

Toward Power-Aware Federated Activity Detection in Cell-Free Massive MIMO-Based mMTC Networks

ALI ELKESHAWY¹ (Student Member, IEEE), WAEL JAAFAR² (Senior Member, IEEE),
HAÏFA FARÈS¹ (Member, IEEE), AND AMOR NAFKHA¹ (Senior Member, IEEE)

¹IETR, UMR CNRS 6164, CentraleSupélec, Rennes, 35576 Cesson-Sévigné, France

²Department of Software and Information Technology Engineering, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada

CORRESPONDING AUTHOR: A. ELKESHAWY (ali-fekry-ali-hassan.elkeshawy@centralesupelec.fr)

This work was supported in part by French National Research Agency within the framework of the Power-efficient and IOW complexity approaches for cEll-free massive MIMO Networks (POSEIDON) Project under Grant ANR-22-CE25-0015, in part by the Collège Doctoral Bretagne through an International, and in part by the Mobility Centre national de la recherche scientifique (CNRS) Research Group [Groupement de recherche (GDR)] Information Apprentissage Signal Image viSion (IASIS)

ABSTRACT The rapid growth of massive machine-type communications (mMTC), combined with advances in edge intelligence, is paving the way for low-latency, low-overhead connectivity. However, the sporadic nature of device activity in mMTC scenarios calls for efficient methods to determine which devices are active at any given time. This motivates collaborative learning within a cell-free massive multiple-input multiple-output (CF-mMIMO) architecture, where the wide geographical distribution of access points (APs) and their joint coordination make distributed learning efficient and secure. Consequently, federated learning (FL) emerges as a promising solution. Indeed, FL enables participants to train a shared model without exchanging raw local data, thereby enhancing data privacy at the AP side and lowering the fronthaul load while leveraging heterogeneous, location-dependent data. The present study proposes a novel FL framework where the CF-mMIMO participants are APs. Due to differences in device behavior, mobility patterns, and environmental factors across the network, the data collected at each AP is often non-independent and non-identically distributed (non-IID). This heterogeneity slows down the convergence of standard FL training and increase variability among client updates, particularly under heterogeneous radio feature distributions. To address this, we propose a client selection strategy that prioritizes APs based on their average received signal power. Our approach shows competitive performance compared to baseline methods, while also addressing the scalability and privacy requirements of mMTC systems. Furthermore, our study analyzes the fairness achieved by APs across devices and presents a representative, percentage-scale analysis of power-consumption gains relative to detection performance when some APs are dropped (i.e., taken out of service), examining two AP-dropping strategies. These results bring valuable insights and set guidelines towards the implementation of FL-based activity detection in CF-mMIMO networks.

INDEX TERMS Activity detection, cell-free massive MIMO, federated learning, grant-free random access, massive machine-type communications.

I. INTRODUCTION

A. BACKGROUND

THE rapid growth of the Internet of Things (IoT) and the emergence of massive machine-type communications (mMTC) have introduced unprecedented challenges in the design of next-generation wireless networks [1]. Massive MTC scenarios are characterized by a large number of

sporadically active devices, with stringent requirements on energy efficiency, scalability, and low signaling overhead. Although cellular massive multiple-input multiple-output (mMIMO) has shown strong potential for supporting such large-scale connectivity [2], [3], its practical deployment in ultra-dense environments presents significant challenges in real-world implementations.

In this context, cell-free massive multiple-input multiple-output (CF-mMIMO) has emerged as a promising architecture [4]. In CF-mMIMO, a large number of distributed access points (APs) jointly serve numerous devices across a wide area without the traditional notion of cell boundaries. This architecture provides high spectral efficiency and improved coverage. Leveraging the device-centric concept with coherently coordinated transmission and power control, it reduces inter-device interference and improves reliability, making it particularly well-suited to the irregular traffic patterns and spatial distribution of devices in mMTC systems. A key task in mMTC-enabled CF-mMIMO is device activity detection (AD), i.e., identifying which devices are active in a given time slot. Traditional grant-based random access (GB-RA) faces limitations in this context due to the scarcity of orthogonal resources. In particular, the finite coherence time and limited pilot sequence length often lead to access collisions when multiple devices select the same resources, resulting in failed transmissions [5]. To overcome these limitations, recent works explored grant-free random access (GF-RA) [6]. In GF-RA, devices can access the network without explicit scheduling requests. Instead, they are assigned non-orthogonal pilot sequences, thereby avoiding the strict constraints of orthogonal pilot assignments and enabling efficient support for large-scale connectivity despite the shortage of coherence time [7].

Achieving high detection performance in mMTC-enabled CF-mMIMO systems is particularly challenging due to the heterogeneous nature of device distributions, varying channel conditions, and the large-scale nature of the network. These factors naturally give rise to non-independent and identically distributed (non-IID) data in the network across access points, making centralized learning approaches difficult to generalize effectively. Although centralized solutions may achieve strong performance, they typically incur excessive fronthaul overhead and raise serious data privacy concerns, especially when sensitive device-related data must be shared with a central processor.

To address these challenges, it is essential to adopt methods that ensure both high detection performance and data privacy within CF-mMIMO networks. Traditional machine learning (ML) frameworks in telecommunications rely on a centralized cloud architecture, where raw data from devices is transmitted, stored, and processed in a central data center. However, such approaches not only increase latency and signaling overhead but also raise concerns regarding data privacy, particularly if data are exposed to unauthorized parties. Edge computing has emerged as a partial solution by enabling data processing at distributed nodes [8], yet it does not eliminate the need for secure and collaborative learning across distributed datasets.

A promising solution is Federated Learning (FL), a recently developed paradigm that enables collaborative model training without requiring the exchange of raw data [9], [10]. In the context of CF-mMIMO, FL allows each AP to locally train models using its own observations and

only share model updates with a coordinating server. This decentralized approach reduces fronthaul requirements and the impact of non-IID data distributions, and preserves data privacy, making it fit for large-scale heterogeneous wireless networks.

B. LITERATURE REVIEW

Accurate device activity detection is pivotal in mMTC systems, as it conditions subsequent channel estimation and, ultimately, reliable data transmission. Since only a small subset of devices is active at any given time, the detection task is inherently linked to the sparse activity patterns of machine-type devices. To address this challenge, sparse support recovery methods have been widely investigated, as they allow efficient identification of the few active devices within a massive pool of potential transmitters. Broadly, these methods fall into three main families: compressive sensing (CS)-based techniques, covariance-based (CV) methods, and data-driven (DD) approaches.

In the CS framework, AD is modeled as the recovery of a sparse signal from limited observations [11], [12]. In GF-RA, the received signals at APs serve as measurements, while the sparse vector represents sporadic device activity. Greedy pursuit strategies and approximate message passing (AMP) are widely used for sparse recovery [13], [14]. Alternatively, optimization-based approaches such as least absolute shrinkage and selection operator (LASSO) [15] have led to iterative solvers like iterative soft-thresholding algorithm (ISTA) and its accelerated version, FISTA [16]. More advanced schemes, such as the ones in [17], extend CS methods to support non-coherent mMIMO systems without scheduling or explicit grants.

Within the CV framework, AD in CF-mMIMO has been modeled as an inference problem using the covariance of received pilots. In [18], the authors investigated this direction by formulating device detection under GF-RA as a maximum likelihood problem. Their earlier work [19] introduced a simple baseline using the dominance of a single AP. Then, they extended it in [18] with a clustering-based method that exploits spatial correlation to improve detection accuracy in dense networks, though at the cost of higher computational complexity. However, under imperfect channel estimation the performance deteriorates and can even collapse in the presence of practical impairments, as proven in our previous work [20].

For centralized DD approaches, several studies have investigated deep learning (DL) solutions for AD. In the cellular (co-located) scenario, authors of [21] introduced a dual-strategy framework with two DL algorithms for sparse support recovery. Their approach included a feature-selection step that exploits pilot sequences with favorable cross-correlation to guide the neural network design, and it evaluates co-located configurations to enhance detection performance. From co-located to cell-free (CF) networks, several works have been proposed. For instance, authors of [22] studied CF networks under a fully grant-free scheme

with non-orthogonal pilots, employing a centralized convolutional neural network (CNN) approach to aggregate signals from distributed APs, thus outperforming CV-based methods. Previously, we introduced in [23] a lightweight centralized learning approach for CF networks and examined alternative deployment strategies (centralized vs. decentralized inference). A weighting-based post-detection fusion up to the global decision improves performance, confirming CF gains over cellular (better connectivity, lower outage). Empirical evidence indicates that DD detectors outperform iterative baselines in both detection performance and response time, supporting the use of DL in mMTC.

Despite the progress achieved by these methods, several challenges remain open in mMTC-enabled CF-mMIMO systems, including non-IID data distributions, network heterogeneity, and data-privacy concerns. Moreover, recent works highlighted the limitations of Gaussian-based AD methods under non-Gaussian noise conditions. For instance, the authors in [24] proposed robust covariance-based AD methods for massive random access by replacing the Gaussian loss function with robust ones, thus improving detection reliability under heavy-tailed and impulsive noise. Similarly, the authors in [25] proposed a robust covariance-learning-based matching pursuit algorithm for grant-free random access. Their method leveraged a robust loss function into the Gaussian negative log-likelihood formulation and used a matching pursuit framework to greedily select active users.

From the learning-based perspective, FL represents a promising distributed paradigm for handling non-IIDness and heterogeneous settings in CF-mMIMO. FL and CF-mMIMO share a common foundation in distributed architectures. Indeed, while CF-mMIMO leverages geographically distributed APs for network coverage, FL relies on distributed learning across clients. This conceptual synergy makes FL well-suited for CF-mMIMO, as both frameworks are inherently aligned to deal with decentralized environments. Moreover, in practical deployments, APs may belong to different operators, in which a centralized approach would require cross-domain raw-data sharing, often infeasible due to contractual, regulatory, and privacy constraints. This further highlights the importance of distributed learning, which preserves data locality and mitigates the limitations associated with centralized data sharing.

In recent years, FL has been integrated into CF-mMIMO in several ways to address diverse challenges. For instance, FL has been applied to accelerate energy-efficient transmission strategies [26], optimize the trade-off between training time and energy consumption, and improve resource allocation under practical system constraints [10]. Beyond communications, FL has also been exploited for malicious behavior detection and defense mechanisms, where client selection strategies play a key role in strengthening robustness and data privacy [27]. Moreover, outside the wireless domain, healthcare applications such as patient monitoring, gait analysis, and cognitive disease detection have

demonstrated the utility of FL in human activity recognition tasks [28]. Its successful use in analogous problems across other domains demonstrates the relevance of FL to wireless communications, especially for the device AD problem. In the wireless network context, only our previous work [29] has proposed an FL-based framework for AD in CF-mMIMO. We opted for a classical FedAvg-based FL bringing insights about the performances under a homogeneous IID setting only.

Motivated by the need for distributed processing that mitigates centralized bottlenecks and single points of failure in a secure manner, this work explores the integration of FL into CF-mMIMO for device AD. In the proposed framework, APs learn from their local data and share only model updates with the central server for aggregation, without transmitting raw local data. Unlike standard approaches [29], our proposed method introduces a power-aware client selection strategy, where APs are prioritized based on criteria on their average received signal power. This mechanism ensures that the most informative APs participate in each federated round, thus improving detection accuracy while reducing communication overhead. By leveraging the distributed nature of CF-mMIMO and the collaborative learning of FL, our approach strengthens robustness under non-IID conditions while also meeting the scalability and privacy requirements of mMTC systems. Table 1 summarizes the differences between the related works and ours.

