

# Vertical Federated Learning for Failure-Cause Identification in Disaggregated Microwave Networks

Fatih Temiz<sup>1\*</sup>, Memedhe Ibrahimli<sup>1</sup>, Francesco Musumeci<sup>1</sup>, Claudio Passera<sup>2</sup>, and Massimo Tornatore<sup>1</sup>

<sup>1</sup>*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*

<sup>2</sup>*SIAE-Microelettronica, Cologno Monzese, Italy*; \*Corresponding author: *fatih.temiz@mail.polimi.it*

**Abstract**—Machine Learning (ML) has proven to be a promising solution to provide novel scalable and efficient fault management solutions in modern 5G-and-beyond communication networks. In the context of microwave networks, ML-based solutions have received significant attention. However, current solutions can only be applied to monolithic scenarios in which a single entity (e.g., an operator) manages the entire network. As current network architectures move towards disaggregated communication platforms in which multiple operators and vendors collaborate to achieve cost-efficient and reliable network management, new ML-based approaches for fault management must tackle the challenges of sharing business-critical information due to potential conflicts of interest. In this study, we explore the application of Federated Learning in disaggregated microwave networks for failure-cause identification using a real microwave hardware failure dataset. In particular, we investigate the application of two Vertical Federated Learning (VFL), namely using Split Neural Networks (*SplitNNs*) and Federated Learning based on Gradient Boosting Decision Trees (*FedTree*), on different multi-vendor deployment scenarios, and we compare them to a centralized scenario where data is managed by a single entity. Our experimental results show that VFL-based scenarios can achieve F1-Scores consistently within at most a 1% gap with respect to a centralized scenario, regardless of the deployment strategies or model types, while also ensuring minimal leakage of sensitive-data.

**Index Terms**—Vertical Federated Learning, Disaggregated Microwave Networks, Failure Cause Identification

## I. INTRODUCTION

Microwave networks are an integral part of today's and future mobile transport networks, offering cost-efficient and high-capacity transmission to meet the rising demand for backhaul capacity [1]. In microwave networks, effective and efficient failure management is essential. Previously, failure management was manually carried out by network experts using monitoring data collected by the Network Management System (NMS). This human-based approach results in a time-consuming and error-prone Root Cause Analysis(RCA) that can cause, e.g., costly unnecessary site visits. Machine Learning (ML)-based network failure management in microwave networks has hence gained traction in the last years [2], [3], as it promises to ensure cost-efficient and accurate failure management [4].

However, current solutions can only be applied to monolithic scenarios, with tightly integrated components managed by a single vendor, while next-generation 5G-and-beyond networks are moving towards disaggregated architectures where multiple operators and vendors manage different interoperable

NEs [5]. This shift, supported by groups like Open Radio Access Network (O-RAN) and the Telecom Infra Project (TIP), aims to mitigate vendor lock-in and foster competition and innovation. Data sharing between different vendors or operators is often restricted in disaggregated networks due to confidentiality concerns which create a barrier to centralized ML training.

Federated Learning (FL) is a collaborative ML technique, in which each participating client, such as a vendor or an operator, trains a local model and shares only model updates with a central server that aggregates the local models as a third-party collaborator and then produces a global model that is again shared with local participating clients. By training models locally and simultaneously, and by only sharing model updates rather than raw data, the exposure of risk-sensitive data is minimized and a certain degree of data privacy can be guaranteed. The two main FL categories, based on how data is partitioned across clients, are Horizontal Federated Learning (HFL) and Vertical Federated Learning (VFL), where data is partitioned sample-wise and feature-wise, respectively. These categories are illustrated in Fig. 1. HFL is appropriate in case features across clients are homogeneous [6] such as in a single-vendor multi-operator scenario where each operator uses the same type of equipment. Conversely, VFL is appropriate for disaggregated networks where several vendors manage different interoperable Network Elements (NEs), e.g., Indoor Unit Hardware (HW-IDU), Outdoor Unit Hardware (HW-ODU), and Network Operating System Software (NOS-SW) in a microwave network. Also, VFL is mostly applied in cross-silo scenarios [7] where participants (e.g. entities such as banks, hospitals, vendors etc.) seek privacy protection without compromising performance. Among the other privacy-preserving algorithms (e.g., Secure-Multi Party Computation, Differential Privacy (DP), etc.), Homomorphic Encryption (HE) is a promising approach to be applied in conjunction with VFL due to its additive property that allows secure aggregation of client outputs [7] without significant performance degradation and communication overhead.

