Lessons Learned from Implementation of Multi Institutional Collaborative Platform for Lung Disease Screening by means of Federated Learning in Indonesia

Agung Alfiansyah [1], Laurent Bobelin [2], Helena Widiarti [1]

[1]Universitas Prasetiya Mulya, Tangerang, Banten 15339, Indonesia
[2]INSA Centre Val de Loire, 18022 Bourges, France
agung.alfiansyah, helena.widiarti@prasetiyamulya.ac.id, lbobelin@insa-cvl.fr

Abstract

Purpose: This study aims to reduce lung disease mortality by developing an automatic screening system that analyzes X-ray images and utilizes distributed image data storage, creating a collaborative platform among Indonesian hospitals while ensuring patient data privacy. **Methods:** Using Federated Learning, a decentralized machine learning approach, hospitals build local models with their data, which are aggregated into a global model on a central server without compromising confidentiality. An innovative system for archiving medical images is also introduced, which anonymizes, secures, and curates data for training marchine learning based diagnosis systems. **Results:** The Federated Learning implementation resulted in a privacy preserved detection model that aggregates models from multiple hospitals while keeping patient data secure and remains in their silos. The archiving system successfully stores anonymized medical images and created a valuable dataset for CAD development. **Conclusion:** This work advances machine learning in healthcare, prioritizing patient privacy while enhancing X-ray analysis and collaborative model development. By addressing technical and ethical challenges, this framework sets a new standard for responsible AI in healthcare, with potential for application in other imaging modalities and diseases, aiming to revolutionize medical diagnostics.

Introduction

Indonesia, a significant nation in Southeast Asia, holds notable demographic and geographic prominence. It ranks fourth globally in population and fifth in Asia by land area. The country comprises 38 provinces, with a population exceeding 270 million distributed across 1,904,569 km^2. Notably, Indonesia's archipelagic nature encompasses over 17,000 islands, with Java being particularly densely populated at approximately 1,100 persons per km^2. The nation has experienced substantial improvements in life expectancy, increasing from 68.68 years in 2010 to 72.32 years in 2023. This progress correlates with economic advancements, as evidenced by Indonesia's ascension to upper-middle-income status in 2021. Average household income now approximates 10,089 U.S. dollars annually, reflecting the country's economic growth.

According to data from the Ministry of Health, noncommunicable diseases, specifically cardiovascular ailments, malignancies, and chronic respiratory conditions, still constitute primary morbidity and mortality factors. Concurrently, infectious diseases such as tuberculosis, COVID-19, dengue fever, and malaria persist as substantial health threats. Moreover, Indonesia confronts challenges in healthcare accessibility, predominantly attributed to socioeconomic disparities and geographical variations. Of particular concern is Indonesia's alarmingly high smoking prevalence among men, which exceeds 70% and ranks as the highest globally. This statistic underscores the significant threat lung cancer poses to the nation's public health. The high smoking rate exacerbates the already challenging health landscape, further complicating efforts to address both communicable and noncommunicable diseases effectively (UNICEF 2020).

The other significant challenge in the diagnosis and treatment of lung diseases in Indonesia is the inadequate distribution of pulmonologists, particularly in rural areas (of Public Health 2024). The data reveals a pronounced geographical imbalance, with pulmonologists predominantly concentrated in the western regions, specifically Sumatra and Java islands. However, even in these areas, the adequacy rate remains suboptimal. This distribution pattern is particularly concerning given that the highest incidence of pneumonia occurs in the eastern regions of the country, where pulmonologist availability is most scarce. This mismatch between healthcare resource allocation and disease burden underscores a critical issue in Indonesia's healthcare system, potentially impacting timely diagnosis and effective treatment of respiratory diseases in underserved areas.

Background

In the this modern context of medical diagnosis and treatment, deep learning models exhibit great potential, particularly in the interpretation of lung and pulmonary images and the diagnosis of conditions such as pneumonia (Rauf et al. 2023; Dalhoumi et al. 2015; Albahli et al. 2021). Nonetheless, their effectiveness hinges upon the availability of extensive and diverse datasets. A recent investigation has unveiled a critical concern—deep learning models tend to overfit to subtle biases present within institutional data, yielding suboptimal performance when tested against data from unseen institutions. Notably, these models might inadvertently in-

corporate confounding factors tied to institutional biases, diverting their predictive focus away from the intended pathological assessment. This phenomenon engenders high accuracy when assessed against internal data but falls short in generalizing outcomes across external institutions or even within different departments of the same institution.

