APPLICATION OF EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) AND DOMAIN-ADAPTATION (DA) FOR DIAGNOSIS AND TREATMENT IN FEDERATED LEARNING-BASED HEALTHCARE SYSTEM

Chau The Vi

**University of Information Technology**

What ?

Explainable AI (XAI) and Domain Adaptation (DA) are advanced methods in the field of AI, designed to enhance the interpretability and adaptability of AI models in healthcare applications. A federated healthcare system uses Federated Learning (FL) to protect data privacy and optimize AI models on data from various sources without directly sharing sensitive information.

Why ?

- XAI makes AI decisions clear and trustworthy, ensuring that the decision-making process is understandable.
- Federated learning trains AI models without the need to share sensitive patient data, maintaining privacy.
- DA adjusts AI models to different medical data sources.
- AI aids in early disease detection, accurate diagnosis, and better patient care, improving overall healthcare quality.

Overview

1 Federated Learning **2** Domain Adaptation **3** Explainable AI
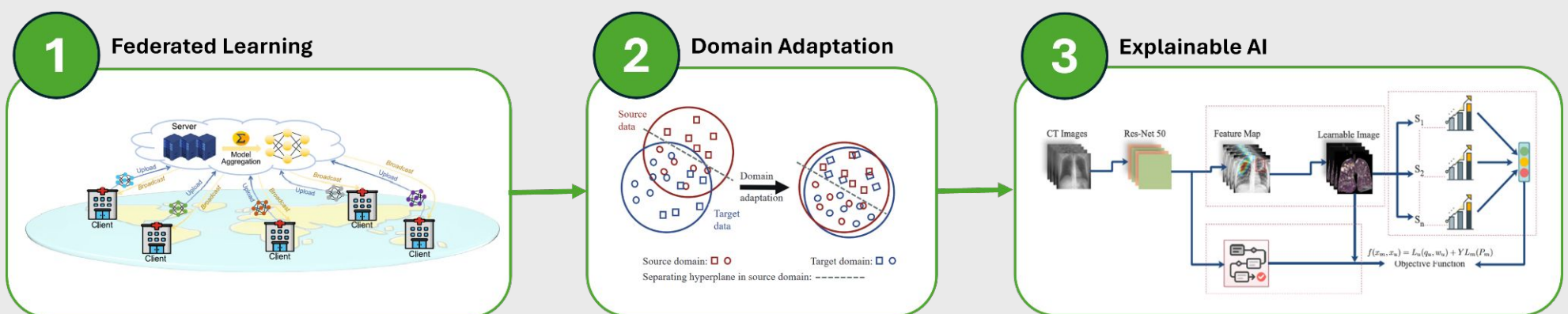
Description

1. Federated Learning

- Federated Learning trains models without sharing sensitive data between locations, ensuring individual privacy.
- Combines local models from various locations to create a global model, which is then redistributed for refinement and deployment.

How it work?

- Local models train on local data and send encrypted parameters to a central server.
- The central server aggregates these parameters to create a global model and sends it back for further refinement.
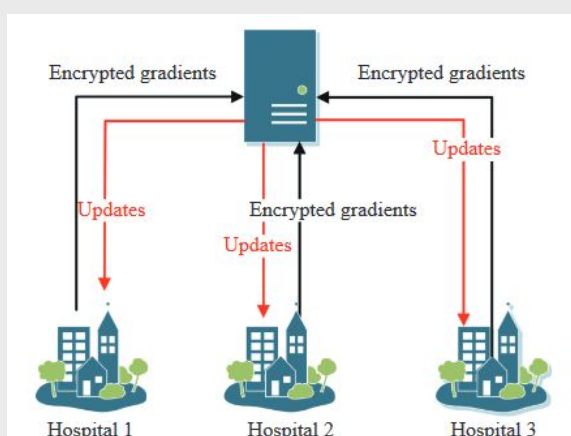

Figure 1. Example of federated learning based on horizontally.

2. Domain Adaptation

- Handles differences between datasets from various sources.
- Enhances the accuracy and reliability of the model on new data.
- Minimizes distribution differences between source and target data.
- Utilizes knowledge from source data to apply to target data.

Techniques:

- Data is split into source data (related to the current task) and target data (may be unlabeled or sparsely labeled).
- Aim to bridge the gap between these datasets, improving the accuracy and reliability of AI models in medical diagnosis and treatment.
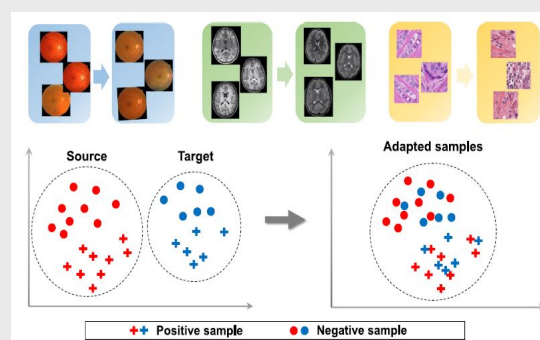

Figure 2. Domain shift among different medical sites (domains).

3. Explainable AI

- XAI makes AI models more transparent and understandable.
- It provides clear explanations of the processes and reasons behind AI model decisions.
- This enhances trust and allows users to comprehend how AI arrives at specific conclusions.

Techniques:

- Techniques like activation mapping and highlighting important features make AI models more interpretable.
- Methods like Class Activation Mapping (CAM) highlight infected areas in medical images, improving visual interpretability.
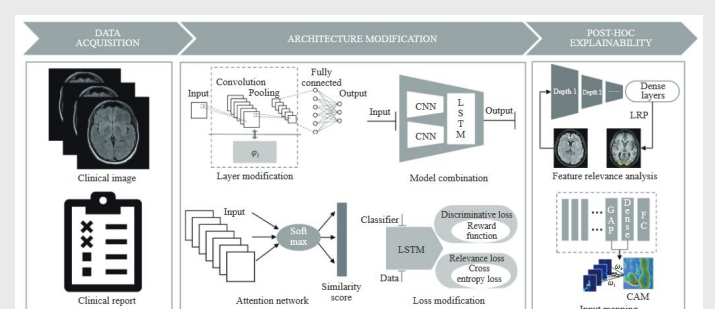

Figure 3. Flowchart for the explainable artificial intelligence model.