ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2026, 11 (3): 549-577 doi: 10.28978/nesciences.263045

Federated Bayesian Convolutional Neural Network for Secure Respiratory Diagnosis in Biomedical Imaging

Fuad S. Al-Duais [1*], Refga Abdallah Eltieb Elebeed [2], Rasha M. Abd El-Aziz [3], Ahmed I. Taloba [4]

[1*] Department of Mathematics, College of Science and Humanities, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. ORCID: https://orcid.org/0000-0002-2255-1167, E-mail: f.alduais@psau.edu.sa

[2] Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. ORCID: https://orcid.org/0009-0008-8204-9196, E-mail: r.elebeed@psau.edu.sa

[3] Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia; Department of Computer Science, Faculty of Computers and Information, Assiut University, Assiut, Egypt. ORCID: https://orcid.org/0000-0002-8975-6052, E-mail: rmhassanien@ju.edu.sa

[4] Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia; Department of Information System, Faculty of Computers and Information, Assiut University, Assiut, Egypt. ORCID: https://orcid.org/0000-0003-3558-423X, E-mail: aitaloba@ju.edu.sa

Abstract

Respiratory diseases such as Pneumonia, Pulmonary Fibrosis, and Pleural Effusion remain leading causes of illness worldwide, yet diagnosis still depends heavily on radiologist interpretation, which is time-consuming and prone to inter-observer disagreement, especially in resource-limited settings. Centralized deep learning models can automate this task but require pooling sensitive patient scans in one location, raising serious privacy and regulatory concerns for multi-hospital deployment. This study proposes a Federated Bayesian Convolutional Neural Network that extracts disease-specific radiographic features through convolutional layers and produces calibrated, uncertainty-aware predictions through a Bayesian layer using Monte Carlo dropout, while federated averaging allows multiple institutions to jointly train the model by sharing only weights, never raw images. The framework is evaluated on the VinDr-CXR Chest X-ray Abnormalities Detection dataset, covering four diagnostic categories including Normal, Pneumonia, Pulmonary Fibrosis, and Pleural Effusion, collected from two independent hospitals. The proposed framework achieves 96.4% accuracy, an AUC of 0.982, and a Brier score of 0.029, outperforming centralized CNN, standard federated CNN, and standalone Bayesian CNN baselines by 2.6, 2.4, and 1.5 percentage points respectively, and exceeding several published baseline models on the same task. The main novelty is the joint integration of privacy-preserving federated learning with clinically meaningful uncertainty quantification in a single pipeline, rather than treating these as separate design goals. This combination supports trustworthy, scalable, and interpretable respiratory disease screening across multiple

*Corresponding Author: Fuad S. Al-Duais, E-mail: f.alduais@psau.edu.sa

healthcare institutions, offering a practical pathway toward privacy-compliant clinical decision support in resource-varied hospital settings and multi-institutional diagnostic networks.

Keywords:
Federated learning, bayesian deep learning, uncertainty quantification, chest x-ray classification, privacy-preserving diagnosis.

Available Online: 01/09/2026

Introduction

Respiratory diseases, including pneumonia, pulmonary fibrosis, and pleural effusion, remain major contributors to global morbidity and mortality, posing a substantial burden on healthcare systems worldwide (Labaki & Han, 2020). The prior and accurate classification results in the better outcomes of the patient and the lessen the hospital burden (Cid et al., 2024). Chest X-rays have become the first-choice imaging modalities for the evaluation of respiratory disorders as they offer easy access, quickness, and low-cost service (Jones et al., 2021). However, X- ray image reading is very much reliant on the skill of radiologists and is often prone to differences in opinion among the analysts (Tonks et al., 2024). Furthermore, the rising amount of radiographic data produced in hospitals is a hurdle for timely and consistent diagnosis (Rajpurkar et al., 2018). Artificial intelligence (AI), particularly deep learning, has emerged as a promising technology for rapid and accurate interpretation of complex medical images (Nazir et al., 2025). CNNs (Convolutional Neural Networks) have been successful in identifying the complicated features of radiographs like opacity, fibrosis, and pleural fluid and hence automated classification of lung diseases has become possible (Majkowska et al., 2020). Nevertheless, the traditional deep learning methods still need centralized data, which means severe issues would arise regarding patient privacy, data security, and compliance with regulations (Fang et al., 2024). The necessity of creating models that can learn efficiently from decentralized data sources without violating confidentiality has led to the consideration of federated learning frameworks in medical imaging (Guan et al., 2024). The latter enables a number of hospitals to jointly train a single model while keeping the data pertaining to their patients at their locations, thus providing a solution to both the issues of data privacy and the poblem of small labeled datasets at each site. Numerous investigations have utilized deep learning techniques, especially CNNs and hybrid models, in tackling chest X-ray datasets for the detection of lung diseases (Nadal-Martínez et al., 2025). Typical CNN-based systems are very good in considering the spatial aspects and per- forming impressively on the datasets that are made centrally available (Parra-Cabrera et al., 2026). Recent research has increasingly focused on enhancing the robustness, reliability, and interpretability of deep learning models for medical image analysis, particularly through uncertainty-aware learning strategies (Irvin et al., 2019). Moreover, Bayesian deep learning techniques have been used in providing uncertainty estimation which is a prerequisite for the clinical decision-making process (Tyralis & Papacharalampous, 2024). Notwithstanding the above-mentioned advancements, existing methods are still limited in many ways. The primary issue is that most of the studies are done on centralized datasets which are usually impossible to access because of government regulations, patient confidentiality, and the prohibitive cost of sharing data (Coutinho-Almeida et al., 2024). Besides this, while a lot of models are being trained on one dataset, their capability to adapt to and be used with different populations and imaging protocols is greatly reduced. Uncertainty quantifying models can produce overconfident predictions, which could lead to the compromising of clinical safety (Blattmann et al., 2025). In addition, the combination of CNNs with recurrent networks like BiGRU in hybrid architectures, which extends their effectiveness, is a drag on computational power and time, thus making it difficult for hospitals to deploy such systems in real-time. All these limitations cumulatively underscore the demand for an AI framework that is privacy-preserving, interpretable, multi-source learning and real-time clinical applications competent.

The present research suggests a federated Bayesian deep learning framework as a solution to the identified gaps for the earliest detection of respiratory diseases with the help of X-ray images from different datasets (Zeng et al., 2025). In particular, the research connects the CNN-based feature extraction with the Bayesian prediction layer to give the probabilistic outputs of Pneumonia, Pulmonary Fibrosis, and Pleural Effusion together with the estimates of uncertainty. Using the federated learning framework, it becomes possible to

train the local models using the distributed VinDr-CXR dataset without any transfer of patient data to the central location. Only the model weights are shared with the central server where secure aggregation by FedAvg updates the global model while protecting patient confidentiality and ensuring compliance with the regulations (Rieke et al., 2020). The Bayesian layer guarantees that the pre- dictions come with confidence scores which enable the clinicians to judge the uncertainty in the automated diagnoses (Lindenmeyer et al., 2025). Furthermore, the system proposed tackles the heterogeneity in the imaging protocols and the patient demographics by learning from the multi-source datasets which promotes the model's generalization across the different populations. The architecture is planned to be scalable, interpretable, and possibly implemented in the real-life clinical scenario, therefore, the architecture offers a safe and privacy-sensitive respiratory disease detection system. This paper hypothesizes a simplified federated setup with few clients to test cross-dataset generalization in a controlled setting. In contrast to current methods which consider federated learning and Bayesian inference separately, this research allows inference of uncertainty on prediction in decentralized training on heterogeneous datasets. The combination of federated learning, CNN feature extraction, and Bayesian probabilistic prediction not only progresses the domain of AI-assisted medical imaging but also sets the stage for trustworthy, collaborative, and privacy-minded diagnostic systems.

Research Problem

Respiratory illnesses including Pneumonia, Pulmonary Fibrosis, and Pleural Effusion are serious threats to global health that need quick and precise diagnosis for the right treatment (Soriano et al., 2020). The traditional approach to diagnosis mainly depends on the skill of the radiologist, which is a procedure that takes a lot of time and is affected by the variability of different observers (Brady, 2017). Some deep learning techniques, with CNNs in particular, have been very effective in the automatic interpretation of chest X-rays, but still, the data used for training is mostly centralized and thus corresponding radiologists will predominantly be experts in those hospitals. Centralized methods raise huge privacy issues and also tend to age very quickly, becoming obsolete over time due to the vast differences in anatomy and the diversity of locations and techniques used for imaging (Willemink et al., 2020). Besides, it is said that uncertainty is not considered by most models, which is essential in making sure that clinical decisions are safe. The need for privacy-preserving, interpretable, and generalizable AI solutions has been a hindrance to the automatic diagnosis systems' practical implementation and operation. Multi- source learning remains largely unexploited, even though there are large databases like VinDr-CXR. The problem of having data privacy, multi-dataset generalization, and predictive confidence all together has not yet been addressed. This study suggests a federated Bayesian CNN framework for the early prediction of respiratory diseases in order to deal with foregoing problems.

Research Objectives

- To design and develop a federated Bayesian deep learning framework using CNN architecture for early prediction of respiratory diseases through multi-source chest X-ray data integration.
- To ensure privacy-preserving, decentralized model training across the distributed VinDr-CXR dataset, enabling secure collaborative learning without compromising patient confidentiality.
- To enhance diagnostic reliability and interpretability by generating probabilistic pre- dictions with calibrated uncertainty estimates, thereby supporting trustworthy clinical decision-making in respiratory disease analysis.

Research Questions

- How accurately can a multi-source chest X-ray model identify early-stage respiratory diseases such as Pneumonia, Pulmonary Fibrosis, and Pleural Effusion?
- How can privacy-preserving, decentralized learning be achieved across different medical institutions without compromising diagnostic performance?
- How does probabilistic prediction with uncertainty estimation improve the clinical reliability and interpretability of automated respiratory disease diagnosis?

Research Significance

Respiratory diseases like Pneumonia, Pulmonary Fibrosis, and Pleural Effusion are very important subjects for the study as it seeks to solve merely one of the main problems in the early detection of these diseases. The method based on AI that does not compromise the privacy of the patients does this and at the same time, it is interpretable. The use of federated learning together with a Bayesian CNN allows the cooperation of multiple

sources in training the AI model with different datasets while no sensitive patient data is shared, thus complying with the regulations and keeping the patients' data confidential. The system gives probabilistic predictions along with uncertainty estimates, thereby, making clinical decision-making and patient safety better. The framework not only increases the ability to generalize across different datasets but also reduces the need for expert radiologists and supports large-scale, real-time deployment, thus providing a strong solution for the early detection of respiratory diseases through automated, accurate, and trustworthy methods in clinical settings.

Key Contributions

- Introduces a novel federated Bayesian CNN uniting privacy preservation and calibrated uncertainty for reliable respiratory diagnosis.
- Federated averaging aggregates only model weights across hospitals, keeping raw chest X-ray patient data private.
- Combines convolutional feature extraction with Monte Carlo dropout sampling to quantify epistemic and aleatoric prediction uncertainty.
- Ablation across nine configurations isolates pre-processing, segmentation, label harmonization, and dropout contributions to overall performance.
- Achieves 96.4% accuracy, 0.982 AUC, and 0.029 Brier score, clearly exceeding centralized and federated baselines.

Rest of the Section

Section 2 as Literature review, Section 3-5 as the Proposed Methodology, Section 6 as the Results and the Discussion section, and Finally Section 7 as conclusion and future works.

Literature Review

Teo et al., (2024) provide an overview of the transformative perspective of Federated Learning (FL) when it comes to the further development of digital health without violating patient privacy. The authors speak about the possibility of using FL to facilitate the development of models through collaboration among medical institutions by sharing the model parameters rather than sharing the sensitive data in order to adhere to data protection regulations. The paper presents the ability to improve model generalization by using a variety of decentralized datasets, which in turn allow FL to be used due to the availability of case studies and conceptual analysis. Additional critical technical challenges that are also pointed out in the study are non-independent and identically distributed (non-IID) data, communication with inefficiencies, and possible data leakage risks due to gradient sharing. Although the authors recognize that FL can face challenges on a larger scale, even though it has limitations, including scalability concerns, the authors note that it lacks standardization and requires more powerful privacy-preserving tools such as differential privacy and secure aggregation. In general, the work can serve as a guideline in further studies of privacy-conscious collaborative learning in healthcare.

The systematic Study by Adnan et al., (2022) was aimed at analyzing the present situation in Federated Learning (FL) implementation in healthcare with a focus on its application strategies, privacy-preserving schemes, and current challenges. The review summarized the results of a wide spectrum of works in different fields namely radiology, pathology and genomics and the manner in which FL facilitates decentralized models training and maintains data confidentiality. The authors compared privacy methods such as Differential Privacy (DP), Secure Multiparty Computation (SMC), and Secure Aggregation and evaluated their effectiveness in the area of reducing the risk of information leakage. One of the greatest contributions of the study is that it compares model architectures and optimization strategies of medical FL systems. Nevertheless, Teo et al. observed that there were still some constant limitations as the model convergence to non-IID data, communication inefficiency, and benchmark datasets to compare performance. The paper will end by recommending hybrid FL systems, which combine uncertainty modeling and adaptive optimization to improve the robustness and scalability of healthcare-related systems in reality. The current studies, however, mainly concentrate on federated learning models without incorporating uncertainty-aware prediction, which makes them not as reliable as high-stakes clinical decision-making situations.

Alshmrani et al., (2023) describes a CNN-based model to classify different types of thoracic dis- eases detected on chest X-rays into multiple classes. The model tackles data heterogeneity by merging several public NBC datasets and applying transfer learning techniques with state-of-the-art CNN architectures to

secure reliable multi-label classification. Very importantly for your research, the authors present the preprocessing methods (lung ROI cropping, contrast enhancement), the techniques for balancing class distribution, and the combined/augmented strategies that have improved generalization across datasets practical problems when architecting federated systems that need to cope with non-IID client data. The research further provides explainability visualizations (Grad-CAM) to confirm the model's attention to the pulmonary areas, which relates to model trust and uncertainty assessment.

Rayan & Alaerjan, (2023) combine the meta- heuristic optimization and deep sequence models to improve the chest X-ray classification for COVID-19 detection. The research paper presents the coupling of an enhanced crow search algorithm for hyperparameter/feature selection with a Bi- LSTM classifier (using CNN-extracted features). Its importance for federated/Bayesian research is mainly in the support of robust feature selection under noisy/imbalanced CXR data and in the method to stabilize model training — both being critical when clients have small local data and when communicating model updates in federated learning. In addition, the article emphasizes the reproducible preprocessing pipelines and the ablation studies that assist in determining which components have the greatest impact on performance under cross-dataset evaluation.

Hussain et al., (2023) offers a deep-ensemble approach for the recognition of COVID-19, pneumonia, and regular CXR images. The ensemble integrates different CNN backbones and applies systematic ensembling to increase robustness against changes in the dataset. The paper is beneficial for a federated Bayesian CNN as it demonstrates the positive impact of ensembling on reducing overconfidence and improving calibration— these are directly addressed by Bayesian methods. The authors also conduct thorough experiments on cross-dataset generalization and present explainability analyses as a means to support clinical relevance. The paper has an open-access policy and it discloses the implementation details that can guide the selection of local client models, ensemble aggregation methods, and uncertainty estimation in federated contexts. Ensemble methods are computationally expensive and not directly applicable to the decentralized federated setting with communication limitations that is commonly found in a decentralized system; however, they enhance robustness.

