Enhancing Clinical Decision Support Systems with a Active Federated Learning Framework

B. Tech. Project Mid Sem Report

Submitted by

Shubham Manish Gandhi	**112003135**
Bakliwal Aagam	**112003013**

Under the guidance of

Prof. Trishna Ugale

College of Engineering, Pune

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COEP TECHNOLOGICAL UNIVERSITY, PUNE-5

March 2024

DEPARTMENT OF COMPUTER ENGINEERING

AND

INFORMATION TECHNOLOGY,

COLLEGE OF ENGINEERING, PUNE

CERTIFICATE

Certified that this project, titled "Enhancing Clinical Decision Support Systems with a Active Federated Learning Framework" has been successfully completed by

Shubham Manish Gandhi	**112003135**
Bakliwal Aagam	**112003013**

and is approved for the partial fulfillment of the requirements for the degree of "B.Tech. Computer Engineering".

SIGNATURE
Trishna Ugale
Project Guide
Department of Computer Engineering
and Information Technology,
COEP Technological University Pune,
Shivajinagar, Pune - 5.

SIGNATURE
Pradeep Deshmukh
Head
Department of Computer Engineering
and Information Technology,
COEP Technological University Pune,
Shivajinagar, Pune - 5.

Abstract

Federated Learning (FL) has rapidly evolved into a pivotal decentralized methodology for the development of machine learning models, especially in the healthcare sector. It offers a novel approach for enabling collaborative training across multiple data sources. This approach is particularly advantageous in contexts requiring stringent data privacy measures, such as in medical settings where patient confidentiality is paramount. By facilitating the aggregation of insights from diverse datasets without necessitating the direct exchange of sensitive information, FL aligns with current privacy regulations, ensuring the protection of patient data. This research introduces a sophisticated federated learning framework designed to enhance collaborative efforts among various medical institutions. The objective is to improve the diagnostic accuracy of lesion detection in Chest X-ray images by employing advanced deep learning techniques and leveraging the YOLOv9 architecture, all while circumventing the need to share patient-specific data. The study delves into several critical aspects of the federated learning environment, including the distribution of datasets across clients, the efficacy of different model averaging methods, and the impact of client dropout scenarios on the overall training process. Experimental results from this study underscore the viability of the proposed FL framework, demonstrating that it achieves competitive performance metrics comparable to traditional models trained on pooled data. This consistency in performance across different model architectures suggests that medical institutions could benefit from adopting a collaborative approach, leveraging the wealth of private data to quickly develop robust medical machine learning models and accelerate progress in medical research.

Contents

Abstract **2**

List of Tables **ii**

List of Figures **iii**

List of Symbols **iv**

1 Introduction **1**

1.1 Introduction 1

2 Literature Review **4**

3 Research Gaps and Problem Statement **14**

3.1 Research Gaps 14

3.2 Problem Statement 15

4 Proposed Methodology **16**

4.1 Lesion Identification on the Edge Device 16

4.1.1 YOLOv9 Architecture for Lesion Detection 19

4.1.2 Client-Side Model Update 20

4.1.3 Server-Side Model Aggregation 21

4.2 Experimental Setup 22

4.2.1 Dataset 22

4.2.2 Data Augmentation 23
4.2.3 Training of the Detection Models 25
4.2.4 Evaluation Metrics 25

5 Results and Discussion 28

5.1 Results and Discussion 28
5.1.1 Comparative Analysis of Federated and Centralized Training Approaches 28
5.1.2 Comparative Analysis of Federated Averaging and Weighted Federated Averaging 30
5.1.3 Results on IID Data with Varying Client Dropout . . . 32
5.1.4 Comparative Analysis of Performance on IID versus non-IID Datasets 34

A Algorithm for Client-Side Training 37

B Algorithms for Server-Side Aggregation 38

List of Tables

4.1 The specifics of the hyperparameters used for training the YOLOv9 models 26

5.1 Average mAPs 30

5.2 Average mAPs for Centralized and Federated Models 32

5.3 Average mAPs over varying client dropout on IID dataset . . 34

5.4 Average mAPs over varying client dropout on IID dataset . . 36

List of Figures

4.1 The proposed framework for the lesion detection system 17

4.2 Proposed Architecture of the Federated Learning Model 19

4.3 YOLOv9 detailed architecture [13] 20

4.4 Examples of CXRs with radiologist's annotations 24

4.5 Distribution of Findings and pathologies on the training set of the VinDr-CXR Dataset 24

4.6 Image Augmentations 25

5.1 Comparison of Federated Learning to Centralized Training using original and augmented dataset for learning 30

5.2 Comparison of Federated Averaging methods on IID and non-IID datasets 32

5.3 Effect of the client fraction C on the test accuracy of YOLOv9 models. Note $C = 1$ corresponds to all clients are selected at each round, $C = 0.5$ corresponds to half clients and $C = 0.25$ corresponds to only one client per round 34

5.4 Comparison of Federated Learning results on IID data and non-IID data partitions with C = 1 (all clients are considered at each round) 36

List of Symbols

Chapter 1

Introduction

1.1 Introduction

In the domain of healthcare, the accuracy and timeliness of disease diagnosis are critical determinants of the efficacy of subsequent treatments and, ultimately, patient outcomes. Radiological imaging, such as X-rays, constitutes an essential diagnostic tool, offering valuable insights into the patient's condition without invasive procedures. The detection and analysis of lesions from these images are pivotal for diagnosing a myriad of conditions, ranging from benign growths to malignancies indicative of cancer. Traditional diagnostic processes, heavily reliant on the expertise and experience of radiologists, are subject to variability in diagnostic accuracy and efficiency [3].

The advent of artificial intelligence (AI) and machine learning (ML) technologies heralds a new era in medical diagnostics, promising significant enhancements in the speed, accuracy, and consistency of lesion detection in radiological images [7]. Algorithms like YOLO (You Only Look Once), have shown remarkable success in object detection tasks by identifying and localizing objects with high precision. The application of such algorithms in healthcare are revolutionizing how medical imagery is analyzed, leading to more accurate diagnoses [6, 5, 14].

Despite these advancements, the deployment of AI in healthcare, particularly in sensitive areas such as radiology, faces significant hurdles. Centralized training of AI models on medical imagery raises profound concerns over privacy, data security, and the ethical use of patient data. The sensitivity and confidentiality of patient medical records are of utmost importance, governed by strict regulations such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States [12]. Moreover, the variability and heterogeneity of medical data across different healthcare institutions pose additional challenges in training AI models that are robust, generalizable, and capable of performing accurately across diverse populations.

Federated learning emerges as a novel solution to these challenges, which was introduced by Google [9]. This decentralized approach to machine learning allows for the collaborative training of AI models across multiple institutions, without the need to exchange or centralize sensitive data. By keeping the data localized and only sharing model updates or weights, federated learning ensures the privacy and security of patient data, aligning with regulatory requirements and ethical standards [11].

The objective of this research is to implement a federated learning framework for the identification of lesions in CXRs using the advanced capabilities of the latest YOLO model, YOLOv9. The proposed methodology innovatively combines the precision of AI in lesion detection with the expertise of medical practitioners, who validate and refine the AI-generated bounding boxes, ensuring the high accuracy and reliability of the diagnostic process. These local refinements are then integrated into the model through federated learning.

The main contributions of this research are threefold. First, we propose a decentralized and collaborative framework that enables clinicians to benefit from the advantages of rich, private data sharing, while simultaneously preserving the privacy of this data. Second, our research demonstrates that, despite the challenges posed by the decentralized nature of the data, such as its non-IID (independent and identically distributed) and unbalanced properties, the proposed federated learning framework remains robust and offers competitive results when compared to traditional centralized learning processes. Third, we explore the impact of client dropout rates on the overall performance of the model. Through extensive experimentation and comparison across various scenarios, we underscore the effectiveness and importance of our strategy, proving its particular utility in applications like lesion identification from CXRs.

The potential impact of this research extends beyond the immediate benefits of improved diagnostic accuracy and efficiency in lesion detection. By leveraging federated learning, this work paves the way for a scalable, privacy-preserving, and ethically responsible framework for the deployment of AI technologies in healthcare. It addresses critical concerns regarding patient privacy and data security, showcasing a viable pathway for the integration of advanced AI diagnostics in a manner that respects the confidentiality and sensitivity of patient data. Furthermore, this research contributes to the broader field of machine learning by demonstrating the practical application and benefits of federated learning in a high-stakes, real-world context.

