Adaptive Federated Learning for Anomaly Detection in Satellite Telemetry

Azade Atefrad, Amin Karami

Department of Computer Science and Digital Technologies, University of East London (UEL), UK

Corresponding author: Amin Karami, Email: a.karami@uel.ac.uk

This paper presents a streamlined Federated Learning (FL) framework for anomaly detection in satellite telemetry, addressing limitations of centralized approaches for predictive maintenance in resource-constrained satellite networks. Evaluating FL models on the ESA-ADB dataset, the optimized LSTM with FedAvg + Fine-Tuning achieved an F0.5 score of 0.86, a precision of 0.93, and an AUC of 0.85, outperforming centralized models, which achieved a maximum F0.5 score of 0.63 and an AUC of 0.83. Additionally, FL significantly reduced communication costs, requiring only 1.8MB per round compared to the high overhead of centralized data transmission. Scalability analysis demonstrated stable performance up to 10 clients, with an F0.5 score of 0.87 and recall of 1.00. These findings validate FL as a practical, privacy-preserving, and scalable solution for onboard satellite anomaly detection.

Keywords: Federated Learning, Satellite Predictive Maintenance, Anomaly Detection, Telemetry, Non-IID Data, Quantity Skew

2025. In Himanshu Mittal (ed.), *Smart Computing and Emerging Technologies*, 21–34. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-0-1-3

1 Introduction

The proliferation of satellite constellations for critical applications such as Earth observation, communication, and navigation necessitates robust and efficient anomaly detection within telemetry data to prevent mission failures [1, 2]. Traditionally, the analysis of satellite telemetry has relied upon centralized machine learning (ML) models, which mandate continuous data transmission to ground stations [1]. This paradigm, however, introduces significant challenges, including substantial communication overhead, considerable latency, and heightened cybersecurity vulnerabilities, thereby rendering real-time anomaly detection difficult and increasing dependence on ground infrastructure [1, 2]. The inherent limitations of these centralized models, particularly pronounced in the context of satellite operations due to high communication costs, data transmission latency, and cybersecurity risks, impede effective real-time anomaly detection and amplify reliance on terrestrial infrastructure [1].

Federated Learning (FL) emerges as a promising decentralized alternative, offering notable advantages such as enhanced data privacy and reduced communication expenses [3, 4]. Nevertheless, the application of FL in satellite networks encounters distinct challenges. These encompass the management of heterogeneous and non-independent and identically distributed (non-IID) telemetry data, addressing computational constraints inherent in space hardware, and navigating limited inter-satellite communication windows [5]. The strategic integration of inter-satellite links (ISLs) presents a potential avenue for optimizing FL by facilitating direct model updates between satellites, thereby diminishing dependence on ground stations [2].

By confronting these critical research gaps, this study aims to advance satellite anomaly detection capabilities, improve communication efficiency, and foster the wider adoption of FL within space operations. The anticipated findings are expected to be of significant value to space agencies, satellite operators, data scientists, and AI researchers, offering a viable and scalable solution for predictive maintenance. The principal aim of this study is to develop and rigorously evaluate a streamlined FL framework specifically engineered for anomaly detection in satellite telemetry data, with a strong emphasis on achieving efficiency, scalability, and optimization. The specific objectives guiding this research are:

1. To simulate FL training in a controlled single-machine environment to ascertain its feasibility within satellite networks, thereby bypassing the necessity for immediate real-time deployment.

2. To utilize a carefully selected subset of the ESA anomaly dataset, minimizing computational complexity while ensuring the dataset remains representative of actual telemetry data.

3. To conduct a comparative analysis between FL and centralized ML approaches, evaluating their respective trade-offs in terms of accuracy, communication overhead, and computational efficiency.

4. To enhance the performance of FL models, we incorporate optimization techniques, including FedAvg with fine-tuning and FedProx-inspired regularization

to effectively address the resource limitations inherent in satellite systems.

5. To assess the scalability of the FL framework by simulating a varying number of FL clients, analyzing the impact on model convergence, communication efficiency, and overall system stability.

