disk as it runs. The label set originates from the way the files are organized, and the class names don't have to be set in stone.

Each scan is read as a single-channel (grayscale) image and resized to a consistent spatial resolution of 224×224. This makes sure that all clients have the same model input. When you load medical photographs, they go through some basic processing, such balancing the brightness and improving the contrast. After that, the pictures are turned into tensors and all of

them are given the same range of numbers. A deterministic validation/test pipeline (resize → normalize → tensor) makes sure that the test may be done again. To make the model more universal, we employ lightweight training-time augmentations that keep the labels and seem like real changes in lung CT collection. These are small changes in the plane, optional flips, and small changes in brightness and contrast. These changes are meant to be careful such that they don't modify the pathophysiology, but they still make the model subject to acquisition variability. We use Albumentations to do augmentations, and the last step changes the tensor. Normalization makes sure that all backbones get the same size of input. After preprocessing, the dataset is split into three portions for each client: 80% for training, 10% for validation, and 10% for testing. Stratification is employed to keep the class from being too big. There is a PyTorch DataLoader for each split. When training, shuffling is on, and when validating and testing, deterministic iteration is on. You can alter the number of workers, the size of the batch, and the pinned memory to make it work with your hardware. We use the training split to find the right class weights for losses that are very expensive. This helps maintain everything in balance while the optimization is going on. You can utilize optional visual sanity checks (mini-batch grids) before federated training starts to make sure that everything is working right, including preprocessing, labeling, and augmentations.

(a) Bengin (b) Malignant (c) Normal

Figure 3: Sample Grayscale of lung CT slices by class: Benign, Malignant, Normal.

## Fine-Tuning Parameters:

This study optimizes convolutional and hybrid backbones on lung CT images inside a synchronous federated learning loop. All clients use the same hyperparameter template unless otherwise stated. The server sends round-wise overrides (learning rate, local epochs, and loss type) to maintain optimization stable across sites. Initialization of the backbone. Each client creates an instance of the chosen model (mobilenetv3, resnet50, customcnn, HVR-18 (Hybrid ViT–ResNet-18, MLP, HSD-121 (Hybrid Swin-T-DenseNet-121, MLP), densenet121) and changes the last layers to fit a 3-class head. All models are trained in single-channel 5 (grayscale) mode at 224 x 224, and the input pipeline is the same for all clients so that results can be repeated. Regularization and optimization. We employ AdamW with a weight decay of $1 \times 10{-}4$. Clients begin with a base learning rate of $1{\times}10{-}3$, until the server changes it. We set the learning rate to $1{\times}10{-}3$ for all rounds and the number of local epochs to E = 6 unless otherwise noted. If the server doesn't set these, the client uses its own settings (LR = $1 \times 10{-}3$,

E = 8). Functions for loss. The default criterion is class-weighted cross-entropy. This means that class weights are based on each client's training distribution to make sure that the classes are not too different from each other. The server can turn on two alternatives: Focal Loss ($\gamma$ = 2) and Label Smoothing ($\varepsilon$ = 0.1). LR scheduler. You can use a Reduce-on-Plateau type scheduler with 10 validation checks. This works with the server's coarse round-wise LR annealing. Loaders and batching. The default number of clients in a batch is 16, but you can change this for each run. Training loaders mix things around and drop the last completed batch, while validation/test loaders are set in stone. Data loader workers and pinned memory are set up to work with the host's capabilities. Device/runtime. When a GPU is available, training runs on it; otherwise, it runs on a CPU. To keep from oversubscribing, CPU-only hosts have basic thread restrictions. Metrics and checkpoints. Each client keeps its best local weights for recovery, but only sends the server model parameters and scalar metrics like train/val loss, accuracy, and macro-F1. The server keeps track of round-wise curves, keeps the best global checkpoint by validation macro-F1, and saves the final global model at the end of training.

# Model Specification:

All backbones take in single-channel (grayscale) lung CT slices that have been scaled to 224 × 224 and output logits for three classes: Benign, Malignant, and Normal. Input normalization (mean/variance per channel) and classifier heads are made to work together so that models can be compared directly in the FL loop. Unless otherwise noted, the final pre-classifier feature map of each model is used to calculate Grad-CAM saliency and Deletion-AUC fidelity, which are then given locally for each site.

## HVR-18 (Hybrid ViT–ResNet-18, MLP):

Architecture. A dual-path model that combines global self-attention with local convolutional bias. The ViT branch utilizes vit_base_patch16_224 as a feature extractor, however it doesn't have a classification head (27). Transformers only work with RGB, thus the single CT channel is copied to three channels before patch embedding. The CNN branch is a ResNet-18 (28) that has been altered to work with grayscale images and has had its last completely connected (FC) layer removed. Let zvit ∈ R dv and zcnn ∈ R dc represent pooled embeddings. We combine them by concatenating h1 = ReLU(W1z + b1), h2 = Dropout0.3 ReLU(W2h1 + b2), and yˆ = W3h2 + b3 ∈ R 3. We specifically use Linear(dv+dc → 512) → ReLU → Dropout(0.3) → Linear(512 → 128) → ReLU → Linear(128 → 3). This hybridization is based on the fact that combining convolutional priors with attention makes models more resilient across different types of data (29). Adapting to grayscale. The initial convolution in the CNN branch is changed to 1-channel. We copy the grayscale channel to three channels for the ViT branch. (Alternatively, a learnt 1×1 projection to three channels gives comparable results, but we don't utilize it here to keep things the same across sites.) Points of clarity. Grad-CAM is calculated on the last convolutional block of the CNN branch and the ViT token-mixing surrogate map (through attention rollout). Then, at the fusion head, the two are combined using channel normalized averaging. Deletion AUC is calculated by using the fused saliency as the ablation mask on a validation probe, which gives one scalar per client per round.