An intuitive approach to surmount this challenge involves augmenting the scale and diversity of training data through collaborative learning paradigms, wherein multi-institutional datasets converge within a unified model during the machine learning process. This conventional practice of constructing models via data collection on a single server, however, introduces privacy concerns when transmitting data over the internet. Extensive evidence substantiates the viability of the re-identification process within anonymized datasets, achieved through the linking of such data with auxiliary datasets (Narayanan and Shmatikov 2006, 2008).

Data privacy is a fundamental concern in today's digital landscape, particularly in specific fields like healthcare. It involves safeguarding sensitive and personal information from unauthorized access, use, or disclosure. In the context of medical image processing for diagnosis, we need a platform that ensures patient-related information remains within the boundaries of each hospital's local infrastructure.

For instance, in a scenario where multiple hospitals collaborate to improve pneumonia detection using chest X-rays through Federated Learning, data privacy ensures that the actual X-ray images and associated patient identifiers never leave the individual hospital's network. Instead, only the model updates are shared exclusively across the collaborating hospitals. This policy prevents the risk of data leakage, where confidential patient information could inadvertently become accessible to unauthorized parties, thus upholding patient confidentiality and compliance with data protection regulations like HIPAA or GDPR.

One well-known example of a privacy infringement case in healthcare is the "UCLA Health Data Breach" (Alder 2015; Reddy et al. 2022) incident. The University of California, Los Angeles (UCLA) Health suffered a data breach that compromised the personal and medical information of approximately 4.5 million patients. The breach exposed sensitive information such as names, birth dates, social security numbers, medical record numbers, and even some medical conditions of patients. This incident had serious implications for patient privacy, as the compromised data contained personal and medical details that could be used for identity theft, insurance fraud, and other malicious activities. UCLA Health faced legal and regulatory consequences, including lawsuits and fines and affected patients also had to deal with the potential long-term consequences of their personal information being exposed.

Proposed Method

As outlined in the previous discussion, several challenges must be addressed through Federated Learning(McMahan et al. 2023). One significant issue is the collection, curation, and management of X-ray image data from various collaborating hospitals. This requires a systematic approach to ensure that the data is properly organized and maintained across institutions.

Additionally, there is a need to collaboratively build a lung disease detection model using the gathered data while prioritizing patient privacy. It's essential that the model development process safeguards sensitive information, allowing hospitals to benefit from shared knowledge without compromising patient confidentiality.

Finally, it's crucial to create a secure diagnostic platform that delivers anonymized and explainable results. Such a platform would enhance trust in the system, providing healthcare professionals with insights that they can understand and apply in their decision-making processes. Together, these efforts can effectively leverage Federated Learning to improve lung disease detection while addressing critical privacy and security concerns.

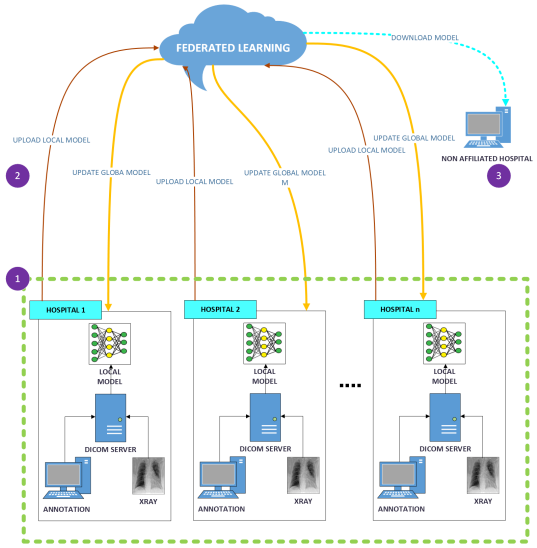

Figure 1: Federated Learning (Alfiansyah et al. 2024) model aggregation cycle

1 Starting with an initial model, each computational node constructs its local neural radiance network using its local data. **2** The local model is then transferred to an aggregation server for fusion with models from other nodes. **3** the updated model is sent back to the local node servers for further retraining by iterating process to step 1.

1. **Building Local Model:** Each hospital involved makes their own deep learning model using the X-ray pictures they collected from their patients. They train these local models using CNN inside the hospital. This way, patient data stays safe inside the hospital and doesn't go out.
2. **Combine and Integrate:** Hospitals send the weights of their models to a central server. This server is like a mixer; it combines the weights from all hospitals to create one big global model. In our implementation, we simply averaged the total CNN weights from each hospital. This model learns from all the hospitals without sharing any sensitive patient information.

3. **Share and Update Global Model:** The centralized global model goes back to each hospital to help start more local training. Hospitals then use this global model to improve their own models with local data and what they learn from the global model. This sharing and updating keeps going until everyone agrees it works well.
4. **Safe and Privacy Preserving Predictions:** Hospitals and authorized third parties that are not directly involved can use the trained model to make predictions on new X-ray pictures. They can give new images to the model and get results, but they cannot access actual patient data, keeping privacy safe and following rules.