Chapter 2

Literature Review

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
1	Vu-Thu-Nguyet Pham & Quang-Chung Nguyenand & Quang-Vu Nguyen (2022).). "Chest X-Rays Abnormalities Localization and Classification Using an Ensemble Framework of Deep Convolutional Neural Networks". Vietnam Journal of Computer Science	• Utilizes YOLOv5 in conjunction with CSPDarknet as the backbone, addressing gradient issues and reducing the model size. PANet refines the architecture with FPN and adaptive feature pooling, enabling multi-scale predictions and disease detection through a binary classifier. • User-uploaded images are transformed into 640x640 JPEGs using YOLOv5 and ResNet50. Classification depends on confidence levels, utilizing regular detection results if confidence exceeds the threshold. YOLOv5 detection output is employed when confidence falls below the threshold. • WBF technique is implemented for label gathering, constructing a common bounding box. Tasks are separated for images with and without findings using a classifier CNN, achieving a max F1-score of 0.76 at a threshold of 0.203.	• The adoption of a Convolutional Neural Network (CNN) as a classifier for disease detection in images introduces computational challenges due to the model's resource-intensive nature. The heavy reliance on this CNN model demands a substantial allocation of computational resources, posing potential limitations in terms of scalability and operational efficiency. • The conversion of DICOM images to heavily compressed JPEG format results in significant information loss, adversely impacting both data quality and model performance.

Sr. No.	**Publication Details**	**Methodology & Observations**	**Challenges & Limitations**
2	Nhan Ngo, Toi Vo, Lua Ngo (2022). "Application of deep learning in chest X-ray abnormality detection". Life Sciences — Medicine, Biomedical Applications	• Enhanced bounding box efficiency using WBF, surpassing Soft-NMS. Stratified dataset into standard (R8, R9, R10 - 94.35% images) and Additional (R11 through R17) categories for improved organization. • Applied extensive data augmentation post Final data generation, incorporating strategies like padding, horizontal flip, and image rotation to enhance model training. • Calculated optimal IoU threshold for different diseases, with the ideal value being 0.4 while achieving a mAP@0.5 of 0.2174 with ResNet50 as the backbone. • Utilized augmented dataset with ResNet50 as the backbone, achieving mAP at IoU 0.5 of 0.3052 for five diseases. With ResNet101 as the backbone, mAP at IoU 0.5 improved to 0.3193.	• Using heavily scaled-down 3x compressed JPEG images instead of full-quality DICOM images has reduced image quality, diminishing clarity, and hence possibly lowering the overall mAP in the pipeline. • The use of the cosine warm-up function resulted in extended training times, in this case it comprises a significant portion (20%) of the total training duration, increasing the overall training duration.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
3	Heyang Huang, Yijung Long, Li Wei (2021). “Chest X-ray Abnormalities Detection”. Stanford Winter Report	• Removed the no-observation class to address class imbalance. Resized images to a standardized 512x512 format, and utilized YOLOv5 for efficient bounding box creation. • Avoided random cropping to while augmentations to preserve dataset integrity and aid experienced doctors in abnormality detection. • Explored training set augmentation at between 10% and 50%, settling on 20% due to no discernible improvement in mAP beyond this threshold. • Achieved best mAP@0.5 of 0.34 with the YOLOv5x after additional hyperparameter tuning. • Model exhibited notable success in predicting Aortic Enlargement and Cardiomegaly but faced challenges with Other Lesion, Infiltration, and Atelectasis. Attempts to address this involved oversampling minority cases, yet the coexistence of these cases with common abnormalities hindered effective oversampling.	• The model struggles to detect certain abnormalities, often misclassifying them as background due to limitations in medical expertise for error interpretation. • To address this, future efforts may involve seeking assistance from medical experts to gain insights into the underlying patterns of these errors, acknowledging that some may surpass the understanding of even top doctors.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
4	Ngoc Huy Nguyen, et. al (2022). “Deployment and validation of an AI system for detecting abnormal chest radiographs in clinical settings”. Frontiers in Digital Health.	• Utilizes a Posterior-Anterior (PA) classifier with the ResNet-18 and EfficientNet-B6 architecture as the backboneone to ensure that only PA-view CXRs are passed to the abnormality classifier, trained exclusively on this image type. The PA classifier outputs the probability of the input being a PA-view CXR, with images surpassing a threshold of 0.5 forwarded to the abnormality classifier. • The lesion detector localizes and classifies different types of lesions on abnormal CXRs using bounding boxes, attaining a mAP@0.4 of 0.365 • The method’s F1 score drops from 0.831 to 0.653 during the transition from training to clinical deployment, potentially influenced by shifts in CXR image distribution or additional clinical information for radiologists.	• The combined use of ResNet-18 and EfficientNet-B6 architectures in training amplifies computational demands, making the process more resource-intensive. • This hybrid approach, while enhancing feature extraction, extends training duration and adds complexity to system orchestration, impacting efficiency and manageability.

Sr. No.	**Publication Details**	**Methodology & Observations**	**Challenges & Limitations**
5	Ines Feki, Sourour Ammar, Yousri Kessentini, Khan Muhammad (2021). "Federated learning for COVID-19 screening from Chest X-ray images". Applied Soft Computing 106	• Applies a CNN with FedAvg, where a central server maintains a global model shared with clients for collaborative updates, resulting in a powerful model from private datasets. • Each round, local models use the global weight. After local epochs and SGD iterations, clients update the model for privacy. The server oversees training, distributing the initial model. • The federated VGG16 showed a 0.4% accuracy gain over the centralized model, while the federated ResNet50 had a 0.5% improvement. However, the federated learning model took two to three times longer to converge compared to the centralized model. • The federated learning model with a 0.75 client dropout rate exhibited a 5% lower accuracy compared to the model with a 0 dropout rate, reflecting realistic scenarios.	• The paper doesn't address communication costs or data transmission efficiency, critical in scenarios with limited bandwidth or cross-geographical implementations. • The paper overlooks potential algorithmic bias or fairness issues when aggregating models trained on diverse local datasets, essential for ethical application across different demographics or data distributions.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
6	Adrian Nilsson et. al (2018). A Performance Evaluation of Federated Learning Algorithms. DIDL '18: Proceedings of the Second Workshop on Distributed Infrastructures for Deep Learning	• Conducted analysis on three federated averaging strategies for enhancing distributed machine learning models. FedAvg aggregates weights from local models to update the global model efficiently. Federated Stochastic Variance Reduced Gradient (FSVRG) reduces variance in stochastic gradient updates for stable convergence. CO-OP uses an asynchronous update mechanism for real-time integration of client models with the global model, promoting continuous learning in dynamic environments. • FedAvg excels with i.i.d. datasets, proving reliable in uniform data distribution scenarios. Its effectiveness wanes with non-i.i.d. data, presenting a challenge in federated learning with diverse client data. • Despite this, FedAvg remains competitive in heterogeneous data settings, showcasing versatility and robustness across scenarios.	• Although FSVRG is more resource-effcient as compared to FedAvg, the comparision between the two highlights the trade-off between computational demands and performance in federated learning settings. • The CO-OP Method, while offering unique advantages in certain scenarios, has been associated with a notable drawback – the potential for deadlocks, introducing a level of uncertainty and operational risk where the CO-OP Method is employed.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
7	Akhil Vaid et. al (2020). "Federated Learning of Electronic Health Records Improves Mortality Prediction in Patients Hospitalized with COVID-19". JMIR Med Inform.	• A centralized federated model with randomly initialized parameters was deployed. After one training epoch at each site, model parameters were sent back for federated averaging. This process, involving scaled site parameters and layer-wise aggregation, ensured secure data transmission without raw data sharing. • Both federated LASSO and federated MLP models outperformed their local counterparts in the hospitals implemented hospitals. • The federated MLP model consistently outperformed the federated LASSO model across all hospitals, showcasing the potential of federated learning in COVID-19 EHR data for robust predictive models while preserving patient privacy. • Introduction of Gaussian noise into federated MLP resulted in decreased performance at all sites, where they achieved an AUC-ROC of 0.822. AUC-ROCs ranged from 0.796 to 0.834 in federated MLP without noise and 0.767-0.830 in federated MLP with noise	• The study could not establish an operational framework for immediate deployment, leaving aspects like load balancing, convergence, and scaling unexplored for patient EHR data. • While identical MLP architectures were employed for direct comparisons across all learning strategies, there exists potential for further optimization in these architectures.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
8	Binhang Yuan, Song Ge, Wenhui Xing (2022). "A Federated Learning Framework for Healthcare IoT devices". Arxiv	• The intermediate weights are decomposed such that each device includes a local version of the first shallow component, while the remaining part is located on the centralized server. • The findings reveal an insignificant delay in convergence between the vanilla SGD and the proposed algorithm across 16, 32, and 64 devices, with the final accuracy loss being less than 2%. • The first convolution layer in both the edge device and the proposed framework are assigned, significantly reducing network traffic, showing a reduction of 99.8% and 90% compared to FedAvg and SplitNN, respectively.	• While the study employed only 64 devices for training, it is imperative to consider the likelihood of a more extensive network of hospitals in real-world scenarios, enhancing our understanding of scalability and generalizability. • Not having client dropout rates hinders result validity. Recognizing challenges with clients intermittently participating in federated learning would enhance realistic model evaluation.