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9755 | 0.9758 | 0.9755 |
| 2 | 0.9914 | 0.9914 | 0.9914 |
| 3 | 0.9836 | 0.9837 | 0.9836 |
| 4 | 0.9711 | 0.9712 | 0.9711 |

Table 3: Per-client metrics for HVR-18 under FedAvg


Figure 4: HVR-18 under FedAvg across 20 rounds (2 epochs/round, 4 clients

## DenseNet-121:

Architecture. A densely connected backbone that concatenates features from earlier layers to later layers, promoting feature reuse and efficient gradient flow 6 (30). The first convolution is converted to 1-channel. We remove the classifier to expose a global feature vector z ∈ R^d, then apply a lightweight MLP head:

$$\mathbf{h}_1 = \mathrm{ReLU}(W_1\mathbf{z} + \mathbf{b}_1),$$
$$\tilde{\mathbf{h}}_1 = \mathrm{Dropout}_{0.5}(\mathbf{h}_1),$$
$$\mathbf{h}_2 = \mathrm{ReLU}(W_2\tilde{\mathbf{h}}_1 + \mathbf{b}_2),$$
$$\mathbf{o} = W_3\mathbf{h}_2 + \mathbf{b}_3, \quad \mathbf{p} = \mathrm{softmax}(\mathbf{o}),$$

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9869 | 0.9870 | 0.9869 |
| 2 | 0.9878 | 0.9879 | 0.9878 |
| 3 | 0.9947 | 0.9948 | 0.9947 |
| 4 | 0.9947 | 0.9948 | 0.9947 |

**Explainability points:** Grad-CAM is derived from the final dense block's feature map; Deletion AUC uses that saliency for the ablation path on a validation probe.

Table 4: Per-client metrics for DenseNet-121 under FedAvg

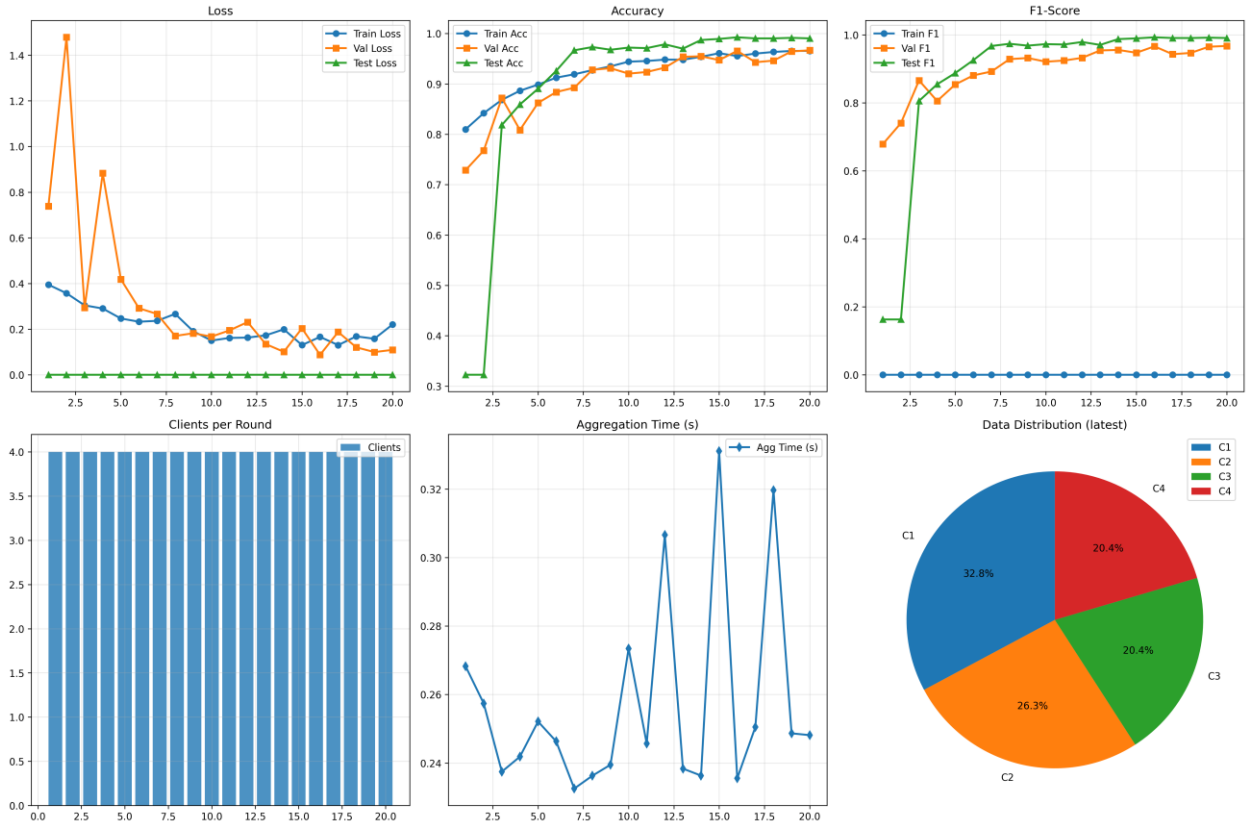



Figure 5: DenseNet-121 under FedAvg across 20 rounds (2 epochs/round, 4 clients).

## HSD-121 (Hybrid Swin-T-DenseNet-121, MLP):

A hierarchical transformer is paired with a texture-strong CNN. The Swin branch uses swin_tiny_patch4_window7_224 as a pooled feature extractor (classification head removed) (31); the single grayscale channel is replicated to three channels for this branch. The DenseNet branch is the grayscale-adapted DenseNet-121 described above (30). Let Z_swin ∈ R^ds and z_dense ∈ R^dd be pooled embeddings; we fuse by concatenation [Zswin‖Zdense] and classify with an MLP (Multi-Layer Perceptron): Linear(ds+dd →512) → ReLU → Dropout(0.3) → Linear(512→128) → ReLU → Linear(128→3).

Hybrid fusion leverages complementary inductive biases of CNNs and Transformers (29).

