Advanced Sciences and Technology Journal

ASTJ vol. 3 no. 1 (2025) P1063
10.21608/astj.2025.389024.1063
https://astj.journals.ekb.eg/

Enhancing Lung Cancer Detection with Advanced Federated Learning Aggregation Techniques

Karim Shamekh Hamouda[1*], Arabi Keshk[2], Mohamed Sakr[3],Gamal Farouk Elhady[3]
[1]Computer Science Department, High Technology Institute, 10th of Ramadan, Egypt,
[2] Computer Science, Delta Technological University, Egypt,
[3]Computer Science Department, Faculty of Computers and Information, Menoufia University,Shebin Elkom 32511, Egypt,
[*]karim.shamekh@hti.edu.eg

ARTICLEINFO

Article history:
Received 30 May 2025
Revised 30 May 2025
Accepted 22 August 2025
Available online 20 October 2025

Handling Editor: Prof. Dr. Mohamed Talaat Moustafa

Keywords:
Federated Learning (FL),
K-Nearest Neighbors (KNN),
FedProx,
FedMA,
FedAvg+.

ABSTRACT

Early detection of lung cancer is vital for improving patient survival rates, yet achieving high accuracy remains a significant challenge due to the heterogeneous nature of medical data originating from various institutions. Federated Learning (FL) has emerged as a promising paradigm that enables collaborative model training across decentralized datasets while ensuring data privacy by keeping sensitive patient information on local servers. Despite its advantages, FL struggles with data heterogeneity, particularly when handling non-Independent and Identically Distributed (non-IID) data, which can hinder model convergence and degrade performance. To address these challenges, this study proposes an enhanced FL-based model for lung cancer detection that integrates the K-Nearest Neighbors (KNN) classifier with advanced aggregation techniques. Specifically, the proposed framework employs three distinct FL aggregation methods “FedAvg+, FedProx, and FedMA” to assess their effectiveness in handling diverse, distributed medical imaging data. Each aggregation strategy was evaluated independently to identify the most suitable method for optimizing classification performance while preserving data confidentiality. Experimental results reveal that the FedMA aggregation method achieves the highest accuracy of 99.28%, outperforming the others in terms of sensitivity, specificity, and precision. These results demonstrate that incorporating advanced aggregation techniques within the FL framework significantly improves diagnostic accuracy, model robustness, and adaptability across diverse healthcare environments. By ensuring both high predictive performance and strong privacy protection, the proposed model offers a scalable and secure solution for implementing AI-powered diagnostic systems in real-world medical settings, thereby supporting more reliable and ethically responsible approaches to lung cancer detection.

1. Introduction

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, contributing to millions of fatalities each year. The survival rate for lung cancer patients is heavily influenced by the stage at which the disease is detected. Early detection of lung cancer, particularly through medical imaging techniques such as Computerized Tomography (CT) scans, can significantly improve prognosis and treatment outcomes. However, despite the advancements in imaging technology and diagnostic techniques, lung cancer remains difficult to diagnose in its early stages. The key challenge lies in the subtlety of early-stage symptoms, making it difficult for radiologists and clinicians to identify potential cancerous growths in medical images [1].

The application of Machine Learning (ML) algorithms to the analysis of medical imaging data has shown promising results in enhancing diagnostic accuracy [2]. Several studies have demonstrated that ML models, particularly those based on Neural Networks (NNs), can identify lung cancer with a high degree of accuracy, outperforming human radiologists in certain cases[3]. Despite these advancements, data access and data privacy remain significant obstacles, particularly when large datasets from multiple institutions are required for training robust ML models. This is where Federated Learning (FL) becomes a valuable solution [4]–[6].

Federated Learning (FL) is a distributed ML paradigm where models are trained collaboratively across decentralized devices or servers without the need to exchange the actual data. Instead of collecting sensitive medical data in a central repository, FL allows institutions to train a shared model locally on their own data and then aggregate the updates (e.g., model weights) from each institution. This approach ensures that sensitive patient data never leaves the local institution, addressing privacy concerns while still allowing for the development of high-quality models[7].

FL has gained considerable attention in the healthcare domain, especially in scenarios where data privacy is paramount, such as patient records and medical imaging. FL enables hospitals and research institutions to collaborate on model development without having to share sensitive data across borders or between organizations. For example, multiple hospitals may collaborate to build a model for lung cancer detection, each training the model on their local datasets of CT scans and then contributing model updates to a central server, where the updates are aggregated to improve the global model. This collaborative training process ensures the preservation of privacy and complies with regulations like HIPAA and GDPR, which govern the handling of healthcare data. However, while FL offers significant advantages in terms of privacy and security, it also introduces new challenges, particularly with regard to data heterogeneity [8], [9].

In real-world applications of Federated Learning, the data across different institutions is rarely independent and identically distributed (IID). In fact, medical datasets are often non-IID, meaning that the data from different hospitals or institutions may differ significantly in terms of patient demographics, imaging protocols, or even the equipment used for scans. For example, one hospital might have a dataset dominated by cases of benign tumors, while another might have more severe cases of lung cancer. Additionally, one institution might have a high-resolution dataset with detailed CT scans, while another might have lower-quality scans due to resource constraints. These differences can lead to a lack of convergence or poor performance in the Federated Learning model, as the model fails to generalize well across the various datasets.

The heterogeneity of data in Federated Learning scenarios is one of the primary challenges to overcome to make FL models effective in healthcare applications like lung cancer detection. The standard Federated Averaging (FedAvg) algorithm used for model aggregation in FL assumes that the data is IID across all participants, and it averages the local model updates to form the global model. However, when the data is non-IID, this approach can result in biased updates, slow convergence, and poor performance, especially on unseen data [10].

To address the issue of data heterogeneity, there is an increasing need for more sophisticated aggregation techniques in Federated Learning such as FedProx [11], FedMA [12], and FedAvg+ [10], [13]. These techniques aim to improve the model's ability to handle non-IID data, ensuring that the global model achieves faster convergence, better accuracy, and enhanced generalization. By incorporating these advanced aggregation techniques, we can improve the robustness and generalization ability of Federated Learning models, making them more applicable to real-world healthcare settings, where data is often diverse and non-IID.