To improve the practical use of the proposed platform, we have established a two-layer server system that includes separate local servers in each hospital and a central aggregator server. The **local servers** manage tasks such as storing, organizing, and labeling medical data, which helps in handling the information effectively. The **aggregator server** serves as the main hub that combines the local models from various hospitals in the network. It also stores the combined model, which can be accessed for diagnosis whenever needed. This server setup allows for smooth cooperation between the local servers and the central aggregator, enhancing the overall efficiency of the platform.

Each local node in this platform is an extension of XNAT (Marcus et al. 2007; Alfiansyah et al. 2024), allowing each hospital to curate and manage patient data while conducting local learning. Physically, each node is equipped with a GPU to facilitate efficient learning.

Implemented Federated Learning

Federated learning (FL) is a collaborative approach to machine learning that addresses data privacy and governance concerns. It allows multiple parties to train algorithms together without directly sharing their data. While initially developed for applications like mobile and edge computing, FL has recently gained popularity in healthcare settings.

The key advantage of FL in healthcare is that it enables institutions to collectively develop insights and create a shared model without transferring sensitive patient data beyond their own secure networks. Instead, the machine learning process occurs independently at each participating organization, with only model-related information (such as parameters or gradients) being exchanged.

On our platform, we employ three methods for lung disease screening: chest X-ray (CXR) classification using Densenet, object detection to identify specific lung disease locations with YOLO, and tuberculosis detection utilizing a combination of attention mechanisms and CNN.

At first during the initial phase of our study to develop x-ray classification, we evaluated the performance of several advanced neural network architectures for this tasks using a meticulously curated dataset. The architectures examined included Inception, VGG, ResNet, and DenseNet, each bringing unique strengths to the table. Inception excels at processing information across multiple scales, VGG is recognized for its straightforward yet effective deep learning implementation, ResNet employs residual connections to address the vanishing gradient problem, and DenseNet is notable for its densely connected layers that enhance feature propagation and reuse. Among these, DenseNet achieved the highest F1-score of 91.90, reflecting its superior balance between precision and recall, making it a prime candidate for pneumonia detection (Alfiansyah et al. 2024).

We enhance our capacity to accurately identify subtle patterns associated with pneumonia in medical images. This decision aligns with our strategic goal of integrating distributed data sources while prioritizing data privacy, thereby safeguarding sensitive health information as we advance our diagnostic capabilities. Furthermore, adopting DenseNet showcases our commitment to leveraging state-of-the-art methodologies in medical image analysis, ultimately pushing the envelope of current diagnostic standards.

For second method we implemented in our detection model is YOLO (You Only Look Once) method. The process of aggregating a model involves sophisticated parameter management due to the architecture's inherent complexity, comprising multiple interconnected components such as the *backbone*, *neck*, and *detection head*. Each client independently trains its local YOLO model on its private dataset, updating the model weights via stochastic gradient descent or similar optimization techniques. These local weight updates are then transmitted—typically as tensors representing the model parameters—to the central server. The core aggregation mechanism relies on the Federated Averaging (FedAvg) algorithm (McMahan et al. 2023), which computes a weighted average of these local parameters, where the weights are proportional to each client's dataset size, ensuring that clients with more data have a proportionally greater influence on the global model.

Implementing this in practice requires careful handling of model parameter tensors, often necessitating their flattening into vectors to streamline element-wise operations, or maintaining a structured approach that respects different layer types. Given YOLO's architecture, aggregating layer-wise parameters separately can help mitigate issues inherent to heterogeneity—for example, the backbone, which extracts features, may require different aggregation considerations compared to the detection head responsible for bounding box predictions.

To maintain model stability and improve convergence, several advanced techniques are employed. Layer-wise aggregation can assign different importance weights across layers, especially for layers more sensitive to local data distribution variations. Regularization methods such as FedProx (Sahu et al. 2018) introduce a proximal term to prevent excessive divergence between local updates and the current global model:

$$L_k(w) = f_k(w) + \frac{\mu}{2}||w - w_t||^2 \tag{1}$$

where:

- $f_k(w)$ is the original loss function for client k.
- w is the model parameter.
- w_t is the global model parameter at iteration t.

- μ is a non-negative regularization parameter that controls the impact of the proximal term that encourages local updates to remain close to the current global parameters.

To ensure convergence and robustness, iterative training proceeds over multiple communication rounds, with periodic validation against hold-out datasets. This process involves updating the central global model with aggregated parameters, distributing it back to the clients, and repeating the cycle until satisfactory performance is achieved.