C. PAPER CONTRIBUTIONS

Motivated by the above discussion, this paper proposes a novel FL-based AD in GF-RA CF-mMIMO mMTC networks. Our contributions can be summarized as follows:

- *Power-aware FL framework with client selection:* A power-aware FL setup is designed in which APs act as clients, tailored to GF-RA with non-orthogonal pilots and sporadic device activity. A client-selection mechanism chooses APs above or below the per-round mean of the APs' average received power, yielding two complementary selection policies.
- *IID vs. non-IID scenarios:* Developed FL models are tested in both IID and non-IID settings to capture the effect of spatial heterogeneity (dense vs. sparse regions) and device distributions, creating realistic deployments and highlighting FL robustness and generalizability.
- *Fairness analysis:* Fairness across APs/devices is evaluated using Jain's index to quantify how equitably devices are served for each detection algorithm and under different IID/non-IID settings.
- *Energy-performance trade-off via AP deactivation:* Recognizing fairness considerations, our study examines AP "drop-off" mechanisms to illustrate how switching off APs can reduce network consumption at the cost of detection degradation, offering guidance on policy choices under energy constraints.

TABLE 1. Summary of related work.

References	Topology	Dataset Setting	Model Type	Deployment	Client Select.	Data Privacy	Detection Strategy
[13]–[16]	–	–	Model-driven	Centralized	–	Not addressed	Compressive sensing
[18], [19]	CF-mMIMO	IID	Model-driven	Centralized	–	Not addressed	Covariance-based (maximum-likelihood)
[24], [25]	Cellular	Non-IID	Model-driven	Centralized	–	Not addressed	Robust covariance-based AD under heavy-tailed/impulsive noise
[20]	CF-mMIMO	IID	Data-driven	Centralized	–	Not addressed	DMLP
[21]	Cellular	IID	Data-driven	Centralized	–	Not addressed	DL (CNN and DMLP)
[22]	CF-mMIMO	IID	Data-driven	Centralized	–	Not addressed	CNN
[23]	CF-mMIMO	IID	Data-driven	Centralized	–	Not addressed	Lightweight learning (SLP)
[29]	CF-mMIMO	IID	Data-driven	Distributed	–	Achieved	FedAvg-based FL framework
Our work	CF-mMIMO	IID and non-IID	Data-driven	Distributed	✓	Achieved	Power-aware FL framework

The remainder of this paper is structured as follows: Section II presents the network and communication models and highlights the constraints that induce heterogeneity in the system. Section III details the proposed FL framework, including the client data structures, model architecture, FL process, and AP-client selection criteria. Section IV presents the numerical results, demonstrating the effectiveness of the proposed framework under different deployment scenarios. Finally, Section V concludes the paper.

II. NETWORK AND COMMUNICATION MODELS

We consider a CF-mMIMO network with GF-RA deployed over a $D \times D$ km² square area, divided into four equal sections, as illustrated in Fig. 1. The network consists of M APs, each equipped with N antennas, collaboratively serving K single-antenna devices. The latter follow the mMTC paradigm, characterized by a large number of low-complexity, sporadically active devices. For later reference, the four sections are denoted as the first, second, third, and fourth sections, corresponding respectively to the lower-left, upper-left, upper-right, and lower-right quadrants in Fig. 1, respectively.

In our modeling, two scenarios are considered, namely the IID and non-IID ones. In the IID case, both APs and devices are drawn from a uniform spatial distribution, which directly reflects the statistical meaning of IID homogeneity (all locations have the same probability, and no spatial structure or imbalance is introduced). In contrast, the non-IID scenario (e.g., Fig. 1) captures spatial heterogeneity typical of real deployments, where some regions exhibit higher

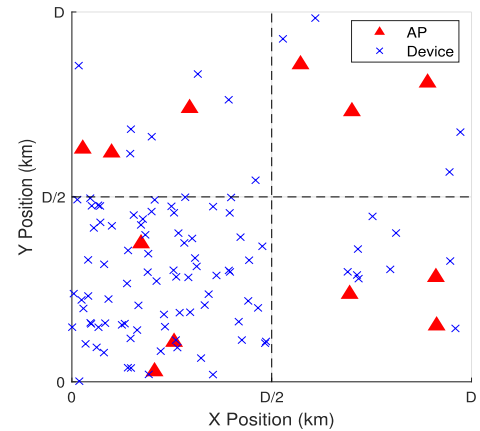


FIGURE 1. Spatial distribution of devices (blue crosses) and APs (red triangles) in a $D \times D$ km² area. Axes indicate X and Y positions in km. Each quadrant represents one-fourth of the deployment area.

device concentration (80% of the devices are concentrated in the first section (bottom left), while the remaining 20% are randomly distributed across the other three sections). This setup reflects real-world conditions in which certain areas (e.g., city centers) may experience higher device density than others at specific times of day. Specifically, the non-IID nature of the data in our study is characterized primarily in terms of feature skew. The features, which correspond to the signal power received by each AP, exhibit significant variation across APs due to the geographical distribution of devices. APs located in dense areas tend to receive

stronger signals from nearby active devices, while APs in less dense areas receive weaker signals from devices located farther away. This disparity in feature distribution caused by differences in device proximity to APs leads to feature skew, which we consider the primary source of non-IIDness in our scenario. It is worth noting that while devices may change their positions over time, the density pattern is preserved, i.e., the first section maintains approximately 80% of the total device population. However, the specific devices located in each section can vary over time. A device initially located in the dense section may move to a different section, while still preserving the overall density proportions. Furthermore, all APs are connected to a central processing unit (CPU) via high-speed, infinite-capacity, error-free fronthaul links [30]. Devices are synchronized and are simultaneously served by a cluster of APs of finite size T , using the same time-frequency resources.

Due to the sporadic nature of traffic in mMTC, only a small subset of devices is active at any given time. The activity of the k -th device is represented by $a_k \in \{0, 1\}$, where $a_k = 1$ indicates activity and $a_k = 0$ inactivity. The activation probability $\epsilon \ll 1$ follows a Bernoulli distribution, resulting in a sparse activity vector $\mathbf{a} = [a_1, a_2, \dots, a_K]$. Considering independent channel realizations between devices and access points, the channel gain $g_{mk}^{(n)}$ between the n -th antenna of the m -th access point and the k -th device is modeled as:

$$g_{mk}^{(n)} = \beta_{mk}^{1/2} \times h_{mk}^{(n)}, \quad (1)$$

where β_{mk} denotes the large-scale fading (LSF) coefficient, modeled according to the micro-cell propagation model specified in the 3GPP technical report [31]. The small-scale fading (SSF) component is represented by $h_{mk}^{(n)} \sim \mathcal{CN}(0, 1)$, where the magnitude $|h_{mk}^{(n)}|$ corresponds to Rayleigh fading.

Each active device transmits a non-orthogonal pilot sequence $\mathbf{s}_k \in \mathbb{C}^{L \times 1}$, where L is the pilot sequence length. The received signal $\mathbf{y}_{mn} \in \mathbb{C}^{L \times 1}$ at the n -th antenna of the m -th access point, is represented as [20]:

$$\mathbf{y}_{mn} = \sum_{k=1}^K a_k \rho_k^{1/2} g_{mk}^{(n)} \mathbf{s}_k + \mathbf{w}_{mn}, \quad (2)$$

where ρ_k is the transmit power of the k -th device, and $\mathbf{w}_{mn} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_L)$ is the additive white Gaussian noise (AWGN) vector, with σ^2 denoting the noise variance and $\mathbf{I}_L$ the $L \times L$ identity matrix. For the remainder of this work, we consider the signal received at the m -th AP during the uplink phase across all antenna elements, denoted as $\mathbf{Y}_m = [\mathbf{y}_{m1}, \mathbf{y}_{m2}, \dots, \mathbf{y}_{mN}] \in \mathbb{C}^{L \times N}$.

Note that all data generations throughout the rest of this work is based on the received signal structure $\mathbf{Y}_m$ described above. This formulation is used to generate data across multiple random access slots for each AP, where variations in device and AP distributions are introduced.

III. PROPOSED FL-BASED AD IN CF-MMIMO

The primary objective of the device AD model is to estimate the activity indicator vector $\mathbf{a} = [a_1, \dots, a_K]$

TABLE 2. Notations.

Notation	Description
$\tilde{\mathbf{a}}$	Predicted activity output vector of the SLP
$\hat{\mathbf{a}}$	Estimated activity descriptor in its binary domain
$\mathbf{a}$	True activity descriptor of the K devices
a_k	Activity status of the k -th device (1 = active, 0 = inactive)
$\hat{a}_k$	Weighted prediction for device k after weighting process
$\tilde{\mathbf{a}}_k^{(M)}$	M predicted activity values for device k from all APs
$\tilde{\mathbf{a}}_k^T$	T best predicted activity values for device k selected based on clustering
$\mathcal{C}_r$	Set of all r -chunks across the M AP-clients
$\mathcal{D}_m$	Full local training dataset for AP-client m , composed of N_c chunks
$\mathcal{D}_{m,n_c}$	Chunk n_c extracted from AP _{m} dataset $\mathcal{D}_m$
$\mathcal{D}^{(r)}$	Collection of all data samples from selected AP-clients participating in round r
$g_{mk}^{(n)}$	Channel gain between device k and antenna n of AP _{m}
$h_{mk}^{(n)}$	Small-scale fading (SSF) component, $h_{mk}^{(n)} \sim \mathcal{CN}(0, 1)$
I	Size of the local training dataset $\mathcal{D}_m$
J	Number of random access slots per chunk (chunk size)
N_c	Number of chunks in each AP dataset
$\bar{P}_m^{(j)}$	Average received power at AP _{m} at random access slot j (Eq. 3)
$\bar{P}_m^{(r)}$	Average received power score at AP _{m} over all slots in chunk r (Eq. 4)
$\bar{P}^{(r)}$	Set of all AP-client power scores in round r , i.e., $\{\bar{P}_1^{(r)}, \dots, \bar{P}_M^{(r)}\}$
$\bar{P}_{\text{avg}}^{(r)}$	Threshold computed as mean of $\bar{P}^{(r)}$ for client selection in round r
$\mathbf{s}_k$	Pilot sequence of device k
$\mathcal{S}$	Set of all AP-clients in the network
$\mathcal{S}_{\text{above}}^{(r)}$	Subset of APs with average received power scores $\geq \bar{P}_{\text{avg}}^{(r)}$
$\mathcal{S}_{\text{below}}^{(r)}$	Subset of APs with average received power scores $< \bar{P}_{\text{avg}}^{(r)}$
$\mathcal{S}_{\text{sel}}^{(r)}$	Subset of AP-clients selected to participate in local training during round r
$\mathbf{w}_{mn}$	AWGN noise vector at antenna n of AP _{m}
$W_{m,r}$	Local model parameters trained at AP _{m} in round r
W_r	Global model parameters obtained by aggregating selected $W_{m,r}$ at the end of round r
$\mathbf{y}_{mn}$	Signal received at antenna n of AP _{m}
$\mathbf{Y}_m$	Signal matrix at AP _{m} , stacked over N antennas (columns correspond to antennas)
α_{tk}	Weight associated with AP t and device k in the cluster-based weighting
ρ_k	Transmit power of device k
β_{mk}	Large-scale fading (LSF) coefficient between AP _{m} and device k

from the received signal $\mathbf{Y}_m$. To accomplish this, we adopt an FL-based approach as follows. As presented in Fig. 2, multiple APs act as clients, coordinated by a CPU, and collaboratively train a ML model while keeping their local data private. This approach is advantageous in mMIMO networks, where data is inherently heterogeneous and non-IID due to varying channel conditions, traffic patterns, and the spatial distribution of both APs and devices.