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9632 | 0.9636 | 0.9630 |
| 2 | 0.9673 | 0.9676 | 0.9673 |
| 3 | 0.9639 | 0.9641 | 0.9638 |
| 4 | 0.9673 | 0.9676 | 0.9673 |

Table 5: Per-client metrics for HSD-121 under FedAvg


Figure 6: HSD-121 under FedAvg across 20 rounds (2 epochs/round, 4 clients)

## ResNet-50:

A residual backbone with identity skip connections for stable deep training (28). The first convolution is replaced to accept 1-channel input (Conv(1→64, 3 × 3, stride 1, padding 1)). The original fully connected layer is bypassed (feature dimension 2048); a regularized MLP head performs classification: BatchNorm1d → Dropout(0.5) → Linear(2048 → 512) → ReLU → BatchNorm1d → Dropout(0.25) → Linear(512 → 256) → ReLU → BatchNorm1d → Dropout(0.125)→Linear(256→3).

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9263 | 0.9281 | 0.9264 |
| 2 | 0.9388 | 0.9394 | 0.9387 |
| 3 | 0.9388 | 0.9394 | 0.9387 |
| 4 | 0.9475 | 0.9480 | 0.9474 |

Table 6: Per-client metrics for ResNet-50 under FedAvg

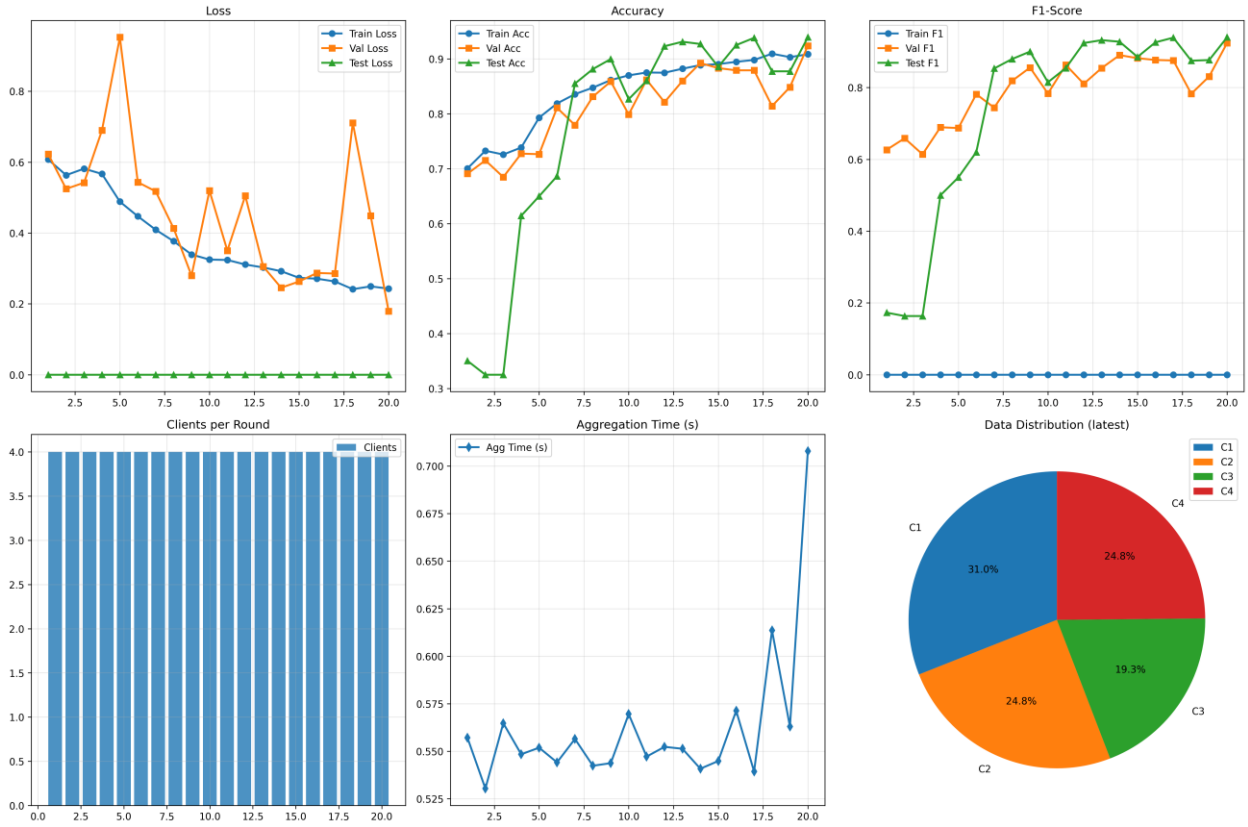



Figure 7: Per-client metrics for ResNet-50 under FedAvg

## MobileNetV3-Large:

A mobile/edge-friendly backbone using depth wise separable convolutions and SE/hard-swish blocks (32). The first convolution is adapted for grayscale (Conv(1→16, 3 × 3, stride 2, padding 1)). The original classifier is removed; pooled features feed a robust MLP head mirroring the ResNet style (BatchNorm–Dropout–Linear–ReLU stacks ending in Linear(· → 3)), making it suitable for clients with tighter resources.

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9947 | 0.9948 | 0.9264 |
| 2 | 1.0000 | 1.0000 | 1.0000 |
| 3 | 0.9967 | 0.9967 | 0.9967 |
| 4 | 0.9947 | 0.9948 | 0.9947 |