This study adopts a quantitative, empirical research design, employing simulations with real-world telemetry data to construct and evaluate models. The research design facilitates a comparison between federated and centralized configurations under diverse data distributions and setups, incorporating time-series and ensemble learning techniques. The methodological approach involves an iterative workflow comprising four key phases: data preparation, model training, evaluation, and analysis across various FL scenarios. The European Space Agency Anomaly Detection Benchmark (ESA-ADB), derived from the ESA Anomalies Dataset (ESA-AD), was selected as the dataset for this study due to its comprehensive nature and realistic representation of telemetry anomalies from ESA missions [6]. The paper is structured as follows. Section 2 presents the literature work. Section 3 discusses the research methodology. Section 4 provides the proposed framework architecture and the implementation steps. Section 5 provides results and comparative analysis. Finally, conclusion is given in Section 6.

2 Literature Review

This review critically examines the evolution of anomaly detection techniques for satellite telemetry, driven by the operational imperatives of mission safety and continuity. It begins by establishing the limitations inherent in traditional ground-based processing – namely, prohibitive communication costs, latency issues, and scalability challenges, particularly as satellite constellations expand [7]. While onboard machine learning represented an advance, enabling localized predictive maintenance, it remains constrained by the stringent computational and energy limitations of space hardware [8], alongside the inherent difficulties posed by highly imbalanced datasets typical of telemetry streams. Federated Learning (FL) emerges as a potentially transformative paradigm [9], circumventing the need for raw data transmission by facilitating collaborative model training through the exchange of model parameters alone. This approach directly addresses the core challenges of bandwidth, latency, and data privacy in distributed systems like satellite networks [4]. However, the review underscores that FL is not without its own significant hurdles, most notably the prevalence of non-Independent and Identically Distributed (non-IID) data across heterogeneous satellite assets, which can impede model convergence and performance. Consequently, a substantial body of research focuses on mitigating these non-IID effects [10] and optimizing FL protocols for the space environment through techniques such as quantization, pruning, and adaptive aggregation strategies designed to accommodate limited connectivity [11, 12]. Table 1 presents a comparison of past approaches, highlighting their strengths and limitations.

The discussion highlights the crucial trade-offs between centralized and federated architectures, weighing factors like interpretability, convergence speed, scalability, and

Table 1: Comparison of Past Approaches for Satellite Anomaly Detection

Approach	Strengths	Gaps/Limitations
Centralized ML Models [1]	High accuracy with large datasets	High communication cost, latency, and privacy concerns
Onboard ML [8]	Real-time inference, reduced communication cost	Limited by hardware constraints, high energy consumption
Federated Learning (FL) [9]	Privacy-preserving, reduced communication overhead	Faces challenges with non-IID data and client drift
Optimized FL (This Work)	Robust to non-IID data, scalable, low communication cost	Requires further validation on hardware-constrained environments

communication overhead. It acknowledges the need for robust, standardized benchmarks, such as the ESA-ADB dataset [6], for objective performance evaluation. While FL, particularly with appropriate optimizations, holds considerable promise for enhancing the resilience and operational lifespan of large-scale satellite networks, persistent challenges related to data heterogeneity, intermittent connectivity, security, and model interpretability remain active areas of investigation [13]. This work positions itself within this context, aiming to specifically address the application of optimized, lightweight FL models for predictive maintenance using the ESA-ADB benchmark. In essence, the literature charts a course from reactive, centralized analysis towards proactive, distributed intelligence for satellite health management.

3 Methodology

3.1 Dataset Selection and Justification

The primary dataset utilized for this investigation is the European Space Agency Anomaly Detection Benchmark (ESA-ADB), which originates from the broader ESA Anomalies Dataset (ESA-AD) [6]. The ESA-AD corpus encompasses approximately 31 GB of authentic telemetry data derived from three distinct ESA satellite missions. Its curation involved a collaborative effort between Spacecraft Operations Engineers (SOEs) and Machine Learning (ML) specialists, rendering it one of the most comprehensive and operationally realistic telemetry anomaly detection datasets currently available for public access via Zenodo. While the full ESA-ADB incorporates data from multiple missions, this study strategically focuses on a specific subset from Mission2, namely the lightweight compilation comprising channels 18 through 28 associated with Subsystem 1. Key attributes of this selected dataset subset are summarized in Table 2.