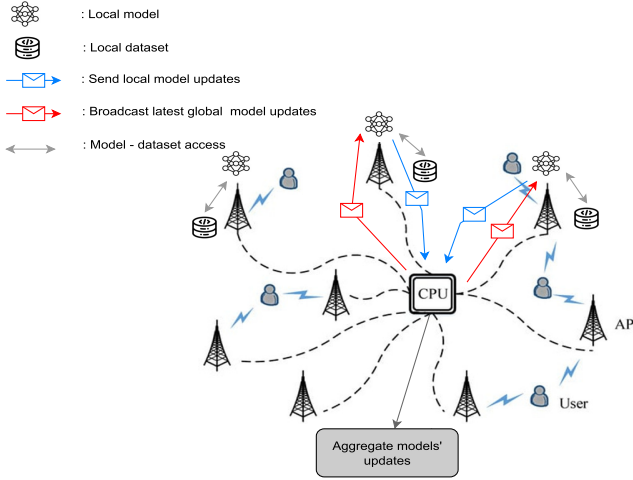


FIGURE 2. CF-mMIMO federated learning framework with secure aggregation updates.

A. CLIENT DATA STRUCTURE

Let $\mathcal{D}_m$ denote the training dataset associated with client m , i.e., AP_m . Each AP in the network, indexed by $m \in \mathcal{M}$ where $\mathcal{M}$ represents the set of APs, holds a local dataset of size $I = |\mathcal{D}_m|$. This dataset is divided into N_c chunks, such that each chunk has size $J = |\mathcal{D}_m|/N_c$. Accordingly, the full dataset for AP_m can be expressed as the union of its chunks $\mathcal{D}_m = \mathcal{D}_{m,1} \cup \mathcal{D}_{m,2} \cup \dots \cup \mathcal{D}_{m,N_c}$.

Each data chunk indexed by $n_c \in \{1, \dots, N_c\}$, denoted $\mathcal{D}_{m,n_c}$, consists of the signal samples received at AP_m over $|\mathcal{D}_m|/N_c$ random access slots, where each sample corresponds to a received signal matrix $\mathbf{Y}_m$. The signals correspond to eq.(2), where the variations of both fading components depend on the spatial distribution of AP m and all devices within chunk n_c . This structure ensures that the dataset reflects temporal and spatial variability observed across access slots.

Unlike conventional FL settings, where each client's dataset remains fixed throughout training and data heterogeneity is solely modeled by a global non-IID dataset distributed across clients, our approach introduces dynamic variability both within and across clients. This reflects a more realistic and complex non-IID scenario. Specifically, each data chunk captures a distinct spatial distribution of devices and APs, thereby modeling device mobility and environmental fluctuations. Specifically, within a given chunk, the spatial positions of devices and APs remain fixed, but their locations are randomized from one chunk to another. Although APs are constrained to remain within their designated sections of the area, their locations inside those sections may vary across chunks. In contrast, device locations are allowed to change freely between chunks without preserving their previous section assignments. Nevertheless, the overall density constraint is maintained. For instance, ensuring that 80% of the devices are concentrated in a single section of the area, as illustrated in Fig. 1.

In each chunk, data is generated across multiple random access slots. In every slot, the device activity pattern and small-scale fading vary. This chunk-based generation strategy is applied under both non-IID conditions (e.g., as in Fig. 1) and IID conditions, where devices and APs are uniformly distributed. The same structure is adopted for generating the testing dataset, which is set to 40% of the training dataset size. This approach closely imitates real-world conditions by capturing device mobility between chunks while maintaining consistency within each chunk, which is important for effective model learning and assessment.

B. FL MODEL ARCHITECTURE

As depicted in Fig. 2, all APs are connected to the CPU, which acts as the aggregator. Each AP is initialized with the same generic FL model created by the CPU. The model architecture, inspired by [23], follows a lightweight single-layer perceptron (SLP) structure, designed for efficiency and fast convergence in large-scale networks. Specifically, the model consists of an input layer representing the received signals $\mathbf{Y}_m$, followed by a single hidden layer ($Z = 1$) with $V = 512$ artificial neurons using the ReLU activation function. The output layer consists of K neurons with a Sigmoid activation function, predicting the target activity descriptor $\hat{\mathbf{a}} = [\hat{a}_1, \dots, \hat{a}_K]$, where each output value represents a probability in the interval $[0, 1]$.

We denote by $W_{m,r}$ the local model parameters of AP_m at FL round r . Furthermore, we assume that the CPU leverages a power-aware selection policy, which performs model aggregation by combining the local updates from the selected APs into a global model.

C. CLIENT SELECTION AND FL AGGREGATION

In the considered framework, each FL round is divided into two sub-rounds. The first sub-round r_1 consists of a power-aware client selection process, where the clients participating in the training phase are selected. In the second sub-round r_2 , local training is performed at the selected APs, assuming that the APs are equipped with edge-computing capabilities or connected to local processing units. The resulting local model updates are then transmitted to the central server for aggregation, as summarized in Algo. 1. Although FL avoids transmitting raw local data, the exchanged model updates may still leak information about the training data, for instance through gradient inversion attacks. Hence, secure transmission and additional privacy-preserving mechanisms, such as secure aggregation [32], differential privacy, or update perturbation, can be incorporated in practical deployments to further protect the shared model updates.

Each client dataset is organized into chunks, and training is scheduled so that each chunk n_c corresponds to one FL round r , ensuring that all chunks from all datasets are considered during training. The total number of rounds is equal to the total number of chunks, i.e., $R = N_c$. Following

Algorithm 1 Power-Aware Federated Learning Procedure

```

1: Input: Local datasets  $\{\mathcal{D}_m\}_{m=1}^M$  partitioned into chunks  $\mathcal{D}_{m,r}$ ; number of rounds  $R$ ; selection mode SelMode  $\in \{\text{above}, \text{below}\}$ 
2: Initialization:
3: - Initialize global model  $W_0$  at the CPU
4: Federated Learning Rounds:
5: for each round  $r = 1$  to  $R$  do
6:   Sub-round  $r_1$ : Client Selection
7:   for each client  $m = 1$  to  $M$  do
8:     - Compute average received power  $\bar{P}_m^{(r)}$  on  $\mathcal{D}_{m,r}$  (4)
9:   end for
10:  - Each AP sends its score  $\bar{P}_m^{(r)}$  to the CPU
11:  - CPU computes global threshold  $\bar{P}_{\text{avg}}^{(r)}$  (5)
12:  if SelMode = above then
13:    - CPU selects clients  $\{m : \bar{P}_m^{(r)} \geq \bar{P}_{\text{avg}}^{(r)}\}$  (6)
14:  else {SelMode = below}
15:    - CPU selects clients  $\{m : \bar{P}_m^{(r)} < \bar{P}_{\text{avg}}^{(r)}\}$  (7)
16:  end if
17:  Sub-round  $r_2$ : Local Training and Aggregation
18:  for each selected client  $m \in \mathcal{S}_{\text{sel}}^{(r)}$  do
19:    - Train local AP model on  $\mathcal{D}_m$ 
20:    - Send local AP update to CPU
21:  end for
22:  - CPU aggregates local AP updates to form  $W_r$  (9)
23: end for
24: Output: Final global model  $W_R$ 

```

the standard FL convention, an FL round is indexed by r rather than n_c so that r can be used to denote both the chunk index and the FL round. Let $\mathcal{C}_r$ denote the set of data chunks between all clients in round r , defined as $\mathcal{C}_r = \{\mathcal{D}_{1,r}, \mathcal{D}_{2,r}, \dots, \mathcal{D}_{M,r}\}$. Then, the sub-rounds are executed as follows.

1) SUB-ROUND 1: POWER-AWARE FL CLIENT SELECTION

In the first sub-round r_1 of round r , each client loads its corresponding chunk $\mathcal{D}_{m,r}$. The average received power at AP m over a single random access slot j is given by:

$$\bar{P}_m^{(j)} = \frac{1}{LN} \sum_{n=1}^N (\mathbf{y}_{mn}^{(j)})^H \mathbf{y}_{mn}^{(j)}. \quad (3)$$

where $\mathbf{y}_{mn}^{(j)} \in \mathbb{C}^{L \times 1}$ denotes the signal received at antenna n of AP m during random access slot j , and $(\cdot)^H$ represents the Hermitian (conjugate) transpose operation. Here, $\bar{P}_m^{(j)}$ provides an estimate of the expected received power at AP m during the slot, averaged over all antennas and pilot symbols. Dividing by LN ensures dimensional consistency, yielding a power value per symbol that is comparable across APs with different antenna counts and pilot lengths. This quantity is used as a received-energy score for client selection, rather than as a noise-free estimate of the useful signal power.

Indeed, it serves as a low-overhead received-energy indicator computed directly from the observations available at each AP before activity detection. This choice is motivated by the fact that client selection is performed before activity detection, when the active-device set and the clean signal component are not yet available. Under the considered AWGN model, all APs are assumed to experience the same noise variance σ^2 and use the same pilot length L and number of antennas N . Therefore, the expected noise contribution introduces a common offset in the AP scores and has limited impact on their relative comparison within the same FL round.

In each data chunk r of client m , the data is composed of J random access slots. For each random access slot $j \in \{1, \dots, J\}$, the signal received at AP m is denoted by $\mathbf{Y}_m^{(j)} = [\mathbf{y}_{m1}^{(j)}, \dots, \mathbf{y}_{mN}^{(j)}]$. The average received power score at AP m over all J random access slots within chunk r can be then expressed by:

$$\begin{aligned} \bar{P}_m^{(r)} &= \frac{1}{J} \sum_{j=1}^J \left[\frac{1}{LN} \sum_{n=1}^N (\mathbf{y}_{mn}^{(j)})^H \mathbf{y}_{mn}^{(j)} \right] \\ &= \frac{1}{J} \sum_{j=1}^J \bar{P}_m^{(j)}, \quad \forall m = 1, \dots, M. \end{aligned} \quad (4)$$

The arithmetic mean in eq.(4) provides a stable chunk-level score by averaging the slot-level received-energy measurements over J random access slots. This averaging reduces the impact of instantaneous variations due to device activity and small-scale fading, while remaining consistent with the local training process, where all chunk samples are used equally. Weighted averaging would require additional side information, such as instantaneous activity, slot-level signal-to-noise ratio (SNR), or channel reliability, which are unavailable before activity detection, besides increasing complexity.

Once the chunk-level received-energy scores are computed, they are used by the CPU to determine the participating APs in the next sub-round, resulting in a set of scores for round r given by $\bar{P}^{(r)} = \{\bar{P}_1^{(r)}, \bar{P}_2^{(r)}, \dots, \bar{P}_M^{(r)}\}$. Using these scores, the CPU selects the participants for training in round r (i.e., sub-round r_2) by applying a threshold $\bar{P}_{\text{avg}}^{(r)}$ equal to the mean score across all clients, as follows:

$$\bar{P}_{\text{avg}}^{(r)} = \text{mean}(\bar{P}^{(r)}) = \frac{1}{M} \sum_{m=1}^M \bar{P}_m^{(r)}. \quad (5)$$

Using eq.(5), the CPU can apply one of two participant selection strategies. In the first strategy, the CPU selects the subset of APs whose average received power scores in round r are greater than or equal to the threshold, defined as:

$$\mathcal{S}_{\text{above}}^{(r)} = \{m \in \mathcal{S} \mid \bar{P}_m^{(r)} \geq \bar{P}_{\text{avg}}^{(r)}\}, \quad (6)$$

where $\mathcal{S}$ denotes the set of all AP-clients in the network. This strategy tends to prioritize APs with comparatively stronger average link conditions.