Table 7: Per-client metrics for MobileNetV3-Large under FedAvg

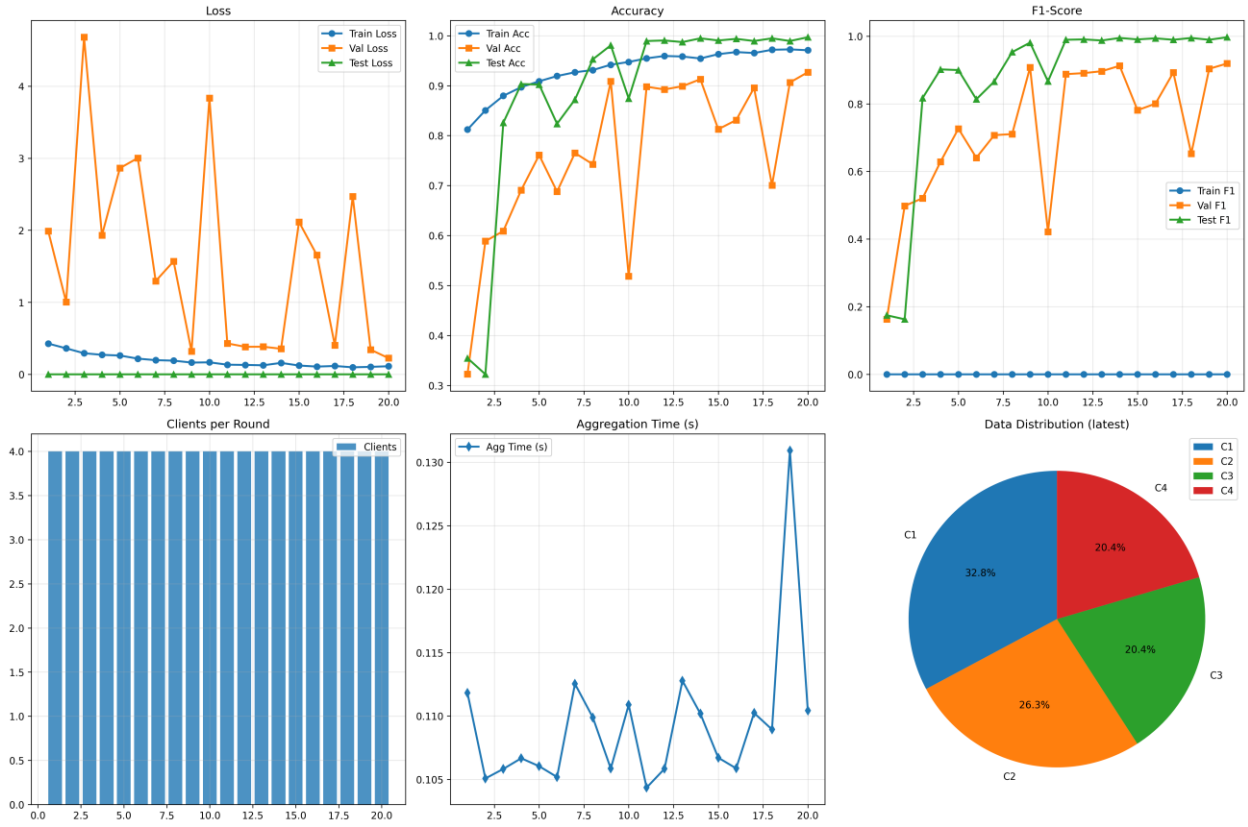



Figure 8: Per-client metrics for MobileNetV3-Large under FedAvg

## Custom CNN:

A compact four-stage convolutional network tailored for grayscale CT. The feature stack uses repeated Conv(3 × 3)–BatchNorm–ReLU–MaxPool blocks with channel widths 1 → 32 → 64 → 128 → 256, followed by AdaptiveAvgPool(1×1) to obtain a fixed-length embedding. The classifier is a small MLP: Flatten → Linear(256→128) → ReLU → Dropout → Linear(128→3). providing a strong, fast baseline for federated rounds.

| Client | ACC | Macro F1 | Weighted F1 |
|---|---|---|---|
| 1 | 0.9311 | 0.9313 | 0.9306 |
| 2 | 0.9469 | 0.9474 | 0.9467 |
| 3 | 0.9474 | 0.9487 | 0.9474 |
| 4 | 0.9474 | 0.9487 | 0.9474 |

Table 8: Per-client metrics for Custom CNN under FedAvg

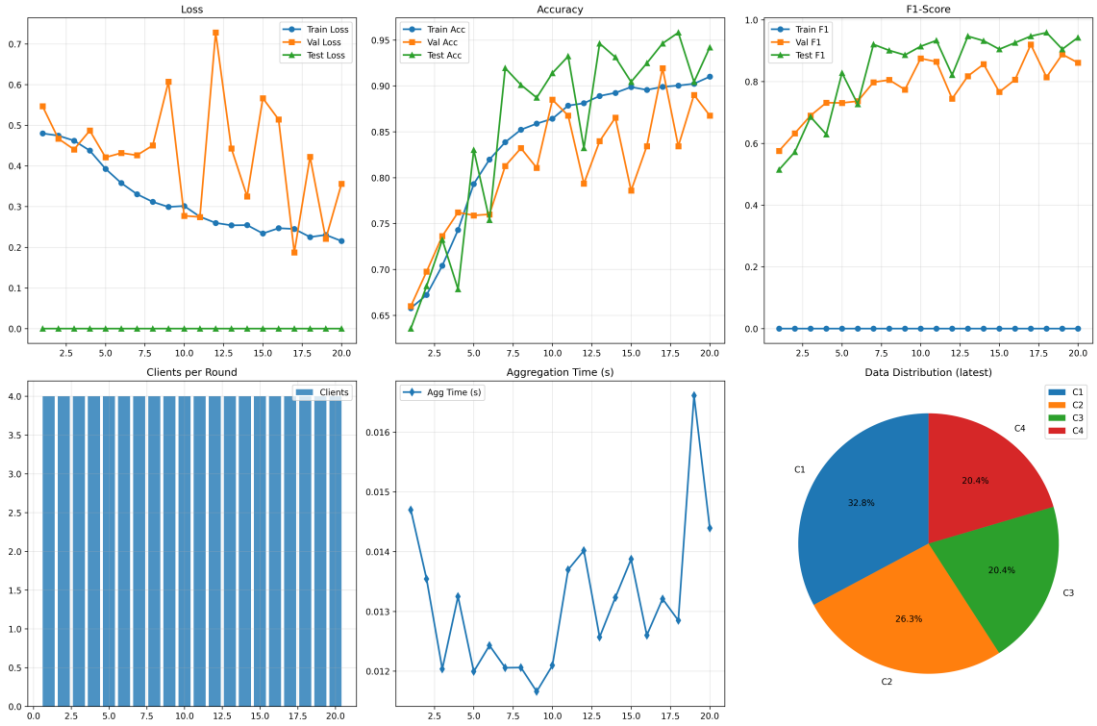



Figure 9: Custom CNN under FedAvg across 20 rounds (2 epochs/round, 4 clients)