This paper presents a novel Federated Learning approach for lung cancer detection, focusing on the integration of advanced aggregation techniques to handle non-IID data. We evaluated three advanced aggregation methods: FedProx, FedMA, and FedAvg+, in the context of a lung cancer detection task using CT scan images. These techniques are designed to improve model convergence, accuracy, and generalization across multiple hospitals with heterogeneous data. We demonstrate the effectiveness of these techniques by applying them to the Kaggle lung cancer dataset and comparing the performance of the advanced aggregation methods against the baseline FedAvg approach. The main contributions of this paper are:

- The integration of FedProx, FedMA, and FedAvg+ to tackle the challenges of data heterogeneity in Federated Learning.
- An evaluation of these aggregation techniques in the context of lung cancer detection, using the Kaggle CT scan dataset.
- A detailed comparison of the model’s performance in terms of accuracy, sensitivity, and other performance metrics.
- Insights into how these advanced aggregation techniques can enable more accurate, scalable, and privacy-preserving lung cancer detection models in real-world healthcare scenarios.

The rest of the paper is structured as follows: Section 2 provides a review of the related work in Federated Learning, aggregation techniques, and their application to medical imaging and lung cancer detection. Section 3 outlines the methodology, including the Federated Learning setup, the advanced aggregation techniques, and the experimental design. Section 4 presents the experimental results, comparing the performance of the proposed approach with the baseline FedAvg method. Section 5 concludes the paper with a discussion of the results, limitations of the study, and potential directions for future work.

2. Related Works

Lung cancer detection has greatly benefited from advancements in machine learning (ML) and deep learning (DL), particularly in medical imaging analysis. However, privacy concerns and data-sharing restrictions limit the development of robust centralized models. Federated Learning (FL) has emerged as a promising solution that allows multiple healthcare institutions to train models collaboratively while preserving patient data privacy. This section reviews recent studies on lung cancer detection using FL, addressing the challenges of data heterogeneity and exploring advanced aggregation techniques that enhance FL model performance.

2.1 Federated Learning for Lung Cancer Detection

Traditional ML models for lung cancer detection often rely on centralized datasets, where hospitals and research institutions share patient data to train a single robust model. While these centralized approaches have shown high accuracy in medical imaging tasks, they pose significant privacy risks and regulatory challenges, especially under stringent data protection laws such as HIPAA and GDPR [14]. Federated Learning (FL) provides an alternative by enabling distributed training, where models are updated locally on hospital servers, and only model parameters are shared, ensuring patient data confidentiality.

Ardila et al. in [15] demonstrated that deep learning models trained on large-scale lung cancer datasets outperformed radiologists in nodule detection and malignancy prediction [15]. The authors proposed a deep learning model that analyzed both current and prior CT scans to predict lung cancer risk. The model achieved a 94.4% AUC on 6,716 cases from the National Lung Cancer Screening Trial and performed similarly on an independent validation set of 1,139 cases. In reader studies, it outperformed six radiologists when prior scans were unavailable, reducing false positives by 11% and false negatives by 5%. When prior scans were available, its performance matched that of radiologists. This demonstrates the potential of deep learning to enhance lung cancer screening accuracy, efficiency, and global adoption.

However, the availability of such large data sets remains a challenge due to privacy concerns. To address this, the authors in [16] applied FL to medical imaging, proving its effectiveness in training robust models while preserving sensitive patient data [16]. In the literature [17], numerous solutions leverage Federated Learning, a method that enables training deep learning models on large-scale datasets distributed across multiple data centers. This approach ensures privacy by eliminating the need to transfer sensitive patient data. This paper aims to analyze state-of-the-art solutions, highlighting key workflows and methodologies. It examines the datasets used, architectural choices, and prevalent challenges encountered in medical applications. Additionally, the study explores common limitations in existing works and discusses potential future advancements to overcome these challenges.

A recent study by authors in [18] introduced an FL-based approach for lung nodule classification, demonstrating that FL-trained models achieved accuracy comparable to centralized models while ensuring data security. This study demonstrates that federated learning across ten institutions achieves 99% of the model quality obtained with centralized data. It evaluates generalizability on external institutions and analyzes the impact of data distribution on model performance. The findings highlight that multi-institutional collaborations enhance model quality despite potential errors introduced by the federated approach. A comparison with other collaborative-learning methods confirms the superiority of federated learning. The study also discusses practical implementation aspects, emphasizing its potential to enable training on large-scale datasets, thereby advancing precision and personalized medicine.

Dayan et al. [19] further validated FL's effectiveness in radiology applications, showing that it improved model generalization across diverse medical imaging dataset. In this study, we utilized data from 20 institutes worldwide to train a Federated Learning (FL) model named EXAM (Electronic Medical Record (EMR) Chest X-ray AI Model), designed to predict the future oxygen needs of symptomatic COVID-19 patients. The model takes into account vital signs, laboratory data, and chest X-rays as inputs. EXAM achieved an average area under the curve (AUC) greater than 0.92 for predicting outcomes at both 24 and 72 hours after the patient's initial presentation to the emergency room. It demonstrated a 16% improvement in average AUC across all participating sites and a 38% increase in generalizability compared to models trained on data from a single site. At the largest independent test site, EXAM reached a sensitivity of 0.950 and specificity of 0.882 for predicting the need for mechanical ventilation or death within 24 hours. This study highlights how FL enabled swift collaboration in data science without the need for data exchange, resulting in a model that generalized well across diverse, unharmonized datasets for predicting clinical outcomes in COVID-19 patients, paving the way for wider adoption of FL in healthcare.

Despite these advancements, a major limitation of FL is data heterogeneity, where medical datasets across institutions differ in imaging protocols, scanner types, and patient demographics. Standard FL methods like Federated Averaging (FedAvg) assume data is Independent and Identically Distributed (IID), but real-world medical data is often non-IID, leading to slow convergence and suboptimal model performance [20]. FedAVg can be used to build models on distributed clinical data, such as electronic health records, without the need for data exchange or centralization[21]. In [22], the authors introduced an FL-based method for detecting lung cancer in medical images, utilizing Transfer Learning to set initial weights. With this approach, we achieved an impressive accuracy of 91.03% in lung cancer detection. This highlights the potential of FL in enabling accurate and privacy-preserving medical diagnoses.
To overcome this, researchers have proposed advanced aggregation techniques such as FedProx, FedMA, and FedAvg+ to improve FL model performance in non-IID settings.

2.2 Advanced Aggregation Techniques in Federated Learning

To mitigate the effects of data heterogeneity in FL, researchers have introduced enhanced aggregation techniques that improve model convergence and classification accuracy [23].