In third method, we aggregate attention mechanisms within a FL framework involves the consolidation of attention representations—such as attention weight matrices or activation maps—across multiple client devices. Unlike standard parameter aggregation, attention aggregation requires a focused approach to preserve the interpretability and spatial relevance of attention maps. During local training, each client computes attention weights or activation maps, typically represented as tensors with dimensions corresponding to spatial features (e.g., ($H \times W \times C$)), which highlight salient regions relevant to the task. These attention outputs are transmitted to the central server either directly or via distillation techniques, where they serve as interpretable features for aggregation.

On the server side, aggregation methods such as element-wise averaging are employed if attention maps are spatially aligned and share the same dimensions; this yields a mean attention map that reflects regions consistently attended across clients. To incorporate varying client data sizes or reliability, weighted averaging schemes can be utilized, assigning importance based on dataset size, data quality, or trust scores:

$$A_{global} = \frac{\sum_{k=1}^{K} n_k \cdot A_k}{\sum_{k=1}^{K} n_k} \tag{2}$$

where A_k denotes the attention map from client (k), and n_k is its dataset size.

In summary, attention aggregation in FL involves collecting spatial attention tensors, applying weighted or unweighted averaging, optionally training dedicated fusion models to generate consensus attention maps, and implementing regularization strategies during local training to enhance cross-client attention alignment. This paradigm addresses the need for interpretability, robustness, and spatial focus preservation in decentralized models tackling tasks such as medical image analysis and object detection.

To ensure convergence and robustness, iterative training proceeds over multiple communication rounds, with periodic validation against hold-out datasets. This process involves updating the central global model with aggregated parameters, distributing it back to the clients, and repeating the cycle until satisfactory performance is achieved.

Local Data Storage

The implementation of this server is important across all participating project-affiliated hospitals, serving a dual role in managing patient data curation within hospital and annotating chest training data. Beyond these functions, the server

Algorithm 1: The κ clients are indexed by k; β is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server side execution

Initialize neural network weight w_0
for each round $t = 1, 2, 3...$ **do**
  $m \leftarrow (C\dot{K}, 1)$
  $S_t \leftarrow$ random set of m clients
  for each client $k \epsilon S_t$ **in parallel do**
    $w_{t+1}^{w_k} \leftarrow$ ClientUpdate (k, w_t^k)
    $m_t \leftarrow \sum_{k \epsilon S} n_k$
    $m_{t+1} \leftarrow \sum_{k \epsilon S} \frac{n_k}{m_t} w_{t+1}^k$
  end for
end for

ClientUpdate$(k.w)$ //Run on client k

$B \leftarrow$ split data P_k into batches of the size B
for each local epoch i form 1 to E **do**
  for each local epoch i form 1 to E **do**
    $w \leftarrow w - \eta \nabla \ell(w; b)$
  end for
end for
return w to server

assumes a critical role in the initial development of a localized pneumonia detection model, which precedes the aggregation phase executed on the federated learning server. To fulfill its multifaceted tasks effectively, the server demands significant computational power. In this project, each server uses an open source database provided by XNAT (Marcus et al. 2010), (Marcus et al. 2007), (Schwartz et al. 2012) to build an image storage database.

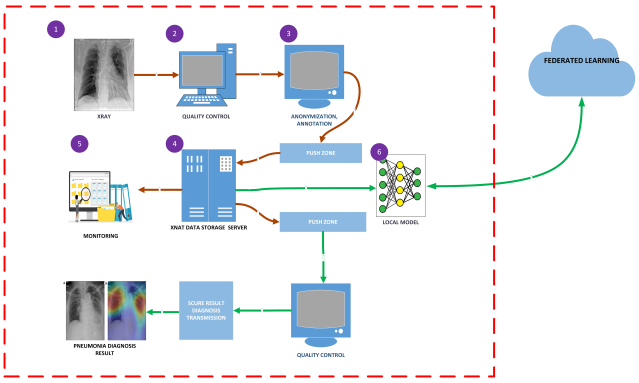

Figure 2: Local database server that is used to store patient data in hospitals, perform anonymization and annotation and build local models to be aggregated with models on other servers

All studies conducted within this framework adhere to a standardized workflow, as illustrated in Figure 1. Prior to commencing data acquisition, each constituent center form-

ing the imaging network (depicted as Figure 1 part 1) undergoes a meticulous setup procedure aimed at optimizing data harmonization within the hospital environment. In the context of pneumonia detection, this procedure is tailored specifically to the chest X-ray imaging modality. The setup entails follow-up visits to ensure sustained harmonization over time. Once all participating centers have completed the initial setup, patient inclusion initiates, and subjects are scanned across the entire network, adhering rigorously to acquisition protocols outlined by medical guidelines.

