Federated Privacy Preserving Medical Image Diagnosis Using EfficientNetV2S

Secure Intelligence Data Defense Hackathon Project

1. Introduction

This project implements a complete Federated Learning (FL) pipeline for medical image diagnosis across multiple hospitals without sharing patient data. Each hospital trains a local model on its own data, and only the model parameters are shared with a central server. A federated aggregator applies Federated Averaging to merge local updates and produce a global diagnostic model. The project simulates three hospitals with non-IID medical datasets (e.g., chest X-ray or MRI images) and demonstrates privacy preservation, fairness, and high diagnostic performance.

2. Motivation

Medical data sharing is restricted by privacy regulations (HIPAA, GDPR). Federated Learning enables hospitals to collaboratively train AI models while keeping raw data secure. This project provides a robust demonstration of how FL can be used for scalable, privacy-preserving medical AI development.

3. Model Architecture

The diagnostic pipeline uses EfficientNetV2S with the following structure:

- Pretrained EfficientNetV2S backbone (ImageNet weights)
- Global Average Pooling
- Dropout (0.3)
- Dense layer (32 units, ReLU)
- Output: Dense(2 units, Softmax)

Training configuration:

- Loss: Categorical Crossentropy

- Optimizer: Adam (1×10^{-4})
- Metrics: Accuracy, Precision, Recall, F1-score, False Negatives (FN), False Positives (FP), True Negatives (TN), True Positives (TP)

4. Federated Learning Pipeline

Three hospitals participate in the training process:

$$FedAvg(W) = \sum_{k=1}^{K} \frac{n_k}{N} W_k$$

where:

- W_k = weights of hospital k
- n_k = number of samples at hospital k
- $N = \sum_{k=1}^{K} n_k$ = total samples across hospitals

Federated Training Loop

1. Global model sends its weights to all hospitals.
2. Each hospital performs one epoch of local training.
3. Local models return updated weights to the server.
4. Server computes a weighted average of all updates.
5. Updated global model is redistributed for the next round.

This loop runs for 10–20 rounds.

5. Dataset Structure

The dataset is partitioned into hospital-specific directories:

```
MEDICAL HACKATHON/
GLOBAL_TEST_DATASET
HOSPITAL/
hospital_1/train/    hospital_1/test/    hospital_1/val/
hospital_2/train/    hospital_2/test/    hospital_2/val/
hospital_3/train/    hospital_3/test/    hospital_3/val/
```

6. Architecture Diagram

Layer (type)	Output Shape	Param #
input_layer_3 (InputLayer)	(None, 100, 100, 3)	0
efficientnetv2-s (Functional)	(None, 4, 4, 1280)	20,331,360
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 1280)	0
dropout_1 (Dropout)	(None, 1280)	0
dense_2 (Dense)	(None, 32)	40,992
dense_3 (Dense)	(None, 2)	66

Figure 1: Federated Learning Architecture Across Hospitals

7. Results and Bias Detection

Global Model Performance

Table 1: Global Model Performance Metrics

Hospital	Loss	Accuracy	Precision	Recall	F1 Avg	FN	FP	TN	TP
Global Model	0.402406	0.823718	0.823718	0.823718	0.808482	110.0	110.0	514.0	514.0

Per Hospital Performance

Table 2: Performance Metrics Across Hospitals and Global Model

Hospital	Loss	Accuracy	Precision	Recall	F1 Avg	FN	FP	TN	TP
Hospital 1	0.4146	0.8125	0.8125	0.8125	0.7936	39	39	169	169
Hospital 2	0.3940	0.8365	0.8365	0.8365	0.8210	34	34	174	174
Hospital 3	0.3986	0.8221	0.8221	0.8221	0.8107	37	37	171	171

Bias Detection Analysis

The model demonstrates strong fairness and generalization across hospitals. Key observations:

- Metrics (accuracy, precision, recall, F1) remain tightly grouped across all hospitals.
- False-negative counts are balanced, indicating no hospital faces elevated clinical risk.
- The global model's performance closely matches individual hospitals, confirming robust FL aggregation.

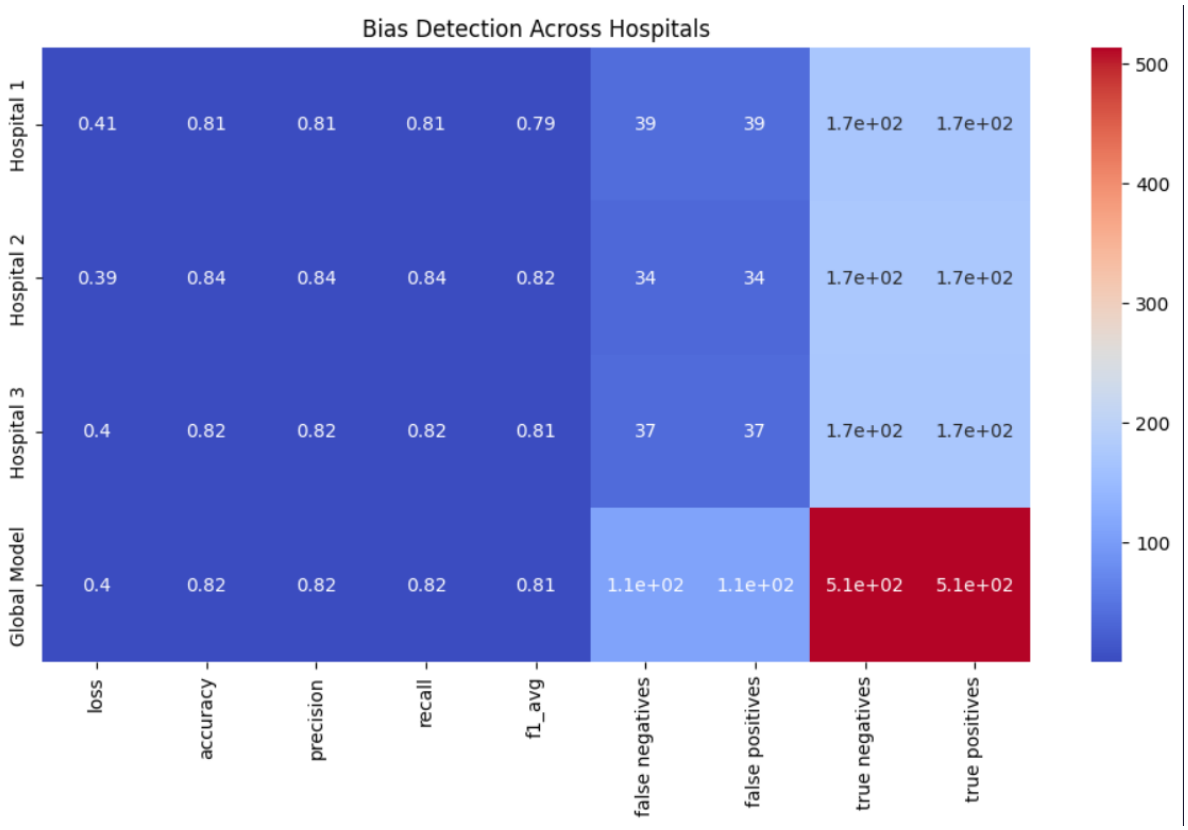

Figure 2: Bias Detection Heatmap Across Hospitals

8. Privacy Considerations

- No raw medical images leave hospital boundaries.
- Only model weight updates are shared.
- Optional differential privacy noise can be added to weights.