

PREDICTING BREAST CANCER SUBTYPE FOR PERSONALIZED TREATMENT USING FEDERATED LEARNING

Tayyaba Khalil, Dr. Kevin Bryson, Dr. Christos Anagnostopoulos



Motivation

- Medical data is highly **sensitive and distributed** across hospitals
- Data sharing is limited due to:
 - Privacy concerns
 - Legal regulations (e.g., GDPR)
- Traditional Machine Learning (ML) requires centralized datasets

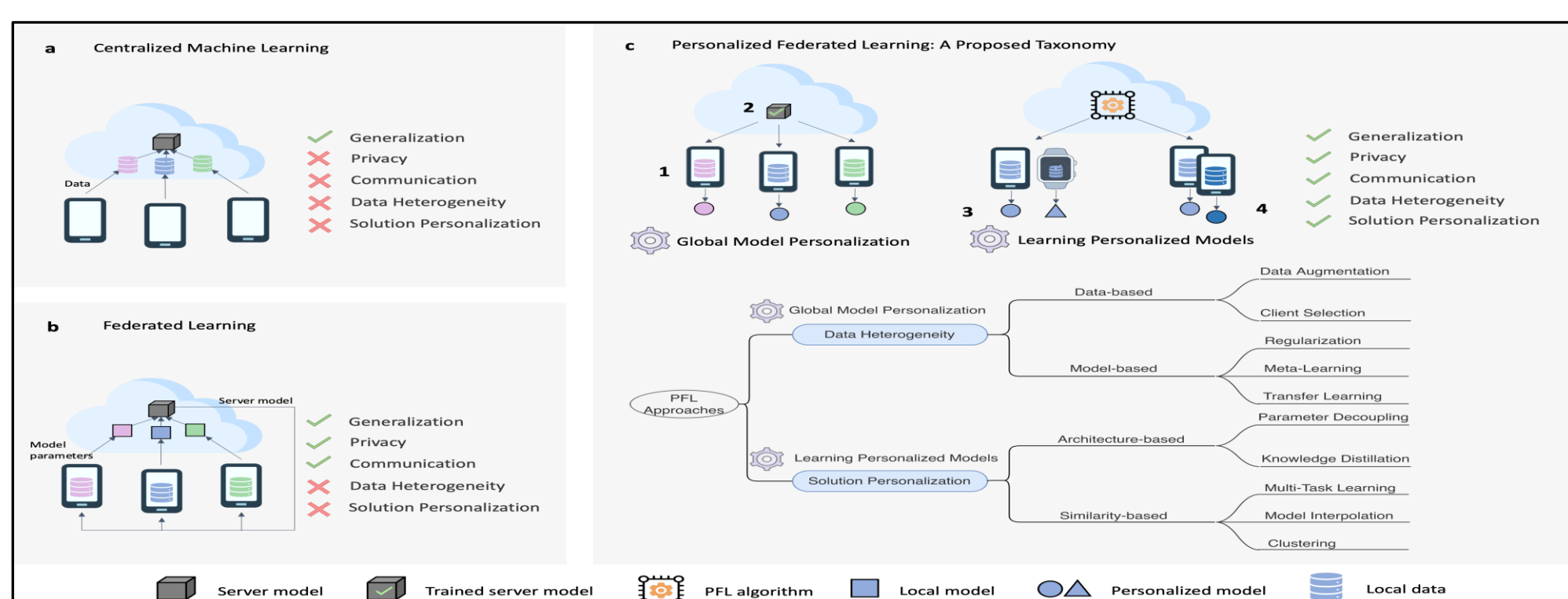
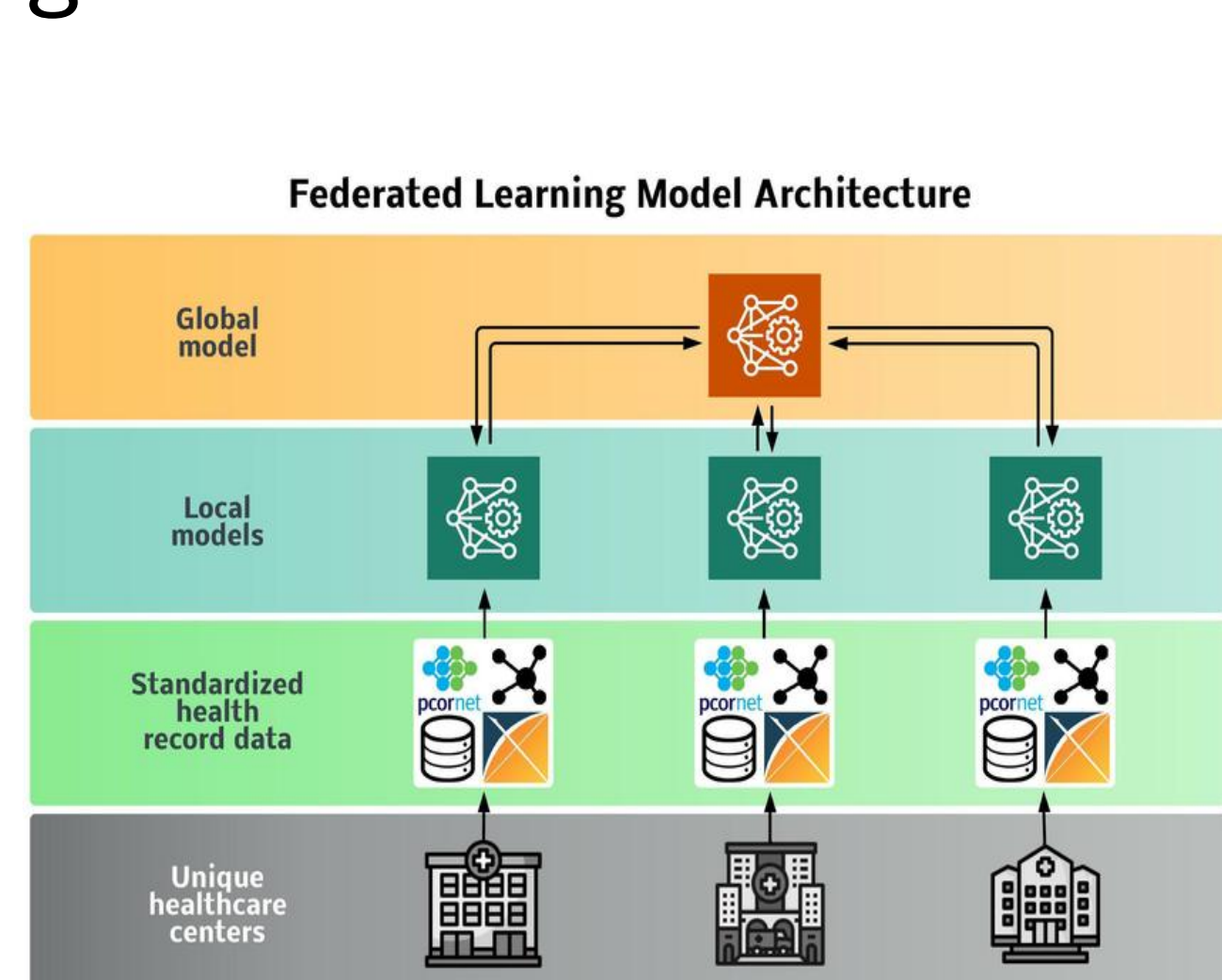


