Privacy Preserving Federated Learning for Multi -Disease Chest X -Ray Classification

Sakshi Bhushan
Department of Electronics and Communication Engineering, IGDTUW, Delhi, India

Samridhi Tripathi
Department of Electronics and Communication Engineering, IGDTUW, Delhi, India

Shanti Kumari
Department of Electronics and Communication Engineering, IGDTUW, Delhi, India

Greeshma Arya
Department of Electronics and Communication Engineering, IGDTUW, Delhi, India

Abstract

The rapid growth of deep learning in medical diagnostics has significantly enhanced clinical workflows, particularly in radiology. However, centralized machine learning introduces substantial risks related to patient confidentiality, data governance, and regulatory compliance. Federated Learning (FL) offers a decentralized training approach that eliminates the need to pool data across institutions. Nevertheless, raw gradient updates may still leak sensitive information. To address this limitation, this study proposes a hybrid framework that integrates Federated Averaging (FedAvg) with Differential Privacy (DP) through custom gradient clipping and noise injection mechanisms.

The framework incorporates performance-based weighted aggregation, optimized noise calibration ($\sigma = 0.05$), gradient clipping ($C = 1.2$), and adaptive learning rate scheduling to effectively balance privacy preservation and model utility. This paper presents a five-class chest X-ray classification system that achieves 80% accuracy while maintaining differential privacy with $\varepsilon = 5.0$. The dataset comprises five disease categories (COVID-19, Normal, Pneumonia, Pneumothorax, and Tuberculosis) distributed across five simulated hospital clients, where local Convolutional Neural Networks (CNNs) are trained independently, and a central server aggregates model parameters over multiple communication rounds. Differential Privacy enhances local training by injecting statistically calibrated noise to safeguard individual patient identities. Experimental evaluation confirms that the proposed system achieves robust classification performance while providing strong privacy guarantees. The analysis further examines convergence behaviour, training dynamics, and the trade-off between privacy noise levels and predictive accuracy.