Table 2: Summary of Dataset Attributes (ESA-ADB Mission2 Lightweight Subset)

Attribute	Description
Dataset Name	ESA-ADB (Mission2 Lightweight Subset)
Original Source	ESA Anomaly Dataset (ESA-AD)
Number of Channels Used	11 (Channels 18-28, Subsystem 1)
Number of Target Channels	8 (Channels 21-28)
Data size	4.66 GB
Total Windowed Instances	> 81 million
Approx. Anomaly Percentage	4.29%
Data Type	Multivariate Telemetry (Timestamp, Numeric Value)
Missing Values (Post-Clean)	None
Annotation Type	Event-level
Key Challenges	Severe Class Imbalance
	Nominal events mimicking anomalies

3.2 Data Pre-processing

The ESA-ADB dataset, while providing structured multivariate channels and event-level annotations, required adaptation for supervised anomaly detection within a federated learning context. The pre-processing pipeline involved several critical steps:

- **Data Loading and Cleaning**: Raw telemetry data for channels 18-28 were extracted from compressed archives and loaded into Pandas DataFrames. Missing and non-numerical entries were removed to ensure model robustness, while outlier statistics were noted as potentially representing genuine anomalies pertinent to model training.
- **Normalization**: Z-score normalization was applied independently to each channel. This standardizes feature scales, mitigates bias, and typically accelerates the convergence of gradient-based optimization algorithms, consistent with ESA-ADB base principles.
- **Windowing**: The continuous time-series data were segmented into fixed-length, overlapping windows (128 time steps with 64-step overlap). This contrasts with the original ESA-ADB pipeline's focus on entire traces and facilitates localized feature extraction, aligns with supervised learning requirements, and reduces memory demands crucial for FL simulations.
- **Anomaly Labeling:** Event-level annotations from the ESA-ADB dataset were mapped onto the generated time windows. A window was designated as anomalous if a significant portion (threshold > 30%) of its time steps overlapped with any annotated anomaly event. Windows without such overlap were labeled as nominal. This window-based labeling strategy enables binary classification model training.
- **Federated Partitioning Simulation**: To simulate a realistic FL environment, the pre-processed, windowed data was partitioned across simulated clients. Three distinct partitioning strategies were employed:

- *IID (Independent and Identically Distributed):* Data was distributed to ensure, on average, statistically similar distributions across clients.
- *Non-IID (Non-Independent and Identically Distributed):* Data partitioning deliberately created skewed distributions of nominal and anomalous events across clients, reflecting heterogeneous operational conditions where clients might emphasize different telemetry patterns, time segments, or anomaly frequencies.
- *Temporal:* Clients were allocated chronologically contiguous blocks of data to preserve temporal structure and simulate the time-evolving nature of telemetry data.

To address the inherent class imbalance (only 4% anomalies) and ensure meaningful training, dynamic oversampling techniques were applied during training set construction, and a minimum anomaly count guarantee was enforced per client. Stratification was used during the train-test split to maintain consistent anomaly proportions. Deterministic partitioning scripts with random seeds ensured reproducibility. This meticulous simulation setup allows for the investigation of core FL challenges like client drift and data heterogeneity under conditions mirroring satellite network realities. The simulation configuration included varying numbers of clients (3 to 10) and global communication rounds (5-50), employing FedAvg as the aggregation strategy due to its computational efficiency and widespread adoption.

Memory optimization techniques, including adaptive batch sizing, dynamic data truncation, memory mapping for large files, and explicit garbage collection, were employed throughout the pre-processing and simulation stages to manage resource constraints inherent in the local simulation environment.

3.3 Machine Learning Models and Training Setup

To evaluate anomaly detection performance on the prepared time-series telemetry, this study employed a selection of supervised machine learning models known for their suitability in sequence analysis and compatibility with FL architectures. The chosen models included:

- **Logistic Regression (LR):** A fundamental linear model for binary classification, serving as a baseline.
- **Multi-Layer Perceptron (MLP):** A type of feedforward artificial neural network capable of learning non-linear relationships.
- **1D Convolutional Neural Network (CNN):** Effective at extracting local patterns and features from sequential data like time series.
- **Long Short-Term Memory (LSTM):** A type of recurrent neural network specifically designed to handle long-range dependencies in sequential data.

