

**Literature Review**

Federated Learning in Healthcare

Full Name: Bhumika Shah

Student Number: 2331422

Course: BSc(hons) Computer Science

University email address: B.Shah6@wlv.ac.uk

Date of submission:

# Introduction

## Aims and Objectives

**Aims**

The primary aim of this review is to explore and critically analyze the application of Federated Learning (FL) in healthcare, focusing on its potential to address data privacy challenges, improve collaborative machine learning, and enhance the efficiency of healthcare systems. This review seeks to assess the current advancements, challenges, and future directions of FL in real-world healthcare scenarios.

**Objectives**

* Understanding the Fundamentals of Federated Learning (FL).
* To evaluate the applicability of Federated Learning in handling sensitive healthcare data
* Applications and Use Cases of FL in Real-World Scenarios.
* To explore advancements in privacy-preserving techniques
* Decentralized Learning and Collaborative Model Training used in healthcare.
* To evaluate future Prospects and Research Opportunities in FL.

# Academic Questions

** What are the key challenges in implementing Federated Learning (FL) in healthcare?**

 **How does Federated Learning ensure data privacy and security while maintaining the accuracy and efficiency of machine learning models in decentralized settings?**

 **What are the future prospects of FL in healthcare, and how can this emerging technologies in IoMT enhance its scalability and applicability?**

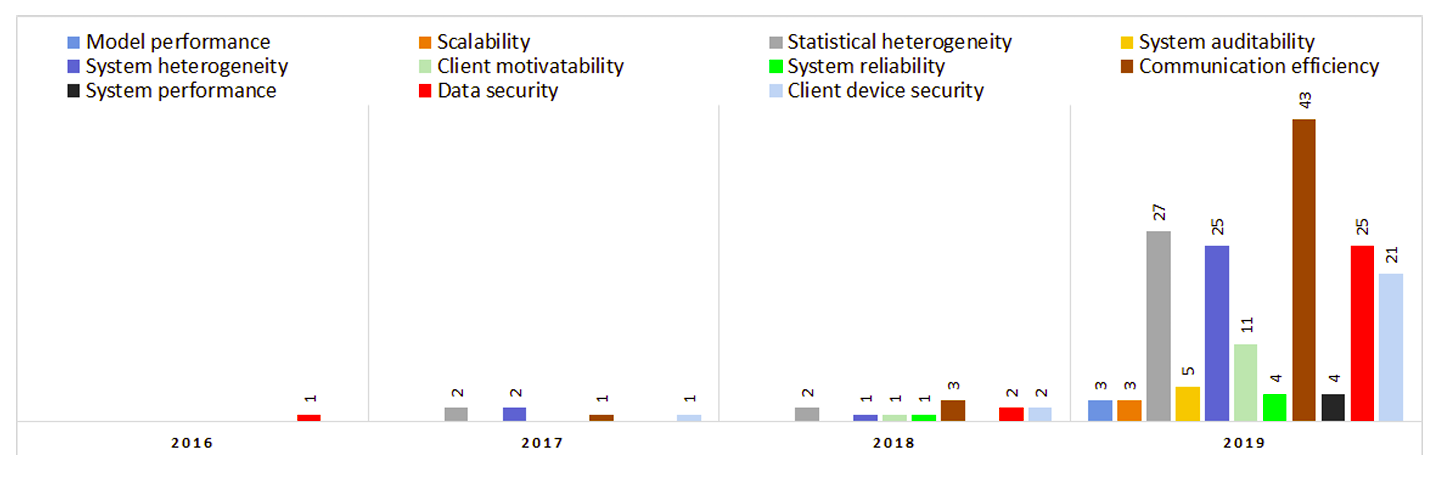
# Review of Literature

**A Systematic Literature Review on Federated Machine Learning: From A Software Engineering Perspective**

The paper by (Sin Kit Lo, 2021) explores the application and challenges of Federated Machine Learning systems. It discusses the integration of Federated Learning (FL) in real-world systems. FL facilitates a collaborative learning across various distributed devices and organizations without data centralization enabling privacy.