The processes of data anonymization and secure transfers, denoted as 1 part 3, remain under the jurisdiction of each local hospital. Following data collection, clinical research associates meticulously scrutinize all raw data acquisitions (represented by Figure 1 part 4)) utilizing a dedicated software platform meticulously designed in alignment with this project's protocol. Notably adaptable to new protocols, this platform facilitates protocol consistency checks, parameter comparisons against initial center settings for each study, and flexible conversion of raw DICOM images to alternative medical formats as required. Each sequence within the protocol subsequently undergoes a comprehensive assessment via a documented series of qualitative and quantitative evaluation metrics. These metrics are engineered to characterize various aspects, including acquisition slab positioning, movement anomalies, spikes, and other artifacts. Additionally, they gauge the overall image quality through assessments of contrast, noise levels, intensity uniformity, and other pertinent parameters (depicted in Figure 2).

The successful completion of this initial quality assessment confers authorization for further analysis that at the end utilizing as dataset for local machine learning training sytems. The validated data is then gathered locally and marked as training data to construct AI models for pneumonia screening. The dataset undergoes initial annotation before being placed into a designated push zone (as shown in Figure 1-4), where it is subjected to automatic sanity checks. Subsequently, the dataset is automatically conveyed to a secure directory, serving as the local dataset repository for our federated learning approach across collaborating hospitals, facilitated by our framework. Despite of duplication across multiple storage sites, the training data remains within their respective local storage in hospital (indicated by Figure 1 part 5). The contents of this directory are stored within a local database, enabling local users to query the data effectively.

Privacy Preserving Screening

The final implementation within our framework centers on creating a user-friendly prototype system designed for healthcare professionals. We developed this tool to provide users with a clear understanding of the effectiveness of AI-driven tools for lung-related diagnoses. This system acts as a secondary opinion platform, enabling users to analyze images for confirmation or assistance in their diagnostic decisions. Our solution is presented in the form of a mobile application specifically aimed at predicting pneumonia from chest X-rays, making it both accessible and practical for healthcare providers.

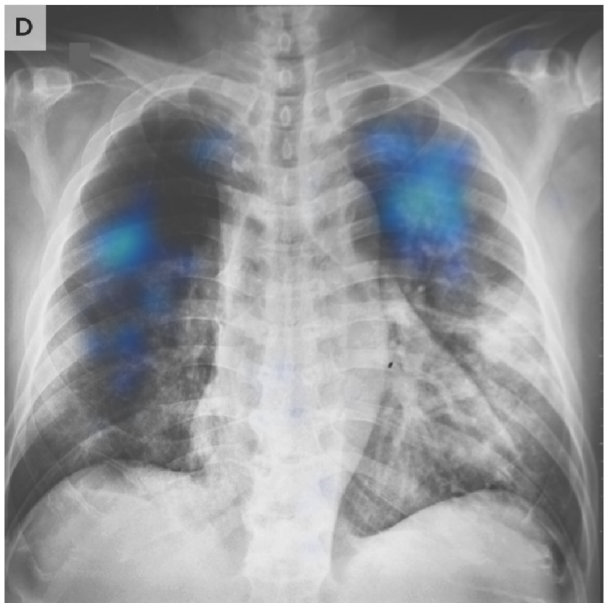

Figure 3: A diagnostic application dashboard capable of providing heatmaps for various lung-related diseases, here is a predictive regions for Mass (can be found at https://xraychest.pnumonai.asia/).

This approach offers several significant benefits. The system is designed to keep patient data entirely on the user's device, ensuring maximum privacy. Furthermore, all data processing is conducted locally, enabling us to improve computational scalability and reduce costs. In comparison, other software deployment methods, such as a desktop application, would involve considerable development effort, which is impractical for a freely available prototype. Rather than sending patient image data to a server, our system adopts a el method. We deploy the pneumonia prediction model directly on the user's device, allowing predictions to be made using only the local model. This architecture eliminates the need to transmit any patient data outside of the device, enhancing privacy and ensuring that no sensitive patient information is shared externally.

Explainability

Providing an explanation for predictions is essential in building trust in the model's accuracy and empowering users to draw their own insights from the tool (as emphasized in section 4). While there is a wide range of techniques available for generating these explanations, we must work within a constrained computational budget. This limitation necessitates the careful selection of methods that effectively balance interpretability and resource efficiency, ensuring that users can gain valuable understanding without placing excessive demands on computational resources.