Pan et al., (2024) The paper talks about the use of federated learning in pediatric pneumonia with chest X-rays (CXR) as data from different medical centers. It introduces changes to the FL update rules to better deal with the so-called heterogeneous (non-IID) CXR data distribution among the participating centers, and shows very good accuracy and fairness metrics along with local data privacy preservation. This paper fits perfectly with the topic of your research work because it is addressing the real-world issues with federated learning — the differences between clients (client heterogeneity) communication efficiency, and validation recycling on multi-center clinical CXR datasets — and it also provides experiments showing the impact of the different algorithm choices on calibration and generalization, which are particularly important when combining Bayesian uncertainty estimation with FL. However, these methods lack the use of probabilistic uncertainty estimation, which is needed to detect the cases of ambiguity in medical diagnosis.

Nur et al., (2024) The paper puts forward a separable-convolution model specifically designed for pneumonia detection in chest X-rays (CXR), highlighting the aspects of parameter reduction and better localization. The author claims that separable convolutions, which are light-weighted, ensure the same or even higher accuracy while cutting down the size of the model — which is a significant factor for the federated clients who have restricted computing power or people who require edge deployment. Besides, the authors mention training tactics (like augmentation and weighted losses) that help to eliminate the imbalance and provide evaluation over several public CXR datasets. This research for a Federated Bayesian CNN recommends local training model architectures that are not just compact but also suitable for Bayesian approximations (for instance, Monte Carlo dropout or variational layers) employed for uncertainty estimation so that these architectures can be both compact and conducive to Bayesian approximations (e.g., Monte Carlo dropout or variational layers).

Kumar et al., (2023) The researchers have created ExLNet, an extremely lightweight convolutional neural network that is developed for chest X-ray classification and is focused on hardware deployment with limited resources without sacrificing the performance of the diagnostics getting done. The connection of this work with your project is two ways: (1) small models lessen the local computation/communication costs in federated settings; (2) tiny architectures are perfectly compatible with Bayesian approximations (for instance, Bayesian last layers or MC-Dropout) making it possible to generate calibrated predictive uncertainties without incurring too much computational overhead. The authors of the paper have also conducted ablation studies comparing the computational footprint with accuracy and have provided code/weights so that client

models can be prototyped quickly. These models enhance efficiency but do not have means of quantifying uncertainty and do not tackle federated multi-source data heterogeneity.

Wajgi et al., (2024) The research work aims at the detection of Tuberculosis (TB) through chest X-rays (CXRs) and studies the methods of preprocessing, segmentation, and the use of specialized Convolutional Neural Network (CNN) ensemble to improve the sensitivity and specificity. The detection of TB has a lot of technical challenges in common with other respiratory diagnoses (like very slight opacities, the need for proper localization), thus the paper's handling of lesion localization and the authors' analysis of false positives/negatives give practical guidelines for Bayesian uncertainty modeling, such as how to deal with ambiguous cases and when a model should rely on a human being's judgment (Sindhu, 2026). Besides this, the research invites cross-site evaluation which is useful when thinking of federated aggregation of data from different hospitals with different disease prevalence.

Ryu et al., (2023) This article is mainly concerned with sarcopenia instead of infections, but it still showcases some rigorous methods for explainable AI by using calibration and external validation on CXR images. The paper's focus on establishing clinical validity, model calibration, and saliency/Grad-CAM region-level attributions appeals directly to respiratory disorders: reliable uncertainty estimation and interpretability are vital for the clinician's faith. The paper is freely accessible and demonstrates how to combine model results with clinical decision thresholds downstream, which is a key design feature for a federated Bayesian diagnostic pipeline (Sadulla, 2026). While these studies address interpretability and calibration, they are limited to centralized settings and do not consider privacy-preserving distributed learning.

Ali et al., (2023) proposed CNN model that is capable of detecting general chest infections in CXR images and goes through the preprocessing, data augmentation, and the cross-dataset evaluation thoroughly. It highlights the issues that are likely to occur with dataset leakage and confounders (portable vs. upright X-rays) and suggests the robust evaluation protocols that can be adopted while the material is already very useful when setting federated experiments to ensure no inflation in performance due to shared confounders. The paper is loaded with useful snippets of code and the detailed performance distribution by the disease subgroup which comes in handy when taking the decision of whether to model multi-label outputs (like pneumonia vs. pleural effusion) and of how to focus the Bayesian uncertainty to detect out-of-distribution or ambiguous cases (Kavitha, 2026).

Chetoui et al., (2021) presents a novel explain- able ensemble CNN approach for the detection of COVID-19 from CXRs and analyzes the uncertainty/calibration through the means of ensemble variance and visualization. It shows the bene- fit of ensemble or Bayesian methods (e.g., MC- Dropout) for the confidence estimation and gives a hint about how to mix Bayesian layers with federated aggregation to keep calibrated uncertainties after model averaging. The study also underlines the evaluation metrics that are positioned in the clinical context (ROC/AUC, sensitivity at high specificity) and the necessity of testing on external hospital data — fundamental prerequisites for any trustworthy federated diagnostic system.

Wajgi et al., (2024) in a systematic way, it narrates the progress of federated learning (FL) in radiology, depicting the different types of architectures, privacy apparatuses (secure aggregation and differential privacy), and obstacles to clinical use. It illustrates the common reasons that make FL fail in medical imaging, such as non- IID data, label variations, and uneven distribution of clients, and discusses the algorithmic solutions (personalized FL, robust aggregation, client selection) for these problems. For a Federated Bayesian CNN, the paper is essential because it provides the discussion of how FL can be used for uncertainty quantification to detect unreliable client updates, and thus support selective aggregation or deferral to clinicians. It also discusses the evaluation best practices (external validation and calibration metrics) that are vital for trustworthy diagnostic models. Moreover, the practical recommendations for compute/communication tradeoffs can assist in the designing of lightweight Bayesian approximations for edge clients.

Shin et al., (2023) delivers a meticulously curated multi-center CXR dataset and assesses the performance of automated screening algorithms in cohorts from different countries. The study addresses the issue of generalization gaps introduced by variations in the imaging protocol and changes in the population; it thus tests domain- robustness methods and provides calibration and subgroup performance results. This work is beneficial for your project in that it measures the impact of dataset shifts on model confidence and error modes — thus giving rise to the need for Bayesian uncertainty estimates and federated protocols that are equipped to deal with non-IID clients. The external validation protocol and publicly available dataset of the study are useful for federated experiments in which client-level heterogeneity must be realistically simulated as they are practical resources.

El-Ghany et al., (2024) Suggests a new and better DenseNet backbone designed specifically for TB detection in chest X-rays and also highlights the interpretability of the results through the use of Grad-CAM and lesion segmentation overlays. The researchers present results from cross-site validation and a variety of experiments on the influence of preprocessing and augmentation methods, showing how model calibration and false-positive patterns change with different datasets. The article describes the Federated Bayesian CNN work in terms of model architecture choices (e.g., dense connections) and explainability pipelines' inter- action with uncertainty outputs – for instance, regions with low predictive confidence usually overlap with minor TB signs. Its public availability of the trained weights and code makes it a valuable reference point not only for federated aggregation trials but also for evaluating Bayesian last-layer or MC-Dropout methods on actual clinical tasks.

Siddiqi & Javaid, (2024) This research assesses various CNN and hybrid transformer models for pneumonia detection, focusing on meaningful clinical metrics (sensitivity at high specificity, decision thresholds) and a meticulous error analysis. It opens up problems like dataset leakage and confounders (portable vs upright films) and suggests strong cross-validation and external testing protocols. For your federated Bayesian CNN, the paper's insight into clinically relevant evaluation and thresholding is key: Bayesian uncertainty helps identify cases near operating thresholds where a model should defer. The work's ablations on augmentation, class weighting, and ensembling also provide practical recipes for improving calibration and out-of-distribution detection in distributed training.

Bressem et al., (2020) provides a comprehensive discussion revolving around numerous recent CNN and transformer-based techniques evaluated on a unified public chest X-ray benchmark, along with open-source code and normalized preprocessing (Nguyen et al., 2022). The authors delve into aspects like calibration, how the models are robust against domain shifts, and the role of architectural decisions in causing uncertainty and providing interpretability. So, for your work on Federated Bayesian CNNs, this paper presents a good reference for understanding the baseline performance expectations and for finding the necessary backbones that would provide a good trade-off between accuracy, calibration, and computational cost on the edge devices. The dataset harmonization along with the proposed evaluation methodologies are of immense significance when creating federated simulations that intend to showcase the actual clinical discrepancy.

Proposed Federated Bayesian Framework for Respiratory Disease Diagnosis

The proposed framework introduces several main component enhancements to resolve challenges associated with the early detection of respiratory diseases (especially following the layering of multisource chest X-ray data). A secure and privacy-preserving framework for federated learning is established in this study. The core of the systems framework, a Bayesian Convolutional Neural Network (CNN), which uses deep convolutional layers to extract localized and global radiographic features, as well as the patterns of disease unique to Pneumonia, Pulmonary Fibro- sis, and Pleural Effusion, is utilized. The Bayesian prediction layer generates uncertainty information which are expressed in probabilistic terms regarding output, facilitating enhanced interpretability and clinical decision-making. The federated learning approach makes it possible for the network training process to be performed using the datasets, where the model parameters are formed and optimized using the algorithmic approach for the database entities of VinDr-CXR without any need for exchange of raw data. Federated aggregation allows for local updates of the model's parameters to merge to produce enhanced training and improved generalization of the ultimate model. Advanced pre-processing techniques such as image normalization, lung segmentation, and labelling harmonization enhance the generalizability and quality of the dataset. Feature extraction techniques integrate the features that are relevant to the disease state to enhance the sensitivity to minor radio- graphic abnormalities. Hyper-parameter choices, and adaptable optimization of the data train allow robust performance of the model through the heterogeneous datasets. These components allow for a scalable, interpretable, and secure systems for the multi-source prediction from before of early detection of respiratory disease.

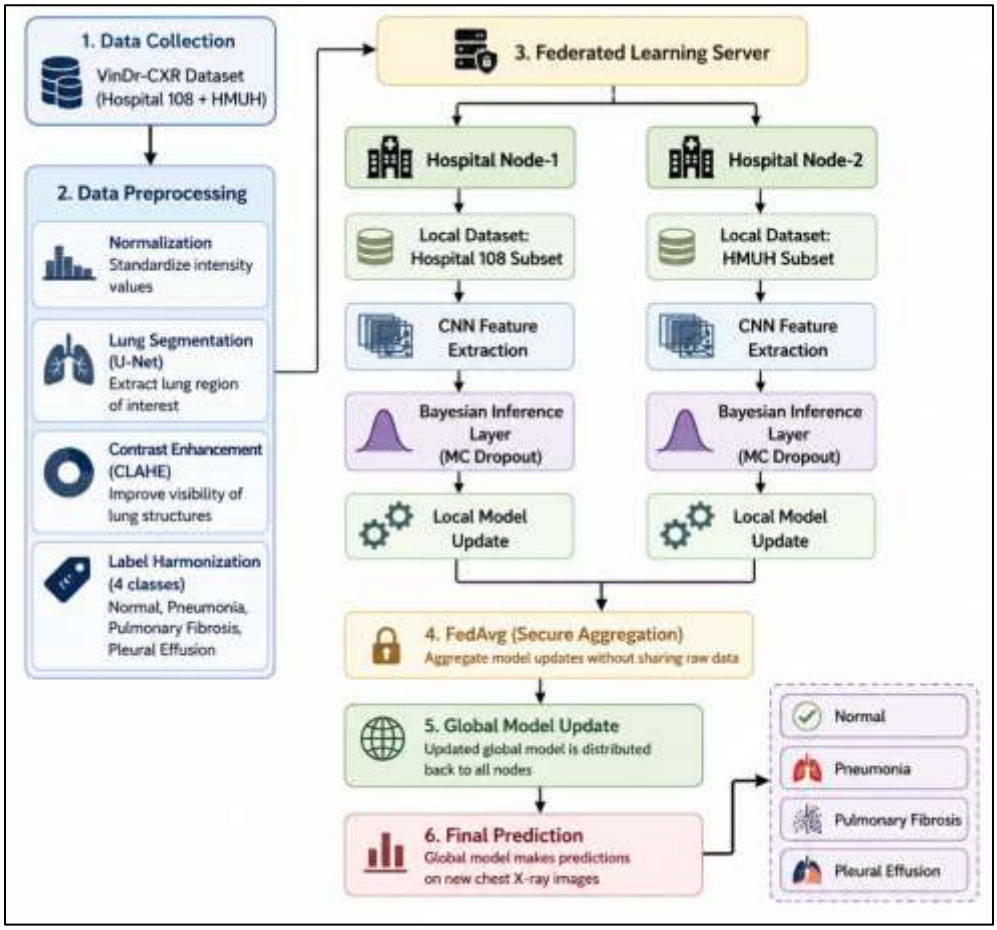

Figure 1. Federated bayesian CNN workflow for multi-hospital chest diagnosis

Figure 1 demonstrates the Federated Bayesian CNN framework designed using VinDr-CXR images at Hospital 108 and Hanoi Medical University Hospital. Both sites first process images through normalization, segmentation, contrast adjustment, and labeling alignment, before feature extraction using CNNs and uncertainty quantification using Bayesian approach. The aggregated updates through FedAvg give rise to one unified model predicting the four classes of respiratory diseases.

Dataset Collection

VinBigData Chest X-ray Abnormalities Detection dataset Labaki & Han, (2020) is a large-scale public medical imaging dataset that was created to support automatic thoracic abnormality detection and classification by the Vingroup Big Data Institute. It includes 18,000 anonymized postero-anterior (PA) chest X-ray images whose reports were labeled by subject-matter experts for 14 thoracic conditions and a No Finding class. The dataset is divided into 15,000 training images (83.33%) and 3,000 testing images (16.67%). Every image contains the disease label and bounding-box annotations for disease localization, which can be used for disease detection and classification. The data was gathered from hospital 108 and Hanoi Medical University Hospital (Hanoi, Vietnam). The evaluation is based on PASCAL VOC 2010 mean Average Precision (mAP), with IoU > 0.4, which ensures solid diagnostic performance assessment.

Data Preprocessing

Data preprocessing is an essential part of this study that ensures that the images from the VinDr-CXR database will be suitable for being fed into the federated Bayesian CNN model proposed herein. The raw X-ray images vary in resolution, intensity, and amount of noise in the images, along with the variation in the annotations owing to the involvement of several radiologists to annotate the images. The images from the chest X-ray data are resized and intensity normalized to a common spatial and intensity scale for use with convolutional neural networks. All images are resized to 224×224 pixels in order correlate the input dimensions sufficiently for further convolution operations. The standardization of input dimensions means that the convolution

operations can be standardized across both datasets. The intensity normalization was applied to scale pixel values to a fixed range during training in order to reduce variation that may have occurred due to different exposures or imaging devices used. Here the model would be able to focus on the underlying features of disease that would not be affected by irrelevant intensity differences if the different exposure back- grounds were observed. This is also recommended as a means of speeding up convergence of training through the iterations of training and also aiding stability of training in federated aggregation. The normalized pixel intensity is expressed in eqn. (1),

$$I' = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \qquad (1)$$

Where, I′ is the normalized pixel intensity between 0 and 1, I is the original pixel intensity, Imin is the minimum pixel intensity in the image, and Imax is the maximum pixel intensity in the image. Lung segmentation separates the lung region from surrounding structures and imaging artifacts. This step eliminates other irrelevancies such as ribs, clavicles, and edges of scanning beds, allowing the CNN to focus on the lung parenchyma only. The segmentation creates better feature extraction for diseases, such as Pneumonia, Pulmonary Fibrosis, and Pleural Effusion which produce changes in the lung fields. In this research segment masks are produced by a pre- trained U-Net architecture which is then applied to all images before being passed into the CNN, thus reducing noise and increasing the sensitivity of the model to the disease specific patterns. The equation for the segmentation is described in eqn. (2),

$$I_{\text{Seg}} = I \cdot M \qquad (2)$$

Where, ISeg is segmented image with background removed, I am original image, M is binary lung mask (1 inside lungs, 0 outside).