Sr. No.	Publication Details	Methodology & Observations	Challenges & Limitations
9	Yiqiang Chen et. al (2021). "FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare". Arxiv	• The proposed FedHealth framework leverages federated transfer learning for precise, personalized healthcare while ensuring data privacy. It starts with a server-trained cloud model using public datasets, followed by user-specific training without data sharing, using encrypted model parameters. The approach incorporates transfer learning for model customization, maintaining privacy through homomorphic encryption. • By exchanging only encrypted parameters, FedHealth facilitates the creation of user-tailored models from combined cloud and personal data, overcoming privacy and data diversity challenges for continual personalization. • FedHealth surpasses traditional and non-federated models in activity recognition, offering a notable 5.3% accuracy improvement. This efficiency highlights its potential to improve healthcare outcomes through federated learning.	• Federated learning with homomorphic encryption increases computational complexity and requires additional hardware capabilities, complicating sytem setup and maintenance. • Scaling FedHealth for a large user base poses challenges in managing efficient model updates and parameter sharing, with the effectiveness of personalization relying on diverse user data, potentially impacting overall performance.

Chapter 3

Research Gaps and Problem Statement

3.1 Research Gaps

While the application of deep learning to medical imaging has seen considerable advancements, the integration of these techniques into the practical workflows of healthcare systems remains fraught with challenges. One of the most significant gaps in current research lies in the area of data privacy and security. The sensitivity of medical data necessitates stringent compliance with privacy regulations, which traditional centralized machine-learning models often struggle to meet [10].

Another problem is the chance of algorithmic bias. AI models, like those using YOLOv9, are usually trained on datasets that don't completely represent the global patient population. Most studies and implementations overlook the variation in medical data based on geography and institutions. This oversight can result in models that work well on their training data but struggle with datasets from diverse demographics or using different imaging equipment [8].

Despite federated learning's potential to uphold data privacy, its adoption in healthcare contexts is limited. The complex nature of medical data, characterized by extensive dimensionality and the requirement for meticulous annotation, poses distinct challenges not fully addressed by existing federated learning solutions [2]. The research community has yet to fully explore and develop methodologies for federated learning that effectively adapt to the intricacies of medical imaging. This includes accommodating diverse imaging modalities and ensuring compatibility with the varied standards employed across different healthcare institutions.

3.2 Problem Statement

The objective of this study is to develop an artificial intelligence (AI) model that is both secure and efficient for diagnosing diseases from X-ray images. This model aims to adhere to stringent data privacy laws, offering performance on par with traditional centralized training approaches. Through the utilization of federated learning, the research seeks to ensure that the AI model performs consistently and effectively across different healthcare institutions. The study also aims to augment medical diagnostic processes by integrating AI with the expertise of medical professionals, thereby enhancing the accuracy of diagnoses. By employing YOLOv9 technology, this research intends to create a scalable and responsive tool that supports radiologists in their work. This initiative is geared towards advancing the field of digital healthcare by introducing advanced, privacy-conscious clinical decision support systems (CDSS).

Chapter 4

Proposed Methodology

4.1 Lesion Identification on the Edge Device

In this study, we introduce a federated learning framework based on a client-server architecture, depicted in Figure 4.2 for the classification of CXRs into categories indicative of any of the following 14 diseases: Aortic enlargement, Atelectasis, Calcification, Cardiomegaly, Consolidation, ILD, Infiltration, Lung Opacity, Nodule/Mass, Other lesion, Pleural effusion, Pleural thickening, Pneumothorax, Pulmonary fibrosis. A centralized parameter server oversees the maintenance of a global model, which is distributed among the clients. These clients, operating on their respective private datasets, collaborate to enhance the model's accuracy in detecting any of the aforementioned chest lesions.

In our approach, we employ a Convolutional Neural Network (CNN) in combination with the advanced capabilities of YOLOv9 for the purpose of for the precise identification of chest lesions within CXRs. Initially, the process begins with the input of an X-ray image into the pre-trained CNN, VGG16, which is designed to detect the presence of a lesion. This detection is based on whether the presence of a lesion exceeds a certain threshold of confidence. Upon surpassing this threshold, the image is then forwarded to the YOLOv9

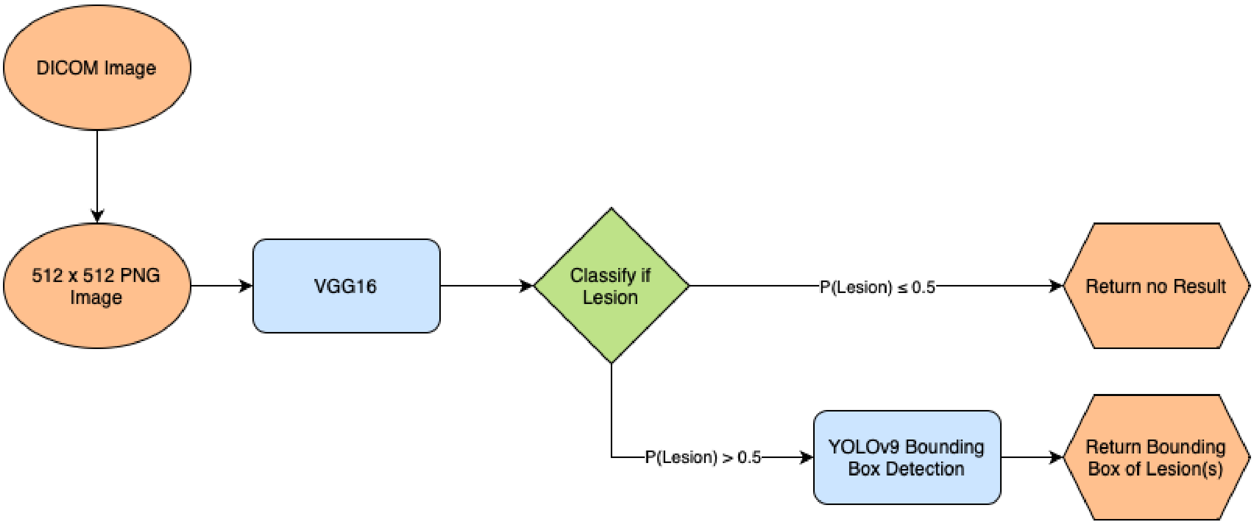


Figure 4.1: The proposed framework for the lesion detection system

model. The role of YOLOv9 in this pipeline is to further analyze the X-ray image, providing a detailed output that includes not only the probability of lesion presence but also an accurately placed bounding box that delineates the lesion. This methodology ensures a comprehensive analysis of the X-ray imagery, facilitating the accurate detection and localization of chest lesions. The sequential flow of this process, from the initial input to the final detection and classification stages, is clearly illustrated in Figure 4.1.

The learning process of the YOLOv9 model involves multiple rounds of communication between the central server and the clients. Initially, the YOLOv9 model is seeded with random weights w_0. Assuming the existence of K clients, each possessing a private collection of n_k CXRs, the learning phase unfolds over several steps in each communication round t:

1. The central server initializes the global model g with weights w_{t-1} and shares it with a randomly selected subset of clients S_t, based on a fraction C, where $C \in [0, 1]$.

2. Upon receiving the initial parameters w_{t-1}, each client $k \in S_t$ proceeds to train on a mini-batch b of its local data. This training is aimed at

minimizing the local objective function F_k with a local learning rate η_{local} over a predetermined number of epochs E. The optimization focuses on reducing the categorical cross-entropy loss for the classification task.