In the second strategy, the CPU selects the complementary set, consisting of APs whose average received power scores fall strictly below the threshold, defined as:

$$\mathcal{S}_{\text{below}}^{(r)} = \mathcal{S} \setminus \mathcal{S}_{\text{above}}^{(r)} = \{m \in \mathcal{S} \mid \bar{P}_m^{(r)} < \bar{P}_{\text{avg}}^{(r)}\}. \quad (7)$$

Both strategies are evaluated and compared in the results section for the IID and non-IID scenarios, in order to assess their effectiveness under data heterogeneity.

2) SUB-ROUND 2: FL AGGREGATION

Following client selection in sub-round r_1 , in sub-round r_2 , the aggregator informs the selected clients, either from the set $\mathcal{S}_{\text{above}}^{(r)}$ or $\mathcal{S}_{\text{below}}^{(r)}$ (depending on the selection policy), that they will participate in the local training of round r . At the beginning of the training process (i.e., first round $r = 1$), the aggregator initializes a generic global model W_0 , which serves as the starting point for all clients' models. In each subsequent round $r > 1$, the selected clients download the most recent global model W_{r-1} from the aggregation server for initialization and synchronization. Then, each selected client $m \in \mathcal{S}_{\text{sel}}^{(r)}$, where $\mathcal{S}_{\text{sel}}^{(r)}$ denotes the subset of clients selected from $\mathcal{S}$ in round r regardless of the selection policy, performs multiple iterations of the adaptive moment estimation optimizer (Adam) [33] over all N_c mini-batches (chunks) in its local dataset $\mathcal{D}_m$. The training objective is the binary cross-entropy loss, since the device AD task is formulated as a binary classification problem. Subsequently, the local model parameters $W_{m,r}$ are updated as follows [27]:

$$W_{m,r} = \text{Adam}(W_{r-1}, \mathcal{D}_m) - W_{r-1}, \quad \forall m \in \mathcal{S}_{\text{sel}}^{(r)}. \quad (8)$$

The subtracted last term in eq.(8) emphasizes the relative change of the weights, following conventions used in prior FL literature [27]. In practice, the Adam optimizer is applied directly to the absolute weights at each client, and the resulting $W_{m,r}$ is communicated to the server for aggregation. The latter executes weights aggregation and produces a new global model W_r as:

$$W_r = W_{r-1} + \sum_{m=1}^{|\mathcal{S}_{\text{sel}}^{(r)}|} \frac{|\mathcal{D}_m|}{|\mathcal{D}^{(r)}|} W_{m,r}, \quad (9)$$

where $\mathcal{D}^{(r)} = \bigcup_{m=1}^{|\mathcal{S}_{\text{sel}}^{(r)}|} \mathcal{D}_m$ represents the collection of all data samples from the selected clients in round r , and $|\mathcal{S}_{\text{sel}}^{(r)}|$ denotes the number of selected clients in round r . For uniform weighting, where all local datasets across clients in the network have the same size, the aggregation rule can be simplified into:

$$W_r = W_{r-1} + \frac{1}{|\mathcal{S}_{\text{sel}}^{(r)}|} \sum_{m=1}^{|\mathcal{S}_{\text{sel}}^{(r)}|} W_{m,r}. \quad (10)$$

This process is repeated in every FL round until convergence or reaching the maximal number of rounds.

D. MODEL INFERENCE

At the end of the training phase, a final global model is obtained, which is then used by all APs in the network for inference. In this study, we aim to detect device activity in the CF-mMIMO architecture. Utilizing all APs for AD in the context of CF-mMIMO can be effective but is computationally expensive. Consequently, we opt for a clustering strategy, which is consistent with the device-centric property of CF-mMIMO, where each device is served by a cluster of T APs.

During a specific access slot, each AP_m applies the trained global model to its received signal $\mathbf{Y}_m$ and predicts the activity vector $\hat{\mathbf{a}}_m$, resulting in M activity descriptor vectors. In other words, for each device k , there are M predicted activity values $\hat{\mathbf{a}}_k^{(M)} = [\hat{a}_{1k}, \hat{a}_{2k}, \dots, \hat{a}_{Mk}]$, where $\hat{a}_{mk} \in [0, 1]$ is the prediction probability of the activity of device k by AP_m , $m = 1, \dots, M$. These vectors are collected at the CPU for post-processing and global device detection by following these steps:

1) For each device k , the CPU selects the T best predicted values $\hat{\mathbf{a}}_k^{(T)} = [\hat{a}_{1k}, \hat{a}_{2k}, \dots, \hat{a}_{Tk}]$, out of the M predictions, based on the large-scale fading coefficients β_{mk} known at the CPU.

2) A weighting process is then performed, where the goal of the weighting function is to compute a final prediction for device k by combining the predictions from the selected cluster of T APs. The weights are assigned according to the quality of the channel between the device and each AP. The weighted prediction for a specific device k , derived from its assigned cluster, can be written as:

$$\hat{a}_k = \sum_{t=1}^T \alpha_{tk} \hat{a}_{tk}, \quad (11)$$

where $\alpha_{tk} \in [0, 1]$ represents the weight assigned to the t -th AP for device k , $\forall t = 1, 2, \dots, T$, and $\hat{a}_k \in [0, 1]$. The weight α_{tk} reflects the LSF quality (signal strength or reliability) between the t -th AP and device k , and is defined by:

$$\alpha_{tk} = \frac{\beta_{tk}}{\sum_{t=1}^T \beta_{tk}}. \quad (12)$$

3) A hard-decision threshold $\tau \in [0, 1]$ is applied to the weighted prediction $\hat{a}_k$ to produce a binary decision for each device k . This threshold is selected according to service-specific Quality-of-Service (QoS) requirements, taking into account the allowable error probability (e.g., target false-alarm and miss-detection rates). The final activity decision $\hat{a}_k$ is a binary value, expressed as:

$$\hat{a}_k = \begin{cases} 0, & \text{if } \hat{a}_k < \tau, \\ 1, & \text{if } \hat{a}_k \geq \tau. \end{cases} \quad (13)$$

Finally, after selecting the APs cluster for each device, applying the weighting, and thresholding, the estimated activity vector for all devices is denoted as $\hat{\mathbf{a}} = [\hat{a}_1, \dots, \hat{a}_K]$.

IV. NUMERICAL RESULTS

In this section, we assess the performance of the developed FL-based approaches for activity detection in CF-mMIMO, called “Selection Above” and “Selection Below” depending on the used selection policy. Our methods are compared to several benchmarks, including the standard Vanilla federated averaging (“FedAvg”) benchmark, the federated proximal optimization method (“FedProx”) [34], the sparse recovery model-driven Approximate Message Passing (“AMP”) algorithm [16], and a distributed learning approach where clients train with local data only, without any collaboration (“Distributed”). All methods are assessed under near-realistic conditions in an uplink mMTC system deployed over a 0.25 km² area ($D = 0.5$ km), comprising $M = 12$ APs, each equipped with $N = 2$ antennas, and $K = 100$ devices, each active during the observation time with probability $\epsilon = 10\%$. The cluster size is fixed to $T = 4$, and the non-orthogonal pilot sequences have length $L = 40$. Two traffic distributions are considered: i) IID where APs and devices are uniformly distributed across the coverage area, and ii) non-IID as illustrated in Fig. 1.

For FL training, each client is assigned a local training dataset of $I = 2 \times 10^5$ samples of received signals. This dataset is partitioned into $N_c = 100$ chunks (mini-batches), with each client performing 2 local iterations per communication round. The total number of FL rounds is fixed as $R = 100$. Finally, the inference phase (i.e., testing) is conducted using a dataset comprising 80,000 samples (40% of the training dataset size).

For the FedProx benchmark, the proximal coefficient is set to $\mu = 10^{-3}$. This parameter controls the strength of the proximal regularization term added to the local objective, which restricts local model updates from moving away from the current global model. As discussed in [34], the choice of μ plays an important role in the convergence behavior of FedProx, i.e., a large value may overly constrain the local updates and slow convergence while a small value may have a negligible effect. In this work, FedProx is reported only in the ROC analysis (Subsection IV-B).

The global performance of device AD is evaluated using the receiver operating characteristic (ROC) curve, which depicts the trade-off between the detection probability (P_D) and the false alarm probability (P_{FA}). In addition, the Jain’s fairness index is evaluated to measure the equitability of performance across devices, with both P_D and P_{FA} considered in the analysis.

A. QUANTITATIVE ANALYSIS OF DATA HETEROGENEITY, FEDAVG CONVERGENCE, AND MODEL DRIFT

This analysis provides a quantitative assessment of the data heterogeneity across APs and its impact on the FedAvg learning process.

Fig. 3 examines the distribution of the average scores across all chunks for each AP during training under the IID and non-IID scenarios. In this study, the non-IIDness

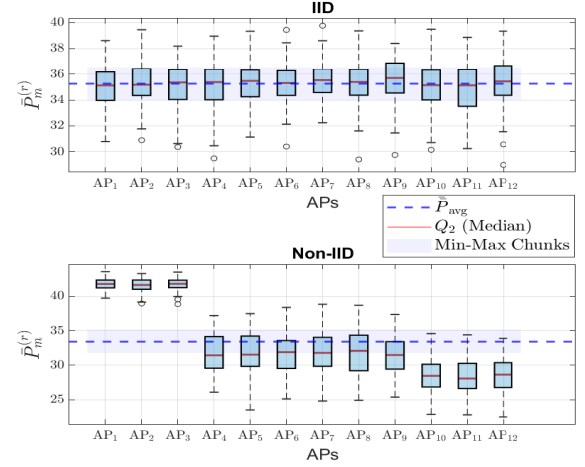


FIGURE 3. Interquartile range (IQR) of chunks’ average score per AP in IID and non-IID settings.

corresponds to feature skewness rather than label or quantity skewness. Indeed, the device-activity labels are generated according to the same activity model across APs, and each AP observes the same number of training chunks. However, due to the non-uniform spatial distribution of active devices, the received-signal statistics differ from one AP to another. Specifically, APs located close to dense device regions receive stronger average signal powers, whereas APs located in less dense regions observe weaker received powers. Hence, the heterogeneity appears in the input feature distribution, i.e., in the received-power-related features used for activity detection. The interquartile range (IQR) [35] is used to represent the variability of the average received-power score per AP, as given by eq. (4), computed over all chunks in the training dataset. The dashed blue line in each subplot of Fig. 3 represents the overall mean of the selection thresholds across all chunks and APs, defined as:

$$\bar{P}_{\text{avg}} \triangleq \frac{1}{RM} \sum_{r=1}^R \sum_{m=1}^M \bar{P}_m^{(r)} = \frac{1}{R} \sum_{r=1}^R \bar{P}_{\text{avg}}^{(r)}. \quad (14)$$

Plotting the threshold for each chunk individually is impractical, so the shaded region around the dashed line depicts the per-chunk variations, which remain approximately within ± 2 units from the overall mean. The red line indicates the second quartile (Q2) median, i.e., the midpoint of the N_c per-AP average received-power scores.

In the IID case (top, Fig. 3), the IQR values are uniform across APs, indicating that all clients exhibit similar dataset characteristics and average power scores. As a result, the selection policies (“Selection Above” and “Selection Below”) yield nearly identical performance in terms of ROC curves, as shown in Fig. 6. In contrast, in the non-IID case (bottom, Fig. 3), the IQR values vary significantly across APs, reflecting higher heterogeneity in client datasets. For example, the first three APs, serving the densest part of the network, consistently exhibit average power chunk scores

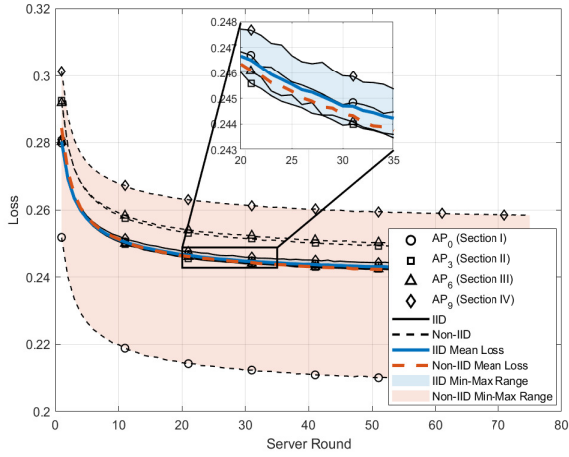


FIGURE 4. FedAvg convergence in terms of training loss across communication rounds and AP clients (under IID and non-IID conditions).