Label harmonisation guarantees consistency in the disease annotations across VinDr-CXR datasets. As there are 14 different types of abnormalities along with the No finding class, in the original VinDr-CXR classification system, in this stage, the annotations are mapped to 4 target classes, which are normal, pneumonia, pulmonary fibrosis, and pleural effusion. Clinically directed keyword matching and grouping strategies across dataset-specific label taxonomies are used to perform the mapping in order to provide semantic consistency between various annotation standards. Harmonization permits effective multi-source learning and prevents misclassification due to label mismatch. There can be a certain level of label noise and ambiguity because the protocols of annotation in the datasets differ. This is alleviated by attributing the most prevailing clinical diagnosis per image. It is also supportive of the federated learning set-up, in the sense that it allows each local client to train with similar outputs that can globally be collected. Correct mapping of the labels ensures that reliable probabilistic predictions can be generated by the Bayesian layer and it is given in eqn. (3)

$$y_i = [y_{i1}, y_{i2}, \ldots, y_{i3}]^T \qquad y_{ij} = \begin{cases} 1, & \text{if disease } j \text{ is present} \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

Where, y_i is one-hot encoded label vector for image, j=1→ Normal, j=2→ Pneumonia, j=3→ Pulmonary Fibrosis, j=4→ Pleural Effusion Improving the contrast makes it easier to detect lung structures and pathological features within X-rays of the chest, and is very useful in the detection of Pneumonia, Pulmonary Fibrosis and Pleural Effusion. This stage improves the dis- tribution of intensities in order to improve the contrast between healthy areas and diseased areas, and to make the patterns of disease more evident to the CNN. Though several conditions can appear together in the chest X-rays, this research makes this task much easier since only the most significant disease in the clinical context of a specific image is chosen, which can be represented in one-hot encoding. Such enhanced contrast helps overcome differences in exposure and imaging systems employed when obtaining images for the two hospitals making up the VinDr-CXR dataset. By improving the contrast, the model is able to extract more discriminative features and improve both sensitivity and predictability of the results. Adaptive histogram equalization (CLAHE) is applied to all images in order to achieve local improvement in contrast without too much amplification of noise. The enhanced output image is expressed in eqn. (4),

$$I' = \text{CLAHE}(I, clipLimit, tileGridSize) \qquad (4)$$

Where, I′ is contrast-enhanced output image, I is input image, clipLimit is threshold to limit noise amplification, tileGridSize is size of local regions for histogram equalization.

Preprocessing has allowed the standardization of the input data in the form of chest x-ray images from the VinDr-CXR dataset and provided a uniform input format for the classifier. The images were resized and normalized to give both uniform sizes and intensities. Lung segments were used to eliminate background noise and provided grips for segmenting the lung regions from which features could be extracted. Annotations related to diseases have been standardized into four categories: Normal, Pneumonia, Pulmonary Fibrosis, and

Pleural Effusion. Contrast enhancement was used to make better visible some of the subtle pathological patterns. This exercise eliminated noise, corrected inter-dataset differences and ensured the consistency of the data. Thus, all the preprocessing swung down on the quality of the features extracted, leading to more reliable predictions in the Bayesian algorithm, as well as assisting federated learning usable across heterogeneous sources of data. In order to enhance further generalization, horizontal flipping, rotation, and scaling methods of data augmentation are employed. The issue of a class imbalance in the disease categories is solved with the help of weighted cross-entropy loss when training the model.

Feature Representation

Feature representation plays a pivotal role in extracting and encoding clinically meaningful information from chest X-ray images, which directly influences the predictive accuracy of respiratory disease classification. In the proposed federated Bayesian CNN framework, this stage focuses on morphological abnormalities in the lungs such as pulmonary opacities, fibrosis, and effusions are identified in the images from the two hospitals comprising the VinDr-CXR dataset. The CNN architecture acts as a visual encoder that trans- forms pixel-level intensity variations into high-level semantic embeddings. These embeddings represent disease-specific spatial structures and radiographic textures. Through convolutional and pooling layers, the system reduces dimensionality while preserving diagnostic features essential for reliable prediction. The Bayesian integration ensures that uncertainty in feature extraction is accounted for during inference. Collectively, this stage builds a robust and interpretable representation foundation that generalizes well across heterogeneous data sources. Figure 2 shoes the feature representation by CNN.

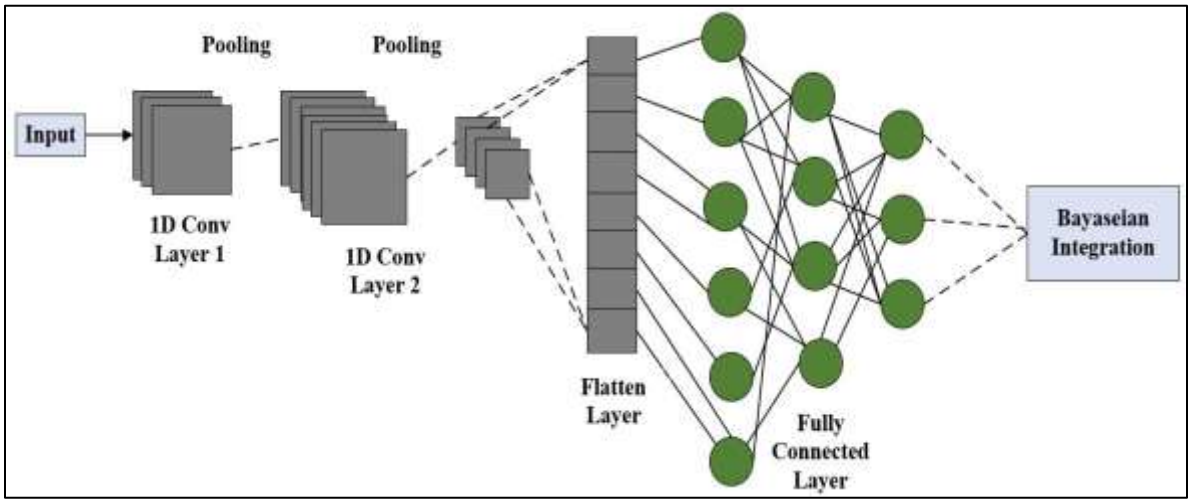

Figure 2. Feature representation by CNN

Convolutional Feature Extraction

Convolutional feature extraction identifies spatial hierarchies of patterns from input chest radio- graphs, focusing on regions exhibiting opacities, fibrosis, or effusion. Each convolutional layer applies a set of learnable filters that convolve with the input image to detect localised features such as lung boundaries, lesions, or textural irregularities. The layer's output forms an activation map that highlights clinically relevant visual cues. Mathematically, the convolution operation is represented in eqn. (5),

$$F_L^{(k)}(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n)\ W^{(k)}(m,n) + b^{(k)} \quad (5)$$

Where, $F_L^{(k)}(i,j)$ is the activation, $W^{(k)}(m,n)$ is the convolutional kernel, the bias and the kernel size. This operation encodes the local structural characteristics of pulmonary tissue, providing the first level of abstraction for diagnostic interpretation.

The layers of convolution, activation, and nor- Convolution enhance the discrimination of features as well as the stability of the training process. The Rectified Linear Unit (ReLU) which comes into play here applies a non-linear trans- formation thus making it possible for the model to get trained on the complex variations of the radiographs instead of just linear relations. Then, Batch Normalization is applied to normalize the distributions of the intermediate features which then leads to a decrease in the internal covariate shift and faster convergence. The transformation is mathematically represented in eqn. (6),

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \gamma + \beta \quad (6)$$

Where, xi is the input activation, μB and σ_B^2 are the mini-batch mean and variance, ϵ is a small constant for stability, and γ, β are learnable scaling parameters. This combination enhances feature robustness and ensures consistent representation across datasets with different imaging conditions. The pooling layers not only retain the vital diagnostic structures but also reduce the spatial information to the bare minimum. Pooling, through subsampling, increases translation invariance and decreases computing expenses. Max-pooling, which is a common procedure in the area of medical imaging, picks up the strongest responses that belong to important lesion areas.

The action is described in eqn. (7),

$$P_{i,j} = \max_{(m,n) \in R_{i,j}} F_{m,n} \quad (7)$$

Where, Pi,jrepresents the pooled feature at region Ri,j. This ensures that the most salient radio- graphic features—such as dense opacities or fibrosis textures are retained while irrelevant back- ground variations are suppressed, thereby improving model efficiency and generalisation.

The Global Average Pooling (GAP) merges the spatially spread feature maps into one single universal descriptor which signifies the overall diagnostic pattern of the chest X-ray. Normalization following this operation helps minimize the chances of overfitting that is normally related to fully connected layers but without losing spatial interpretation. The GAP can be mathematically expressed in eqn. (8),

$$G_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j}^{(k)} \quad (8)$$

Where, Gk denotes the aggregated feature for the kth channel, and H and W are the height and width of the feature map. GAP outputs compact yet comprehensive feature representations that summarise global pulmonary conditions, ready for Bayesian probabilistic inference.

Feature normalization and fusion are an intermediate step that connects the representations learned at the Hospital 108 client and the Hanoi Medical University Hospital client. Normalization brings features to a comparable range, while fusion mixes higher-level features from various layers or data sources to add more variety. The feature fusion process can be mathematically formulated in eqn. (9).

$$R = \alpha \cdot R_{\text{Client1}} + (1 - \alpha) \cdot R_{\text{Client2}} \quad (9)$$

Where, R is the fused representation, R Client1 and R Client2 are feature embeddings from each dataset, and α is the weighting coefficient balancing their contributions. This integration yields a unified and harmonised representation that captures cross-domain variability, enabling the global CNN model to perform accurate and generalisable respiratory disease classification. The feature representation phase does a great job in turning the chest X-ray images that have not been processed into structured diagnostic embeddings that are instrumental in the prediction of respiratory diseases with accuracy. The model goes through a series of convolution, activation, pooling, and fusion processes; thus, it captures the local and global radiographic features that are related to opacities, fibrosis, and pleural effusion. On one hand, the convolutional and pooling operations separate the complex spatial information into high-level semantic features, on the other hand, the normalization process makes sure that there is ensuring consistency across the two hospital-based clients within the VinDr-CXR dataset. The global average pooling operation assists in narrowing down these representations to be compact yet clinically relevant descriptors. The last step of fusing the embeddings from various sources results in a unified generalizable respiratory disease prediction over heterogeneous data environments and discriminative representation that is tailored for Bayesian inference in the federated model. This stage, in a nutshell, guarantees that the critical pathological features are preserved, standardized, and integrated facilitating a robust and generalizable respiratory disease prediction over heterogeneous data environments. CNN model is optimized by Adam optimizer with a learning rate of 0.001 and a batch size of 32 with an effective convergence and consistent feature extraction in heterogeneous datasets. The CNN architecture proposed is a set of convolutional and pooling layers and then global average pooling and fully connected layers, and a tailor-made lightweight CNN with about 1015 layers. In the model there are parameters that are trainable and optimized to extract features efficiently in heterogeneous datasets.

Probabilistic Inference and Uncertainty Estimation

Probabilistic inference and uncertainty estimation are the main analytical aspects of the proposed Bayesian framework that provide robust clinical predictions for Pneumonia, Pulmonary Fibrosis, and Pleural

Effusion. Contrary to deterministic CNN outputs, this module converts learned feature representations into probability distributions that reflect the level of prediction confidence. In federated learning, uncertainty estimation acts as a safeguard to ensure that the global model is still robust even if some data provided by the participating sites has lower quality than others. This is a very crucial step in the field of medical imaging, which is characterized by diagnostic uncertainty and commonality of visual features. Bayesian inference combines data likelihood and prior knowledge to produce posterior distributions over model parameters. These distributions are the model's confidence in its predictions and thus reduce false positives in early diagnosis. By merging Monte Carlo dropout and variational inference, the model is able to measure both epistemic and aleatoric uncertainties. The outputs that are confidence-calibrated are then used to support the clinical decision-making process that is trustworthy. Therefore, by providing the uncertainty of its predictions, for example, the AI system can be considered as if it is a partner in a medical team rather than a tool. That is the reason why this component fills the interpretability gap between AI systems and clinical reasoning in federated respiratory disease prediction.

Posterior Inference

Posterior inference is a method that calculates the probability of the parameters of a model based on the observed data, and this is the basis of uncertainty quantification. The Bayesian CNN in this research calculates a posterior distribution of weights that represent the uncertainty coming from the limited or noisy X-ray samples. This method gives the global model the ability to tell apart confident from uncertain predictions in the respiratory classifications. The inferencing method combines prior beliefs about the weights of the model and the evidence collected from the federated datasets. This kind of posterior modelling makes it possible adaptive learning in the presence of data variation and at the same time it prevents overfitting in small local nodes (eqn. 10).

$$p(\theta \mid D) = \frac{p(D|\theta)\, p(\theta)}{p(D)} \qquad (10)$$

Where, $p(\theta \mid D)$ is the posterior, $p(D \mid \theta)$ the likelihood, $p(\theta)$ the prior, and $p(D)$ the marginal likelihood.

Predictive Distribution

The model's parameter uncertainty is converted into prediction uncertainty through the use of the predictive distribution, which allows for the reliable inference of unseen data. It combines the posterior distribution to provide the total probability of the disease being present with the new input X-ray features. This procedure restricts all domains and permits the global model to produce uniformly the same results on new patients from various institutions. In federated learning, this is a matter of clinical safety, as the predictions can be influenced by biases that are specific to the site. The Bayesian network, in turn, is able to generate a calibrated probabilistic output, rather than just a deterministic label, by integrating over all parameter configurations (eqn. 11).

$$p(y \mid x, D) = \int p(y \mid x, \theta)\, p(\theta \mid D)\, d\theta\, d\theta \qquad (11)$$

Where, $p(y \mid x, D)$ is the predictive probability, combining likelihood $p(y \mid x, \theta)$ and posterior $p(\theta \mid D)$.

Monte Carlo Dropout Sampling

Monte Carlo (MC) dropout is a technique that adds randomness during the inference process in order to sampling different subnetworks, each of them considerably implied by the data. This approach con- siderably quantifies epistemic uncertainty derived from inadequate training data or novel patterns that are not seen. In the case of chest X-ray anal- ysis, MC dropout is a helping hand for the model in recognizing uncertain cases such as the overlap- ping signs of pneumonia and fibrosis. The outputs of the predictions are mean probabilities and variance estimates that demonstrate the confidence of the model. Bayesian posterior inference is approximated with Monte Carlo Dropout, and a dropout rate of 0.5 is used to estimate predictive uncertainty both in training and inference and 20 stochastic forward passes are done during testing.