3. After completing the local training, clients from S_t transmit their updated model weights w_{kt}, with $k \in S_t$, back to the central server.

4. The server aggregates these updates from all participating clients, computing a new average model w_t according to the Equation 4.1 if using Federated Averaging, and Equation 4.2 if using Weighted Federated Averaging to update the global model g's parameters.

$$w_t \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{kt} \tag{4.1}$$

$$w^t \leftarrow \sum_{k=1}^{K} \frac{n_k}{N} w_k^t \tag{4.2}$$

In these equations, the parameter w^t represents the global model weights updated at training round t. The term w_k^t denotes the local model weights sent by client k at round t, capturing the individual contributions of each client's model to the federated learning process. The variable n_k indicates the number of data points stored by client k, while N stands for the total number of data points across all participating clients, with $N = \sum_{k=1}^{K} n_k$. This formula ensures that the global model update at each round takes into account the volume of data contributed by each client, weighted by their respective data size, thus aiming for a more representative and fair aggregation of local model updates across the network.

These steps represent a single round of federated learning for the YOLOv9 model, a process which is iterated through numerous rounds. This process is elucidated in 4.2 It is important to note that with each new round t, the

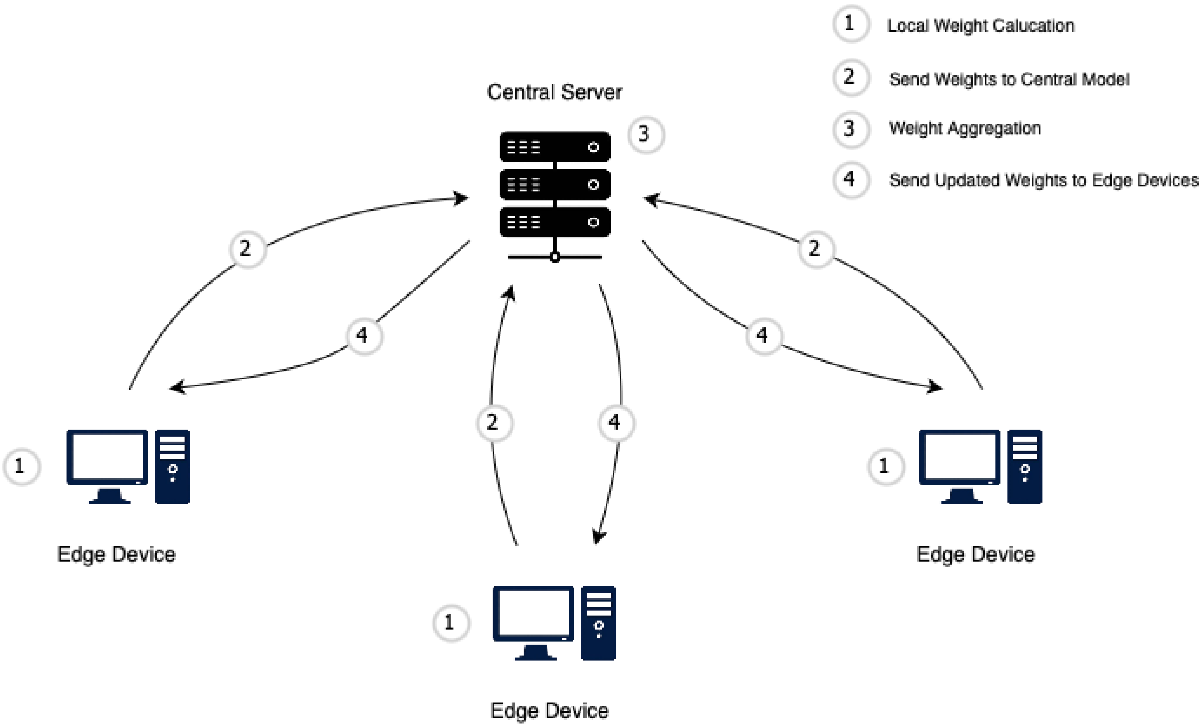


Figure 4.2: Proposed Architecture of the Federated Learning Model

server distributes the newly updated parameters w_{t-1} of the global model g, which were assembled in the preceding round $t-1$. Additionally, the selection of client subsets S_t can vary between rounds to accommodate the availability of multiple clients.

4.1.1 YOLOv9 Architecture for Lesion Detection

In the context of lesion detection, YOLOv9 has been tailored to the demands of medical image analysis. The architecture's novel features, such as Pyramid Gradient Integration (PGI) and the Generalized Erosion Linear Activation Network (GELAN), address the common issue of information degradation in deep neural networks. These features work in concert to enhance the model's capacity to preserve essential diagnostic information throughout the layers, culminating in a model that delivers superior performance metrics in the specialized task of medical lesion detection. Its architecture is elucidated in Figure 4.3

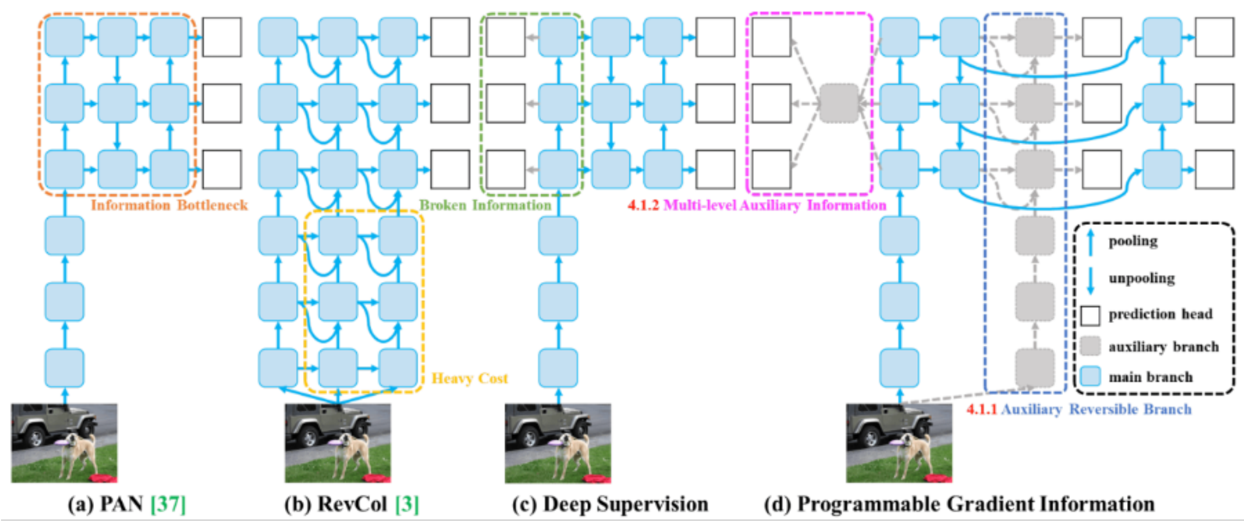


Figure 4.3: YOLOv9 detailed architecture [13]

4.1.2 Client-Side Model Update

The implementation of federated learning in this study is centered around the training activities that occur at the individual client level. Clients, each with a specific dataset and computational capabilities, operate independently to contribute to the development of a global model. This method ensures the utilization of distributed data sets for model improvement while maintaining strict adherence to data privacy standards.

In the proposed system, outlined in Algorithm 1, a network of clients is equipped with the YOLOv9 architecture, all standardized with identical loss functions to ensure consistency in training outcomes. This uniformity facilitates the integration of local updates into the global model managed by the server. At the start of each federated learning round, denoted by t, clients initialize their local models with the global weights w_t received from the server. This step marks the beginning of a series of training epochs at the client side.

Through these training epochs, each client adjusts its model by learning from its unique data set. The training involves a series of optimization steps where the model parameters are updated to minimize the local loss function, reflecting the data's distinct characteristics. The outcome of this training is an updated set of model parameters, representative of the learning that has taken place within the client's domain.

Once the local training epochs are completed, the clients are tasked with sending their updated model parameters to the central server. It is important to highlight that only the parameters of the model are transmitted, not the actual data, thereby preserving the privacy of the data. This process enables the server to integrate the learnings from all clients while ensuring that each client's data remains confidential.

This methodology, at the intersection of local data processing and global model enhancement, allows for the preservation of privacy while leveraging the distributed nature of the data. It contributes to the overall robustness and accuracy of the global model by incorporating diverse insights from multiple clients into a single, enhanced framework.