To explain why a neural network makes certain predictions, gradient saliency maps can be used, as discussed in (Simonyan, Vedaldi, and Zisserman 2014) and (Lo, Cohen, and Ding 2015). These maps are visual tools that pinpoint

which parts of an input, such as an image or text, the network concentrates on during the prediction process. They work like a "heat map," with brighter regions indicating more significant areas for the prediction. To create gradient saliency maps, one calculates the gradient of the model's output (like the predicted class score) concerning the input features. The gradient's magnitude reveals how much a slight alteration in a specific input feature can influence the output.

When given an input image I and the presoftmax output y from the neural network, we can determine the influence of each pixel on a specific output (y_i) or aggregate the effects across all outputs. The computational expense of generating these saliency maps is equivalent to performing a single feedforward pass through the network. The saliency map helps explain the prediction for the task of $\max\left\{0, \frac{\delta_y}{\delta_I}\right\}$. Generally, high gradient values tend to occur in clusters of pixels, highlighting regions that are indicative of the disease. However, one challenge with directly interpreting these gradients is that high gradients appear not only at the exact location of a significant feature, like a nodule, but also in areas that influence the impact of that region, such as the space surrounding a nodule.

However, one challenge in interpreting the gradients directly arises from their behavior at specific locations. High gradient values are not only found at the precise locations of key features—such as a nodule—but also in adjacent areas that influence the network's decision-making process. This phenomenon indicates that while certain pixels may contribute directly to the prediction, surrounding regions can also play a significant role in shaping the outcome, complicating the straightforward interpretation of the saliency maps.

One difficulty in directly interpreting gradients is their presence at particular spots. High gradient values appear not just at the exact locations of important features like nodules, but also in nearby areas that affect the network's decision-making. This means that while some pixels contribute directly to the prediction, the surrounding regions also significantly influence the outcome, making the interpretation of saliency maps less straightforward.

Experiments and result

To asses the performance of the system we developed, we set up local servers in various locations throughout Indonesia, as illustrated in Figure 4 there are 6 hospital contributed in this study. All local servers are located at university hospitals that are geographically distant from one another, with network latencies that are relatively similar. Four of these servers are situated outside of Java, while the other two are on the island of Java. We need an aggregation server for thus Federated Learning approche and the server is installed in the univerity.

We have conducted an initial laboratory simulation study on the use of federated learning for pneumonia screening, along with a comparison to a centralized system, as referenced in (Alfiansyah et al. 2024). This study confirms our previous findings.

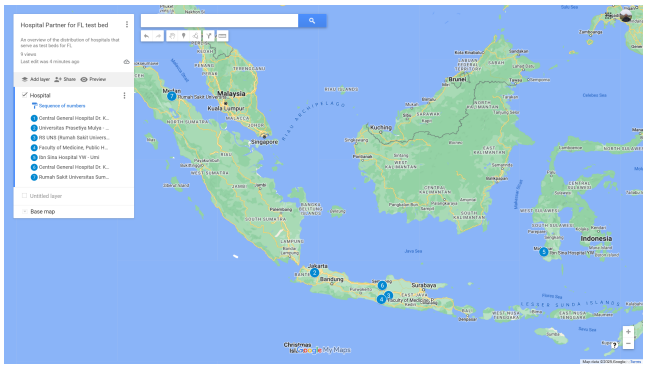

Figure 4: Geographic distribution of hospital host the servers used in the experiments.

The results presented in the table 1 significant achievements in applying FL for lung disease detection across multiple datasets and advanced CNN architectures. The DenseNet model for X-ray classification, trained on both RSNA and Kaggle datasets, achieved a high F1-score of 0.953, indicating a strong synergy between precision and recall, and demonstrating the model's robustness in identifying lung abnormalities with minimal false positives and negatives. This suggests that DenseNet is highly effective in feature extraction and classification in medical imaging, especially under the federated setting where data heterogeneity and privacy constraints are prevalent.

In the case of tuberculosis detection, the combination of attention mechanisms with multilayer perceptrons (MLP) on a multi-national patient cohort yielded an even higher F1-score of 0.960. This reflects the model's capacity to focus on salient regions in the chest X-rays effectively, improving its sensitivity and specificity across diverse populations. The multi-national aspect underscores the model's generalizability, which is critical for real-world deployment across varying demographics and imaging conditions. Attention modules help enhance interpretability and focus the model's capacity on critical features, which likely contributed to the superior performance.

Furthermore, the object detection task utilizing YOLO version 8 on the MIMIC-CXR dataset achieved an F1-score of 0.961, indicating exceptional competence in localizing and detecting lung lesions or abnormalities. YOLO's real-time detection capabilities combined with high accuracy suggest its suitability for clinical workflows requiring rapid, precise identification of pathological regions.