It reviews 231 studies and addresses questions to analyze challenges associated with FL integration in healthcare, IoT and mobile systems to emphasize data privacy. Studies existing tools and frameworks to support the process of design, development and deployment of Federated Machine Learning Systems.

The study highlights challenges of handling non-IID (non-identically distributed) data, a frequent issue in distributed systems such as healthcare and finance. Key challenges that are mentioned include, high communication overhead, security risks – adversarial attacks, issues with scalability while dealing with larger clients, and lack of effective monitoring and testing tools for FL systems. It focuses on techniques for preserving privacy, which includes differential privacy – added noise to data and secure multi-party computation where each original input remains private. For frequent share of updates in FL, use of model compressions to reduce data size sent between clients and the server. The authors also focus on blockchain – a decentralized server ensuring scalability and tamper-proof systems.



1 Research area trend - challenges of federated learning (Sin Kit Lo, 2021)

This paper is relevant for research on foundational FL systems, techniques of preserving data privacy and addressing challenges.

**Federated Learning for Smart Healthcare: A Survey**

This paper (Dinh C. Nguyen, 2022) reviews Federated Learning (FL) in the context of smart healthcare systems with the focus on FL that allows data sharing for learning without compromising patients’ information. The authors also focus on privacy concerns while implementing collaborative learning across various data sources.

It underlines the current implementation of FL in smart healthcare systems also indulging in techniques and frameworks well-suited for healthcare sector. It scopes to highlight applications of FL in smart healthcare, diagnosis, personalized medication and with stress on privacy regulations.

Applications of FL in healthcare spans widely for electronic heath record (EHR) management, disease diagnosis, personalized treatment plans and hospital management systems. However, along with applications multiple challenges are significant, data heterogeneity – a major issue in data formats, and clinical practices worldwide which create barriers to train unified models. Various regulatory compliance such as GDPR and HIPPA are to be followed for data use. Furthermore, communication and security threats create a major concern, this is especially true in resource constrained environments such as those found in Internet of Medical Things (IoMT) devices that must efficiently support frequent model updates.

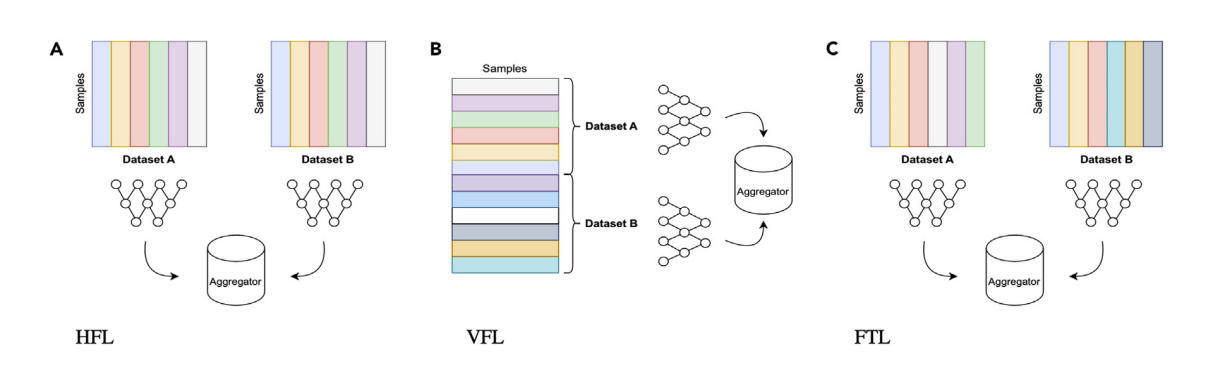
The survey presents FL as a necessary enabler to move smart healthcare systems forward, and identifies areas that need further work. For example, in disease diagnosis, FL facilitates joint training of diagnostic models amongst different institutions that can use varied data to improve prediction accuracy.