4.1.3 Server-Side Model Aggregation

The federated learning paradigm hinges on the intricate process of server-side model aggregation. The server, which has the global model, orchestrates the training progression, ensuring coherent and synchronized model evolution across all clients. At the onset of each federated learning round t, the server dispatches the current global model to the participating clients, catalyzing the local training phase.

Upon the completion of local training across the network, the server's role shifts to receiving updates. Each client, having iterated over its local dataset and computed updates to the model parameters, sends these adjustments back to the server. The task of the server is to aggregate these individual contributions to refine the global model, as expressed in Algorithm ??.

The server initiates the process by collecting the updated weights w_k^t from each client k. The objective is to synthesize a new set of global weights w^t that encapsulate the distributed learnings from all clients, as mathematically defined by Equation 4.1.

The server not only performs aggregation but also monitors the progress of model training. It assesses the convergence and performance metrics, adjusting the learning protocol as necessary to optimize the collective learning trajectory. This ongoing management is crucial for addressing any discrepancies or drifts that might emerge during the federated learning process.

After aggregating the updates, the server propagates the new global model w^t back to the clients, instigating the next round of local learning. This cyclical process of distribution, local training, collection, and aggregation forms the iterative heartbeat of federated learning, driving the model towards improved accuracy and robustness.

The server-side model aggregation is a multifaceted process that involves the collection local model updates, ensuring that the global model reflects a comprehensive understanding of the distributed data.

4.2 Experimental Setup

This sections reports all the details regarding the datasets used and the experimental setup.

4.2.1 Dataset

The dataset employed in this study was sourced from VinBigData, an institution committed to advancing fundamental research in emergent and critical technologies. Their medical imaging division focuses on the acquisition, processing, and analysis of medical data, striving to create scalable, high-accuracy medical imaging solutions that leverage recent advancements in artificial intelligence to enhance clinical workflows [1].

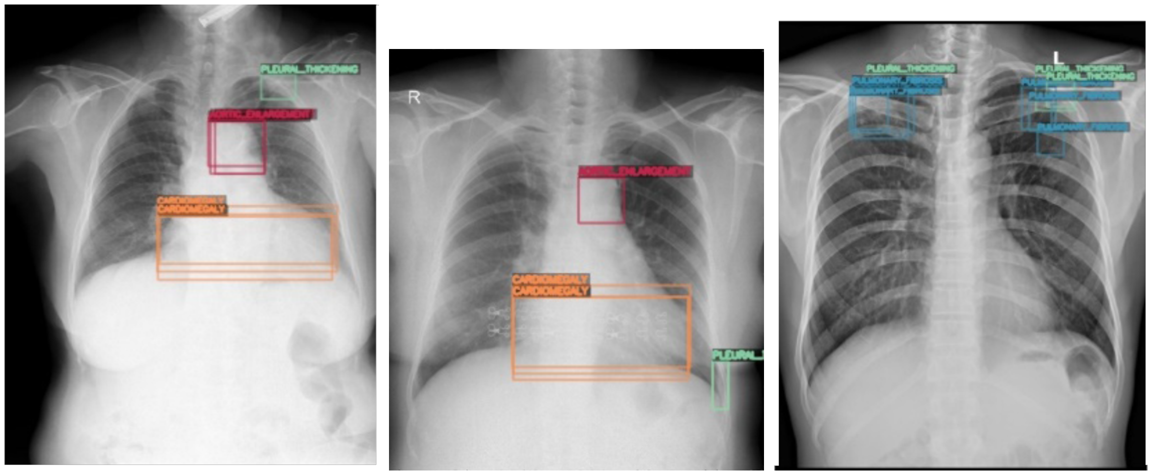
The VinDr-CXR dataset, which was utilized for this research, comprises over 100,000 raw Digital Imaging and Communications in Medicine (DICOM) images retrospectively collected from two prominent Vietnamese healthcare institutions: Hospital 108 and the Hanoi Medical University Hospital [4]. This dataset is distinguished by its inclusion of 18,000 posteroanterior (PA) chest X-ray (CXR) images, which have been meticulously annotated to identify and classify common thoracic conditions. A cohort of 17 seasoned radiologists, each with a minimum of eight years of professional experience, contributed to the annotation process. These experts provided detailed local labels for 22 critical findings, each marked with a bounding box, and 6 global diagnostic labels.

The dataset is bifurcated into a training set, with 15,000 images, and a test set, comprising 3,000 images. Representative examples of the annotated CXRs are illustrated in Figure 4.4.

Figure 4.5 underscores the class imbalance evident in the dataset, particularly the contrast between normal CXRs and those indicative of pathology. To address this imbalance, image augmentation techniques were be employed to enhance the representativeness of the dataset, which are delineated in the following section.

4.2.2 Data Augmentation

To mitigate the risks of overfitting and to bolster the model's generalization to new data, our methodology incorporated a diverse suite of image augmen-



(a) Example CXR 1 (b) Example CXR 2 (c) Example CXR 3

Figure 4.4: Examples of CXRs with radiologist's annotations

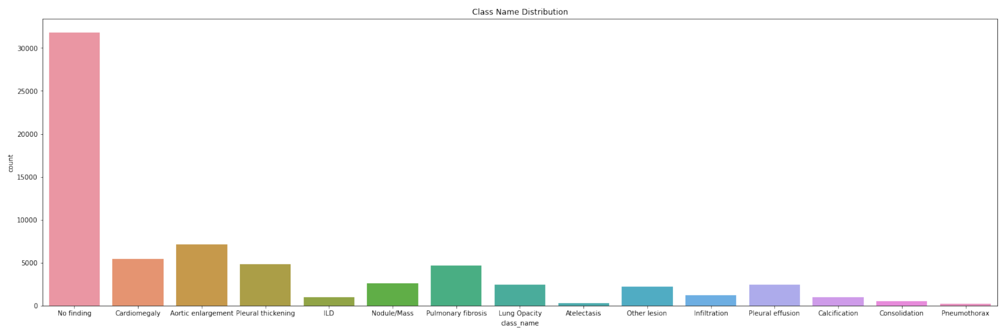


Figure 4.5: Distribution of Findings and pathologies on the training set of the VinDr-CXR Dataset

tation techniques. These techniques were designed to emulate the range of variations that naturally occur in medical settings, including fluctuations in imaging perspectives, illumination, and patient anatomy.

The augmentation protocol extended beyond basic transformations such as rotations, flips, and shifts. We introduced additional augmentations like scaling, contrast modification, color adjustments, and shearing. Collectively, these methods enriched the training dataset, equipping the model to recognize and interpret features across a spectrum of clinical imaging conditions.

We ensured this approach calibrated these augmentations to maintain the clinical integrity of the images, ensuring that the images are clinically pertinent, fostering a model that is both accurate and reliable in practical diagnostic applications.

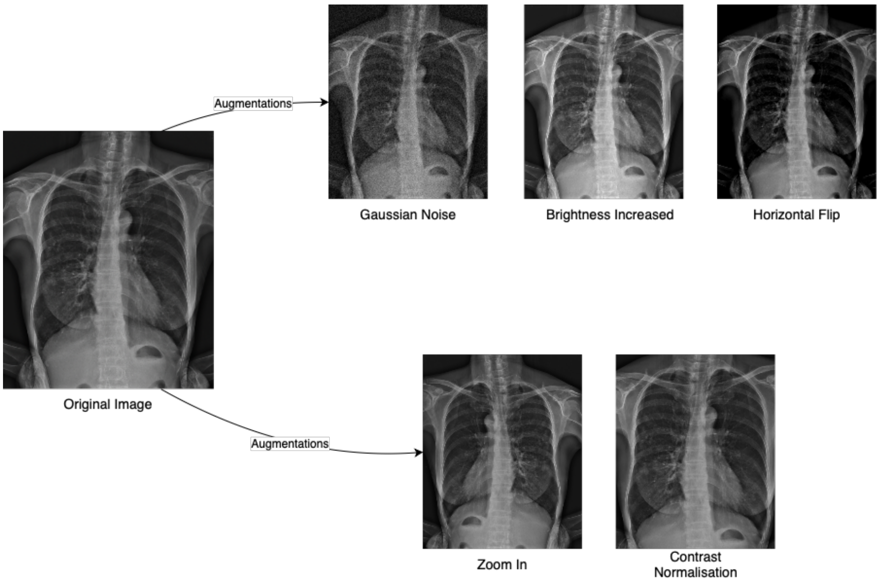


Figure 4.6: Image Augmentations

4.2.3 Training of the Detection Models

The VGG16 convolutional neural network model underwent training utilizing the Vinbigdata dataset, distinguishing between two categories: presence and absence of lesions. The YOLOv9 detection network received training through a federated learning approach, employing GELAN weights. Each participating local machine was outfitted with an NVIDIA Tesla P100 16GB GPU. The model's hyperparameters were maintained uniformly across the training process, as detailed in Table 4.1.