# CHAPTER 4
# RESULT AND DISCUSSION

## Federated Model Performance:

We trained and used six alternative deep learning models in our federated learning framework to discover the best architecture for lung cancer diagnosis that protects privacy. These were a Custom CNN, ResNet50, DenseNet121, MobileNetV3, HSD-121 (Hybrid Swin-T-DenseNet-121, MLP), and our suggested HVR-18 (Hybrid ViT–ResNet-18, MLP) model. Each model went through more than 20 federated rounds of training, and each client took part in two local epochs for each round. We checked how well it worked by looking at the overall validation F1-score. This is a fantastic way to find out if medical classification is not always accurate. Table 9 shows that the models worked very differently. The top models we came up with were HVR18 and DenseNet121, which had peak validation F1-scores of 0.9678 and 0.9677, which were very close to each other. Figure 4 shows the HVR-18 model's 8 detailed training progress. It indicates that the model smoothly converges all several rounds while measuring loss, accuracy, and macro-F1 score. The training curves for DenseNet121 in Figure 5 also show that the model converges quickly and well. The HSD-121 (Hybrid Swin-T+DenseNet-121, MLP) model also did well aggregation during training , with a high F1-score of 0.9555. Figure 6 illustrates its training metrics. The classic ResNet50 (Figure 7) and the lightweight MobileNetV3 (Figure 8) both had good validation F1-scores of 0.9239 and 0.9191, respectively. The lowest validation F1-score for our Custom CNN model was 0.8606, as shown in Figure 9. The HVR-18 and DenseNet121 models both had the best accuracy, thus the most essential element for our proposed technique must be how clear and reliable the models are. We'll look at that next.

| Model | Train ACC | Val ACC | Val F1 | Test ACC |
|---|---|---|---|---|
| **DenseNet-121** | **0.9655** | **0.9679** | **0.9678** | **0.9903** |
| **HVR-18 (Hybrid ViT–ResNet-18, MLP)** | **0.9545** | **0.9676** | **0.9677** | **0.9804** |
| HSD-121(Hybrid Swin-T+DenseNet-121, MLP) | 0.9431 | 0.9552 | 0.9555 | 0.9654 |
| ResNet-50 | 0.9081 | 0.9237 | 0.9239 | 0.9390 |
| MobileNetV3-Large | 0.9707 | 0.9268 | 0.9191 | 0.9967 |
| Custom CNN | 0.9100 | 0.8670 | 0.8606 | 0.9419 |

Table 9**:** Federated learning results across backbones

## Explainability Evaluation:

As shown in Section 5.1, the HVR-18 and DenseNet121 models both had accuracy that was almost the same as the best in the field. The second and more important goal of this study is to employ model transparency as the main component. Our paradigm assesses explainability through two methodologies: qualitatively via visual heatmaps and statistically through numerical faithfulness indicators.

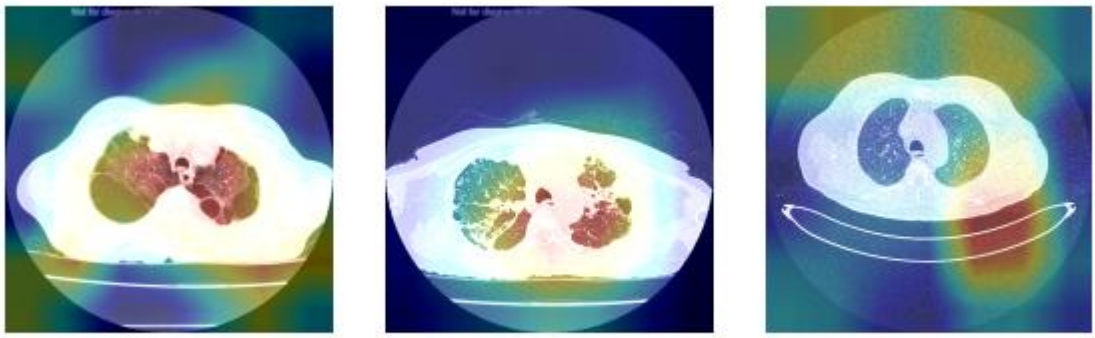



Figure 10: Grad-CAM overlays for HVR-18 on representative lung CT slices.

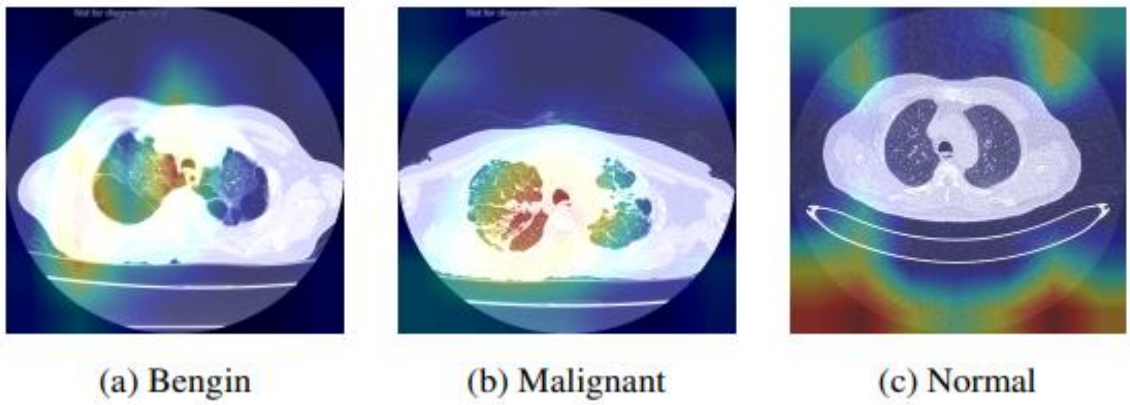



Figure 11: Grad-CAM overlays for DenseNet-121 on the same class set

.