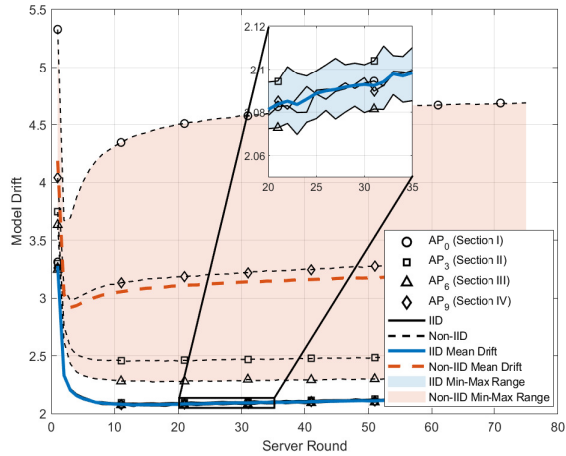


FIGURE 5. FedAvg model drift evolution across communication rounds and AP clients (under IID and non-IID conditions).

significantly above the threshold. Under “Selection Below” approach, these APs are almost never selected, whereas under “Selection Above” method, they participate in almost all FL rounds. Conversely, the last three APs display the opposite pattern, i.e., they are almost always selected in “Selection Below” and rarely in “Selection Above”. This analysis highlights the high impact of the selection threshold on client participation with respect to the scenario conditions.

The impact of feature heterogeneity is also observed in the convergence behavior of the FL framework. To do so, the training-loss evolution and the model drift are analyzed over the server communication rounds in Figs. 4–5. The training loss reflects how effectively the local models learn from their AP-specific data distributions, whereas the model drift measures how far each local model deviates from the global aggregated model.

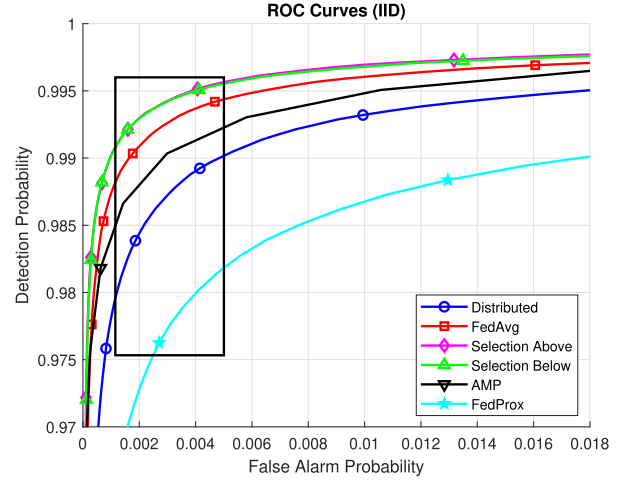


FIGURE 6. ROC analysis under IID conditions.

In Fig. 4, we plot the training loss for a representative number (not all) of APs, identified with markers. In addition, the mean training-loss is shown, together with a shaded min-max envelope that represents the range of loss values observed across all APs. The IID-related results are represented with a solid line, while non-IID results are shown with dashed line. In the IID scenario, the APs exhibit very similar convergence behavior. This is consistent with the uniform IQR distribution observed in Fig. 3, where all APs have comparable received-power score distributions. The shaded envelope around the “IID mean loss” remains narrow, indicating a limited variation among the APs. The training losses converge to similar values, close to approximately 0.24, which confirms that the IID setting leads to stable and balanced FedAvg convergence. In contrast, the non-IID scenario exhibits a wider shaded envelope, indicating a stronger variation among the AP losses. This larger spread reflects the fact that the APs do not learn from statistically equivalent feature distributions. APs located in dense device regions (e.g., Section I) benefit from stronger received-power and thus converge to a better loss value, below 0.22. In contrast, APs located in less denser regions (e.g., Section IV) converge to higher loss values, above 0.25. This difference indicates that the local optimization problems become AP-dependent under feature-skewed non-IID conditions. Although FedAvg still converges globally, the convergence is less uniform across clients, which demonstrates the limitation of standard FedAvg in non-IID settings. Fig. 5 shows the model drift evolution across the server rounds under IID and non-IID scenarios. Model drift measures the deviation between the local model of each AP and the global model obtained after server aggregation. For AP m at communication round r , the model drift is defined as:

$$d_{m,r} = \|\mathbf{W}_{m,r} - \mathbf{W}_r\|_F, \quad (15)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, which quantifies the magnitude of the difference between the local and global

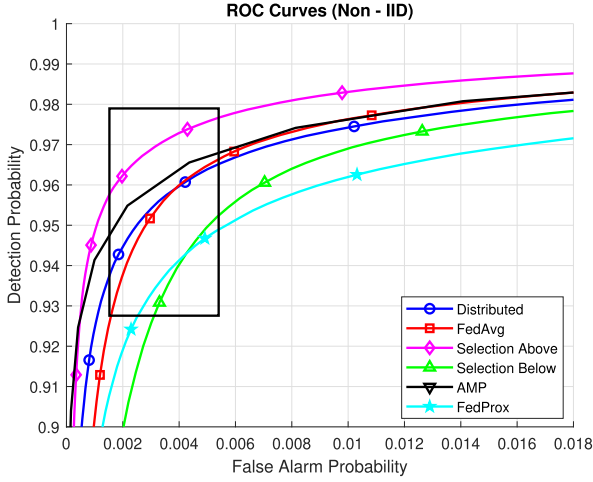


FIGURE 7. ROC analysis under non-IID conditions.

model parameters. Similarly to Fig. 4, only representative APs are plotted to avoid overloading Fig. 5. In the IID case, the model drift remains low with a narrow envelope, meaning that the local AP models remain close to the global aggregated model throughout the FL process. Since the APs observe similar feature distributions, their local updates are also similar, thus leading to model stability through training. In contrast, within the non-IID setting, both the model drift and envelope are significantly larger. This means that AP local models deviate significantly from the global model during training. The mean drift is impacted by the predominance of APs located in less dense regions.

These results prove that feature-skewed non-IIDness affects the behavior of FedAvg in two ways. First, it leads to non-uniform convergence across APs (wide loss envelope). Second, it increases the deviation between local and global models through training (higher model drift). Although FedAvg converges in non-IID conditions, the latter is achieved with unbalanced results for APs. This outcome motivates the adoption of data-aware client-selection strategies. Specifically, selection criteria based on the received-power score can prioritize APs with more informative and reliable feature distributions, thus improving the robustness of the FL framework in non-IID conditions.

B. ROC ANALYSIS

Motivated by the heterogeneity observed in the previous analysis, this section assesses the impact of client selection on the detection performance under both IID and non-IID scenarios. From the selection criteria perspective, Figs. 6 and 7 present the global performance after applying the ponderation strategy in the inference phase (vector $\hat{a}$), in terms of ROC curves, i.e., the detection probability as a function of the false alarm probability, in IID and non-IID scenarios, respectively. In the IID scenario (Fig. 6), results demonstrate the superiority of the proposed methods “Selection Above”

and “Selection Below” over the benchmarks, with the worst performance achieved by “FedProx” method. “FedAvg” performs strongly and surpasses “AMP”, thereby confirming the robustness of FL in device AD tasks. Nevertheless, the selection policies introduced in our approaches allowed superior performance. As mentioned in Fig. 3 (top), this is due to the APs exhibiting similar average score distributions across chunks. It is also observed that FedProx does not outperform FedAvg with the considered empirical setting of the proximal coefficient $\mu = 10^{-3}$, meaning that the proximal constraint does not provide any benefit; worse, it limits useful local adaptation. Fig. 7 illustrates the ROC performance under non-IID conditions. Compared to Fig. 6, the results differ due to the heterogeneity in local datasets. Indeed, the best performances are achieved by “Selection Above”. The superior performance of “Selection Above” is due to the spatial distribution of devices and the resulting signal quality at each AP. For instance, the local models trained at APs with high received power capture more informative patterns for activity detection, which significantly contributes to the global model aggregation. In contrast, low-power-receiving APs often observe sparse or noisy data, resulting in weaker local updates, thus contributing less to the global model.

Moreover, we notice a slight stability in “AMP”, whereas “FedAvg” proves less effective. For instance, when $P_{FA} < 4 \times 10^{-3}$, “Distributed” outperforms “FedAvg”. In contrast, selecting clients with low average power scores (“Selection Below”) leads to the poorest AD, highlighting that, in non-IID settings, low-power receiving clients’ contribution to FL is negligible. Finally, FedProx shows some improvement over “Selection Below” in the low false-alarm region ($P_{FA} < 4 \times 10^{-3}$), but it remains below FedAvg and the proposed power-aware selection strategies. This behavior confirms that proximal regularization alone is not sufficient to compensate for the spatial and channel-induced heterogeneity across APs.

Taking into account both IID and non-IID scenarios, we see that FL, when combined with the proper client selection policy (i.e., “Selection Above”), outperforms state-of-the-art methods. Moreover, detection performance for all studied approaches is consistently higher in the IID case, where the absence of heterogeneity across AP datasets ensures more balanced and representative contributions across clients.

Fig. 8 illustrates the per-AP local model accuracy over the server rounds. For clarity, Fig. 8 adopts the same concept as in Fig. 4, showing a shaded region with a mean curve over rounds for each IIDness setting. In the IID scenario, the AP performance is nearly uniform, reflected by the narrow width of the envelope around the mean accuracy. In the non-IID scenario, the performance varies widely among APs, as represented by the wide shaded region. Indeed, APs in low-density sections (e.g., section IV) exhibit the worst accuracy, whereas APs in high-density sections (section I) achieve the highest accuracy. These results agree with those in Fig. 4.

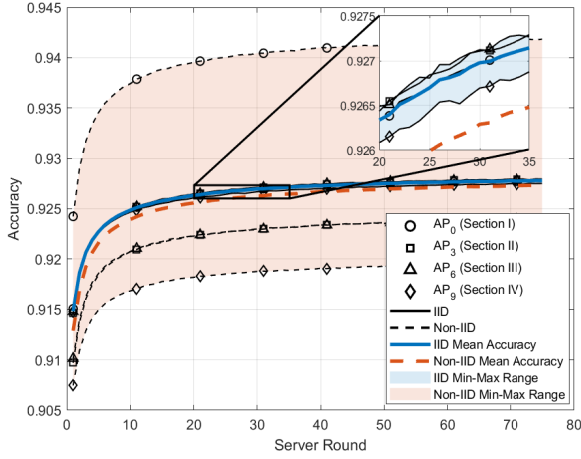


FIGURE 8. Per-AP training accuracy across communication rounds under IID and non-IID conditions.

C. AVERAGE JAIN INDEX ANALYSIS

While achieving high detection performance remains a primary objective, it is important to ensure that such performance is not achieved by disproportionately favoring certain devices. Indeed, a model achieving high global detection performance may still suffer from a lack of fairness among devices, particularly in heterogeneous systems. Incorporating Jain's index into our analysis, therefore, provides an additional perspective, highlighting whether an algorithm achieves its gains by serving all devices equitably or by benefiting only a subset of them.