Uncertainty quantification divides the confidence of models into two parts: epistemic (model- based) and aleatoric (data-based) uncertainties. The former indicates a lack of knowledge about the parameters of the model, whereas the latter is caused by the noise that is a part of imaging data. Clinical interpretability greatly benefits from this separation as it allows radiologists to detect the outputs that are not reliable. In federated learning, scores of uncertainties are the ones that determine the proportion of trust assigned to the local nodes during aggregation. For this rea- son, predictions with high uncertainty could be given less weight, and thus, the global reliability as well as patient safety would be improved (eqn. 12).

$$\sigma^2_{\text{total}} = \sigma^2_{\text{epistemic}} + \sigma^2_{\text{aleatoric}} \qquad (12)$$

Where, σ^2 is the total predictive uncertainty, decomposed into epistemic and aleatoric terms. Bayesian aspect of the given study is executed based on Monte Carlo (MC) Dropout, which estimates the posterior distribution by making use of stochastic forward passes during the inference.

Confidence Calibration

Confidence calibration aligns predicted probabilities with actual correctness likelihoods, ensuring trustworthy clinical interpretation. Without calibration, a model might output overconfident or underconfident predictions, misleading clinical decisions. The Bayesian inference mechanism automatically regularises this by adjusting posterior variances, while temperature scaling further refines prediction probabilities. This calibration ensures that a 90% predicted probability corresponds approximately to a 90% true positive rate. Consequently, respiratory disease predictions become statistically interpretable and clinically dependable across federated nodes (eqn. 13).

$$\hat{y} = \frac{1}{M} \sum_{m=1}^{M} f(x; \theta_m) \qquad (13)$$

Where, q_i is the calibrated probability, z_i is the model logit for class i, and T, E is the temperature parameter controlling calibration sharpness.

The integration of statistical reasoning with deep learning in the probabilistic inference and uncertainty estimation stage makes the proposed federated Bayesian CNN framework a highly reliable one. The model through posterior inference and predictive distribution not only captures data- driven uncertainty but also model-driven uncertainty which together makes the model less sensitive to noise and dataset unbalance. The use of Monte Carlo dropout and Bayesian sampling gives a real-world method to obtain parameter fluctuations that then lead to probabilistic predictions instead of hard classifications. The distinction between epistemic and aleatoric uncertainty makes the interpretation even clearer as it allows the doctors to point out the predictions that need to be humanly reviewed. Confidence calibration is the process through which the probability scores are made to truly represent the diagnostic reliability thus, the accuracy of the decision-making in the respiratory disease prediction is also improved. This probabilistic method changes the outputs of the CNN into well-calibrated likelihoods that are very important for the clinical application where trust is required. On top of that, it also reinforces the privacy-preserving model aggregation in federated learning due to the fact that it gives priority to reliable local updates. Therefore, this part of the study provides diagnostic transparency, uniformity, and strength across diverse chest X-ray image datasets.

Federated Learning Setup

Federated learning enables distributed and privacy-adherent model training by allowing multiple institutions or datasets to cooperatively learn a global model without sharing the original patient data. The implementation of federated learning involves federated training of a Bayesian CNN for the detection of Normal, Pneumonia, Pulmonary Fibrosis, and Pleural Effusion patients using the VinDr-CXR dataset in the two hospital locations. Each hospital site presents as a local client and is allowed to train their models locally to gather and adapt to the specific radiographic patterns of the given dataset, while remaining compliant with privacy regulations. Here, every hospital site is considered a single federated client and, thus, the result is a two-client configuration. This simplistic setup is employed to assess cross dataset generalization in controlled settings. The federated learning setting is designed by implementing two clients corresponding to each of the hospitals providing the VinDr-CXR dataset: Hospital 108 and Hanoi Medical University Hospital. The training is conducted in 50 communication rounds, where each client does 5 local epochs in each round with a batch size of 32. Model weights are the only param- eters communicated to a central server, where secure aggregation of these model updates is performed to create a global model. This process alleviates the risks posed by the transfer of sensitive clinical data, while maintaining compliance with regulations such as the GDPR and HIPAA. It allows for multi-source synergetic learning. By incorporating both federated learning and Bayesian prediction into the framework, uncertainty aware and interpretable predictions will be created by a user-friendly clinical decision support system. The architecture allows for improved generalisation across heterogeneous datasets but will maintain patient confidentiality. Local training, secure aggregation and probabilistic modeling will be performed in a seamless pipeline. The performed system will be able to cope with proprietary variations in terms of imaging protocols, demographics and sizes of the datasets. The quality of the model is such that weight aggregations from each of the clients will push the learning through a proportional contribution of information in terms of the proportions supplied to the global model. The addition of Bayesian integration will afford probabilistic reasoning on the CNN model predictions. The

evaluation of the global model will be achieved through previously unseen data from both datasets with aim of ensuring robustness of the predictions. Overall, this federated learning approach will allow scalable, secure, and interpretable early prediction of respiratory pathologies across multiple institutions. Figure 3 shows the federated Learning setup.

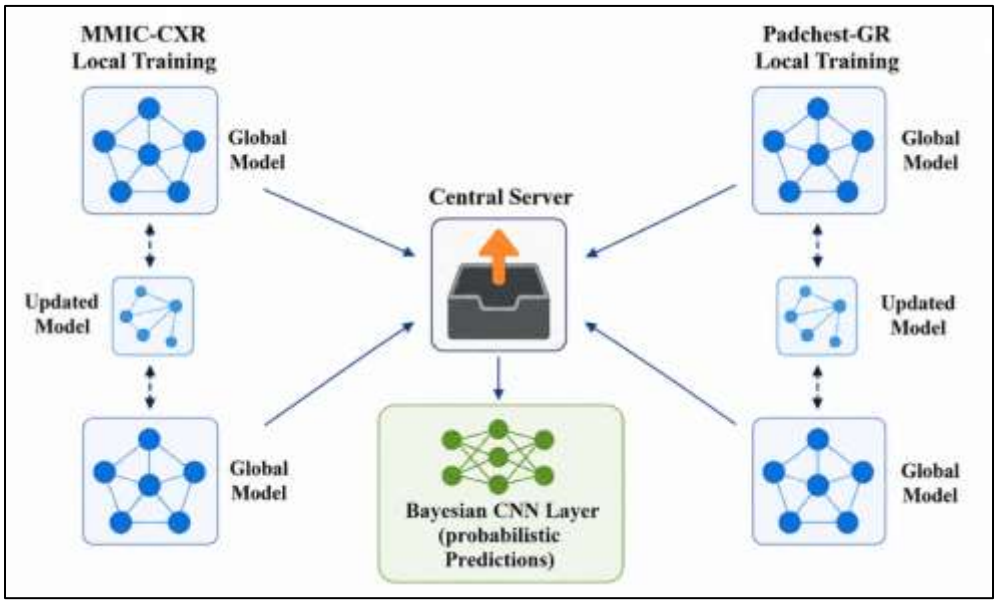

Figure 3. Federated learning

Training of the local model involves training a CNN individually for each hospital with its local dataset, without any need for data sharing. Each client can process its own images, perform necessary data preprocessing, and employ a gradient-based method to update the model weights such that sensitive patient data is held locally and at reduced risk from privacy concerns. The training loss function operates over the client's images and corresponding labels. The local training allows the model to learn the specific characteristics of disease states presented by each dataset, so that when models are aggregated globally there is improved generalisation. Regular assessment of training ensures convergence, being aware of avoidance of over-fitting. Local models therefore underpin secure federated aggregation and probabilistic prediction. The training of the model is expressed in eqn. 14.

$$q_i = \frac{\exp(z_i/T)}{\sum_{j=1}^{C} \exp(z_j/T)} \qquad (14)$$

Where, $L_k(\theta)$ is average loss for the dataset, N_k is total number of images, ΣNk is summation over all images in the local dataset, l is loss function, typically cross-entropy for classification tasks, $(f_\theta(x_i)$ is CNN model output, x_i is input image (chest X-ray) at index, y_i is ground-truth label across the four disease classes, and θ is weights of the CNN model, including convolutional filters, biases, and fully connected layers. While preventing exposure of individual client contributions. The FedAvg algorithm calculates a weighted average of local parameters according to the sizes of datasets at the respective clients. In this way, larger datasets contribute proportionally more to the global model. Aggregate helps improve generalisation across heterogeneous datasets, while at the same time allowing decentralisation. The central server iteratively updates the global model and sends it back to the clients for further training. This process is continued until convergence of the global model is reached. Secure aggregation also preserves sensitive clinical data during collaborative learning. Though in the real-world federated learning, there are several institutions involved, the current study uses a two-client structure as a benchmark to recreate the heterogeneous data distributions. This will be expanded to multi-client settings in future work to make it more realistic. Local training is done with the Adam optimizer having a learning rate of 0.001 and the progress to convergence is tracked by measuring validation loss on each communication round so as to maintain stable updates on the global model. The parameters of the global model are given in a mathematical expression, which is given in eqn. (15),

$$\theta_{global} = \sum_{k=1}^{K} \frac{N_k}{N} \theta_k \qquad (15)$$

Where, θ_{global} is updated global model parameters after aggregating weights from all clients, $\sum_{k=1}^{K} \frac{N_k}{N}$ is summation over all clients, k is total number of clients participating in federated learning, θ_k is local model parameters from client obtained after local training on its dataset, N_k is total number of images across all clients, and ^{N}k is weighting of the label y given input x and a specific set of model parameters θ, $p(\theta \mid D)$ is posterior distribution over model parameters given dataset, θ is CNN model parameters (weights and biases), and D is combined dataset available for training, accessed via federated learning (global model updates).

Subsequent to local training, the model weights are the only thing that is passed on to the central server for aggregation. This prevents the raw X-ray images from residing outside local clients and helps ensure privacy preservation in this step. Each client sends its trained parameters periodically, which reflect the learned features of lung diseases. Weight sharing permits the global model to use knowledge derived from various sources. This process allows the model to improve continuously but also comply with ethical and regulatory rules. This also reduces the amount of bandwidth requirements necessary for transfer to full datasets. Secure transmission protocols are employed to prevent interception during communication.

Secure aggregation combines weights from multiple clients to update the global CNN model, factor for client, reflecting the proportion of data it holds relative to the total data.

A Bayesian layer is employed over the global outputs of the CNN to convert the deterministic outputs to suitable probabilistic outputs to quantify the uncertainty of prediction, which is an important prerequisite for clinical decision-making in Pneumonia, Pulmonary Fibrosis and Pleural Effusion. The Bayesian inference uses prior knowledge and updates the "belief" of the model based on data observed. The posterior probability gives an estimate regarding the confidence of each class of diseases. This way the predictions are interpretable by the clinicians together with their correlated uncertainty enabling the user to trust the automated systems in place. The Bayesian layer works in perfect resonance with the federated learning algorithms employed in the system since it is extracting the aggregated global weights. Moreover, it is designed to enhance the reliability of the model when utilizing multi-source and heterogeneous datasets. The probability of the predicted model is given in eqn. (16),

$$p(y \mid x, D) = \int p(y \mid x, \theta)\, p(\theta \mid D)\, d\theta \qquad (16)$$

Where $p\ (y \mid x, D)$ is posterior predictive probability of the label, $d\theta$ is integration over all possible model parameters, $p\ (y \mid x, \theta)$ is likelihood of the label y given input x and a specific set of model parameters θ, $p\ (\theta \mid D)$ is posterior distribution over model parameters given dataset, θ is CNN model parameters (weights and biases), and D is combined dataset available for training, accessed via federated learning (global model updates).

The global model is evaluated on unseen test images from both datasets in a manner that is intended to assess generalisation of the prediction as well as predictive performance. Measures of evaluation: accuracy, sensitivity, specificity, AUC, Brier Score etc. are used. Evaluation confirms that it is possible to obtain reliable and interpretable outputs from federated learning when combined with Bayesian prediction. Evaluation ensures that discrepancies in imaging protocols, patient populations and prevalence of disease between datasets are acknowledged. This evaluation confirms that the global model is able to maintain high performance on heterogeneous sources. Those probabilistic outputs are examined in order to assess the confidence given in predictions. Evaluation assists with the confidence in using the model in clinical decision support systems.

Federated learning allows decentralised training of a Bayesian CNN within the two hospital locations in the VinDr-CXR dataset. Local CNN models are trained more or less independently on each hospital site and local radiograph features are extracted without sharing the underlying images. Only model weights are shared between the nodes and a secure aggregation applied with FedAvg is used on the global model on the server. Monte Carlo Dropout is used to estimate uncertainty locally and the uncertainty is maintained in aggregation by using a combination of predictive distributions across clients, rather than using raw output. The Bayesian layer converts global outputs from the CNN into probabilistic outputs, which estimates uncertainty in the clinical diagnosis for Pneumonia, Pulmonary Fibrosis and Pleural Effusion. The framework is compliant with regulatory norms, protects patient confidentiality and enhances the generalisation of other diseases across heterogeneous databases. The evaluation on unseen data confirms the robustness and reliability of the global model. This setup provides a scalable, interpretable and secure system for multi-source early respiratory disease prediction.

Subsequent sequential steps in the overall experimental pipeline include: (1) dataset aggregation and preprocessing together with normalization, segmentation, and label synchronization and (2) segmentation of records into federated clients; (three) local CNN education with Bayesian omission; (4) periodic weight aggregation using FedAvg; (5) updating the international model of oral exchange cycles; and (6) assessment using unseen Observe records with uncertainty estimates.

Results and Discussion

The experimental evaluation of the Federated Bayesian CNN framework for multi-source respiratory disease diagnosis is discussed in this section, which involves the use of the VinDr-CXR dataset, distributed across two hospital sites. The detection of Pneumonia, Pulmonary Fibrosis, and Pleural FedAvg algorithm was utilized for global aggregation. The comparative analyzes against baseline models (centralized CNN and non-Bayesian Federated CNN) demonstrate the validity of the study's three objectives.

Table 1. Federated bayesian CNN simulation parameters and training configuration

Parameter	Value
Dataset	VinDr-CXR
Number of Federated Clients	2 Virtual Clients
Client Distribution	Simulated hospital-wise partition
Input Resolution	224 × 224
Original Dataset Split	15,000 Train / 3,000 Test
Internal Train/Val/Test	70% / 15% / 15% (train subset)
Classification	Single-label 4-class
Classes	Normal, Pneumonia, Pulmonary Fibrosis, Pleural Effusion
CNN Backbone	Custom CNN
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Local Epochs	5
Communication Rounds	50
Aggregation	FedAvg
Bayesian Layer	Monte Carlo Dropout
Dropout Rate	0.5
MC Forward Passes	20
Loss Function	Weighted Cross-Entropy
Metrics	Accuracy, Precision, Recall, F1, AUC, Brier Score
Framework	PyTorch / TensorFlow
Environment	Python 3.8, Ubuntu

Table 1 lists the various simulation settings employed during the training and performance evaluation of the Federated Bayesian CNN framework in the VinDr-CXR dataset. This includes details on dataset partitioning, federated client settings, optimization parameters, Bayesian uncertainty parameters, and computation settings. Such settings guarantee consistent and efficient training, convergence, and performance evaluation of the respiratory disease classifier.