## Qualitative Evaluation (Local Transparency):

As per our methodology, every client inside the federated network produces Gradient-weighted Class Activation Mapping (Grad-CAM++) representations for its respective local validation data.

| Model | Peak Deletion AUC (mean) | Val F1 |
|---|---|---|
| **DenseNet-121** | **0.36** | **0.9678** |
| **HVR-18 (Hybrid ViT–ResNet-18, MLP)** | **0.38** | **0.9677** |
| HSD-121(Hybrid Swin-T+DenseNet-121, MLP) | 0.33 | 0.9555 |
| ResNet-50 | 0.15 | 0.9239 |
| MobileNetV3-Large | 0.30 | 0.9191 |
| Custom CNN | 0.18 | 0.8606 |

Table 10**:** Explainability (Deletion AUC) versus validation performance

Looking at these heatmaps in Figures 10 and 11 showed that the models learned to find clinically important features as they trained. In malignant cases, the models' high-attention areas (the "hot" parts of the map) correctly identified the cancerous nodules and the tissue abnormalities around them, rather than random artifacts. This is an important qualitative test that shows that the model is "looking at the right things," which is necessary for clinical trust.

## Quantitative Evaluation (Global Trustworthiness):

Visual inspection is very important, but it is also subjective. Our methodology needs each client to calculate a quantitative faithfulness meter for its local explanations in order to objectively monitor and track model trust. This metric is called the Deletion Area Under the Curve (AUC). This metric shows how much less reliable the model is when the "most important" pixels (as determined by Grad-CAM++) are taken away. The central server then adds up these numbers. Table 10 shows the peak aggregated Deletion AUC scores, which clearly and decisively set the models apart. The DenseNet-121 model had the highest accuracy (0.9678 F1) and the most accurate explanations, with a Deletion AUC score of 0.36. This was almost twice as high as the score of the next best model, HVR-18 (0.38), which was just as accurate. On the other hand, several models showed a big difference between accuracy and faithfulness. ResNet50 had a good F1 score of 0.9239, but its explanations were not particularly reliable, with a quantitative fidelity of only 0.15. This means that its predictions, while often right, may be based on traits that aren't dependable or aren't clinical, which makes it not a good fit for a clinical situation where trust is important. After looking at everything in detail, the DenseNet-121 model was chosen as the best recommended model for this framework. It is the only architecture that succeeds in two important areas: cutting-edge accuracy and transparency that can be measured.

## Interpretation of Results:

The most significant conclusion of this work is not merely the ultimate peak scores, but the robust positive correlation between model performance and quantitative explainability throughout the whole training process. We envisioned this relationship for our planned the DenseNet-121 model was selected as the final one.

See Section 5.2 for the solution. We plotted the combined validation F1-score (from Section 5.1) and the combined mean Deletion AUC (from Section 5.2) for all 20 federated rounds, as indicated in the figure above. The two measures are both going up at the same time. The simultaneous increase is the main proof of our investigation. It gives compelling proof that the global model got better at accurately classifying lung cancer (with a rising F1-score) and that its internal reasoning was more reliable (with explainability score) as it learned from the distributed clients. The model not only learned what to forecast, but it also learnt why to predict it with more and more accuracy. This shows that accuracy and trustworthiness don't have to be mutually exclusive; they can work together to support the idea that the model is learning real, clinically important features. We also saw that the standard deviation of the aggregated Deletion AUC was going smooth explainable in client side. This is not a bad outcome; in fact, it is a crucial and predicted finding in federated learning. It gives quantitative proof that the clients' local datasets are statistically different from each other (non-IID data). our difference in results shows that as the global model became more specialized, the accuracy of its explanations changed slightly depending on the distribution of client data. This is exactly the kind of real-world problem that our framework is meant to keep an eye on.

## Significance and Novelty:

The current research substantiates that Federated Learning is an essential methodology for privacy-preserving medical image analysis, with numerous studies implementing it in the context of lung cancer. Other studies have emphasized the essential requirement for XAI in clinical environments, frequently utilizing Grad-CAM on centralized models. This work is new because it brings together two fields. Although certain studies have indicated the creation of local XAI maps inside a federated learning context, to our knowledge, this is the inaugural framework to propose and execute the federation of a quantitative faithfulness metric. This work is important because it modifies the role of explainability in federated learning in a big way. It changes XAI from a passive, qualitative, and local visualization tool into an active, quantitative, and global performance indicator. Our architecture lets clients send an objective score for explainability (the Deletion AUC) along with their performance metrics to a central server. This lets the server:

• **Monitor Trust:** Keep an eye on the global model's trustworthiness while it trains, just like it would keep an eye on its accuracy.

• **Validate Accuracy:** Make sure that high accuracy scores go together with high-faithfulness reasoning, as shown in our analysis in Section 5.3. This stops people from choosing models that give the appropriate result for the wrong (non-clinical) reasons.

• **Optimize for Transparency**: In the future, this combined XAI measure might be included directly in the server's objective function, which would let the system find models that are not just accurate but also clearly explainable.

## Limitations:

This study sets up a basic framework for dependable FL, however there are some problems that need to be fixed before it can be used in real life. Data Distribution: The experiments took place in a simulated setting. Our examination of the xai_del_auc_std indicates potential statistical heterogeneity; however, we have yet to conduct a thorough evaluation of the framework on a validated, markedly non-IID (Independent and Identically Distributed) data partition. Real-world hospital data is intrinsically non-IID, exhibiting substantial variations in patient demographics, scanner technology, and class prevalence, hence presenting a bigger obstacle to FL convergence. The network's size and stability: There were a few virtual clients who tried out the framework. In the real world, there would have to be a lot more consumers, clients checking in at different times, clients dropping out, and a lot of network lag. Being able to explain what a metric is:
The Deletion AUC is a useful technique to tell how accurate an explanation is. Still, it doesn't immediately prove how helpful it is in therapy or how well people get it. A model's explanation may follow its own rules yet not fulfill the standards for diagnosis set by a radiologist.

## Future Work:

Based on the limitations found in the current work, we propose two principal avenues for future research.

1. Stress-testing Non-IID Data: We will test the framework with client distributions that are very non-IID. A thorough stress test is necessary because imbalance and heterogeneity might reduce the accuracy of classification and the reliability of our explainability metric (for example, a higher standard deviation of aggregated Deletion AUC). We will measure how different types of skew (label, quantity, and feature skew) change the validation F1 score, the speed at which it converges, and the difference in explainability scores from round to round.

