The College of Graduate Studies and the College of Information Technology Cordially Invite You to a
PhD Thesis Defense

Entitled

ENHANCING HEALTH ANALYTICS: SECURE AND PRIVATE FEDERATED LEARNING SOLUTIONS

by

Nisha Thorakkattu Madathil
Faculty Advisor
Dr. Saed Alrabaee
College of Information Technology

Date & Venue

Monday, 15 April 2024
11:00 - 13:50
Room H1-0066

Abstract

Federated Learning (FL) is a collaborative approach permitting individuals to jointly train a model without sharing their local datasets with others. FL utilizes decentralized data sources to train machine learning models while protecting privacy, has emerged as a promising method. This holds especially true in medical contexts, where the confidentiality of data is critical. FL permits the utilization of heterogeneous datasets from a variety of healthcare organizations while maintaining patient confidentiality. It also plays a crucial role in advancing medical research and healthcare services while adhering to data distribution and compliance requirements. The primary challenges within federated healthcare encompass privacy preservation among sensitive distributed data, ensuring efficient communication, addressing data heterogeneity, and ultimately guaranteeing model accuracy. To tackle these issues, this thesis proposes solutions aimed at enhancing efficiency, accuracy, and privacy in FL systems. This research centrally employs synthetic data for generating initial model parameters and utilizes federated feature selection methods, effectively improving model convergence and reducing communication and computational overhead. Additionally, synthetic data augmentation mitigates data heterogeneity issues, enhancing model performance. Finally, this research combines differential privacy with synthetic data and Homomorphic Encryption, ensuring secure computations on encrypted data and maintaining data privacy during model aggregation. The experimental results demonstrate the following: a) the proposed solution for generating the initial model reduces the number of iterations by more than a factor of 4 compared to the baseline FL algorithm, while maintaining model accuracy. b) This thesis proposed a federated feature selection method that improves algorithm efficiency by a factor ranging from 4 (Backward Elimination) to 14 (Threshold variance) and enhances accuracy. c) The proposed solution utilizes synthetic data augmentation in heterogeneous data to significantly boost accuracy. d) Finally, this research presents a hybrid privacy mechanism to enhance data privacy while maintaining efficiency and accuracy. In summary, this thesis offers a comprehensive framework for enhancing FL systems, with a specific focus on advancing privacy, efficiency, accuracy, and overall performance.

Keywords: Federated learning, privacy and security, synthetic data, non-IID data, federated feature selection, homomorphic encryption, differential privacy.