ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

ADVANCED MACHINE LEARNING FRAMEWORKS FOR PNEUMONIA DIAGNOSIS: INTEGRATING MULTIMODAL SIGNALS AND PRIVACY-PRESERVING TECHNIQUES

Faizan Abbas , Medical Student, Samarkand state medical university, Uzbekistan, faizanmughalg96@gmail.com

Arshad, Medical Student, Samarkand state medical university, Uzbekistan, amd21726@gmail.com

Wajiha Batool, Medical Student, Samarkand state medical university, Uzbekistan, Wajihabatool667@gmail.com

Saurabh Kumar Kushwaha , Medical Student, Samarkand state medical university, Uzbekistan Shraykush7777@gmail.com

Abstract

Pneumonia remains a leading cause of morbidity and mortality worldwide, disproportionately affecting vulnerable populations such as young children and the elderly. Traditional diagnostic workflows, which heavily rely on manual triage and resource-intensive imaging assessments, are increasingly strained by rising patient volumes in emergency departments. Recent advancements in artificial intelligence have introduced promising solutions, ranging from acoustic analysis of digital stethoscope recordings to deep learning models for medical image segmentation. This paper synthesizes current research to propose a comprehensive, multimodal diagnostic framework that leverages both audio-visual data and federated learning techniques. By bridging the gap between automated triage screening, weakly supervised image localization, and privacy-preserving model training, this study aims to provide a scalable blueprint for accelerating pneumonia diagnosis while maintaining rigorous ethical and clinical standards.

Introduction

Pneumonia is a critical respiratory illness characterized by the infection of the lung alveoli, representing a substantial burden on global healthcare systems. Environmental, genetic, and family background factors contribute significantly to the spread of this disease, which is particularly devastating for children under twelve years old (Sufahani et al., 2012). In regions like Tawau, Malaysia, as well as in the United States where tens of thousands die annually, rapid and accurate clinical intervention is

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

paramount (Sufahani et al., 2012). Furthermore, the steady rise in emergency department visits has overloaded traditional clinical workflows, leading to prolonged wait times and delayed treatment administration (Lu, 2023). Consequently, there is an urgent need for automated systems capable of accurately diagnosing pneumonia and assessing its severity at the point of care.

The scope of this paper encompasses the development and integration of computational methods for pneumonia detection, focusing on both acoustic and radiological modalities. The primary problem lies in synthesizing diverse biological signals—such as digital stethoscope audio, chest X-rays (CXR), and computed tomography (CT) scans—into a unified, efficient diagnostic pipeline. While deep learning has revolutionized medical data analysis, deploying these technologies in real-world clinical environments remains highly challenging. First, existing clinical workflows rely heavily on manual triage assessments, which are impeded by limited human workload and can lead to invasive over-testing or inaccurate initial diagnoses (Lu, 2023). Second, many modern deep learning solutions demand massive amounts of meticulously annotated, pixel-level data, which is expensive to acquire, while centralized data collection methods inherently compromise sensitive patient privacy (Wang et al., 2025)(Shahi & Bagale, 2025).

To address these critical shortcomings, this paper presents a conceptual synthesis of state-of-the-art machine learning techniques tailored for respiratory disease detection. Specifically, our contributions to the field are defined as follows:

We propose a hypothetical multimodal diagnostic framework that integrates acoustic features from digital stethoscopes with weakly supervised imaging models, aiming to improve triage efficiency without relying on costly pixel-level annotations.We outline a secure deployment strategy utilizing federated few-shot learning and differential privacy mechanisms, ensuring that multi-institutional diagnostic models can be trained without exposing sensitive patient records to gradient leakage.

Related Work

Acoustic-Based Diagnostic Models

The analysis of body sounds via digital stethoscopes has emerged as a non-invasive frontline diagnostic tool. Researchers have proposed methods using Empirical Mode Decomposition (EMD) and spectral analysis to isolate clinically relevant biosignals from cardiovascular and respiratory acoustic data, achieving high accuracy in classifying diseases like pneumonia and chronic obstructive pulmonary disease (Casado et al., 2023). Similarly, systems like VoxMed utilize an Audio Spectrogram

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

Transformer (AST) coupled with a 1-D Convolutional Neural Network (CNN) to provide healthcare professionals with near-instantaneous respiratory assessments (Mundra et al., 2024). Other approaches, such as the Fused Audio Instance and Representation (FAIR) method, construct joint feature vectors from various body sounds—including cough, breath, and speech—using self-attention mechanisms on both waveform and spectrogram representations (Truong et al., 2022). While these acoustic models are highly accessible and fast, their weakness lies in their susceptibility to ambient clinical noise. Compared to these unimodal audio systems, our proposed work seeks to fuse acoustic screening with downstream radiological verification to enhance overall diagnostic confidence.