To address concerns about sharing raw data, we propose to use Split Neural Networks (SplitNN) and Gradient Boosting Decision Tree (GBDT)-based VFL (FedTree) for the hardware failure-cause identification problem in disaggregated microwave networks, considering both a default and HE-assisted implementation, and we test the proposed approaches in a multi-vendor deployment scenario. To the best of our

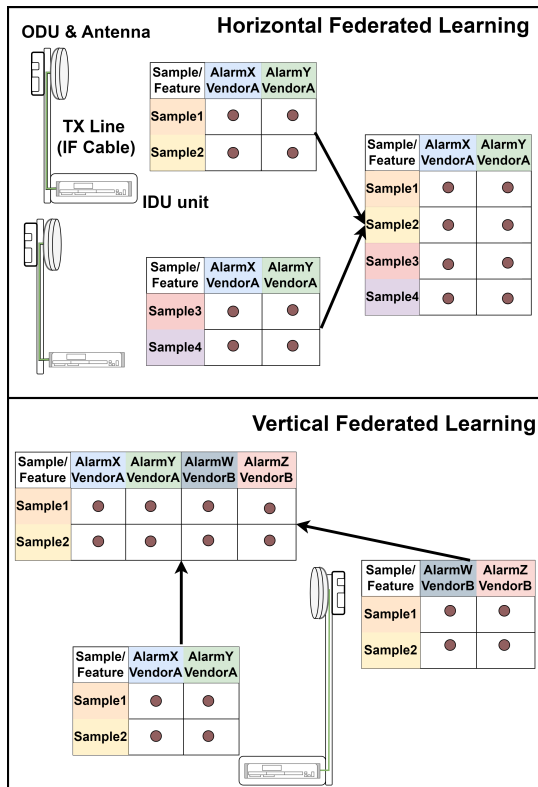


Fig. 1. HFL in Single-Vendor HFL Scenario and VFL Multi-Vendor Scenario knowledge, this is the first time VFL is adopted to address failure-cause identification in disaggregated microwave networks. The main paper contributions are as follows:

- 1) We define the problem of VFL-based failure-cause identification in a multi-vendor disaggregated microwave network and make use of real microwave hardware-failure data [8].
- 2) We propose to use GBDT and SplitNN VFL to perform failure-cause identification in multi-vendor scenarios and compare them to a benchmark comprising a Single Vendor Scenario (SVS), where failure-cause classification is performed in a centralized manner.
- 3) We carried on experimental results that show how the proposed VFL solutions achieve a maximum average F1-Score of 96.44%, closely matching the centralized result of 96.49%.

## II. BACKGROUND AND RELATED WORKS

### A. Background on Disaggregated Microwave Networks

The concept of disaggregation in communication networks has been investigated extensively in the last decade. Disaggregation mainly aims to reduce CapEx and foster innovation thanks to the definition of open interfaces that facilitate competition. The main research initiatives for disaggregation in wireless networks are the O-RAN Alliance and the TIP's Wireless Backhaul Group (WBG), a project group of TIP that investigates disaggregation in wireless backhaul (i.e., microwave networks). WBG particularly works on disaggregated white box solutions for microwave networks and has recently

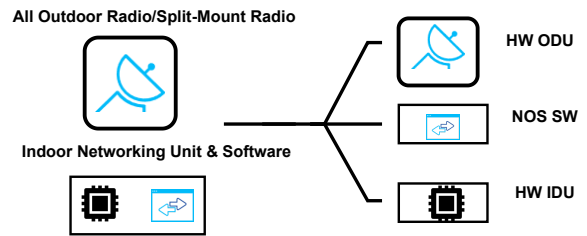


Fig. 2. OSH's proposed disaggregation in microwave networks

presented a solution called OpenSoftHaul (OSH) [9]. As illustrated in Fig. 2, OSH decouples microwave equipment into three distinct interoperable components, full outdoor radios in the microwave and millimetre-wave frequencies (i.e., HW ODU), IDU equipment that has open interfaces (i.e., HW IDU) and operator configurable Network Operating System (i.e., NOS SW). Another significant initiative for disaggregation in the industry is Disaggregated Cell Site Gateways (DCSGs). A cell site gateway, or router, handles the aggregation of mobile data from cell sites and its transport to the service provider's core network. Cell site routers (CSRs), also known as cell site gateways (CSGs), have traditionally been offered as integrated solutions by a limited number of large vendors, but this trend changes through disaggregation to avoid vendor lock-in. The main use cases for DCSG would be to perform 2G/3G/4G/5G mobile transport backhaul, but another DCSG application is acting as IDU for Microwave Backhaul. In the case where IDU is disaggregated, a microwave modem is implemented as part of the ODU, reducing the IDU's role to basic networking functions while ensuring seamless interoperability with other microwave equipment [10].