4.2.4 Evaluation Metrics

To quantitatively assess the performance of the lesion detection model, we employ Mean Average Precision (mAP), a standard metric for evaluating object detection algorithms. mAP provides an aggregate measure of precision

Table 4.1: The specifics of the hyperparameters used for training the YOLOv9 models

Hyperparameters	**Values**
Initial Learning rate	0.01
Final Learning rate	0.1
Momentum	0.937
Batch Size	20
Epochs	300
Optimizer	Stochastic Gradient Descent
IoU training threshold	0.2
Anchors per output layer	4.0
Image Size	512 x 512

across different recall levels, offering a comprehensive view of model accuracy. Specifically, we utilize two variants of mAP: mAP@0.5 and mAP@0.5:0.95.

$$\text{mAP@0.5} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i\text{@0.5} \tag{4.3}$$

Equation (4.3) defines mAP@0.5, where N is the number of queries, and AP_i@0.5 is the average precision at IoU threshold of 0.5 for the i-th query. This metric considers a detection to be correct if the Intersection over Union (IoU) between the predicted bounding box and the ground truth is greater than or equal to 0.5.

For a more rigorous evaluation that accounts for varying levels of detection difficulty, we also calculate mAP at different IoU thresholds, ranging from 0.5 to 0.95 with a step size of 0.05:

$$\text{mAP@0.5:0.95} = \frac{1}{10} \sum_{t=0.5}^{0.95} \text{mAP@}t \tag{4.4}$$

Equation (4.4) details the computation of mAP@0.5:0.95, which averages the mAP computed at each IoU threshold t, offering a more detailed evaluation metric that considers the model's performance across a range of IoU thresholds.

Chapter 5

Results and Discussion

5.1 Results and Discussion

This study of federated learning for multi-class lesion detection from CXRs to highlight the effectiveness of this type of decentralized and collaborative learning in such context where data is private.

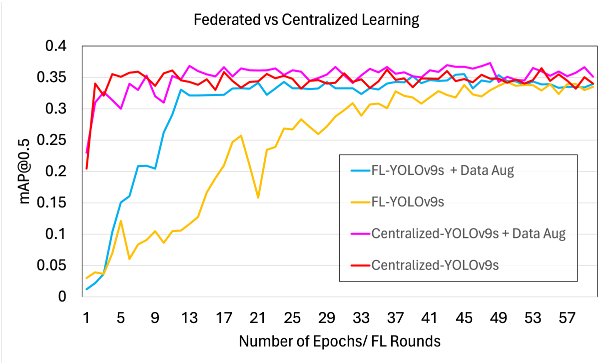
First, we compare our decentralized method with the centralized one. Then, we study the effect of the parameter the fraction of clients participating in each training round C on the model performance after each round when we deal with IID data distribution. Finally, we compare the two distribution settings IID and non-IID data over various weight aggregation strategies.

5.1.1 Comparative Analysis of Federated and Centralized Training Approaches

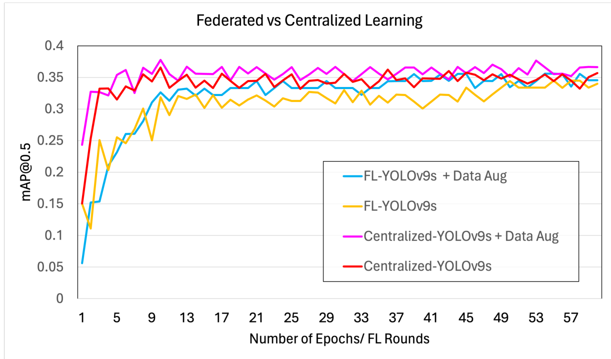
In the exploration of Federated Learning (FL) configurations, C is set as $C = 1$, meaning all clients partake in each training round, while using Federated Averaging as the aggregation method, with the results in Tbl. 5.1. Illustrated in Figure 5.1, the findings from the FL model aggregating for 60 epochs demonstrate that the FL methodology devised in this study attains a classification efficacy on par with traditional approaches, yet it does so without necessitating the exchange of client data. Particularly noteworthy, as depicted in Figure 5.1(a), is the observation that beginning from round 49, the classification performance of our FL framework, when trained on the original dataset, closely aligns with the performance metrics, specifically (mAP@0.5) of a Centralized training model trained on an identical dataset. Moreover, Figure 5.1(b) reveals that the implementation of the more advanced FL architecture, FL-YOLOv9c, permits the attainment of comparative results to the centralized model in significantly fewer epochs.

Through this research, it is observed that the FL-YOLOv9s strategy, augmented with data enhancement techniques, yields results that are not only comparable to those obtained via Centralized-YOLOv9s with Data Augmentation but does so remarkably within just 13 rounds of training. A similar outcome is observed with the FL-YOLOv9c approach, which, despite utilizing the same architectural foundation, achieves equivalence with the centralized methods in terms of performance. Hence, this underscores the critical influence of client-side data volume on the ultimate results.

Another salient finding, as showcased in Figure 5.1(a), is that post approximately 50 rounds, all deployed methods leveraging the YOLOv9s converge in terms of performance, indicating a uniformity in results. This trend is mirrored in the analysis presented in Figure 5.1(b), where a comparable pattern of convergence is observed after round 50. Such outcomes underscore the efficiency and potential of the proposed FL framework. Similarly, the mAPs are also mirrored in the results of mAP@0.5:0.95. This framework not only facilitates the attainment of performance metrics akin to those of centralized approaches through iterative training rounds but does so while steadfastly upholding the principle of data privacy by obviating the need for direct data sharing.



(a) Results using YOLOv9s



(b) Results using YOLOv9c

Figure 5.1: Comparison of Federated Learning to Centralized Training using original and augmented dataset for learning

Table 5.1: Average mAPs

Method	mAP@0.5	mAP@0.5:0.95
Centralized-YOLOv9s	0.3403	0.1433
Centralized-YOLOv9s + data aug	0.3514	0.1597
FL-YOLOv9s	0.3349	0.1337
FL-YOLOv9s + data aug	0.3394	0.1398
Centralized-YOLOv9c	0.3568	0.1489
Centralized-YOLOv9c + data aug	0.3661	0.1612
FL-YOLOv9c	0.3401	0.1367
FL-YOLOv9c + data aug	0.3454	0.1403

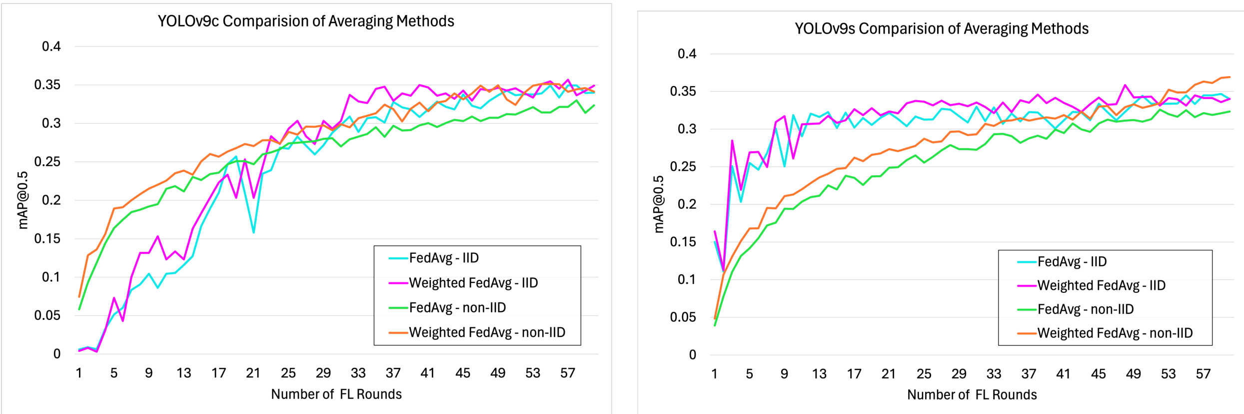
5.1.2 Comparative Analysis of Federated Averaging and Weighted Federated Averaging

This subsection elucidates the differential impacts of two distinct averaging methodologies—Federated Averaging and Weighted Federated Averaging—on both IID and non-IID datasets, as detailed in Table 5.2. The implementation of Federated Averaging follows the formula presented in Eq. 4.1, while the application of Weighted Federated Averaging adheres to the formula outlined in Eq. 4.2. In the context of IID datasets, Weighted Federated Averaging assigned enhanced weights to randomly selected clients, whereas for non-IID datasets, additional weighting was allocated to clients possessing more distinctive or novel data sets. Figure 5.2 illustrates the mAP@0.5 across 60 epochs of federated learning for both FL-YOLOv9s and FL-YOLOv9c models under each averaging method.