Overall, these results demonstrate that employing advanced CNN architectures in a federated learning framework can effectively leverage diverse and sensitive medical datasets without compromising patient privacy. The high performance metrics across different tasks and datasets affirm the potential of federated learning to enable scalable, privacy-preserving, and robust AI models for lung disease screening in hospitals and clinics. They also suggest that such an approach can adapt well to the heterogeneity inherent in multi-institutional data, ultimately facilitating broader deployment of AI-driven diagnostics in resource-

Task	CNN type	Dataset	Precision	Recall	F1-Score
X-ray classification	Dense-net	RSNA database(Wang et al. 2017) and Kaggle	0.921	0.918	0.953
Tubercollosis 2	Attention+MLP 2	Multi-national patient cohort (Rahman et al. 2020)	0.957	0.922	0.960
Task	**CNN type**	**Dataset**	**Precision**	**Recall**	**F1-Score**
Object Detection 2	YOLO ver 8	MIMIC-CXR Database (Johnson et al. 2024)	0.932	0.923	0.961

Table 1: X-ray screening methods implemented in this hospitals collaboration in Indonesia by means of federated learning

constrained settings like Indonesia.

Discussion

Federated Learning represents a significant shift from traditional centralized data storage methods, making it essential to recognize its effects on the various participants within a Federated Learning ecosystem. During the development of this research, we observed that multiple stakeholders could benefit from the project we developed, as outlined in the following part.

Clinicians Clinicians often encounter a limited subset of the population influenced by their geographic and demographic context, which can lead to skewed perceptions of disease probabilities and their relationships. Implementing machine learning (ML) systems, like a second opinion tool, allows them to enhance their skills with insights from other institutions, promoting diagnostic consistency that is currently hard to achieve. While this benefit applies broadly to ML systems, those trained through federated methods may produce even less biased outcomes and greater sensitivity to uncommon cases, as they have access to a more diverse data set. However, this requires initial efforts to establish agreements on data structure, annotation, and reporting protocols to ensure that all collaborators understand the information in a unified way.

Patients Typically, patients receive treatment at local facilities. Implementing Federated Learning (FL) on a global scale could enhance the quality of clinical decision-making, irrespective of where the treatment occurs. This approach would particularly benefit individuals in remote locations by providing them access to high-quality, machine-learning-assisted diagnoses similar to those offered in larger hospitals with extensive case histories. The same applies to rare or geographically specific diseases, where quicker and more precise diagnostics can lead to better outcomes. Additionally, FL could make it easier for patients to share their data, as they would have the reassurance that their information stays within their own healthcare institution and that they can withdraw access at any time.

Hospitals and practices Hospitals and medical practices can maintain complete control over their patient data, ensuring traceability of data access and reducing the risk of misuse by external parties. However, achieving this requires investment in on-site computing infrastructure or private cloud services, along with adherence to standardized data formats to facilitate seamless training and evaluation of machine learning (ML) models. The computing requirements will vary depending on whether a facility is involved solely in evaluation and testing or also in the training process. Even smaller institutions can participate and still reap the benefits of the collaborative models developed.

Researchers and AI developers Researchers and AI developers can gain access to a potentially extensive repository of real-world data, which is especially advantageous for smaller research labs and startups. This allows them to focus their resources on addressing clinical needs and related technical challenges rather than depending on the limited availability of open datasets. However, it will be essential to explore algorithmic strategies for federated training, such as efficient model combination and updates, as well as ensuring robustness to distribution shifts. Additionally, working within a federated learning framework means that researchers or AI developers cannot fully examine or visualize the complete dataset used for training the model, which limits their ability to analyze individual failure cases to understand why the model may not perform well in certain situations.

Healthcare providers In many countries, healthcare providers are experiencing a significant shift from volume-based care, where payment is tied to service quantity, to value-based care, closely linked to the advancement of precision medicine. This transition aims not to advocate for more costly individualized treatments but to achieve improved patient outcomes more efficiently through targeted therapies, ultimately lowering costs. Federated Learning (FL) has the potential to enhance the accuracy and reliability of AI in healthcare, while simultaneously reducing expenses and improving patient care, making it an essential component of precision medicine.

Manufacturers Healthcare software and hardware manufacturers can also benefit from FL, as it allows for the integration of learning from numerous devices and applications while keeping patient-specific data confidential. This approach can support the ongoing validation and enhancement of their machine learning systems. However, achieving this capability may necessitate substantial upgrades to local computing resources, data storage, networking infrastructure, and related software systems.

