LotusFA: a Federated Analytics System for Federated Learning of Watermarked Data

Tao Ling[1], Siping Shi[1], Dan Wang[1], Yifei Zhu[2], Zhu Han[3]

[1]The Hong Kong Polytechnic University , [2]Shanghai Jiao Tong University, [3]University of Houston

• Introduction:

- **Federated Learning (FL)** is a privacy-preserving approach that allows multiple clients to collaboratively train models without sharing their private data with a central server.
- **Digital watermarking** is crucial for data ownership and copyright protection in multimedia, but it can lead to shortcut learning, where models rely on simple features like watermarks instead of complex core features. In federated learning, the problem is exacerbated because each client's watermark features are unknown to the server due to privacy constraints.
- **Challenge: (i)** identifying diverse watermark features is difficult due to client heterogeneity; **(ii)** mitigating negative impacts requires sharing insights to protect data privacy; **(iii)** ensuring compatibility with existing FL frameworks is essential for deployment.

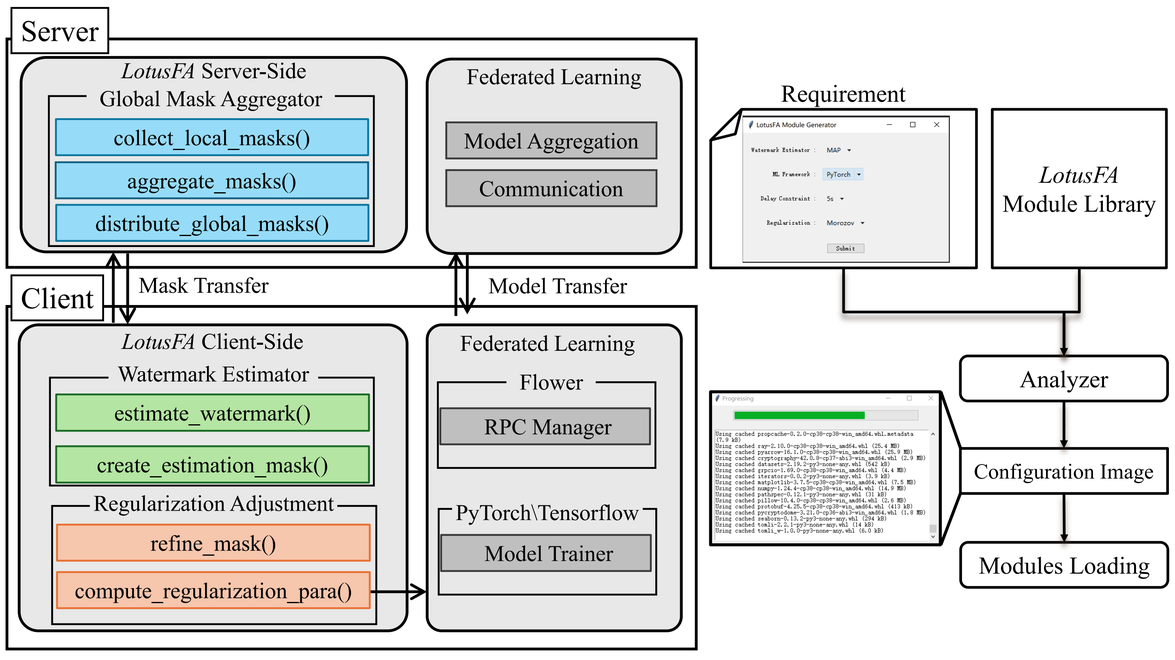

Fig. 1: Overview of *LotusFA* (Left) and MODL system (Right).

• The *LotusFA* System:

- **Client-side.** *Watermark Estimator* module on edge devices identifies watermark patterns without compromising privacy. It produces a *Local Watermark Estimation Mask* using statistical criteria. Data processing remains ephemeral for privacy, with parallel computation that doesn't interfere with training. The mask is encrypted for secure transmission. In Watermark-adapted Training Regularization Adjustment, module use the *Regularization Adjustment* module to blend local and global masks into *Refined Watermark Mask*, with blending ratio based on performance metrics. An adaptive regularization parameter is calculated to balance model fitting and generalization.
- **Server-side.** In privacy-preserving global mask aggregation module, the *Global Mask Aggregator* combines local masks into a *Global Watermark Mask* using gRPC protocols compatible with *Flower*. Masks are aggregated with weighted contributions based on factors like dataset size, with customizable aggregation algorithms. The global mask is distributed alongside model updates to reduce communication overhead.