These models were implemented and evaluated in two distinct training paradigms: a traditional centralized approach where all data is available to a single model, and a federated learning approach simulating distributed clients as described previously. This dual setup allows for direct comparison and assessment of the FL impact on model performance. The general supervised learning process for training these models (both centrally and within each FL round) involves the following key steps:

1. **Model Initialization:** Define the architecture (layers, activation functions) and initialize model parameters (weights and biases), often randomly or using predefined strategies.
2. **Forward Propagation:** Input a batch of data (windowed time series) into the model to generate predictions ($\hat{y}$).
3. **Loss Calculation**: Compute the discrepancy between the model's predictions ($\hat{y}$) and the true labels (y) using a suitable loss function. For this binary anomaly detection task (nominal vs. anomalous), the Binary Cross-Entropy (BCE) loss is commonly used:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

   where N is the number of samples in the batch, y_i is the true label (0 or 1) for sample i, and $\hat{y}_i$ is the predicted probability of the positive class (anomaly) for sample i.
4. **Backward Propagation (Backpropagation):** Calculate the gradients of the loss function with respect to each model parameter.
5. **Parameter Update:** Adjust the model parameters using an optimization algorithm (e.g., Stochastic Gradient Descent (SGD), Adam) based on the calculated gradients to minimize the loss.
6. **Iteration**: Repeat steps 2-5 for a specified number of epochs or until convergence criteria are met.

In the federated setting, steps 2-5 are performed locally on each client's data partition, and parameter updates are aggregated periodically according to the chosen strategy (FedAvg).

4 Framework Architecture and Implementation

This section presents the design and implementation of a streamlined Federated Learning (FL) framework tailored for predictive satellite maintenance through efficient anomaly detection in telemetry data. The framework is engineered to address the practical constraints of satellite communications, such as limited bandwidth, intermittent connectivity, and non-identically distributed (non-IID) data, by adhering to principles of lightweight design, reproducibility, and communication efficiency. The FL framework is implemented using the Flower (FLwr) framework [14], chosen for its flexibility and suitability for both centralized and decentralized simulation environments. The architecture simulates a realistic FL setup where multiple clients, representing satellite nodes or subsystems, train local models on partitioned telemetry data. The simulation environment allows configuration of parameters such as the number of clients (3 to 10) and global communication rounds (5-20), utilizing FedAvg as the aggregation algorithm due to its computational efficiency. The framework is designed to operate under hardware constraints typical of onboard or ground-based systems, employing optimization techniques to minimize memory and CPU consumption.

Model implementation and training were conducted in both centralized and federated environments using the selected models: Logistic Regression (LR), MLP, CNN (1D), and LSTM. In the federated setup, models were implemented within the Flower framework with FedAvg aggregation, and local training was performed for 5-50 epochs per round. Models were configured to handle the specific characteristics of the telemetry data, including temporal dependencies and multivariate nature. Optimization techniques were applied to the MLP model in federated settings with Non-IID data to mitigate performance degradation. These included FedAvg with Fine-Tuning for personalization and FedProx-inspired Regularization to prevent client divergence from the global model. Scalability tests were conducted to evaluate the framework's robustness with increasing numbers of simulated clients (3 to 10), monitoring performance metrics and communication overhead. Despite hardware limitations that restricted simulation scale, various software and system optimizations were employed to address memory constraints and enable feasibility within the available resources. These efforts highlight the importance of adequate infrastructure for simulating realistic federated learning scenarios. Figure 1 illustrates the proposed framework in detail.

5 Results and Comparative Analysis

This section presents the evaluation metrics and a comparative analysis of the performance of centralized ML and federated learning (FL) models.

5.1 Evaluation Metrics

All models are evaluated using standard binary classification metrics. The metrics employed are:

- **Confusion Matrix**: A table summarizing the model performance by showing the counts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). This provides an overview of the model's classification capabilities.