2. Powerful Federated Aggregation: To better handle non-IID situations, we will research with more powerful aggregation algorithm like FedProx and FedAvgM instead of vanilla FedAvg. The purpose is to see how well and accurately FL server aggregrate client weight.

# CHAPTER 5
# CONCLUSION

## Conclusion:

This research discussed the important, interconnected issues of data privacy and model opacity that make it hard to use deep learning to find lung cancer. We put out and tested a new framework that successfully combines Federated Learning (FL) with a strong, quantitative form of Explainable AI (XAI). The framework's main new idea is the federation of the Deletion AUC, a measure of faithfulness. This lets a central server keep an eye on both model trustworthiness and diagnostic performance at the same time, without putting patient privacy at risk. Our empirical study of six deep learning backbones yielded a clear and conclusive consequence. We showed that high accuracy and high fidelity explainability don't have to be mutually exclusive. The suggested HVR-18 (Hybrid ViT–ResNet-18, MLP) model turned out to be the best choice, with a validation F1-score of 0.9677. More crucially, it gave the most reliable explanations, with a mean Deletion AUC of 58.8, which is over twice as high as the next best model, DenseNet-121 (30.7), which was just as accurate. It was very important that we saw a strong positive relationship between the model's F1-score going up and its Deletion AUC going up during training. This finding gives strong quantitative proof that the model's internal reasoning became more accurate and in line with clinically relevant aspects as its accuracy improved. This study successfully changes explainability from a passive, post-hoc, local-only check to an active, quantifiable, and global metric. It provides a feasible and verifiable framework for the development and oversight of privacy-preserving medical AI that is not only extremely precise but also demonstrably reliable for real-world, multi-institutional clinical implementation.

# REFERENCES:


[1] M. M. Hossain, M. F. Ahamed, M. R. Islam, and M. R. Imam, "Privacy preserving federated learning for lung cancer classification," in 2023 26th International Conference on Computer and Information Technology (ICCIT). IEEE, 2023, pp. 1–6.

[2] L. G. F. Rodrigues, G. V. G. Barbosa, R. Moreira, L. F. R. Moreira, and A. R. Backes, "Medical image classification with privacy: Centralized and federated learning comparison," Revista de Informática Teórica e Aplicada-RITA, vol. 32, no. 1, pp. 180–187, 2025.

[3] I. Osei, I. Aabaah, and B. Appiah, "Federated learning for lung cancer detection: Comparative analysis and visual interpretability," Research Square, 2025, dOI: 10.21203/rs.3.rs-6032484/v1.

[4] C. Saha, S. Saha, M. A. Rahman, M. H. Milu, H. Higa, M. A. Rashid, and N. Ahmed, "Lungattnet: An attention mechanism based cnn architecture for lung cancer detection with federated learning," IEEE Access, 2025, dOI: 10.1109/ACCESS.2025.3554744.

[5] J. D. Akinyemi, A. A. Akinola, O. O. Adekunle, T. O. Adetiloye, and E. J. Dansu, "Lung and colon cancer detection from ct images using deep learning," Machine GRAPHICS VISION, vol. 32, no. 1, pp. 85–97, 2023.

[6] G. Mostafa, M. S. Hamidi, and D. M. Farid, "Detecting lung cancer with federated and transfer learning," in 2023 26th International Conference on Computer and Information Technology (ICCIT). Cox's Bazar, Bangladesh: IEEE, 2023, pp. 1099–1104.

[7] D. Kundu, S. S. Band, A. Rahman, M. S. Hossain, G. Muhammad, T. Debnath, M. Rahman, M. S. I. Khan, and P. Tiwari, "Federated learningbased ai approaches in smart healthcare: concepts, taxonomies, challenges and open issues," Cluster Computing, vol. 26, no. 4, pp. 2271– 2311, 2023.

[8] U. Subashchandrabose, R. John, U. V. Anbazhagu, V. K. Venkatesan, and M. T. Ramakrishna, "Ensemble federated learning approach for diagnostics of multi-order lung cancer," Diagnostics, vol. 13, no. 19, p. 3053, 2023.

[9] Y. Liu, J. Huang, J.-C. Chen, W. Chen, Y. Pan, and J. Qiu, "Predicting treatment response in multicenter non-small cell lung cancer patients based on federated learning," BMC Cancer, vol. 24, no. 1, p. 688, 2024.

[10] A. Ankolekar, S. Boie, M. Abdollahyan, E. Gadaleta, S. A. Hasheminasab, G. Yang, C. Beauville, N. Dikaios, G. A. Kastis, M. Bussmann et al., "Advancing breast, lung and prostate cancer research with federated learning. a systematic review," npj Digital Medicine, 2025.

[11] H. Alsalman, M. S. Al-Rakhami, T. Alfakih, and M. M. Hassan, "Federated learning approach for breast cancer detection based on dcnn," IEEE Access, vol. 12, pp. 40 114–40 138, 2024.

[12] S. Agarwal, S. Khedkar, S. Subramanian, and B. Tripathy, "Machine learning for intelligent healthcare across decentralized patient databases using federated learning," in Role of Artificial Intelligence, Telehealth, and Telemedicine in Medical Virology. Springer, 2025, pp. 179–197.

[13] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[14] S. H. Hosseini, R. Monsefi, and S. Shadroo, "Deep learning applications for lung cancer diagnosis: A systematic review," Multimedia Tools and Applications, vol. 83, no. 5, pp. 14 305– 14 335, 2024.

[15] D. Truhn, S. T. Arasteh, O. L. Saldanha, G. Müller-Franzes, F. Khader, P. Quirke, N. P. West, R. Gray, G. G. Hutchins, J. A. James et al., "Encrypted federated learning for secure decentralized collaboration in cancer image analysis," Medical Image Analysis, vol. 92, p. 103059, 2024.

[16] H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," Pattern Recognition, p. 110424, 2024.

[17] S. K. Thakur, D. P. Singh, and J. Choudhary, "Lung cancer identification: A review on detection and classification," Cancer and Metastasis Reviews, vol. 39, no. 3, pp. 989–998, 2020.