### B. Related Works

Recently, some ML-based solutions for predictive maintenance in microwave networks have been explored. For example, Ref. [11] illustrates a successful implementation (up to 93% classification Accuracy) of supervised and unsupervised ML approaches for propagation failure identification (e.g., *deep fading*, *interference*). Similarly, Ref. [4], [8] shows successful implementation of ML-based classifiers (up to 96% classification Accuracy) for hardware failure classification. Ref. [2] presents various preventive maintenance tasks, including hardware degradation detection, high-temperature early warning, and many specialized applications. Similarly, Ref. [12] introduces the Proactive Microwave Anomaly Detection System (PMADS), a dynamic system designed for anomaly detection in cellular network microwave links. PMADS leverages network performance metrics and topological information to identify anomalies leading to faults. Other works (e.g., Refs. [13], [14]) have investigated the application of explainable Artificial Intelligence (XAI) frameworks to provide insights into the decision-making process of ML classifiers.

However, these works focus on monolithic scenarios, where data and models are managed by a single entity, ignoring disaggregation and collaborative approaches like FL. Few studies have explored FL-based solutions for microwave networks.

| IDU Group Alarms |     |     |     | ODU Group Alarms |     |     |     | NOS Group Alarms |      |      |      |      |     |      |       |
|------------------|-----|-----|-----|------------------|-----|-----|-----|------------------|------|------|------|------|-----|------|-------|
| A1               | A2  | A3  | ... | A52              | A53 | A54 | A55 | ...              | A116 | A117 | A118 | A119 | ... | A164 | Label |
| 0                | 0   | 0   | ... | 463              | 0   | 0   | 467 | ...              | 0    | 0    | 0    | 900  | ... | 0    | Power |
| 900              | 0   | 0   | ... | 0                | 0   | 159 | 0   | ...              | 0    | 0    | 0    | 0    | ... | 0    | IDU   |
| 0                | 0   | 0   | ... | 0                | 0   | 0   | 0   | ...              | 0    | 0    | 900  | 0    | ... | 0    | Cable |
| 0                | 0   | 0   | ... | 0                | 2   | 0   | 0   | ...              | 0    | 0    | 900  | 0    | ... | 0    | ODU   |
| 0                | 0   | 0   | ... | 0                | 0   | 0   | 0   | ...              | 889  | 0    | 0    | 900  | ... | 900  | Power |
| 0                | 0   | 0   | ... | 0                | 0   | 2   | 0   | ...              | 882  | 0    | 0    | 900  | ... | 900  | Power |
| 900              | 900 | 0   | ... | 117              | 117 | 0   | 0   | ...              | 0    | 0    | 0    | 900  | ... | 0    | IDU   |
| ...              | ... | ... | ... | ...              | ... | ... | ... | ...              | ...  | ...  | ...  | ...  | ... | ...  | ...   |

Fig. 3. Illustration of the Dataset

The most closely related work is Ref. [6], and it shows the application of an HFL-assisted solution for identifying the causes of propagation failures (e.g., extra attenuation, deep fading) in microwave networks within a multi-operator non-disaggregated scenario. While HFL performs effectively, it requires that all clients in the federated setting share the same feature space. This is a strict limitation in disaggregated networks, where managed network elements may possess varying feature sets, as in our scenarios.

While VFL-based solutions have been explored in other contexts, such as O-RAN [15] and Network Function Virtualization (NFV) systems [16], there is no prior work, to the best of our knowledge, that has investigated the application of VFL in disaggregated microwave networks.

### III. VFL-BASED FAILURE CAUSE IDENTIFICATION IN DISAGGREGATED MICROWAVE NETWORKS

#### A. Problem Definition

The basic components of a microwave link are illustrated in Fig. 1. In this study, we consider split-mount configuration, where electronics are split into outdoor and indoor units of microwave radio (ODU and IDU, respectively). We model the hardware failure cause identification as a multi-class ML problem, where each class corresponds to a specific hardware failure (see next subsection). The classifier is trained collaboratively by vendors on partitioned features, as described in Section V-A. In brief, each client has its specific dataset of features, i.e., IDU, ODU, NOS group features or combined versions, and the model predicts one of the four output classes that correspond to failure causes.

#### B. Dataset Description

The NMS records equipment alarms for each microwave link. Our dataset was collected on a real microwave network by SIAE Microelettronica, and it is an open-source dataset [8] comprising 164 equipment alarm features ( $A_1$  to  $A_{164}$ ) and 1669 data points (observations). Each observation in the dataset compiles data from 15-minute (900 seconds) non-overlapping intervals. For each microwave link and for each 15-minute interval, the classification problem is fed with all features, each representing the number of seconds a specific alarm was active within that window.