We evaluate here the average Jain's index computed over the testing dataset for each AP in both IID and non-IID settings. Jain's index is a fairness measure ranging between 0 (worst case) and 1 (perfect fairness), commonly used to quantify how evenly a given resource or performance metric x is distributed among multiple entities. It is defined by

$$\mathcal{J}(x_1, x_2, \dots, x_K) = \frac{\left(\sum_{k=1}^K x_k\right)^2}{K \sum_{k=1}^K x_k^2}, \quad (16)$$

where, in our context, the entities represent the devices, and the Jain's index is computed in the inference phase using two performance metrics x , namely the average probability of detection (P_D), shown in Figs. 9 and 11, and the average probability of false alarm (P_{FA}) shown in Figs. 10 and 12. This allows us to examine how fairly each AP serves all devices in the network under different detection strategies. Given the number of APs M , plotting the Jain index curve for each individual AP would result in cluttered and hard-to-interpret figures. Therefore, for each method, we present the Jain index averaged over all APs and illustrate the dispersion of the M per-AP curves using a shaded band (envelope) around the mean value.¹

¹The Jain index for the AMP algorithm is excluded from the figures as its iterative estimation procedure inherently produces very high fairness values, treating all devices nearly equally. Including it would overshadow the differences among the other approaches.

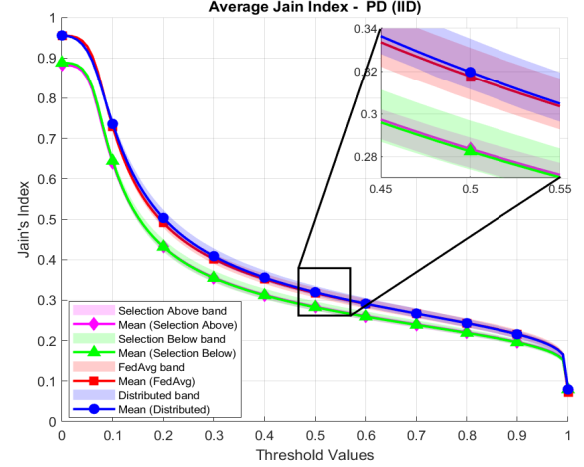


FIGURE 9. Average Jain's index across all clients for detection probability-IID scenario.

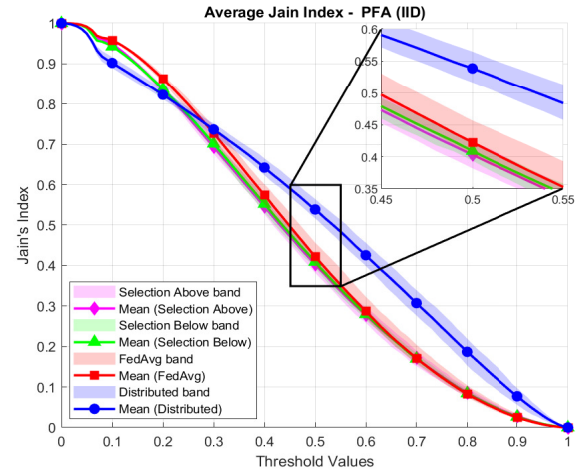


FIGURE 10. Average Jain's index across all clients for false alarm probability-IID scenario.

The curves in Figs. 9–12 illustrate Jain's index as a function of the decision thresholds value τ used in each algorithm. This representation establishes a direct link between the Jain's index and the ROC curves. The portion of the Jain's index curve at low threshold values (e.g., $\tau \ll 0.5$) corresponds approximately to the right-hand side of the ROC curves (i.e., to the right of the highlighted boxes in Figs. 6–7), where both the detection probability P_D and the false alarm probability P_{FA} are significant. Conversely, the portion of the Jain's index curve at high threshold values (τ close to 1) corresponds to the left-hand side of the ROC curves (i.e., to the left of the highlighted boxes in Figs. 6–7), where the detection rate becomes very low and the miss detection rate increases, with false alarms essentially disappearing when detections are absent.

The most interesting region of a Jain's index curve corresponds to the threshold that offers the best trade-off between P_D and P_{FA} , i.e., the closest point to the ideal

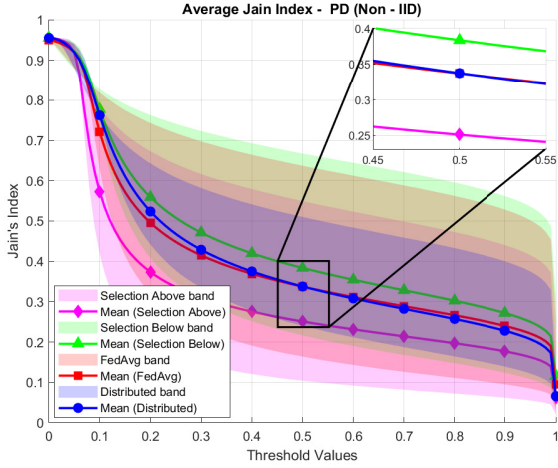


FIGURE 11. Average Jain's index across all clients for detection probability-Non-IID scenario.

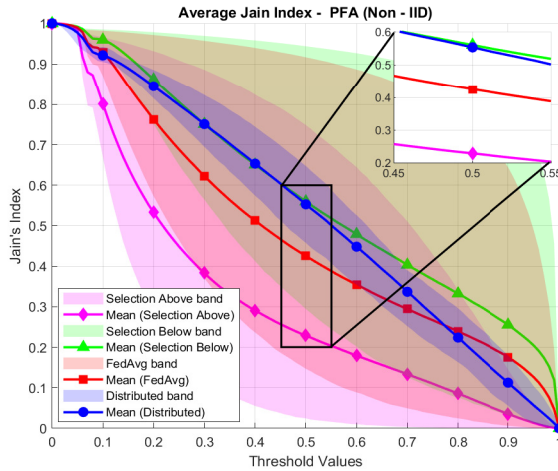


FIGURE 12. Average Jain's index across all clients for false alarm probability-Non-IID Scenario.

corner ($P_{FA} = 0$, $P_D = 1$) of the ROC curve. In this study, we focus on the areas enclosed within the boxes in both ROC and Jain's index figures as the main regions of interest. Nevertheless, this choice is not unique, as the optimal operating point ultimately depends on the desired balance between system performance and fairness across devices.

Figs. 9–10 present the average Jain index across the studied algorithms for P_D and P_{FA} , evaluated under the IID scenario. Fig. 9 shows that the average Jain index exhibits almost the same behavior for any method. Under IID conditions, the shaded band representing the spread of the M AP curves reflects low variance, indicating that the APs behave similarly in promoting equitable resource distribution among devices from the P_D perspective. These observations stem from the homogeneous nature of the IID setup, where all APs have similar views of the devices, given the uniformity of their datasets. Some variations between methods are still observed when comparing each method's average curve and

its band, particularly where “Distributed” and “FedAvg” achieve slightly better Jain index values compared to the proposed methods. In Fig. 10, a similar behavior is observed. Similarly to Fig. 9, the bands remain narrow, and the average Jain index curves of the different methods are close. Both proposed approaches achieve similar Jain index values, with similar fairness to “FedAvg” but lower than “Distributed”. This result is compensated by better P_D and P_{FA} results, as shown in Fig. 6.

Figs. 11–12 represent the Jain index in the non-IID scenario, from the P_D and P_{FA} perspectives, respectively. First, for any method, the shaded bands are significantly large, emphasizing a high variation of the M curves from the mean curve and unbalanced fairness across APs, which is directly linked to the data heterogeneity among them. In Fig. 11, “Selection Above” achieves the lowest fairness index, meaning that some APs tend to prioritize certain devices over others. This bias arises because selecting clients with higher received power typically favors APs located in denser areas in the non-IID setup. Consequently, the global model of “Selection Above” becomes biased toward improving P_D in dense regions while APs in non-dense regions neglect certain devices, thus resulting in a high variation in Jain's index values among APs. In contrast, “Selection Below” realizes the best fairness, as it encourages participation from APs with low received power, leading to an equitable global model. This result comes at the expense of a degraded detection performance, as illustrated in Fig. 7. Finally, “FedAvg” and “Distributed” realize performances between those of “Selection Above” and “Selection Below”, since the former involves all APs all the time, while the latter relies only on the individual contribution to the detection model. Similar results are observed in Fig. 12, where, from the P_{FA} perspective, “Selection Below” outperforms all other methods in terms of fairness, and the shaded bands are significantly larger compared to the results of the IID case.

It is worth noting that the Jain index metric in Figs. 9–12 is computed over devices during the testing phase. It quantifies how equitably the detection performance is distributed among devices. In the non-IID case, the lower Jain index of “Selection Above” confirms that the improvement in global detection performance is mainly driven by devices located in dense regions, where high-power APs provide more informative observations. From an AP-level perspective, the selection policy induces an imbalance during the training phase. Specifically, “Selection Above” tends to select high-power APs frequently, while “Selection Below” favors low-power APs. As illustrated by the received-power distribution in Fig. 3 and the per-AP accuracy dispersion in Fig. 8, APs under heterogeneous conditions do not contribute equally to local training. Hence, AP-level fairness is also related to participation frequency and resource consumption, since frequently selected APs perform more local training and transmit more model updates.

Deployment guidelines: The benefit of FL depends strongly on the adopted client-selection policy and on the considered network conditions. FL is particularly advantageous when the network data are naturally distributed across APs and when collecting all raw observations at a central processor is costly, impractical, or undesirable from a data-privacy perspective. In such cases, each AP can learn from its local observations and transmit only model updates to the server for aggregation, thus reducing raw-data sharing while enabling collaborative learning across the network. Without loss of generality, “Selection Above” is the most suitable method for AD when high-power APs contain more informative data features, i.e., devices are spatially concentrated in dense regions of the network. In contrast, “Selection Below” is more relevant when the objective is not only to maximize AD performance, but also improving fairness among devices. Hence, the selection between the two policies depends on the network conditions and on the desired objective. The selection threshold should also be adapted according to the average received-power distribution across APs, as defined in eq. (5). Nevertheless, FL should not be interpreted as an unconditional replacement for model-driven detectors. Algorithms such as AMP remain important model-driven baselines and can achieve competitive detection performance when the assumed signal model is well matched to the system conditions. However, their performance may degrade under more challenging practical settings, such as high device density and pilot-assignment constraints [36], as well as hardware impairments and memory/computational constraints [37], where iterative recovery procedures may become less stable or more costly. Consequently, the proposed FL framework should be viewed as a complementary distributed-learning approach that can exploit AP-side observations while reducing centralized raw-data collection and enabling power-aware client selection.

D. COMPUTATIONAL COMPLEXITY AND COMMUNICATION OVERHEAD

We analyze here the computational complexity in terms of multiply-accumulate (MAC) operations. During the inference phase, all FL-based methods, namely FedAvg, FedProx, “Selection Above”, and “Selection Below”, use the same final SLP architecture for the APs. Hence, the inference complexity per AP is identical. The used SLP model is based on [23], where a complexity analysis has been conducted, a comparison has been done showing both the execution time and complexity, highlighting the superiority of the SLP-based data-driven detector over the CS-based approach, i.e., AMP. The inference complexity per AP can be expressed as:

$$C_{\text{SLP}} = V(2NL + K), \quad (17)$$

where N is the number of antennas per AP, L is the pilot length, and K is the number of devices. For comparison, the

TABLE 3. Inference computational complexity per AP for the considered simulation setup.

Method	Numerical complexity	Relative complexity
SLP-based detector	133,120 MACs	$1\times$
AMP detector	1,728,000 MACs	$\approx 13\times$

AMP detector requires:

$$C_{\text{AMP}} = 12tLKN, \quad (18)$$

MAC operations, where t denotes the number of AMP iterations. This highlights that AMP complexity depends directly on the number of iterations required for convergence, whereas the SLP inference cost is fixed once the model is trained.