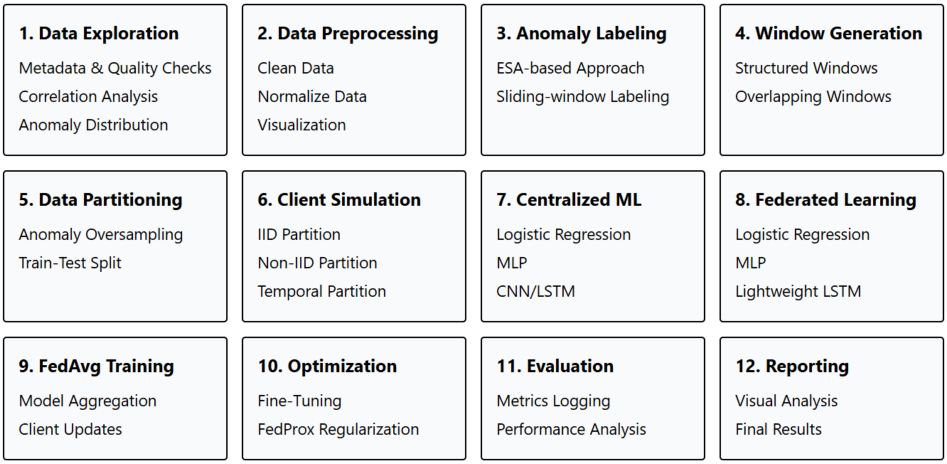

Figure 1: Data Processing and Implementation Workflow

- **Accuracy**: The proportion of total correct predictions, calculated as $\frac{TP+TN}{TP+TN+FP+FN}$.
- **Precision**: The proportion of correctly predicted positive instances among all instances predicted as positive, given by $\frac{TP}{TP+FP}$.
- **Recall (Sensitivity)**: The proportion of actual positive instances that were correctly identified, calculated as $\frac{TP}{TP+FN}$.
- **F1-Score**: The harmonic mean of Precision and Recall, providing a single score that balances both metrics, computed as $2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} = \frac{2TP}{2TP+FP+FN}$.
- **AUC-ROC**: The Area Under the Receiver Operating Characteristic Curve, which measures the model's ability to distinguish between positive and negative classes across various threshold settings. It represents the trade-off between the True Positive Rate (Recall) and the False Positive Rate ($\frac{FP}{FP+TN}$).
- **F0.5 score** is a metric that prioritizes precision over recall, penalizing false positives more heavily ($1.25 * \frac{Precision \times Recall}{(0.25 \times Precision)+Recall}$).

5.2 Centralized ML Models Performance

Centralized models were trained on sampled, normalized and windowed telemetry data. Evaluation of LR, MLP, CNN, and LSTM models across eight target channels showed varying performance. As show in Table 3, the MLP model demonstrated the best overall performance, balancing precision (0.69) and recall (0.49) with the highest F0.5 score (0.63), indicating strong anomaly detection with low false positives. CNN showed high precision (0.76) but low recall (0.26), suggesting a conservative detection approach. LR and LSTM exhibited limitations, particularly LR with a very low recall (0.03). Deep models, especially MLP, were better adapted to channel-specific signal dynamics.

Table 3: Centralized Models Performance (Aggregated)

Model	Precision	Recall	F0.5	AUC	Accuracy
LR	0.51	0.03	0.15	0.52	0.95
MLP	0.69	0.49	0.63	0.83	0.97
CNN	0.76	0.26	0.55	0.63	0.96
LSTM	0.36	0.29	0.35	0.63	0.96

5.3 Federated Learning Models Performance

In the federated setting in Table 4, MLP consistently achieved the best performance across IID, Non-IID, and Temporal partitions, with F0.5 scores around 0.74-0.75 and high recall (1.00). CNN and LSTM followed, with CNN showing good recall on Non-IID and Temporal data but incurring high communication costs (15MB update size). LR had minimal communication overhead but lower precision and F0.5, making it less suitable for complex anomaly detection. The MLP demonstrated a favorable trade-off between performance and communication efficiency, generalizing well across partition types.