Analysis of Figure 5.2(a) indicates that within the YOLOv9s model framework, Federated Averaging yielded the least favorable performance on non-IID datasets, primarily due to the uniform weight distribution across inherently unequal data sets. Conversely, models trained on IID data demonstrated quicker convergence, highlighting the challenges Weighted Federated Averaging faces, requiring extensive data and prolonged training periods to reach convergence. Notably, with the exception of the Federated Averaging model on non-IID data, performances across the lightweight YOLOv9s configurations were relatively comparable, suggesting that the impact of averaging techniques is somewhat muted within models of this category.

In contrast, Figure 5.2(b) reveals that the YOLOv9c model experienced similar challenges with Federated Averaging on non-IID data, suffering from the same performance detriments as its YOLOv9s counterpart. Models operating on IID data achieved earlier convergence, whereas Weighted Federated Averaging models necessitated additional data and training time for convergence. Remarkably, the Weighted Federated Averaging model tailored for non-IID data outperformed its counterparts, including models trained on augmented data as noted in Tbl. 5.1, by prioritizing the inclusion of novel data. The parity observed between the Weighted Federated Averaging model on IID data and the Federated Averaging model on IID data underscores the efficacy of novel data prioritization in enhancing model performance.

In essence, this research underscores the nuanced efficacy of Federated Averaging and Weighted Federated Averaging techniques in different data distribution scenarios. It reveals the inherent limitations of Federated Averaging in non-IID contexts and highlights the potential of Weighted Federated Averaging to improve model performance through strategic weighting, especially in the presence of novel data.



(a) Results on YOLOv9s

(b) Results on YOLOv9c

Figure 5.2: Comparison of Federated Averaging methods on IID and non-IID datasets

Table 5.2: Average mAPs for Centralized and Federated Models

Model	Method	mAP@0.5	mAP@0.5:0.95
FL-YOLOv9s	FedAvg - IID	0.3349	0.1389
	Weighted FedAvg - IID	0.3352	0.1291
	FedAvg - non-IID	0.3235	0.1268
	Weighted FedAvg - non-IID	0.3442	0.1477
FL-YOLOv9c	FedAvg - IID	0.3401	0.1367
	Weighted FedAvg - IID	0.3403	0.1368
	FedAvg - non-IID	0.3359	0.1339
	Weighted FedAvg - non-IID	0.3691	0.1611

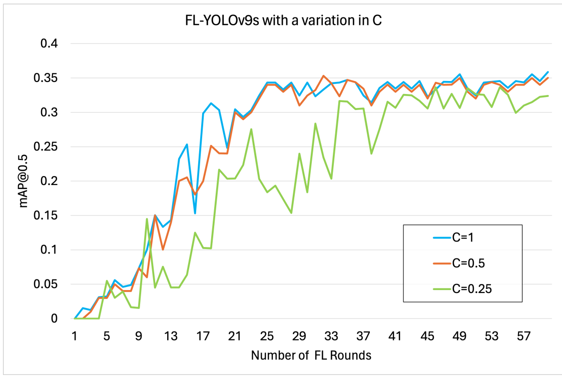
5.1.3 Results on IID Data with Varying Client Dropout

This section investigates the impact of varying client participation rates on the performance of federated learning models, specifically within the context of an IID data partition, with the results being in Tbl. 5.3. The fraction of clients participating in each training round, denoted by C, was manipulated to examine its effect on multi-client parallelism. A value of $C = 1$ indicates full participation of all available clients in the collaborative training for each round, whereas $C = 0.25$ signifies that only a quarter of the clients are selected at each iteration. This study reports on the outcomes associated with three distinct settings of the parameter C: 1, 0.5, and 0.25.

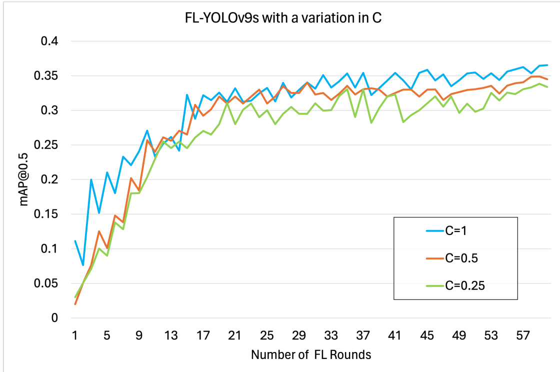
Figure 5.3 displays the $mAP@0.5$ plotted over federated learning epochs up to 60 for both FL-YOLOv9s and FL-YOLOv9c models. According to Figure 5.3(a), the FL-YOLOv9s model demonstrates faster convergence to comparable mAPs when the participation rate is maximal ($C = 1$; blue curve) compared to when only half ($C = 0.5$; brown curve) or a quarter ($C = 0.25$; green curve) of clients are engaged in each round. The diminished rate of client participation results in more variable outcomes and delayed convergence, reflecting the limited scope of collaborative learning under these conditions. Specifically, when fewer clients contribute to the model updates, the aggregated model at the server is primarily influenced by the data from the participating clients, which variably impacts model quality.

In the case of FL-YOLOv9c, as depicted in Figure 5.3(b), a similar pattern of convergence is observed across all client participation rates, with marginally improved accuracy observed for $C = 0.5$ and $C = 0.25$. This research further identifies that the adoption of a more sophisticated YOLOv9c architecture facilitates earlier convergence. Notably, the mAP@0.5 performance metrics exhibit a successive decrease with client dropout rates of 1, 0.5, and 0.25, attributable to the reduced client involvement per round.

The findings from this analysis of IID data partitions suggest that engaging a higher number of clients in each training round enhances the accuracy at convergence, thereby necessitating fewer rounds for the learning process to achieve comparable precision levels.



(a) Results using YOLOv9s



(b) Results using YOLOv9c

Figure 5.3: Effect of the client fraction C on the test accuracy of YOLOv9 models. Note $C = 1$ corresponds to all clients are selected at each round, $C = 0.5$ corresponds to half clients and $C = 0.25$ corresponds to only one client per round

Table 5.3: Average mAPs over varying client dropout on IID dataset

Method	Client Dropout (C)	mAP@0.5	mAP@0.5:0.95
	1.0	0.3585	0.1558
FL-YOLOv9c	0.5	0.3501	0.1512
	0.25	0.3241	0.1211
	1.0	0.3652	0.1624
FL-YOLOv9s	0.5	0.3611	0.1601
	0.25	0.3339	0.1241

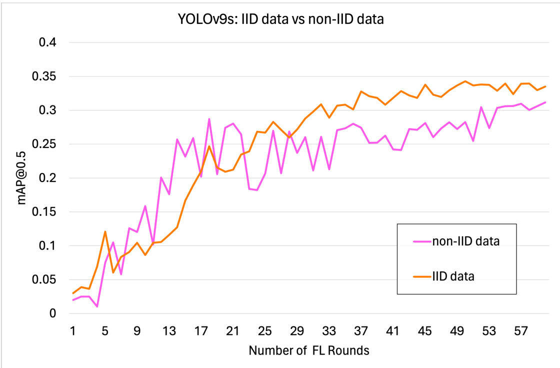
5.1.4 Comparative Analysis of Performance on IID versus non-IID Datasets

This section delves into an evaluative comparison between FL methodologies applied to both IID and non-IID datasets, maintaining a constant client participation rate ($C = 1$) throughout the analysis, with the results in Tbl. 5.4. The empirical outcomes, illustrated in Figure 5.4, underscore the disparate performance dynamics inherent to the two data distribution paradigms.