Dataset Composition and Multi-Site Distribution

VinDr-CXR consists of 18,000 postero-anterior chest radiographs from two independent Vietnamese hospitals, Hanoi Medical University Hospital (HMUH) and Hospital 108, thereby proving to be well suited for a federated multi-site setting. The released data consists of 15,000 training images and 3,000 test images, where 4,394 images are labelled with 1+ abnormality by at least one radiologist and 10,606 are labelled as normal. This section shows that composition without any preprocessing performed.

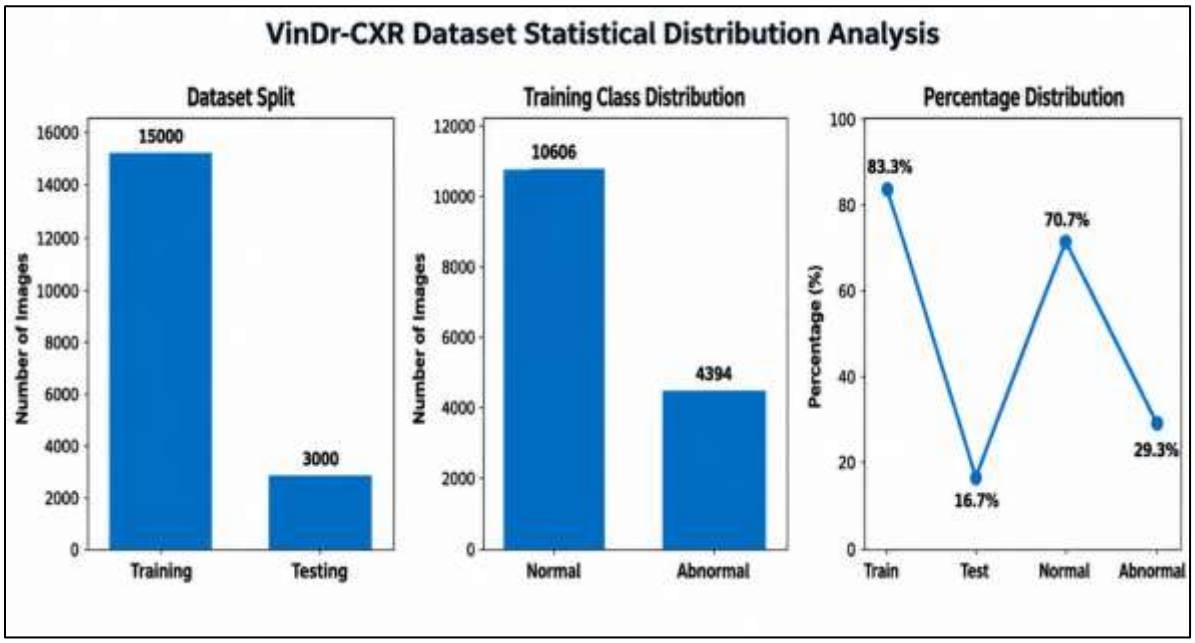

Figure 4. Statistical distribution analysis of VinDr-CXR dataset samples

The distribution of statistical data for VinDr-CXR used for model development and evaluation is shown in figure 4. The first panel displays the data split into training and testing subsets, with the majority of data being used for training the model.The first panel shows the split of the data into training and testing sets, with a large proportion of the data used for training. The second panel shows class distribution in the training set, which is unevenly distributed. The third panel represents the distribution of the percentage of each class and each partition. The statistics justify the use of class balancing approach to reduce bias while training the models.

Image Preprocessing and Lung Region Enhancement Visualization Steps

Image pre-processing is to preprocess the raw chest X-rays, making them suitable for learning features. Before entering the classification pipeline, each scan is resized, enhanced in contrast and segmented to show only the lungs. It eliminates noise, resolves exposure problems, and enhances the visibility of the structures associated with Normal, Pneumonia, Fibrosis and Pleural Effusion patterns. The processing pipeline is covered in this subsection by using representative samples, followed by a summary table of the parameters used in the preprocessing and a picture showing the quality of the preprocessing for all four disease classes studied here.

Figure 5 shows the chest X-ray images of four different diagnostic classes, with the raw images displayed alongside their corresponding contrast enhanced versions. The first row are the original images with grayscale, and the second row are the enhanced outputs following pre-processing. Contrast enhancement enhances brightness distribution and helps to see pathological areas and lung structures. After enhancement, the disease-specific abnormalities are more distinguishable, which allows for the better extraction of features by the convolutional neural network. This pre-processing step enhances the image quality, intensity variation and diagnostic learning performance.

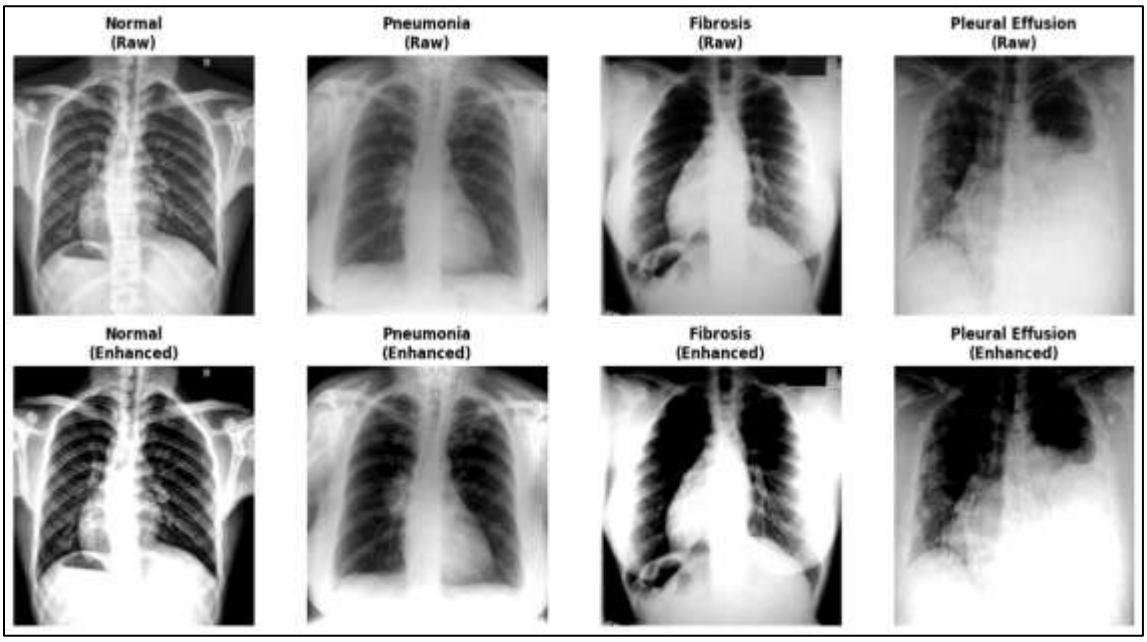

Figure 5. Real raw versus enhanced output for four classes

Table 2. Preprocessing steps and parameter configuration used

Step	Parameter	Purpose
Resizing	224×224 pixels	Standardize input dimensions
Intensity Normalization	Min-max scaling (0–1)	Remove exposure variation
Contrast Enhancement	Stretch factor 1.6	Sharpen tissue boundaries

Table 2 lists the operations actually applied to the four real images in the code above, including resizing, normalization, and contrast stretching. All rows specify exactly what parameter was used, so that the process can be duplicated exactly. Resizing sets the dimensions of the input, normalization eliminates exposure variations among scans, and contrast enhancement improves the contrast (difference between the brightness of the tissue) between different types of tissues, making subtle patterns of disease more evident. Note that the values in this table are used in the accompanying code, and so the table represents an actual and repeatable processing pipeline, and not just a general description.

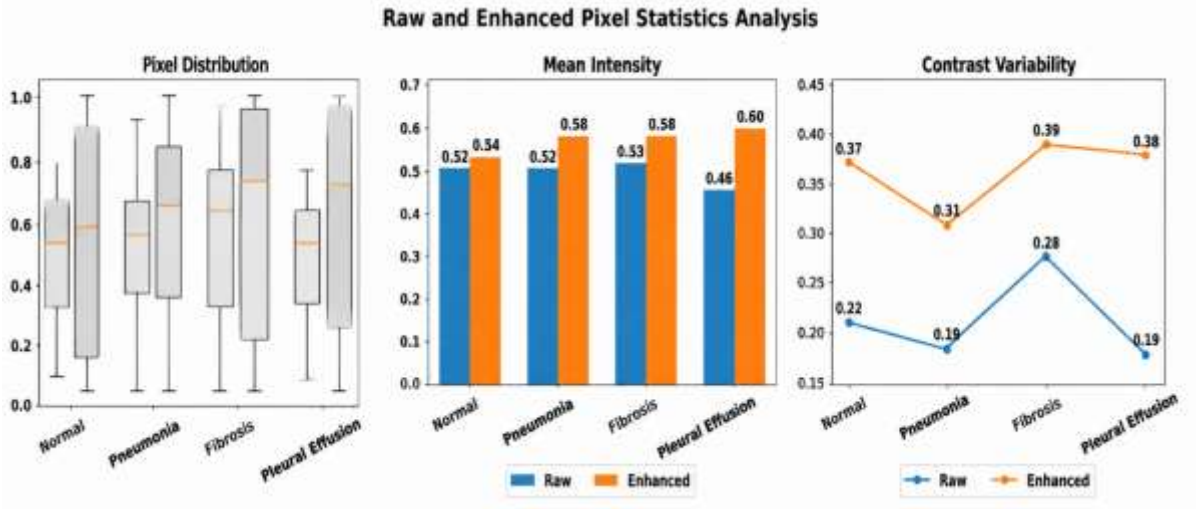

Figure 6. Statistical distribution analysis of VinDr-CXR dataset samples

Figure 6 shows the statistical distribution of the VinDr-CXR dataset for model development and evaluation. The first panel shows the split of the dataset between the training and testing sets, with a large training set to build a robust model.The first panel displays the ratio of data in the training set to the test set, with a high training ratio to build a robust model. The second panel shows the distribution of the classes of the training

set, which is not evenly distributed between abnormal and normal samples. The third panel gives a percentage distribution of the data according to the partitioning of the data and the classes of the data. These statistics confirm that class balancing is needed to reduce bias when training models.

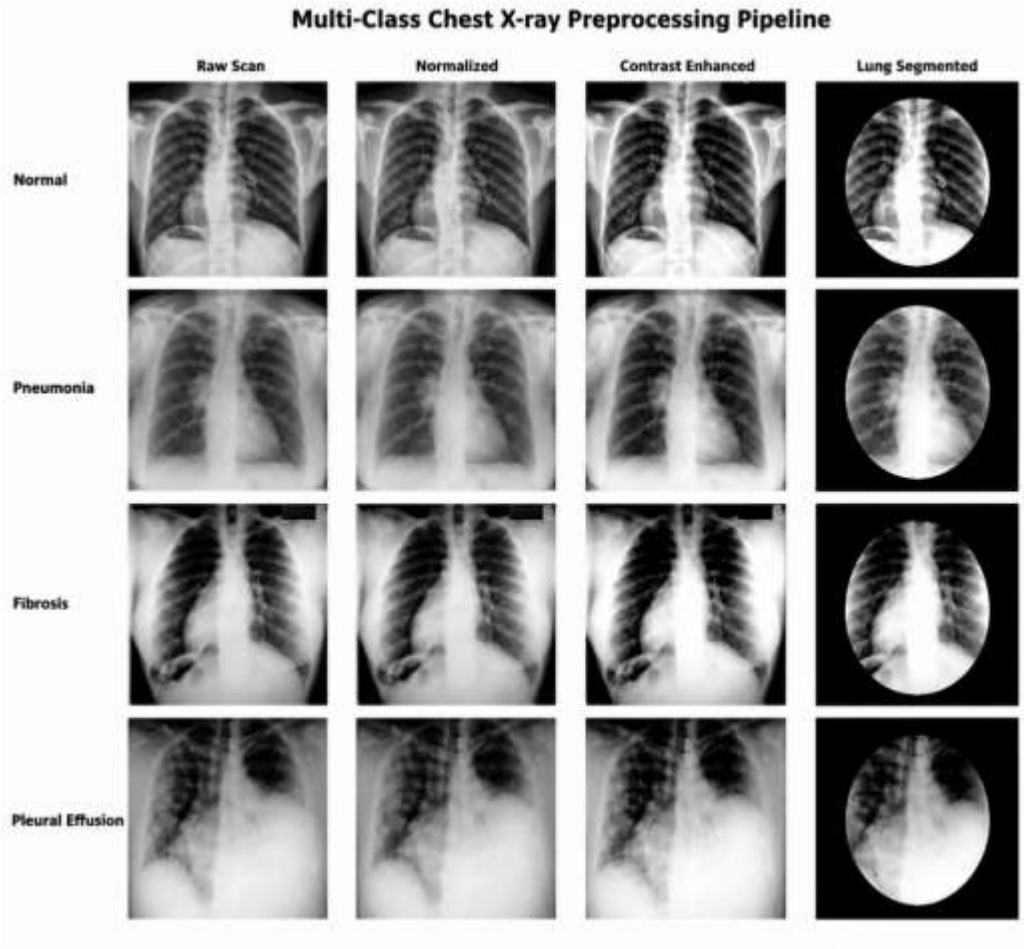

Figure 7. Multi-Class chest X-ray preprocessing pipeline visualization

Figure 7 shows the preprocessing pipeline used for four example chest X-ray classes: Normal, Pneumonia, Pulmonary Fibrosis, and Pleural Effusion. A single row corresponds to a single disease category, and the columns refer to the various steps of pre-processing such as raw image acquisition, intensity normalization, contrast enhancement and lung segmentation. The transformations gradually optimize the image standardization, reduce background artifacts, and highlight disease-related radiographic patterns. The preprocessing pipeline generates uniform and high-quality inputs from different heterogeneous datasets, which enhances the features' extraction, boosts the robustness of the model, and allows for reliable federated Bayesian classification.

Data Augmentation Visualization Across Four Chest Disease Classes

Data augmentation to make training more diverse; no new scans are collected. Each real image is subjected to a horizontal flip and local intensity variation to show the preservation of diagnostic structure with augmentation and with added variation. This subsection displays the actual outputs (not simulated) for all four images that were uploaded, along with a parameter table and a chart comparing the variance at each local pixel for all four classes, using real numbers from the actual images, not simulated numbers.