These features range from 0 to 900, where 0 means the alarm was OFF and 900 means the alarm was ON for the entire 15-minute period. Each observation is labelled and validated

by domain experts. The dataset includes four types of hardware failures (i.e., labels) and the distribution of each class is as follows:

- **Class-1 (IDU failure):** 515 observations (e.g., failure of an electronic IDU component, or a temperature issue due to improper equipment installation and/or a worn fan)
- **Class-2 (ODU failure):** 611 observations (failure causes are similar to those seen in the IDU in the point above)
- **Class-3 (Cable failure):** 207 observations (e.g., damaged connectors)
- **Class-4 (Power failure):** 336 observations (e.g., due to a power outage and/or a battery problem)

#### C. Preprocessing

After removing data for alarms with all zeros in the dataset (as they do not provide information), a total of 119 features are considered. Alarms can be considered as a) binary, i.e., 1 if the alarm is ON and 0 if the alarm is OFF, b) categorical, i.e., 0 if the alarm is OFF, 1 if the alarm is ON for less than 45 seconds in a 15-minute window, 2 if the alarm is ON for less than 450 seconds and more than 45 seconds in the 15-minute window, and 3 if the alarm is ON for more than 450 seconds in a 15-minute window c) raw, i.e., number ranging between 0 and 900. Following the approach in previous work [8], we consider categorical binning.

### IV. VFL APPROACHES FOR MULTI-VENDOR ML-MODEL DEVELOPMENT

VFL is a promising approach for developing ML models based on heterogeneous feature sets across different organizations. Most of the VFL algorithms are based on the Stochastic Gradient Descent (SGD) algorithm, which can leverage logistic regression, SVM, and NN models as potential candidates. NN-based architectures enhance data privacy thanks to the complex nonlinear functions these networks employ by default [17]. Additionally, tree-based models have also been investigated in the context of VFL in the last years [18]. Also, recent studies [19], [20] show that tree-based, especially GBDT-based, models are as powerful as NN-based models on small-sized (i.e. lower than 10k samples) datasets. In this study, we employ both NN-based and GBDT-based VFL approaches.

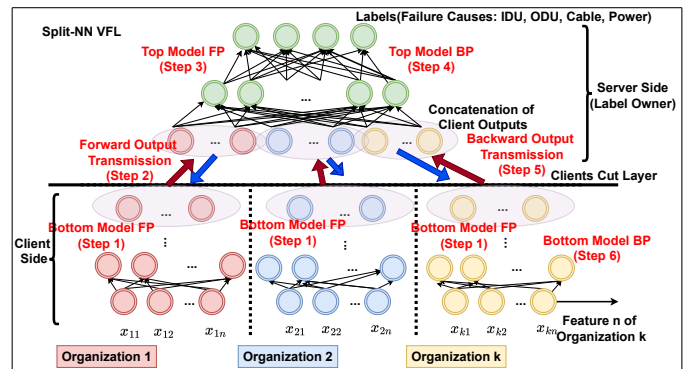


Fig. 4. VFL Architecture Based on Split Neural Networks

### A. Split Neural Networks

SplitNN is a VFL framework [21], consisting of local NN models for  $k$  clients (i.e., vendors) and a central server (i.e., a third-party entity). Each client manages its own local NN model, while the central server manages another NN model where the input layer is constructed by concatenating the local client's outputs. Fig. 4(a) shows the general architecture of SplitNN in the case of  $k$  organizations, where each organization manages its local NN model. The first layer in each local model corresponds to the input layer and includes as many nodes as the number of features at each client; hidden layers follow, where the number of hidden layers and neurons per layer may differ among local models; the last layer is the output of the local NN models. The outputs of the local NN models are then concatenated and serve as inputs to the NN model at the central server, followed by  $h$  hidden layers, and the last layer constituted by  $o$  neurons serving as output. In our scenario, we consider  $o=4$ , as the number of failure causes to discriminate. The main steps of the SplitNN VFL algorithm can be summarized as follows [17]:

- **Step 1: Bottom Model Forward Propagation (FP).** Each client computes the FP of their local model using local data to obtain intermediate representations (*Bottom Model FP* in Fig. 4(a)). This step is similar to conventional training except for loss calculation since labels are held in the global model.
- **Step 2: Forward Output Transmission.** The outputs (intermediate cuts) from the local models (client side) are transmitted to the global model (server side). Advanced security protocols such as HE could be implemented at this step to provide privacy preservation.
- **Step 3: Top Model FP.** The server concatenates the received forwarded outputs and processes them through the global model for the final output of the classification task. There are various merging techniques available, including element-wise max, min, average pooling, product, and sum. While concatenation is the simplest approach, our experiments showed these techniques generally result in only slight differences in performance.
- **Step 4: Top Model Backward Propagation (BP).** Based on the calculated loss from the prediction and the labels, the top model begins the backward propagation to compute gradients.
- **Step 5: Backward Output Transmission.** Gradients computed in the top model are sent back to each participating client.
- **Step 6: Bottom Model BP.** Upon receiving the gradients, each client updates model weights on their local model through backward propagation. This completes the cycle of one training iteration, referred to as *epoch*.