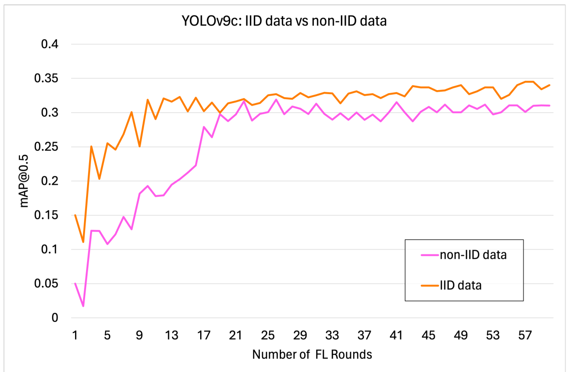
Observations from Figure 5.4(a) reveal a pronounced volatility in the mAP@0.5metrics for the FL-YOLOv9s model when applied to non-IID data, characterized by erratic fluctuations across federated learning rounds. Notably, a semblance of stability in performance metrics emerges beyond the 41st epoch, signaling a delayed adaptation to the heterogeneity of the non-IID data distribution. This contrast starkly with the performance trajectory observed for the same model on IID data, wherein a more consistent and predictable improvement in mAP@0.5 is documented, highlighting the critical impact of data homogeneity on model performance and the inherent challenges posed by non-IID datasets.

Further insights are visible from Figure 5.4(b), which depicts a significantly accelerated convergence for the FL-YOLOv9c model on IID data, achieving a stable performance plateau as early as the 17th epoch. The advanced architectural nuances of the YOLOv9c model facilitate an expedited adaptation to the intricacies of the data, as evidenced by the relatively prompt convergence (around 25 epochs) even in the context of non-IID data. This rapid attainment of convergence, particularly when compared to the FL-YOLOv9s model, demonstrates the enhanced capability of more sophisticated models to navigate the complexities of diverse data distributions effectively.

The juxtaposition of performance metrics across both the FL-YOLOv9s and FL-YOLOv9c models reaffirms the superior efficacy of training on IID datasets, as visible in significantly higher mAP@0.5 outcomes. Moreover, the FL-YOLOv9c model's ability to extract detailed lesion features from non-IID data, achieving commendable mAP@0.5 metrics surpassing those of the FL-YOLOv9s model trained on IID data, underscores the pivotal role of advanced model architectures in optimizing federated learning outcomes. This analysis substantiates the profound influence of data distribution characteristics on the ultimate model performance within federated learning frameworks, accentuating the necessity for robust model design to counteract the performance degradation associated with non-IID datasets.



(a) Results using YOLOv9s



(b) Results using YOLOv9c

Figure 5.4: Comparison of Federated Learning results on IID data and non-IID data partitions with C = 1 (all clients are considered at each round)

Table 5.4: Average mAPs over varying client dropout on IID dataset

Method	Dataset	mAP@0.5	mAP@0.5:0.95
FL-YOLOv9s	IID	0.3349	0.1337
	non-IID	0.3114	0.1211
FL-YOLOv9c	IID	0.3401	0.1367
	non-IID	0.3103	0.1258

Appendix A

Algorithm for Client-Side Training

Algorithm 1 Client-side Training with Lesion Detection

Require: local learning rate η, loss function ℓ, pre-trained VGG16 model M_{VGG16}

procedure CLIENTUPDATE(w^t, P_k) $\triangleright$ Where P_k is the local data of client k
    $w \leftarrow w^t$
    Use M_{VGG16} to predict lesion probability p on P_k
    if $p > 0.5$ **then**
        $B \leftarrow$ Split P_k into batches of size B
        for each local epoch i from 1 to E **do**
            for each b in B **do**
                Compute gradient $g^b \leftarrow \nabla \ell(w; b)$ using YOLOv9 specifics
                Update local model $w \leftarrow w - \eta g^b$
            end for
        end for
    else
        Update w to reflect no lesion detected by YOLOv9
    end if
    return w
end procedure

Appendix B

Algorithms for Server-Side Aggregation

Algorithm 2 Federated learning: server-side aggregation procedure (Federated Avg)

Require: T: num_federated_rounds

procedure AGGREGATING(C, K)
  Initialize global model w^0
  for each round $t = 1, 2, \ldots, T$ **do**
    $m \leftarrow \max(C \times K, 1)$
    $S_t \leftarrow$ (random set of m clients)
    for each client k in S_t **do**
      Send w^{t-1} to client k
      $w_k^t \leftarrow \text{ClientUpdate}(k, w^{t-1})$
    end for
    $w^t \leftarrow \frac{1}{K} \sum_{k=1}^{K} n_k w_k^t$
  end for
  return w^T
end procedure

Algorithm 3 Federated Learning: Server-side Aggregation Procedure (Weighted Federated Averaging)

Require: T: num_federated_rounds, N: Total number of data points across all clients

procedure AGGREGATING(C, K)
    Initialize global model w^0
    for each round $t = 1, 2, \ldots, T$ **do**
        $m \leftarrow \max(C \times K, 1)$
        $S_t \leftarrow$ (random set of m clients)
        $W_t \leftarrow 0$ ▷ Total weight sum for round t
        for each client k in S_t **do**
            Send w^{t-1} to client k
            $w_k^t, n_k \leftarrow \text{ClientUpdate}(k, w^{t-1})$ ▷ Receive updated model and data count
            $W_t \leftarrow W_t + n_k$
        end for
        $w^t \leftarrow \frac{1}{N} \sum_{k \in S_t} n_k w_k^t$ ▷ Update global model with weighted average
    end for
    return w^T
end procedure

Bibliography

[1] Vinbigdata - about us. Accessed: April 5, 2024.

[2] Robert Colgan, David Gutierrez, Jugesh Sundram, and T. Bhaskar. Analysis of medical data using dimensionality reduction techniques. 04 2013.

[3] Shanshan Du, Yingwen Wang, Xinyu Huang, Ruiwei Zhao, X. Zhang, Rui Feng, Quanli Shen, and Jianqiu Zhang. Chest x-ray quality assessment method with medical domain knowledge fusion. *IEEE Access*, 11:22904–22916, 2023.

[4] Ha Q Nguyen DungNB. Vinbigdata chest x-ray abnormalities detection. 2020.

[5] Ghada Hamed, M. Marey, S. El-Sayed, and M. Tolba. Yolo based breast masses detection and classification in full-field digital mammograms. *Computer methods and programs in biomedicine*, page 105823, 2020.

[6] Rong Han, Xiaohong Liu, and Ting Chen. Yolo-sg: Salience-guided detection of small objects in medical images. *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4218–4222, 2022.

[7] F. Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Q. Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence

in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2:230 – 243, 2017.

[8] Joffrey L. Leevy, T. Khoshgoftaar, Richard A. Bauder, and Naeem Seliya. A survey on addressing high-class imbalance in big data. *Journal of Big Data*, 5:1–30, 2018.

[9] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017.

[10] Goutham Ramakrishnan, A. Nori, Hannah Murfet, and Pashmina Cameron. Towards compliant data management systems for healthcare ml. *ArXiv*, abs/2011.07555, 2020.

[11] Adam Sadilek, Luyang Liu, Dung Nguyen, Methun Kamruzzaman, Stylianos Serghiou, Benjamin Rader, Alex Ingerman, Stefan Mellem, Peter Kairouz, Elaine O. Nsoesie, Jamie MacFarlane, Anil Vullikanti, Madhav Marathe, Paul Eastham, John S. Brownstein, Blaise Aguera y. Arcas, Michael D. Howell, and John Hernandez. Privacy-first health research with federated learning. *npj Digital Medicine*, 4(1):132, 09 2021.

[12] Wasim Fathima Shah. Preserving privacy and security: A comparative study of health data regulations - gdpr vs. hipaa. *International Journal for Research in Applied Science and Engineering Technology*, 2023.

[13] Chien-Yao Wang and Hong-Yuan Mark Liao. YOLOv9: Learning what you want to learn using programmable gradient information. 2024.

[14] Zhendong Yao, Tao Jin, Boneng Mao, Bo Lu, Yefei Zhang, Sisi Li, and Weichang Chen. Construction and multicenter diagnostic verification of intelligent recognition system for endoscopic images from early gastric cancer based on yolo-v3 algorithm. *Frontiers in Oncology*, 12, 2022.