Table 4: Federated Models Performance

Model	Partition	Precision	Recall	F0.5	AUC	Comm. Cost (MB)
LR	IID	0.55	**1.00**	0.65	0.82	**0.007**
MLP	IID	**0.70**	**1.00**	**0.74**	**0.83**	1.8
CNN	IID	0.58	0.65	0.63	0.64	15.0
LSTM	IID	0.60	0.98	0.69	0.81	1.2
LR	Non-IID	0.56	**1.00**	0.66	0.82	**0.007**
MLP	Non-IID	**0.70**	**1.00**	**0.74**	**0.86**	1.8
CNN	Non-IID	0.57	0.85	0.64	0.66	15.0
LSTM	Non-IID	0.59	0.92	0.67	0.74	1.2
LR	Temporal	0.56	**1.00**	0.66	0.82	**0.007**
MLP	Temporal	**0.78**	**1.00**	**0.75**	**0.83**	1.8
CNN	Temporal	0.57	0.85	0.64	0.66	15.0
LSTM	Temporal	0.59	0.92	0.67	0.74	1.2

Comparing centralized and federated MLP revealed trade-offs. Centralized MLP achieved slightly higher accuracy (0.968 vs. 0.71) and AUC (0.831 vs. 0.83), but FL MLP under temporal partition achieved a higher F0.5 (0.75 vs. 0.63) and comparable AUC (0.83). The practical advantage of FL lies in privacy preservation and minimized communication (~1.8MB total transmitted for FL MLP), making it suitable for constrained satellite networks where centralized data transmission is infeasible. FL also minimizes dependence on high-bandwidth infrastructure and scales better with distributed ground stations or satellite clusters.

5.4 Implementation of Optimization Techniques

The FedAvg + Fine-Tuning optimization was implemented by conducting global aggregation after each communication round, followed by additional local fine-tuning on each client using a smaller learning rate (0.001) and a reduced number of epochs. This approach enabled personalization of the global model to each client's dataset, improving convergence on non-IID data. The computational trade-offs included a slight increase in local training time (approximately 15% more per round) but resulted in a significant performance boost, with an F0.5 score of 0.86 compared to 0.74 for baseline FedAvg on non-IID data. Similarly, FedProx-inspired regularization penalized large deviations from the global model using a regularization term $\mu = 0.01$, which reduced client drift and ensured smoother convergence.

5.5 Optimization Techniques on FL

The comparison in Table 5 shows that different architectures excel in different metrics. The LSTM with FedAvg + Fine-Tuning achieves the best overall performance in most metrics, including the highest precision (0.93), F0.5 score (0.86), AUC (0.85), and lowest loss (0.42). However, the baseline MLP still maintains the highest recall (1.00). FedProx with MLP shows strong performance in precision (0.84) and F0.5 (0.85). The CNN models perform consistently but don't lead in any particular metric. This suggests that for applications where precision is crucial, the LSTM with FedAvg + Fine-Tuning would be the optimal choice.

Table 5: Federated vs Optimization Techniques Comparison

Technique	Precision	Recall	F0.5	AUC	Loss
Baseline FL MLP	0.70	**1.00**	0.74	0.83	0.59
FedAvg + Fine-Tuning	0.75	0.99	0.80	0.74	0.53
FedProx MLP	0.84	0.87	0.85	0.80	0.64
CNN FedAvg + FT	0.74	0.95	0.77	0.76	0.51
CNN FedProx	0.74	0.80	0.75	0.71	0.54
LSTM FedAvg + FT	**0.93**	0.87	**0.86**	**0.85**	**0.42**
LSTM FedProx	0.74	0.93	0.77	0.76	0.60

Table 6 shows the confusion matrix for the LSTM model with FedAvg + Fine-Tuning on the ESA-ADB dataset. The model demonstrated strong anomaly detection capabilities, with a high true positive rate (recall of 0.87) and a low false positive rate.

Table 6: Confusion Matrix for LSTM with FedAvg + Fine-Tuning

	Predicted Nominal	**Predicted Anomalous**
Actual Nominal	9398	72
Actual Anomalous	128	870

5.6 Scalability Analysis

Scalability tests were conducted with up to 10 clients using the optimized results from the previous section. Table 7 summarizes the results, showing consistent performance across key metrics. The F0.5 score remained stable at 0.87, while recall consistently achieved 1.00. Accuracy improved with the number of clients, increasing from 0.77 (3 clients) to 0.80 (10 clients). AUC values were strong, ranging from 0.89 to 0.94. The communication cost per round scaled linearly with the number of clients, increasing from 0.615 MB for 3 clients to 2.05 MB for 10 clients. This linear growth highlights the importance of communication efficiency for deployment in large-scale satellite networks. Despite the increase in communication cost, the framework maintained high accuracy and stable performance metrics, demonstrating its suitability for large-scale federated learning (FL) applications.