### B. FedTree

FedTree is an alternative implementation of GBDT-based VFL algorithm [22], where histograms (i.e., summation of gradients) are used as exchanged information differently from other implementations.

Within each iteration, the GBDT model trains iteratively and constructs a new tree aiming at minimizing the residuals, which are the differences between the predicted outcomes and the actual target values. In Vertical FedTree (See Alg. 1), clients (i.e., vendors) have their feature-wise partitioned local datasets (i.e., IDU-group alarms, ODU-group alarms, NOS-group alarms) and the active client  $P_1$  also has labels that act as a server. In Alg. 1, at first, the active client proposes cut points and generates public and private keys for HE (lines 1-3). The active client calculates gradients, since it has the labels (i.e., failure causes), and then sends them to the other parties in an encrypted way. For each depth  $d$ , each party (i.e., including  $P_1$ ) computes local histograms and sends local histograms to the active party again in an encrypted way (lines 4-14). The active party receives local histograms and concatenates them after decrypting. Using aggregated histograms, the active client calculates the best-split points and lets all parties update the node. Iteration ends when  $d$  reaches  $max\_depth$  for each tree in each client. (lines 15-18) In this

---

#### Algorithm 1 Vertical FedTree (adapted from [22])

---

**Input:** Clients  $P_1, \dots, P_n$ , number of trees  $T$ , tree depth  $d$ . Assume  $P_1$  has the labels.

**Output:** The final GDBT model  $f$ .

```

1:  $S \leftarrow \text{ProposeSplitCandidates}()$ 
2:  $P_1$  generates HE key pair  $(K_{\text{pub}}, K_{\text{pri}})$ .
3:  $P_1$  sends public keys to the other clients.
4: for  $i = 1$  to  $T$  do
5:   Conduct on client  $P_1$ 
6:    $g, h \leftarrow \text{UpdateGradients}()$ 
7:    $[g], [h] \leftarrow \text{Encrypt}(g, K_{\text{pub}}), (h, K_{\text{pub}})$ 
8:   Send  $[g], [h]$  to the other client.
9:   for  $j = 1$  to  $d$  do
10:    for  $k = 1$  to  $n$  do
11:     Conduct on each client  $P_k$ 
12:      $[G_k], [H_k] \leftarrow \text{ComputeHist}([g], [h], S_k)$ 
13:     Send  $[G_k], [H_k]$  to  $P_1$ .
14:    end for
15:   Conduct on  $P_1$ 
16:    $G, H \leftarrow \text{Decrypt}([G_k], K_{\text{pri}}), ([H_k], K_{\text{pri}})$ 
17:    $\mathbf{H} \leftarrow \bigcup_{k=1}^n \mathbf{H}_k$ 
18:   Update depth  $j$  of tree  $f_i$  using  $G, H$ 
19:   end for
20: end for

```

---

study, we adopt HE to secure exchanged information such as gradients and histograms. Specifically, our HE implementation is additive, based on the Paillier cryptosystem, and considers the *Honest but Curious* threat model [22], where participants strictly follow protocol but may infer sensitive information about others (e.g. gradients and histograms). In the default setting (i.e., FedTree), we protect local data while in the HE-assisted one (i.e., FedTreeHE) both local data and exchanged information are protected assuming a trusted server.

## V. CASE STUDIES AND NUMERICAL RESULTS

### A. Alarm Grouping

Considering disaggregation aspects discussed so far, alarms are grouped into IDU-Alarms (31 features), ODU-Alarms (54

features), and NOS-Alarms (34 features), mimicking a situation where each interoperable NE is managed by a different player in the disaggregated network, e.g., a different vendor for each NE. The alarms from the electronic parts of the IDU are included in IDU-Alarms (e.g., modulator/demodulator), ODU-Alarms include radio link-related alarms as well as electronic components (e.g., voltage-controlled oscillator), whereas alarms that indicate SW-configuration mismatch and system-health like alarms are grouped into NOS-Alarms.