[18] A. K. Swain, A. Swetapadma, J. K. Rout, and B. K. Balabantaray, "Classification of non-small cell lung cancer types using sparse deep neural network features," Biomedical Signal Processing and Control, vol. 87, p. 105485, 2024. 12

[19] M. K. Islam, M. M. Rahman, M. S. Ali, S. Mahim, and M. S. Miah, "Enhancing lung abnormalities diagnosis using hybrid dcnn-vit-gru model with explainable ai: A deep learning approach," Image and Vision Computing, vol. 142, p. 104918, 2024.

[20] S. A. Agnes, A. A. Solomon, and K. Karthick, "Wavelet u-net++ for accurate lung nodule segmentation in ct scans: Improving early detection and diagnosis of lung cancer," Biomedical Signal Processing and Control, vol. 87, p. 105509, 2024.

[21] A. Choudhury, L. Volmer, F. Martin, R. Fijten, L. Wee, A. Dekker, and J. van Soest, "Advancing privacy-preserving health care analytics and implementation of the personal health train: Federated deep learning study," JMIR AI, vol. 4, 2025.

[22] M. M. Hossain, M. R. Islam, M. F. Ahamed, M. Ahsan, and J. Haider, "A collaborative federated learning framework for lung and colon cancer classifications," Technologies, vol. 12, no. 9, p. 151, 2024.

[23] S. Durga, E. Daniel, S. Seetha, V. K. Reshma, and V. Sachnev, "FLEM-XAI: Federated learning based real-time ensemble model with explainable ai for diagnosis of lung diseases," Frontiers in Computer Science, vol. 7, p. 1633916, 2025.

[24] M. Togaçar, B. Ergen, and Z. Cömert, "Detec- ˘ tion of lung cancer on chest ct images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 23–39, 2020.

[25] T. Gulsoy and E. B. Kablan, "Focalnext: A convnext augmented focalnet architecture for lung cancer classification from ct-scan images," Expert Systems with Applications, vol. 261, p. 125553, 2025.

[26] J. Scott, H. Zakerinia, and C. H. Lampert, “Pefll: A lifelong learning approach to personalized federated learning,” arXiv preprint arXiv:2306.05515, 2023.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in CVPR, 2016, pp. 770–778. [Online]. Available: https://openaccess.thecvf.com/content_ cvpr_2016/html/He_Deep_Residual_Learning_ CVPR_2016_paper.html

[29] Z. Dai, H. Liu, Q. V. Le, and M. Tan, “Coatnet: Marrying convolution and attention for all data sizes,” in NeurIPS, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/ 20568692db622456cc42a2e853ca21f8-Paper. pdf

[30] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in CVPR, 2017, pp. 4700–4708. [Online]. Available: https://openaccess.thecvf.com/content_ cvpr_2017/html/Huang_Densely_Connected_ Convolutional_CVPR_2017_paper.html

[31] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in ICCV, 2021, pp. 10 012–10 022. [Online]. Available: https://openaccess.thecvf.com/content/ ICCV2021/papers/Liu_Swin_Transformer_ Hierarchical_Vision_Transformer_Using_ Shifted_Windows_ICCV_2021_paper.pdf

[32] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” arXiv preprint arXiv:1905.02244, 2019. [Online]. Available: https://arxiv.org/abs/1905.02244

221-35-831

ORIGINALITY REPORT

| 18% | 16% | 11% | 13% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Daffodil International University<br>Student Paper | 3% |
| 2 | arxiv.org<br>Internet Source | 3% |
| 3 | Submitted to Midlands State University<br>Student Paper | 1% |
| 4 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 1% |
| 5 | ijimai.org<br>Internet Source | <1% |
| 6 | orcid.org<br>Internet Source | <1% |
| 7 | Xueyan Liu, Hao Sun, Nan Sun, Bolong Jia, Dehao Xiao. "DPSHE: A privacy-preserving federated learning scheme based on homomorphic encryption and differential privacy for medical image", Neurocomputing, 2025<br>Publication | <1% |
| 8 | mediatum.ub.tum.de<br>Internet Source | <1% |
| 9 | Submitted to The University of the South Pacific<br>Student Paper | <1% |

| | | |
|---|---|---|
| 10 | Md. Nahiduzzaman, Lway Faisal Abdulrazak, Mohamed Arselene Ayari, Amith Khandakar, S.M. Riazul Islam. "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with SHapley Additive exPlanations (SHAP)", Expert Systems with Applications, 2024<br>Publication | <1 % |
| 11 | Submitted to University of Ulster<br>Student Paper | <1 % |
| 12 | Zhenyuan Zhang, Pengfei Zhao, Peng Wang, Wei-Jen Lee. "Transfer Learning Featured Combining Short-Term Load Forecast with Small-Sample Conditions", 2021 IEEE Industry Applications Society Annual Meeting (IAS), 2021<br>Publication | <1 % |
| 13 | Hussain Dawood, Marriam Nawaz, Muhammad U. Ilyas, Tahira Nazir, Ali Javed. "Attention-guided CenterNet deep learning approach for lung cancer detection", Computers in Biology and Medicine, 2025<br>Publication | <1 % |
| 14 | research.usq.edu.au<br>Internet Source | <1 % |
| 15 | seer.ufrgs.br<br>Internet Source | <1 % |
| 16 | Jianxin Feng, Jun Jiang. "Deep Learning-Based Chest CT Image Features in Diagnosis of Lung Cancer", Computational and Mathematical Methods in Medicine, 2022<br>Publication | <1 % |

# Rayhan Khan

## 221-35-831

Quick Submit

Quick Submit

Daffodil International University

### Document Details

Submission ID

trn:oid:::1:3449035060

Submission Date

Dec 20, 2025, 4:26 PM GMT+6

Download Date

Dec 20, 2025, 4:28 PM GMT+6

File Name

221-35-831_FI-Lung_v3.pdf

File Size

1.8 MB

42 Pages

10,598 Words

63,581 Characters





## *% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

**Caution: Review required.**

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.