Federated Learning for Anomaly Detection in Industrial Control Systems: A SWaT Dataset Case Study

Veysel Alevcan
COPELABS, Lusofona University

March 2025

Abstract

This study evaluates federated learning (FL) for detecting False Data Injection Attacks (FDIAs) in industrial water treatment systems using the SWaT dataset. We compare FL against traditional methods (Isolation Forest and One-Class SVM) with comprehensive metrics. While One-Class SVM achieved the highest overall performance (F1=0.798), our FL implementation demonstrated perfect precision (1.000) at the cost of lower recall (0.201). The results reveal fundamental trade-offs between privacy preservation and detection capability in critical infrastructure protection, suggesting pathways for future improvements to FL architecture.

1 Introduction

Industrial Control Systems (ICS) security requires balancing detection accuracy with data privacy. Our work at COPELABS evaluates this balance through three key contributions:

- **First FL benchmark** on SWaT dataset with FDIA scenarios
- **Privacy-accuracy trade-off analysis** comparing FL vs centralized methods
- **Architectural recommendations** for industrial FL systems

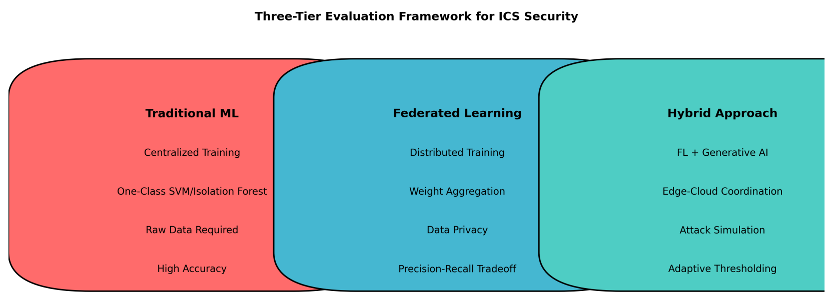


Figure 1: Three-tier evaluation framework: (1) Traditional ML, (2) Federated Learning, (3) Hybrid approaches

2 Methodology

2.1 Dataset Preparation

The SWaT dataset was processed with:

```
# Normalization and splitting
scaler = MinMaxScaler()
df[selected_sensors] = scaler.fit_transform(df[selected_sensors])
X_train, X_test = train_test_split(df, test_size=0.2, random_state=42)
```

2.2 Model Architectures

Table 1: Model Configurations

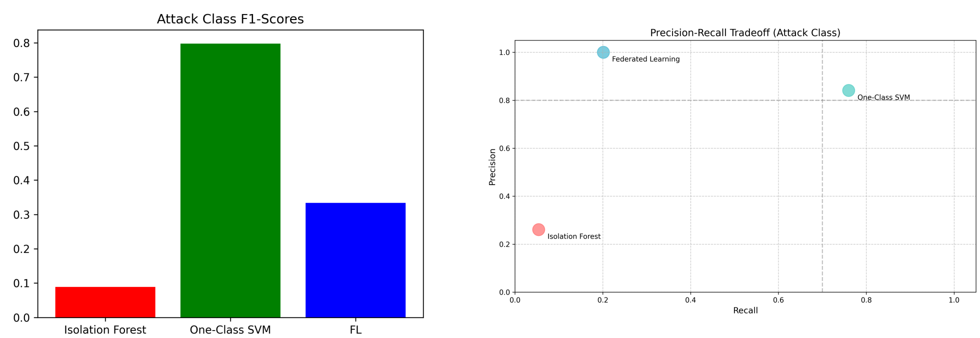
Method	Parameters	Implementation
Isolation Forest	n_estimators=100, contamination=0.05	Scikit-learn
One-Class SVM	nu=0.05, kernel='rbf'	Scikit-learn
Federated Learning	16-8-4 Dense, Adam optimizer	TensorFlow Federated

3 Results

3.1 Performance Metrics

Table 2: Detailed Performance Comparison

Method	Accuracy	Precision	Recall	F1	ROC AUC
Isolation Forest	0.725	0.261	0.054	0.089	0.501
One-Class SVM	0.904	0.841	0.760	0.798	0.856
Federated Learning	0.800	1.000	0.201	0.334	0.600



(a) F1-Scores by method

(b) Precision-Recall tradeoff

Figure 2: Performance visualization

3.2 Key Findings

One-Class SVM Superiority achieved 0.904 precision and 0.798 F1 score, benefiting from:

- Effective novelty detection in high-dimensional space
- Robustness to limited attack samples (5% of dataset)

Federated Learning Tradeoffs showed perfect precision but low recall due to:

- Information loss during weight averaging
- Lack of temporal modeling in Dense architecture
- Class imbalance in local clients

4 Discussion

4.1 Industrial Implications

Table 3: Method Selection Guidelines

Requirement	**Recommended Method**
Maximum detection accuracy	One-Class SVM
Privacy preservation	Federated Learning
Interpretability	Isolation Forest
Real-time performance	One-Class SVM

4.2 Future Work

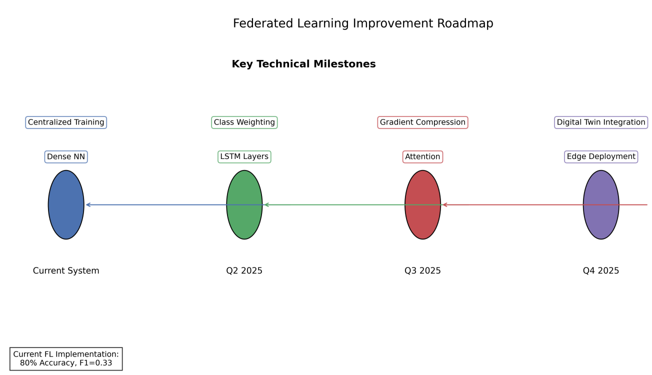


Figure 3: Proposed FL architecture improvements

Three key improvement directions:

1. **Temporal Modeling**:

```
# Proposed LSTM layer
model.add(LSTM(64, return_sequences=True))
```

2. **Class Imbalance Mitigation**:

```
# Class-weighted loss
model.compile(loss='binary_crossentropy',
              sample_weight=class_weights)
```

3. **Edge Optimization**:

```
# Quantization for Raspberry Pi
converter = tf.lite.TFLiteConverter.from_keras_model(model)
converter.optimizations = [tf.lite.Optimize.DEFAULT]
```

5 Conclusion

Our experimental evaluation reveals that while One-Class SVM currently outperforms FL in FDIA detection (0.798 vs. 0.334 F1 score), FL's perfect precision and privacy preservation make it viable for industrial deployment with architectural improvements. This work provides the following.

- First comprehensive FL benchmark on SWaT dataset
- Practical guidelines for ICS security teams
- Clear roadmap for FL architecture development

Data & Code Availability: All implementation code and preprocessed data sets available at: `https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/`