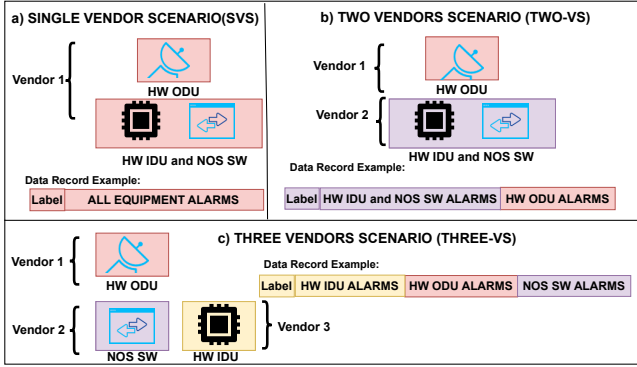


Fig. 5. Single and Multi-Vendor Scenarios in Microwave Networks

## B. Deployment Scenarios

We consider three different deployment scenarios, as shown in Fig. 5: 1) Single Vendor Scenario (SVS), 2) Two Vendor Scenario (2-VS), and 3) Three Vendor Scenario (3-VS).

- **SVS:** A single vendor manages the whole network elements. All equipment alarm features (e.g., 119 alarm features) belong to one vendor.
- **2-VS:** HW-IDU and NOS-SW belong to one vendor, and HW-ODU belongs to another vendor. Vendor 1 has only alarm information related to ODU equipment (e.g., radio power setting alarms), while Vendor 2 has both HW-IDU and NOS-SW alarms (e.g., IDU temperature, software unit mismatch alarms).
- **3-VS:** Equipment is fully disaggregated, and each interoperable NE is managed by a vendor. 3-VS complies with the OpenSoftHaul architecture, where HW-IDU, HW-ODU, and NOS-SW are managed by different vendors.

**Baseline and FL methodologies:** We consider *six baselines*: two centralized and four non-cooperative and six FL scenarios, obtained by varying the model type (i.e., NN-based or GBDT-based), the single vs. multi-vendor deployment (i.e., SVS, 2-VS, 3-VS), the privacy-preserving method (i.e., using HE or not), and whether vendors are cooperative or not.

Briefly, FL settings focus on collaboratively developing ML models across multiple vendors, allowing them to work together without sharing sensitive data. The centralized scenario (i.e., SVS) involves a single vendor developing the ML model using centralized data. In contrast, non-cooperative scenarios (i.e., 2VS and 3VS that use XGB or NN for local models) indicate that each data owner (e.g., IDU vendor) independently develops local ML models using only their data (e.g., 31 features for IDU vendor) without cooperation.

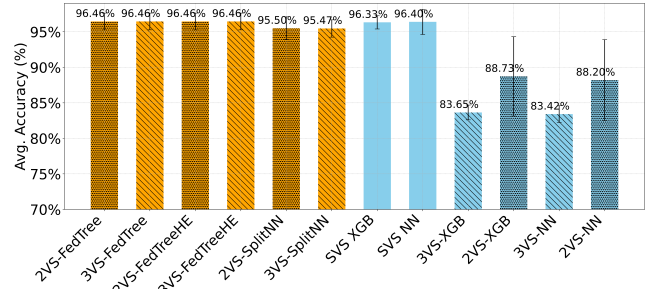


Fig. 6. Avg. Accuracy in % for VFL scenarios and baselines

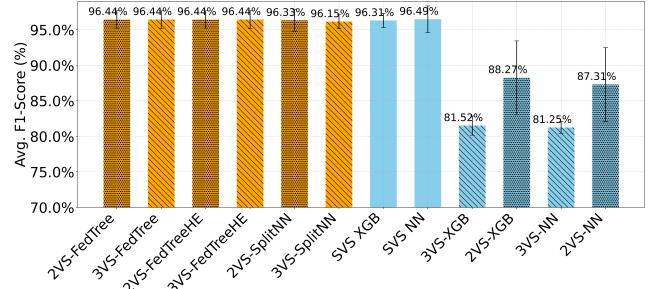


Fig. 7. Avg. F1-Score in % for VFL scenarios and baselines

## C. Evaluation Settings

We perform stratified 5-fold cross-validation on both baselines and FL settings to find the best-performing model and report the mean Accuracy and F1-score, and standard deviation, over the 5 folds. Moreover, we consider hyperparameter optimization Grid Search (GS), for a fixed number of combinations, and select the best based on Accuracy. Additionally, for non-cooperative settings, we trained and found the best local models for each owner in the same way and averaged the best results over different owners (i.e., IDU, ODU, NOS vendors) and standard deviation capturing variation across these different owners' local models.