Medical Imaging and Weak Supervision

Radiological imaging remains the gold standard for confirming pneumonia and assessing lung infection severity. Extensive research has been dedicated to deep learning models for segmenting lung infections from CT images, employing techniques such as boundary-guided semantic learning to handle blurred boundaries and complex background interference (Cong et al., 2022). Multi-task encoder-decoder networks have also been utilized to overcome data shortages by segmenting both the lung regions and the specific infections (Elharrouss et al., 2020). Furthermore, Transformer-based architectures with cross-gated attention mechanisms have demonstrated robust performance in predicting infection severity across both CT scans and CXRs (Slika et al., 2025). However, obtaining the pixel-level annotations required for these segmentation tasks is prohibitively expensive. To mitigate this, weakly supervised frameworks utilizing Gradient-weighted Class Activation Mapping (Grad-CAM) have been developed to localize pneumonia using only image-level labels (Shahi & Bagale, 2025). Our approach builds upon this weakly supervised paradigm, adopting it as the core imaging module to bypass the prohibitive costs associated with pixel-level labeling while maintaining clinical explainability.

Automated Triage and Privacy-Preserving Systems

The integration of machine learning into emergency department triage represents a critical step in alleviating hospital congestion. Models such as TriNet have been developed to automate first-line screening for conditions like pneumonia and urinary tract infections, achieving high positive predictive values and reducing the risk of over-testing (Lu, 2023). Concurrently, the deployment of such models across various institutions raises severe patient privacy concerns regarding direct data sharing. To counter this, researchers have proposed federated few-shot learning frameworks equipped with meta-stochastic gradient descent and differential privacy noise (Wang

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

et al., 2025). This approach prevents the reconstruction of medical images from gradient leakage while handling the scarcity of high-quality labeled datasets. In contrast to standalone triage or privacy algorithms, our work theoretically combines the triage efficiency of models like TriNet with differential privacy mechanisms, ensuring that early screening tools can be continuously improved across hospital networks without compromising data security.

Method/Approach

To realize a robust and scalable pneumonia diagnostic system, we propose a structured, multimodal framework that sequentializes patient screening, diagnosis confirmation, and privacy-preserving model updates. The system leverages both acoustic signals from the triage stage and radiological images from downstream testing. By doing so, it acts as a comprehensive pipeline capable of operating in heavily burdened clinical environments.

The proposed framework consists of the following numbered pipeline:

Initial Triage Screening (Audio Module): Digital stethoscope recordings are captured during the initial nursing assessment. These audio signals are processed using an Audio Spectrogram Transformer (AST) for robust feature extraction, identifying immediate respiratory anomalies in seconds.

Severity and Localization (Imaging Module): If the acoustic screening indicates high pneumonia probability, a CXR or CT scan is ordered. This image is processed by a Vision Transformer relying on a weakly supervised learning paradigm, generating Grad-CAM heatmaps to highlight infected regions using only image-level training labels.

Federated Privacy-Preserving Update: Local inferences and minimal parameter updates are aggregated. Standard Gaussian noise is injected into the local gradients before they are transmitted to a central server, ensuring differential privacy and preventing data reconstruction.

Key design choices in this framework address specific clinical bottlenecks. The choice of an AST for audio processing is driven by its ability to capture complex temporal-frequency patterns in respiratory sounds, which outpaces traditional auscultation (Mundra et al., 2024). Furthermore, incorporating weakly supervised learning for the imaging module circumvents the labor-intensive nature of medical data annotation, a major hurdle in resource-constrained settings (Wang et al., 2025)(Shahi & Bagale, 2025). Finally, the federated learning architecture with differential privacy is an

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

essential structural choice, as patient privacy concerns and strict healthcare regulations make centralized medical data aggregation impractical (Wang et al., 2025).

The evaluation plan for this proposed framework involves simulating a multi-institutional clinical trial using distinct datasets. For the acoustic triage module, we hypothesize utilizing the publicly available ICBHI dataset to measure classification accuracy and processing speed (Mundra et al., 2024). For the imaging module, hypothetical evaluation would involve large-scale, image-level annotated CXR datasets, measuring the intersection-over-union (IoU) of Grad-CAM bounding boxes against a small expert-annotated validation set. The primary metrics for overall system success will include Positive Predictive Value (PPV) at triage, Area Under the Receiver Operating Characteristic Curve (AUC) for disease classification, and the reduction in average patient wait times in a simulated emergency department environment.

Discussion

The practical implications of deploying this multimodal framework in clinical settings are substantial. By shifting the initial diagnostic burden to automated acoustic screening, emergency departments can drastically reduce the time it takes for a patient to move from the waiting room to receiving targeted care (Lu, 2023). Portable digital stethoscopes linked to UI-assisted classifiers can empower triage nurses to make faster, evidence-based decisions, minimizing unnecessary downstream radiological testing for low-risk patients. Consequently, this leads to optimized hospital resource allocation and improved patient outcomes during periods of high clinical demand, such as viral pandemics.

Despite its potential, the proposed system faces several limitations and failure modes. First, domain shift remains a significant challenge; acoustic models trained on data from specific regions or devices may experience performance degradation when deployed in hospitals with different digital stethoscope hardware or ambient noise levels. Second, clinical environments, particularly emergency departments, are highly chaotic, and overlapping background noises could severely corrupt the audio spectrograms, leading to false triage predictions. Third, the reliance on weakly supervised Grad-CAM heatmaps for image localization can sometimes yield diffuse or imprecise boundary definitions compared to fully supervised segmentation models, potentially causing clinicians to misjudge the exact extent of the lung infection.