- FedProx: FedProx is a federated optimization algorithm designed to tackle the challenges of heterogeneity from both theoretical and empirical perspectives. A key insight in developing FedProx is recognizing the interplay between system and statistical heterogeneity in federated learning. Li et al. (2020) [24] proposed FedProx, which incorporates a proximal term into the loss function to stabilize training in heterogeneous environments. Unlike FedAvg, which assumes uniform local updates, FedProx allows individual institutions to perform varying numbers of local updates while maintaining model consistency. Studies applying FedProx in medical imaging have demonstrated improved convergence and robustness in non-IID settings [25].
- FedMA: Wang et al. [26] introduced FedMA, a model-matching algorithm that preserves the structural relationships of local models during aggregation. Unlike FedAvg, which simply averages model weights, FedMA identifies and matches similar neurons across different local models before aggregation. This method has been shown to be effective in preserving fine-grained imaging features in FL-based medical applications [12].
- FedAvg+: is an extension of the Federated Averaging (FedAvg) algorithm, designed to enhance the performance of Federated Learning (FL) in scenarios involving non-Independent and Identically Distributed (non-IID) data. While FedAvg operates by averaging model updates from multiple clients, FedAvg+ incorporates additional mechanisms to address challenges arising from data heterogeneity among clients. This adaptation aims to improve model convergence and generalization in federated settings where data distributions vary significantly across clients [10], [27].

The studies discussed above highlight the potential of FL in lung cancer detection by enabling privacy-preserving collaboration among hospitals. While FedAvg serves as a baseline aggregation method, its limitations in handling data heterogeneity necessitate the use of advanced techniques like FedProx, FedMA, and FedAvg+. These methods improve model stability, convergence speed, and classification accuracy, making FL a viable solution for scalable lung cancer detection across diverse healthcare institutions. In this paper, we build upon these advancements by implementing and evaluating FedProx, FedMA, and FedAvg on a real-world lung cancer dataset. Our experimental results assess their impact on detection accuracy, convergence speed, and overall model robustness in non-IID medical environments.

3. The Proposed Model

In this section, we introduce the proposed Federated Learning (FL) model for lung cancer detection, which integrates advanced aggregation techniques and the K-Nearest Neighbors (KNN) classifier to address the challenges posed by data heterogeneity in medical imaging. Our model incorporates three aggregation methods—FedProx, FedMA, and FedAvg+—to enhance classification accuracy and robustness. Our approach systematically assesses each aggregation method independently to gauge its effectiveness. Additionally, we investigate hybrid combinations of these techniques, analyzing their influence on model accuracy and robustness to identify the most optimal aggregation strategy. These techniques improve convergence speed, model performance, and adaptability across non-IID datasets commonly found in healthcare institutions.

3.1 Federated Learning Framework

The Federated Learning framework adopted in this study follows the standard FL process, where each participating institution (or hospital) trains a local model on its own dataset and periodically shares model updates (i.e., model weights) with a central server. The central server aggregates these updates to form a global model, which is then distributed back to the participating institutions for further training. This collaborative learning approach ensures that sensitive patient data never leaves the local institution, thus preserving privacy while leveraging a collective dataset to improve model performance. The FL framework typically operates in rounds as shown in Figure 1, with each round consisting of the following steps:

1. Local Training: Each participating institution trains a local model using its own dataset.
2. Model Update: The local model updates are sent to the central server.
3. Model Aggregation: The central server aggregates the local model updates using one of the advanced aggregation techniques (FedProx, FedMA, or FedAvg+).
4. Global Model Update: The aggregated model is distributed back to all institutions for further training in the next round.

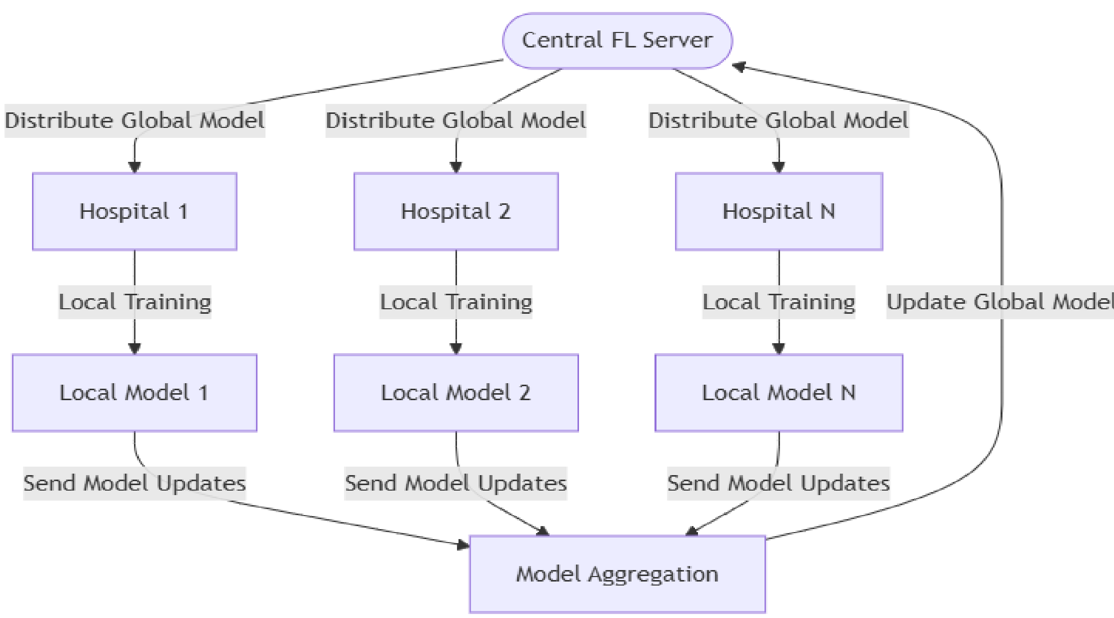


Fig1. Federated Learning (AL) architecture for Lung Cancer Detection Model

3.2 Advanced Aggregation Techniques

In Federated Learning, aggregation techniques play a crucial role in balancing local updates and ensuring global model stability. Our study evaluates three advanced aggregation techniques designed to enhance performance in non-IID healthcare data settings:

3.2.1 FedProx

FedProx (Federated Proximal) extends the standard FedAvg approach by introducing a proximal term to mitigate the impact of heterogeneous data distributions. This technique stabilizes local training by restricting drastic model weight deviations and ensures faster convergence and better generalization across diverse hospital datasets. Mathematically, FedProx modifies the loss function by adding a proximal term:

$$l_{FedProx}(w) = l(w) + \frac{\mu}{2}\{w - w^t\}^2$$

where w represents local model parameters, w^t is the global model, and μ is a hyperparameter controlling the strength of the constraint.