## D. Results Discussion

Figures 6 and 7 show the Accuracy and F1-score (with Std Dev) for all proposed VFL and baseline scenarios. For the non-cooperative scenario, the results reflect the mean performance of the best local models from each data owner. Specifically, the 3-VS-XGB metrics represent the mean performance across individual models from IDU, ODU, and NOS, with the standard deviation capturing variation across these different owners. Due to space constraints, per-class metrics are not detailed, but class imbalance on the dataset causes no significant degradation, with performance variation across classes under 5% in all cases. Note that XGB and FedTree are addressed together as GBDT-based when discussing results.

**Multi-Vendor vs. Single Vendor.** We observe that both 2-VS and 3-VS FL scenarios reach up to 96.46% in Accuracy and 96.44% in F1-Score and SVS scenarios reach up to 96.40% in Accuracy and 96.49% in F1-Score. SVS-NN and SVS-XGB centralized models perform similarly on their best-performing models (e.g., the difference is only 0.18% in F1-Score). Both SplitNN and FedTree models show slight differences across all scenarios compared to centralized baselines, confirming VFL effectiveness. Additionally, FL scenarios perform significantly better compared to non-cooperative scenarios (e.g., up to

TABLE I  
TRAINING TIME (SECONDS)

| 2VS     |         |                     | 3VS     |         |                     | SVS  |      |
|---------|---------|---------------------|---------|---------|---------------------|------|------|
| SplitNN | FedTree | FedTreeHE           | SplitNN | FedTree | FedTreeHE           | NN   | XGB  |
| 9.05    | 0.62    | 64.84<br>(14.82 HE) | 3.53    | 0.46    | 77.41<br>(23.55 HE) | 1.70 | 0.46 |

13.04% higher Accuracy and 15.19% higher F1-Score for 3-VS, and up to 8.26% higher Accuracy and up to 9.13% higher F1-Score for 2-VS) where vendors only use their data to develop a local model without collaborating.

**SplitNN vs. FedTree.** GBDT-based models outperform NN-based models consistently both on Accuracy and F1-Score, except for the centralized scenario where SVS-NN performs slightly better than SVS-XGB with higher variance over folds. Moreover, FedTree model performance is not affected even if HE is integrated, which demonstrates that privacy can be guaranteed with no performance penalty.

**2-VS vs. 3-VS.** Figures 6 and 7 show a negligible performance difference between 2-VS-SplitNN and 3-VS-SplitNN, which demonstrates that the proposed approaches apply to two vendors and three vendors. In the case of non-cooperative, 2-VS performs better than 3-VS regardless of the model type used. This aligns with expectations since the IDU-NOS Vendor has a larger data chunk that drives performance metrics above (e.g., up to 5.31% in Accuracy and 7.02% in F1-Score).

**Training Time.** Tab. I reports the training time in seconds for all considered scenarios. Depending on the model hyperparameters, VFL scenarios take longer training time compared to SVS (0.62s vs 0.46s in best case, and 77.41s vs. 0.46s in worst case). Among NN-Based scenarios, 2-VS takes the longest training time due to its more complex architecture (i.e., several hidden layers). Across the FedTree-based scenarios, 2-VS training time is longer compared to 3-VS. As a larger number of partitions results in smaller clients' data sizes, hence 3-VS has lower training time than 2-VS. Higher privacy protection increases training time, with FedTreeHE requiring up to 168 times more than its baseline, confirming that encryption increases complexity substantially. Encryption and decryption alone account for 15–30% of total training time, even on smaller datasets.

## VI. CONCLUSION

We investigated the application of SplitNN-based and FedTree-based VFL for hardware failure-cause identification in disaggregated microwave networks. We consider equipment alarms from a real microwave network and compare multi-vendor (2-VS and 3-VS) to single-vendor deployment scenarios. The difference in Accuracy and F1-Score between centralized and federated scenarios is consistently within 1%, regardless of the deployment strategies or model types used, proving that the proposed FL-based scenarios preserve classification performance while minimizing sensitive data leakage. In future work, a tailored privacy-preserving protocol will be developed to secure intermediate cuts in SplitNN. Additionally, DP will be applied to histogram concatenation on the server side in FedTreeHE scenarios to mitigate leakage risks, as the server has access to the global model and labels. These enhancements will strengthen the overall threat model.