Ethical considerations and risks must also be carefully managed before real-world deployment. First, although federated learning with standard Gaussian noise mitigates gradient leakage, sophisticated adversarial attacks could theoretically still extract

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025

sensitive demographic or pathological information from the aggregated model weights, posing a risk to patient confidentiality (Wang et al., 2025). Second, there is a risk of algorithmic bias; if the federated nodes predominantly represent specific demographics, the model may underperform on underrepresented racial, age, or socioeconomic groups, leading to inequitable healthcare delivery.

Future work should aim to refine and expand upon this baseline architecture. One immediate avenue for future research is the integration of longitudinal electronic health records (EHR) into the multimodal fusion module, allowing the model to weigh acoustic and visual data against a patient's historical health background and genetic predispositions (Sufahani et al., 2012). Additionally, future engineering efforts should focus on optimizing these complex Transformer-based models for edge-computing devices, enabling real-time, offline pneumonia diagnosis in rural or resource-constrained clinics lacking stable internet connectivity for federated server communication.

Conclusion

This paper has explored the pressing need for advanced, automated systems in diagnosing pneumonia and assessing lung infection severity. By reviewing cutting-edge acoustic classifiers, weakly supervised deep learning models for medical imaging, and triage automation techniques, we highlighted the limitations of existing isolated and data-heavy approaches. In response, we proposed a comprehensive, multimodal diagnostic framework that logically bridges automated audio triage with weakly supervised image localization, all governed by a privacy-preserving federated learning architecture.

The integration of artificial intelligence into respiratory disease diagnostics holds transformative potential for modern healthcare. By alleviating emergency department bottlenecks and minimizing the necessity for exhaustive pixel-level data annotations, the proposed methodologies offer a scalable path forward. Ultimately, refining these machine learning tools and rigorously addressing their limitations and ethical implications will be critical in ensuring they serve as reliable, equitable aids that meaningfully improve patient survival and care quality.

References

Sufahani, Suliadi F., Razali, Siti N. A. Mohd, Mormin, Mohammad F., & Khamis, Azme (2012). An Analysis of the Prevalence of Pneumonia for Children under 12 Year Old in Tawau General Hospital, Malaysia. https://arxiv.org/pdf/1205.2109v1

ISSN: 3060-4737 Volume 3, Issue 5 IF(Impact Factor) 10.98 / 2025
Lu, Stephen Z. (2023). Screening of Pneumonia and Urinary Tract Infection at Triage using TriNet. https://arxiv.org/pdf/2309.02604v1

Wang, Ming, Duan, Zhaoyang, Xue, Dong, Liu, Fangzhou, & Zhang, Zhongheng (2025). An Enhanced Privacy-preserving Federated Few-shot Learning Framework for Respiratory Disease Diagnosis. https://arxiv.org/pdf/2507.08050v1

Shahi, Kiran, & Bagale, Anup (2025). Weakly Supervised Pneumonia Localization from Chest X-Rays Using Deep Neural Network and Grad-CAM Explanations. https://doi.org/10.54364/JAIAI.2024.1126

Casado, Constantino Álvarez, Cañellas, Manuel Lage, Pedone, Matteo, Wu, Xiaoting, Nguyen, Le, & López, Miguel Bordallo (2023). Respiratory Disease Classification and Biometric Analysis Using Biosignals from Digital Stethoscopes. https://arxiv.org/pdf/2309.07183v2

Mundra, Paridhi, Sharma, Manik, Chaudhuri, Yashwardhan, Phukan, Orchid Chetia, & Buduru, Arun Balaji (2024). VoxMed: One-Step Respiratory Disease Classifier using Digital Stethoscope Sounds. https://arxiv.org/pdf/2407.18926v1

Truong, Tuan, Lenga, Matthias, Serrurier, Antoine, & Mohammadi, Sadegh (2022). Fused Audio Instance and Representation for Respiratory Disease Detection. https://arxiv.org/pdf/2204.10581v4

Cong, Runmin, Zhang, Yumo, Yang, Ning, Li, Haisheng, Zhang, Xueqi, Li, Ruochen, Chen, Zewen, Zhao, Yao, & Kwong, Sam (2022). Boundary Guided Semantic Learning for Real-time COVID-19 Lung Infection Segmentation System. https://arxiv.org/pdf/2209.02934v1

Elharrouss, Omar, Subramanian, Nandhini, & Al-Maadeed, Somaya (2020). An encoder-decoder-based method for COVID-19 lung infection segmentation. https://arxiv.org/pdf/2007.00861v2

Slika, Bouthaina, Dornaika, Fadi, Bougourzi, Fares, & Hammoudi, Karim (2025). Lung Infection Severity Prediction Using Transformers with Conditional TransMix Augmentation and Cross-Attention. https://arxiv.org/pdf/2510.06887v1