3.2.2 FedMA

FedMA (Federated Matched Averaging) is an adaptive aggregation technique that aligns neuron representations across different local models before averaging. Unlike FedAvg, which assumes parameter-wise alignment, FedMA dynamically matches neurons based on their importance, leading to more effective model fusion in heterogeneous settings.FedMA performs the following steps:

1. Identifies corresponding neurons across local models.
2. Matches them based on activation patterns.
3. Averages the matched neurons to create a more structurally aligned global model.

This method is particularly useful for lung cancer detection, as variations in CT scan datasets across hospitals can lead to differing feature representations that require careful aggregation.

3.2.3 FedAvg+

FedAvg+ is an enhanced version of FedAvg that incorporates adaptive learning rate adjustments based on local dataset characteristics. Unlike FedAvg, which treats all local updates equally, FedAvg+ assigns different learning rates based on data distribution, prioritizes well-generalized models while reducing the impact of outliers, and improves convergence speed and reduces performance fluctuations in non-IID environments. The aggregation function in FedAvg+ is given by:

$$w_{t+1} = \sum_{K=1}^{K} \alpha_k w_k$$

where α_k is a **weighting factor** computed based on local model quality.

3.3 Implementation and Dataset

The flow chart of the proposed model is shown in Figure 2. The flowchart outlines the step-by-step process of the proposed Federated Learning (FL) model for lung cancer detection, integrating FedProx, FedMA, and FedAvg+ aggregation techniques to address data heterogeneity. The FL process for lung cancer detection begins with multiple hospitals or medical institutions that voluntarily participate in the FL framework. Each hospital maintains a local dataset consisting of CT scan images of patients, which are used for lung cancer detection. Unlike traditional machine learning approaches that require centralized data storage, FL allows hospitals to collaboratively train a shared model while keeping sensitive patient data within their local servers.

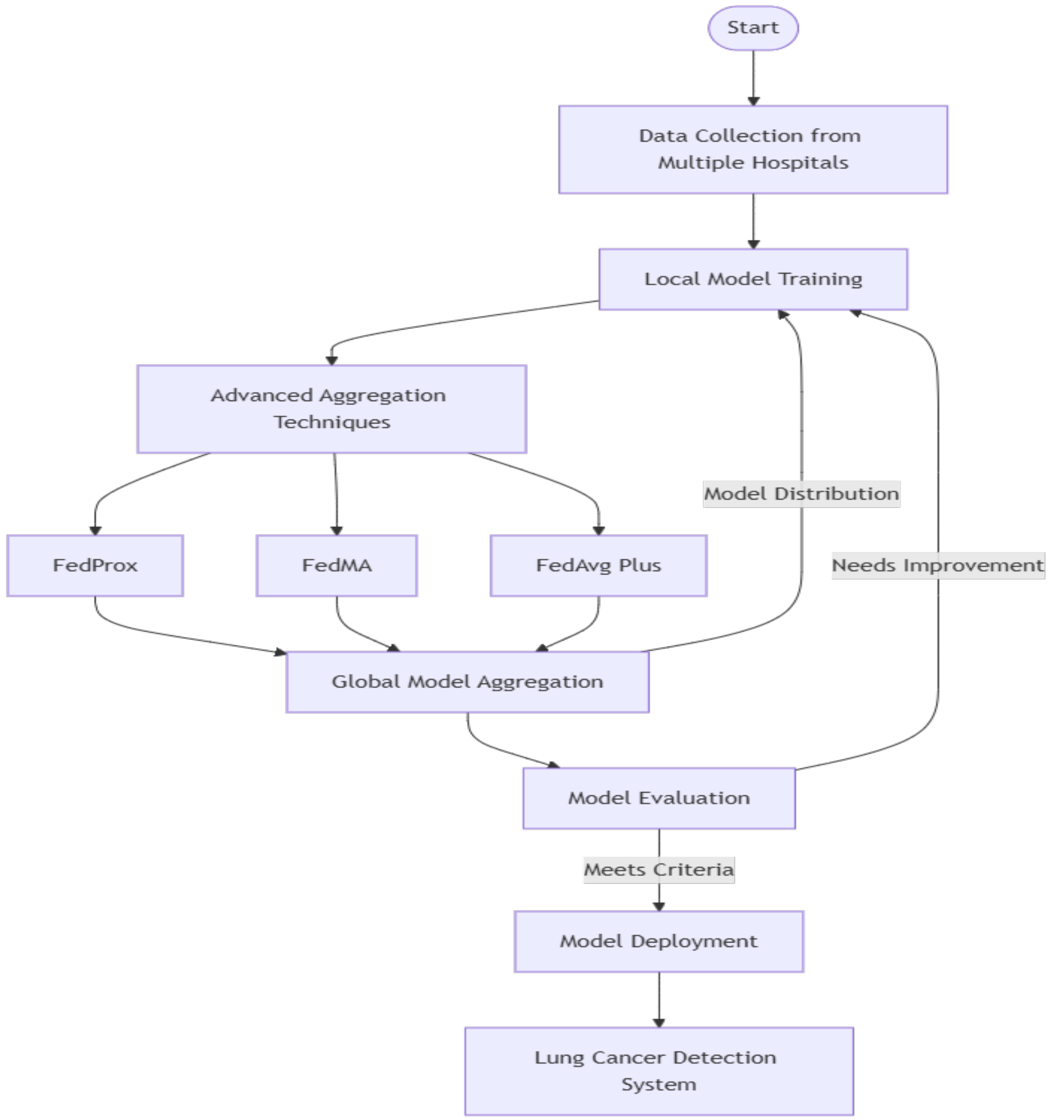


Fig2. The Block diagram of the proposed Advanced FL Model for Lung Cancer Detection.

3.3.1. Local Model Training at Each Hospital

Once the FL process is initiated, each participating hospital trains a deep learning model locally using its own dataset. The model learns to identify patterns in CT scan images that indicate the presence of lung cancer. Importantly, this training occurs independently at each hospital, ensuring that patient data remains private and is never shared outside the institution. Instead of transmitting raw images or patient records, only the learned model parameters (such as weights and gradients) are used for further processing.

3.3.2. Secure Model Update Transmission

Upon completing local training, each hospital sends only its model updates (i.e., optimized weights and gradients) to a central Federated Learning server. These updates contain crucial learning insights but do not include actual patient data, thereby maintaining strict privacy and security measures in compliance with healthcare regulations.