## REFERENCES

- [1] Ericsson, "Microwave outlook report 2022," Ericsson, Technical Report, 2022, accessed: 2024-08-11. [Online]. Available: <https://www.ericsson.com/en/reports-and-papers/microwave-outlook/reports/2022>
- [2] —, "Microwave outlook report 2023," Ericsson, Technical Report, 2023, accessed: 2024-11-01. [Online]. Available: <https://www.ericsson.com/en/reports-and-papers/microwave-outlook>
- [3] F. Lateano, O. Ayoub, F. Musumeci, and M. Tornatore, "Machine-learning-assisted failure prediction in microwave networks based on equipment alarms," in *2023 19th International DRCN Conference*, 2023, pp. 1–7.
- [4] N. Cicco, M. Ibrahim, O. Ayoub, F. Bruschetta, M. Milano, C. Passera, and F. Musumeci, "Asap hardware failure-cause identification in microwave networks using venn-abers predictors," *IEEE TNSM*, vol. PP, pp. 1–1, 10 2024.
- [5] M. Polese, M. Dohler, F. Dressler, M. Erol-Kantarci, R. Jana, R. Knopp, and T. Melodia, "Empowering the 6g cellular architecture with open ran," *IEEE JSAC*, vol. 42, no. 2, pp. 245–262, 2024.
- [6] T. Tandel, O. Ayoub, F. Musumeci, C. Passera, and M. Tornatore, "Federated-learning-assisted failure-cause identification in microwave networks," in *2022 12th RNDM*, 2022, pp. 1–7.
- [7] L. Yang, D. Chai, J. Zhang, Y. Jin, L. Wang, H. Liu, H. Tian, Q. Xu, and K. Chen, "A survey on vertical federated learning: From a layered perspective," 2023. [Online]. Available: <https://arxiv.org/abs/2304.01829>
- [8] N. Di Cicco, M. Ibrahim, F. Musumeci, F. Bruschetta, M. Milano, C. Passera, and M. Tornatore, "Machine learning for failure management in microwave networks: A data-centric approach," *IEEE TNSM*, vol. 21, no. 5, pp. 5420–5431, 2024.
- [9] Telecom Infra Project, "Wireless backhaul," 2024, accessed: 2024-11-01. [Online]. Available: <https://telecominfraproject.com/wirelessbackhaul/>
- [10] —, *Disaggregated Cell Site Gateway: Technical Specification*, Telecom Infra Project, Sep. 2019, accessed: 2024-11-02. [Online]. Available: <https://telecominfraproject.com/oopt/>
- [11] F. Musumeci, L. Magni, O. Ayoub, and et al., "Supervised and semi-supervised learning for failure identification in microwave networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 2, pp. 1934–1945, 2021.
- [12] L. Pan, J. Zhang, P. P. Lee, M. Kalander, J. Ye, and P. Wang, "Proactive microwave link anomaly detection in cellular data networks," *Computer Networks*, vol. 167, p. 106969, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1389128619307662>
- [13] O. Ayoub and et al., "Explainable artificial intelligence in communication networks: A use case for failure identification in microwave networks," *Computer Networks*, vol. 219, p. 109466, 2022.
- [14] O. Ayoub, F. Musumeci, F. Ezzeddine, C. Passera, and M. Tornatore, "On using explainable artificial intelligence for failure identification in microwave networks," in *2022 ICIN Conference*, 2022, pp. 48–55.
- [15] H. Zhang, H. Zhou, M. Elsayed, M. Bavand, R. Gaigalas, Y. Ozcan, and M. Erol-Kantarci, "On-device intelligence for 5g ran: Knowledge transfer and federated learning enabled ue-centric traffic steering," *IEEE TCCN*, vol. 10, no. 2, pp. 689–705, 2024.
- [16] F. Ezzeddine, O. Ayoub, D. Andreoletti, M. Tornatore, and S. Giordano, "Vertical split learning-based identification and explainable deep learning-based localization of failures in multi-domain nfv systems," in *2023 IEEE Conference on Nfv-SDN*, 2023, pp. 46–52.
- [17] K. Wei, J. Li, and et al., "Vertical federated learning: Challenges, methodologies and experiments," 2024. [Online]. Available: <https://arxiv.org/abs/2202.04309>
- [18] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, and Q. Yang, "Secureboost: A lossless federated learning framework," 2021. [Online]. Available: <https://arxiv.org/abs/1901.08755>
- [19] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?" 2022. [Online]. Available: <https://arxiv.org/abs/2207.08815>
- [20] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," 2021. [Online]. Available: <https://arxiv.org/abs/2106.03253>
- [21] I. Ceballos, V. Sharma, E. Mugica, A. Singh, A. Roman, P. Vepakomma, and R. Raskar, "Splitnn-driven vertical partitioning," 2020.
- [22] Q. Li, Z. Wu, Y. Cai, Y. Han, C. M. Yung, T. Fu, and B. He, "Fedtree: A federated learning system for trees," in *Proceedings of MLSys*, D. Song, M. Carbin, and T. Chen, Eds., vol. 5. Curran, 2023, pp. 89–103.