Chapter 7
Privacy-Preserving Federated Learning for Multi-Hospital Patient Data Integration

Arul Selvam P.
http://orcid.org/0000-0002-5822-4360
Hindusthan College of Engineering and Technology, India

Tamije Selvy P.
Hindusthan College of Engineering and Technology, India

ABSTRACT

The exponential growth of medical data across hospitals presents significant opportunities for predictive healthcare analytics, yet stringent privacy laws and data fragmentation hinder collaborative research. This chapter introduces a Privacy-Preserving Federated Learning (PP-FedAvg) framework that enables multiple hospitals to train machine-learning models collectively without sharing raw patient records. The system architecture integrates secure aggregation, homomorphic encryption, and differential privacy to ensure confidentiality during model updates, while blockchain-based accountability enhances transparency and trust. Experiments conducted on the MIMIC-III and MedMNIST datasets demonstrate that the proposed framework achieves near-centralized accuracy while significantly reducing privacy loss and communication cost. These results confirm that privacy-preserving federated learning can enable secure, regulation-compliant, and scalable collaboration among healthcare institutions.

DOI: 10.4018/979-8-3373-7426-0.ch007

1. INTRODUCTION

The integration of medical data across multiple hospitals has emerged as a cornerstone of next-generation healthcare analytics, promising to enhance diagnostic accuracy, treatment personalization, and population-level disease surveillance. In recent years, healthcare institutions have transitioned toward data-centric operations, where patient records, medical imaging, laboratory results, and sensor data from wearables are continuously collected and digitized. The volume, velocity, and variety of this information have created immense opportunities for artificial intelligence (AI) to transform clinical decision-making. However, despite technological progress, most healthcare data remain confined within isolated institutional silos, hindering the realization of collective intelligence that could revolutionize patient care (Rieke et al., 2020).

The fragmentation of healthcare data arises from both technical and organizational barriers. Hospitals often employ incompatible Electronic Health Record (EHR) systems developed under disparate standards, leading to inconsistent data schemas, missing fields, and interoperability bottlenecks. Furthermore, competitive and ethical considerations discourage open data sharing. Institutions may fear reputational risk, regulatory scrutiny, or the potential misuse of sensitive patient information by third parties. Consequently, the majority of healthcare data are underutilized, with only a fraction contributing to collaborative research or AI model development. The challenge, therefore, is not merely collecting more data but enabling secure, privacy-preserving integration of existing data sources across diverse clinical environments.

This challenge is compounded by stringent legal frameworks that govern data privacy. Regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union impose strict constraints on data handling, secondary use, and cross-border transfer (Brisimi et al., 2018). These laws are designed to protect patients' rights to confidentiality and consent, yet they inadvertently restrict collaborative model development. For instance, under HIPAA, any transfer of identifiable patient data requires explicit authorization, while GDPR mandates data minimization and transparency obligations that limit centralized data aggregation. Similar frameworks are emerging globally, including India's Digital Personal Data Protection Act (DPDP 2023), which further emphasizes informed consent and lawful data processing. Together, these frameworks have created a regulatory environment that prioritizes data sovereignty but poses operational challenges for AI-driven collaboration.

Traditional centralized learning architectures, in which all data are consolidated into a single repository for model training, directly conflict with these privacy constraints. Beyond compliance issues, centralized approaches introduce serious security