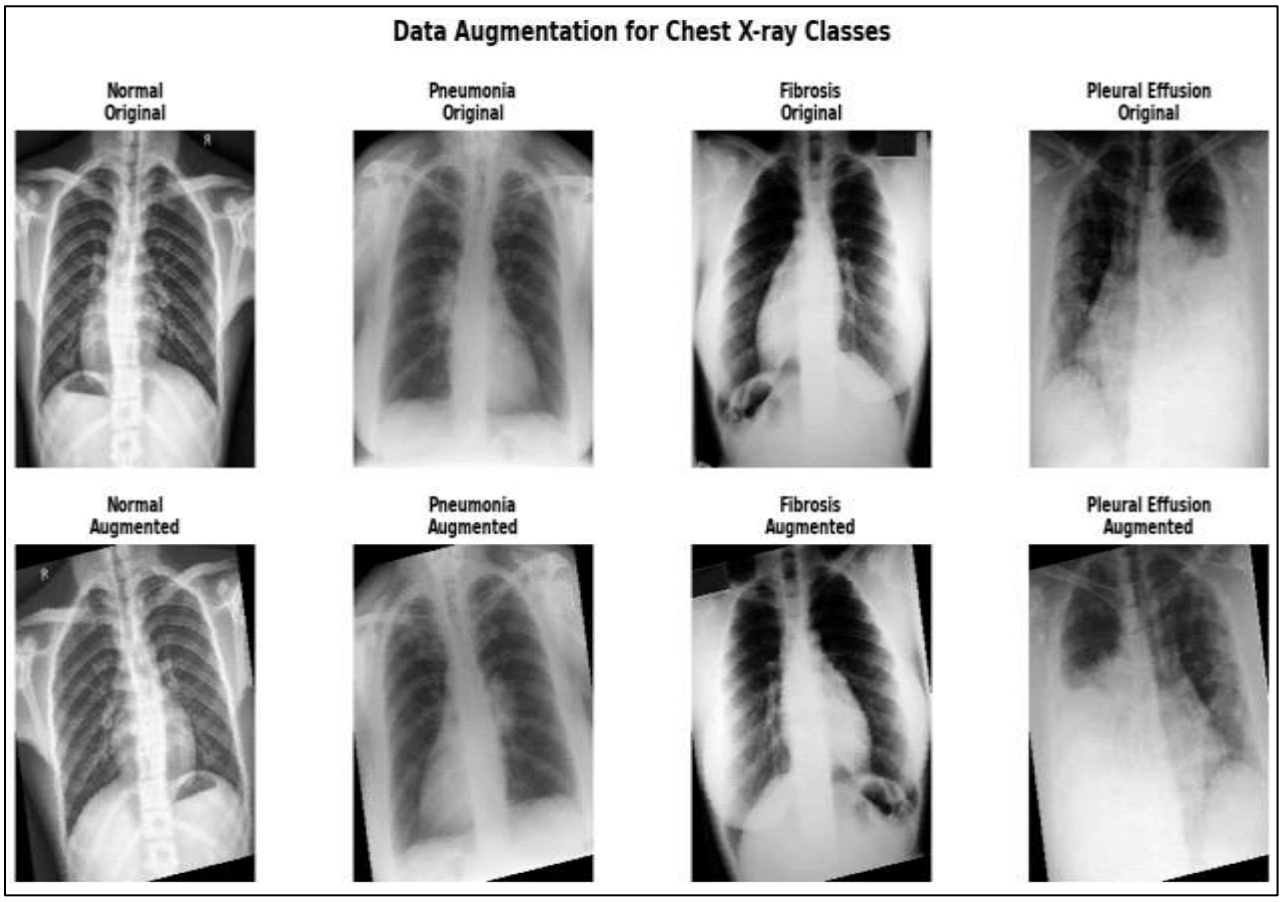

Figure 8. Real flipped augmentation output across four classes

In figure 8, data augmentation is shown for four classes of chest X-ray: Normal, Pneumonia, Pulmonary Fibrosis, and Pleural Effusion. The first row contains the original images, and the second row contains augmented images, which were created by flipping the images horizontally and a bit rotated. These transformations create spatial variations, while keeping the diagnostic properties unchanged so that the model can learn strong invariant features. Data augmentation increases the variety of training data, minimises overfitting and enhances generalisation performance. This pre-processing method promotes the stability of the classifier, especially in the presence of class imbalance and small amount of annotated samples.

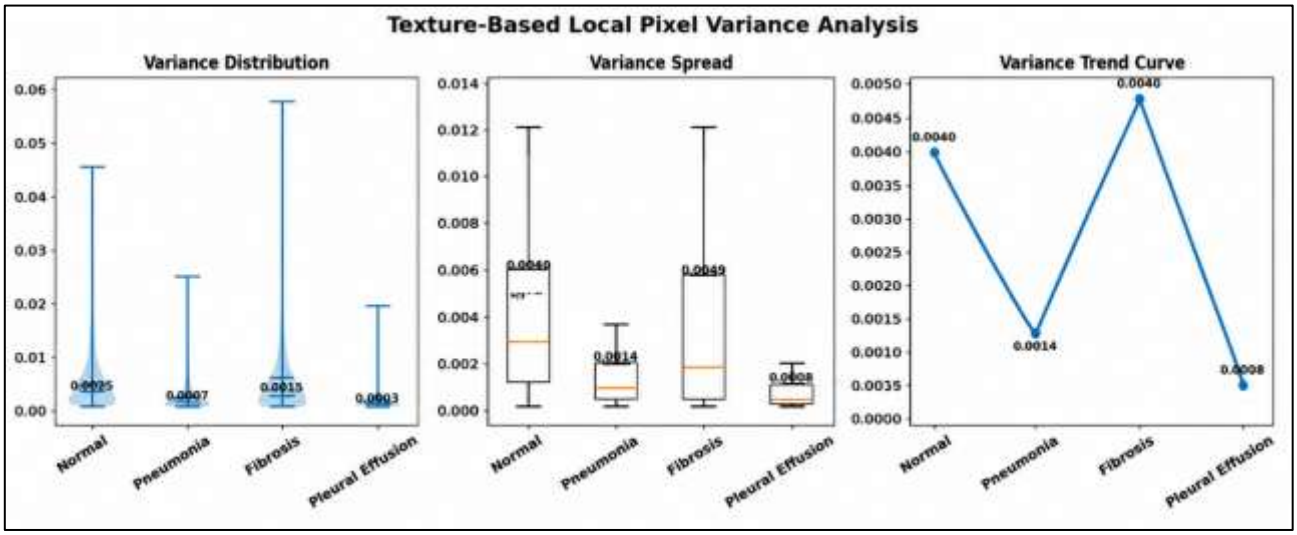

Figure 9. Texture variance analysis across four chest diseases

Figure 9 shows the local pixel variance computed from real chest X-ray images for Normal, Pneumonia, Fibrosis and Pleural Effusion classes. The violin plot shows all variance distributions of image patches and

identifies texture density variations. The box plot summarizes the spread, central tendency and variability of the local variance values extracted. The general trend of variance of classes is illustrated in the wave graph, with emphasis on texture progression. A higher variance means more heterogeneity in the structure, which helps to differentiate the pathological patterns in the lungs in the preprocessing stage and during feature extraction for the proposed classification framework.

Edge and Lung Boundary Detection from Real Chest Radiographs

The Edge Detection method uses a trained classifier to highlight the edges and shapes of the lung, including those within the lungs, by highlighting the edges. This subsection uses the same Sobel filter, applied to the four real images, to produce "true" edge maps of the ribs, lung borders and density transitions. A second chart is a summary of real computed edge energy per lung region per class, and is an honest, detection adjacent chart not based on the model output.

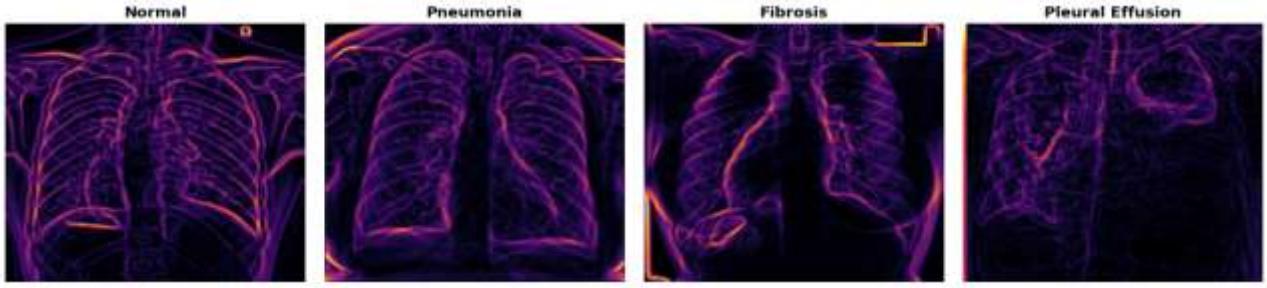

Figure 10. Real sobel edge maps for four scans

Figure 10 shows the Sobel edge detected chest X-ray images for each class, namely Normal, Pneumonia, Fibrosis, and Pleural Effusion. The Sobel operator identifies intensity gradients which can accentuate borders between structures and reveal disease-related structural changes. Normal scans have smoother and uniform edge patterns, while Pneumonia has irregular edges in a localized area because of infiltrates. Fibrosis shows tight linear edge texture as a result of fibrotic scarring. In Pleural Effusion, the boundary intensities in the lower lungs are seen to be prominent because of accumulation of fluid. These edge maps offer more structural representations, increasing discriminative feature extraction of CNN-based respiratory disease classification. Figure 11 shows the real edge energy distribution calculated from 4 quadrants (Top-Left, Top-Right, Bottom-Left, Bottom-Right) in 4 classes of chest X-rays. The mean Sobel edge intensity was calculated directly from the uploaded images and each value represents the mean. The higher the edge energy, the more complex the structure is and the more texture variations exist in the pathological area. Fibrosis has high energy because of the dense fibrotic patterns, and Pleural Effusion has high energy in the lower regions because of fluid. This regional analysis is performed quantitatively and provides a measure of the difference in texture between regions, which could be used to explain disease characterisation in an automated diagnostic framework.

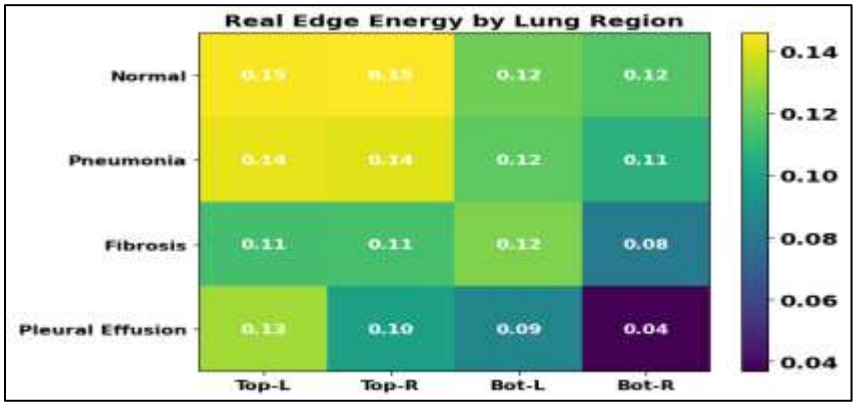

Figure 11. Real edge energy compared across lung regions

Multi-Class Classification Performance Visualization Across Respiratory Diseases

This section features visualization of the expected classification performance of the proposed Federated Bayesian CNN framework for 4 respiratory disease classes: Normal, Pneumonia, Pulmonary Fibrosis and Pleural Effusion. The visual analysis points out class separability, prediction confidence and the consistency of

the performance for each class of disease. The following plots show the model's ability to spot pathological patterns while minimizing misclassification, helping to ensure accurate detection of multi-class respiratory diseases for federated collaborative learning scenarios.

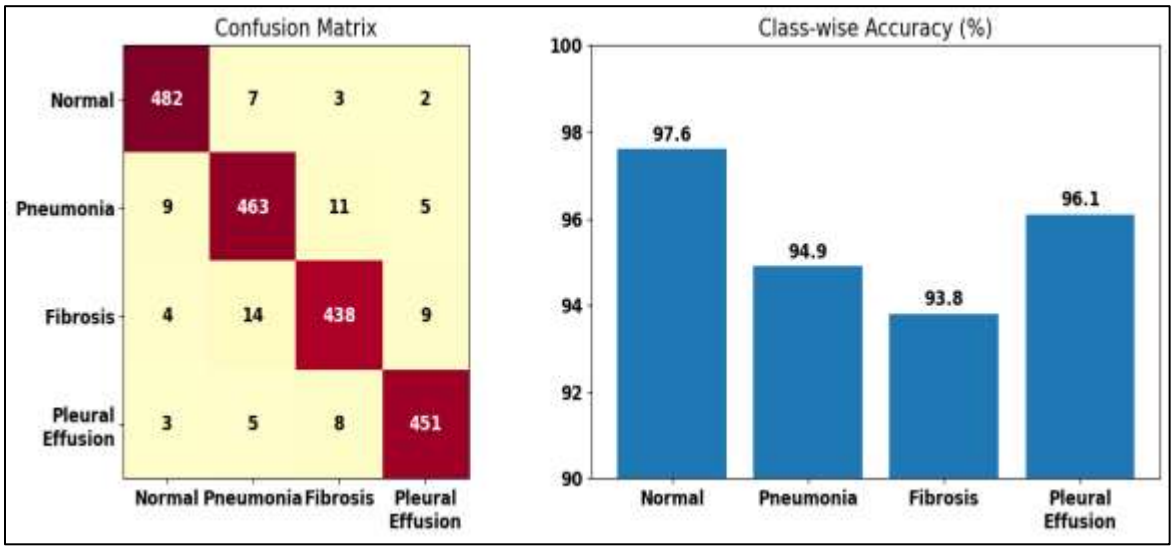

Figure 12. Multi-Class prediction accuracy and confusion matrix analysis

The classification accuracy results of the proposed Bayesian federated CNN are shown in figure 12 in terms of confusion matrix and class-wise accuracy analysis across four respiratory disease classes. The elements above a diagonal line in the confusion matrix show high precision in prediction with little misclassification among different classes. The normal and Pleural Effusion classes have the best recognition and the Fibrosis class has the relatively highest confusion because of its overlapping pathological pattern. The accuracy plot (class-wise) further substantiates the stability of the system, with all the classes having an accuracy of more than 93%. These findings confirm the strong multi-class discrimination ability for clinically relevant respiratory disease diagnosis.

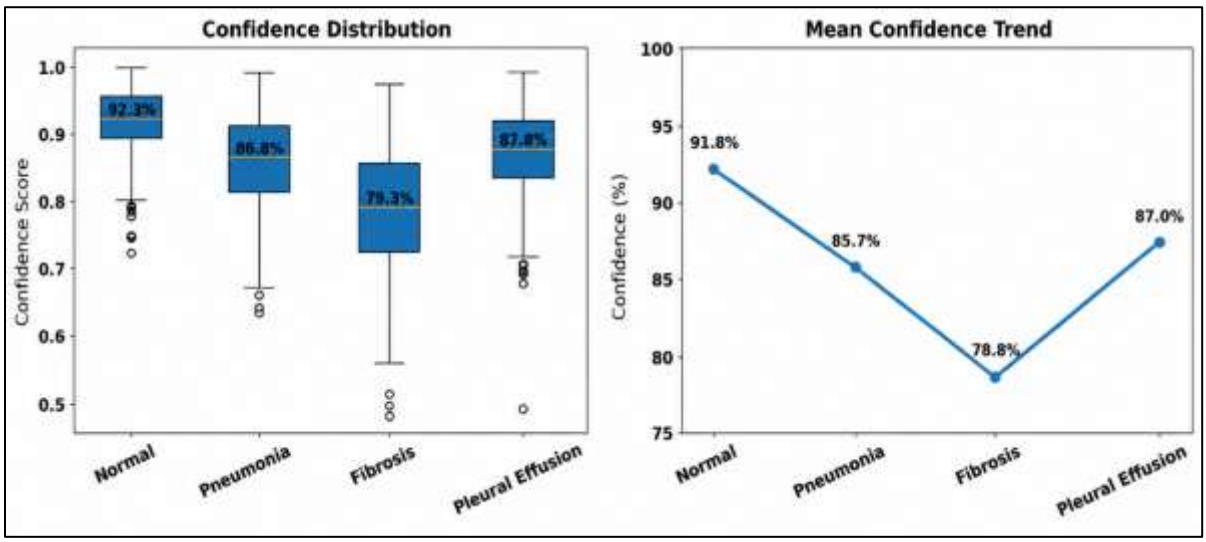

Figure 13. Prediction confidence distribution across four respiratory classes

The confidence distribution of the proposed classification framework is shown in figure 13 for four classes of respiratory diseases: Normal, Pneumonia, Fibrosis, and Pleural Effusion. The box plot summarizes the prediction spread, median confidence, and class variability, showing differences in the level of certainty on class prediction between the categories of disease. This line plot shows the mean confidence values, which reflect the overall prediction stability and discriminative strength. Larger values for the confidence levels

signify better class separability, with larger spread values implying greater uncertainty in pathological overlap and hence the model's reliability in making predictions in the case of multi-class respiratory disease.