Figure 1

Introduction

- Apply Federated Learning (FL) and Personalized FL (PFL) to breast cancer subtype classification.
- Reproduce EMBER in a federated setting.
- Evaluate performance on heterogeneous non-IID multi-cohort data.
- Compare FL and PFL for accuracy and generalization.



Source: <https://ar5iv.labs.arxiv.org/html/2103.00710>

Figure 2

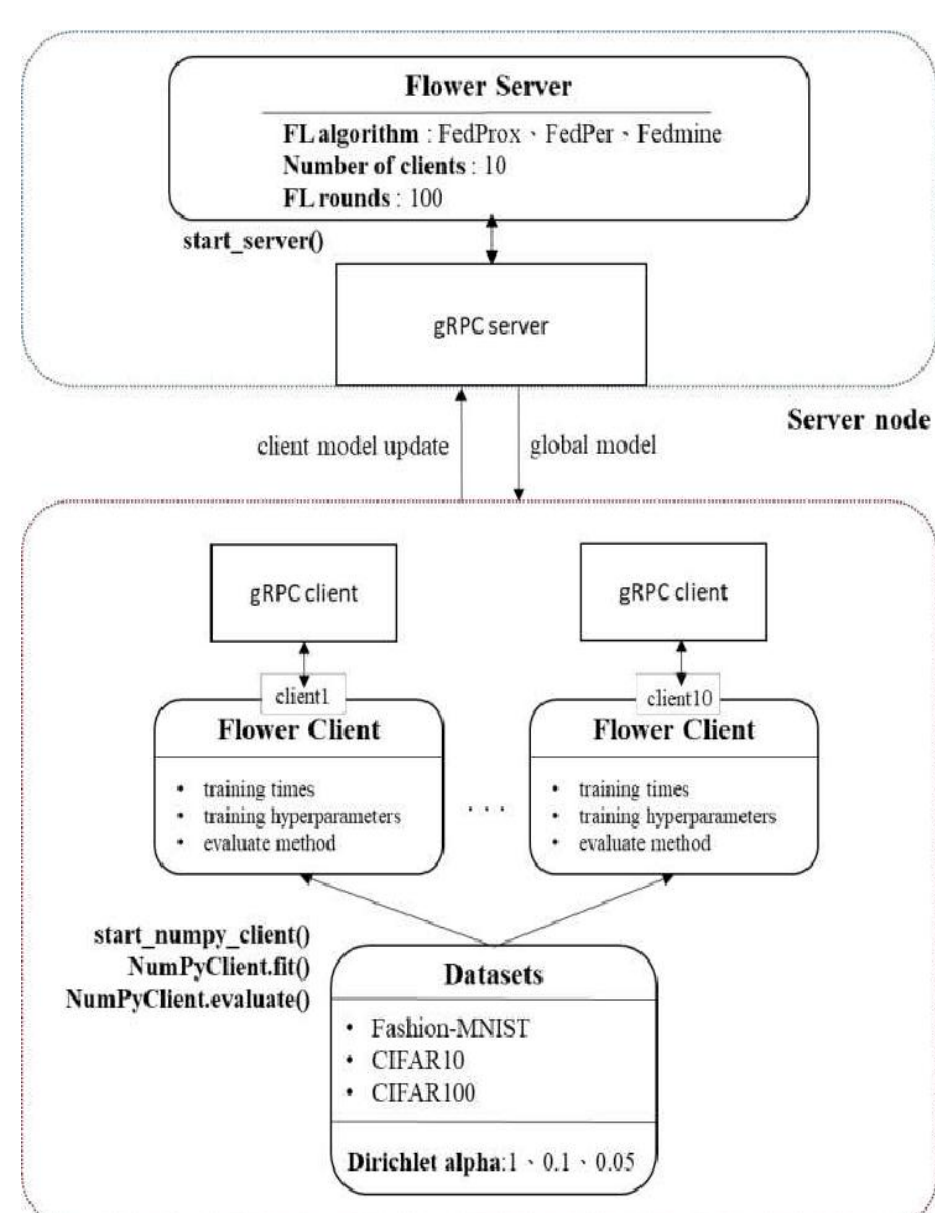


Figure 3

Future Work

Future work includes extending the framework to multi-modal data integration and improving model generalization across heterogeneous clinical datasets.

Methods

1. Reproducing EMBER Using an Autoencoder

- Development and training of local denoising autoencoders for TCGA and METABRIC cohorts.
- Extraction of 128-dimensional latent embeddings.
- Validation using PAM50 breast cancer subtype classification.
- Cross-cohort transfer experiments between TCGA and METABRIC.

2. Reproducing EMBER Using Federated Learning

Gene expression profiles from SCANB, TCGA, and METABRIC were harmonized using 14,036 common genes and standardized independently. A federated personalized autoencoder was trained across cohorts, producing latent embeddings that were subsequently evaluated using subtype classification, reconstruction error, and visualization analyses.

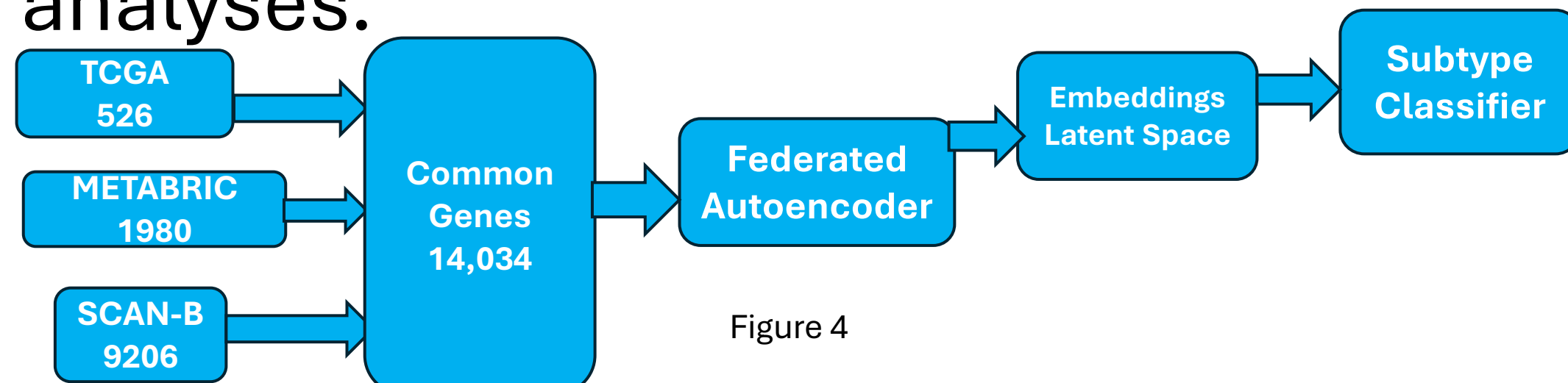


Figure 4

Results

- Denoising autoencoder trained on 11,712 breast cancer samples using 14,036 genes.
- Achieved 79.2% subtype classification accuracy on TCGA. (Local TCGA = 33.6%)
- Generalized across cohorts with 77.3% accuracy (TCGA → METABRIC).
- Survival prediction = 0.60 C-index
- Latent representation retained both subtype and prognostic information.