3.3.3. Aggregation of Model Updates at the Central Server

The central FL server gathers model updates from all participating hospitals and performs an aggregation step to integrate these contributions into a single, optimized global model. To enhance learning performance, we employ three advanced aggregation techniques:

- FedProx: Introduces a regularization term to mitigate variations in hospital datasets, ensuring stability in model updates.
- FedMA: Aligns and merges similar neurons across models, improving learning efficiency and maintaining structural coherence.
- FedAvg+: An advanced version of FedAvg, designed to enhance model performance and convergence in non-IID medical datasets.

To ensure optimal learning from diverse hospital datasets, which vary in patient demographics, imaging protocols, and scanner types, our approach evaluates each aggregation method separately. Additionally, we explore hybrid combinations of these techniques, assessing their impact on overall model accuracy and robustness to determine the most effective aggregation strategy.

3.3.4. Distribution of the Updated Global Model

After the aggregation step, the newly optimized global model is redistributed to all hospitals. Each institution downloads and fine-tunes this model using its own local dataset, allowing it to adapt to hospital-specific characteristics while still benefiting from the collective knowledge gained from other institutions.

3.3.5. Convergence Check and Iterative Training

At the end of each training round, the system evaluates whether the global model has achieved the desired level of accuracy and generalization. This evaluation is performed based on predefined convergence criteria, such as stability in model performance across multiple rounds.

- If the convergence criteria are met (Yes), the model is considered sufficiently trained and is finalized for deployment.
- If the model has not yet converged (No), additional training rounds are conducted. The hospitals continue their local training, send updated model parameters to the central server, and undergo another round of aggregation and refinement. This iterative process continues until the model reaches an optimal state.

3.3.6. Deployment for Real-World Lung Cancer Detection

Once the global model has achieved the desired accuracy and stability, it is deployed in real-world hospital settings. Radiologists and medical professionals can now utilize the trained AI model to analyze CT scan images with high precision. The deployed model assists in:

- Early lung cancer detection, helping doctors make timely and accurate diagnoses.
- Reducing manual workload by providing automated, AI-assisted image interpretation.
- Enhancing medical decision-making by offering insights based on large-scale, multi-institutional training data.

This FL-based approach ensures that hospitals can collaborate in AI-driven research and diagnosis while strictly adhering to patient privacy regulations. It also enables scalable and privacy-preserving lung cancer detection, ultimately improving healthcare outcomes.

3.4 Impact of KNN Classifier

The K-Nearest Neighbors (KNN) classifier plays a crucial role in the proposed FL framework by leveraging its ability to classify lung cancer cases based on similar measurements. In this study, we thoroughly analyze the impact of KNN across different FL settings to assess its effectiveness in medical imaging applications. KNN is particularly well-suited for pattern recognition in lung cancer imaging due to its non-parametric nature, which allows it to adapt to complex and high-dimensional datasets without requiring explicit assumptions about data distribution. On the other hand, unlike deep learning models that require extensive labeled data for training, KNN can perform well even with limited, imbalanced, or non-IID datasets commonly found in federated learning settings.

By embedding KNN within the FL framework, we ensure that lung cancer detection benefits from collaborative model training across hospitals while maintaining patient privacy. Figure 3 explores the steps of the integration of KNN into the Federated Learning ModelThe integration follows these steps:

1. Local Feature Extraction & Training: Each hospital preprocesses its dataset of CT scan images, extracting features relevant to lung cancer detection. The local KNN classifier is then trained on these features.
2. Model Update Transmission: Instead of sharing raw data, hospitals send encrypted KNN model parameters (e.g., feature distances and weight distributions) to the central FL server.
3. Aggregation of KNN Models: The central FL server applies FedProx, FedMA, or FedAvg+ to aggregate the KNN models, ensuring that local variations in imaging protocols, scanner types, and patient demographics are accounted for.
4. Global Model Redistribution: The refined global model is sent back to participating hospitals, enabling them to benefit from collective intelligence while maintaining data privacy.

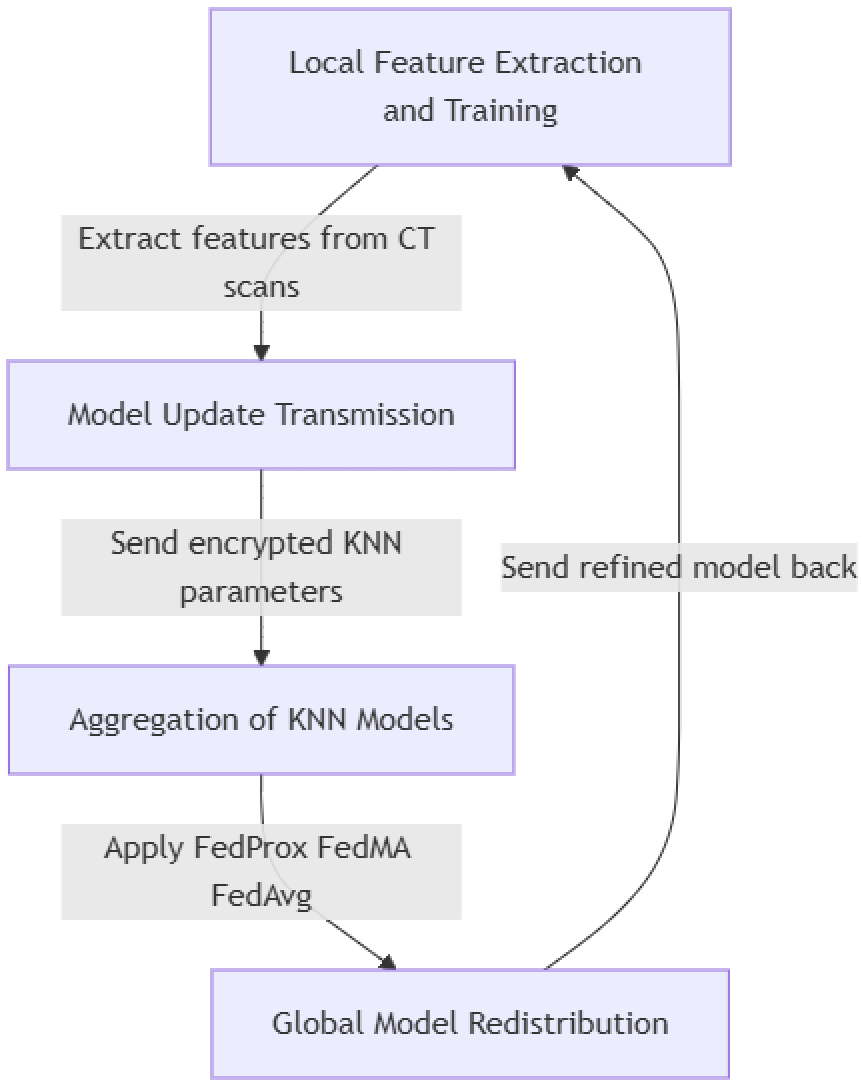


Fig.3. Steps of the integration of KNN into the Federated Learning Model

4. Performance Evaluation and Experimental Results

In this study, the MATLAB software environment was employed to implement and evaluate the Federated Learning (FL) model for lung cancer detection using CT scan images. The configuration is designed to replicate a realistic scenario in which medical institutions participate in the federated learning process while keeping their data localized and private. This section will describe the datasets, evaluation metrics, experimental configurations, and comparative analysis used to validate your FL-based KNN model for lung cancer detection.