Conclusion

In conclusion, this study demonstrates the effectiveness of federated learning in advancing medical diagnostics, particularly for pneumonia detection in Indonesia. By facilitating collaborative machine learning across multiple insti-

tutions and prioritizing patient data privacy through decentralized data storage and shared model updates, the research addresses crucial ethical and technical challenges in healthcare AI. The development of a multi-institutional collaborative platform, enhanced by federated learning and incorporating data anonymization and explainability techniques, sets a new standard for responsible AI in healthcare. This approach enhances the accuracy and generalization of diagnostic models, providing clinicians with valuable insights and contributing to improved patient outcomes and a more equitable healthcare landscape. Federated learning signifies a significant shift from traditional centralized data methods, benefitting clinicians, patients, hospitals, practices, researchers, and AI developers alike.

Acknowledgment

This project is supported by ISIF Asia Grant, Application ID 202204-00759.

References

Albahli, S.; Rauf, H. T.; Algosaibi, A. A.; and Balas, V. E. 2021. AI-driven deep CNN approach for multi-label pathology classification using chest X-Rays. *PeerJ Computer Science*, 7: 1–17.

Alder, S. 2015. UCLA Health System Hacked: 4.5 Million Patient Records Exposed. https://www.hipaajournal.com/ucla-health-system-hacked-4-5-million-patient-records-exposed-8033/. [Accessed 29-03-2025].

Alfiansyah, A.; Widiarti, H.; Reynaldo, V.; Widjaja, W.; and Bobelin, L. 2024. A Framework for Multi Institutional Collaboration for Pneumonia Screening by Means of Federated Learning. In *Electronics, Communications and Networks*, 135–142. IOS Press.

Dalhoumi, S.; Dray, G.; Montmain, J.; Derosière, G.; and Perrey, S. 2015. An adaptive accuracy-weighted ensemble for inter-subjects classification in brain-computer interfacing. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 126–129.

Johnson, A.; Pollard, T.; Mark, R.; Berkowitz, S.; and Horng, S. 2024. MIMIC-CXR Database.

Lo, H. Z.; Cohen, J. P.; and Ding, W. 2015. Prediction gradients for feature extraction and analysis from convolutional neural networks. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, 1–6.

Marcus, D. S.; Fotenos, A. F.; Csernansky, J. G.; Morris, J. C.; and Buckner, R. L. 2010. Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*, 22(12): 2677–2684.

Marcus, D. S.; Olsen, T. R.; Ramaratnam, M.; and Buckner, R. L. 2007. The extensible neuroimaging archive toolkit. *Neuroinformatics*, 5(1): 11–33.

McMahan, H. B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2023. Communication-Efficient Learning of Deep Networks from Decentralized Data. arXiv:1602.05629.

Narayanan, A.; and Shmatikov, V. 2006. How To Break Anonymity of the Netflix Prize Dataset. *ArXiv*, abs/cs/0610105.

Narayanan, A.; and Shmatikov, V. 2008. Robust Deanonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 111–125.

of Public Health, F. 2024. Peta Sebaran Tenaga Kesehatan Indonesia. https://pkmk-ugm.shinyapps.io/sdmkesehatan/. [Accessed 29-01-2025].

Rahman, T.; Khandakar, A.; Kadir, M. A.; Islam, K. R.; Islam, K. F.; Mazhar, R.; Hamid, T.; Islam, M. T.; Kashem, S.; Mahbub, Z. B.; Ayari, M. A.; and Chowdhury, M. E. H. 2020. Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization. *IEEE Access*, 8: 191586–191601.

Rauf, H. T.; Lali, M. I. U.; Khan, M. A.; Kadry, S.; Alolaiyan, H.; Razaq, A.; and Irfan, R. 2023. Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks. *Personal and Ubiquitous Computing*, 27(3): 733–750.

Reddy, J.; Elsayed, N.; ElSayed, Z.; and Ozer, M. 2022. Data Breaches in Healthcare Security Systems. arXiv:2111.00582.

Sahu, A. K.; Li, T.; Sanjabi, M.; Zaheer, M.; Talwalkar, A.; and Smith, V. 2018. On the Convergence of Federated Optimization in Heterogeneous Networks. *CoRR*, abs/1812.06127.

Schwartz, Y.; Barbot, A.; Thyreau, B.; Frouin, V.; Varoquaux, G.; Siram, A.; Marcus, D.; and Poline, J. 2012. PyXNAT: XNAT in Python. *Frontiers in Neuroinformatics*, 6(MARCH).

Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2014. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv:1312.6034.

UNICEF. 2020. Every child's right to survive. https://data.unicef.org/resources/every-childs-right-to-survive-an-agenda-to-end-pneumonia-deaths/. [Accessed 29-01-2025].

Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; and Summers, R. M. 2017. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3462–3471.