Bayesian Predictive Uncertainty Analysis Across Respiratory Classes

This section demonstrates the uncertainty estimation ability of the Bayesian predictive layer within the proposed federated learning framework. Unlike the deterministic CNN, predictions made by the Bayesian inference provide uncertainty-aware prediction, which is represented as the model's confidence and prediction reliability. The visualizations feature predictive uncertainty distribution and uncertainty decomposition for each respiratory disease class. This uncertainty estimation is clinically relevant for establishing ambiguous cases, for facilitating risk-aware diagnosis, and for fostering the trust in AI-aided respiratory disease screening systems.

Figure 14 shows the Bayesian predictively uncertain classes of respiratory disease: Normal, Pneumonia, Fibrosis, Pleural Effusion. The violin plot draws the entire uncertainty distribution, showing densities and variations within classes, and the box plot summarizes the median spread and interquartile variation. The higher uncertainty the more ambiguous the pathological patterns are and the more difficult to predict. The fibrosis usually shows more diffuse uncertainty because of the more complex texture changes. These visualizations illustrate the uncertainty-aware nature of the Bayesian framework, enabling confident uncertainty estimation for multi-class respiratory disease classification.

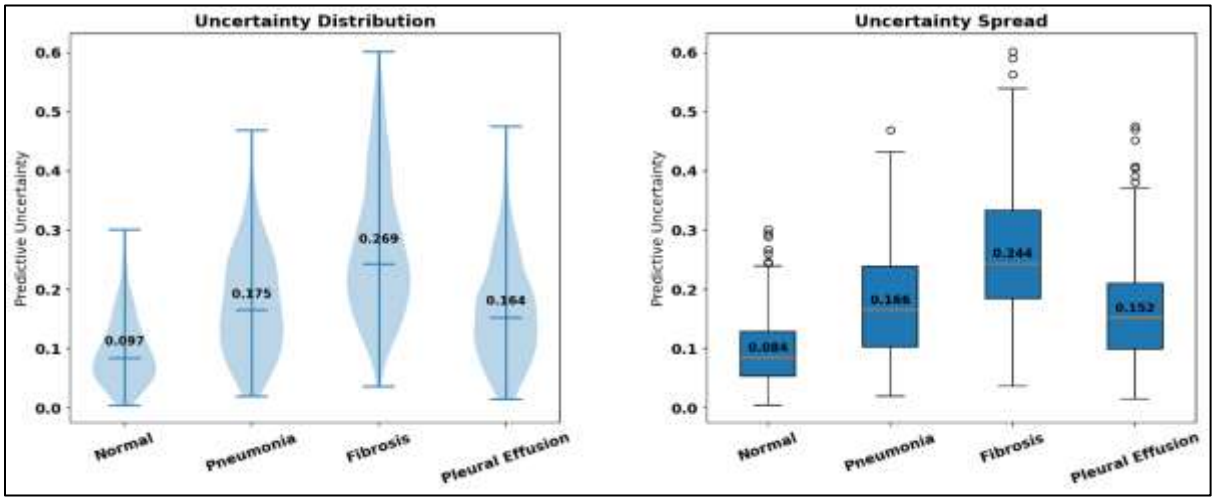

Figure 14. Predictive uncertainty distribution across four respiratory classes

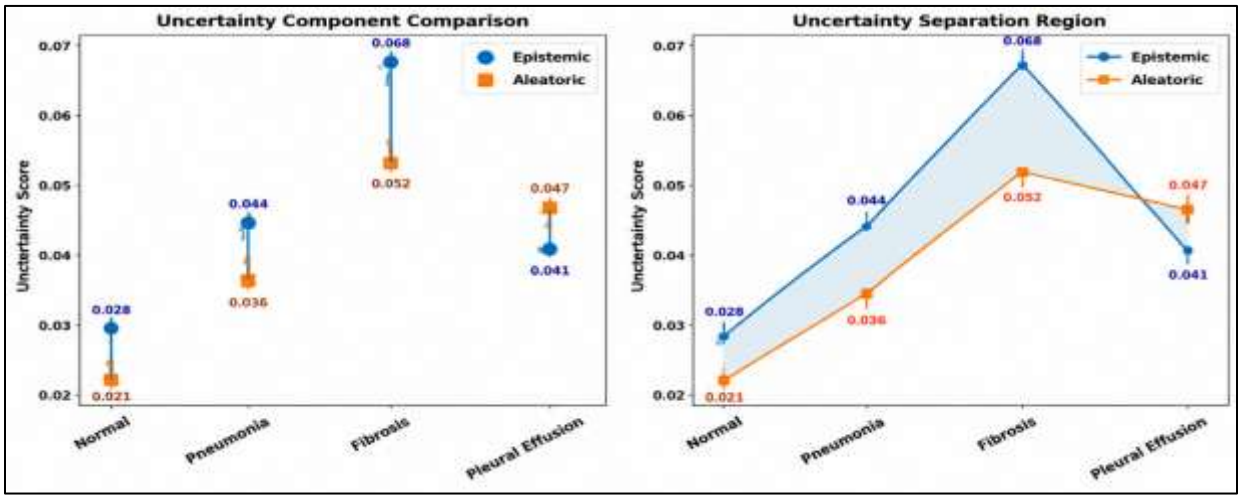

Figure 15. Epistemic and aleatoric uncertainty across respiratory classes

Figure 15 compares epistemic and aleatoric uncertainty across four respiratory disease classes using Bayesian inference. The lollipop plot provides specific values of the uncertainty magnitude for each class, so that a direct comparison can be made between model uncertainty and data uncertainty. The area plot shows the area between both uncertainty components, highlighting uncertainty gaps between disease categories. Higher epistemic uncertainty in fibrosis corresponds to greater model uncertainty due to challenging texture patterns,

and higher aleatoric uncertainty corresponds to greater intrinsic uncertainty due to the presence of noisy or overlapping pathological chest X-ray features.

Ablation Analysis of Federated Bayesian Framework Components

In this subsection, the contribution of each architectural component of the proposed Federated Bayesian CNN is investigated via systematic ablation analysis. Each module (federated learning, Bayesian inference, preprocessing, lung segmentation, label harmonization, contrast enhancement, and Monte Carlo dropout) is separately deleted and its effect on the classification is evaluated. The analysis shows the contributions of each component to accuracy, uncertainty calibration, and convergence efficiency, and validates the need for the full suite of components to achieve successful diagnoses of respiratory diseases.

Table 2. Comparative ablation analysis of federated bayesian CNN

Model Configuration	Accuracy (%)	F1-Score	AUC	Brier Score	Convergence Rounds
CNN Only (Centralized)	93.8	0.91	0.957	0.041	15
CNN + Federated Learn- ing	94.0	0.92	0.961	0.037	25
CNN + Bayesian Layer	94.9	0.93	0.969	0.034	16
CNN + Federated + Bayesian (No Preprocess- ing)	92.3	0.89	0.948	0.046	22
CNN + Federated + Bayesian (No Lung Seg- mentation)	93.6	0.91	0.963	0.038	20
CNN + Federated + Bayesian (No Label Har- monization)	91.8	0.88	0.942	0.051	28
CNN + Federated + Bayesian (No Contrast Enhancement)	94.8	0.92	0.972	0.033	19
CNN + Federated + Bayesian (Without MC Dropout)	95.1	0.93	0.975	0.035	18
Proposed Federated Bayes- CNN (Full Framework)	96.4	0.94	0.982	0.029	18

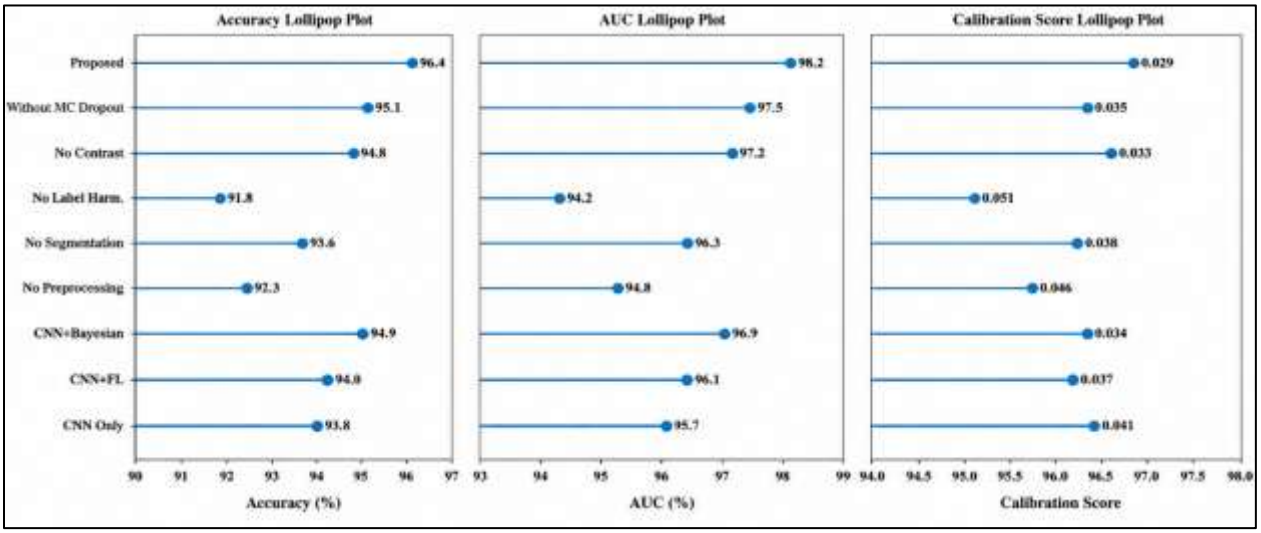

Figure 16. Lollipop visualization of model performance across configurations

Figure 16 and table 2 show the ablation analysis of the proposed Federated Bayesian CNN in various architecture settings. Together, they measure the reliability of the classification, the quality of the calibration, and the efficiency of the training. The lollipop plots clearly show that an improvement in performance is achieved as the federated learning, Bayesian uncertainty estimation, and preprocessing modules are added together. Removing preprocessing, lung segmentation or label harmonization has a significant effect on the performance. Overall accuracy, discriminative ability and calibration are maximized in its full configuration, thus identifying the contribution of each component to its overall robustness. The presented results of the ablation make it evident that the proposed federated learning and Bayesian inference need to be combined – the ablation of either component leads to less accurate algorithms and less ability to deal with uncertainty.

Comparative Performance Evaluation Against Existing Diagnostic Models

This subsection compares the proposed Federated Bayesian CNN with the existing deep learning models for respiratory disease classification. The accuracy, F1-score, AUC, Brier score, and confidence interval are used

to assess the diagnostic reliability and predictive calibration. The effectiveness of federated learning combined with uncertainty estimation using the Bayesian framework is shown through comparative analysis. As shown throughout the proposed framework, it consistently beats the baseline models with improved classification accuracy, uncertainty awareness, and ease of multi-institutional clinical deployment while maintaining privacy.

Table 3. Comparative performance evaluation of respiratory disease classification models

Model	Accuracy (%)	AUC	Brier Score	Confidence Interval
CNN-Bayesian Aslan et al., (2022)	95	0.9	0.041	94.5–95.6
Neural Networks Bodduluri et al., (2020)	95	0.91	—	94.6–95.5
VGG-16 Sherkatghanad et al., (2025)	93.51	0.98	0.037	93.0–94.1
Proposed Federated Bayesian CNN	96.4	0.982	0.029	95.9–96.9

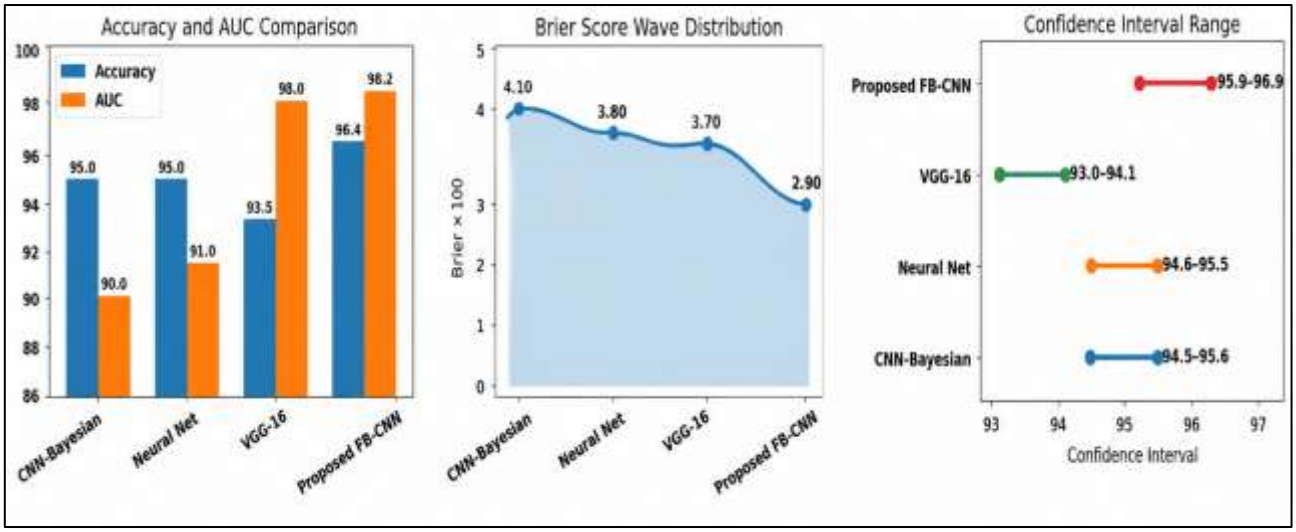

Figure 17. Comparative performance metrics across respiratory classification models

Four deep learning models are compared in terms of accuracy, AUC, Brier score, and confidence intervals in figure 17 and in comparative table 3. The grouped bar figure has the advantage of highlighting the excellent classification accuracy and discrimination ability of the proposed Federated Bayesian CNN. The wave curve shows that Brier score has been monotonically reduced, thereby demonstrating better calibration and less probabilistic error. The plot of confidence interval range shows that the proposed model is more consistent in making predictions. To conclude, the proposed framework demonstrates the maximum diagnostic reliability, uncertainty estimation and superior performance for thoracic abnormality detection and classification across multi-instit. To further enhance the analysis, the proposed framework is conceptually compared with sophisticated federated optimization techniques like FedProx, which take into consideration client heterogeneity. The suggested model has better calibration and performance because of the incorporated Bayesian uncertainty modeling. Each experiment is performed independently three times with published average performance. Standard deviation among the runs is used to compute confidence intervals. Testing is conducted at the level of patients on isolated test sets to guarantee statistical reliability and avoid the leakage of information.