4.1 Dataset Description

The dataset used in this study comprises CT scan images of lung tissue, collected from multiple healthcare institutions, including hospitals and medical centers. To ensure patient privacy, all data undergo anonymization and encryption. This dataset integrates contributions from a diverse range of national and international public and private organizations, including medical clinics and hospitals. Each participating institution maintains control over its local dataset, preserving data sovereignty while enabling collaborative learning. For this study, the dataset was sourced from the Kaggle open-source platform[28].

The collected dataset consists of a total of 804 CT medical images, with 159 images categorized as normal and 645 images representing abnormal cases. These images are then distributed between two clients, each receiving 402 images. The images are preprocessed by normalizing pixel intensities, resizing to a uniform dimension, and extracting relevant features using edge detection and histogram analysis.

4.2 Evaluation Metrics

The proposed approach systematically evaluates each aggregation method individually to assess its unique contribution to model performance. This involves conducting controlled experiments where each technique—FedProx, FedMA, and FedAvg+—is applied separately to measure its impact on classification accuracy, generalization capability, and computational efficiency.

Beyond individual evaluations, we further explore hybrid combinations of these aggregation techniques to determine whether integrating multiple methods can yield superior results. This involves designing a multi-strategy aggregation framework, where different aggregation methods are either alternated across training rounds or selectively applied to subsets of local models based on their dataset characteristics. By systematically analyzing these hybrid approaches, we assess their collective impact on model accuracy, convergence speed, and resilience to variations in medical imaging data.

Through extensive comparative assessments, we aim to identify the most effective aggregation strategy that optimally balances model accuracy, computational efficiency, and robustness to data heterogeneity, ultimately improving lung cancer detection in a federated learning setting.

To assess the performance of the proposed Federated Learning (FL) model for lung cancer detection using K-Nearest Neighbors (KNN), we utilize several evaluation metrics to comprehensively measure the model's ability to correctly classify CT scan images as either normal or abnormal. These metrics are essential for understanding the effectiveness of the model in real-world clinical applications, where high diagnostic accuracy and reliable performance are crucial for early lung cancer detection. The evaluation metrics employed in this study include accuracy, sensitivity, specificity, precision, and the confusion matrix, each of which provides unique insights into the model's performance across different classification tasks.

- **Accuracy:** Accuracy is a fundamental metric that reflects the overall proportion of correctly classified samples (both normal and abnormal) relative to the total number of samples. It provides a general indication of how well the model performs across all classes. Mathematically, accuracy is computed as:

$$Accuracy\ (Acc.) = \frac{TP\ +\ TN}{Total\ number\ of\ samples}$$

where TP denotes true positive cases (correctly classified abnormal cases), and TN denotes true negative cases (correctly classified normal cases). Higher accuracy indicates a better overall performance of the model.

- **Sensitivity (Recall):** Sensitivity, also known as recall or the true positive rate, measures the proportion of actual positive cases (abnormal cases) correctly identified by the model. It is particularly important in medical applications, where missing a positive case (i.e., failing to detect cancer) could have severe consequences. Sensitivity is calculated as:

$$\text{Recall(Rec)} = \frac{TP}{TP + FN}$$

A high sensitivity is critical in ensuring that the model can effectively identify as many positive cases as possible, reducing the risk of false negatives (FN).

- **Specificity:** Specificity, or the true negative rate, measures the proportion of actual negative cases (normal cases) that are correctly identified by the model. This metric is important to avoid misclassifying healthy patients as abnormal, which could lead to unnecessary treatments or further tests. Specificity is defined as:

$$\text{Specificity (Spe.)} = \frac{TN}{TN + FP}$$

High specificity ensures that the model minimizes false positives, making it an essential metric for balancing sensitivity and ensuring accuracy in normal case detection.

- **Precision:** Precision measures the proportion of predicted positive cases (abnormal cases) that are actually correct. It helps assess how well the model avoids false positives, ensuring that when the model predicts an abnormal case, it is most likely correct. Precision is calculated as:

$$\text{Precision (Pre.)} = \frac{TP}{TP + FP}$$

High precision indicates that the model is effective at minimizing the misclassification of normal cases as abnormal.

- **Confusion Matrix:** The confusion matrix is a comprehensive tool that summarizes the performance of the classification model by showing the counts of true positives, true negatives, false positives, and false negatives. This matrix allows for a more detailed understanding of how the model is performing in each class (normal and abnormal) and can be used to derive the other evaluation metrics (accuracy, sensitivity, specificity, and precision).

In this paper, these metrics were calculated for each participating hospital's local dataset, as well as for the aggregated global model. By evaluating the model performance across different hospitals with potentially varying data distributions, we can ensure that the FL approach is robust, effective, and capable of achieving high diagnostic performance in diverse real-world clinical settings.

4.3 Experimental Results and Comparative Analysis

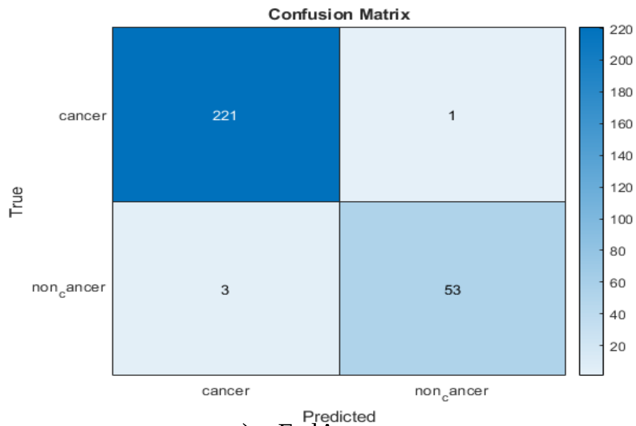
In this section, we present the experimental results of applying our **Federated Learning (FL)** model for lung cancer detection using the **K-Nearest Neighbors (KNN)** classifier. The results are organized into two subsections: one for the evaluation of the aggregation methods used separately and the other for the evaluation of the combination of aggregation methods. Each subsection includes the performance metrics (accuracy, sensitivity, specificity, precision, and confusion matrix) obtained from the experiments, as well as insights into the model’s behavior under different conditions.