Using the considered parameters $N = 2$, $L = 40$, $K = 100$, $V = 512$, and $t = 18$, Table 3 reports the numerical inference complexity per AP. The SLP-based detector requires 133,120 MAC operations, whereas AMP requires 1,728,000 MAC operations. Therefore, AMP requires approximately 13 times more MAC operations than the SLP-based detector during inference. This confirms the lower inference complexity of the SLP-based data-driven detector compared with the CS-based AMP approach.

During training, the proposed client-selection mechanism introduces only a negligible additional cost. Each AP computes one scalar average received-power score per round, and the server applies a simple threshold comparison with the per-round mean score. Therefore, the additional computational burden of “Selection Above” and “Selection Below” compared with FedAvg is limited to scalar score computation and comparison.

For the communication overhead, for both selection strategies, since only the selected APs participate in local training and transmit model updates in each round, these proposed selection strategies can reduce the training communication overhead and local training cost compared with standard FedAvg, where all APs participate at every round.

E. IMPACT OF APS' DEACTIVATION

In practical CF-mMIMO deployments, network operators may deactivate or temporarily switch off certain APs during low-traffic periods to reduce energy consumption and operational costs. However, this adaptive management strategy can inadvertently affect network coverage and degrade system performance, particularly during inference. Hence, analyzing the impact of APs' deactivation provides valuable insights into the trade-off between energy efficiency and activity detection performance across different network conditions.

We assess here the impact of deactivating some APs on the ROC curves of the evaluated methods and the realized power savings, under the IID and non-IID settings. The total power consumption of the detection procedure spans the receive-to-decision chain and includes multiple components: baseband processing (e.g., the local oscillator and

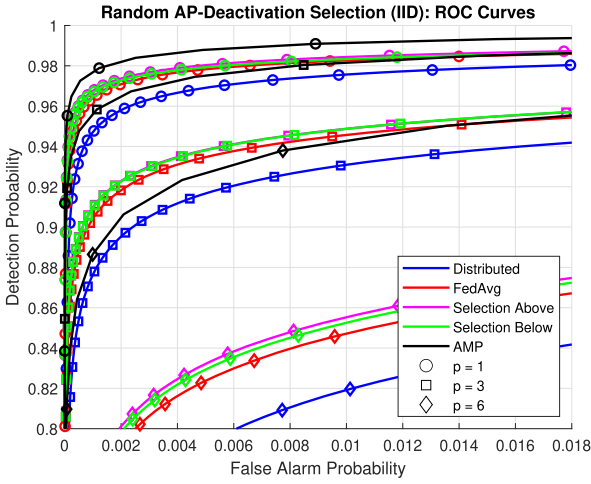


FIGURE 13. ROC under random AP deactivation-IID scenario.

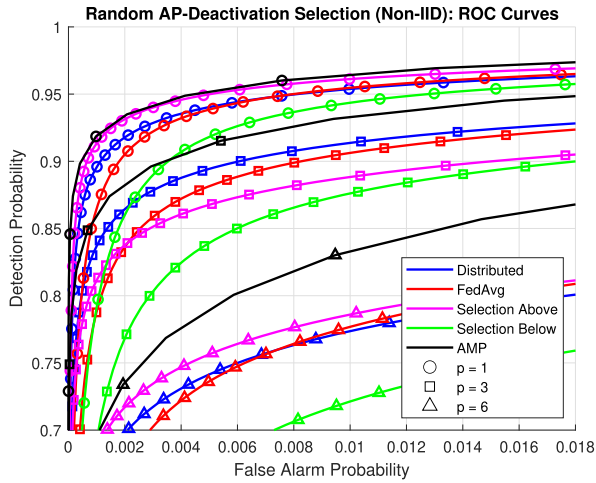


FIGURE 14. ROC under random AP deactivation-Non-IID scenario.

signal-detection operations whose complexity depends on the model distributed across the APs), RF-chain consumption, backhaul consumption due to exchanges between the central server (CPU) and the active APs, and the consumption at the central server itself associated with post-processing functions required to finalize the overall detection procedure. For simplicity, a representative power analysis on a percentage scale is provided, qualitatively inferred from the observed ROC behavior.

In Figs. 13–15, we report the detection performance under AP deactivation, where $p \in \{1, 3, 6\}$ APs are switched off in the network. Two switching-off policies are adopted, namely (i) random (applied in both IID and non-IID scenarios) and (ii) density-aware (applied only in the non-IID scenario).

For the random policy, p APs to deactivate are drawn uniformly at random from the M deployed ones, which yields $\binom{M}{p}$ as the total number of possible combinations. Due to the time complexity, for a specific p , the ROC curve for a given detection approach is obtained by averaging over

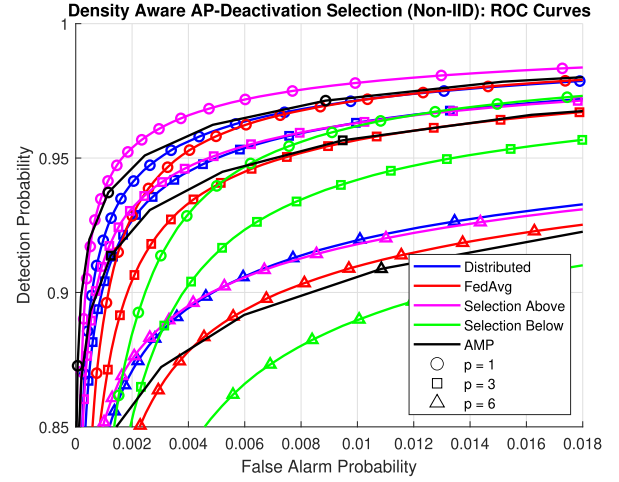


FIGURE 15. ROC under density-aware AP deactivation-Non-IID scenario.

$C < \binom{M}{p}$ combinations, i.e., $C = 20$ ($p \in \{3, 6\}$) and $C = M$ ($p = 1$).

For the density-aware policy, APs in the dense section are always kept active, while p APs to deactivate are selected from the remaining less-dense sections. This results in $\binom{3M/4}{p}$ possible combinations for each value p . In this setting, we set $C = 20$ ($p > 1$) and $C = 3M/4$ ($p = 1$). All experiments use the same models trained with no AP deactivation ($p = 0$), and no retraining is performed when some APs are switched off.

Fig. 13 presents the effect of randomly selecting and dropping p APs in the IID scenario. We notice that “AMP” is the most robust approach, outperforming all methods for any p . Indeed, “AMP” treats all devices equally, yielding to similarly distributed AP predictions. Hence, even if some APs are dropped, it ensures prediction fairness among all devices. For the proposed “Selection Above” and “Selection Below” methods, the ROC curves are slightly lower than that of “AMP”, with a loss between 1% ($p = 1$) and 10% ($p = 6$). Although each AP is equipped with the same global model, dropping APs modifies the device serving cluster. This change affects the performance since accurate device activity detection is bound to the serving cluster for each device. Finally, “FedAvg” and “Distributed” achieve lower performances when $p > 0$, showcasing their low robustness against AP deactivation.

In Table 4, we present the corresponding power consumption of APs under the event of APs’ deactivation, for “AMP” only.² For a target false-alarm rate $P_{FA} = 0.2\%$, the overall network power consumption is 100% when $p = 0$ (all APs active), and “AMP” achieves a detection rate of $P_D = 0.988$ (from Fig. 6). When p increases to 3 (resp. to 6), P_D drops to 0.964 (resp. 0.904), but power consumption

²We chose to show the results only for “AMP” as it is the most robust method to random AP deactivation, and similar results can be obtained for the other methods.

TABLE 4. Power consumption analysis (“AMP”, Target $P_{FA} = 0.2\%$, IID).

p	Power Consumption	Active APs	P_D	P_E
0	100%	12	0.988	0.988
3	75%	9	0.964	1.28
6	50%	6	0.904	1.80

TABLE 5. Power consumption analysis (“Selection Above”, Target $P_{FA} = 0.2\%$, Non-IID).

p	Power Consumption	Active APs	P_D	P_E	Deactivation Policy
0	100%	12	0.961	0.961	Random
3	75%	9	0.835	1.113	
6	50%	6	0.72	1.44	
0	100%	12	0.961	0.961	Density-aware
3	75%	9	0.93	1.24	
6	50%	6	0.875	1.75	

is also reduced by 25% (resp. 50%). Looking at the power efficiency, defined as $P_E = P_D/\text{Power Consumption}$ and referring to the amount of successful activity detection per unit of power consumed, we see that APs’ deactivation improves P_E to 1.28 ($p = 3$) and 1.8 ($p = 6$), respectively.

Fig. 14 plots the same performance as in Fig. 13, but for the non-IID setting. Similarly, “AMP” outperforms all other methods, and is the most robust one in front of APs’ deactivation. “Selection Below” achieves the worst results for any p . Indeed, its unsuitability to non-IID conditions and the randomness of AP deactivation significantly impact its performance. In contrast, “Selection Above” slightly resists APs’ deactivation, but still realizes worse results than “AMP”, “FedAvg”, and “Distributed”. Finally, “Distributed’s” performance depends on the sampled combination C , thus outperforming sometimes “Selection Above”, while “FedAvg” has a similar performance.

Clearly, random AP deactivation is unsuited for the non-IID scenario. To tackle this issue, we plot in Fig. 15 the ROC curves for the density-aware policy. With this policy, activity detection is enhanced for all methods and any p . “Selection Above” now demonstrates the best results, while “Selection Below” is still the least preferred one due to its inherent AP selection criterion. In between, “AMP”, “FedAvg” and “Distributed” achieve similar results with “Distributed” being favored for most p values. Indeed, the latter trains APs independently, which limits the impact of APs’ deactivation.

In Table 5, we present the corresponding power consumption of APs under the event of APs’ deactivation (random and density-aware), for “Selection Above” only.³ For a target false-alarm rate $P_{FA} = 0.2\%$ and $p = 0$, “Selection Above” achieves a detection rate of $P_D = 0.961$ (from Fig. 7). Using

³We prefer to present the results only for “Selection Above” as it is the most robust method to density-aware AP deactivation, and similar results can be obtained for the other methods.

the random deactivation policy, when p grows to 3 (resp. to 6), P_D drops drastically to 0.835 (resp. 0.72), but power consumption is also reduced by 25% (resp. 50%). From the power efficiency perspective, P_E enhances to 1.113 ($p = 3$) and 1.44 ($p = 6$), respectively. Under the density-aware deactivation policy, we notice that the power efficiency is better, reaching 1.24 ($p = 3$) and 1.75 ($p = 6$), respectively, which are close to those achieved by “AMP” in the IID setting.

AP Deactivation guidelines: The optimal AP deactivation ratio is not a fixed value, but depends on the desired trade-off between detection performance and power efficiency. In the IID scenario, deactivating $p = 3$ APs out of 12, corresponding to 25% of the APs, provides a good balance, since it reduces power consumption while maintaining limited performance degradation. Specifically, for a fixed QoS constraint given by $P_{FA} = 0.2\%$, all detectors still achieve a detection probability higher than 0.88. In contrast, deactivating 6 APs, corresponding to 50% of the APs, further improves power efficiency, but leads to a more noticeable degradation, with $P_D < 0.82$ for some detectors. In the non-IID scenario, the deactivation policy becomes more critical. In fact, random AP deactivation may severely degrade the detection performance, whereas density-aware deactivation better preserves the most informative APs. Hence, 25% AP deactivation can be considered a safe operating point, while 50% density-aware deactivation may be adopted when power saving is prioritized.