Fig. 3: Testbed experiment setup on edge devices

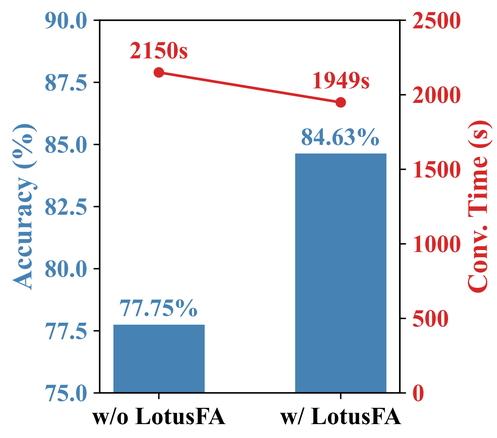

TABLE I
FL PERFORMANCE COMPARISON

Data Setting	Acc.	Conv. Time
Clean Data	85.15	1775s
Watermarked Data	77.75	2150s
W.M. + *LotusFA*	84.63	1949s

• Demonstration:

- We provide a Module-On-Demand Loading (MODL) system with a user interface for module generation. Developers can select parameters and algorithms and export customized modules with a single click.
- Using the chest X-ray dataset, we compared federated learning performance on clean and watermarked datasets (Tab. I). Watermarks caused degradation in FL accuracy, increased convergence time, and prompted models to focus on watermark regions rather than intended features, as shown by saliency maps (Fig. 2). The trained models returned to normal behavior with acceptable delay increase (see Tab. II).

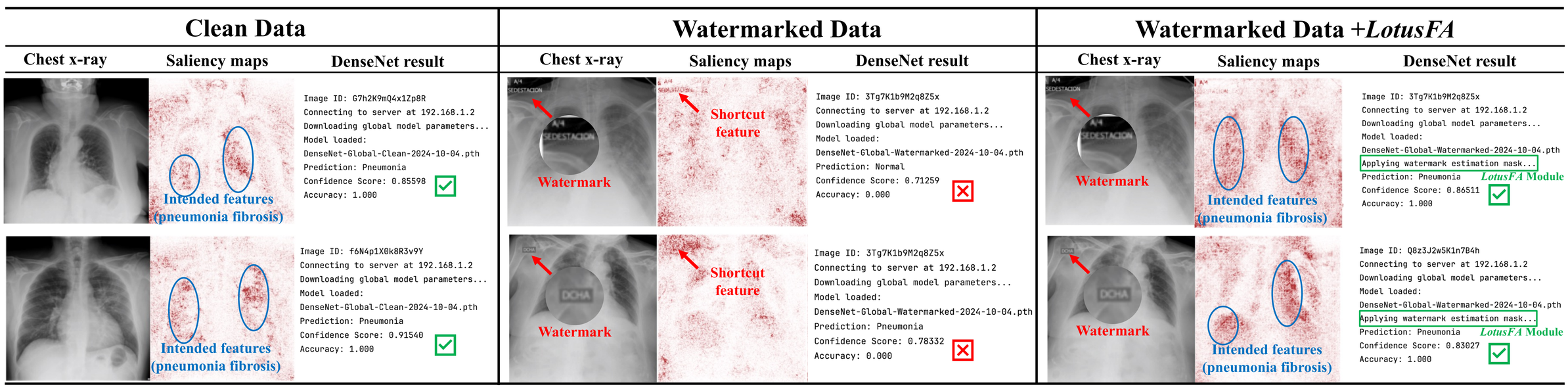

Fig. 2: Model identifies pneumonia (Left). Watermarks mislead model (Middle). Fixed the misleading with *LotusFA* (Right).