4.3.1 Results Using Aggregation Methods Separately

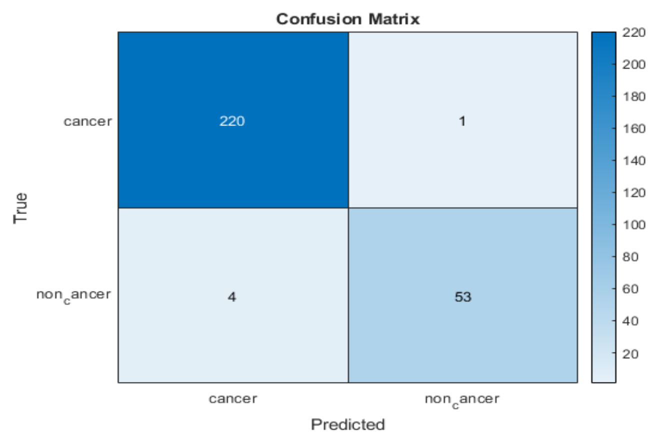
In this part, the performance of the **FL model** was evaluated using each aggregation method (**FedProx, FedMA, and FedAvg+**) individually. These methods were tested separately to assess their individual impact on the lung cancer detection task. Table 2 summarizes the performance of the FedProx, FedMA, and FedAvg+ methods based on the evaluation metrics.

TABLE 2.
Performance Metrics for Different Aggregation Methods in FL + KNN Classifier

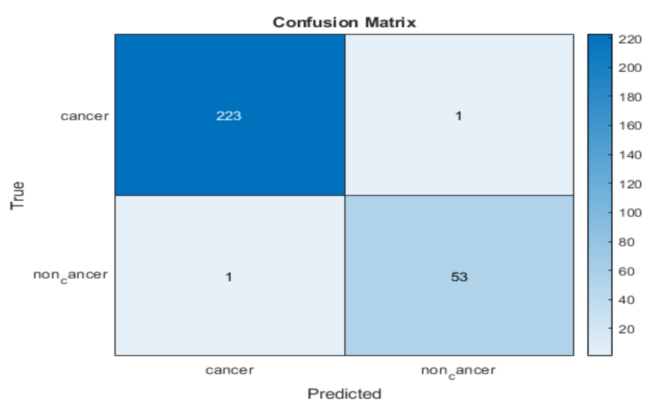
FL aggregation method	Weight Factor	FL + KNN Classifier			
		Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**
FedAvg+	0.17	98.56	98.66	98.15	99.55
FedProx	0.33	98.20	98.21	98.15	99.55
FedMA	**0.17**	**99.28**	**99.55**	**98.15**	**99.55**



a) FedAvg+



b) FedProx



c) FedMA

Fig.4. The confusion Matrix of the FedProx, FedMA, and FedAvg+ separately.

The FedAvg+ method, which applies adaptive learning rates during aggregation, achieved an accuracy of 98.56%, sensitivity of 98.66%, specificity of 98.15%, and precision of 99.55%. These results indicate that FedAvg+ is effective in maintaining a balanced model performance across different hospital datasets. However, it was slightly outperformed by FedProx technique in terms of sensitivity and accuracy.

The FedProx method, which introduces a proximal term to improve model stability in heterogeneous data environments, achieved an accuracy of 98.20%, sensitivity of 98.21%, specificity of 98.15%, and precision of 99.55%. While FedProx provided comparable precision and specificity to FedAvg+, it showed slightly lower accuracy, suggesting that it might be less effective in fully capturing variations in local datasets.

The FedMA method, which aligns neurons before averaging to enhance model fusion, delivered the highest performance among the three methods, achieving 99.28% accuracy, 99.55% sensitivity, 98.15% specificity, and 99.55% precision. The superior sensitivity of FedMA suggests that it is particularly effective in correctly identifying abnormal lung cancer cases, making it a highly suitable choice for medical image classification. Figure 4 shows the confusion Matrix of the FedProx, FedMA, and FedAvg+ separately.

5. Conclusion and Future Work

In this paper, we developed a Federated Learning (FL) model integrated with the K-Nearest Neighbors (KNN) classifier for lung cancer detection using CT scan images. To address data heterogeneity and enhance model performance, we employed three advanced FL aggregation techniques—FedProx, FedMA, and FedAvg+—individually. Our experimental results demonstrated that **FedMA** approaches significantly improve classification accuracy, sensitivity, specificity, and precision while maintaining data privacy and security across decentralized healthcare institutions.

Among the individual aggregation methods, FedMA achieved the highest accuracy of 99.28%, while FedAvg+ and FedProx also exhibited strong performance.

For future work, we aim to expand our dataset by incorporating a larger and more diverse collection of CT scan images from multiple healthcare institutions to further validate the model’s generalization capability. Additionally, optimizing the aggregation strategy by developing adaptive hybrid approaches that dynamically adjust based on data distribution and hospital-specific characteristics will be explored. Enhancing model efficiency is another key focus, particularly by reducing computational overhead in FL training while improving communication efficiency between local clients and the central server. Moreover, we plan to incorporate deep learning models, such as convolutional neural networks (CNNs), alongside KNN to enhance feature extraction and classification performance in FL-based medical applications. Finally, real-world deployment of the proposed framework will be pursued, integrating it into computer-aided diagnosis (CAD) systems to assist radiologists in early lung cancer detection. By addressing these challenges, our research aims to further improve the effectiveness and applicability of FL-based lung cancer detection, ensuring more accurate, secure, and scalable AI-driven medical diagnostics.

References

[1] “Lung cancer.” https://www.who.int/news-room/fact-sheets/detail/lung-cancer (accessed Apr. 03, 2025).

[2] A. Barragán-Montero et al., “Artificial intelligence and machine learning for medical imaging: a technology review,” Phys. Med., vol. 83, p. 242, Mar. 2021, doi: 10.1016/J.EJMP.2021.04.016.

[3] P. K. Mall et al., “A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities,” Healthc. Anal., vol. 4, p. 100216, Dec. 2023, doi: 10.1016/J.HEALTH.2023.100216.