**Recent methodological advances in federated learning for healthcare** (Fan Zhang, 2024)

The paper explores innovative approaches in Federated Learning (FL) designed to tackle healthcare challenges such as privacy, data heterogeneity, and class imbalance. The study reviews 89 works, focusing on advancements in key FL components: local data processing, optimization, communication, aggregation, and redistribution.

It highlights recent advancements, and how these advancements also address the challenges in healthcare. It scopes to understand the application of FL in diagnostics, remote monitoring and medical research including innovations for scalability and privacy in these systems

Some of the important contributions include; differential privacy as well as encryption to enhance security of sensitive health information. It also presents methods for managing data heterogeneity through more sophisticated aggregation approaches and enhancements to them to cope with restricted resources. Some of the real-world applications relate to disease diagnosis, multi-institutional collaborations include remote monitoring in healthcare field. The paper recognizes some gaps such as scarce discussions about Vertical FL and the lack of research outcomes on the implementation of FL in real-world scenarios and suggests some practical advices to standardize and improve the FL frameworks.



2 Types of FL (Fan Zhang, 2024)

This review offers significant information towards addressing systematic hurdles and offers a direction to achieving large-scale and resilient healthcare solutions with FL.

**Federated and Distributed Learning Applications for Electronic Health Records and Structured Medical Data** (Li, 2023)

Structured clinical data is explored in the paper as a use case for Federated Learning (FL). The authors’ analysis of 34 studies showed that FL could solve privacy problems while facilitating collaborative learning between healthcare institutions. There were only two studies involved in performing vertical FL, where different datasets have the same samples but different features, and most of the studies involved horizontal FL.

Key challenges mentioned by authors include data heterogeneity, limited model evaluation against local analyses and limited exploration of clinical benefits. FL is shown to have applications in prediction tasks, association studies and phenotyping which can be used in clinical decision making. Data harmonization and standardization is required for effective implementation. In the paper it is also advocated for integration of engineering and statistic-based FL frameworks for generalizability.

This FL can be expanded in a resource scarce environment through adaptive optimization and lightweight frameworks as highlighted by the study. Even though the paper presents a good background for the application of FL enhanced systems in healthcare applications, it has some limitations: There is a lack of empirical performance reviews and a deeper investigation of the area between other disciplines. As a result, it is a key resource to promote FL methodologies for addressing practical and clinical questions arising in healthcare research.

**Survey of Medical Applications of Federated Learning** (Geunho Choi, 2024)

Federated Learning (FL) is transforming healthcare by addressing data privacy concerns while enabling multi-institutional collaboration in medical research. A survey investigating FL’s medical applications addressing the data privacy concern in medical research and allowing coordination. FL's medical applications are surveyed and its findings categorized with respect to data types, diseases, and methodologies from 58 studies. In privacy sensitive domains, FL allows institutions to learn machine learning models collaboratively without sharing sensitive patient data, making it particularly relevant. The majority of work has been on image data (e.g. radiology) and diseases (cancer, COVID-19) using neural networks as the model type. This emphasizes the early-stage FL applications in healthcare based on the heavy dependence on open datasets.

Critical challenges of the study include data heterogeneity among institutions and security risks including data leaks and system attacks. To deal with these problems, proposed solutions are differential privacy, encryption, and secure multi-party computation. Even now, FL is not fully utilized for non-disease common data types, which leaves vast territories of FL's application to be mined. For example, such research is limited for structured data, wearable devices, or rare diseases to explore in the future.

The paper acknowledges the transformative potential of FL in healthcare by reviewing current applications and challenges in FL, and shows how it can provide secure, collaborative medical research across institutions. More research is requested in order to broaden FL's impact aims, and to fill gaps in healthcare applications.

# References