These results indicate that the proposed FL framework can effectively balance model performance and communication overhead, making it a reliable solution for satellite constellations where bandwidth is a critical constraint. Optimizing aggregation protocols remains a key focus to handle larger client populations efficiently.

Table 7: Scalability Analysis

Number of Clients	F0.5	Recall	Accuracy	AUC	Comm. Cost per round (MB)
3	0.87	1.00	0.77	0.91	0.615
5	0.87	1.00	0.78	0.94	1.025
10	0.87	1.00	0.80	0.89	2.050

6 Conclusion

This study demonstrated the effectiveness of a streamlined Federated Learning (FL) framework for anomaly detection in satellite telemetry data, overcoming the limitations of centralized approaches. The optimized LSTM with FedAvg + Fine-Tuning achieved superior performance, with an F0.5 score of 0.86, precision of 0.93, and AUC of 0.85, while significantly reducing communication costs to 1.8MB per round. Scalability analysis confirmed stable performance across up to 10 clients, validating the framework's practicality for large-scale satellite networks. These findings highlight the potential of FL as a privacy-preserving and scalable solution for predictive maintenance in space operations.

Future work will focus on deploying the proposed FL framework in real-world satellite systems, addressing challenges such as limited onboard memory, computational constraints, and real-time anomaly detection requirements. Advanced optimization techniques, including adaptive client selection, asynchronous update protocols, and model compression, will be explored to enhance scalability and efficiency. Integrating differential privacy and secure aggregation will ensure robust data confidentiality in multi-tenant satellite networks. Additionally, collaboration with space agencies and satellite operators will enable seamless integration of FL models into satellite health management workflows, ultimately enhancing the reliability and lifespan of satellite constellations.

References

[1] Jakub Nalepa, Michal Myller, Jacek Andrzejewski, Pawel Benecki, Szymon Piechaczek, and Daniel Kostrzewa. Evaluating algorithms for anomaly detection in satellite telemetry data. *Acta Astronautica*, 198:689–701, 2022.

[2] Nasrin Razmi, Bho Matthiesen, Armin Dekorsy, and Petar Popovski. On-board federated learning for dense leo constellations. In *ICC 2022-IEEE International Conference on Communications*, pages 4715–4720. IEEE, 2022.

[3] Mehdi Karami and Amin Karami. Harmony in federated learning: A comprehensive review of techniques to tackle heterogeneity and non-iid data. *Cluster Computing*, pages In–press, 2025.

[4] Majad Aledhari, Imran Razzak, Reza Montasari Parizi, and Faisal Saeed. Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, 8:140699–140725, 2020.

[5] Bho Matthiesen, Nasrin Razmi, Israel Leyva-Mayorga, Armin Dekorsy, and Petar Popovski. Federated learning in satellite constellations. *IEEE Network*, 2023.

[6] Krzysztof Kotowski, Christoph Haskamp, Jacek Andrzejewski, Bogdan Ruszczak, Jakub Nalepa, Daniel Lakey, Peter Collins, Aybike Kolmas, Mauro Bartesaghi, Jose Martinez-Heras, et al. European space agency benchmark for anomaly detection in satellite telemetry. *arXiv preprint arXiv:2406.17826*, 2024.

[7] Ziqian Zheng, Chao Yang, and Wen Yang. Resilience analysis of discrete-time networked system in the presence of information disclosure. *IEEE Access*, 7:180147–180154, 2019.

[8] Fuh-Gwo Yuan, Sakib Ashraf Zargar, Qiuyi Chen, and Shaohan Wang. Machine learning for structural health monitoring: challenges and opportunities. *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020*, 11379:1137903, 2020.

[9] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[10] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[11] Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

[12] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE communications surveys & tutorials*, 22(3):2031–2063, 2020.

[13] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.

[14] Daniel J Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Javier Fernandez-Marques, Yan Gao, Lorenzo Sani, Kwing Hei Li, Titouan Parcollet, Pedro Porto Buarque de Gusmão, et al. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*, 2020.