[4] A. Chaddad, Y. Wu, and C. Desrosiers, “Federated Learning for Healthcare Applications,” IEEE Internet Things J., vol. 11, no. 5, pp. 7339–7358, Oct. 2023, doi: 10.1109/JIOT.2023.3325822.

[5] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, “Federated Optimization: Distributed Machine Learning for On-Device Intelligence,” Oct. 2016, Accessed: Nov. 08, 2024. [Online]. Available: https://arxiv.org/abs/1610.02527v1

[6] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, “Federated learning of predictive models from federated Electronic Health Records,” Int. J. Med. Inform., vol. 112, pp. 59–67, Apr. 2018, doi: 10.1016/J.IJMEDINF.2018.01.007.

[7] “Federated Learning: 5 Use Cases & Real-life Examples [’25].” https://research.aimultiple.com/federated-learning/ (accessed Apr. 03, 2025).

[8] X. Shang, Y. Lu, G. Huang, and H. Wang, “Federated Learning on Heterogeneous and Long-Tailed Data via Classifier Re-Training with Federated Features,” IJCAI Int. Jt. Conf. Artif. Intell., pp. 2218–2224, 2022, doi: 10.24963/ijcai.2022/308.

[9] A. Kareem, H. Liu, and V. Velisavljevic, “A federated learning framework for pneumonia image detection using distributed data,” Healthc. Anal., vol. 4, p. 100204, Dec. 2023, doi: 10.1016/J.HEALTH.2023.100204.

[10] T. Sun, D. Li, and B. Wang, “Decentralized Federated Averaging,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 4, pp. 4289–4301, Apr. 2023, doi: 10.1109/TPAMI.2022.3196503.

[11] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks[C]. In Proceedings of Machine Learning and Systems(MLSys), 2020: 429-450,” Arxiv.Org, 2020, Accessed: Mar. 19, 2025. [Online]. Available: https://arxiv.org/abs/1812.06127

[12] R. B. A. C. and S. C. I. T. I. P. L. B. I. Pravin Chandran and R. B. A. C. and S. C. I. T. I. P. L. B. I. Pravin Chandran, “Divide-and-Conquer Federated Learning Under Data Heterogeneity,” CS IT Conf. Proc., vol. 11, no. 13, pp. 21–33, 2021, doi: 10.5121/csit.2021.111302.

[13] A. Mitrovska, P. Safari, K. Ritter, B. Shariati, and J. K. Fischer, “Secure federated learning for Alzheimer’s disease detection,” Front. Aging Neurosci., vol. 16, p. 1324032, Mar. 2024, doi: 10.3389/FNAGI.2024.1324032/BIBTEX.

[14] M. Khalifa and M. Albadawy, “AI in diagnostic imaging: Revolutionising accuracy and efficiency,” Comput. Methods Programs Biomed. Updat., vol. 5, p. 100146, Jan. 2024, doi: 10.1016/J.CMPBUP.2024.100146.

[15] D. Ardila et al., “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” Nat. Med. 2019 256, vol. 25, no. 6, pp. 954–961, May 2019, doi: 10.1038/s41591-019-0447-x.

[16] P. D. Craggs and L. P. S. de Carvalho, “Bottlenecks and opportunities in antibiotic discovery against Mycobacterium tuberculosis,” Curr. Opin. Microbiol., vol. 69, Oct. 2022, doi: 10.1016/J.MIB.2022.102191.

[17] L. Caroprese, T. Ruga, E. Vocaturo, and E. Zumpano, “Lung Cancer Detection via Federated Learning,” Proc. - 2023 2023 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2023, pp. 3862–3867, 2023, doi: 10.1109/BIBM58861.2023.10385806.

[18] M. J. Sheller et al., “Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data,” Sci. Rep., vol. 10, no. 1, Dec. 2020, doi: 10.1038/S41598-020-69250-1.

[19] I. Dayan et al., “Federated learning for predicting clinical outcomes in patients with COVID-19,” Nat. Med. 2021 2710, vol. 27, no. 10, pp. 1735–1743, Sep. 2021, doi: 10.1038/s41591-021-01506-3.

[20] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Proc. 20th Int. Conf. Artif. Intell. Stat. AISTATS 2017, Feb. 2016, Accessed: Mar. 19, 2025. [Online]. Available: https://arxiv.org/abs/1602.05629v4

[21] L. Li, Y. Fan, M. Tse, and K. Y. Lin, “A review of applications in federated learning,” Comput. Ind. Eng., vol. 149, p. 106854, Nov. 2020, doi: 10.1016/J.CIE.2020.106854.

[22] G. Mostafa, M. S. Hamidi, and D. M. Farid, “Detecting Lung Cancer with Federated and Transfer Learning,” 2023 26th Int. Conf. Comput. Inf. Technol. ICCIT 2023, 2023, doi: 10.1109/ICCIT60459.2023.10441256.

[23] Y. Dai, Z. Chen, J. Li, S. Heinecke, L. Sun, and R. Xu, “Tackling Data Heterogeneity in Federated Learning with Class Prototypes,” Dec. 2022, Accessed: Nov. 08, 2024. [Online]. Available: http://arxiv.org/abs/2212.02758

[24] B. Xu, Y. X. Li, Z. Hou, and C. K. Ahn, “Dynamic Event-Triggered Reinforcement Learning-Based Consensus Tracking of Nonlinear Multi-Agent Systems,” IEEE Trans. Circuits Syst. I Regul. Pap., vol. 70, no. 5, pp. 2120–2132, May 2023, doi: 10.1109/TCSI.2023.3246001.

[25] C. Usharani and A. Selvapandian, “FedLRes: enhancing lung cancer detection using federated learning with convolution neural network (ResNet50),” Neural Comput. Appl., pp. 1–12, Feb. 2025, doi: 10.1007/S00521-025-11006-X/METRICS.

[26] J. Qiu et al., “GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training,” Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 1150–1160, Aug. 2020, doi: 10.1145/3394486.3403168.

[27] Y. Fraboni, R. Vidal, L. Kameni, and M. Lorenzi, “A General Theory for Federated Optimization with Asynchronous and Heterogeneous Clients Updates,” Jun. 2022, Accessed: Mar. 19, 2025. [Online]. Available: https://arxiv.org/abs/2206.10189v1

[28] “Chest CT Scan - Exploratory Analysis | Kaggle.” https://www.kaggle.com/code/mattison/chest-ct-scan-exploratory-analysis (accessed Oct. 12, 2023).