V. CONCLUSION

This work introduced a power-aware client selection strategy for federated activity detection in cell-free massive MIMO-based mMTC networks, addressing the dual objectives of performance and fairness. By leveraging the average received signal power at each AP as a selection criterion, we demonstrated that detection performance can be significantly improved, particularly under non-IID conditions, while also revealing the inherent trade-off between global detection accuracy and fairness across devices. Through extensive evaluations in both IID and non-IID scenarios, we showed that the proposed “Selection Above” method consistently outperforms benchmark strategies in terms of detection probability. However, its fairness degrades in highly heterogeneous networks. In contrast, the proposed “Selection Below” method enhances fairness but at the cost of reduced detection performance, underscoring the need to balance these competing objectives in practical deployments. Our findings highlight that optimizing for detection rate alone is insufficient in mMTC systems, where equitable service delivery is equally critical for robustness and inclusivity. Moreover, we explored the impact of AP deactivation policies on power efficiency and detection performance. Results indicate that density-aware deactivation offers the best compromise in non-IID settings, enabling substantial power savings with minimal performance loss. Among all strategies, “Selection Above” emerged as the most robust,

offering a favorable balance between accuracy, fairness, and energy efficiency. These insights pave the way for more adaptive and sustainable FL frameworks in CF-mMIMO wireless networks.

APPENDIX: LIST OF ACRONYMS

TABLE 6. List of acronyms.

3GPP	3rd generation partnership project
AD	Activity Detection
AMP	Approximate Message Passing
APs	Access Points
CF-mMIMO	Cell-Free Massive Multiple-Input Multiple-Output
CNN	Convolutional Neural Network
CS	Compressive Sensing
CPU	Central Processing Unit
CV	Covariance-Based
DMLP	Deep Multilayer Perceptron
PFA	Probability of False Alarm
FedAvg	Federated Averaging
FL	Federated Learning
FISTA	Fast Iterative Shrinkage-Thresholding Algorithm
GB-RA	Grant-Based Random Access
GF-RA	Grant-Free Random Access
IID	Independent and Identically Distributed
IoT	Internet of Things
IQR	Interquartile range
ISTA	Iterative Soft Thresholding Algorithm
LASSO	Least Absolute Shrinkage and Selection Operator
LSF	Large-Scale Fading
mMIMO	Massive Multiple-Input Multiple-Output
mMTC	Massive Machine-Type Communication
ML	Machine Learning
QoS	Quality of Service
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristics
SLP	Single-Layer Perceptron
SSF	Small-Scale Fading

REFERENCES

- [1] C. Bockelmann et al., "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
- [2] B. M. Lee and H. Yang, "Massive MIMO for industrial Internet of Things in cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2641–2652, Jun. 2018.
- [3] B. M. Lee, "Massive MIMO for massive industrial Internet of Things networks: Operation, performance, and challenges," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 6, pp. 2119–2135, Dec. 2024.
- [4] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.
- [5] J. Ding, D. Qu, and J. Choi, "Analysis of non-orthogonal sequences for grant-free RA with massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 150–160, Jan. 2020.
- [6] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 3rd Quart., 2020.
- [7] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.
- [8] S. Vidya and R. Gopalakrishnan, "Dynamic task offloading in edge computing for computer access point selection based on adaptive deep reinforcement learning with meta-heuristic optimization," *Appl. Soft Comput.*, vol. 176, May 2025, Art. no. 113105.
- [9] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.
- [10] T. T. Vu, D. T. Ngo, N. H. Tran, H. Q. Ngo, M. N. Dao, and R. H. Middleton, "Cell-free massive MIMO for wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6377–6392, Oct. 2020.
- [11] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 4, pp. 389–400, Jun. 2013.
- [12] F. Monsees, M. Woltering, C. Bockelmann, and A. Dekorsy, "Compressive sensing multi-user detection for multicarrier systems in sporadic machine type communication," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [13] Z. Gao, L. Dai, Z. Wang, S. Chen, and L. Hanzo, "Compressive-sensing-based multiuser detector for the large-scale SM-MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8725–8730, Oct. 2016.
- [14] Y. Du et al., "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.
- [15] D. L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, Jul. 2006.
- [16] Q. Zou and H. Yang, "A concise tutorial on approximate message passing," 2022, *arXiv:2201.07487*.
- [17] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.
- [18] U. K. Ganesan, E. Björnson, and E. G. Larsson, "Clustering-based activity detection algorithms for grant-free random access in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7520–7530, Nov. 2021.
- [19] U. K. Ganesan, E. Björnson, and E. G. Larsson, "An algorithm for grant-free random access in cell-free massive MIMO," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.
- [20] A. Elkeshawy, H. Farès, and A. Nafkha, "Robust learning-based sparse recovery for device activity detection in grant-free random access cell-free massive MIMO: Enhancing resilience to impairments," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2025, pp. 2150–2155.
- [21] J. H. I. de Souza and T. Abrão, "Deep learning-based activity detection for grant-free random access," *IEEE Syst. J.*, vol. 17, no. 1, pp. 940–951, Mar. 2023.
- [22] M. U. Khan, E. Testi, M. Chiani, and E. Paolini, "Blind user activity detection for grant-free random access in cell-free mMIMO networks," in *Proc. IEEE 8th Forum Res. Technol. Soc. Ind. Innov. (RTSI)*, Sep. 2024, pp. 432–436.
- [23] A. Elkeshawy, H. Farès, and A. Nafkha, "Lightweight learning for grant-free activity detection in cell-free massive MIMO networks," *IEEE Access*, vol. 13, pp. 182872–182889, 2025.
- [24] X. Wang, E. Ollila, and S. A. Vorobyov, "Robust activity detection for massive random access," *IEEE Trans. Signal Process.*, vol. 73, pp. 3513–3527, 2025.
- [25] X. Wang, E. Ollila, and S. A. Vorobyov, "Robust activity detection for massive access using covariance-based matching pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2025, pp. 1–5.
- [26] A. Mahmoudi, M. Xiao, and E. Björnson, "Accelerating energy-efficient federated learning in cell-free networks with adaptive quantization," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 3, pp. 761–778, 2025.
- [27] Y. Cheriguen, W. Jaafar, H. Yanikomeroglu, and C. A. Kerrache, "Towards reliable participation in UAV-enabled federated edge learning on non-IID data," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 125–141, 2024.
- [28] S. Kalabakov et al., "Federated learning for activity recognition: A system level perspective," *IEEE Access*, vol. 11, pp. 64442–64457, 2023.
- [29] A. Elkeshawy et al., "Federated learning for secure and efficient device activity detection in mMTC networks," in *Proc. EuCNC & 6G Summit*, Jun. 2025, pp. 1–2.

- [30] U. T. Demir, E. Bjernson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, Jan. 2021.
- [31] *3rd Generation Partnership Project, Study on Channel Model for Frequencies From 0.5 to 100 GHz, 3GPP*, document TR 38.901 V17.0.0 (2022-03), 3GPP, Valbonne, France, Mar. 2022, release 17.
- [32] F. Alalyan, M. Awad, W. Jaafar, and R. Langar, "Secure distributed federated learning for cyberattacks detection in B5G open radio access networks," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 3067–3081, 2025.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
- [34] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, I. Dhillon, D. Papailiopoulos, and V. Sze, Eds., Cambridge, MA, USA: Proceedings of Machine Learning Research (PMLR), 2020, pp. 429–450.
- [35] H. Wu et al., "Probabilistic automatic outlier detection for surface air quality measurements from the China national environmental monitoring network," *Adv. Atmos. Sci.*, vol. 35, no. 12, pp. 1522–1532, Dec. 2018.
- [36] A. Elkeshawy, H. Farès, and A. Nafkha, "Which is more suitable for pilot assignment: K-agnostic or K-dependent activity detection?" in *Proc. IEEE Veh. Technol. Conf. Fall*, Boston, MA, USA, Mar. 2026.
- [37] A. Elkeshawy, B. Tegin, H. Farès, and A. Nafkha, "Hardware-aware federated activity detection under analog in-memory computing non-idealities in CF-mMIMO," in *Proc. IEEE Veh. Technol. Conf. Fall*, Boston, MA, USA, Mar. 2026.



ALI ELKESHAWY (Student Member, IEEE) received the Diploma degree in telecommunication, communication, and computer engineering from the Faculty of Engineering III (ULFG), Lebanese University, Beirut, Lebanon, in 2021, and the M.Sc. degree in telecommunication, networks, and security from Saint Joseph University (USJ-ESIB), Mansouriyeh, Lebanon, in 2022. He is currently pursuing the Ph.D. degree with CentraleSupélec, Rennes, France.

He is a member of the SIGNAL Research Group, CNRS, Institute of Electronics and Digital Technologies (IETR). His research focuses on artificial intelligence applications in telecommunications, particularly physical layer techniques for massive machine-type communications and grant-free random access in cell-free massive MIMO systems. His expertise includes AI-driven wireless communications for IoT intrusion detection, natural language processing for speech recognition and machine translation, and embedded system deployment and practical network design.



WAEEL JAAFAR (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from Polytechnique Montréal, Montreal, QC, Canada.

He has been an Associate Professor with the Software and Information Technology Engineering Department, École de Technologie Supérieure (ÉTS), University of Quebec, Montreal, since September 2022. From 2019 to 2022, he was an NSERC Post-Doctoral Fellow with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada. From 2014 to 2018, he pursued a career in the telecommunications industry, where he was involved in the design of telecommunication solutions for projects in Canada and abroad. During his academic career, he held a Visiting Researcher position with Khalifa University, Abu Dhabi, United Arab Emirates, in 2019; Keio University, Tokyo, Japan, in 2013; and the University of Québec at Montréal (UQAM), Canada, in 2007. His research interests include wireless communications, integrated terrestrial and non-terrestrial networks, resource allocation, edge caching and computing, and machine learning for communication and networking. He has received several prestigious awards and grants, including the Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Scholarship; the Fonds de Recherche du Québec—Nature et Technologies (FRQNT) Scholarship; and the Best Paper Awards at CIoT 2024, IEEE ISCC 2023, and IEEE ICC 2021.



HAÏFA FARÈS (Member, IEEE) received the bachelor's and M.Sc. degrees in telecommunication engineering from the Higher School of Communications of Tunis (Sup'Com), in 2007 and 2008, respectively, and the Ph.D. degree in digital communications from IMT-Atlantique, France, in 2011.

She is currently an Associate Professor with CentraleSupélec, Rennes. She is a member of the SIGNAL Research Group, CNRS, Institute of Electronics and Digital Technologies (IETR), Rennes Laboratory. Her research interests include the area of communication theory, including nonlinear modulations, indexed modulations, energy-efficient communications, non-orthogonal multiple access, physical layer security, and integration of machine learning into future communication systems.



AMOR NAFKHA (Senior Member, IEEE) received the B.Sc. degree in information and communications technology engineering from the Higher School of Communications (Sup'Com), Tunis, Tunisia, in 2001, the Ph.D. degree in information and communications technology from the University of South Brittany (UBS), Lorient, France, in 2006, and the Habilitation à Diriger des Recherches degree from the University of Rennes 1 (UR1), Rennes, France, in November 2021.

From 2006 to 2007, he was a Post-Doctoral Researcher with the Signal, Communication, and Embedded Electronics (SCEE-IETR) Research Group, CentraleSupélec, France, where he has also been an Associate Professor since January 2008. He is promoted to a Full Professor with CentraleSupélec in 2024. He was actively involved in the reconfigurable hardware platform implementation for software-defined radio. He has co-authored several contributions on FPGA dynamic partial reconfiguration. He has published more than 90 papers in international peer-reviewed journals and conferences. His research interests include multi-user and MIMO detection, optical mimo multimode fiber, information theory, and spectrum sensing techniques. His current research interests include physical layer security, mmWave and sub-terahertz communications, cell-free massive MIMO, and securing embedded AI.