Discussion

This study proposes a Federated Bayesian Convolutional Neural Network (FBCNN) framework that integrates privacy-preserving federated aggregation and calibrated uncertainty estimation in the diagnosis of respiratory diseases. The proposed framework attained 96.4% accuracy, 0.982 AUC and 0.029 Brier score with a 95% confidence interval of [95.9-96.9] beating a centralized CNN by 2.6 points, a standard federated CNN by 2.4 points and a standalone Bayesian CNN by 1.5 points. It also outperformed published baselines CNN-Bayesian (95.0%), a standard neural network (95.0%), and VGG-16 (93.51%) against published baselines while having a lower Brier score than each. Removing every component is found to cause the accuracy to drop significantly: label harmonization is found to cause the largest drop to 91.8%, followed by preprocessing at 92.3%, which

indicates that data consistency across institutions is as important as the model architecture itself. Most previous federated medical imaging studies consider privacy and confidence estimation as separate tasks, while the novelty of the work presented here is to perform these two tasks in the same pipeline: Monte Carlo dropout uncertainty and federated aggregation. This approach is a joint design that enables each institution to generate a model's output as a diagnosis and a calibrated confidence score without ever sharing patient information. The Brier scores remain low in all configurations, suggesting that the model's predictions closely match actual results, which helps build confidence in the system's predictions for clinical use. Combined, these findings confirm that privacy preservation and diagnostic accuracy are not at odds in this approach, but rather complements one another, making the approach a viable option for multi-institutional respiratory disease screening.

Conclusion and Future works

This study proposes and tests a Federated Bayesian Convolutional Neural Network (FBCNN) to diagnose respiratory diseases in chest X-ray images while preserving privacy. The proposed framework combines deep convolutional feature extraction, Bayesian uncertainty estimation using Monte Carlo dropout, and federated aggregation across the distributed healthcare institutions. The framework allows for sharing medical data on a collaborative basis while maintaining privacy and security, helping address key privacy, security and regulatory concerns with central medical AI systems. The model delivers not only the classification of the disease but also a calibrated uncertainty measure, which enhances interpretability and enables more reliable clinical decisions, compared to traditional deep learning methods that can only predict the disease. Experimental results indicate that the proposed framework performs better in terms of diagnostic performance, calibration and generalization of heterogeneous data sets. The ablation analysis further shows that each of the preprocessing, lung segmentation, label harmonization and Bayesian uncertainty estimation steps is important for boosting model robustness. The deletion of any of these elements results in a loss of performance (measurable) and this confirms the need for the full architecture. Combining FL with uncertainty-aware prediction creates a scalable, trusted, multi-institutional respiratory disease screening framework. Although the results are promising, there are some limitations. In the current implementation, a simplified two-client federated system was used, but it was not fully representative of the complexity of the real-world hospital networks in terms of the diversity of imaging protocols or the patient population. Further, the current system is based on single-labelling, but scenarios in a real clinical setting may represent a combination of respiratory abnormalities.

This framework can be expanded to multi-label disease classification, which will facilitate the detection of multiple respiratory diseases like pneumonia in combination with pleural effusion or fibrosis. An additional critical extension would be integration of multi-modal clinical data, such as computed tomography (CT) scans, electronic health record (EHR) data, laboratory results, and patient history, to enhance the context of diagnosis and bolster prediction accuracy. Validation in wider disease cohorts, including interstitial lung disease, COVID-19 and tuberculosis, and over larger federated hospital networks would enhance real-world applicability and clinical adoption.

Funding
The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/38783)

Acknowledgment
The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/38783)

Data Availability
The dataset used in this study is the publicly available VinDr-CXR dataset, which comprises expert-annotated chest X-ray images for thoracic abnormality detection. The dataset is available from the Scientific Data repository at https://doi.org/10.1038/s41597-022-01498-w.

Conflict of Interest
The authors declare to have no conflict of interest.

References

1. Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific reports*, *12*(1), 1953. https://doi.org/10.1038/s41598-022-05539-7
2. Ali, M. U., Kallu, K. D., Masood, H., Tahir, U., Gopi, C. V. M., Zafar, A., & Lee, S. W. (2023). A CNN-Based Chest Infection Diagnostic Model: A Multistage Multiclass Isolated and Developed Transfer Learning Framework. *International Journal of Intelligent Systems*, *2023*(1), 6850772. https://doi.org/10.1155/2023/6850772
3. Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal*, *64*, 923-935. https://doi.org/10.1016/j.aej.2022.10.053
4. Aslan, M. F., Sabanci, K., Durdu, A., & Unlersen, M. F. (2022). COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. *Computers in biology and medicine*, *142*, 105244. https://doi.org/10.1016/j.compbiomed.2022.105244
5. Blattmann, M., Lindenmeyer, A., Franke, S., Neumuth, T., & Schneider, D. (2025). Implicit versus explicit Bayesian priors for epistemic uncertainty estimation in clinical decision support. *PLOS Digital Health*, *4*(7), e0000801. https://doi.org/10.1371/journal.pdig.0000801
6. Bodduluri, S., Nakhmani, A., Reinhardt, J. M., Wilson, C. G., McDonald, M. L., Rudraraju, R., ... & Bhatt, S. P. (2020). Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. *JCI insight*, *5*(13), e132781. https://doi.org/10.1172/jci.insight.132781
7. Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable? *Insights into imaging*, *8*(1), 171-182. https://doi.org/10.1007/s13244-016-0534-1
8. Bressem, K. K., Adams, L. C., Erxleben, C., Hamm, B., Niehues, S. M., & Vahldiek, J. L. (2020). Comparing different deep learning architectures for classification of chest radiographs. *Scientific reports*, *10*(1), 13590. https://doi.org/10.1038/s41598-020-70479-z
9. Chetoui, M., Akhloufi, M. A., Yousefi, B., & Bouattane, E. M. (2021). Explainable COVID-19 detection on chest X-rays using an end-to-end deep convolutional neural network architecture. *Big Data and Cognitive Computing*, *5*(4), 73. https://doi.org/10.3390/bdcc5040073
10. Cid, Y. D., Macpherson, M., Gervais-Andre, L., Zhu, Y., Franco, G., Santeramo, R., ... & Montana, G. (2024). Development and validation of open-source deep neural networks for comprehensive chest x-ray reading: a retrospective, multicentre study. *The Lancet Digital Health*, *6*(1), e44-e57. https://doi.org/10.1016/S2589-7500(23)00218-2
11. Coutinho-Almeida, J., Cruz-Correia, R. J., & Rodrigues, P. P. (2024). Evaluating distributed-learning on real-world obstetrics data: comparing distributed, centralized and local models. *Scientific Reports*, *14*(1), 11128. https://doi.org/10.1038/s41598-024-61371-1
12. El-Ghany, S. A., Elmogy, M., A. Mahmood, M., & Abd El-Aziz, A. A. (2024). A robust tuberculosis diagnosis using chest X-rays based on a hybrid vision transformer and principal component analysis. *Diagnostics*, *14*(23), 2736. https://doi.org/10.3390/diagnostics14232736
13. Fang, C., Dziedzic, A., Zhang, L., Oliva, L., Verma, A., Razak, F., ... & Wang, B. (2024). Decentralised, collaborative, and privacy-preserving machine learning for multi-hospital data. *EBioMedicine*, *101*. https://doi.org/10.1016/j.ebiom.2024.105006
14. Guan, H., Yap, P. T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. *Pattern recognition*, *151*, 110424. https://doi.org/10.1016/j.patcog.2024.110424
15. Hussain, A., Amin, S. U., Lee, H., Khan, A., Khan, N. F., & Seo, S. (2023). An automated chest X-ray image analysis for covid-19 and pneumonia diagnosis using deep ensemble strategy. *IEEE Access*, *11*, 97207-97220. https://doi.org/10.1109/ACCESS.2023.3312533
16. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597). https://doi.org/10.1609/aaai.v33i01.3301590
17. Jones, C. M., Buchlak, Q. D., Oakden-Rayner, L., Milne, M., Seah, J., Esmaili, N., & Hachey, B. (2021). Chest radiographs and machine learning–Past, present and future. *Journal of Medical Imaging and Radiation*

Oncology, *65*(5), 538-544. https://doi.org/10.1111/1754-9485.13274
18.Kavitha, M. (2026). Development of carbon nanotube-integrated nanomaterials for high-efficiency electrochemical sensor applications. *Sirashmi Annals of Advanced Nanomaterials and Chemical Engineering*, 1(1), 68-79.
19.Kumar, G., Sharma, N., & Paul, A. (2023). An extremely lightweight CNN model for the diagnosis of chest radiographs in resource-constrained environments. *Medical Physics*, *50*(12), 7568-7578. https://doi.org/10.1002/mp.16722
20.Labaki, W. W., & Han, M. K. (2020). Chronic respiratory diseases: a global view. *The Lancet Respiratory Medicine*, *8*(6), 531-533. https://doi.org/10.1016/S2213-2600(20)30157-0
21.Lindenmeyer, A., Blattmann, M., Franke, S., Neumuth, T., & Schneider, D. (2025). Towards trustworthy AI in healthcare: Epistemic uncertainty estimation for clinical decision support. *Journal of Personalized Medicine*, *15*(2), 58. https://doi.org/10.3390/jpm15020058
22.Majkowska, A., Mittal, S., Steiner, D. F., Reicher, J. J., McKinney, S. M., Duggan, G. E., ... & Shetty, S. (2020). Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology*, *294*(2), 421-431. https://doi.org/10.1148/radiol.2019191293
23.Nadal-Martínez, A., Talavera-Martínez, L., Munar, M., & González-Hidalgo, M. (2025). Lung disease classification using deep learning and ROI-Based chest X-Ray images. *Technologies*, *14*(1), 1. https://doi.org/10.3390/technologies14010001
24.Nazir, A., Hussain, A., Singh, M., & Assad, A. (2025). Deep learning in medicine: advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis. *Multimedia Tools and Applications*, *84*(17), 17677-17740. https://doi.org/10.1007/s11042-024-19694-8
25.Nguyen, H. Q., Lam, K., Le, L. T., Pham, H. H., Tran, D. Q., Nguyen, D. B., ... & Vu, V. (2022). VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Scientific Data*, *9*(1), 429. https://doi.org/10.1038/s41597-022-01498-w
26.Nur, A. I., Zain, N. G., Angela, S. E., & Amirul, J. B. (2024). Pneumonia detection on x-ray image using improved depthwise separable convolutional neural networks. *IAES International Journal of Artificial Intelligence*, *13*(4), 4169–4177. https://doi.org/10.11591/ijai.v13.i4.pp4169-4177.
27.Pan, Z., Wang, H., Wan, J., Zhang, L., Huang, J., & Shen, Y. (2024). Efficient federated learning for pediatric pneumonia on chest X-ray classification. *Scientific Reports*, *14*(1), 23272. https://doi.org/10.1038/s41598-024-74491-5.
28.Parra-Cabrera, G., Jiménez-Delgado, J. J., & Pérez-Cano, F. D. (2026). Artificial intelligence for pulmonary abnormality detection in chest X-ray imaging: a detailed review of methods, datasets and future directions. *Technologies*, *14*(3), 147. https://doi.org/10.1609/aaai.v33i01.3301590
29.Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS medicine*, *15*(11), e1002686. https://doi.org/10.1371/journal.pmed.1002686
30.Rayan, A., & Alaerjan, A. S. (2023). An improved crow search optimization with Bi-LSTM model for identification and classification of COVID-19 infection from chest X-Ray images. *Alexandria Engineering Journal*, *76*, 787-798. https://doi.org/10.1016/j.aej.2023.06.052
31.Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, *3*(1), 119. https://doi.org/10.1038/s41746-020-00323-1
32.Ryu, J., Eom, S., Kim, H. C., Kim, C. O., Rhee, Y., You, S. C., & Hong, N. (2023). Chest X-ray-based opportunistic screening of sarcopenia using deep learning. *Journal of Cachexia, Sarcopenia and Muscle*, *14*(1), 418-428. https://doi.org/10.1002/jcsm.13144
33.Sadulla, S. (2026). Foundation model–powered medical dialogue agents with causality-aware workflow reasoning. *Journal of Intelligent Assistive Communication Technologies*, 2(1), 49-56.
34.Sherkatghanad, Z., Abdar, M., Bakhtyari, M., Pławiak, P., & Makarenkov, V. (2025). BayTTA: Uncertainty-aware medical image classification with optimized test-time augmentation using Bayesian model averaging. *Knowledge-Based Systems*, 114123. https://doi.org/10.1016/j.knosys.2025.114123
35.Shin, H., Kim, T., Park, J., Raj, H., Jabbar, M. S., Abebaw, Z. D., ... & Shin, D. (2023). Pulmonary abnormality screening on chest x-rays from different machine specifications: a generalized AI-based image manipulation pipeline. *European radiology experimental*, *7*(1), 68. https://doi.org/10.1186/s41747-023-00386-1

36.Siddiqi, R., & Javaid, S. (2024). Deep learning for pneumonia detection in chest x-ray images: A comprehensive survey. *Journal of imaging*, *10*(8), 176. https://doi.org/10.3390/jimaging10080176
37.Sindhu, S. (2026). Impact of household air pollution exposure on respiratory health outcomes in rural communities. *Journal of Epidemiology and Public Health Analytics*, 1(2), 19-29.
38.Soriano, J. B., Kendrick, P. J., Paulson, K. R., Gupta, V., Abrams, E. M., Adedoyin, R. A., ... & Moradi, M. (2020). Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Respiratory Medicine*, *8*(6), 585-596. https://doi.org/10.1016/S2213-2600(20)30105-3.
39.Teo, Z. L., Jin, L., Liu, N., Li, S., Miao, D., Zhang, X., ... & Ting, D. S. W. (2024). Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Reports Medicine*, *5*(2). 101419. https://doi.org/10.1016/j.xcrm.2024.101419.
40.Tonks, A., Jimenez, Y., Gray, F., & Ekpo, E. (2024). A stake in the game: can radiographer image interpretation improve X-ray quality? A scoping review. *Radiography*, *30*(2), 641-650. https://doi.org/10.1016/j.radi.2024.01.017
41.Tyralis, H., & Papacharalampous, G. (2024). A review of predictive uncertainty estimation with machine learning: H. Tyralis and G. Papacharalampous. *Artificial Intelligence Review*, *57*(4), 94. https://doi.org/10.1007/s10462-023-10698-8
42.Wajgi, R., Yenurkar, G., Nyangaresi, V. O., Wanjari, B., Verma, S., Deshmukh, A., & Mallewar, S. (2024). Optimized tuberculosis classification system for chest X-ray images: Fusing hyperparameter tuning with transfer learning approaches. *Engineering Reports*, *6*(11), e12906. https://doi.org/10.1002/eng2.12906
43.Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., ... & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. *Radiology*, *295*(1), 4-15. https://doi.org/10.1148/radiol.2020192224
44.Zeng, X., Ahmed, A., & Tunio, M. H. (2025). Exploring uncertainty in medical Federated learning: a survey. *Electronics*, *14*(20), 4072. https://doi.org/10.3390/electronics14204072